Technische Universität München

Department of Mathematics

Master's Thesis

Distorted Risk Measures and Backward Stochastic Differential Equations

Hannes Hagen Haferkorn

Supervisor: Prof. Dr. Claudia Klüppelberg

Advisor: Dr. Jianing Zhang

Submission Date: 28^{th} of November 2013

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

Garching, 28^{th} of November 2013

Zusammenfassung

Eine spezielle Klasse statischer kohärenter Risikomaße wird durch die sog. Choquet Integrale, die "statische Version" verzerrter Risikomaße dargestellt. Will man diese auf "naive" Weise zu dynamischen Risikomaßen verallgemeinern, so stellt man fest, dass das Resultat nicht Zeit-konsistent ist. Allerdings ist es möglich, ein dynamisches Risikomaß in diskreter Zeit durch Aneinanderreihung von Choquet Integralen auf den Subintervallen rekursiv zu erzeugen – das sog. verzerrte Risikomaß. Man kann zeigen, dass dieses Risikomaß in einem immer feiner werdenden Bernoulli Random Walk Approximationsschema (für die Brownsche Bewegung) nach entsprechender Reskalierung gegen die Lösung einer BSDE konvergiert. Ein numerischer BSDE-Lösungsalgorithmus wurde implementiert und in einem Beispiel angewandt. Im eindimensionalen Fall kann sogar noch mehr gezeigt werden, nämlich, dass die Lösung der BSDE für pfadweise steigende Positionen sich zu einer klassischen bedingten Erwartung unter einem äquivalenten Wahrscheinlichkeitsmaß reduzieren lässt. Ein großer Nachteil des Ansatzes ist, dass die im Grenzwert auftauchende BSDE vom angewendeten Approximationsschema abhängt. Deshalb wird am Ende der Arbeit ein anderer Ansatz, der ohne Approximation der Brownschen Bewegung auskommt, vorgestellt.

Abstract

A special case of static coherent risk measures is given by Choquet integrals, the "static version" of distorted risk measures. If one tries to generalize them to dynamic risk measures in a "naive" way, one faces the problem that the result will in general not be time-consistent. However, it is possible to construct a time-consistent dynamic risk measure – the so called distorted risk measure – in discrete time recursively, which corresponds to the Choquet integrals on the time increments. It can be shown that after rescaling this risk measure converges under a Bernoulli random walk approximation scheme to the solution of a backward SDE (BSDE) as the number of time increments tends to infinity and the maximum size of the time increments tends to zero. A numerical solver for BSDEs is used to calculate the limiting dynamic risk measure in an example. In one dimension, for pathwise increasing claims even more will be shown, namely that the limiting dynamic risk measure reduces to a classical conditional expectation under a particular equivalent probability measure. One big disadvantage of this approach is that the limiting BSDE depends on the approximation scheme for Brownian motion that was chosen. For that reason, a different approach that does not need an approximation for Brownian motion is proposed in the conclusion.

Acknowledgement

I want to thank Prof. Klüppelberg for giving me the opportunity to write my Master's Thesis under her supervision and for providing me with this topic that perfectly fits my research interests.

Moreover I express my gratitude to Dr. Jianing Zhang for his excellent advice and guidance. Especially the inspiring discussions we had always turned out to be an initial point for new considerations and perceptions that really boosted my knowledge.

Contents

1 Introduction

When a bank (or another regulated financial institution) takes a financial position, it has to quantify the corresponding risk somehow and it has to hold an appropriate capital reserve. This is achieved by so called risk measures. Since a financial position often has a longer time to maturity and market conditions can change significantly within this time, it is not enough to measure its risk only at the beginning, when the contract is settled. It should rather be tracked over the whole period until maturity (updating the information continuously and using at any point in time only the information available at this time). That leads to an adapted stochastic process, a so called dynamic risk measure. Of course it is of interest for the bank to study such dynamic risk measures carefully. For example the bank (financial institution) or the regulator could have defined a boundary which should not be exceeded at any time within the term. If one knows enough properties of the dynamic risk measure it might be possible to compute the probability of exceeding the boundary (analytically or by simulation). It turns out that the generalization of risk measures to dynamic risk measures is not trivial, since in the dynamic setting, more conditions have to be fulfilled to get something meaningful.

We will study a particular class of dynamic risk measures, the so called distorted risk measures. First we consider risk measures ρ in the static case, that can be written as a so called Choquet integral or Choquet expectation, i.e. $\rho(X) = \mathcal{C}^{\Psi}[-X]$ where

$$\mathcal{C}^{\Psi}[X] = \int_{-\infty}^{0} (\Psi(P(X > x)) - 1)\mathrm{d}x + \int_{0}^{\infty} \Psi(P(X > x))\mathrm{d}x.$$

Furthermore we require Ψ to be a so called distortion. Note that in the case $\Psi = \mathrm{id}$ this is nothing else than the usual expectation, so the Choquet expectation can be interpreted as a usual expectation with probabilities that are transformed by Ψ. In practice, the distortion Ψ will be chosen to be concave which means that for any fixed level the probability for a loss to exceed this level is increased which in the end means that risk will be evaluated higher than just with negative usual expectation. I.e. the distortions lead to an additional safety buffer. Furthermore, Choquet integrals lead to coherent risk measures when Ψ is a concave distortion. As a consequence these risk measures have a supremum representation

$$\mathcal{C}^{\Psi}[-X] = \sup_{Q \in \mathcal{D}^{\Psi}} E_Q[-X],$$

where $\mathcal{D}^{\Psi}$ is a set of measures depending on Ψ. The generalization of these risk measures to the dynamic case will be called distorted risk measures.

Dynamic risk measures are closely related to solutions of backwards stochastic differential equations (BSDEs). For distorted risk measures under a Bernoulli random walk approximation scheme we will prove some interesting properties of this BSDE. Furthermore examples of how to use distorted risk measures are given and a numerical solution-scheme for BSDEs is implemented to simulate the dynamical $AVaR$ in these examples.

The content of this thesis is organized as follows: In Section 2, the static case is considered. First the definition of a risk measure in the static setting is recalled, then it is defined what is meant by a distortion and a Choquet integral and a representation and coherence theorem is stated. All the definitions and properties are presented together with some examples of risk measures that are useful in practice. Section 3 is about the dynamic case. It is shown that the canonical way of generalizing the supremum representation to the dynamic case leads to time-inconsistency and is therefore not successful. Instead another approach is introduced, namely to devide the time-interval which is of interest into sub-intervals and to establish static risk measures on these sub-intervals. Then the static risk measures are concatenated recursively in a time-consistent way which yields a dynamic risk measure (that is indeed time-consistent), which is called distorted risk measure. This can only be done in discrete time, therefore it is an interesting question what happens if the number of sub-intervals is increased to infinity while the mesh tends to zero. It can be shown that under a Bernoulli random walk approximation, these distorted risk measures in discrete time converge to a dynamic risk measure in continuous time, when some proper rescaling is done. This continuous-time risk measure is strongly related to the solution of a particular BSDE, depending on the distortion Ψ. Section 4 contains a larger example about evaluating the risk of a hedging cost process in an incomplete market (under cross hedging, to be precise) when a variance minimizing hedging strategy is applied. A numerical scheme is implemented to solve the two-dimensional BSDE which appears in this example. In Section 5 the 1-dimensional case is studied and for an arbitrary but fixed distortion Ψ, the corresponding BSDE is given in explicit form. Furthermore it is shown that for a pathwise increasing claim which is Malliavin and Fréchet differentiable, the dynamic risk measure simplifies to a usual conditional expectation under some equivalent measure $Q^{\#}$ that depends on the distortion Ψ. Finally, in Section 6 a conclusion is given, where also the problem that the limiting BSDE depends on the chosen approximation scheme for the Brownian motion is discussed and an idea of improvement is given.

2 Static Case

In this chapter we consider the static case. First we give a short introduction to risk measures. Then we introduce the concept of a distortion and take a look on the "static version" of the distorted risk measures that are of interest for us. Finally we state the coherence theorem that deals with the supremum representation which was already mentioned in the introduction.

2.1 Basic Definitions

First we define what we mean by a risk measure in the static setting:

Definition 2.1 (Static Risk Measures).
$(\Omega, \mathcal{F}, P)$ probability space, $\mathcal{H}$ the space of random variables $X : \Omega \to \mathbb{R}$, representing the possible financial positions.

1. *A **monetary risk measure** is a function $\rho : \mathcal{H} \to \mathbb{R}$ such that (s.t.):*

   $$\forall X, Y \in \mathcal{H} : X \leq Y \Rightarrow \rho(X) \geq \rho(Y) \qquad \textit{(Monotonicity)}$$
   $$\forall X \in \mathcal{H}, m \in \mathbb{R} : \rho(X + m) = \rho(X) - m \qquad \textit{(Cash-Invariance)}$$

2. *A **convex risk measure** is a monetary risk measure ρ that fulfills for all $X, Y \in \mathcal{H}$ and all $\lambda \in [0, 1]$:*

   $$\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y) \qquad \textit{(Convexity)}$$

3. *If a convex risk measure ρ satisfies for all $X, Y \in \mathcal{H}$ and all $\lambda \geq 0$:*

   $$\rho(\lambda X) = \lambda\rho(X) \qquad \textit{(Positive Homogeneity)}$$

   *then it is called a **coherent risk measure**.*

For a better understanding, some examples might be helpful:

Example 2.1 (Some Risk Measures).

1. *The negative expectation $\rho(X) := E[-X]$ is a coherent risk measure.*
2. *The Value at Risk at level p, VaR_p, which is defined as:*

   $$\rho(X) = VaR_p(X) := q^-_{-X}(1 - p),$$

   where $q^-_X(p) := \inf\{x \in \mathbb{R} : P(X \leq x) \geq p\}$, is unfortunately not coherent. For $X \sim \mathcal{N}(\mu, \sigma^2)$, it takes the value $VaR_p(X) = -\mu + \sigma\Phi^{-1}(1 - p)$.

3. *The Average Value at Risk (or Conditional Value at Risk) at level λ*

   $$\rho(X) = AVaR_\lambda(X) := \frac{1}{\lambda}\int_0^\lambda VaR_p(X)dp,$$

   is a coherent risk measure. For $X \sim \mathcal{N}(\mu, \sigma^2)$, it takes the value $AVaR_\lambda(X) = -\mu + \frac{\sigma}{\lambda}\varphi(\Phi^{-1}(1 - \lambda))$.

2.2 Static Version of Distorted Risk Measures: Choquet Integrals

Since this work deals with distorted risk measure, we first have to define what is meant by a "distortion".

Definition 2.2 (Distortion)**.**
A ***probability distortion*** *(or just distortion) is a continuous increasing surjective function* $D : [0, 1] \to [0, 1]$*. To a given probability distortion* D *define the corresponding probability distortion* $\hat{D}$ *as* $\hat{D} := 1 - D \circ (1 - id)$*. Since requiring* D *to be a distortion is sometimes to strict, we define an increasing function* $D : [0, 1] \to [0, 1]$*, with* $D(0) = 0$*,* $D(1) = 1$ *to be a* ***generalized (probability) distortion****.*

Now let $(\Omega, \mathcal{F}, P)$ be a probability space. We want to define the so called Choquet integral of a position $X : \Omega \to \mathbb{R}$, therefore we first have to generalize the concept of a measure to a set function called "capacity":

Definition 2.3 (Capacity)**.**
A ***measure capacity*** *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is a monotone set function* $c : \mathcal{B}(\mathbb{R}_+) \to \mathbb{R}_+$ *with* $c(\emptyset) = 0$*,* $c(\mathbb{R}_+) = 1$ *and* $\forall A, B \in \mathcal{B}(\mathbb{R}_+)$ *with* $A \subset B : c(A) \leq c(B)$*.*

Example 2.2. *Let* D *be a (generalized) distortion and* P *be a probability measure on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*. Then* $D \circ P$ *given by*

$$(D \circ P)(A) := D(P(A)),\ \forall A \in \mathcal{B}(\mathbb{R}_+)$$

is a capacity. Pay attention to the fact that $D \circ P$ *does not denote the image measure of* P *under* D*,* P_D*.*

This enables us now to define the Choquet integral which will directly lead to a static version of a distorted risk measure.

Definition 2.4 (Choquet integral)**.**
Let X *be a random variable on the probability space* $(\Omega, \mathcal{F}, P)$ *and let* Ψ *be a generalized distortion. Then the* ***Choquet integral*** *of* X *under* $\Psi \circ P$*,* $\mathcal{C}^{\Psi}[X]$ *is defined as:*

$$\begin{aligned} \mathcal{C}^{\Psi}[X] &:= -\int_0^{\infty} (\hat{\Psi} \circ P)(X \leq x)\,dx + \int_0^{\infty} (\Psi \circ P)(X > x)\,dx \\ &= \int_{-\infty}^{0} (\Psi(P(X > x)) - 1)\,dx + \int_0^{\infty} \Psi(P(X > x))\,dx. \end{aligned} \tag{1}$$

Remark 2.1. *This is a generalization of the expectation of* X*, which can always be written as*

$$E[X] = \int_{-\infty}^{0} (P(X > x) - 1)\,dx + \int_0^{\infty} P(X > x)\,dx.$$

Therefore the Choquet integral is also called Choquet expectation.

Now we consider VaR and $AVaR$ and show that they can be expressed as a Choquet integral.

Example 2.3 (VaR is a Choquet integral)**.**
The Value at Risk at level p, VaR_p, can be expressed as

$$VaR_p(X) = \mathcal{C}^{\Psi}[-X],$$

*with the **generalized distortion*** $\Psi(x) = \mathbb{1}_{\{x>p\}}$.

This can be seen as follows: Let $\Psi(x) = \mathbb{1}_{\{x>p\}}$. Then

$$\begin{aligned}
\mathcal{C}^{\Psi}[-X] &= \int_{-\infty}^{0} \underbrace{(\mathbb{1}_{\{P(-X>x)>p\}} - 1)}_{=-\mathbb{1}_{\{1-F_{-X}(x)\leq p\}}}\mathrm{d}x + \int_{0}^{\infty} \underbrace{\mathbb{1}_{\{P(-X>x)>p\}}}_{\mathbb{1}_{\{1-F_{-X}(x)>p\}}}\mathrm{d}x \\
&= \int_{-\infty}^{0} \underbrace{-\mathbb{1}_{\{F_{-X}(x)\geq 1-p\}}}_{=-\mathbb{1}_{\{x\geq \inf\{y:F_{-X}(y)\geq 1-p\}\}}} \mathrm{d}x + \int_{0}^{\infty} \underbrace{\mathbb{1}_{\{F_{-X}(x)<1-p\}}}_{=\mathbb{1}_{\{x<\inf\{y:F_{-X}(y)\geq 1-p\}\}}} \mathrm{d}x \\
&= \int_{\min\{0,q^-_{-X}(1-p)\}}^{0} -1\mathrm{d}x + \int_{0}^{\max\{0,q^-_{-X}(1-p)\}} 1\mathrm{d}x \\
&= -(0 - \min\{0, q^-_{-X}(1-p)\}) + (\max\{0, q^-_{-X}(1-p)\} - 0) \\
&= q^-_{-X}(1-p) = VaR_p(X)
\end{aligned}$$

□

Example 2.4 ($AVaR$ is a Choquet integral)**.**
The Average Value at Risk at level λ, can be expressed as

$$AVaR_\lambda(X) = \mathcal{C}^{\Psi}[-X],$$

*with the **distortion*** $\Psi(x) = \min\{1, \frac{x}{\lambda}\}$.

From the representation of VaR as a Choquet integral and the definition of $AVaR$, we get:

$$AVaR_\lambda(X) = \frac{1}{\lambda}\int_0^\lambda \left(\int_{-\infty}^{0} (\mathbb{1}_{\{P(-X>x)>p\}} - 1)\mathrm{d}x + \int_0^\infty \mathbb{1}_{\{P(-X>x)>p\}}\mathrm{d}x\right)\mathrm{d}p$$

By Fubini's theorem, we have for the right double integral:

$$\begin{aligned}
\int_0^\lambda \int_0^\infty \mathbb{1}_{\{P(-X>x)>p\}}\mathrm{d}x\mathrm{d}p &= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_{[0,\lambda]}(p)\mathbb{1}_{[0,P(-X>x))}(p)\mathbb{1}_{[0,\infty)}(x)\mathrm{d}x\mathrm{d}p \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_{[0,\lambda]}(p)\mathbb{1}_{[0,P(-X>x))}(p)\mathbb{1}_{[0,\infty)}(x)\mathrm{d}p\mathrm{d}x \\
&= \int_0^\infty \int_0^{\min\{\lambda,P(-X>x)\}} 1\mathrm{d}p\mathrm{d}x \\
&= \int_0^\infty \min\{\lambda, P(-X>x)\}\mathrm{d}x
\end{aligned}$$

And for the left double integral:

$$
\begin{aligned}
&\int_0^\lambda \int_{-\infty}^0 (\mathbb{1}_{\{P(-X>x)>p\}} - 1)\mathrm{d}x\mathrm{d}p \\
&\quad = \int_0^\lambda \int_{-\infty}^0 -\mathbb{1}_{\{P(-X>x)\leq p\}}\mathrm{d}x\mathrm{d}p \\
&\quad = \int_{\mathbb{R}} \int_{\mathbb{R}} -\mathbb{1}_{[0,\lambda]}(p)\mathbb{1}_{[P(-X>x),\infty)}(p)\mathbb{1}_{(-\infty,0)}(x)\mathrm{d}x\mathrm{d}p \\
&\quad = \int_{-\infty}^0 \int_{P(-X>x)}^{\max\{\lambda, P(-X>x)\}} -1\mathrm{d}p\mathrm{d}x \\
&\quad = \int_{-\infty}^0 \underbrace{-\max\{\lambda, P(-X>x)\} + P(-X>x)}_{=\min\{\lambda, P(-X>x)\}-\lambda} \mathrm{d}p\mathrm{d}x.
\end{aligned}
$$

Combining both integrals and multiplying with the factor $1/\lambda$ yields

$$
AVaR_\lambda(X) = \int_{-\infty}^0 \min\left\{1, \frac{P(-X>x)}{\lambda}\right\} - 1\mathrm{d}x + \int_0^\infty \min\left\{1, \frac{P(-X>x)}{\lambda}\right\}\mathrm{d}x
$$

which completes the proof.

□

Now we are able to formulate a theorem that gives us a representation of risk measures induced by Choquet integrals which will be very useful when trying to generalize this concept to dynamic cases. Actually the theorem consists of two parts – one representation part which was shown in Madan et al. (2013), [1], Proposition 1, Remark 1 and one coherence part which was shown for all risk measures with this representation, compare Gianin (2006) [3], Theorem 3.

Theorem 2.1 (Coherence of Choquet integrals)**.**
Let Ψ be a concave probability distortion, $X \in L^2(P)$ and suppose that

$$
\mathcal{D}^\Psi := \{Q \in \mathcal{P}_{2,P}^{ac} : Q(A) \leq \Psi(P(A)) \forall A \in \mathcal{F}\} \neq \emptyset.
$$

Then $\rho(X) := \mathcal{C}^\Psi[-X]$ defines a ***coherent risk measure which has the representation***

$$
\mathcal{C}^\Psi[-X] = \sup_{Q \in \mathcal{D}^\Psi} E_Q[-X]. \tag{2}
$$

Remark 2.2. *It is not enough to require Ψ to be only a* ***generalized*** *probability distortion as the example of VaR shows, which can be written as a Choquet integral (as shown before) but which is not coherent.*

Remark 2.3. *Since* ***concavity*** *of the distortion is needed as well, we will from now on mean a* ***concave probability distortion****, whenever we speak about a* ***distortion****. As already justified in the introduction this is not really a constraint when dealing with risk measures.*

3 Dynamic Case

In this chapter we study the dynamic case. First we define dynamic risk measures, afterwards we try to generalize the concept of Choquet expectations making use of the supremum representation. We will see that the canonical idea of just computing at any time point t the static risk measure making use of all the information up to t does not work. Instead we will discuss another concept that bases upon the concatenation of one-period static valuations and point out its relation to BSDEs.

3.1 Basic Definitions

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ be a filtrated probability space with $\mathcal{F}_T = \mathcal{F}$. First of all let us define what we mean by a dynamic risk measure:

Definition 3.1 (Dynamic Risk Measure)**.**

1. *$(\rho_t)_{t\in[0,T]}$, $\rho_t : L^2(\mathcal{F}_T) \to L^2(\mathcal{F}_t)$ is called **dynamic risk measure**, if it fulfills for all $X, Y \in L^2(\mathcal{F}_T)$, $m \in L^\infty(\mathcal{F}_t)$, $A \in \mathcal{F}_t$ and $0 \le s \le t \le T$:*

   $X \le Y \Rightarrow \rho_t(X) \ge \rho_t(Y)$ *(Monotonicity)*

   $\rho_t(X + m) = \rho_t(X) - m$ *($\mathcal{F}_t$-Cash-Invariance)*

   $\rho_t(\mathbb{1}_A X + \mathbb{1}_{A^c} Y) = \mathbb{1}_A \rho_t(X) + \mathbb{1}_{A^c}\rho_t(Y)$ *($\mathcal{F}_t$-Local Property)*

   $\rho_t(X) \le \rho_t(Y) \Rightarrow \rho_s(X) \le \rho_s(Y)$ *(Time-Consistency)*

2. *A dynamic risk measure is **normalized** if*

   $\rho_t(0) = 0$ *(Normalization)*

3. *A dynamic risk measure is $\mathcal{F}_t$**-convex** if $\forall X, Y; \lambda \in L^\infty(\mathcal{F}_t) \cap [0, 1]$:*

   $\rho_t(\lambda X + (1 - \lambda)Y) \le \lambda\rho_t(X) + (1 - \lambda)\rho_t(Y)$ *($\mathcal{F}_t$-Convexity)*

Remark 3.1. *For a normalized dynamic risk measure, time-consitency is equivalent to the so called **dynamic programming principle** or **tower property** (compare [6] Definition 2.2 and following):*

$\forall X \in L^2(\mathcal{F}_T), 0 \le s \le t \le T : \rho_s(X) = \rho_s(-\rho_t(X))$ *(Tower Property).*

Since we will now consider families of random variables and want to take the "sup" we have to generalize this concept since we can only compare random variables up to almost sure uniqueness. Therefore we define an "essential supremum" (compare, for example [4], Definition A.1 or [5], Theorem A.32 and Definition A.33):

Definition 3.2. *(The essential supremum of a family of random variables)*
Let $(X_i)_{i\in I}$ be a family of random variables with some index set I. The essential supremum of this family of random variables is defined as a random variable X^* which fulfills

1. *$\forall i \in I : X_i \le X^*$ a.s.*

2. if X^{**} is a r.v. with $\forall i \in I : X_i \leq X^{**}$ a.s., then $X^* \leq X^{**}$ a.s..

We write $\text{esssup}_{i\in I} X_i := X^*$. The essential infimum essinf is defined in an analogous manner.

Remark 3.2.

1. *This definition is not a generalization of the "usual" definition of an essential supremum, which would be as follows: Let $(S, \mathcal{A}, \mu)$ be a measure space and $f : S \to [-\infty, \infty]$ a measurable function. Then*

$$\text{esssup}_{x\in S} f(x) := \inf\{y \in [-\infty, \infty] : \mu(f > y) = 0\}.$$

2. *In the following we will take essential suprema of the type "$\text{esssup}_{Q\in\mathcal{D}} E_Q[...|\mathcal{F}_t]$". Here the conditional expectations are the random variables indexed by Q and $\mathcal{D}$ is the index set. With the usual definition of essential suprema one would need a measure on the set of measures $\mathcal{D}$.*

3. *The esssup is not the same as the* sup*-almost-surely. Consider $I = \Omega = [0, 1]$, $X_i(\omega) = \mathbb{1}_{\{\omega=i\}}$, $P = \lambda$ (Lebesgue measure):*

$$\sup X_i = 1 \text{ a.s. but } \text{esssup}\, X_i = 0 \text{ a.s.}$$

4. *All essential supremas and infimas are defined with respect to the measure P, which can be understood either as the "physical/real-world measure" or just as some "reasonable" reference measure.*

3.2 Construction of Dynamic Version: First Attempt

Now we want to generalize the concept of those risk measures that can be represented as a Choquet integral which we just studied in the chapter before to the dynamic case. Recall the supremum representation from Theorem 2.1 for Choquet integrals under distortions in the static case:

$$\rho(X) = \sup_{Q\in\mathcal{D}^\Psi} E_Q[-X], \text{ where } \mathcal{D}^\Psi := \{Q \in \mathcal{P}^{ac}_{2,P} : Q(A) \leq \Psi(P(A))\forall A \in \mathcal{F}\}$$

A natural generalization of this concept to the dynamic case would now be

$$\mathcal{C}^\Psi[-X|\mathcal{F}_t] := \text{esssup}_{Q\in\mathcal{D}^\Psi_t} E_Q[-X|\mathcal{F}_t], t \leq T, \quad (3)$$

where

$$\mathcal{D}^\Psi_t := \{Q \in \mathcal{P}^{ac}_{2,P} : Q(A|\mathcal{F}_t) \leq \Psi(P(A|\mathcal{F}_t))\forall A \in \mathcal{F}\}. \quad (4)$$

The interpretation of $\mathcal{C}^\Psi[-X|\mathcal{F}_t]$ is the risk one would attribute to X at time t, given the information one has up to then (of course at time 0, when the position is written, this is a random variable, depending on events in future). The problem with this definition is that it leads to a dynamic risk measurement which is in general **not time-consistent** (and which is therefore not even a dynamic risk measure by our definition).

Counterexample 3.1. *An example where time-consistency fails is the Average (or Conditional) Value at Risk, i.e.* $\Psi(x) = \min\{1, \frac{x}{\lambda}\}$.

To show this, we first have to establish some lemmas.

Lemma 3.1. *Let* $t < T$ *and define* $\tilde{\mathcal{D}}_t^\Psi := \{Q \in \mathcal{D}_t^\Psi : Q|\mathcal{F}_t = P|\mathcal{F}_t\}$. *Then it holds that*

$$\textit{esssup}_{Q\in\mathcal{D}_t^\Psi} E_Q[-X|\mathcal{F}_t] = \textit{esssup}_{Q\in\tilde{\mathcal{D}}_t^\Psi} E_Q[-X|\mathcal{F}_t]$$

Proof: Let $Q \in \mathcal{D}_t^\Psi$. Then Q has a Radon-Nikodym density $\frac{dQ}{dP} =: Z(T)$ w.r.t. P. Let $Z(t) = E[Z(T)|\mathcal{F}_t]$. Then $(Z(t))_{t\in[0,T]}$ is a nonnegative martingale with expectation 1 and it is easy to show that the following Bayes rule holds on $\{Z(t) \neq 0\}$:

$$E_Q[-X|\mathcal{F}_t] = \frac{E_P[-Z(T)X|\mathcal{F}_t]}{Z(t)}.$$

On the other hand, we know that $\{Z(T) = 0\}$ is a Q-null-set (since $Q(Z(T) = 0) = E_P[Z(T)\mathbb{1}_{\{Z(T)=0\}}] = 0$) and $\{Z(t) = 0\} \subseteq \{Z(T) = 0\}$ P-a.s. (where we define $A \subset B$ P-a.s. $:\Leftrightarrow P(A \cap B^c) = 0$) since

$$E_P[Z(T)\mathbb{1}_{\{Z(t)=0\}}] = E_P[E_P[Z(T)\mathbb{1}_{\{Z(t)=0\}}|\mathcal{F}_t]] = E_P[Z(t)\mathbb{1}_{\{Z(t)=0\}}] = 0,$$

i.e. $Z(T) = 0$ P-a.s. on $\{Z(t) = 0\}$ if $P(Z(t) = 0) > 0$, because $Z(T) \geq 0$. And if $P(Z(t) = 0) = 0$, then automatically $\{Z(t) = 0\} \subseteq \{Z(T) = 0\}$ P-a.s., because then $P(\{Z(t) = 0\} \cap \{Z(T) \neq 0\}) = 0$. Since $E_Q[-X|\mathcal{F}_t]$ is only defined Q-a.s. we are free to define what should happen on $\{Z(t) = 0\}$. Note that there exists a nonnegative $\mathcal{F}_T$-measurable random variable ξ with $E_P[\xi|\mathcal{F}_t] = 1$ such that $Z(T) = Z(t)\xi$. It must be of the form

$$\xi = \frac{Z(T)}{Z(t)}\mathbb{1}_{\{Z(t)>0\}} + b\mathbb{1}_{\{Z(t)=0\}}$$

where b is a random variable such that $E_P[b|\mathcal{F}_t] = 1$. The easiest choice would be $b \equiv 1$. Now define $\tilde{Q}$ via $\frac{d\tilde{Q}}{dP} := \xi$. Then of course $\tilde{Q} \in \tilde{\mathcal{D}}_t^\Psi$ (since $\frac{d(\tilde{Q}|\mathcal{F}_t)}{d(P|\mathcal{F}_t)} = \frac{d\tilde{Q}}{dP}|\mathcal{F}_t = E_P[\xi|\mathcal{F}_t] = 1$, i.e. $Q|\mathcal{F}_t = P|\mathcal{F}_t$). And on $\{Z(t) > 0\}$

$$\begin{aligned} E_Q[-X|\mathcal{F}_t] &= \frac{E_P[-Z(T)X|\mathcal{F}_t]}{Z(t)} \\ &= E_P[-\xi X|\mathcal{F}_t] \\ &= \frac{E_P[-\xi X|\mathcal{F}_t]}{1} \\ &= E_{\tilde{Q}}[-X|\mathcal{F}_t]. \end{aligned}$$

The set $\{Z(t) = 0\}$ is a Q-null-set as already shown and therefore we just *define* $E_Q[-X|\mathcal{F}_t] := E_{\tilde{Q}}[-X|\mathcal{F}_t] := E_P[-\xi X|\mathcal{F}_t]$.

□

Especially the last definition is of interest, since this conditional expectation is defined P-a.s. and we always use P-esssup's. For this reason we will from now on always mean this definition when writing $E_Q[...]$.

In the following we consider $AVaR$ as a special case of a risk measure coming from a distorted expectation and show that its generalization in the way given above is not time consistent. That means that from now on, for the rest of this section, we consider Ψ given by $\Psi(x) = \min\{1, \frac{x}{\lambda}\}$.

Lemma 3.2. *Define* $\tilde{\mathcal{Q}}_t := \{Q \in \mathcal{P}^{ac}_{2,P} : Q|\mathcal{F}_t = P|\mathcal{F}_t, \frac{dQ}{dP} \leq \frac{1}{\lambda}\}$. *Then*

$$\tilde{\mathcal{D}}^{\Psi}_t = \left\{ Q \in \mathcal{P}^{ac}_{2,P} : Q|\mathcal{F}_t = P|\mathcal{F}_t, Q(A|\mathcal{F}_t) \leq \min\left\{1, \frac{P(A|\mathcal{F}_t)}{\lambda}\right\}\right\} = \tilde{\mathcal{Q}}_t.$$

Proof: Let $Q \in P^{ac}_{2,P}$ with $Q|\mathcal{F}_t = P|\mathcal{F}_t$. We will show $\frac{dQ}{dP} \leq \frac{1}{\lambda} \Leftrightarrow Q(A|\mathcal{F}_t) \leq \min\left\{1, \frac{P(A|\mathcal{F}_t)}{\lambda}\right\}$.

Step 1: "$\Leftarrow$"

We show: $Q \notin \tilde{\mathcal{Q}}_t \Rightarrow Q \notin \tilde{\mathcal{D}}^{\Psi}_t$. Let A be a $(P-)$non-null set, i.e. $P(A) > 0$ with $\frac{dQ}{dP} > \frac{1}{\lambda}$ on A (exists by assumption). Since $Q(A|\mathcal{F}_t)$ is $\mathcal{F}_t$-measurable and $Q|\mathcal{F}_t = P|\mathcal{F}_t$, we get

$$\begin{aligned} E_P[Q(A|\mathcal{F}_t)] &= E_Q[Q(A|\mathcal{F}_t)] = E_Q[E_Q[\mathbb{1}_A|\mathcal{F}_t]] \\ &= E_Q[\mathbb{1}_A] = E_P\left[\frac{dQ}{dP}\mathbb{1}_A\right] \\ &> E_P\left[\frac{1}{\lambda}\mathbb{1}_A\right] = E_P\left[\frac{1}{\lambda}E_P[\mathbb{1}_A|\mathcal{F}_t]\right] \\ &= E_P\left[\frac{P(A|\mathcal{F}_t)}{\lambda}\right]. \end{aligned}$$

Because $\frac{P(A|\mathcal{F}_t)}{\lambda} \geq 0$ and $Q(A|\mathcal{F}_t) \geq 0$, this implies that there is a non-null set in $\mathcal{F}$ on which $Q(A|\mathcal{F}_t) > \min\left\{1, \frac{P(A|\mathcal{F}_t)}{\lambda}\right\}$, i.e. $Q \notin \tilde{\mathcal{D}}^{\Psi}_t$ which concludes the proof of "$\Leftarrow$".

Step 2: "$\Rightarrow$"

Let $Q \notin \tilde{\mathcal{D}}^{\Psi}_t$, i.e. there exists an $A \in \mathcal{F}$ with $P(A) > 0$ and a P-non-null-set $B \in \mathcal{F}_t$ such that $Q(A|\mathcal{F}_t) > \min\left\{1, \frac{P(A|\mathcal{F}_t)}{\lambda}\right\}$ on B ($B \in \mathcal{F}_t$ and not just $\in \mathcal{F}$ since both $Q(A|\mathcal{F}_t)$ and $\min\left\{1, \frac{P(A|\mathcal{F}_t)}{\lambda}\right\}$ are $\mathcal{F}_t$-measurable). Set $A' := A \cap B \cap \{P(A|\mathcal{F}_t) < \lambda\} \in \mathcal{F} \cap \mathcal{F}_t \cap \mathcal{F}_t \subseteq \mathcal{F}$.

There are two possible cases:

1. **First case:** $P(A') > 0$. Then by assumption

$$Q(A \cap B \cap \{P(A|\mathcal{F}_t) \geq \lambda\}) + Q(\underbrace{A \cap B \cap \{P(A|\mathcal{F}_t) < \lambda\}}_{=A'})$$
$$\begin{aligned} &= Q(A \cap B) \\ &= E_Q[\mathbb{1}_A \mathbb{1}_B] \\ &= E_Q[E_Q[\mathbb{1}_A \underbrace{\mathbb{1}_B}_{\mathcal{F}_t\text{-msble}} |\mathcal{F}_t]] \\ &= E_Q[\underbrace{\mathbb{1}_B Q(A|\mathcal{F}_t)}_{\mathcal{F}_t\text{-msble}}]. \end{aligned}$$

And since $Q|\mathcal{F}_t = P|\mathcal{F}_t$, this is equal to $E_P[\mathbb{1}_B Q(A|\mathcal{F}_t)]$, i.e.

$$\begin{aligned} &Q(A \cap B \cap \{P(A|\mathcal{F}_t) \geq \lambda\}) + Q(A') \\ &\quad = E_P[\mathbb{1}_B Q(A|\mathcal{F}_t)] \\ &\quad > E_P\left[\mathbb{1}_B \min\left\{1, \frac{P(A|\mathcal{F}_t)}{\lambda}\right\}\right] \\ &\quad = E_P\left[\mathbb{1}_B \mathbb{1}_{\{P(A|\mathcal{F}_t)\geq\lambda\}} + \mathbb{1}_B \frac{P(A|\mathcal{F}_t)}{\lambda} \mathbb{1}_{\{P(A|\mathcal{F}_t)<\lambda\}}\right] \\ &\quad = P(\underbrace{B \cap \{P(A|\mathcal{F}_t) \geq \lambda\}}_{\in\mathcal{F}_t}) + \frac{1}{\lambda} E_P[\underbrace{\underbrace{\mathbb{1}_B}_{\mathcal{F}_t\text{-msble}} E_P[\mathbb{1}_A|\mathcal{F}_t] \underbrace{\mathbb{1}_{\{P(A|\mathcal{F}_t)<\lambda\}}}_{\mathcal{F}_t\text{-msble}}}_{=E_P[\mathbb{1}_B \mathbb{1}_A \mathbb{1}_{\{P(A|\mathcal{F}_t)<\lambda\}}|\mathcal{F}_t]}] \\ &\quad = Q(B \cap \{P(A|\mathcal{F}_t) \geq \lambda\}) + E_P\left[\frac{1}{\lambda} \mathbb{1}_{A\cap B\cap\{P(A|\mathcal{F}_t)<\lambda\}}\right] \\ &\quad \geq Q(A \cap B \cap \{P(A|\mathcal{F}_t) \geq \lambda\}) + E_P\left[\frac{1}{\lambda} \mathbb{1}_{\underbrace{A\cap B\cap\{P(A|\mathcal{F}_t)<\lambda\}}_{=A'}}\right]. \end{aligned}$$

The last inequality is due to the fact that in general $Pr[U \cap V] \leq Pr[U]$ (for any probability measure Pr). Subtracting $Q(A \cap B \cap \{P(A|\mathcal{F}_t) \geq \lambda\})$ on both sides yields

$$E_P\left[\frac{dQ}{dP} \mathbb{1}_{A'}\right] = E_Q[\mathbb{1}_{A'}] = Q(A') > E_P\left[\frac{1}{\lambda} \mathbb{1}_{A'}\right].$$

Since $\frac{1}{\lambda} > 0$, $\frac{dQ}{dP} \geq 0$ and $P(A') > 0$ it follows that $\frac{dQ}{dP} > \frac{1}{\lambda}$ on a non-null subset of A'.

2. **Second case:** $P(A') = 0$. In this case the above inequality becomes $Q(A \cap \{P(A|\mathcal{F}_t) \geq \lambda\}) > Q(A \cap \{P(A|\mathcal{F}_t) \geq \lambda\})$ which is a contradiction.

This completes the proof.

□

Lemma 3.3. *Let $t < T$. Then*

$$esssup_{Q\in\tilde{\mathcal{Q}}_t} E_Q[-X|\mathcal{F}_t] = esssup_{Q\in\mathcal{Q}} E_Q[-X|\mathcal{F}_t],$$

where $\mathcal{Q} := \{Q \in \mathcal{P}^{ac}_{2,P} : \frac{dQ}{dP} \leq \frac{1}{\lambda}\}$, which is the same set as $\mathcal{D}^\Psi$ from the static setting when $\Psi(x) = min\{1, x/\lambda\}$.

The proof of this lemma is analogue to the proof of the two lemmas before and thus skipped. Now we are able to prove the following counterexample:

Proof of the counterexample: We consider the case of $AVaR$, i.e. $\Psi(x) = \min\{1, x/\lambda\}$, with $X = W(T)$ (W a Brownian motion) and $\mathcal{F}_t = \sigma(W_s, s \leq t)\forall t \in [0,T]$, in particular, X is a *continuous* random variable and thus we know that

$$E[-X| - X > q_{-X}(1-\lambda)] = AVaR_\lambda(X) = C^\Psi[-X] = \text{esssup}_{Q\in\mathcal{D}^\Psi} E_Q[-X].$$

We will now show that the tower property does not hold, i.e. it does not hold that $\mathcal{C}^\Psi[\mathcal{C}^\Psi[-X|\mathcal{F}_t]] = \mathcal{C}^\Psi[-X]$. The idea behind is that this is equivalent to time-inconsistency, provided that all the other properties of a dynamic risk measure are fulfilled. First notice that from Lemmas 3.1, 3.2 and 3.3 it follows that

$$C^\Psi[-X|\mathcal{F}_t] := \text{esssup}_{Q\in\mathcal{D}^\Psi_t} E_Q[-X|\mathcal{F}_t] = \text{esssup}_{Q\in\mathcal{Q}(=\mathcal{D}^\Psi)} E_Q[-X|\mathcal{F}_t]$$

Thus, we have

$$\begin{aligned}
\mathcal{C}^\Psi[\mathcal{C}^\Psi[-W(T)|\mathcal{F}_t]] &= \mathcal{C}^\Psi[\text{esssup}_{Q\in\mathcal{Q}} E_Q[-W(T)|\mathcal{F}_t]] \\
&= \mathcal{C}^\Psi[\text{esssup}_{Q\in\mathcal{Q}} E_Q[-(W(T)-W(t)) - \underbrace{W(t)}_{\mathcal{F}_t\text{-msble}} |\mathcal{F}_t]] \\
&= \mathcal{C}^\Psi[\text{esssup}_{Q\in\mathcal{Q}}\{-W(t) + E_Q[-(W(T)-W(t))|\mathcal{F}_t]\}] \\
&= \mathcal{C}^\Psi[-W(t) + \text{esssup}_{Q\in\mathcal{Q}} E_Q[-(W(T)-W(t))|\mathcal{F}_t]].
\end{aligned}$$

At this point we cannot assume that $(W(T) - W(t))$ is Q-independent of $\mathcal{F}_t$, but what we can do is to choose a particular member of $\mathcal{Q}$ knowing that the expectation in this case is smaller or equal the supremum over all measures in $\mathcal{Q}$:

Choose Q^* such that

$$\frac{dQ^*}{dP} = \frac{1}{\lambda}\mathbb{1}_{\{-(W(T)-W(t))>q_{-(W(T)-W(t))}(1-\lambda)\}}. \tag{5}$$

This clearly defines a measure with

$$Q^*(\Omega) = \int_\Omega \frac{dQ^*}{dP}dP = \frac{1}{\lambda}P(-(W(T)-W(t)) > q_{-(W(T)-W(t))}(1-\lambda)) = \frac{\lambda}{\lambda} = 1,$$

i.e., Q^* is a probability measure and since $\frac{dQ^*}{dP} \leq \frac{1}{\lambda}$ (and thus of course also in L^2), we clearly have $Q^* \in \mathcal{Q}$. Therefore,

$$
\begin{aligned}
&\mathcal{C}^\Psi[\mathcal{C}^\Psi[-W(T)|\mathcal{F}_t]] \\
&\quad \geq \mathcal{C}^\Psi[-W(t) + E_{Q^*}[-(W(T)-W(t))|\mathcal{F}_t]] \\
&\quad = \mathcal{C}^\Psi\left[-W(t) + \frac{E_P[-(W(T)-W(t))\frac{1}{\lambda}\mathbb{1}_{\{-(W(T)-W(t))>q_{-(W(T)-W(t))}(1-\lambda)\}}|\mathcal{F}_t]}{E_P[\frac{1}{\lambda}\mathbb{1}_{\{-(W(T)-W(t))>q_{-(W(T)-W(t))}(1-\lambda)\}}|\mathcal{F}_t]}\right]
\end{aligned}
$$

where we use the Bayes rule we already used in the proof of Lemma 3.1. Since $(W(T) - W(t))$ is P-independent of $\mathcal{F}_t$ both conditional expectations become ”normal” expectations and the denominator reduces to 1, such that we get:

$$
\begin{aligned}
&\mathcal{C}^\Psi[\mathcal{C}^\Psi[-W(T)|\mathcal{F}_t]] \\
&\quad \geq \mathcal{C}^\Psi\left[-W(t) + \underbrace{E_P\left[-(W(T)-W(t))\frac{1}{\lambda}\mathbb{1}_{\{-(W(T)-W(t))>q_{-(W(T)-W(t))}(1-\lambda)\}}\right]}_{\in\mathbb{R}\text{ (deterministic)}}\right] \\
&\quad = \operatorname{esssup}_{Q\in\mathcal{Q}} E_Q[-W(t)] \\
&\qquad + E_P\left[-(W(T)-W(t))\frac{1}{\lambda}\mathbb{1}_{\{-(W(T)-W(t))>q_{-(W(T)-W(t))}(1-\lambda)\}}\right] \\
&\quad \geq E_P\left[-W(t)\frac{1}{\lambda}\mathbb{1}_{\{-W(t)>q_{-W(t)}(1-\lambda)\}}\right] \\
&\qquad + E_P\left[-(W(T)-W(t))\frac{1}{\lambda}\mathbb{1}_{\{-(W(T)-W(t))>q_{-(W(T)-W(t))}(1-\lambda)\}}\right].
\end{aligned}
$$

The last inequality can be shown exactly the same way as the one before. Now let $t = 1$ and $T = 2$, then we get

$$
\begin{aligned}
\mathcal{C}^\Psi[\mathcal{C}^\Psi[-W(2)|\mathcal{F}_1]] &\geq 2E_P\left[-W(1)\frac{1}{\lambda}\mathbb{1}_{\{-W(1)>q_{-W(1)}(1-\lambda)\}}\right] \\
&= \frac{2}{\lambda}\int_{\Phi^{-1}(1-\lambda)}^{\infty} z\varphi(z)dz \\
&> \frac{\sqrt{2}}{\lambda}\int_{\Phi^{-1}(1-\lambda)}^{\infty} z\varphi(z)dz \\
&= AVaR_\lambda(W(2)) \\
&= \mathcal{C}^\Psi[-W(2)]
\end{aligned} \tag{6}
$$

This completes the proof.

□

3.3 Construction of Dynamic Version: A New Approach

In the last section, we have shown that ρ defined by $\rho_t(X) := \mathcal{C}^\Psi[-X|\mathcal{F}_t] := \operatorname{esssup}_{Q\in\mathcal{D}_t^\Psi} E_Q[-X|\mathcal{F}_t]$ where $\mathcal{D}_t^\Psi := \{Q \in \mathcal{P}_{2,P}^{ac} : Q(A|\mathcal{F}_t) \leq \Psi(P(A|\mathcal{F}_t))\forall A \in$

$\mathcal{F}\}$ is in general not time-consistent, i.e. it does not yield a dynamic risk measure. Therefore we need to make another approach. Instead of defining a dynamic risk measure in **continuous time** directly, we will construct a time-consistent dynamic risk measure in **discrete time** and let the time increments go to zero. As discribed in Stadje (2010) [2], without some proper rescaling this will in general lead to an "explosion" of the risk valuation in the limit.

We start by defining the so called "valuations", because they are related to BSDEs as shown later.

Definition 3.3 (Valuations).
*Let ρ be a risk measure (static or dynamic). The corresponding **valuation** ϕ is defined as $\phi = -\rho$.*

We need to set up a discrete time framework first and we introduce the same notation as in [2].

Definition 3.4 (Discrete Time Setup).
Let again $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ be a filtrated probability space, $\mathcal{F}_T = \mathcal{F}$. Assume $0 = t_0 < t_1 < ... < t_k = T$ and define for $i \in \{0, ..., k-1\}$ the set of i^{th} one-period transition densities

$$D_{t_{i+1}} := \{\zeta \in L^1_+(\mathcal{F}_{t_{i+1}}) : E[\zeta|\mathcal{F}_{t_i}] = 1\}.$$

Furthermore let $\xi := (\xi_{t_j})_{j=1,...,k} \in D := D_{t_1} \times ... \times D_{t_k}$. We now define equivalent probability measures Q^ξ by

$$\frac{dQ^\xi}{dP}|\mathcal{F}_{t_r} := \prod_{j=1}^{r} \xi_{t_j}. \tag{7}$$

But instead of looking at valuations of the form $\operatorname{essinf}_{Q\in\mathcal{D}_t^\Psi} E_Q[X|\mathcal{F}_t]$ (which are coming from the above time-inconsitent risk measures), we now consider the so called **one-period valuations**.

Definition 3.5 (One-period valuations).
*The **one-period valuations** $F_{t_i} : L^\infty(\mathcal{F}_{t_{i+1}}) \to L^\infty(\mathcal{F}_{t_i})$, are defined by*

$$F_{t_i}(X) := \mathit{essinf}_{\xi_{t_{i+1}}\in D_{t_{i+1}}, Q^\xi(A|\mathcal{F}_{t_i})\leq\Psi(P(A|\mathcal{F}_{t_i}))\forall A\in\mathcal{F}_{t_{i+1}}} E_P[\xi_{t_{i+1}} X|\mathcal{F}_{t_i}]. \tag{8}$$

This is, by an argument like in the proof of Lemma 3.1, only a P-measurable version of

$$F_{t_i}(X) := \operatorname{essinf}_{Q^\xi\in\mathcal{D}^{t_i}} E_{Q^\xi}[X|\mathcal{F}_{t_i}],$$

where $\mathcal{D}^{t_i} := \{Q^\xi : \xi_{t_{i+1}} \in D_{t_{i+1}}, Q^\xi(A|\mathcal{F}_{t_i}) \leq \Psi(P(A|\mathcal{F}_{t_i}))\forall A \in \mathcal{F}_{t_{i+1}}\}$. Remember that we defined the esssup and essinf w.r.t. P so the second version would not be sufficient since the conditional expectation inside is only Q^ξ-measurable. But nonetheless we will use both representations later, always meaning the P-measurable version.

An important observation is that F_{t_i} is nothing else than the valuation based on the dynamic risk measure introduced in the last section (eq. (3)) **but only**

evaluating at the timepoint t_i and restricted on the probability space $(\Omega, \mathcal{F}_{t_{i+1}}, P|\mathcal{F}_{t_{i+1}})$. It is thus a **static** valuation, which means that we do not run into trouble with time-inconsistency.

As in Definition 4.2 from Stadje (2010) [2], the so called **penalty functions** are given by

$$\varphi_{t_i}^{F_{t_i}}(Q^\xi) := \varphi_{t_i}^{F_{t_i}}(\xi_{t_{i+1}}) := \text{esssup}_{X \in L^\infty(\mathcal{F}_{t_{i+1}})}\{F_{t_i}(X) - E_P[\xi_{t_{i+1}} X|\mathcal{F}_{t_i}]\}. \quad (9)$$

There are only two values wich can be taken by the esssup:

1. **Case 1:** $\xi_{t_{i+1}}$ is such that $Q^\xi(A|\mathcal{F}_{t_i}) \leq \Psi(P(A|\mathcal{F}_{t_i})) \forall A \in \mathcal{F}_{t_{i+1}}$ a.s.
   In this case we clearly have for all $X \in L^\infty(\mathcal{F}_{t_{i+1}})$ that (by definition of F_{t_i}) $F_{t_i}(X) \leq E_P[\xi_{t_{i+1}} X|\mathcal{F}_{t_i}]$. And we know that equality holds for $X = 0$. Thus, in this case $\varphi_{t_i}^{F_{t_i}}(\xi_{t_{i+1}}) = 0$.

2. **Case 2:** $\exists A \in \mathcal{F}_{t_{i+1}}$ with $P(A) > 0$ and $P(Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i}))) > 0$
   In this case, we have

$$\begin{aligned} 0 < P(Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i}))) &= P(\underbrace{Q^\xi(A|\mathcal{F}_{t_i}) - \Psi(P(A|\mathcal{F}_{t_i}))}_{=:Y} > 0) \\ &= 1 - P(Y \leq 0) = 1 - F_Y(0) \\ \Leftrightarrow F_Y(0) < 1. \end{aligned}$$

   Because of right-continuity of distribution functions we know that there exists an $\varepsilon > 0$ such that $F_Y(\varepsilon) < 1$, i.e. (by the above calculations backwards) $P(Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon) > 0$. Now fix ε and consider the set of random variables $(X_n)_{n \in \mathbb{N}}$, given by

$$X_n = -n \mathbb{1}_{\{Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\} \cap A} \in L^\infty(\mathcal{F}_{t_{i+1}}).$$

   Notice that $\{Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\} \cap A$ is not a null-set since otherwise $\xi_{t_{i+1}} \mathbb{1}_{\{Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\} \cap A}$ would be 0 a.s. but we have

$$\begin{aligned} &E_P[\xi_{t_{i+1}} \mathbb{1}_{\{Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\} \cap A}] \\ &\quad = E_P[E_P[\xi_{t_{i+1}} \mathbb{1}_{\{Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\} \cap A} | \mathcal{F}_{t_i}]] \\ &\quad = E_P[E_{Q^\xi}[\mathbb{1}_{\{Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\} \cap A} | \mathcal{F}_{t_i}]] \\ &\quad = E_P[\mathbb{1}_{\{Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\}} \underbrace{Q^\xi(A|\mathcal{F}_{t_i})}_{> \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon \geq \varepsilon}] \\ &\quad > \varepsilon P(Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon) \\ &\quad > 0, \end{aligned}$$

   since $\{Q^\xi(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\}$ is not a null-set by assumption. Then (by definition of F_{t_i})

$$
\begin{aligned}
&F_{t_i}(X_n) - E_P[\xi_{t_{i+1}} X_n | \mathcal{F}_{t_i}] \\
&\quad = \text{essinf}_{\tilde{\xi}_{t_{i+1}} \in D_{t_{i+1}}, Q^{\tilde{\xi}}(\tilde{A}|\mathcal{F}_{t_i}) \leq \Psi(P(\tilde{A}|\mathcal{F}_{t_i})) \forall \tilde{A} \in \mathcal{F}_{t_{i+1}}} \\
&\qquad E_P[\underbrace{-n \mathbb{1}_{\{Q^{\xi}(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\}}}_{\mathcal{F}_{t_i}\text{-msble}} (\tilde{\xi}_{t_{i+1}} - \xi_{t_{i+1}}) \mathbb{1}_A | \mathcal{F}_{t_i}] \\
&\quad = \text{essinf}_{\tilde{\xi}_{t_{i+1}} \in D_{t_{i+1}}, Q^{\tilde{\xi}}(\tilde{A}|\mathcal{F}_{t_i}) \leq \Psi(P(\tilde{A}|\mathcal{F}_{t_i})) \forall \tilde{A} \in \mathcal{F}_{t_{i+1}}} \\
&\qquad \{-n \mathbb{1}_{\{Q^{\xi}(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\}} \underbrace{E_P[(\tilde{\xi}_{t_{i+1}} - \xi_{t_{i+1}}) \mathbb{1}_A | \mathcal{F}_{t_i}]}_{= E_{Q^{\tilde{\xi}}}[\mathbb{1}_A|\mathcal{F}_{t_i}] - E_{Q^{\xi}}[\mathbb{1}_A|\mathcal{F}_{t_i}]} \} \\
&\quad = \text{essinf}_{\tilde{\xi}_{t_{i+1}} \in D_{t_{i+1}}, Q^{\tilde{\xi}}(\tilde{A}|\mathcal{F}_{t_i}) \leq \Psi(P(\tilde{A}|\mathcal{F}_{t_i})) \forall \tilde{A} \in \mathcal{F}_{t_{i+1}}} \\
&\qquad \{n \mathbb{1}_{\{Q^{\xi}(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\}} \underbrace{(\underbrace{Q^{\xi}(A|\mathcal{F}_{t_i})}_{> \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon} - \underbrace{Q^{\tilde{\xi}}(A|\mathcal{F}_{t_i})}_{\leq \Psi(P(A|\mathcal{F}_{t_i}))})}_{> \varepsilon} \} \\
&\quad > n\varepsilon \mathbb{1}_{\{Q^{\xi}(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\}} \\
&\quad \to \infty \mathbb{1}_{\{Q^{\xi}(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})) + \varepsilon\}}, \text{ a.s. for } n \to \infty \\
&\quad \to \infty \mathbb{1}_{\{Q^{\xi}(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i}))\}}, \text{ a.s. for } \varepsilon \to 0.
\end{aligned}
$$

In particular for every $y \in \mathbb{R}$ there exists an $X \in L^{\infty}(\mathcal{F}_{t_{i+1}})$, such that $F_{t_i}(X) - E_P[\xi_{t_{i+1}} X|\mathcal{F}_{t_i}] > y$ on $\{Q^{\xi}(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i}))\}$. This means that in this case $\varphi_{t_i}^{F_{t_i}}(\xi_{t_{i+1}}) = \infty$ on $\{Q^{\xi}(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i}))\}$.

Now, it follows from Proposition 4.5 in Stadje 2010 [2], that we can define a valuation (in particular time-consistent) in discrete time $(\phi_s)_{s \in t_0, \ldots t_k}$, $\phi_s : L^{\infty}(\mathcal{F}_T) \to L^{\infty}(\mathcal{F}_s)$ by

$$
\begin{aligned}
\phi_{t_i}(X) &:= \text{essinf}_{Q \in D} E_Q[X + \underbrace{\sum_{j=i}^{k-1} \varphi_{t_j}^{F_{t_j}}(Q)}_{= \infty \mathbb{1}_{\cup_{j=i}^{k} \{\exists A_j \in \mathcal{F}_{t_{j+1}} : Q^{\xi}(A_j|\mathcal{F}_{t_j}) > \Psi(P(A_j|\mathcal{F}_{t_j}))\}}} | \mathcal{F}_{t_i}] \\
&= \text{essinf}_{Q \in \hat{\mathcal{D}}^{t_i}} E_Q[X|\mathcal{F}_{t_i}],
\end{aligned}
$$

where $\hat{\mathcal{D}}^{t_i} = \{Q \in \mathcal{P}_{2,P}^{ac} : Q(A_j|\mathcal{F}_{t_j}) \leq \Psi(P(A_j|\mathcal{F}_{t_j})) \forall A_j \in \mathcal{F}_{t_{j+1}} \forall j \geq i\}$ (w.l.o.g. it suffices to take these measures into account since the $\infty \mathbb{1}$...-part disappears then). Again, by the Bayes argument which was already used before, the essinf can also be taken over $\hat{\mathcal{D}} = \{Q \in \mathcal{P}_{2,P}^{ac} : Q(A_j|\mathcal{F}_{t_j}) \leq \Psi(P(A_j|\mathcal{F}_{t_j})) \forall A_j \in \mathcal{F}_{t_{j+1}} \forall j \geq 1\}$, so that we get the following equivalent definition of ϕ.

Definition 3.6 (Discrete Valuations and Distorted Risk Measure**).**
Let

$$
\hat{\mathcal{D}} = \{Q \in \mathcal{P}_{2,P}^{ac} : Q^{\xi}(A_j|\mathcal{F}_{t_j}) \leq \Psi(P(A_j|\mathcal{F}_{t_j})) \forall A_j \in \mathcal{F}_{t_{j+1}} \forall j \geq 1\}.
$$

Then

$$
\phi_{t_i}(X) = \mathit{essinf}_{Q \in \hat{\mathcal{D}}} E_Q[X|\mathcal{F}_{t_i}], \tag{10}
$$

defines a valuation in discrete time. The corresponding risk measure $\rho = -\phi$ is called ***distorted risk measure****.*

This definition is used in Madan et al. (2013) [1] (Definition 5). On the other hand, Proposition 4.5 in Stadje (2010) [2] states that this time consistent discrete valuation can be cunstructed recursively using the one-period valuations:

Lemma 3.4 (Recursive Definition)**.**
Let $F_{t_0}, ..., F_{t_{k-1}}$ be the one-period valuations. Then the valuation ϕ from Definition 3.6 can be defined recursively by

$$\begin{cases} \phi_T(X) &= X \\ \phi_{t_i}(X) &= F_{t_i}(\phi_{t_{i+1}}(X)). \end{cases} \tag{11}$$

This recursive definition corresponds exactly to Proposition 2 (ii) in [1]. It **fulfills the tower property by construction** (as $F_{t_i} = \phi_{t_i}|\mathcal{F}_{t_{i+1}}$).

It can be shown (see for example Proposition 2 (iii) in [1]) that the one-period valuations F_{t_i} can be written as a Choquet integral:

Lemma 3.5 (Choquet Representation of One-Period Valuations)**.**
For $X \in L^2(\mathcal{F}_{t_{i+1}})$, it holds that

$$F_{t_i}(X) = \int_{-\infty}^{0} (\Psi(P(X > x|\mathcal{F}_{t_i})) - 1)dx + \int_{0}^{\infty} \Psi(P(X > x|\mathcal{F}_{t_i}))dx. \tag{12}$$

These were the basic definitions and properties in the discrete-time case. Now we want to see if we can construct a valuation in continuous time by increasing the number of time-points, such that the length of the subintervals goes to zero. To do so, we will first define the discrete time valuations in a Bernoulli random walk setting which converges to a Brownian motion setting as it was practiced in [2]. But we have to be careful, since the resulting limiting valuation in continuous time will depend on the approximating setting we chose at the beginning. In fact, if one compares the results within this work with the ones in [1] where a multinomial random walk was chosen to approximate Brownian motion, one can see that the limiting valuations differ. **So every result from now on has to be seen as a result one gets when using a Bernoulli random walk for approximation**. The consequences will be discussed later in the conclusion.

3.4 Bernoulli Random Walk Setting: Rescaling and Convergence towards a BSDE

Consider a sequence $(\pi_N)_{N\in\mathbb{N}}$ of partitions $\pi_N = \{t_0, t_1, ..., t_{k(N)} : 0 = t_0 < t_1 < ... < t_{k(N)} = T\}$ whose mesh converges to 0 as $N \to \infty$ and define $(B_j^{N,l})_{j=1,...,k(N);l=1,...,d}$ to be i.i.d. translated and dilated Bernoulli random variables, such that $P(B_j^{N,l} = 1) = P(B_j^{N,l} = -1) = 0.5$. For every N, the d-dimensional Bernoulli random walk R^N is now defined as the process which is constant on any interval $[t_i, t_{i+1})$ and

$$R^{N,l}(t_i) = \sum_{j=1}^{i} \sqrt{\Delta t_j} B_j^{N,l}, \; i = 1, ..., k(N), \; l = 1, ..., d.$$

Because we know (by a generalization of Donsker's theorem) that R^N converges in distribution against a d-dimensional Brownian motion W, we can use Theorem I.2.7 in Ikeda and Watanabe [13] to show that there exists a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ and processes $\tilde{R}^N \overset{d}{=} R^N$ and $\tilde{W} \overset{d}{=} W$, such that

$$\sup_{t\in[0,T]} |\tilde{R}^N(t) - \tilde{W}(t)| \to 0 \text{ in } a.s.$$

For uniform integrability of the sequence $\sup_{t\in[0,T]} |\tilde{R}^N(t) - \tilde{W}(t)|^2$ (compare Cheridito and Stadje (2013) [14]), we also get:

$$\sup_{t\in[0,T]} |\tilde{R}^N(t) - \tilde{W}(t)| \to 0 \text{ in } L^2.$$

From now on let us work on this probability space and let us skip the tilde for simplicity, i.e. $(\Omega, \mathcal{F}, P) := (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, $R^N := \tilde{R}^N$ and $W := \tilde{W}$. Denote by $(\mathcal{F}_t)_{t\in[0,T]}$ the filtration generated by W and by $(\mathcal{F}_t^N)_{t\in[0,T]}$ the filtration generated by R^N. One of the main conclusions of [2] is, that in this setting, valuations cunstructed as in (11) "explode" for $N \to \infty$. Concretely it states the following:

Remark 3.3 ("Explosion" of Valuations, Stadje (2010))**.**
Define $F_{t_i}^N$ as in Definition 3.5 (eq. (8)) and ϕ^N as in Lemma 3.4 (eq. (11)), w.r.t. the filtration $(\mathcal{F}_{t_i}^N)_{i=0,\dots,k(N)}$. Then, under some weak conditions, it exists a payoff $X \in L^2(\mathcal{F}_T)$ (from the Brownian setting) and a sequence of $\mathcal{F}_T^N$-measurable payoffs X^N (i.e. from the Bernoulli random walk setting) such that $X^N \to X$ in L^2 as $N \to \infty$ and for all $t \in [0,T)$, $\phi_t^N(X^N) \to -\infty$ a.s. as $N \to \infty$.

This means that the discrete valuations constructed this way are "too conservative" and in particular it means that if one tries to construct a continuous time valuation by letting $N \to \infty$, $\max_{i=0,\dots,k(N)-1} |t_{i+1} - t_i| \to 0$ this can not be done directly but some proper rescaling is needed.

The rescaling suggested in [2] is the following: For $X \in L^2(\mathcal{F}_{t_{i+1}}^N)$, set

$$\phi_{t_i,t_{i+1}}^N(X) := (1 - \sqrt{\Delta t_{i+1}})E[X|\mathcal{F}_{t_i}^N] + \sqrt{\Delta t_{i+1}}\hat{F}_{t_i}^N(X), \tag{13}$$

where

$$\hat{F}_{t_i}^N(X) = \sqrt{\Delta t_{i+1}} F_{t_i}^N\left(\frac{X}{\sqrt{\Delta t_{i+1}}}\right). \tag{14}$$

Now a discrete dynamic valuation is defined by:

$$\begin{cases} \phi_T^N(X^N) &= X^N \\ \phi_{t_i}^N(X^N) &= \phi_{t_i,t_{i+1}}^N(\phi_{t_{i+1}}^N(X^N)). \end{cases} \tag{15}$$

Another way of interpretation is that instead of rescaling the one-period valuations directly the underlying **distortion** is rescaled. This is shown in the following lemma:

Lemma 3.6 (Square root scaling)**.**
Consider the following ***family of distortions*** $\tilde{\Psi} : [0,1] \to [0,1]$*:*

$$\tilde{\Psi}(p,\delta) := p + \sqrt{\delta}(\Psi(p) - p). \tag{16}$$

Set $\tilde{\Psi}_i^N := \tilde{\Psi}(\cdot, \Delta t_{i+1})$ *and define* $F_{t_i}^{\tilde{\Psi}_i^N}(X)$ *as in (8) but with the distortion* $\tilde{\Psi}_i^N$ *instead of* Ψ*. Then for* $\phi_{t_i,t_{i+1}}$ *as just derived it holds that* $\phi_{t_i,t_{i+1}} = F_{t_i}^{\tilde{\Psi}_i^N}$*, i.e. the rescaled risk measure from (15) is the same as in (11) with* $F_{t_i}^{\tilde{\Psi}_i^N}$ *instead of* F_{t_i}*.*

Proof:

From (12) we know that $F_{t_i}^{\tilde{\Psi}_i^N}(X)$ has the representation

$$\begin{aligned}
&F_{t_i}^{\tilde{\Psi}_i^N}(X)\\
&= \int_{-\infty}^{0} (\tilde{\Psi}_i^N(P(X > x|\mathcal{F}_{t_i}^N)) - 1)\mathrm{d}x + \int_0^{\infty} \tilde{\Psi}_i^N(P(X > x|\mathcal{F}_{t_i}^N))\mathrm{d}x\\
&= \int_{-\infty}^{0} P(X > x|\mathcal{F}_{t_i}^N) + \sqrt{\Delta t_{i+1}}[\Psi(P(X > x|\mathcal{F}_{t_i}^N)) - P(X > x|\mathcal{F}_{t_i}^N)] - 1\mathrm{d}x\\
&\quad + \int_0^{\infty} P(X > x|\mathcal{F}_{t_i}^N) + \sqrt{\Delta t_{i+1}}[\Psi(P(X > x|\mathcal{F}_{t_i}^N)) - P(X > x|\mathcal{F}_{t_i}^N)]\mathrm{d}x\\
&= \underbrace{\int_{-\infty}^{0} (P(X > x|\mathcal{F}_{t_i}^N) - 1)\mathrm{d}x + \int_0^{\infty} P(X > x|\mathcal{F}_{t_i}^N)\mathrm{d}x}_{E[X|\mathcal{F}_{t_i}]}\\
&\quad + \int_{-\infty}^{0} \sqrt{\Delta t_{i+1}}[\Psi(P(X > x|\mathcal{F}_{t_i}^N)) - 1]\mathrm{d}x\\
&\quad - \int_{-\infty}^{0} \sqrt{\Delta t_{i+1}}[P(X > x|\mathcal{F}_{t_i}^N) - 1]\mathrm{d}x\\
&\quad + \int_0^{\infty} \sqrt{\Delta t_{i+1}}\Psi(P(X > x|\mathcal{F}_{t_i}^N))\mathrm{d}x\\
&\quad - \int_0^{\infty} \sqrt{\Delta t_{i+1}}P(X > x|\mathcal{F}_{t_i}^N)\mathrm{d}x,
\end{aligned}$$

provided that all these integrals exist. Note that the term $0 = \sqrt{\Delta t_{i+1}} - \sqrt{\Delta t_{i+1}}$ was added and pulled into the second and third integral. Furthermore note that the integrals with the minus sign in front form together the term $-\sqrt{\Delta t_{i+1}}E[X|\mathcal{F}_{t_i}^N]$. Thus, for $T(x) = x\sqrt{\Delta t_{i+1}}$ (note that $T((-\infty,0]) = (-\infty,0]$ and $T([0,\infty)) = [0,\infty)$) we get by substitution rule

$$
\begin{aligned}
F_{t_i}^{\tilde{\Psi}_i^N}(X) &= (1-\sqrt{\Delta t_{i+1}})E[X|\mathcal{F}_{t_i}^N] + \int_{-\infty}^{0} \sqrt{\Delta t_{i+1}}[\Psi(P(X > x|\mathcal{F}_{t_i}^N)) - 1]\mathrm{d}x \\
&\quad + \int_{0}^{\infty} \sqrt{\Delta t_{i+1}}\Psi(P(X > x|\mathcal{F}_{t_i}^N))\mathrm{d}x \\
&= (1-\sqrt{\Delta t_{i+1}})E[X|\mathcal{F}_{t_i}^N] \\
&\quad + \int_{-\infty}^{0} \sqrt{\Delta t_{i+1}}[\Psi(P(X > T(x)|\mathcal{F}_{t_i}^N)) - 1]|T'(x)|\mathrm{d}x \\
&\quad + \int_{0}^{\infty} \sqrt{\Delta t_{i+1}}\Psi(P(X > T(x)|\mathcal{F}_{t_i}^N))|T'(x)|\mathrm{d}x \\
&= (1-\sqrt{\Delta t_{i+1}})E[X|\mathcal{F}_{t_i}^N] \\
&\quad + \int_{-\infty}^{0} \sqrt{\Delta t_{i+1}}\left[\Psi\left(P\left(\frac{X}{\sqrt{\Delta t_{i+1}}} > x \middle| \mathcal{F}_{t_i}^N\right)\right) - 1\right]\sqrt{\Delta t_{i+1}}\mathrm{d}x \\
&\quad + \int_{0}^{\infty} \sqrt{\Delta t_{i+1}}\Psi\left(P\left(\frac{X}{\sqrt{\Delta t_{i+1}}} > x \middle| \mathcal{F}_{t_i}^N\right)\right)\sqrt{\Delta t_{i+1}}\mathrm{d}x \\
&= (1-\sqrt{\Delta t_{i+1}})E[X|\mathcal{F}_{t_i}^N] + \Delta t_{i+1}F_{t_i}^N\left(\frac{X}{\sqrt{\Delta t_{i+1}}}\right).
\end{aligned}
$$

This is just the same as $\phi_{t_i,t_{i+1}}$ (compare (13) and (14)).

□

It is shown in Stadje (2010) [2] that instead of "exploding", the valuation after rescaling converges under some weak conditions to the solution of a BSDE when $N \to \infty$. The main idea of this proof shall only be sketched here. It consists of five steps:

1. The first step is to show – by use of predictable representation theory for Bernoulli random walks – that $\phi_{t_{i+1}}^N(X^N)$ can be represented as

$$
\phi_{t_{i+1}}^N(X^N) = \beta_{t_i} + \gamma_{t_i}\Delta R_{t_{i+1}}^N + \hat{\gamma}_{t_i}\Delta\hat{R}_{t_{i+1}}^N,
$$

   where $\hat{R}^N$ is a $2^d - d - 1$-dimensional Bernoulli random walk with pairwise independent components and independent increments which is orthogonal to R^N, i.e. $(R^N\hat{R}^N)$ is a martingale (constructed like R^N with translated and dilated Bernoulli r.v.s $(\hat{B}_j^{N,l})_{j=1,\dots,k(N);l=1,\dots,2^d-d-1}$).

2. In a second step one uses the recursive property (15) and the $\mathcal{F}_t$-cash-invariance of $\phi_{t_i,t_{i+1}}^N$ (this follows from $\mathcal{F}_t$-cash-invariance of $F_{t_i}^N$ which is preserved under rescaling) to show that

$$
\begin{aligned}
\Delta\phi_{t_{i+1}}^N(X^N) &= \phi_{t_{i+1}}^N(X^N) - \phi_{t_i}^N(X^N) \\
&= -\phi_{t_i,t_{i+1}}^N(\gamma_{t_i}\Delta R_{t_{i+1}}^N + \hat{\gamma}_{t_i}\Delta\hat{R}_{t_{i+1}}^N) + \gamma_{t_i}\Delta R_{t_{i+1}}^N + \hat{\gamma}_{t_i}\Delta\hat{R}_{t_{i+1}}^N.
\end{aligned}
$$

3. Then one defines for $z_1 \in \mathbb{R}^d$ and $z_2 \in \mathbb{R}^{2^d-d-1}$

$$
\begin{aligned}
g^N(t_i, z_1, z_2) &:= -\frac{1}{\Delta t_{i+1}} \phi^N_{t_i,t_{i+1}}(z_1 \Delta R^N_{t_{i+1}} + z_2 \Delta \hat{R}^N_{t_{i+1}}) \\
&= -F^N_{t_i}\left(\frac{z_1 \Delta R^N_{t_{i+1}} + z_2 \Delta \hat{R}^N_{t_{i+1}}}{\sqrt{\Delta t_{i+1}}}\right) \qquad (17) \\
&= -F^N_{t_i}(z_1 B^N_{i+1} + z_2 \hat{B}^N_{i+1}).
\end{aligned}
$$

4. Plugging this in and using a telescope sum argument one can show in a third step that ϕ^N fulfills the backwards stochastic *difference* equation (BSΔE)

$$
\begin{aligned}
\phi^N_{t_i}(X^N) = X^N - &\sum_{j=i}^{k(N)-1} g^N(t_j, \gamma^N_{t_j}, \hat{\gamma}^N_{t_j})\Delta t^N_{j+1} \\
&- \sum_{j=i}^{k(N)-1} (\gamma_{t_j} \Delta R^N_{t_{j+1}} + \hat{\gamma}_{t_j} \Delta \hat{R}^N_{t_{j+1}}).
\end{aligned}
$$

5. Finally it is shown that if $X^N \to X$ in L^2 and if either $g^N(t_i, z_1, z_2)$ is deterministic and does not depend on N or $g^N(t, z_1, 0)$ converges uniformly in L^2 to a (random) function $g(t, z_1)$, then ϕ^N converges uniformly in L^2 to the solution of the BSDE

$$
Y(t) = X - \int_t^T g(s, Z(s))\mathrm{d}s - \int_t^T Z(s)\mathrm{d}W(s),
$$

which is therefore defined to be the continuous time valuation $\phi_t(X) := Y(t)$.

That is a powerful result, because under weak conditions on the *driver* g, solutions of BSDEs automatically fulfill all properties of a valuation (see appendix or, e.g. [11] for more information on that topic).

This offers a possibility to approximate a (time-consistent) dynamic risk measure (actually a valuation) in continuous time if the filtration is driven by a d-dimensional Brownian motion W and one has constructed a series X^N of r.v.s such that $X^N \to X$ in L^2 as $N \to \infty$:

1. Approximate the $\mathcal{F}_T$-measurable risk X by the $\mathcal{F}^N_T$-measurable random variable X^N (remember that $(\mathcal{F}_t)_{t\in[0,T]}$ is generated by W and $(\mathcal{F}^N_t)_{t\in[0,T]}$ by R^N),

2. Use (15) to define the N-th approximation of the continuous valuation,

3. Repeat this for several N and stop if some breaking criterion fulfilled.

Unfortunately this algorithm is very extensive on calculation power at least if it is implemented this way. Since the algorithm works recursively and there are 2^d possible combinations of the random variables $B^{N,1}_i, ..., B^{N,d}_i$, the algorithm has to call itself 2^d times at the i-th timestep, which means that for one fixed N, the algorithm needs a number of operations which is in the range of $2^{k(N)d}$. And since a Brownian motion must be approximated by a Binomial random walk, it might be a large number of timesteps $k(N)$ needed.

Therefore, in this work a different approach is used. Theoretically Equation (17) gives us the possibility to write the driver in explicit form for any valuation scheme that works in discrete time (it is not said that this is easy). However, in [2] for some particular risk valuation schemes (one-period valuations) the driver is explicitely given. This offers the opportunity to calculate the continuous dynamical valuation directly by solving the BSDE. However it is very unusual that a BSDE can be solved analytically, but there are several numerical schemes to solve a BSDE approximately. One of these schemes which is a generalization of the Picard iteration and should be very efficient in computation time is the Bender-Denk Algorithm which can be found in the appendix and was introduced and proven in [10]. It is the one which will be used later in an example.

An interesting question is in how far the rescaled discrete risk measure and the continuous dynamic risk measure which comes from the solution of the limiting BSDE correspond to the (static) risk measures we started from (i.e. with the actual problem we wanted to solve).

The next section gives an example of how to use the dynamic risk measures just derived.

4 Risk Valuation of Hedging Cost in an Incomplete Market

In this chapter we apply locally risk minimizing (cross-)hedging to a Futures contract in an incomplete market. First we develop the locally risk minimizing strategy, secondly we calculate the dynamic $AVaR$ that corresponds to the hedging cost process.

4.1 Problem and Market Model

Consider a multivariate Black-Scholes type market consisting of one riskless asset S_0, one non-tradable asset S_N (for example wheat) and one tradable asset S_1 (for example stocks of a big agriculture company):

$$\begin{cases} \mathrm{d}S_0(t) & = rS_0(t)\mathrm{d}t, \ S_0(0) = 1, \\ \mathrm{d}S_N(t) & = \mu_N S_N(t)\mathrm{d}t + \sigma_N S_N(t)\mathrm{d}W_1(t), \\ \mathrm{d}S_1(t) & = \mu_1 S_1(t)\mathrm{d}t + \sigma_{11} S_1(t)\mathrm{d}W_1(t) + \sigma_{12} S_1(t)\mathrm{d}W_2(t). \end{cases} \tag{18}$$

The associated *discounted* market model would be (*where we write S instead of $\tilde{S}$ for the discounted values since we will only use this model in the following*)

$$\begin{cases} S_0(t) & = 1, \ \forall t \in [0, T] \\ \mathrm{d}S_N(t) & = (\mu_N - r)S_N(t)\mathrm{d}t + \sigma_N S_N(t)\mathrm{d}W_1(t), \\ \mathrm{d}S_1(t) & = (\mu_1 - r)S_1(t)\mathrm{d}t + \sigma_{11} S_1(t)\mathrm{d}W_1(t) + \sigma_{12} S_1(t)\mathrm{d}W_2(t). \end{cases} \tag{19}$$

We will consider a Futures contract F on S_N with strike K and maturity T and want to hedge it only using the riskless bank account S_0 and the tradable asset S_1. The discounted value $\tilde{F}$ of F at time T (i.e. $e^{-rT}F$) is given by

$$\tilde{F} = S_N(T) - e^{-rT}K. \tag{20}$$

Since this market is *incomplete*, it will not be possible to hedge the Futures perfectly, but there will be either a *hedging error* (e.g. minimum variance hedging) or a *cost process* (e.g. locally risk minimizing hedging). We decide for the latter version. We want to measure the risk which is caused by the hedging cost.

4.2 Locally Risk Minimizing Hedge

In this section we follow the papers of Föllmer and Schweizer (1990) [7] and Schweizer (2001) [8], more information can be found in the appendix.

Stucture Condition and Minimal Martingale Measure

From the market model we know that S_1 is a local square integrable semimartingale, i.e. S_1 can be decomposed into

$$S_1 = S_1(0) + M + A$$

where M is a local square integrable martingale and A is a predictable process of bounded variation. M and A can be represented explicitely:

$$\begin{cases} M(s) & = \int_0^s \sigma_{11} S_1(t) \mathrm{d}W_1(t) + \int_0^s \sigma_{12} S_1(t) \mathrm{d}W_2(t), \\ A(s) & = \int_0^s (\mu_1 - r) S_1(t) \mathrm{d}t. \end{cases}$$

Following [8] we now represent A as

$$\begin{aligned} A(s) &= \int_0^s \hat{\lambda}(t) \mathrm{d}\langle M, M\rangle_t \\ &= \int_0^s \hat{\lambda}(t)[\sigma_{11}^2 S_1^2(t) + \sigma_{12}^2 S_1^2(t)]\mathrm{d}t \\ &\Rightarrow \hat{\lambda}(t) = \frac{\mu_1 - r}{S_1(t)(\sigma_{11}^2 + \sigma_{12}^2)}. \end{aligned}$$

Therefore,

$$\begin{aligned} \hat{K}(s) &:= \int_0^s \hat{\lambda}^2(t) \mathrm{d}\langle M, M\rangle_t \\ &= \int_0^s \frac{(\mu_1 - r)^2}{S_1^2(t)(\sigma_{11}^2 + \sigma_{12}^2)^2} S_1^2(t)(\sigma_{11}^2 + \sigma_{12}^2)\mathrm{d}t \\ &= \frac{s(\mu_1 - r)^2}{\sigma_{11}^2 + \sigma_{12}^2} < \infty, \end{aligned}$$

i.e. the *structure condition (SC)* from [8] is fulfilled, see also Definition A.6 in the appendix. Now define the minimal martingale measure $\hat{P}$ by $\frac{\mathrm{d}\hat{P}}{\mathrm{d}P} = \hat{Z}(T)$ where $\hat{Z}$ is given by

$$\mathrm{d}\hat{Z}(t) = -\hat{Z}(t)\hat{\lambda}(t)\mathrm{d}M(t),$$

i.e.

$$\begin{aligned} \hat{Z}(t) &= \exp\left(- \int_0^t \hat{\lambda}(u) \mathrm{d}M(u) - \frac{1}{2} \underbrace{\langle \int \hat{\lambda} \mathrm{d}M \rangle_t}_{=\hat{K}(t)} \right) \\ &= \exp\Big(- \int_0^t \frac{(\mu_1 - r)\sigma_{11}}{\sigma_{11}^2 + \sigma_{12}^2} \mathrm{d}W_1(u) - \int_0^t \frac{(\mu_1 - r)\sigma_{22}}{\sigma_{11}^2 + \sigma_{12}^2} \mathrm{d}W_2(u) \\ &\qquad - \frac{1}{2} \int_0^t \frac{(\mu_1 - r)^2}{\sigma_{11}^2 + \sigma_{12}^2} \mathrm{d}u \Big) \\ &= \exp\left(- \int_0^t \psi^t \mathrm{d}W(u) - \frac{1}{2} \int_0^t |\psi|^2 \mathrm{d}u \right) \end{aligned}$$

where $\psi = \begin{pmatrix} \frac{(\mu_1 - r)\sigma_{11}}{\sigma_{11}^2 + \sigma_{12}^2} \\ \frac{(\mu_1 - r)\sigma_{22}}{\sigma_{11}^2 + \sigma_{12}^2} \end{pmatrix}$, $dW = \begin{pmatrix} dW_1 \\ dW_2 \end{pmatrix}$.

By the multivariate Girsanov Theorem, we know that S_1 is a martingale under $\hat{P}$ (so $\hat{P}$ is one possible pricing measure) and

$$\hat{W}(t) = W(t) + \int_0^t \psi(u)\mathrm{d}u$$

defines a 2-dimensional $\hat{P}$-Brownian motion. Furthermore, we know that all martingales under P which are orthogonal to M (the P-martingale part of S_1) are martingales w.r.t $\hat{P}$.

The Pseudo-Optimal Strategy

Schweizer [8] defines a *pseudo-optimal strategy* as an L^2-trading strategy $\varphi = (\varphi_0, \varphi_1)$ on S_0 and S_1, which fulfills $V(\varphi, T) := V^{\tilde{F},\hat{P}}(\varphi, T) = \tilde{F}$ and φ is *mean-self-financing* and the martingale of costs $C = C(\varphi)$ is strongly orthogonal to M (i.e. CM is a martingale). It is shown in [8] Thm. 3.5 and can be referred in Theorem A.7 in the appendix, that then

$$V(\varphi, t) = E_{\hat{P}}[\tilde{F}|\mathcal{F}_t]. \tag{21}$$

Let D_1 and D_2 be the Malliavin derivatives w.r.t. W_1 and W_2, respectively (compare Remark A.1 in the appendix). Since ψ is deterministic, we have $D_{1,t}\psi = D_{2,t}\psi = 0\ \forall t \in [0, T]$. And because $\tilde{F}$ only depends on W_1 (via S_N), it holds that $D_{2,t}\tilde{F} = 0\ \forall t \in [0, T]$. Thus, the Clark-Ocone representation of $E_{\hat{P}}[\tilde{F}|\mathcal{F}_t]$ (under change of measure from P to $\hat{P}$):

$$E_{\hat{P}}[\tilde{F}|\mathcal{F}_t] = E_{\hat{P}}[\tilde{F}] + \int_0^t E_{\hat{P}}[D_{1,s}\tilde{F}|\mathcal{F}_s]\mathrm{d}\hat{W}_1(s). \tag{22}$$

Step 1 - computing $E_{\hat{P}}[\tilde{F}]$: We have $E_{\hat{P}}[\tilde{F}] = E_{\hat{P}}[S_N(T)] - e^{-rT}K$ and

$$\begin{aligned} E_{\hat{P}}[S_N(T)] &= S_N(0)\exp([(\mu_N - r) - \frac{1}{2}\sigma_N^2]T)E_{\hat{P}}[\exp(\sigma_N W_1(T))] \\ &= S_N(0)\exp([(\mu_N - r) - \frac{1}{2}\sigma_N^2]T) \\ &\quad \times E_{\hat{P}}[\exp(\sigma_N(\underbrace{\hat{W}_1(T)}_{\sim\mathcal{N}(0,T)} - \underbrace{\int_0^T \psi_1(u)\mathrm{d}u}_{=\psi_1 T \text{ determ.}}))] \\ &= S_N(0)\exp([(\mu_N - r) - \sigma_N\psi_1]T), \end{aligned}$$

so we get

$$
\begin{aligned}
E_{\hat{P}}[\tilde{F}] &= S_N(0)\exp([(\mu_N - r) - \sigma_N\psi_1]T) - \exp(-rT)K \\
&= S_N(0)\exp([(\mu_N - r) - \frac{\sigma_{11}\sigma_N}{\sigma_{11}^2 + \sigma_{12}^2}(\mu_1 - r)]T) - \exp(-rT)K.
\end{aligned}
$$

Step 2 - computing $E_{\hat{P}}[D_{1,s}\tilde{F}|\mathcal{F}_s]$**:** Because $\tilde{F} = S_N(T) - e^{-rT}K$ and $e^{-rT}K$ is deterministic, we get $D_{1,s}\tilde{F} = D_{1,s}S_N(T)$. Furthermore, $S_N(T) = S_N(0)\exp([(\mu_N - r) - \frac{1}{2}\sigma_N^2]T + \sigma_N W_1(T))$, therefore, by the chain rule

$$
\begin{aligned}
D_{1,s}\tilde{F} &= \underbrace{S_N(0)\exp([(\mu_N - r) - \frac{1}{2}\sigma_N^2]T + \sigma_N W_1(T))}_{=S_N(T)} \\
&\quad \times \underbrace{D_{1,s}([(\mu_N - r) - \frac{1}{2}\sigma_N^2]T + \sigma_N W_1(T))}_{=\sigma_N} \\
&= \sigma_N S_N(T).
\end{aligned}
$$

Thus

$$
\begin{aligned}
E_{\hat{P}}[D_{1,s}\tilde{F}|\mathcal{F}_s] &= E_{\hat{P}}[\sigma_N S_N(T)|\mathcal{F}_s] \\
&= \sigma_N S_N(0)\exp([(\mu_N - r) - \frac{1}{2}\sigma_N^2]T)E_{\hat{P}}[\exp(\sigma_N W_1(T))|\mathcal{F}_s] \\
&= \sigma_N S_N(0)\exp([(\mu_N - r) - \frac{1}{2}\sigma_N^2]T) \\
&\quad \times E_{\hat{P}}[\exp(\sigma_N(\hat{W}_1(T) - \underbrace{\int_0^T \psi_1(u)\mathrm{d}u}_{=\psi_1 T,\ \text{determ.}}))|\mathcal{F}_s] \\
&= \sigma_N S_N(0)\exp([(\mu_N - r) - \frac{1}{2}\sigma_N^2 - \sigma_N\psi_1]T)\exp(\sigma_N\hat{W}_1(s)) \\
&\quad \times E_{\hat{P}}[\exp(\sigma_N(\underbrace{\hat{W}_1(T) - \hat{W}_1(s)}_{\sim\mathcal{N}(0,T-s),\ \text{ind. of } \mathcal{F}_s}))|\mathcal{F}_s] \\
&= \sigma_N S_N(s)\exp([(\mu_N - r) - \frac{1}{2}\sigma_N^2 - \sigma_N\psi_1](T - s)) \\
&\quad \times \exp(\frac{1}{2}\sigma_N^2(T - s)) \\
&= \sigma_N S_N(s)\exp([(\mu_N - r) - \sigma_N\psi_1](T - s)).
\end{aligned}
$$

Inserting this result into (22) yields

$$
\begin{aligned}
E_{\hat{P}}[\tilde{F}|\mathcal{F}_t] &= E_{\hat{P}}[\tilde{F}] + \int_0^t \underbrace{\sigma_N S_N(s)\exp([(\mu_N - r) - \sigma_N\psi_1](T - s))}_{=:Y(s)}\mathrm{d}\hat{W}_1(s) \\
&\overset{!}{=} V^{\tilde{F},\hat{P}}(\varphi, t),
\end{aligned} \tag{23}
$$

or in stochastic dynamic terms:

$$\begin{cases} \mathrm{d}V(\varphi,t) &= Y(t)\mathrm{d}\hat{W}_1(t), \\ V(\varphi,0) &= E_{\hat{P}}[\tilde{F}]. \end{cases} \tag{24}$$

But due to Theorem A.7 in the appendix (compare [8]), the dynamics of V should be

$$\mathrm{d}V(\varphi,t) = \varphi_1(t)\mathrm{d}S_1(t) + \mathrm{d}L(t), \tag{25}$$

where L is a martingale strongly orthogonal to S_1, i.e. $\langle L, S_1\rangle = 0$. Therefore we try the attempt

$$\begin{aligned} \mathrm{d}V(\varphi,t) &= Y(t)\mathrm{d}\hat{W}_1(t) \\ &= aY(t)\mathrm{d}S_1(t) + bY(t)\mathrm{d}\hat{W}_1(t) + cY(t)\mathrm{d}\hat{W}_2(t). \end{aligned}$$

And since we know that the dynamic of S_1 is

$$\begin{aligned} \mathrm{d}S_1(t) &= (\mu_1 - r)S_1(t)\mathrm{d}t + \sigma_{11}S_1(t)\mathrm{d}W_1(t) + \sigma_{12}S_1(t)\mathrm{d}W_2(t) \\ &= \sigma_{11}S_1(t)\mathrm{d}\hat{W}_1(t) + \sigma_{12}S_1(t)\mathrm{d}\hat{W}_2(t) \end{aligned}$$

(S_1 is a martingale under $\hat{P}$), we get

$$\begin{aligned} Y(t)\mathrm{d}\hat{W}_1(t) = {} & aY(t)\mathrm{d}(\sigma_{11}S_1(t)\mathrm{d}\hat{W}_1(t) + \sigma_{12}S_1(t)\mathrm{d}\hat{W}_2(t)) \\ & + bY(t)\mathrm{d}\hat{W}_1(t) + cY(t)\mathrm{d}\hat{W}_2(t). \end{aligned}$$

Comparison of coefficients leads to the following system of equations:

$$\begin{cases} a\sigma_{11}S_1(t) + b &= 1 \\ a\sigma_{12}S_1(t) + c &= 0 \\ b\sigma_{11} + c\sigma_{12} &= 0 \end{cases}$$

(the third equation is due to the orthogonality condition). This system of equations has the solution

$$\begin{cases} a &= \frac{\sigma_{11}}{S_1(t)(\sigma_{11}^2+\sigma_{12}^2)} \\ b &= \frac{\sigma_{12}^2}{\sigma_{11}^2+\sigma_{12}^2} \\ c &= -\frac{\sigma_{11}\sigma_{12}}{\sigma_{11}^2+\sigma_{12}^2}. \end{cases}$$

Inserting this result yields

$$
\begin{aligned}
\mathrm{d}V(\varphi,t) = &\frac{\sigma_{11}}{S_1(t)(\sigma_{11}^2+\sigma_{12}^2)}Y(t)\mathrm{d}S_1(t) \\
&+ \underbrace{\frac{\sigma_{12}^2}{\sigma_{11}^2+\sigma_{12}^2}Y(t)\mathrm{d}\hat{W}_1(t) - \frac{\sigma_{11}\sigma_{12}}{\sigma_{11}^2+\sigma_{12}^2}Y(t)\mathrm{d}\hat{W}_2(t)}_{=\mathrm{d}L(t)}.
\end{aligned}
$$

So in summary, inserting the definition of Y, we get

$$
\begin{cases}
V^{\tilde{F},\hat{P}}(\varphi,T) & = \tilde{F}, \\
V^{\tilde{F},\hat{P}}(\varphi,0) & = E_{\hat{P}}[\tilde{F}] \\
& = S_N(0)\exp([(\mu_N - r) - \frac{\sigma_{11}\sigma_N}{\sigma_{11}^2+\sigma_{12}^2}(\mu_1 - r)]T) - e^{-rT}K, \\
V^{\tilde{F},\hat{P}}(\varphi,t) & = V^{\tilde{F},\hat{P}}(\varphi,0) + \int_0^t \varphi_1(s)\mathrm{d}S_1(s) + L(t), \text{ where} \\
\varphi_1(s) & = \frac{\sigma_{11}\sigma_N}{\sigma_{11}^2+\sigma_{12}^2}\frac{S_N(s)}{S_1(s)}\exp([(\mu_N - r) - \sigma_N\psi_1](T-s)).
\end{cases} \tag{26}
$$

This is the *Kunita-Watanabe decomposition* of $\tilde{F}$ under $\hat{P}$, since S_1 and L are $\hat{P}$-martingales with $\langle L, S_1\rangle = 0$. Since L is also a P-martingale, which can be shown very easily,

$$
\begin{aligned}
L(t) &= \int_0^t \frac{\sigma_{12}^2}{\sigma_{11}^2+\sigma_{12}^2}Y(s)\underbrace{\mathrm{d}\hat{W}_1(s)}_{=\mathrm{d}W_1(s)+\psi_1\mathrm{d}s} - \int_0^t \frac{\sigma_{11}\sigma_{12}}{\sigma_{11}^2+\sigma_{12}^2}Y(s)\underbrace{\mathrm{d}\hat{W}_2(s)}_{=\mathrm{d}W_2(s)+\psi_2\mathrm{d}s} \\
&= \underbrace{\int_0^t (\frac{\sigma_{12}^2}{\sigma_{11}^2+\sigma_{12}^2}\psi_1 - \frac{\sigma_{11}\sigma_{12}}{\sigma_{11}^2+\sigma_{12}^2}\frac{\sigma_{12}}{\sigma_{11}}\psi_1)Y(s)\mathrm{d}s}_{=0} \\
&\quad + \int_0^t \frac{\sigma_{12}^2}{\sigma_{11}^2+\sigma_{12}^2}Y(s)\mathrm{d}W_1(s) - \int_0^t \frac{\sigma_{11}\sigma_{12}}{\sigma_{11}^2+\sigma_{12}^2}Y(s)\mathrm{d}W_2(s),
\end{aligned}
$$

(26) is also the *Föllmer-Schweizer decomposition* of $\tilde{F}$ under P.

4.3 Numerical Simulation of the Hedging Risk under $AVaR$

Now we want to measure the risk due to the hedging cost C or rather due to L (because $C = V(\varphi,0)+L$ and $V(\varphi,0)$ is fix and can be passed immediately to the customers as part of the Futures' price) via a dynamical $AVaR$-scheme. We use this particular scheme because in the static setting $AVaR$ is a recommended risk measure since it is coherent and focusses on the tail risk, i.e. the risk of extreme losses. Following Stadje (2010) [2] the one-period valuations or "generators" are given by

$$
F_{t_i}(X) = AVaRV_{t_i,\lambda,\alpha}(X) := E[X|\mathcal{F}_{t_i}] - \lambda AVaR_{t_i,\alpha}(X - E[X|\mathcal{F}_{t_i}]),
$$

for $X \in L^\infty(\mathcal{F}_{t_{i+1}})$, where

$$
AVaR_{t_i,\alpha}(X) := \frac{1}{\alpha}\int_0^\alpha VaR_{t_i,\gamma}(X)\mathrm{d}\gamma,
$$

and

$$VaR_{t_i,\alpha}(X) := \text{essinf}\{m : m\ \mathcal{F}_{t_i} - \text{msble.}, P(X + m < 0|\mathcal{F}_{t_i}) \leq \alpha\}.$$

We choose $\lambda = 1$, i.e. the one-period valuations are just the usual $AVaR_\alpha$ on one period. The associated dynamic $AVaR_\alpha$ in continuous time is given by $AVaR_{t,\alpha}(X) = -Y(t)$, where (Y, Z), $Z = (Z_1, Z_2)$ is the solution to the BSDE

$$Y(t) = X - \int_t^T g(Z(s))\mathrm{d}s - \int_t^T Z_1(s)\mathrm{d}W_1(s) - \int_t^T Z_2(s)\mathrm{d}W_2(s). \tag{27}$$

The driver is (see, for example [2])

$$g(z) = -\frac{1}{\alpha}\left(\chi_{5-\lceil 4\alpha\rceil}(z)\left(\alpha - \frac{\lceil 4\alpha\rceil - 1}{4}\right) + \frac{1}{4}\sum_{j=1}^{\lceil 4\alpha\rceil - 1}\chi_{5-j}(z)\right), \tag{28}$$

and $\chi_p(z)$ the p-th largest value of the set $\{z_1+z_2, z_1-z_2, -z_1+z_2, -z_1-z_2\}$.

Since L is a loss process, i.e. positive values of L are our loss and negative ones are our gains, but we usually consider gains processes (the other way around) we will consider the process $-L$ in the following. Furthermore we make the following **simplifying assumption**: We have infinite financial resources outside the portfolio (bank account plus tradable asset) to pay the costs up to expiry of the Futures contract and we only have to cover the losses at the end. This could for example be achieved by an account at the central bank where we can borrow as much money as we need for the riskless rate r (so that the bank account can take any negative or positive value) until expiry – we just have to settle the account at the end. This means, that the risk is determined by $X = -L(T)$.

In the following we consider the above problem under concrete parameters: $T = 1$, $r = 0.01$, $S_N(0) = 100$ EUR, $S_1(0) = 50$ EUR, $\mu_N = 0.05$, $\mu_1 = 0.04$, $\sigma_N = 0.25$, $\sigma_{11} = 0.18$, $\sigma_{12} = 0.12$ and $K = 100$ EUR. The number of timesteps used for simulation was 100. Figure 1 shows a path of S_N and S_1 (upper left) as well as the Clark-Ocone decomposition (22) (upper right) and the hedging strategy $V^{\tilde{F},\hat{P}}$ (middle left), both in comparison with the Futures value process. Furthermore it shows the hedging strategy without the cost part ($V^{\tilde{F},\hat{P}} - L$), also in comparison with the Futures value process (middle right) to see if the good fitting is only achieved by the cost component (the permanent adjustment part of the portfolio). This would of course indicate that we had done something wrong, but it looks quite okay. Eventually, Figure 1 includes 20 paths of the cost process L (bottom left) plus 20 paths of the unhedged Futures process (bottom right). One can observe that the Clark-Ocone decomposition and the hedging strategy seem to work right (the quadratic deviation of the Clark-Ocone decomposition at T from $\tilde{F}$ is about 0.21 EUR^2 and the quadratic deviation of $V^{\tilde{F},\hat{P}}(T)$ from $\tilde{F}$ is about 0.11 EUR^2). Also it can be observed that the variance of $L(t)$ at any $t > 0$ is far less than the one of the unhedged Futures process at time t (at time T, the variance of L is about 209 EUR^2, whereas the one of the unhedged Futures is about 711 EUR^2). If that was not the case, it would clearly indecate a mistake in our considerations. This example was calculated in Matlab and the seed which was used was 100, from the generator "mt19937ar".

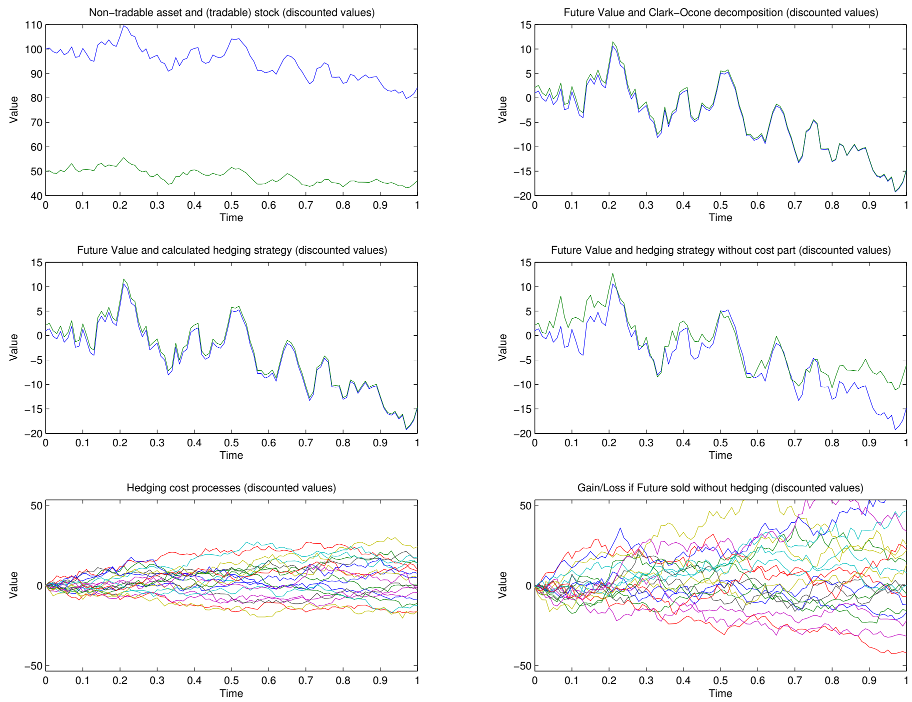

Figure 1: top left: S_N (blue) and S_1 (green), top right: Futures value evolution (blue) and Clark-Ocone decomposition of $\tilde{F}$ (green), middle left: Futures value evolution (blue) and $V^{\tilde{F},\hat{P}}$ (green), middle right: Futures value evolution (blue) and hedging strategy without the cost part $(V^{\tilde{F},\hat{P}} - L)$ (green), bottom left: 20 plots of the hedging cost process L, bottom right: 20 plots of the Futures value without hedging

In a second step the dynamic $AVaR$ was calculated for $\alpha = 0.1$ using the algorithm of Bender and Denk [10]. The number of copies used for this simulation was 100000, the number of time steps was 50, the projection basis was $\eta_k^i := e_k(-L(t_i))$, $k = 1, ..., 6$,

$$e_k(x) = x^{k-1}, \; k = 1, ..., 6,$$

and the breaking criterion was $Y^{(n+1)}(0) - Y^{(n)}(0) < 0.001$. For reason of computational time the maximal number of iterations was set to $n_{Stop} = 6$. Figure 2 shows on the top the hedging cost process L and the calculated dynamic $AVaR$ (which is $-\hat{Y}$ with $(\hat{Y}, \hat{Z})$ numerical solution of BSDE (27) with terminal condition $X = -L(T)$) at the level $\alpha = 0.1$ – it is around 23.44 EUR. One can observe that the dynamic $AVaR$ increases whenever L increases and decreases whenever L decreases, which is clearly what we would expect (remember that we measure the risk of $-L(T)$). Furthermore, as time goes by, the difference between the dynamic $AVaR$ and the process L decreases and becomes eventually 0. This is very intuitive, since the less time remains until expiry, the less can happen. So the risk has some "time-value" that decreases in time (or rather increases in time remaining).

On the bottom, Figure 2 shows the first component $\hat{Y}$ of the calculated numerical solution $(\hat{Y}, \hat{Z})$ of BSDE (27) together with a "forward solution" which is calculated as follows: starting at $\hat{Y}(0)$, use the BSDE dynamics forward in time (by applying a simple Euler-scheme), plugging in $\hat{Z}$, which is also estimated in the Bender-Denk algorithm. If $\hat{Y}$ would be the true solution of the BSDE, both graphs would match perfectly. Since $\hat{Y}$ is only a numerical solution, the graphs do not match, but are close which is an indication that the algorithm works well and $\hat{Y}$ is close to the true solution. Additionally, the static $AVaR$ at level $\alpha = 0.1$ was estimated and its value is around 27.65 EUR. Again this shows that the static $AVaR$ is not its dynamic version, evaluated at 0, but it somehow "has something to do with it".

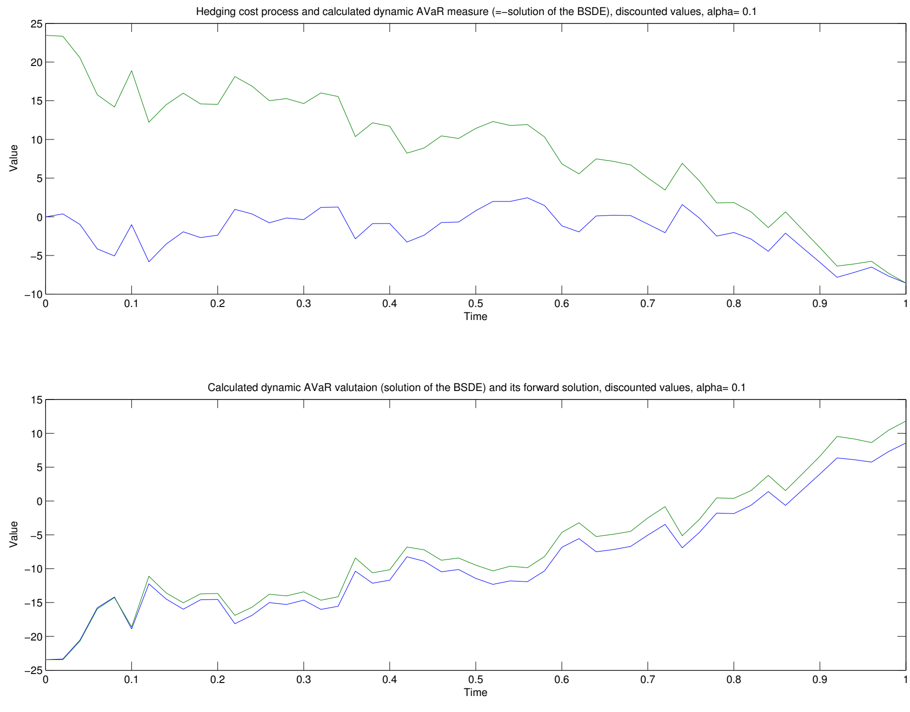

Figure 2: top: Hedging cost process (blue) and calculated dynamic $AVaR_{0.1}$ (green), bottom: calculated dynamic $AVaR_{0.1}$ (blue) and calculated forward solution (green)

Eventually, a small statistical analysis of this numerical solution was done. To do so, 29 more versions (seeds 101 to 129) of the numerical BSDE solution were computed (every version with its own 100000 copies). The computed time-zero values of the BSDE solution, $\hat{Y}(0)$, have empirical mean -23.37 EUR and empirical variance 0.91 EUR^2. The difference between the time-T values of the BSDEs, $\hat{Y}(T)$, and their corresponding forward solutions have empirical mean 1.10 EUR and variance 52.60 EUR^2. Histograms of both values can be found in Figure 4. There is one outlier, where the absolute difference of numerical BSDE solution and forward solution at time T is larger than 30 whereas in all other cases it is smaller than 10. This is the version with seed 118. Without the outlier, the empirical mean and empirical variance of $\hat{Y}(0)$ are -23.29 EUR and 0.72 EUR^2 and the empirical mean and variance of the difference between $\hat{Y}(T)$ and its forward solution at time T are -0.09 EUR and 10.53 EUR^2. Moreover, four more versions of what can be seen in Figure 2 can be found in Figure 5 in the appendix. The seeds were 101, 102, 103 and 118 – the last one was to show what can go wrong.

Now let us return to the example from the introduction, where the bank or the regulator defines a boundary that should not be hitten by the risk process at any time (otherwise the position must be cancelled, for example). Now the bank could define a maximal probability of hitting the boundary it is willing to accept, say 0.05. To see if the position is acceptable in this context, the bank would now simulate a sufficiently large amount of paths. If more than 5% of the simulated paths are hitting the boundary, the contract will not be closed or an additional amount of money will be demanded, s.t. less than 5% of the paths hit the boundary afterwards. But one has to be careful, paths like the one achieved with seed 118 (at the bottom of Figure 5) must be excluded (or the maximal number of time steps in the BSDE solving algorithm must be increased for those paths), since they do not fit the solution of the BSDE well. Therefore a comparison with the forward solution of the BSDE like done here is strongly recommended.

5 The 1-dimensional Case

This chapter deals with the results that were just developed in the 1-dimensional case. We show that the limiting BSDE is of a special type and that for pathwise increasing claims which fulfill some weak additional conditions, the dynamic risk measure is given by a usual conditional expectation w.r.t. some measure that depends only on the distortion. Then we check our results on an example.

5.1 Investigation of the Limiting BSDE

In this section we take a closer look at the BSDE which appears in the limit of the rescaled distorted valuations. Recall the setting from before where W is a Brownian motion generating a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ and let again $((R^N)_N, ((\mathcal{F}_t^N)_{t\in[0,T]})_N)$ be a sequence of Bernoulli random walks approximating W (as $N \to \infty$) and their canonical filtrations. But now let $d = 1$, i.e. W and R^N are one-dimensional processes.

We will show that in this particular case, the driver of the BSDE which appears in the limit of the distorted valuations is of a quite simple type. Furthermore it will be shown that if the claim to be evaluated is pathwise increasing and some auxiliary conditions are fulfilled, the valuation (and thus also the dynamic risk measure) is just a conditional expectation with respect to a certain measure which depends only on the distortion. The following theorem illustrates the general type of the limiting BSDE in one dimension:

Theorem 5.1. *Recall the Bernoulli random walk approximation from before, but now let $d = 1$ and let X be an $\mathcal{F}_T$-measurable random variable. Then, as $N \to \infty$, the dynamic valuation scheme $(\phi_{t_i}^N)_{i=0,\ldots,k(N)}$ converges to the solution of the BSDE*

$$\phi_t(X) = X - \int_t^T 2\left[\Psi\left(\frac{1}{2}\right) - \frac{1}{2}\right]|Z(s)|ds - \int_t^T Z(s)dW(s). \qquad (29)$$

Remark 5.1. *This implies that the corresponding* ***risk measure*** *$\rho_t(X) = -\phi_t(X)$ solves the BSDE*

$$\begin{aligned}\rho_t(X) &= -X + \int_t^T 2\left[\Psi\left(\frac{1}{2}\right) - \frac{1}{2}\right]|Z(s)|ds + \int_t^T Z(s)dW(s)\\ &= -X - \int_t^T 2\left[\frac{1}{2} - \Psi\left(\frac{1}{2}\right)\right]|\tilde{Z}(s)|ds - \int_t^T \tilde{Z}(s)dW(s),\end{aligned}$$

where $\tilde{Z} = -Z$.

Proof:
By definition of the one-period valuations F_{t_i} (8), we have

$$F_{t_i}^N(X) = \text{essinf}_{\xi_{t_{i+1}}\in D_{t_{i+1}}^N, Q^\xi(A|\mathcal{F}_{t_i}^N)\le\Psi(P(A|\mathcal{F}_{t_i}^N))\forall A\in\mathcal{F}_{t_{i+1}}^N}\; E_{Q^\xi}[X|\mathcal{F}_{t_i}^N].$$

Furthermore using (17), the construction of the BSDE driver from before, and (since $d = 1$) $z := z_1 \in \mathbb{R}^d = \mathbb{R}$ and $z_2 \in \mathbb{R}^{2-1-1} = \emptyset$ we get

$$
\begin{aligned}
g^N(t_i, z) &= -F^N_{t_i}(zB^N_{i+1}) \\
&= -\operatorname{essinf}_{\xi_{t_{i+1}} \in D^N_{t_{i+1}}, Q^\xi(A|\mathcal{F}^N_{t_i}) \leq \Psi(P(A|\mathcal{F}^N_{t_i})) \forall A \in \mathcal{F}^N_{t_{i+1}}} E_{Q^\xi}[zB^N_{i+1}|\mathcal{F}^N_{t_i}] \\
&= -\begin{cases} z \operatorname{essinf}_{Q^\xi \in \mathcal{D}^{t_i,N}} E_{Q^\xi}[B^N_{i+1}|\mathcal{F}^N_{t_i}], & \text{for } z \geq 0 \\ z \operatorname{esssup}_{Q^\xi \in \mathcal{D}^{t_i,N}} E_{Q^\xi}[B^N_{i+1}|\mathcal{F}^N_{t_i}], & \text{for } z < 0 \end{cases} \\
&= -z^+ \operatorname{essinf}_{Q^\xi \in \mathcal{D}^{t_i,N}} E_{Q^\xi}[B^N_{i+1}|\mathcal{F}^N_{t_i}] + z^- \operatorname{esssup}_{Q^\xi \in \mathcal{D}^{t_i,N}} E_{Q^\xi}[B^N_{i+1}|\mathcal{F}^N_{t_i}],
\end{aligned}
$$

where we define $\mathcal{D}^{t_i,N} := \{Q^\xi : \xi_{t_{i+1}} \in D^N_{t_{i+1}}, Q^\xi(A|\mathcal{F}^N_{t_i}) \leq \Psi(P(A|\mathcal{F}^N_{t_i})) \forall A \in \mathcal{F}^N_{t_{i+1}}\}$ for convenience (again, we mean the P-a.s. defined versions of the conditional expectations and the $P-\operatorname{essinf}$ and $P-\operatorname{esssup}$). Note that since B^N_{i+1} only takes the values -1 and 1,

$$
B^N_{i+1} + 1 = 2 \cdot \mathbb{1}_{\{B^N_{i+1}=1\}} - 0 \cdot \mathbb{1}_{\{B^N_{i+1}=-1\}} = 2 \cdot \mathbb{1}_{\{B^N_{i+1}=1\}},
$$

and thus, since $Q^\xi \in \mathcal{D}^{t_i,N}$,

$$
\begin{aligned}
E_{Q^\xi}[B^N_{i+1}|\mathcal{F}_{t_i}] &= E_{Q^\xi}[B^N_{i+1} + 1|\mathcal{F}^N_{t_i}] - 1 \\
&= 2E_{Q^\xi}[\mathbb{1}_{\{B^N_{i+1}=1\}}|\mathcal{F}^N_{t_i}] - 1 \\
&\leq 2\Psi(P(B^N_{i+1} = 1|\mathcal{F}^N_{t_i})) - 1 \\
&= 2\left(\Psi\left(\frac{1}{2}\right) - \frac{1}{2}\right).
\end{aligned} \tag{30}
$$

Hence, also

$$
\operatorname{esssup}_{Q^\xi \in \mathcal{D}^{t_i,N}} E_{Q^\xi}[B^N_{i+1}|\mathcal{F}^N_{t_i}] \leq 2\left(\Psi\left(\frac{1}{2}\right) - \frac{1}{2}\right)
$$

is true. Analogously we have

$$
B^N_{i+1} - 1 = 0 \cdot \mathbb{1}_{\{B^N_{i+1}=1\}} - 2 \cdot \mathbb{1}_{\{B^N_{i+1}=-1\}} = -2 \cdot \mathbb{1}_{\{B^N_{i+1}=-1\}},
$$

and

$$
\begin{aligned}
E_{Q^\xi}[B^N_{i+1}|\mathcal{F}_{t_i}] &= E_{Q^\xi}[B^N_{i+1} - 1|\mathcal{F}^N_{t_i}] + 1 \\
&= -2E_{Q^\xi}[\mathbb{1}_{\{B^N_{i+1}=-1\}}|\mathcal{F}^N_{t_i}] + 1 \\
&\geq -2\Psi(P(B^N_{i+1} = -1|\mathcal{F}^N_{t_i})) + 1 \\
&= -2\left(\Psi\left(\frac{1}{2}\right) - \frac{1}{2}\right)
\end{aligned} \tag{31}
$$

and also that

$$
\operatorname{essinf}_{Q^\xi \in \mathcal{D}^{t_i,N}} E_{Q^\xi}[B^N_{i+1}|\mathcal{F}^N_{t_i}] \geq -2\left(\Psi\left(\frac{1}{2}\right) - \frac{1}{2}\right).
$$

Therefore we can conclude that

$$
\begin{aligned}
g^N(t_i,z) &= -z^+ \operatorname{essinf}_{\mathcal{D}^{t_i,N}} E_{Q^\xi}[B_{i+1}^N|\mathcal{F}_{t_i}^N] \\
&\quad + z^- \operatorname{esssup}_{\mathcal{D}^{t_i,N}} E_{Q^\xi}[B_{i+1}^N|\mathcal{F}_{t_i}^N] \\
&\le -\left(-2\left(\Psi\left(\frac{1}{2}\right)-\frac{1}{2}\right)\right)z^+ + 2\left(\Psi\left(\frac{1}{2}\right)-\frac{1}{2}\right)z^- \\
&= 2\left(\Psi\left(\frac{1}{2}\right)-\frac{1}{2}\right)(z^+ + z^-) \\
&= 2\left(\Psi\left(\frac{1}{2}\right)-\frac{1}{2}\right)|z|.
\end{aligned} \tag{32}
$$

To show that this inequality is actually an equality, we just have to find measures $Q^{\xi^1} \in \mathcal{D}^{t_i,N}$ and $Q^{\xi^2} \in \mathcal{D}^{t_i,N}$ such that (30) and (31) are fulfilled with equality. This can be done by choosing $\xi_{t_0}^1 = ... = \xi_{t_i}^1 = 1 = \xi_{t_0}^2 = ... = \xi_{t_i}^2$ (i.e. $Q^{\xi^1}|\mathcal{F}_{t_i} = Q^{\xi^2}|\mathcal{F}_{t_i} = P|\mathcal{F}_{t_i}$) and

$$
\begin{aligned}
\xi_{t_{i+1}}^1 &= 2\left(1-\Psi\left(\frac{1}{2}\right)\right)\mathbb{1}_{\{B_{i+1}^N=-1\}} + 2\Psi\left(\frac{1}{2}\right)\mathbb{1}_{\{B_{i+1}^N=1\}}, \\
\xi_{t_{i+1}}^2 &= 2\Psi\left(\frac{1}{2}\right)\mathbb{1}_{\{B_{i+1}^N=-1\}} + 2\left(1-\Psi\left(\frac{1}{2}\right)\right)\mathbb{1}_{\{B_{i+1}^N=1\}},
\end{aligned}
$$

because then it can be seen immediately that $Q^{\xi^1} \in \mathcal{D}^{t_i,N}$ and $Q^{\xi^2} \in \mathcal{D}^{t_i,N}$ and

$$
\begin{aligned}
E_{Q^{\xi^1}}[B_{i+1}^N|\mathcal{F}_{t_i}^N] &= E[\xi_{t_{i+1}}^1 B_{i+1}^N|\mathcal{F}_{t_i}^N] \\
&= (-1)\frac{1}{2}2\left(1-\Psi\left(\frac{1}{2}\right)\right) + \frac{1}{2}2\Psi\left(\frac{1}{2}\right) \\
&= 2\Psi\left(\frac{1}{2}\right) - 1 \\
&= 2\left(\Psi\left(\frac{1}{2}\right)-\frac{1}{2}\right),
\end{aligned}
$$

and analogously

$$
\begin{aligned}
E_{Q^{\xi^2}}[B_{i+1}^N|\mathcal{F}_{t_i}^N] &= E[\xi_{t_{i+1}}^2 B_{i+1}^N|\mathcal{F}_{t_i}^N] \\
&= (-1)\frac{1}{2}2\Psi\left(\frac{1}{2}\right) + \frac{1}{2}2\left(1-\Psi\left(\frac{1}{2}\right)\right) \\
&= 1 - 2\Psi\left(\frac{1}{2}\right) \\
&= -2\left(\Psi\left(\frac{1}{2}\right)-\frac{1}{2}\right).
\end{aligned}
$$

Note that $\xi_{t_{i+1}}^1$ and $\xi_{t_{i+1}}^2$ really are probability densities, as they are nonnegative for $\Psi : [0,1] \to [0,1]$ concave and

$$
\begin{aligned}
E[\xi^1_{t_{i+1}}] &= \frac{1}{2}2\left(1-\Psi\left(\frac{1}{2}\right)\right)+\frac{1}{2}2\Psi\left(\frac{1}{2}\right) \\
&= 1,
\end{aligned}
$$

and analogously $E[\xi^2_{t_{i+1}}] = 1$. This shows that (32) is fulfilled with equality. Finally we note that g^N does not depend on N, i.e. $g^N(t,z) := g(t,z)$ is the driver of the limiting BSDE and everything is proven.

□

One very important point of criticism is that the driver of the BSDE depends only on Ψ evaluated at one point $(\frac{1}{2})$ so that most of the information of Ψ is actually wasted. As one can see in the proof of the previous theorem, this clearly comes from the fact that we use a Bernoulli random walk scheme for approximation and the Bernoulli random variables B^N_{i+1} only take values 1 and -1. In Madan et al. [1] a trinomial random walk approximation is considered and there, the limiting BSDE depends on $\Psi(\frac{1}{6})$ and $\Psi(\frac{5}{6})$ – still, a lot of information is lost. That is a big disadvantage of the random walk approximation approach. Therefore, in the conclusion, a different method will be proposed.

The next theorem shows that for *pathwise increasing* claims X, the BSDE can be simplified to a linear BSDE, i.e. the dynamic risk measure is a conditional expectation with respect to a particular measure (if some additional conditions are fulfilled). First of all some notation: We will consider $\Omega = C_0([0,T])$, the space of continuous functions starting in 0 at 0, representing the possible paths of the Brownian motion W, i.e. $\omega = \omega(\cdot) = W(\cdot)(\omega)$. If the reader does not want so much structure on the original probability space, think of the path $(W(t)(\omega))_{t\in[0,T]}$ whenever ω appears in the following. The theorem is now:

Theorem 5.2. *Let $\mathcal{X}$ be Malliavin and Fréchet differentiable and pathwise increasing, i.e. for $\omega' \geq \omega$, it holds that $\mathcal{X}(\omega') \geq \mathcal{X}(\omega)$ and let Δ_+ and Δ_- be deterministic constants. Consider the solution of the BSDE*

$$
Y(t) = \mathcal{X} - \int_t^T Z^+(s)\Delta_+ + Z^-(s)\Delta_- ds - \int_t^T Z(s)dW(s), \tag{33}
$$

where $Z^+ = \max(0,Z)$ and $Z^- = \max(0,-Z)$. Define the measure $Q^\#$ via its Radon-Nikodym derivative w.r.t. P,

$$
\frac{dQ^\#}{dP}|\mathcal{F}_t := \exp\left(-\int_0^t \Delta_+ dW(s) - \frac{1}{2}\int_0^t \Delta_+^2 ds\right). \tag{34}
$$

Then, for all $t\in[0,T]$, $Y(t) = E_{Q^\#}[\mathcal{X}|\mathcal{F}_t]$.

Remark 5.2. *Together with Theorem 5.1, this implies, that the risk valuation $\phi_t(\mathcal{X})$ of a pathwise increasing claim $\mathcal{X}$ can be written as a conditional expectation $\phi_t(\mathcal{X}) = E_{Q^\#}[\mathcal{X}|\mathcal{F}_t]$ with $\Delta_+ = \Delta_- = 2\left[\Psi\left(\frac{1}{2}\right)-\frac{1}{2}\right]$.*

Proof:

Let $\mathcal{X}$ pathwise increasing, i.e. for $\omega' \geq \omega$, it holds that $\mathcal{X}(\omega') \geq \mathcal{X}(\omega)$. The claim is now that $Y(t) = E_{Q^\#}[\mathcal{X}|\mathcal{F}_t]$ where $W^\#(s) = W(s) + \int_0^s \Delta_+ \mathrm{d}r$, i.e. $Y \equiv Y^\#$ for the two BSDEs

$$Y(t) = \mathcal{X} - \int_t^T Z^+(s)\Delta_+ + Z^-(s)\Delta_- \mathrm{d}s - \int_t^T Z(s)\mathrm{d}W(s)$$

and

$$\begin{aligned} E_{Q^\#}[\mathcal{X}|\mathcal{F}_t] = Y^\#(t) &= \mathcal{X} - \int_t^T Z^\#(s)\mathrm{d}W^\#(s) \\ &= \mathcal{X} - \int_t^T Z^\#(s)\Delta_+\mathrm{d}s - \int_t^T Z^\#(s)\mathrm{d}W(s). \end{aligned} \tag{35}$$

If we can show that $Z^\# \geq 0$ $(\lambda \otimes P)$-a.s. then BSDE (35) becomes the same as BSDE (33), which means that, for uniqueness of a solution of a BSDE with Lipschitz driver, it holds $(Y, Z) = (Y^\#, Z^\#)$ and everything is shown. As an easy application of the Clark-Ocone Theorem under change of measure (compare Theorem A.4 in the appendix or [9], Theorem 4.5), we see that

$$\begin{aligned} Z^\#(s) &= E_{Q^\#}[(D_s\mathcal{X} + \mathcal{X}\int_s^T \underbrace{D_s(-\Delta_+)}_{=0,\text{ since }\Delta_+\text{ determ.}} \mathrm{d}W^\#(r))|\mathcal{F}_s] \\ &= E_{Q^\#}[D_s\mathcal{X}|\mathcal{F}_s], \end{aligned}$$

where D is the Malliavin derivative. Now take a look at the directional derivative D^γ (compare Definition A.3 in the appendix) for a nonnegative direction $\gamma \in H^1$, $\gamma \geq 0$ where

$$\gamma(\cdot) = \int_0^\cdot g(t)\mathrm{d}t,$$

it holds

$$D^\gamma\mathcal{X}(\omega) = \lim_{\varepsilon \to 0} \frac{\mathcal{X}(\omega + \varepsilon\gamma) - \mathcal{X}(\omega)}{\varepsilon} \geq 0,$$

because with $\omega' := \omega + \varepsilon\gamma$ we have $\omega' \geq \omega$ if $\varepsilon > 0$ and $\omega' \leq \omega$ if $\varepsilon < 0$ and thus, for $\mathcal{X}$ pathwise increasing, the numerator and the denominator in the limes have the same sign. On the other hand, we know (compare Definitions A.3 and A.4 and Remark A.4 in the appendix) that

$$\int_0^T D_t\mathcal{X} \cdot g(t)\mathrm{d}t = \langle D\mathcal{X}, g\rangle_{L^2([0,T])} = D^\gamma\mathcal{X} \geq 0.$$

Since this is true for *all nonnegative* γ, it holds in particular for *all nonnegative* g (since the integral of a nonnegative function g is nonnegative). Thus, we get that $D\mathcal{X} \geq 0$, i.e. $D_s\mathcal{X} \geq 0$ for λ-a.a. $s \in [0,T]$ a.s. and therefore also $Z^{\#}(s) = E_{Q^{\#}}[D_s\mathcal{X}|\mathcal{F}_s] \geq 0$ (for λ-a.a. $s \in [0,T]$ a.s.). This completes the proof.

□

The next subsection shows by an example, that in general the pathwise increasing assumption is necessary. I.e. equality of the valuation with the conditional expectation w.r.t. $Q^{\#}$ fails, if the claim is not pathwise increasing.

5.2 Risk Valuation of Call Option and Straddle (Unhedged)

Consider a Black-Scholes market with one riskless bank account and one risky asset. Let w.l.o.g. be the riskless rate $r = 0$ (otherwise consider the discounted setting):

$$\begin{cases} S_0(t) &= 1,\ t \in [0,T] \\ \mathrm{d}S_1(t) &= \mu S_1(t)\mathrm{d}t + \sigma S_1(t)\mathrm{d}W(t). \end{cases} \tag{36}$$

Furthermore, consider a European Call option and a Straddle on S_1, both with maturity T and Strike K:

$$\begin{aligned} X_{Call} &:= (S_1(T) - K)^+, \\ X_{Straddle} &:= |S_1(T) - K|. \end{aligned} \tag{37}$$

It is easy to see that the Call is pathwise increasing, while the Straddle is not. We take the $AVaR_\alpha$-scheme with $\alpha < \frac{1}{2}$ as an example for a risk valuation. Our aim is now to show, that for the Call, the solution of the limiting BSDE is the same as the conditional expectation w.r.t. $Q^{\#}$, as stated in Theorem 5.2, while for the Straddle the two processes differ.

First we take a look at the limiting BSDE for the dynamic $AVaR_\alpha$ in one dimension:

$$\phi_t(X) = X - \int_t^T g(Z(s))\mathrm{d}s - \int_t^T Z(s)\mathrm{d}W(s). \tag{38}$$

Due to Stadje (2010) [2], the driver is (where $\chi_p(z)$ the p-th largest value of the set $\{z, -z\}$)

$$\begin{aligned} g(z) &= -\frac{1}{\alpha}\left(\chi_{2-\lceil 2\alpha\rceil+1}(z)\left(\alpha - \underbrace{\frac{\lceil 2\alpha\rceil - 1}{2}}_{=0}\right) + \underbrace{\frac{1}{2}\sum_{j=1}^{\lceil 2\alpha\rceil-1}\chi_{2-j+1}(z)}_{=0}\right) \\ &= -\min\{z, -z\} \\ &= |z|. \end{aligned}$$

On the other hand, Theorem 5.2 states, that in case of a pathwise increasing Malliavin and Fréchet differentiable claim, the valuation can be calculated as

$$\phi_t(X) = E_{Q^\#}[X|\mathcal{F}_t]. \tag{39}$$

By Equation (35) and using that $\Delta_+ = 2\left[\Psi\left(\frac{1}{2}\right) - \frac{1}{2}\right] = 1$ (since $\Psi(\frac{1}{2}) = \min\{1, \frac{1}{2\alpha}\} = 1$ for $\alpha < \frac{1}{2}$), $E_{Q^\#}[X|\mathcal{F}_t]$ solves the BSDE

$$E_{Q^\#}[X|\mathcal{F}_t] = X - \int_t^T Z^\#(s)\mathrm{d}s - \int_t^T Z^\#(s)\mathrm{d}W(s).$$

The question is now how to calculate $E_{Q^\#}[X|\mathcal{F}_t]$. Fortunately, the Black-Scholes formula can be used to achieve the answer. It states that the fair value of a Call option on S_1 with strike K and maturity T is given by

$$Call(S_1, t) := S_1(t)\Phi(d_1(S_1(t), T-t)) - Ke^{-r(T-t)}\Phi(d_2(S_1(t), T-t)),$$

where

$$\begin{aligned} d_1(s,t) &= \frac{\log(s/K) + (r + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}, \\ d_2(s,t) &= \frac{\log(s/K) + (r - \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}. \end{aligned}$$

Besides, we know that

$$\begin{aligned} Call(S_1, t) &= e^{rt}E_Q[e^{-rT}X_{Call}|\mathcal{F}_t] \\ &= e^{-r(T-t)}\left(X - \int_t^T Z^Q(s)\mathrm{d}W^Q(s)\right) \\ &= e^{-r(T-t)}\left(X - \int_t^T Z^Q(s)\frac{\mu - r}{\sigma}\mathrm{d}s - \int_t^T Z^Q(s)\mathrm{d}W(s)\right) \end{aligned}$$

where Q is the risk-neutral measure, given by

$$\frac{\mathrm{d}Q}{\mathrm{d}P} = \exp\left(-\int_0^T \frac{\mu - r}{\sigma}\mathrm{d}t - \frac{1}{2}\int_0^T \left(\frac{\mu - r}{\sigma}\right)^2 \mathrm{d}W(t)\right).$$

The BSDE appearing in the formula for $Call(S_1, t)$ looks quite similar to the one for $E_{Q^\#}[X|\mathcal{F}_t]$, the only difference is the term $\frac{\mu - r}{\sigma}$ in it. Fortunately the riskless rate r did not play a role in the calculation of $E_{Q^\#}[X|\mathcal{F}_t]$ and therefore we can **define a pseudo-riskless rate**

$$r^\# := \mu - \sigma, \tag{40}$$

so that $\frac{\mu - r^\#}{\sigma} = 1$ and the two BSDEs become the same. Thus, we get:

$$E_{Q^\#}[X_{Call}|\mathcal{F}_t] = e^{r^\#(T-t)}S_1(t)\Phi(d_1^\#(S_1(t), T-t)) - K\Phi(d_2^\#(S_1(t), T-t)), \tag{41}$$

where

$$d_1^\#(s,t) = \frac{\log(s/K) + (r^\# + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}, \tag{42}$$

$$d_2^\#(s,t) = \frac{\log(s/K) + (r^\# - \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}. \tag{43}$$

To calculate $E_{Q^\#}[X_{Straddle}|\mathcal{F}_t]$, we first use the Put-Call parity and then that $X_{Straddle} = X_{Put} + X_{Call}$ plus the linearity of the conditional expectation. This yields

$$E_{Q^\#}[X_{Put}|\mathcal{F}_t] = -e^{r^\#(T-t)}S_1(t)\Phi(-d_1^\#(S_1(t), T-t)) + K\Phi(-d_2^\#(S_1(t), T-t))$$

and thus

$$\begin{aligned} E_{Q^\#}[X_{Straddle}|\mathcal{F}_t] = e^{r^\#(T-t)}S_1(t)\{\Phi(d_1^\#(S_1(t), T-t)) - \Phi(-d_1^\#(S_1(t), T-t))\} \\ + K\{\Phi(-d_2^\#(S_1(t), T-t)) - \Phi(d_2^\#(S_1(t), T-t))\}. \end{aligned} \tag{44}$$

Now let us compare the solution of BSDE (38) for the Call and the Straddle with the solutions for $E_{Q^\#}[X_{Call}|\mathcal{F}_t]$ and $E_{Q^\#}[X_{Straddle}|\mathcal{F}_t]$ given in Equations (41) and (44). Therefore we choose the following market parameters: the (annual) drift rate $\mu = 0.05$, (annual) volatility $\sigma = 0.2$, starting value of the stock and strike $S_1(0) = K = 100$ EUR and maturity $T = 1$ (year). This was implemented in Matlab using again the seed 100 from the generator "mt19937ar" and again, the BSDEs were solved numerically by use of the Bender-Denk algorithm. The number of timesteps used for the simulation of S_1 was 100 and the number of timesteps in the Bender-Denk algorithm was 50. Moreover, a polynomial basis (with one exception: $e_1(x) = \max\{x - K, 0\}$ was used to achieve the replicating property at $t = T$ for both claims) and 100000 copies of S_1 where used in the Bender-Denk scheme. The results can be seen in Figure 3. One can clearly observe that for the Straddle, the solution of BSDE (38), which should approximate the true path of the risk valuation differs very much from the solution of Equation (44). This shows that the assumption in Theorem 5.2 that the claim is pathwise increasing, really is necessary. For the Call, as expected, both solutions do not differ much. Theoretically they are the same, but remember that we solved the BSDE only numerically, so we cannot expect that they totally match (actually this gives us an idea about how big the approximation error really is).

Again, 29 more paths of the BSDE solution and the conditional Expectations of Call and Straddle were simulated (each time with 100000 new copies of S_1). The seeds that were used are 101 to 129. The results shall only be touched on briefly. For the Call option, the empirical mean of the difference between BSDE solution and conditional expectation solution at 0, $\hat{Y}^{BSDE}_{Call} - E_{Q^\#}[X_{Call}]$

was about -0.51 EUR and the empirical variance was about $5.62 \cdot 10^{-4}$ EUR^2. For the Straddle, the empirical mean of this difference was -9.69 EUR and the empirical variance was about $2.5 \cdot 10^{-3}$ EUR^2. A histogram for these differences is to be found in Figure 6 in the appendix. In Figure 7 in the appendix the solutions for the seeds 101, 102, 103 and 126 are plotted.

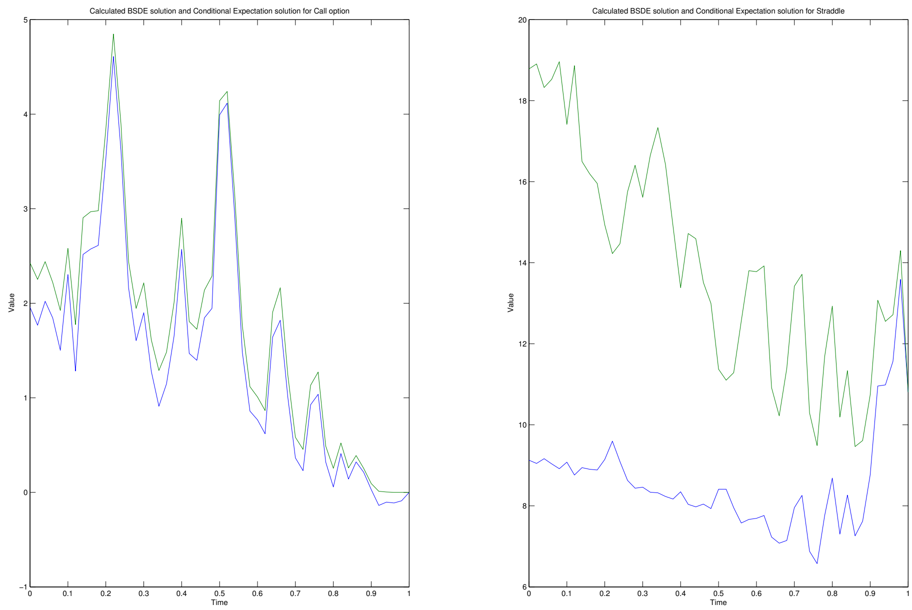

Figure 3: left: BSDE-solution (blue) and Expectation solution (green) for the Call (one path), right: BSDE-solution (blue) and Expectation solution (green) for the Straddle (the same one path)

6 Conclusion

This chapter consists of two parts: First, a short summary of what we have learned is given. Afterwards the problem that the limiting BSDE of the valuations depends on the approximation scheme for the Brownian motion that was chosen is discussed and a proposal for solution is made.

6.1 Summary

Let us briefly recall the main steps in this thesis and point out the most important perceptions. The first insight was that one cannot simply generalize the concept of a static risk measure which is given by a Choquet integral w.r.t. a distortion Ψ canonically to a dynamic risk measure because this leads to time-inconsistency. Instead we defined a time-consistent valuation in discrete time by concatenating static valuations recursively, which led to the so called distorted risk measure. But some rescaling was needed. We showed that this can be done either at the one-step valuation level or at the distortion level. Afterwards a brief summary of the proof given in Stadje (2010) [2] showed that after rescaling the valuations defined in a Bernoulli random walk setting tend to the solution of a BSDE. This means that the properties of a risk valuation are automatically fulfilled, so the outcome is indeed a risk valuation in continuous time. As an example of practical relevance a case study concerning the risk related to cross-hedging in incomplete markets was made. Eventually the 1-dimensional case was investigated in detail. It was shown how the limiting BSDE simplifies in this case and that for pathwise increasing claims the risk valuation can always be written as a usual conditional expectation w.r.t. some measure $Q^{\#}$ that depends on the distortion Ψ. The example of an unhedged straddle showed that the pathwise increasing property cannot be dropped.

6.2 Outlook

As mentioned before, the convergence of the discrete-time valuations against a limiting BSDE was shown particularly in a Bernoulli random walk setting approximating the Brownian motion. Nothing is said about other approximation schemes. Indeed it was shown in [1] that for a Trinomial random walk approximation, the valuations also converge to the solution of a BSDE, but this BSDE differs from the one of the Bernoulli approximation scheme. This fact was also pointed out in Stadje (2010) [2], where it was mentioned that even if the approximation schemes converge against the same Brownian motion, the drivers of the limiting BSDEs and thus the resulting continuous time risk measures could differ. Well, this is somehow dissatisfactory. Of course if one only wants to construct a dynamic risk measure that is "in some sense" a generalization of a static risk measure induced by a distortion, then one can choose a scheme in which convergence towards a solution of a BSDE is shown and use for example a Bender-Denk scheme to solve the BSDE. But this approach seems quite haphazardly.

An intuitive approach is the following: Instead of approximating the whole continuous time setting driven by a Brownian motion by a discrete time setting, for example driven by a Bernoulli random walk, one could rather stay in the

Brownian setting but devide the time-interval into finitely many sub-intervals and establish the one-period valuations on these intervals. Then one increases the frequence of measurement, i.e. the number of sub-intervals. This would correspond to the following interpretation: The financial institution measures the risk dynamically via the recursive definition first weekly, then daily then at every hour, and so on and so forth. These risk measurements converge to the continuous time risk measure which corresponds to the valuation given by the limiting BSDE (if there is a limiting BSDE). We will now show heuristically that there should indeed exist a limiting BSDE in this setting. The idea of the proof is quite similar to the one for the Bernoulli random walk approximation. It consists of five steps:

1. Use the Clark-Ocone formula:

$$\phi^N_{t_{i+1}}(X) = \underbrace{E[\phi^N_{t_{i+1}}(X)|\mathcal{F}_{t_i}]}_{=:\beta^N_{t_i}\ldots\mathcal{F}_{t_i}\text{-msble}} + \int_{t_i}^{t_{i+1}} \underbrace{E[D_s\phi^N_{t_{i+1}}(X)|\mathcal{F}_s]}_{=:Z^N_{t_i}(s)} \, \mathrm{d}W(s).$$

2. Use the recursive definition (15), i.e. use the same rescaling as in the Bernoulli random walk case (this is because $W(t)$ "scales like" $\sqrt{t}$) to get for the time-increments:

$$\begin{aligned}\phi^N_{t_{i+1}}(X) - \phi^N_{t_i}(X) &= \phi^N_{t_{i+1}}(X) - \phi^N_{t_i,t_{i+1}}(\phi^N_{t_{i+1}}(X)) \\ &= \beta_{t_i} + \int_{t_i}^{t_{i+1}} Z^N_{t_i}(s)\mathrm{d}W(s) \\ &\quad - \phi^N_{t_i,t_{i+1}}(\beta_{t_i} + \int_{t_i}^{t_{i+1}} Z^N_{t_i}(s)\mathrm{d}W(s)).\end{aligned}$$

   Here we used the first step of the proof. Because of the cash invariance, we can pull β_{t_i} out of $\phi^N_{t_i,t_{i+1}}$ such that it cancels out and it remains

$$\phi^N_{t_{i+1}}(X) - \phi^N_{t_i}(X) = \int_{t_i}^{t_{i+1}} Z^N_{t_i}(s)\mathrm{d}W(s) - \phi^N_{t_i,t_{i+1}}(\int_{t_i}^{t_{i+1}} Z^N_{t_i}(s)\mathrm{d}W(s)).$$

3. Using this result and a telescope sum argument yields:

$$\begin{aligned}\phi^N_{t_i}(X) = X - &\left[\sum_{j=i}^{k(N)-1} -\phi^N_{t_j,t_{j+1}}(\int_{t_j}^{t_{j+1}} Z^N_{t_j}(s)\mathrm{d}W(s))\right] \\ - &\underbrace{\sum_{j=i}^{k(N)-1} \int_{t_j}^{t_{j+1}} Z^N_{t_j}(s)\mathrm{d}W(s)}_{=\int_{t_i}^T \sum_{j=0}^{k(N)-1} Z^N_{t_j}(s)\mathbb{1}_{[t_j,t_{j+1})}(s)\mathrm{d}W(s)}.\end{aligned}$$

   Now define

$$Z^N(s) := \sum_{j=0}^{k(N)-1} Z^N_{t_j}(s)\mathbb{1}_{[t_j,t_{j+1})}(s), \tag{45}$$

to get

$$\phi^N_{t_i}(X) = X - \left[\sum_{j=i}^{k(N)-1} -\phi^N_{t_j,t_{j+1}}(\int_{t_j}^{t_{j+1}} Z^N(s)\mathrm{d}W(s)) \right] \\ - \int_{t_i}^{T} Z^N(s)\mathrm{d}W(s). \tag{46}$$

4. The idea is now to figure out whether (or rather to find conditions under which) it is justified to substitute $\int_{t_j}^{t_{j+1}} Z^N(s)\mathrm{d}W(s)$ by $Z^N(t_j)\Delta W(t_{j+1})$ in the last equation, where we define $\Delta W(t_{j+1}) := W(t_{j+1}) - W(t_j)$.

 Assumption 1: Assume that Ψ is Lebesgue-a.e. differentiable and let Ψ' be bounded. Let us first focus on the case where Ψ is differentiable everywhere, to see the idea behind this assumption. Let $Q \in \mathcal{D}^{t_i}$ with $Q|\mathcal{F}_{t_i} = P|\mathcal{F}_{t_i}$. Then we have for any $A \in \mathcal{F}_{t_{i+1}}$ that

$$\begin{aligned} E_P[Q(A|\mathcal{F}_{t_i})] &= E_Q[Q(A|\mathcal{F}_{t_i})] \\ &= E_Q[E_Q[\mathbb{1}_A|\mathcal{F}_{t_i}]] \\ &= E_Q[\mathbb{1}_A] \\ &= E_P\left[\frac{\mathrm{d}Q}{\mathrm{d}P}\mathbb{1}_A\right]. \end{aligned}$$

 where we recall that we defined $E_Q[...]$ actually P-almost surely. Now assume for contradiction that it exists a set $A \in \mathcal{F}_{t_{i+1}}$ such that $\frac{\mathrm{d}Q}{\mathrm{d}P} > \max_{x\in[0,1]} \Psi'(x)$. Then, by the mean value theorem, we know that for any $x, y \in [0,1]$, $x < y$ it exists a $z \in (x, y)$ such that

$$\frac{\Psi(y) - \Psi(x)}{y - x} = \Psi'(z).$$

 Therefore we have that for $M \geq \max_{x\in[0,1]} \Psi'(x)$, $M(y-x) \geq \Psi(y)-\Psi(x)$ for all $x, y \in [0,1]$, $x < y$ and thus

$$\begin{aligned} E_P[Q(A|\mathcal{F}_{t_i})] &= E_P\left[\frac{\mathrm{d}Q}{\mathrm{d}P}\mathbb{1}_A\right] \\ &> E_P[\max_{x\in[0,1]} \Psi'(x)\mathbb{1}_A] \\ &= E_P[\max_{x\in[0,1]} \Psi'(x)E_P[\mathbb{1}_A|\mathcal{F}_{t_i}]] \\ &= E_P[\max_{x\in[0,1]} \Psi'(x)(P(A|\mathcal{F}_{t_i}) - 0)] \\ &\geq E_P[\Psi(P(A|\mathcal{F}_{t_i})) - \underbrace{\Psi(0)}_{=0}]. \end{aligned}$$

 So there must be an $\mathcal{F}_{t_i}$-measurable set B, s.t. $Q(A|\mathcal{F}_{t_i}) > \Psi(P(A|\mathcal{F}_{t_i})$ on B which is a contradiction to $Q \in \mathcal{D}^{t_i}$. Therefore we know that $\frac{\mathrm{d}Q}{\mathrm{d}P} \leq \max_{x\in[0,1]} \Psi'(x)$ i.e. it is bounded. This result should be easy to generalize to the case where Ψ is differentiable only Lebesgue almost everywhere.

Assumption 2: A quite strong assumption is the following:

$$\text{esssup}_{j\in\{1,\dots,k(N)\}} E\left[\int_{t_j}^{t_{j+1}} \frac{(Z^N(s)-Z^N(t_j))^2}{\Delta t_{j+1}}\mathrm{d}s\,\middle|\,\mathcal{F}_{t_j}\right] \to 0 \text{ a.s.} \quad (47)$$

as $N\to\infty$. Let these assumptions be fulfilled. We have that

$$\sum_{j=i}^{k(N)-1}\left[\phi^N_{t_j,t_{j+1}}\left(\int_{t_j}^{t_{j+1}} Z^N(s)\mathrm{d}W(s)\right) - \phi^N_{t_j,t_{j+1}}(Z^N(t_j)\Delta W(t_{j+1}))\right]$$
$$= \sum_{j=i}^{k(N)-1}\Delta t_{j+1}\left[F^N_{t_j}\left(\int_{t_j}^{t_{j+1}} \frac{Z^N(s)}{\sqrt{\Delta t_{j+1}}}\mathrm{d}W(s)\right) - F^N_{t_j}\left(\frac{Z^N(t_j)\Delta W(t_{j+1})}{\sqrt{\Delta t_{j+1}}}\right)\right].$$

By definition, it holds

$$F^N_{t_j}\left(\int_{t_j}^{t_{j+1}} \frac{Z^N(s)}{\sqrt{\Delta t_{j+1}}}\mathrm{d}W(s)\right)$$
$$= F^N_{t_j}\left(\int_{t_j}^{t_{j+1}} \frac{Z^N(s)-Z^N(t_j)}{\sqrt{\Delta t_{j+1}}}\mathrm{d}W(s) + \frac{Z^N(t_j)\Delta W(t_{j+1})}{\sqrt{\Delta t_{j+1}}}\right)$$
$$= \text{essinf}_{Q^N\in\mathcal{D}^{t_j}} E_{Q^N}\left[\int_{t_j}^{t_{j+1}} \frac{Z^N(s)-Z^N(t_j)}{\sqrt{\Delta t_{j+1}}}\mathrm{d}W(s) + \frac{Z^N(t_j)\Delta W(t_{j+1})}{\sqrt{\Delta t_{j+1}}}\,\middle|\,\mathcal{F}_{t_j}\right].$$

For the first summand inside the essinf, using Hölder's inequality for conditional expectations and the fact that w.l.o.g. we can assume that $Q^N|\mathcal{F}_{t_j} = P|\mathcal{F}_{t_j}$, we get

$$E_{Q^N}\left[\int_{t_j}^{t_{j+1}} \frac{Z^N(s)-Z^N(t_j)}{\sqrt{\Delta t_{j+1}}}\mathrm{d}W(s)\,\middle|\,\mathcal{F}_{t_j}\right]$$
$$\leq \underbrace{\sqrt{E\left[\left(\frac{\mathrm{d}Q^N}{\mathrm{d}P}\right)^2\,\middle|\,\mathcal{F}_{t_j}\right]}}_{\text{bounded by Assumption 1}}\underbrace{\sqrt{E\left[\left(\int_{t_j}^{t_{j+1}} \frac{Z^N(s)-Z^N(t_j)}{\sqrt{\Delta t_{j+1}}}\mathrm{d}W(s)\right)^2\,\middle|\,\mathcal{F}_{t_j}\right]}}_{\to 0 \text{ a.s. by Ito's isometry and Assumption 2}}$$
$$\to 0 \text{ a.s.}$$

Analoguously (by intersecting an essinf and an esssup) one shows that

$$\text{esssup}_{j\in\{1,\dots,k(N)\}} E_{Q^N}\left[\int_{t_j}^{t_{j+1}} \frac{Z^N(s)-Z^N(t_j)}{\sqrt{\Delta t_{j+1}}}\mathrm{d}W(s)\,\middle|\,\mathcal{F}_{t_j}\right] \to 0 \text{ a.s.}$$

which means that

$$\text{esssup}_{j\in\{1,\dots,k(N)\}} F^N_{t_j}\left(\int_{t_j}^{t_{j+1}} \frac{Z^N(s)}{\sqrt{\Delta t_{j+1}}}\mathrm{d}W(s)\right) - F^N_{t_j}\left(\frac{Z^N(t_j)\Delta W(t_{j+1})}{\sqrt{\Delta t_{j+1}}}\right)$$
$$\to 0 \text{ a.s.}$$

But then we also have that

$$\sum_{j=i}^{k(N)-1} \left[\phi^N_{t_j,t_{j+1}} \left(\int_{t_j}^{t_{j+1}} Z^N(s) \mathrm{d}W(s) \right) - \phi^N_{t_j,t_{j+1}} (Z^N(t_j) \Delta W(t_{j+1})) \right]$$

$$= \sum_{j=i}^{k(N)-1} \Delta t_{j+1} \left[\underbrace{F^N_{t_j} \left(\int_{t_j}^{t_{j+1}} \frac{Z^N(s)}{\sqrt{\Delta t_{j+1}}} \mathrm{d}W(s) \right) - F^N_{t_j} \left(\frac{Z^N(t_j)\Delta W(t_{j+1})}{\sqrt{\Delta t_{j+1}}} \right)}_{\to 0} \right]$$

$\to 0$ a.s.

This legitimates the approximation

$$\phi^N_{t_i}(X) \approx X - \left[\sum_{j=i}^{k(N)-1} -\Delta t_{j+1} F^N_{t_j} \left(\frac{Z^N(t_j)\Delta W(t_{j+1})}{\sqrt{\Delta t_{j+1}}} \right) \right] - \int_{t_i}^{T} Z^N(s) \mathrm{d}W(s),$$

since in the limit, it is the same as in Equation (46). Now define

$$g^N(t_j, z) := -F^N_{t_j} \left(z \frac{\Delta W(t_{j+1})}{\sqrt{\Delta t_{j+1}}} \right), \tag{48}$$

then we get

$$\phi^N_{t_i}(X) \approx X - \sum_{j=i}^{k(N)-1} g^N(t_j, Z^N(t_j)) \Delta t_{j+1} - \int_{t_i}^{T} Z^N(s) \mathrm{d}W(s). \tag{49}$$

5. The rest of the proof should work quite similar to the Bernoulli random walk case. In particular, if $F^N_{t_j}\left(z \frac{\Delta W(t_{j+1})}{\sqrt{\Delta t_{j+1}}} \right)$ is deterministic and does not depend on j and N (as for example in case of an $AVaR$-scheme), the limiting BSDE should exist.

Since this proof was not very strict but rather heuristically, and the result is quite interesting, it probably deserves a more thorough investigation.

A Appendix

A.1 Malliavin Calculus

Through this whole section, let $(\Omega, \mathcal{F}_T, (\mathcal{F}_t)_{t\in[0,T]}, P)$ be a filtrated probability space and let the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ be generated by a Brownian motion W. All definitions and theorems within this section are taken from [9]. This section is only meant to be a short reminder of the most important results from Malliavin calculus and no complete derivation of theory. For a derivation of the theory, reading [9] is recommended.

Definition A.1 (Iterated and Multiple Ito-Integrals).
Define $S_n = \{(t_1, ..., t_n) \in [0,T]^n : 0 \le t_1 \le ... \le t_n \le T\}$.

1. *Let* f *be a deterministic function on* S_n*, such that*

$$\int_{S_n} f^2(t_1, ..., t_n)dt_1...dt_n < \infty.$$

Then the n*-fold iterated Ito-integral of* f *is defined as*

$$J_n(f) := \int_0^T \int_0^{t_n} ... \int_0^{t_2} f(t_1, ..., t_n)dW(t_1)...dW(t_n).$$

2. *Let* $f \in L^2([0,T]^n)$ *be symmetric, i.e.* $f(t_1, ..., t_n) = f(t_{\sigma(1)}, ..., t_{\sigma(n)})$ *for any permutation* σ *of* $(1, ..., n)$*. Then the* n*-fold multiple Ito-integral is defined as*

$$I_n(f) := \int_{[0,T]^n} f(t_1, ..., t_n)dW(t_1)...dW(t_n) := n!J_n(f)$$

and for a constant f, $I_0(f) := f$.

Theorem A.1 (Wiener Chaos Expension).
Let $F \in L^2(\mathcal{F}_T)$*. Then there exists a unique sequence* $(f_n)_n$ *of symmetric functions* $f_n \in L^2([0,T])$ *such that (with convergence in* $L^2(P)$*)*

$$F = \sum_{n=0}^{\infty} I_n(f_n).$$

Definition A.2 (Malliavin Derivative).
Let $F \in L^2(\mathcal{F}_T)$ *with Wiener chaos expension as in Theorem A.1. Then*

1. $F \in \mathbb{D}_{1,2}$ *if*

$$||F||^2_{\mathbb{D}_{1,2}} := \sum_{n=1}^{\infty} nn!||f||^2_{L^2([0,T]^n)} < \infty.$$

If this is the case, we call F *Malliavin differentiable.*

2. *For* $F \in \mathbb{D}_{1,2}$ *we define the Malliavin Derivative of* F, D_tF *as*

$$D_tF = \sum_{n=0}^{\infty} nI_{n-1}(f_n(\cdot, t)),\ t \in [0,T],$$

where I_{n-1} *integrates w.r.t. the first* $n-1$ *arguments of* f.

Remark A.1 (Partial Malliavin Derivatives).
If W is a d-dimensional Brownian motion $W = (W_1, ..., W_d)^{tr}$, generating the filtration $(\mathcal{F}_t)_{t\in[0,T]}$, then $D_{1,t}F$,..., $D_{d,t}F$ denote the partial Malliavin derivatives, i.e. $D_{i,t}F$ is the Malliavin derivative of F w.r.t. W_i, where the W_j, $j \neq i$ are treated like constants.

Theorem A.2 (Product Rule).
Let $F_1, F_2 \in \mathbb{D}_{1,2}$ each with a Wiener chaos expension with only finitely many terms. Then also $F_1F_2 \in \mathbb{D}_{1,2}$ with

$$D_t(F_1F_2) = F_1D_tF_2 + F_2D_tF_1.$$

Theorem A.3 (Chain Rule).
Let $F \in \mathbb{D}_{1,2}$, $g \in C^1(\mathbb{R})$ with bounded derivative. Then $g(F) \in \mathbb{D}_{1,2}$ with

$$D_tg(F) = g'(F)D_tF.$$

The next theorem includes a very important representation formula, namely the **Clark-Ocone formula under change of measure**. First of all let u be an $(\mathcal{F}_t)_{t\in[0,T]}$-adapted stochastic process satisfying the Novikov condition

$$E[\exp(\frac{1}{2}\int_0^T u^2(s)\mathrm{d}s)] < \infty,$$

and define an equivalent measure Q by

$$\frac{\mathrm{d}Q}{\mathrm{d}P}|\mathcal{F}_t = Z(t),$$

where

$$Z(t) = \exp(-\int_0^T u(s)\mathrm{d}W(s) - \frac{1}{2}\int_0^T u^2(s)\mathrm{d}s).$$

Then, by Girsanov's Theorem,

$$\hat{W}(t) := \int_0^t u(s)\mathrm{d}s + W(t)$$

is an $(\mathcal{F}_t)_{t\in[0,T]}$-Brownian motion under Q. Now the following representation holds:

Theorem A.4 (Clark-Ocone Formula under Change of Measure).
Let $F \in \mathbb{D}_{1,2}$ with existing mean under Q and

$$E_Q[\int_0^T |D_tF|^2 dt] < \infty.$$

Furthermore let $Z(T)F \in \mathbb{D}_{1,2}$, $u(s) \in \mathbb{D}_{1,2}$ for λ-a.e. s and

$$E_Q[|F|\int_0^T (\int_0^T D_tu(s)dW(s) + \int_0^T u(s)D_tu(s)ds)^2 dt] < \infty.$$

Then

$$F = E_Q[F] + \int_0^T E_Q[D_t F - F \int_t^T D_t u(s) d\hat{W}(s) | \mathcal{F}_t] d\hat{W}(t). \tag{50}$$

Remark A.2 (Original Clark-Ocone Formula)**.**
Let $Q = P$ and all assumptions fulfilled. Then we have the usual Clark-Ocone formula:

$$F = E[F] + \int_0^T E[D_t F | \mathcal{F}_t] dW(t).$$

From now on we set $\Omega := C_0([0,T])$ (the space of continuous functions ω with $\omega(0) = 0$). Equipped with the supremum norm $||\omega||_\infty := \sup_{t\in[0,T]} |\omega(t)|$, this becomes a Banach space. Let $\mathcal{F}$ be the Borel σ-algebra generated by the $||\cdot||_\infty$-open sets. It is possible to define a probability measure P on $(\Omega, \mathcal{F})$ so that W given by $W(\cdot)(\omega) = \omega(\cdot)$ is a Brownian motion. The resulting probability space $(\Omega, \mathcal{F}, P)$ is called **Wiener space**.

Moreover, we define the so called **Cameron-Martin space** by $H^1 := \{h \in C_0([0,T]), h' \text{ exists and } h' \in L^2([0,T])\}$, equipped with the scalar product $\langle g, h\rangle_{H^1} := \int_0^T g'(t)h'(t)\mathrm{d}t$. Note that

$$\gamma \in H^1 \Leftrightarrow \exists g \in L^2([0,T|) : \gamma(\cdot) = \int_0^\cdot g(t)\mathrm{d}t$$

For a r.v. $F : \Omega \to \mathbb{R}$ we will now define a directional derivative for directions in the Cameron-Martin space.

Definition A.3 (Directional Derivative)**.**
Let $F : \Omega \to \mathbb{R}$ and $\gamma \in H^1$ (in particular, γ deterministic). Then, the directional derivative of F in $L^2(P)$ in direction $\gamma \in H^1$ is defined as

$$D^\gamma F(\omega) = \lim_{\varepsilon\to 0} \frac{F(\omega + \varepsilon\gamma) - F(\omega)}{\varepsilon}, \tag{51}$$

provided this limit exists in $L^2(P)$.

Now we define the Fréchet derivative as a stochastic gradient.

Definition A.4 (Fréchet Derivative)**.**
Let $F : \Omega \to \mathbb{R}$ have a directional derivative in all directions $\gamma \in H^1$,

$$\gamma(\cdot) = \int_0^\cdot g(t)\,dt, \;\; g \in L^2([0,T]),$$

and suppose that there exists a function $\psi : \mathbb{R} \to L^2(P \times \lambda)$ such that the scalar product $\langle\psi(\cdot,\omega), g\rangle_{L^2([0,T])} := \int_\mathbb{R} \psi(t,\omega)g(t)\,dt$ converges in $L^2(P)$ and

$$D^\gamma F(\omega) = \langle\psi(\cdot,\omega), g\rangle_{L^2([0,T])} = \int_\mathbb{R} \psi(t,\omega)g(t)\,dt, \tag{52}$$

then we write $F \in \mathcal{D}_{1,2}$, call F Fréchet differentiable in $L^2(P)$ and write

$$\mathbb{D}_t F(\omega) = \psi(t,\omega), \;\; t \in \mathbb{R}. \tag{53}$$

This can be seen as a gradient w.r.t. g, since now

$$D^\gamma F(\omega) = \langle\mathbb{D}F(\omega), g\rangle_{L^2([0,T])}.$$

Remark A.3. *Pay attention to the fact that in this definition we really mean* ***all*** $\gamma \in H^1$ *and not just P-a.a. (this would not make sense at all, because* $H^1 \subset \Omega$ *is a P-null-set).*

Remark A.4 (Relation between Malliavin and Fréchet Derivative)**.**
If $F \in \mathbb{D}_{1,2} \cap \mathcal{D}_{1,2}$*, then we call F Malliavin and Fréchet differentiable. In this case, both derivatives coincide:*

$$\mathbb{D}F = DF. \tag{54}$$

A.2 BSDE Theory

Let $g : \Omega \times [0,T] \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$. Write $g(t,y,z) := g(\cdot,t,y,z)$. The following assumptions are as in [11]:

(A1) $(g(t,y,z))_{t\leq 0}$ for any $(y,z) \in \mathbb{R} \times \mathbb{R}^d$ continuous progressively measurable with $E[\int_0^T |g(s,y,z)|^2 ds] < \infty$.

(A2) g is Lipschitz in (y,z) uniformly in (ω,t).

(A3) $g(t,y,0) = 0$ a.s. for all $(t,y) \in [0,T] \times \mathbb{R}$.

Definition A.5 (BSDE)**.**
A ***BSDE*** *with* ***terminal condition*** *X and* ***driver*** *g is an equation*

$$\begin{cases} dY(t) &= g(t,Y(t),Z(t))dt + Z(t)dW(t); t \in [0,T] \\ Y(T) &= X, \end{cases} \tag{55}$$

or alternatively

$$Y(t) = X - \int_t^T g(t,Y(t),Z(t))dt - \int_t^T Z(t)dW(t);\ t \in [0,T] \tag{56}$$

where X is $\mathcal{F}_T$*-measurable and Z predictable process s.t. stochastic integral is well-defined. The tuple* (Y,Z) *is called* ***solution of the BSDE (56)****.*

The following theorem was first stated by Pardoux and Peng (1990) [12]:

Theorem A.5 (Existence and Uniqueness of Adapted Solutions)**.**
Let $T > 0$, $X \in L^2(\mathcal{F}_T)$ *and the driver g fulfill (A1) and (A2). Then (56) has a* ***unique solution*** (Y,Z) *s.t. Y is adapted and*

$$E\left[\sup_{t\in[0,T]} |Y(t)|^2 + \int_0^T |Z(t)|^2 dt\right] < \infty.$$

The following property justifies the argumentation and conclusion in the section about the dynamic case:

Property A.1 (Relation between Dynamic Risk Measures and BSDEs)**.**
Assume g fulfills (A1)-(A3). It can be shown (see, for example [11]) that the Y-part of the solution of a BSDE with driver g fulfills all the properties of a valuation, i.e. $\rho := -Y$ *defines a dynamic risk measure.*

Next we introduce a numerical scheme for the solution of a BSDE (or more precisely: a forward-backwards SDE) of the following type:

$$\begin{aligned}
\mathrm{d}X(t) &= b(t, X(t))\mathrm{d}t + \sigma(t, X(t))\mathrm{d}W(t) \\
\mathrm{d}Y(t) &= f(t, X(t), Y(t), Z(t))\mathrm{d}t + Z(t)\mathrm{d}W(t) \\
X(0) &= x \\
Y(T) &= \xi.
\end{aligned}$$

The forward part does not lead to problems, it can easily be solved numerically by a Euler or Milstein scheme. But since we are looking for an **adapted** solution (Y, Z) of the backward part, we cannot simply apply an Euler or Milstein scheme backwards there. The following algorithm is proposed in Bender, Denk (2007) [10] and it is basically a Picard iteration scheme, i.e. the idea is the same as in ODE theory to define:

$$\begin{aligned}
\mathrm{d}Y^{(n)}(t) &= f(t, X(t), Y^{(n-1)}(t), Z^{(n-1)}(t))\mathrm{d}t + Z^{(n)}(t)\mathrm{d}W(t) \\
Y^{(n)}(T) &= \xi,
\end{aligned}$$

and increase n. Then, by rearranging the terms and using the Clark-Ocone formula, we get

$$Y^{(n)}(t) = E\left[\xi - \int_t^T f(s, X(s), Y^{(n-1)}(s), Z^{(n-1)}(s))\mathrm{d}s \,\middle|\, \mathcal{F}_t \right],$$

which projects at any time point t the future evolution (from the integrals from t to T) onto $L^2(\mathcal{F}_t)$ and thus gives us really an adapted solution. Forthermore we have that

$$\int_t^T Z^{(n)}(t)\mathrm{d}W(t) = \xi - Y^{(n)}(t) - \int_t^T f(t, X(t), Y^{(n-1)}(t), Z^{(n-1)}(t))\mathrm{d}t.$$

Therefore heuristically, for $d = 1$ (but can be generalized canonically) we get by use of Ito's isometry

$$\begin{aligned}
&Z^{(n)}(t_i)\Delta t_{i+1} \\
&= E\left[\int_{t_i}^{t_{i+1}} Z^{(n)}(t_i)\mathrm{d}W(s) \underbrace{(W(t_{i+1}) - W(t_i))}_{=:\Delta W_{i+1}} \,\middle|\, \mathcal{F}_{t_i} \right] \\
&\approx E\left[\int_{t_i}^{t_{i+1}} Z^{(n)}(s)\mathrm{d}W(s)\Delta W_{i+1} \,\middle|\, \mathcal{F}_{t_i} \right] \\
&= E\left[\left(\xi - \underbrace{Y^{(n)}(t_i)}_{\mathcal{F}_{t_i}\text{-msble}} - \int_{t_i}^{t_{i+1}} f(s, X(s), Y^{(n-1)}(s), Z^{(n-1)}(s))\mathrm{d}s \right) \Delta W_{i+1} \,\middle|\, \mathcal{F}_{t_i} \right] \\
&= E\left[\left(\xi - \int_{t_i}^{t_{i+1}} f(s, X(s), Y^{(n-1)}(s), Z^{(n-1)}(s))\mathrm{d}s \right) \Delta W_{i+1} \,\middle|\, \mathcal{F}_{t_i} \right]
\end{aligned}$$

For a fixed partition $\pi = \{t_0, ..., t_N : 0 = t_0 \leq ... \leq t_N = T\}$ and discretizations $X^{(\pi)}$ and $\xi^{(\pi)}$ of the forward process and the terminal condition this yields the approximation

$$Y^{(n,\pi)}(t_i) = E\left[\xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X^{(\pi)}(t_j), Y^{(n-1,\pi)}(t_j), Z^{(n-1,\pi)}(t_j))\Delta t_{j+1} \middle| \mathcal{F}_{t_i}\right]$$

$$Z_m^{(n,\pi)}(t_i)$$

$$= E\left[\frac{\Delta W_{i+1}^m}{\Delta t_{i+1}}\left(\xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X^{(\pi)}(t_j), Y^{(n-1,\pi)}(t_j), Z^{(n-1,\pi)}(t_j))\Delta t_{j+1}\right) \middle| \mathcal{F}_{t_i}\right],$$

where $m = 1, ..., d$, $\Delta t_{i+1} = t_{i+1} - t_i$ and $\Delta W_{i+1}^m = W_m(t_{i+1}) - W_m(t_i)$, $\Delta W_{N+1}^m := 0$. Since we have to calculate (conditional) expectations, the idea is now to use regression for approximating these expectations. First we replace projection on $L^2(\mathcal{F}_{t_i})$ (which is done by the conditional expectation) by projection on a finite-dimensional subspace of $L^2(\mathcal{F}_{t_i})$. Therefore we need a basis of a subspace of $L^2(\mathcal{F}_{t_i})$ to project on this subspace via least squares Monte-Carlo regression. Since in our case it is natural to choose the process X to be the financial position whose risk we want to evaluate, i.e. in our case $\xi := X(T) := \mathcal{X}$, we will define our projection basis $\{\eta_1^i, ..., \eta_K^i\}$ by some functions of $X^{(\pi)}(t_i)$:

$$\eta_k^i := e_k(X^{(\pi)}(t_i))$$

For example $e_1, ..., e_K$ could be the first K monomials $(x^k)_{k=0,...,K-1}$ or some indicator functions. Of course the goodness of the approximation depends on the choice of these functions.

Finally we have to estimate the projections on the span of $\{\eta_1^i, ..., \eta_K^i\}$ (which are actually the conditional expectations from above but now no longer conditioned on $\mathcal{F}_{t_i}$ but on $\{\eta_1^i, ..., \eta_K^i\}$). This is done using regression techniques. To do so, for every path of (X, W) we calculate L independent copies and use these for regression.

The resulting algortithm is:

Algorithm A.1 (Numerical Solution Scheme for BSDEs; Bender, Denk (2007))**.**

1. *Choose a time-grid $\pi = \{t_0, ..., t_N : 0 = t_0 \leq ... \leq t_N = T\}$. Simulate on π the processes $X^{(\pi)}$, W_m, $m = 1, ..., d$ and the terminal condition $\xi^{(\pi)}$ and L independent copies $(X_\lambda^{(\pi)})_{\lambda=1,...,L}$, $(W_{m,\lambda})_{\lambda=1,...,L}$ and $(\xi_\lambda^{(\pi)})_{\lambda=1,...,L}$ of $X^{(\pi)}$ and W_m and $\xi^{(\pi)}$. Calculate η_k^i and $\eta_{k,\lambda}^i$ for any $i \in \{0, ..., N\}$ and any $k \in \{1, ..., K\}$. Denote by $(\Delta\mathbb{W}_i^1, ..., \Delta\mathbb{W}_i^d, \mathbb{x}^{(\pi)}, \mathbb{X}^{(\pi)}(t_i), \mathbb{e}_k^i)$ the column vectors of these copies (i.e. $\mathbb{e}_k^i = (\eta_{k,1}^i, ..., \eta_{k,L}^i)^T$).*

2. *Define for $i = 0, ..., N$: $\mathcal{A}_i^L := \frac{1}{\sqrt{L}}(\eta_{k,\lambda}^i)_{\lambda=1,...,L;k=1,...,K}$ with pseudo-inverse $(\mathcal{A}_i^L)^+$.*

3. *Set n_{max}, define a breaking criterion and set for $i = 0, ..., N$ and $k = 1, ..., K$:*

$$\alpha_{i,k}^{(0,\pi,L)} = 0$$
$$\tilde{\alpha}_{m,i,k}^{(0,\pi,L)} = 0, \ m = 1, ..., d.$$

4. *For $n = 1$ to n_{max} or breaking criterion fulfilled*
   - $(\hat{Y}_\lambda^{(n-1,\pi)}(t_i))_{\lambda=1,...,L} := \sum_{k=1}^K \alpha_{i,k}^{(n-1,\pi,L)} \mathbb{e}_k^i$.
   - *For* $m = 1, ..., d$: $(\hat{Z}_{m,\lambda}^{(n-1,\pi)}(t_i))_{\lambda=1,...,L} := \sum_{k=1}^K \tilde{\alpha}_{m,i,k}^{(n-1,\pi,L)} \mathbb{e}_k^i$.
   - *For* $j \geq i$: $\mathbb{f}(t_j) := (f(t_j, X^{(\pi)}(t_j), \hat{Y}_\lambda^{(n-1,\pi)}(t_j), (\hat{Z}_{m,\lambda}^{(n-1,\pi)}(t_i))_{m=1}^d))_{\lambda=1}^L$.
   - $\alpha_{i,k}^{(n,\pi,L)} = \frac{1}{\sqrt{L}}(\mathcal{A}_i^L)^+(\mathbb{x}^{(\pi)} - \sum_{j=i}^{N-1} \mathbb{f}(t_j)\Delta t_{j+1})$.
   - $\tilde{\alpha}_{m,i,k}^{(n,\pi,L)} = \frac{1}{\sqrt{L}}(\mathcal{A}_i^L)^+(\frac{\Delta \mathbb{W}_{i+1}^m}{\Delta t_{i+1}} \cdot (\mathbb{x}^{(\pi)} - \sum_{j=i}^{N-1} \mathbb{f}(t_j)\Delta t_{j+1}))$, *where "$\cdot$" means componentwise multiplication.*
   - $\hat{Y}^{(n,\pi)}(t_i) = \sum_{k=1}^K \alpha_{i,k}^{(n-1,\pi,L)} \eta_k^i$
   - $\hat{Z}_m^{(n,\pi)}(t_i) = \sum_{k=1}^K \tilde{\alpha}_{m,i,k}^{(n-1,\pi,L)} \eta_k^i$

5. *Return the processes* $(\hat{Y}^{(n,\pi)}, (\hat{Z}_m^{(n,\pi)})_{m=1,...,d})$.

For a strict derivation of the algorithm, see Bender, Denk (2007) [10].

A.3 Locally Risk Minimizing Hedging

This section follows the paper of Schweizer (2001) [8]. All random variables are defined in a filtrated probability space $(\Omega, \mathcal{F}_T, (\mathcal{F}_t)_{t\in[0,T]}, P)$. Consider a financial market consisting of a riskless bank account S_0 and a (possibly multi-dimensional) risky asset S_1 which is supposed to be a semi-martingale. Without loss of generality we consider a discounted framework, i.e. $S_0 \equiv 1$ (otherwise devide every asset by S_0). Let H be an $\mathcal{F}_T$-measurable random variable representing some claim. In a complete market, we would now have an $\mathcal{F}_0$-measurable starting value $V(0)$ and a predictable self-financing strategy $\varphi = (\varphi_0, \varphi_1)$, such that V given by

$$V(\varphi, t) = V(\varphi, 0) + \int_0^t \varphi_1(s)\mathrm{d}S_1(s),$$

is a replicating strategy, i.e. $V(\varphi, T) = H$. Here, φ is completely determined by φ_1, since $\varphi_0(t) = V(\varphi, t) - \varphi_1(t)S_1(t)$ (in the discounted setting).

However in the incomplete case, there are claims H, so that such a replicating strategy does not exist. We either have to relax the self-financing condition or the replication condition. The former leads to so called "locally risk minimizing hedging", the latter to the so called "mean-variance hedging". We will use the locally risk minimizing hedging approach, i.e. we insist on the replication property $V(\varphi, T) = H$, but φ does no longer need to be self-financing. Instead, there will be an adapted process C, the so called **cost process**, s.t.

$$V(\varphi, t) = \int_0^t \varphi_1(s)\mathrm{d}S_1(s) + C(\varphi, t). \tag{57}$$

Note that in case of a complete market, C would be constant and equal to $V(\varphi, 0)$. The idea is now to minimize at any time-point $t \in [0, T]$ the second moment of the remaining cost up to maturity:

$$\min_\varphi E[(C(\varphi, T) - C(\varphi, t))^2|\mathcal{F}_t]. \tag{58}$$

This explains the name "risk minimizing hedging", although second moments are no risk measures. Since we later want to evaluate the risk due to this cost process, the name might be misleading, since the risk measured by the particular risk measure we use is not necessarily minimized (that would also be an interesting approach, but we do not pursue this ansatz here). However, if S_1 is not a martingale under P, then in general there exists no risk-minimizing strategy φ^*, so we have to weaken the concept. We start with some basic definitions.

Define $\mathbb{P} := \{Q \sim P : S_1 \text{ local martingale under } Q\}$ and assume $\mathbb{P} \neq \emptyset$ (which means that S_1 is a semimartingale under P). In an incomplete market, we have $|\mathbb{P}| > 1$. Furthermore, for simplicity assume, that S_1 is 1-dimensional (but everything can be generalized canonically to the multi-dimensional case).

Definition A.6 (Structure Condition).
Assume that $S_1 \in \mathcal{S}^2_{loc}(P)$, i.e. S_1 can be decomposed as

$$S_1 = S_1(0) + M + A, \tag{59}$$

where M is a square integrable local P-martingale with $M(0) = 0$ and A a predictable process of finite variation, $A(0) = 0$. Suppose that it exists a predictable process $\hat{\lambda}$, s.t.

$$A(s) = \int_0^s \hat{\lambda}(t)\, d\langle M, M\rangle_t, \;\; s \in [0, T], \tag{60}$$

such that the mean-variance tradeoff process

$$\hat{K}(s) := \int_0^s \hat{\lambda}^2(t)\, d\langle M, M\rangle_t, \tag{61}$$

*is finite P-a.s. Then we say that S_1 fulfills the **structure condition (SC)**.*

Remark A.5. *The structure condition is automatically satisfied if S_1 is continuous.*

Definition A.7 (L^2-Strategy).
A strategy φ, where $\int \varphi_1\, dS_1 \in \mathcal{S}^2$ such that V is right-continuous and $V(\varphi, t) \in L^2(P)$ for all $t \in [0, T]$ is called L^2-strategy.

Definition A.8 (Mean-Self-Financing Strategy).
A strategy φ is mean-self-financing, if its cost process $C(\varphi, \cdot)$ is a P-martingale.

The idea is now to define a strategy φ to be locally risk minimizing, if for small enough pertubations of the strategy, asymptotically the risk does not decrease, which matches with the concept of a local minimum of a function. The true definition is a little bit more complicated, see [8]. Unfortunately the assumption of a locally risk minimizing strategy is still too strong to calculate such a strategy explicitely. However it can be shown that under some weak conditions on S_1, a replicating L^2-strategy φ is locally risk minimizing if and only if φ is mean self-financing and $C(\varphi)$ is strongly orthogonal to M. This motivates the following definition:

Definition A.9 (Pseudo-Locally Risk-Minimizing Strategy).
*A replicating L^2-strategy φ is called **pseudo-locally risk-minimizing** or **pseudo-optimal**, if φ is mean self-financing and $C(\varphi)$ is strongly orthogonal to M.*

This leads straight forward to the following theorem.

Theorem A.6 (Pseudo-Optimality and Föllmer-Schweizer Decomposition)**.** *Let $H \in L^2(\mathcal{F}_T, P)$ be a contingent claim. Then H admits a pseudo-locally risk-minimizing strategy φ^* if and only if H can be represented as*

$$H = H_0 + \int_0^T \xi^H(t)dS_1(t) + L^H(T), \tag{62}$$

where $H_0 \in L^2(\mathcal{F}_0, P)$, ξ^H is such that $\int \xi^H dS_1 \in \mathcal{S}^2$ and L^H is a square-integrable martingale starting at 0 which is strongly orthogonal to M. The strategy φ^* is then determined by

$$\varphi_1^* = \xi^H \tag{63}$$

$$C(\varphi^*, \cdot) = H_0 + L^H. \tag{64}$$

Remark A.6. *Equation (62) is called "Föllmer-Schweizer decomposition" of H. One sufficient condition for existence is that the process $\hat{K}$ is bounded uniformly in (ω, t).*

The question is now how to calculate this decomposition. Therefore, define a new measure $\hat{P}$ by $\frac{\mathrm{d}\hat{P}}{\mathrm{d}P} = \hat{Z}(T)$ where $\hat{Z}$ is given by

$$\mathrm{d}\hat{Z}(t) = -\hat{Z}(t)\hat{\lambda}(t)\mathrm{d}M(t) \tag{65}$$

$$\hat{Z}(0) = 1, \tag{66}$$

i.e.

$$\begin{aligned} \hat{Z}(t) &= \exp(-\int_0^t \hat{\lambda}(u)\mathrm{d}M(u) - \frac{1}{2}\langle \int \hat{\lambda}\mathrm{d}M\rangle_t) \\ &= \exp(-\int_0^t \hat{\lambda}(u)\mathrm{d}M(u) - \frac{1}{2}\hat{K}(t)), \end{aligned} \tag{67}$$

and suppose that $\hat{Z}$ is strictly positive. This new measure $\hat{P}$ has four important properties:

1. By Girsanov's Theorem, S_1 is a local martingale under $\hat{P}$, i.e. $\hat{P} \in \mathbb{P}$.
2. $\int \xi \mathrm{d}S_1$ is a local $\hat{P}$-martingale.
3. For every local martingale L which is strongly orthogonal to M (the P-local-martingale part of S_1), L remains a local martingale under $\hat{P}$.
4. For a pseudo-optimal strategy φ^* for H, $V(\varphi^*, \cdot)$ is a local martingale under $\hat{P}$.

Due to the third property, we call $\hat{P}$ the **minimal equivalent local martingale measure** for S_1. Now after transforming into a martingale setting and supposing that S_1 is continuous, one can use the usual Kunita-Watanabe decomposition to get the following theorem:

Theorem A.7. *Under the conditions from above, it holds*

$$V(\varphi^*, t) = V^{H,\hat{P}}(t) := E_{\hat{P}}[H|\mathcal{F}_t], \tag{68}$$

with the Kunita-Watanabe decomposition

$$V^{H,\hat{P}}(t) = V^{H,\hat{P}}(0) + \int_0^t \xi^{H,\hat{P}}(u)dS_1(u) + L^{H,\hat{P}}(u), \tag{69}$$

which is also the Föllmer-Schweizer decomposition of $V(\varphi^*, t)$ ***under*** P*. Then,*

$$\varphi_1^* = \xi^{H,\hat{P}} \tag{70}$$

$$\varphi_0^* = V(\varphi^*, \cdot) - \xi^{H,\hat{P}} S_1. \tag{71}$$

B Additional Plots

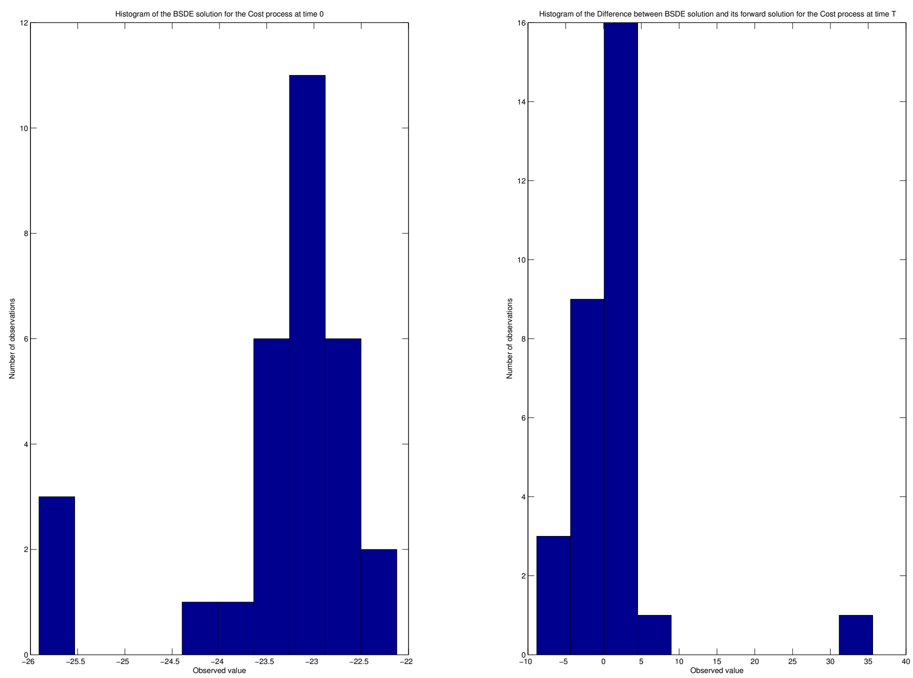

Figure 4: left: Histogram for BSDE solution at time 0 for the Hedging cost, right: Histogram for the difference between the BSDE solution and the corresponding forward solution at time T. The plot on the bottom shows an example of bad fitting, maybe because the maximal number of time steps was chosen too small.

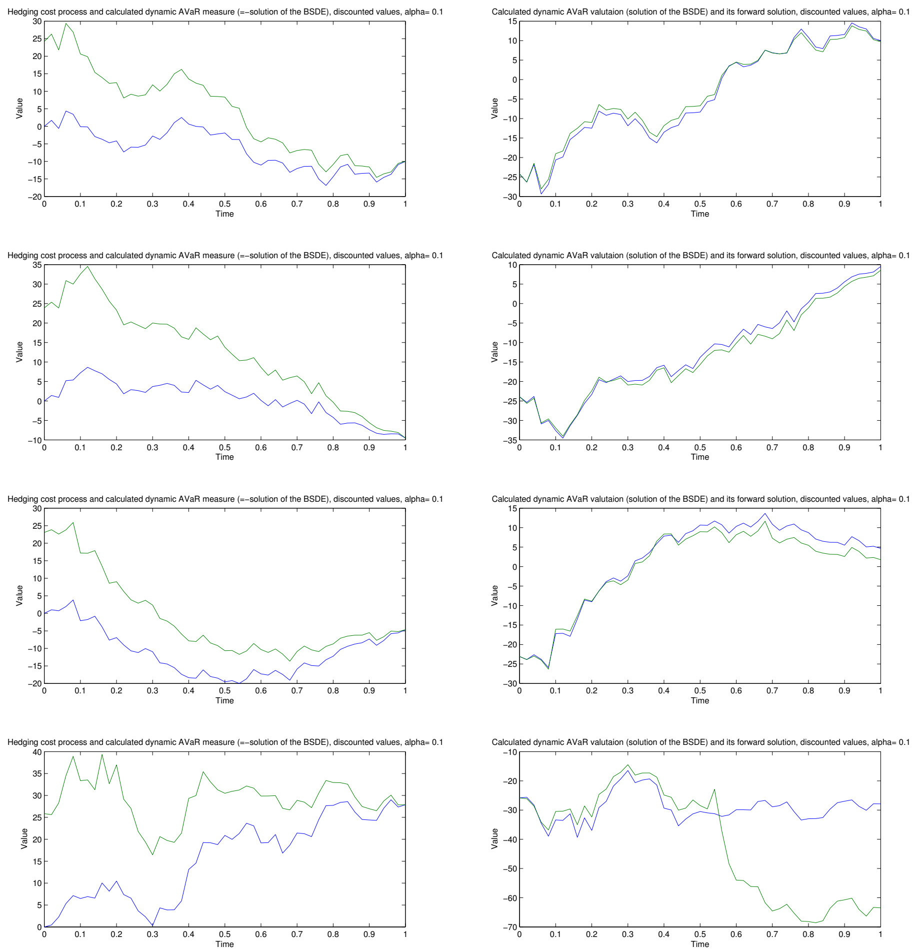

Figure 5: left side: Hedging cost process (blue) and calculated dynamic $AVaR_{0.1}$ (green), right side: calculated dynamic $AVaR_{0.1}$ (blue) and calculated forward solution (green), corresponding to the left hand side

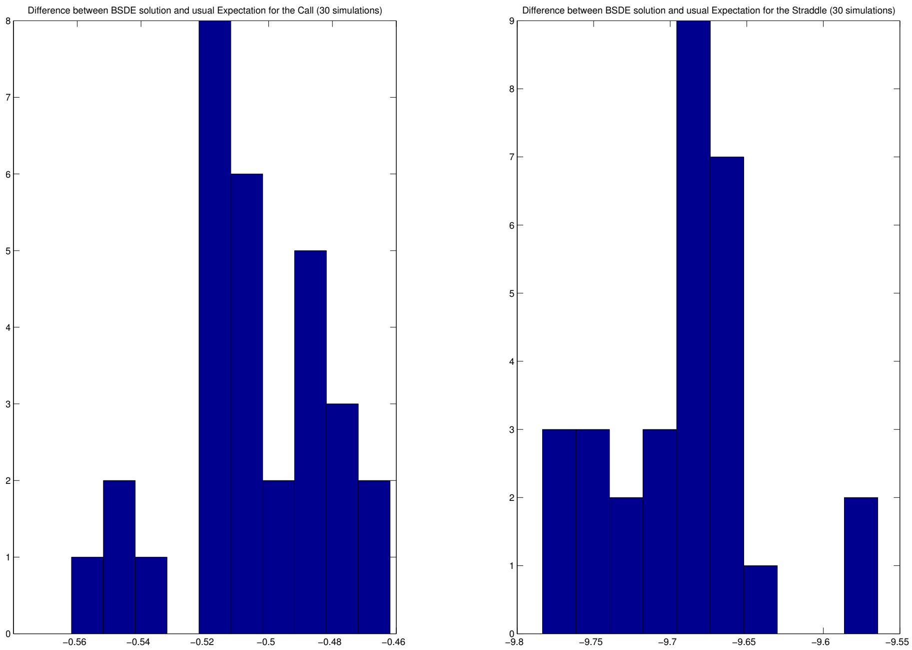

Figure 6: left: Histogram for the difference between the BSDE solution and the usual conditional expectation for the Call, right: Histogram for the difference between the BSDE solution and the usual conditional expectation for the Straddle

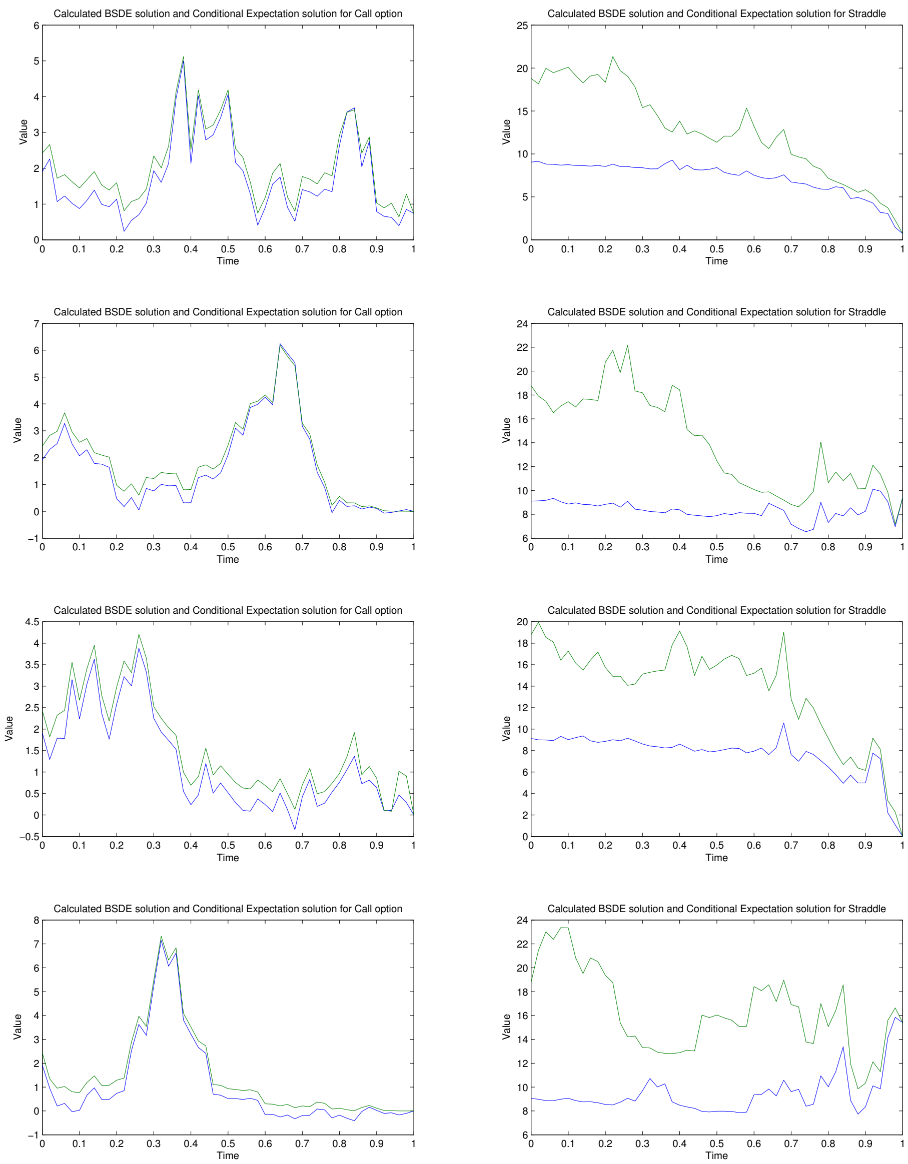

Figure 7: left side: BSDE-solutions (blue) and Expectation solutions (green) for the Call, right side: BSDE-solutions (blue) and Expectation solutions (green) for the Straddle (corresponding to the left hand side)

C Notation

$\circ$	Composition of functions.
$\sim$	Distributed according to
a.s.	Almost surely.
-msble	measurable.
r.v.	Random variable.
s.t.	Such that.
w.l.o.g.	Without loss of generality.
w.r.t.	With respect to.
$\stackrel{d}{=}$	Equality in distribution.
$\mathbb{1}_A$	Indicator function
$C_0([0,T])$	Space of continuous functions f on $[0,T]$, $f(0)=0$.
H^1	Cameron-Martin space.
$L^1_+(\mathcal{F})$	Space of nonnegative $\mathcal{F}$-measurable r.v.s.
$\mathcal{P}^{ac}_{2,P}$	The space probability measures $Q << P$ with $\frac{\mathrm{d}Q}{\mathrm{d}P} \in L^2(P)$.
Pr	Arbitrary probability measure.
$W(\cdot)$	Brownian Motion.
$\mathcal{N}(\mu,\sigma^2)$	Normal distribution with mean μ and variance σ^2.
Φ	Normal distribution function.

References

[1] Dilip Madan, Martijn Pistorius, Mitja Stadje: On consistent valuations based on distorted expectations: from multinomial random walks to Lévy Processes; arXiv 2013

[2] Mitja Stadje: Extending dynamic convex risk measures from discrete time to continuous time: A convergence approach; Insurance: Mathematics and Economics 2010

[3] Emanuela Rosazza Gianin: Risk measures via g-expectations; Insurance: Mathematics and Economics 2006

[4] Ioannis Karatzas, Steven E. Shreve: Methods of Mathematical Finance; Springer 1998

[5] Hans Föllmer, Alexander Schied: Stochastic Finance: An Introduction in Discrete Time; de Gruyter 2004

[6] Patrick Cheridito, Michael Kupper: Composition of Time-Consistent Dynamic Monetary Risk Measures in Discrete Time; 2010

[7] Hans Föllmer, Martin Schweizer: Hedging of Contingent Claims under Incomplete Information; Applied Stochastic Analysis, Stochastics Monographs, Vol. 5 (1990)

[8] Martin Schweizer: A Guided Tour through Quadratic Hedging Approaches; Cambridge University Press (2001)

[9] Giulia Di Nunno, Bernt Øksendal, Frank Proske: Malliavin Calculus for Lévy Processes with Applications in Finance; Springer 2009

[10] Christian Bender, Robert Denk: A forward scheme for backward SDEs; Stochastic Processes and their Applications 117 (2007) 1793-1812

[11] Zengjing Chen, Kun He, Reg Kulperger: Nonlinear expectations and nonlinear pricing

[12] E. Pardoux, S. G. Peng: Adapted solution of a backward stochastic differential equation; Systems & Control Letters 14 (1990)

[13] Nobuyuki Ikeda, Shinzo Watanabe: Stochastic Differential Equations and Diffusion Processes (2nd Edition); North Holland Publishing Company (1989)

[14] Patrick Cheridito, Mitja Stadje: BSΔEs and BSDEs with non-Lipschitz drivers: Comparison, convergence and robustness; arXiv 2013

[15] Beatrice Acciaio, Irina Penner: Dynamic risk measures; arXiv (2010)

[16] E. Serena, E. Bronshtein, S. Rachev, F. Fabozzi, W. Sun, S. Stoyanov: Distortion Risk Measures in Portfolio Optimization; Handbook of Portfolio Construction, Springer (2010)

[17] Philippe Briand, Bernard Delyon, Jean Mémin: On the robustnesss of backward stochastic differential equations; Stochastic Processes and their Applications 97 (2002) 229-253