

Projektbereich B
Discussion Paper No. B-247

**Variance-Optimal Hedging
in Discrete Time**

by

Martin Schweizer *)

February 1993

*) Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged.

(Mathematics of Operations Research 20 (1995), 1–32)

This version: 04.05.1995

Variance-Optimal Hedging in Discrete Time

*Martin Schweizer
Universität Göttingen
Institut für Mathematische Stochastik
Lotzestraße 13
D-37083 Göttingen
Germany*

Abstract: We solve the problem of approximating in \mathcal{L}^2 a given random variable H by stochastic integrals $G_T(\vartheta)$ of a given discrete-time process X . We interpret H as a contingent claim to be paid out at time T , X as the price evolution of some risky asset in a financial market, and $G(\vartheta)$ as the cumulative gains from trade using the hedging strategy ϑ . As an application, we determine the variance-optimal strategy which minimizes the variance of the net loss $H - G_T(\vartheta)$ over all strategies ϑ .

Key words: hedging, financial mathematics, contingent claims, mean-variance criterion, signed martingale measures

AMS 1980 subject classifications: 90A09, 49A60, 60K30

0. Introduction

Consider a stochastic process $X = (X_t)_{t \in \mathcal{T}}$ defined on a probability space (Ω, \mathcal{F}, P) and adapted to a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{T}}$, with a time index set $\mathcal{T} \subseteq [0, T]$ for some $T > 0$, and an \mathcal{F}_T -measurable random variable H . Think of X_t as the price at time t of some risky asset (e.g., a stock) in a financial market and of H as a *contingent claim*, i.e., a random loss suffered at time T by a small agent in this market. The standard example is given by a European call option written on X with expiration date T and strike price K , where $H = (X_T - K)^+$. One of the central problems in financial mathematics is the *pricing* and *hedging* of contingent claims by means of dynamic *trading strategies* based on X . Such a strategy is described by an \mathbb{F} -predictable process $\vartheta = (\vartheta_t)_{t \in \mathcal{T}}$, with ϑ_t representing the number of shares of X held by the agent at time t . Predictability of ϑ is a mathematical formulation of the obvious informational constraint that ϑ is not allowed to anticipate the movement of X . To any trading strategy ϑ then corresponds a process $G(\vartheta) = (G_t(\vartheta))_{t \in \mathcal{T}}$ of cumulative gains from trade. If c denotes the agent's initial capital, a good hedging strategy should thus have the property that the total gain $c + G_T(\vartheta)$ is in some sense as close to H as possible. Moreover, the initial capital c^* of an optimal pair (c^*, ϑ^*) provides a natural candidate for a price of H based on the chosen criterion of closeness.

In this paper, we solve this problem under three assumptions: frictionless trading, a quadratic criterion for measuring closeness, and discrete time. The first assumption implies that the gains process $G(\vartheta)$ of any *self-financing* trading strategy is given by the stochastic integral

$$G_t(\vartheta) = \int_0^t \vartheta_s dX_s \quad \text{for } t \in \mathcal{T};$$

see Harrison/Kreps (1979) and Harrison/Pliska (1981) for motivation and precise definitions. Since we are working in discrete time with $\mathcal{T} = \{0, 1, \dots, T\}$, the above integral is actually a finite sum of the form

$$G_t(\vartheta) = \sum_{j=1}^t \vartheta_j \Delta X_j \quad \text{for } t = 0, 1, \dots, T,$$

where $\Delta X_j := X_j - X_{j-1}$ denotes the increment of X at time j . We assume that each X_t is square-integrable, and we define the set Θ of admissible strategies to consist of all those \mathbb{F} -predictable processes ϑ such that $G(\vartheta)$ is square-integrable. Finally, using a quadratic criterion means that we want to solve, for fixed $c \in \mathbb{R}$ and $H \in \mathcal{L}^2(P)$, the optimization problem

$$(0.1) \quad \text{Minimize } E \left[(H - c - G_T(\vartheta))^2 \right] \text{ over all } \vartheta \in \Theta.$$

Put differently, we are looking for an approximation of H in $\mathcal{L}^2(P)$ by stochastic integrals of X .

Unfortunately, it turns out that (0.1) has no solution in general; this is shown by an explicit counterexample in section 5. We therefore introduce in section 1 a nondegeneracy condition on X . We assume that X has a *bounded mean-variance tradeoff* in the sense that

$$(0.2) \quad \frac{(E[\Delta X_t | \mathcal{F}_{t-1}])^2}{\text{Var}[\Delta X_t | \mathcal{F}_{t-1}]} \text{ is } P\text{-a.s. bounded, uniformly in } t \text{ and } \omega.$$

Under the much stronger assumption that this ratio is *deterministic*, (0.1) was previously solved by Schäl (1994). For an informal discussion of (0.1), see also Hipp (1993). In the present paper, we prove the existence of an optimal strategy $\xi^{(c)}$ for (0.1) under the assumption (0.2). In fact, we prove in section 2 that (0.2) implies that $G_T(\Theta)$ is closed in $\mathcal{L}^2(P)$. Moreover, this result is sharp; a counterexample in section 5 shows that condition (0.2) is in general indispensable. Having established existence, we then proceed to analyze the structure of $\xi^{(c)}$ in more detail. The main tool for this purpose is provided by the *adjustment process* $\beta = (\beta_t)_{t=1, \dots, T}$ of X ; this process is defined in section 2 and studied in detail in Schweizer (1993a). A first structural result on $\xi^{(c)}$ is given in Theorem 2.4 where we obtain an expression for $\xi^{(c)}$ in feedback form. As a corollary, we deduce explicit formulae for the mean and the variance of the net loss $H - c - G_T(\xi^{(c)})$ under the optimal strategy. The second structural result is Theorem 2.8 which links $\xi^{(c)}$ to the locally risk-minimizing strategy for H introduced in Schweizer (1988). This yields a *decomposition* of $\xi^{(c)}$ into three components, namely a pure hedging demand, a demand for mean-variance purposes and a demand for hedging against stochastic fluctuations in the mean-variance ratio (0.2). In particular, this extends previous results by Schäl (1994) and Schweizer (1992).

Section 3 contains applications to several optimization problems with quadratic criteria. We give explicit expressions for the optimal choice of strategy and initial capital, and we determine the strategy which minimizes the variance of the net loss $H - G_T(\vartheta)$, both with and without the constraint of a fixed mean. In section 4, we discuss the simplification of our results in special cases. If X is a *martingale*, problem (0.1) reduces to a simple application of the Kunita-Watanabe projection theorem. If H is *attainable* in the sense that it can be written as the sum of a constant and a (discrete-time) stochastic integral of X , we provide a closed-form expression for $\xi^{(c)}$ in terms of the adjustment process β and the integrand in the representation of H . Finally we examine the case where the mean-variance tradeoff in (0.2) is *deterministic*. Under this assumption, we rederive the results of Schäl (1994) as simple consequences of our approach, and we also provide an alternative proof which can be generalized to a continuous-time framework. Section 5 concludes the paper with some explicit examples and counterexamples.

1. The basic problem

Let (Ω, \mathcal{F}, P) be a probability space, $T > 0$ a fixed natural number and $\mathbb{F} = (\mathcal{F}_k)_{k=0,1, \dots, T}$ a filtration, i.e., an increasing family of sub- σ -algebras of \mathcal{F} . We shall assume that $\mathcal{F} = \mathcal{F}_T$. Let $X = (X_k)_{k=0,1, \dots, T}$ be a real-valued, \mathbb{F} -adapted, square-integrable process, i.e., each X_k is \mathcal{F}_k -measurable and in $\mathcal{L}^2(P)$. We denote by

$$\Delta X_k := X_k - X_{k-1} \quad \text{for } k = 1, \dots, T$$

the increments of X . A process $\vartheta = (\vartheta_k)_{k=1, \dots, T}$ is called *predictable* if ϑ_k is \mathcal{F}_{k-1} -measurable for each k .

Definition. We denote by Θ the set of all predictable processes ϑ such that $\vartheta_k \Delta X_k \in \mathcal{L}^2(P)$ for $k = 1, \dots, T$. For $\vartheta \in \Theta$, $G(\vartheta)$ is the process defined by

$$G_k(\vartheta) := \sum_{j=1}^k \vartheta_j \Delta X_j \quad \text{for } k = 0, 1, \dots, T.$$

We shall use throughout the conventions that a sum over an empty set is defined to be 0, a product over an empty set is defined to be 1, and $\frac{0}{0} = 0$.

The *basic problem* addressed in this paper is the following:

$$(1.1) \quad \begin{aligned} &\text{Given } H \in \mathcal{L}^2(P) \text{ and } c \in \mathbb{R}, \\ &\text{minimize } E \left[(H - c - G_T(\vartheta))^2 \right] \text{ over all } \vartheta \in \Theta. \end{aligned}$$

Interpretation. Think of X_k as the (discounted) *price* at time k of some risky asset (e.g., a stock) in a financial market. The process ϑ describes the *trading strategy* of a small agent in this market, where “small” means that his actions do not influence prices. The random variable ϑ_k is thus interpreted as the number of shares held during the time interval $(k-1, k]$, and predictability is imposed so that the decision about the choice of ϑ_k at time $k-1$ must be made without exact knowledge of the evolution of X in the next time interval. If we also assume the existence of some riskless asset whose discounted value is 1 at all times, then any ϑ determines a unique *self-financing* trading strategy, and $G(\vartheta)$ describes the process of (discounted) *gains* from trade. Roughly speaking, “self-financing” means that a strategy neither requires nor generates funds between dates 0 and T . For a detailed exposition, see Harrison/Kreps (1979), Harrison/Pliska (1981) or Duffie (1988). If we now interpret the *contingent claim* H as a random loss suffered by the agent at time T , then $H - c - G_T(\vartheta)$ is the agent’s net loss if he starts with *initial capital* c and follows the strategy ϑ . As an example, H could be a financial obligation resulting from the sale of some financial instrument with expiration date T . A European call option on X with strike price K , for instance, would lead to $H = (X_T - K)^+$. The goal in the basic problem is then to minimize the expected net square loss by choosing a trading strategy.

Remark. As pointed out by the referees, the assumption of frictionless trading is crucial here. Any inclusion of transaction costs would destroy the linearity of the gains $G(\vartheta)$ in ϑ and thus make (1.1) intractable in the present generality. It would be an interesting and challenging problem to find processes X for which (1.1) can also be solved in the presence of transaction costs, but this is beyond the scope of the present paper.

Without additional assumptions on X , we obtain in general neither uniqueness nor existence of a solution to (1.1). Nonuniqueness obviously prevails for instance if X is constant, and an explicit example in section 5 shows that nonexistence can also occur. In the sequel, we shall be concerned with the existence and structure of an optimal strategy for (1.1). To this end, we introduce the following condition on X :

Definition. We say that X satisfies the *nondegeneracy condition* (ND) if there exists a constant $\delta \in (0, 1)$ such that

$$(E[\Delta X_k | \mathcal{F}_{k-1}])^2 \leq \delta E[\Delta X_k^2 | \mathcal{F}_{k-1}] \quad P\text{-a.s. for } k = 1, \dots, T.$$

Note that by Jensen’s inequality, we always have

$$(E[\Delta X_k | \mathcal{F}_{k-1}])^2 \leq E[\Delta X_k^2 | \mathcal{F}_{k-1}] \quad P\text{-a.s.}$$

The point of condition (ND) is to ensure a strict inequality uniformly in ω .

To obtain other formulations of (ND), we now write X in its *Doob decomposition* as

$$X = X_0 + M + A,$$

where $M = (M_k)_{k=0,1,\dots,T}$ is a square-integrable (P, \mathcal{F}) -martingale with $M_0 = 0$ and $A = (A_k)_{k=0,1,\dots,T}$ is a square-integrable predictable process with $A_0 = 0$. It is well known that this decomposition is unique and given by

$$\begin{aligned}\Delta A_k &:= E[\Delta X_k | \mathcal{F}_{k-1}], \\ \Delta M_k &:= \Delta X_k - \Delta A_k.\end{aligned}$$

The predictable process $\tilde{\lambda} = (\tilde{\lambda}_k)_{k=1,\dots,T}$ is then defined by

$$(1.2) \quad \tilde{\lambda}_k := \frac{\Delta A_k}{E[\Delta X_k^2 | \mathcal{F}_{k-1}]} \quad \text{for } k = 1, \dots, T;$$

by our conventions, $\tilde{\lambda}$ is well-defined due to Jensen's inequality. Finally we recall the notation

$$\text{Var}[\Delta X_k | \mathcal{F}_{k-1}] := E[\Delta X_k^2 | \mathcal{F}_{k-1}] - (E[\Delta X_k | \mathcal{F}_{k-1}])^2$$

for the conditional variance of ΔX_k given \mathcal{F}_{k-1} and note that

$$\text{Var}[\Delta X_k | \mathcal{F}_{k-1}] = E[\Delta M_k^2 | \mathcal{F}_{k-1}] \quad P\text{-a.s.}$$

Definition. The process $\widehat{K} = (\widehat{K}_j)_{j=0,1,\dots,T}$ defined by

$$\widehat{K}_j := \sum_{\ell=1}^j \frac{(E[\Delta X_\ell | \mathcal{F}_{\ell-1}])^2}{\text{Var}[\Delta X_\ell | \mathcal{F}_{\ell-1}]} \quad \text{for } j = 0, 1, \dots, T$$

is called the *mean-variance tradeoff process* of X .

Using the above definitions, it is straightforward to check that (ND) is equivalent to each of the following conditions:

$$(1.3) \quad \tilde{\lambda}_k \Delta A_k \leq \delta \quad P\text{-a.s. for } k = 1, \dots, T, \text{ with a constant } \delta \in (0, 1).$$

$$(1.4) \quad \text{The process } \widehat{K} \text{ is } P\text{-a.s. bounded, uniformly in } \omega \text{ and } j.$$

$$(1.5) \quad E[\Delta M_k^2 | \mathcal{F}_{k-1}] \geq (1 - \delta) E[\Delta X_k^2 | \mathcal{F}_{k-1}] \quad P\text{-a.s. for } k = 1, \dots, T, \\ \text{with a constant } \delta \in (0, 1).$$

$$(1.6) \quad \frac{(E[\Delta X_k | \mathcal{F}_{k-1}])^2}{\text{Var}[\Delta X_k | \mathcal{F}_{k-1}]} \text{ is } P\text{-a.s. bounded, uniformly in } \omega \text{ and } k.$$

Condition (ND) was introduced by Schweizer (1988) and also used by Schäl (1994) in the equivalent form (1.6). In a continuous-time version of (1.1), a condition of the form (1.4)

plays an important role; see Schweizer (1993c). The term “nondegeneracy” is explained by the equivalent formulation of (ND) that

$$E [\Delta M_k^2 | \mathcal{F}_{k-1}] \geq L \Delta A_k^2 \quad P\text{-a.s. for } k = 1, \dots, T, \text{ with a constant } L < \infty,$$

which intuitively states that on each time interval $(k-1, k]$, any drift A of X must be counterbalanced by a sufficiently strong diffusive behaviour of M . For a similar condition in a continuous-time framework, see Schweizer (1991).

Example 1. Suppose that X is an *event tree* in the sense that X_0 is a constant, each X_k takes only finitely many values and \mathcal{F} is the filtration generated by X (i.e., $\mathcal{F}_k = \sigma(X_0, X_1, \dots, X_k)$ for $k = 0, 1, \dots, T$). This corresponds to a situation where at each time for each price, there are only finitely many possible values for the next price. Intuitively, an event tree may be pictured as a graph whose nodes are given by the date-price pairs $(k, X_k(\omega))$; see section 5 for a graphical illustration. Call X *nondegenerate* if for each k , the conditional distribution of X_k given \mathcal{F}_{k-1} is P -a.s. not concentrated in one point. This means that from each node, there are at least two branches going to the right. Then it is easy to see that every nondegenerate event tree X satisfies (ND). In fact, we may take Ω finite without loss of generality, and since nondegeneracy of X implies

$$\text{Var} [\Delta X_k | \mathcal{F}_{k-1}] (\omega) > 0,$$

hence

$$(E [\Delta X_k | \mathcal{F}_{k-1}] (\omega))^2 < E [\Delta X_k^2 | \mathcal{F}_{k-1}] (\omega)$$

for each ω and each k , we can choose $\delta \in (\delta', 1)$ with

$$\delta' := \sup_{k, \omega} \frac{(E [\Delta X_k | \mathcal{F}_{k-1}] (\omega))^2}{E [\Delta X_k^2 | \mathcal{F}_{k-1}] (\omega)} < 1,$$

since both k and ω run through a finite set only.

2. Existence and structure of an optimal strategy

The basic result for proving the existence of a solution for (1.1) is

Theorem 2.1. *If X satisfies (ND), then $G_T(\Theta)$ is closed in $\mathcal{L}^2(P)$.*

Proof. Let $(\vartheta^n)_{n \in \mathbb{N}}$ be any sequence in Θ such that $(G_T(\vartheta^n))_{n \in \mathbb{N}}$ is a Cauchy sequence in $\mathcal{L}^2(P)$, and denote by Y the limit in $\mathcal{L}^2(P)$ of $(G_T(\vartheta^n))$. For any $\vartheta \in \Theta$,

$$(2.1) \quad G_T(\vartheta) = (G_{T-1}(\vartheta) + \vartheta_T \Delta A_T) + \vartheta_T \Delta M_T;$$

since ϑ, A are predictable and M is a martingale, the two terms on the right-hand side of (2.1) are orthogonal in $\mathcal{L}^2(P)$, and therefore

$$E [(G_T(\vartheta))^2] = E [(G_{T-1}(\vartheta) + \vartheta_T \Delta A_T)^2] + E [(\vartheta_T)^2 E [\Delta M_T^2 | \mathcal{F}_{T-1}]].$$

But this implies that the sequence of \mathcal{F}_{T-1} -measurable random variables

$$\psi_T^n := \vartheta_T^n \sqrt{E [\Delta M_T^2 | \mathcal{F}_{T-1}]}$$

is also a Cauchy sequence in $\mathcal{L}^2(P)$, hence convergent in $\mathcal{L}^2(P)$ to some ϑ_T^∞ which is again \mathcal{F}_{T-1} -measurable. Set

$$\vartheta_T^\infty := I_{\{E[\Delta M_T^2 | \mathcal{F}_{T-1}] > 0\}} \frac{\psi_T^\infty}{\sqrt{E[\Delta M_T^2 | \mathcal{F}_{T-1}]}}.$$

Since X satisfies (ND), we have by (1.5)

$$\begin{aligned} E \left[(\vartheta_T^\infty \Delta X_T - \vartheta_T^n \Delta X_T)^2 \right] &= E \left[(\vartheta_T^\infty - \vartheta_T^n)^2 E[\Delta X_T^2 | \mathcal{F}_{T-1}] \right] \\ &\leq \frac{1}{1-\delta} E \left[(\vartheta_T^\infty - \vartheta_T^n)^2 E[\Delta M_T^2 | \mathcal{F}_{T-1}] \right], \end{aligned}$$

and the last term tends to 0 as $n \rightarrow \infty$ by the definition of ϑ_T^∞ . This shows that $\vartheta_T^\infty \Delta X_T$ is in $\mathcal{L}^2(P)$, since $\vartheta^n \in \Theta$, and that

$$\vartheta_T^n \Delta X_T \longrightarrow \vartheta_T^\infty \Delta X_T \quad \text{in } \mathcal{L}^2(P) \text{ as } n \rightarrow \infty.$$

Hence $G_{T-1}(\vartheta^n)$ converges to $Y - \vartheta_T^\infty \Delta X_T$ in $\mathcal{L}^2(P)$ as $n \rightarrow \infty$, and by iterating the preceding arguments we conclude that

$$Y = \sum_{k=1}^T \vartheta_k^\infty \Delta X_k \quad P\text{-a.s.}$$

for some $\vartheta^\infty \in \Theta$. Thus $G_T(\Theta)$ is indeed closed in $\mathcal{L}^2(P)$.

q.e.d.

Remark. It is natural to ask if $G_T(\Theta)$ is still closed in $\mathcal{L}^2(P)$ if one abandons the assumption that X satisfies (ND). A counterexample due to W. Schachermayer shows that the answer is negative in general; the same counterexample also shows that (1.1) will not have a solution in general. For a detailed account, see section 5.

Theorem 2.2. *Suppose that X satisfies (ND). For any $H \in \mathcal{L}^2(P)$ and any $c \in \mathbb{R}$, there exists a strategy $\xi^{(c)} \in \Theta$ which solves (1.1).*

Proof. By Theorem 2.1, $G_T(\Theta)$ is a closed linear subspace of the Hilbert space $\mathcal{L}^2(P)$; hence we can project $H - c$ on $G_T(\Theta)$.

q.e.d.

Let us pause here a moment to emphasize the generality of Theorem 2.2. Under the sole assumption that X satisfies (ND), we can solve (1.1) for any contingent claim H in $\mathcal{L}^2(P)$. Previous work by Schäl (1994) relied crucially on the additional assumption that the mean-variance tradeoff process \widehat{K} is deterministic. Moreover, Theorem 2.2 is essentially the best possible result: the counterexample in section 5 shows that condition (ND) is in general indispensable. Notice also that our existence argument does not require any structural assumptions on X or H , in contrast to the corresponding continuous-time problem studied in Duffie/Richardson (1991), Schweizer (1992) and Schweizer (1993b).

In order to describe the structure of the optimal strategy $\xi^{(e)}$ in more detail, we first recall from Schweizer (1993a) the definition of the *adjustment process* associated to X . This is the predictable process $\beta = (\beta_k)_{k=1, \dots, T}$ defined by

$$(2.2) \quad \beta_k := \frac{E \left[\Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right]}{E \left[\Delta X_k^2 \prod_{j=k+1}^T (1 - \beta_j \Delta X_j)^2 \middle| \mathcal{F}_{k-1} \right]} \quad \text{for } k = 1, \dots, T.$$

The next result summarizes those properties of β we shall use in the sequel; see Schweizer (1993a) for a proof.

Proposition 2.3. *β is well-defined by (2.2) and has the property that for $k = 1, \dots, T$*

$$(2.3) \quad \prod_{j=k}^T (1 - \beta_j \Delta X_j) \in \mathcal{L}^2(P),$$

$$(2.4) \quad \Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \in \mathcal{L}^2(P),$$

$$(2.5) \quad \beta_k \Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \in \mathcal{L}^2(P)$$

and

$$(2.6) \quad E \left[\prod_{j=k}^T (1 - \beta_j \Delta X_j)^2 \middle| \mathcal{F}_{k-1} \right] = E \left[\prod_{j=k}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right] \leq 1 \quad P\text{-a.s.}$$

In particular, the random variable

$$(2.7) \quad \tilde{Z}^0 := \prod_{j=1}^T (1 - \beta_j \Delta X_j)$$

is in $\mathcal{L}^2(P)$ and satisfies

$$(2.8) \quad 0 \leq E[\tilde{Z}^0] = E[(\tilde{Z}^0)^2] \leq 1.$$

Furthermore, \tilde{Z}^0 has the property that

$$(2.9) \quad E \left[\tilde{Z}^0 G_T(\vartheta) \right] = 0 \quad \text{for all } \vartheta \in \Theta$$

or equivalently

$$(2.10) \quad E \left[\tilde{Z}^0 \Delta X_k \middle| \mathcal{F}_{k-1} \right] = 0 \quad P\text{-a.s. for } k = 1, \dots, T.$$

Remark. Our subsequent arguments rely heavily on computations involving conditional expectations. We should like to point out here that the integrability properties of β summarized in (2.3) – (2.5) allow us to verify the existence of all these conditional expectations, and to justify rigorously all manipulations below. For a detailed example of the type of reasoning to be used, we refer to the proofs in Schweizer (1993a).

Throughout the rest of this section, we shall assume that X satisfies (ND), and we fix a contingent claim $H \in \mathcal{L}^2(P)$ and a constant $c \in \mathbb{R}$. In order to elucidate the structure of the corresponding optimal strategy $\xi^{(c)}$, we introduce the predictable process $\varrho = (\varrho_k)_{k=1, \dots, T}$ defined by

$$(2.11) \quad \varrho_k := \frac{E \left[H \Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right]}{E \left[\Delta X_k^2 \prod_{j=k+1}^T (1 - \beta_j \Delta X_j)^2 \middle| \mathcal{F}_{k-1} \right]} \quad \text{for } k = 1, \dots, T.$$

Note that ϱ is well-defined due to (2.4) and the Cauchy-Schwarz inequality, and that

$$(2.12) \quad E \left[H \beta_k \Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right] = E \left[\varrho_k \Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right]$$

P -a.s. for each k by (2.2) and (2.11).

Theorem 2.4. For $k = 1, \dots, T$, we have P -a.s.

$$(2.13) \quad \begin{aligned} & H - c - G_T(\xi^{(c)}) \\ &= H - \sum_{j=k}^T \varrho_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) - \left(c + G_{k-1}(\xi^{(c)}) \right) \prod_{\ell=k}^T (1 - \beta_\ell \Delta X_\ell) \end{aligned}$$

and

$$(2.14) \quad \xi_k^{(c)} = \varrho_k - \beta_k \left(c + G_{k-1}(\xi^{(c)}) \right).$$

Proof. We show (2.13) and (2.14) simultaneously by backward induction. By the projection theorem (see for instance Luenberger (1969), Theorem 3.3.1), a strategy $\xi \in \Theta$ solves (1.1) if and only if

$$E \left[(H - c - G_T(\xi)) G_T(\vartheta) \right] = 0 \quad \text{for all } \vartheta \in \Theta$$

or equivalently

$$(2.15) \quad E \left[(H - c - G_T(\xi)) \Delta X_k \middle| \mathcal{F}_{k-1} \right] = 0 \quad P\text{-a.s. for } k = 1, \dots, T.$$

Since $\xi^{(c)}$ is in Θ and optimal, we obtain for $k = T$

$$\begin{aligned} 0 &= E \left[\left(H - c - G_T(\xi^{(c)}) \right) \Delta X_T \middle| \mathcal{F}_{T-1} \right] \\ &= E \left[H \Delta X_T - \xi_T^{(c)} \Delta X_T^2 - \left(c + G_{T-1}(\xi^{(c)}) \right) \Delta X_T \middle| \mathcal{F}_{T-1} \right] \end{aligned}$$

and therefore (2.14) with $k = T$ by (2.2) and (2.11). This yields in turn

$$\begin{aligned} H - c - G_T(\xi^{(c)}) &= H - \xi_T^{(c)} \Delta X_T - \left(c + G_{T-1}(\xi^{(c)}) \right) \\ &= H - \varrho_T \Delta X_T - \left(c + G_{T-1}(\xi^{(c)}) \right) (1 - \beta_T \Delta X_T) \end{aligned}$$

which is (2.13) for $k = T$. Suppose now that (2.13) and (2.14) hold for $j = k + 1, \dots, T$. Then (2.15) implies

$$\begin{aligned} 0 &= E \left[\left(H - c - G_T(\xi^{(c)}) \right) \Delta X_k \middle| \mathcal{F}_{k-1} \right] \\ &= E \left[\Delta X_k \left(H - \sum_{j=k+1}^T \varrho_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) \right) \middle| \mathcal{F}_{k-1} \right] \\ &\quad - E \left[\Delta X_k \left(c + G_k(\xi^{(c)}) \right) \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{k-1} \right] \\ &= E[H \Delta X_k | \mathcal{F}_{k-1}] - \sum_{j=k+1}^T E \left[\Delta X_k \varrho_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{k-1} \right] \\ &\quad - \xi_k^{(c)} E \left[\Delta X_k^2 \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{k-1} \right] \\ &\quad - \left(c + G_{k-1}(\xi^{(c)}) \right) E \left[\Delta X_k \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{k-1} \right] \\ &= E \left[H \Delta X_k \left(1 - \sum_{j=k+1}^T \beta_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) \right) \middle| \mathcal{F}_{k-1} \right] \\ &\quad - \xi_k^{(c)} E \left[\Delta X_k^2 \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{k-1} \right] \\ &\quad - \left(c + G_{k-1}(\xi^{(c)}) \right) E \left[\Delta X_k \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{k-1} \right] \\ &= E \left[H \Delta X_k \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{k-1} \right] - \xi_k^{(c)} E \left[\Delta X_k^2 \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{k-1} \right] \\ &\quad - \left(c + G_{k-1}(\xi^{(c)}) \right) E \left[\Delta X_k \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{k-1} \right] \end{aligned}$$

by (2.13) with $k + 1$ instead of k , (2.12) with j instead of k (after conditioning on $\mathcal{F}_{j-1} \supseteq \mathcal{F}_k$ for each summand) and the identity

$$(2.16) \quad \prod_{\ell=k+1}^T (1 - c_\ell) = 1 - \sum_{j=k+1}^T c_j \prod_{\ell=j+1}^T (1 - c_\ell).$$

Using (2.6), (2.2) and (2.11), we obtain (2.14) for k . This implies

$$\begin{aligned} c + G_k(\xi^{(c)}) &= c + G_{k-1}(\xi^{(c)}) + \xi_k^{(c)} \Delta X_k \\ &= \varrho_k \Delta X_k + \left(c + G_{k-1}(\xi^{(c)}) \right) (1 - \beta_k \Delta X_k) \end{aligned}$$

and therefore by (2.13) for $k + 1$ instead of k

$$\begin{aligned} H - c - G_T(\xi^{(c)}) &= H - \sum_{j=k+1}^T \varrho_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) - \left(c + G_k(\xi^{(c)}) \right) \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \\ &= H - \sum_{j=k}^T \varrho_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) - \left(c + G_{k-1}(\xi^{(c)}) \right) \prod_{\ell=k}^T (1 - \beta_\ell \Delta X_\ell) \end{aligned}$$

which is (2.13) for k .

q.e.d.

Corollary 2.5. *We have*

$$(2.17) \quad E \left[H - c - G_T(\xi^{(c)}) \right] = E \left[(H - c) \prod_{j=1}^T (1 - \beta_j \Delta X_j) \right] = E[H\tilde{Z}^0] - cE[\tilde{Z}^0]$$

and

$$(2.18) \quad E \left[\left(H - c - G_T(\xi^{(c)}) \right)^2 \right] = c^2 E[\tilde{Z}^0] - 2cE[H\tilde{Z}^0] + E \left[\left(H - \sum_{j=1}^T \varrho_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) \right)^2 \right].$$

Proof. By (2.13) with $k = 1$, we have

$$\begin{aligned} E \left[H - c - G_T(\xi^{(c)}) \right] &= E \left[H - \sum_{j=1}^T \varrho_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) - c \prod_{\ell=1}^T (1 - \beta_\ell \Delta X_\ell) \right] \\ &= E \left[H \left(1 - \sum_{j=1}^T \beta_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) \right) - c \prod_{\ell=1}^T (1 - \beta_\ell \Delta X_\ell) \right] \end{aligned}$$

from (2.12) after conditioning the j -th summand on \mathcal{F}_{j-1} . Using (2.16) and (2.7) then gives (2.17). Again by (2.13) with $k = 1$,

$$\begin{aligned} E \left[\left(H - c - G_T(\xi^{(c)}) \right)^2 \right] &= c^2 E \left[\prod_{j=1}^T (1 - \beta_j \Delta X_j)^2 \right] + E \left[\left(H - \sum_{j=1}^T \varrho_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) \right)^2 \right] \\ &\quad - 2cE \left[\left(H - \sum_{j=1}^T \varrho_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) \right) \prod_{k=1}^T (1 - \beta_k \Delta X_k) \right]. \end{aligned}$$

But the first term equals $c^2 E[\tilde{Z}^0]$ by (2.8), and for each j ,

$$\begin{aligned} & E \left[\varrho_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) \prod_{k=1}^T (1 - \beta_k \Delta X_k) \middle| \mathcal{F}_{j-1} \right] \\ &= E \left[\Delta X_j (1 - \beta_j \Delta X_j) \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell)^2 \middle| \mathcal{F}_{j-1} \right] \varrho_j \prod_{k=1}^{j-1} (1 - \beta_k \Delta X_k) \\ &= \varrho_j E \left[\Delta X_j \prod_{k=1}^T (1 - \beta_k \Delta X_k) \middle| \mathcal{F}_{j-1} \right] = 0 \quad P\text{-a.s.} \end{aligned}$$

by (2.6) (after conditioning on \mathcal{F}_j) and (2.10). Thus the third term equals $-2cE[H\tilde{Z}^0]$ by (2.7), and this proves (2.18).

q.e.d.

Remark. Theorem 2.4 suggests an alternative way to solve (1.1): we could define a predictable process $\xi = (\xi_k)_{k=1, \dots, T}$ recursively by

$$(2.19) \quad \xi_k = \varrho_k - \beta_k (c + G_{k-1}(\xi))$$

and then try to show that ξ is optimal. *If* (and this is the crucial point) ξ is in Θ , then it is not too hard to show that ξ solves (1.1). One first proves by induction as in Theorem 2.4 that

$$H - c - G_T(\xi) = H - \sum_{j=1}^T \varrho_j \Delta X_j \prod_{\ell=j+1}^T (1 - \beta_\ell \Delta X_\ell) - c \prod_{\ell=1}^T (1 - \beta_\ell \Delta X_\ell).$$

A similar argument as in the proof of Corollary 2.5 then shows by using (2.10), (2.12) and (2.16) that

$$E[(H - c - G_T(\xi))\Delta X_k | \mathcal{F}_{k-1}] = 0 \quad P\text{-a.s. for } k = 1, \dots, T$$

which implies optimality as above. However, we have so far not been able to prove that (2.19) automatically implies $\xi \in \Theta$. In the special case where the mean-variance tradeoff process \hat{K} is deterministic, this is indeed true, as was shown by Schäl (1994); see also section 4. In general, however, we do not know if (1.1) can be solved by this approach.

We can obtain additional information on the structure of $\xi^{(c)}$ if we examine H in more detail. The following result essentially goes back to Schweizer (1988); see Schäl (1994) or Schweizer (1993b) for a proof. Recall the assumption that X satisfies (ND).

Proposition 2.6. *Every contingent claim $H \in \mathcal{L}^2(P)$ admits a decomposition*

$$(2.20) \quad H = H_0 + \sum_{j=1}^T \xi_j^H \Delta X_j + L_T^H \quad P\text{-a.s.},$$

where H_0 is a constant, $\xi^H \in \Theta$ and $L^H = (L_k^H)_{k=0,1, \dots, T}$ is a square-integrable (P, \mathbb{F}) -martingale with $E[L_0^H] = 0$ which is strongly orthogonal to M (i.e., the product $L^H M$ is a

(P, \mathbb{F})-martingale null at 0). The processes ξ^H and L^H can be obtained as

$$\xi_k^H := \frac{\text{Cov} \left(H - \sum_{j=k+1}^T \xi_j^H \Delta X_j, \Delta X_k \middle| \mathcal{F}_{k-1} \right)}{\text{Var}[\Delta X_k | \mathcal{F}_{k-1}]} \quad \text{for } k = 1, \dots, T$$

(with the obvious notation for the conditional covariance) and

$$L_k^H := E \left[H - \sum_{j=1}^T \xi_j^H \Delta X_j \middle| \mathcal{F}_k \right] - E \left[H - \sum_{j=1}^T \xi_j^H \Delta X_j \right] \quad \text{for } k = 0, \dots, T.$$

The constant H_0 in (2.20) can be obtained explicitly. To that end, we define the process $\widehat{Z} = (\widehat{Z}_k)_{k=0,1,\dots,T}$ by

$$(2.21) \quad \widehat{Z}_k := \prod_{j=1}^k \frac{1 - \widetilde{\lambda}_j \Delta X_j}{1 - \widetilde{\lambda}_j \Delta A_j} = \prod_{j=1}^k \left(1 - \frac{\widetilde{\lambda}_j}{1 - \widetilde{\lambda}_j \Delta A_j} \Delta M_j \right).$$

Since X satisfies (ND), (1.3) implies that \widehat{Z} is a square-integrable (P, \mathbb{F})-martingale. If we define the signed measure \widehat{P} on (Ω, \mathcal{F}) by setting

$$\frac{d\widehat{P}}{dP} := \widehat{Z}_T = \prod_{j=1}^T \frac{1 - \widetilde{\lambda}_j \Delta X_j}{1 - \widetilde{\lambda}_j \Delta A_j},$$

we thus obtain

$$(2.22) \quad \begin{aligned} E \left[\frac{d\widehat{P}}{dP} \Delta X_k \middle| \mathcal{F}_{k-1} \right] &= E \left[\widehat{Z}_k \Delta X_k \middle| \mathcal{F}_{k-1} \right] \\ &= \frac{\widehat{Z}_{k-1}}{1 - \widetilde{\lambda}_k \Delta A_k} E \left[(1 - \widetilde{\lambda}_k \Delta X_k) \Delta X_k \middle| \mathcal{F}_{k-1} \right] \\ &= 0 \quad P\text{-a.s. for } k = 1, \dots, T \end{aligned}$$

by the definition of $\widetilde{\lambda}$. If we assume that $E[\widetilde{Z}^0] > 0$, we can define another signed measure \widetilde{P} on (Ω, \mathcal{F}) by setting

$$(2.23) \quad \frac{d\widetilde{P}}{dP} := \widetilde{Z} := \frac{\widetilde{Z}^0}{E[\widetilde{Z}^0]}.$$

By (2.22) and (2.10), both \widehat{P} and \widetilde{P} are then signed martingale measures for X in the sense of the following

Definition. A signed measure Q on (Ω, \mathcal{F}) is called a *signed martingale measure* for X if $Q[\Omega] = 1$, $Q \ll P$ on \mathcal{F} with $\frac{dQ}{dP} \in \mathcal{L}^2(P)$, and X is a (Q, \mathbb{F}) -martingale in the sense that

$$E \left[\frac{dQ}{dP} \Delta X_k \middle| \mathcal{F}_{k-1} \right] = 0 \quad P\text{-a.s. for } k = 1, \dots, T.$$

Q is called an *equivalent martingale measure (with square-integrable density)* if in addition, Q is a probability measure and $Q \approx P$ on \mathcal{F} .

We point out that in contrast to most situations encountered in financial mathematics, we are working here with signed instead of equivalent martingale measures. This notion was introduced in Müller (1985) and appears here quite naturally in the study of the solution of (1.1). An explicit example in section 5 shows that \hat{P} and \tilde{P} will typically be signed measures, i.e., \hat{Z} and \tilde{Z} will also take negative values. \tilde{P} is called the *variance-optimal signed martingale measure* for X due to Theorem 3 in Schweizer (1993a), and \hat{P} is the discrete-time version of the *minimal martingale measure* introduced by Föllmer/Schweizer (1991). If X has a deterministic mean-variance tradeoff process \hat{K} , we shall see in section 4 that β coincides with $\tilde{\lambda}$ and \tilde{P} coincides with \hat{P} . In general, however, the two measures are different; this will be shown by an explicit example in section 5.

Now introduce the process \hat{V} by setting

$$(2.24) \quad \hat{V}_k := H_0 + G_k(\xi^H) + L_k^H \quad \text{for } k = 0, 1, \dots, T.$$

Then

$$\hat{V}_k = \hat{E}[H|\mathcal{F}_k] \quad P\text{-a.s. for } k = 0, 1, \dots, T$$

in the sense that

$$E \left[\frac{d\hat{P}}{dP} \Delta \hat{V}_k \middle| \mathcal{F}_{k-1} \right] = 0 \quad P\text{-a.s. for } k = 1, \dots, T$$

(i.e., \hat{V} is a “ (\hat{P}, \mathcal{F}) -martingale”) and $\hat{V}_T = H$ P -a.s. In fact, (2.24), the martingale property of \hat{Z} and (2.21) imply that

$$\begin{aligned} & E \left[\frac{d\hat{P}}{dP} \Delta \hat{V}_k \middle| \mathcal{F}_{k-1} \right] \\ &= E \left[\hat{Z}_k (\xi_k^H \Delta X_k + \Delta L_k^H) \middle| \mathcal{F}_{k-1} \right] \\ &= \xi_k^H E \left[\frac{d\hat{P}}{dP} \Delta X_k \middle| \mathcal{F}_{k-1} \right] + \hat{Z}_{k-1} \left(E [\Delta L_k^H | \mathcal{F}_{k-1}] - \frac{\tilde{\lambda}_k}{1 - \tilde{\lambda}_k \Delta A_k} E [\Delta L_k^H \Delta M_k | \mathcal{F}_{k-1}] \right) \\ &= 0 \quad P\text{-a.s.}, \end{aligned}$$

since \hat{P} is a signed martingale measure for X and L^H and $L^H M$ are (P, \mathcal{F}) -martingales. For an economic interpretation of \hat{V} as the intrinsic value process associated to H , we refer to Hofmann/Platen/Schweizer (1992).

Lemma 2.7. *In the decomposition (2.20), the constant H_0 is given by*

$$(2.25) \quad H_0 = E \left[H \prod_{j=1}^T \frac{1 - \tilde{\lambda}_j \Delta X_j}{1 - \tilde{\lambda}_j \Delta A_j} \right] = \hat{E}[H].$$

If $E[\tilde{Z}^0] > 0$, then

$$(2.26) \quad V_0 := \tilde{E}[H] = H_0 + \frac{E[L_T^H \tilde{Z}^0]}{E[\tilde{Z}^0]}.$$

Proof. Since L^H is a martingale and strongly orthogonal to M , we have for each k by (2.24)

$$0 = E[\Delta L_k^H \Delta M_k | \mathcal{F}_{k-1}] = E[\Delta L_k^H \Delta X_k | \mathcal{F}_{k-1}] = E[(\Delta \hat{V}_k - \xi_k^H \Delta X_k) \Delta X_k | \mathcal{F}_{k-1}]$$

and therefore

$$\xi_k^H = \frac{E[\Delta \hat{V}_k \Delta X_k | \mathcal{F}_{k-1}]}{E[\Delta X_k^2 | \mathcal{F}_{k-1}]}.$$

Since L^H is a martingale, this implies

$$\begin{aligned} 0 &= E[\Delta L_k^H | \mathcal{F}_{k-1}] \\ &= E[\Delta \hat{V}_k | \mathcal{F}_{k-1}] - \xi_k^H E[\Delta X_k | \mathcal{F}_{k-1}] \\ &= E[\Delta \hat{V}_k (1 - \tilde{\lambda}_k \Delta X_k) | \mathcal{F}_{k-1}] \\ &= E[\hat{V}_k (1 - \tilde{\lambda}_k \Delta X_k) | \mathcal{F}_{k-1}] - \hat{V}_{k-1} (1 - \tilde{\lambda}_k \Delta A_k) \end{aligned}$$

and therefore

$$\hat{V}_{k-1} = E\left[\hat{V}_k \frac{1 - \tilde{\lambda}_k \Delta X_k}{1 - \tilde{\lambda}_k \Delta A_k} \middle| \mathcal{F}_{k-1}\right].$$

Since $\hat{V}_T = H$ and

$$H_0 = E[\hat{V}_0 - L_0^H] = E[\hat{V}_0],$$

we obtain (2.25) by iteration. Since

$$H = H_0 + G_T(\xi^H) + L_T^H$$

by (2.20), (2.9) implies

$$E[H \tilde{Z}^0] = H_0 E[\tilde{Z}^0] + E[L_T^H \tilde{Z}^0],$$

hence (2.26) by (2.23).

q.e.d.

Remark. We shall show in section 4 that H_0 and V_0 coincide if the mean-variance tradeoff process \hat{K} is deterministic; an example in section 5 will show that H_0 and V_0 differ in general.

To relate the optimal strategy $\xi^{(c)}$ for (1.1) to the strategy ξ^H , we now introduce the predictable process $\gamma = (\gamma_k)_{k=1, \dots, T}$ defined by

$$\gamma_k := \frac{E\left[(L_T^H - L_{k-1}^H) \Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1}\right]}{E\left[\Delta X_k^2 \prod_{j=k+1}^T (1 - \beta_j \Delta X_j)^2 \middle| \mathcal{F}_{k-1}\right]} \quad \text{for } k = 1, \dots, T.$$

Due to (2.4) and the Cauchy-Schwarz inequality, γ is indeed well-defined.

Theorem 2.8. *For every fixed $c \in \mathbb{R}$, the solution $\xi^{(c)}$ of (1.1) satisfies*

$$(2.27) \quad \xi_k^{(c)} = \xi_k^H + \beta_k \left(\widehat{V}_{k-1} - c - G_{k-1}(\xi^{(c)}) \right) + \gamma_k \quad P\text{-a.s. for } k = 1, \dots, T.$$

Proof. By Theorem 2.4, $\xi^{(c)}$ satisfies for each k

$$\xi_k^{(c)} = \varrho_k - \beta_k \left(c + G_{k-1}(\xi^{(c)}) \right) \quad P\text{-a.s.}$$

By (2.11), (2.20) and (2.24), the numerator of ϱ_k is given by

$$\begin{aligned} E \left[H \Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right] &= \widehat{V}_{k-1} E \left[\Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right] \\ &\quad + \xi_k^H E \left[\Delta X_k^2 \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right] \\ &\quad + \sum_{\ell=k+1}^T E \left[\xi_\ell^H \Delta X_\ell \Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right] \\ &\quad + E \left[(L_T^H - L_{k-1}^H) \Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right]. \end{aligned}$$

But every summand in the third term on the right-hand side equals 0 P -a.s. by (2.10) (after conditioning on $\mathcal{F}_{\ell-1} \supseteq \mathcal{F}_k$), and dividing by the denominator of ϱ_k implies by (2.6) that

$$\varrho_k = \beta_k \widehat{V}_{k-1} + \xi_k^H + \gamma_k \quad P\text{-a.s.},$$

hence (2.27).

q.e.d.

Theorem 2.8 has a very interesting and intuitive interpretation. To explain this, we rewrite (2.27) as

$$\xi_k^{(c)} = \xi_k^H + \tilde{\lambda}_k \left(\widehat{V}_{k-1} - c - G_{k-1}(\xi^{(c)}) \right) + \left((\beta_k - \tilde{\lambda}_k) \left(\widehat{V}_{k-1} - c - G_{k-1}(\xi^{(c)}) \right) + \gamma_k \right).$$

It is known from the results of Schweizer (1988, 1991) that ξ^H determines a unique locally risk-minimizing strategy for the contingent claim H . The first term in the above decomposition of $\xi^{(c)}$ can therefore be interpreted as a *pure hedging* demand. In analogy to Schweizer (1992), the second term can be viewed as a demand for *mean-variance purposes*. Finally, the third term corresponds to a demand for hedging against the *stochastic fluctuations in the mean-variance ratio*

$$\frac{(E[\Delta X_k | \mathcal{F}_{k-1}])^2}{\text{Var}[\Delta X_k | \mathcal{F}_{k-1}]}.$$

We shall see in section 4 that the third term vanishes if this ratio (or equivalently \widehat{K}) is a deterministic process. In general, however, the interplay between the non-hedgeable part L^H of H and the stochastic mean-variance tradeoff induces an additional term to the solutions given by Schweizer (1992) and Schäl (1992).

3. Applications

Throughout this section, we assume that X satisfies (ND), and we consider a fixed contingent claim $H \in \mathcal{L}^2(P)$. With the help of the solution of (1.1), we shall solve several optimization problems with quadratic criteria. It should be mentioned that some of the techniques used are very similar to those in Duffie/Richardson (1991); one major difference is that our computations do not depend on the claim H under consideration and do not require a particular structure for X .

3.1. The optimal choice of strategy and initial capital

The first problem we consider is

$$(3.1) \quad \text{Minimize } E \left[(H - V_0 - G_T(\vartheta))^2 \right] \text{ over all } \vartheta \in \Theta \text{ and all } V_0 \in \mathbb{R}.$$

This can be interpreted as choosing an initial capital V_0 and a self-financing strategy ϑ in such a way that the expected net quadratic loss is minimized. We first prove a general result which holds for any set $\Theta \neq \emptyset$ and any mapping $G_T : \Theta \rightarrow \mathcal{L}^2(P)$. In particular, it uses neither the discrete-time structure of X nor the precise definition of Θ and can thus also be applied in a continuous-time framework.

Lemma 3.1. *If (1.1) has a solution $\xi^{(c)} \in \Theta$ for every $c \in \mathbb{R}$, and if c^* minimizes the function $c \mapsto E \left[(H - c - G_T(\xi^{(c)}))^2 \right]$, then the pair $(c^*, \xi^{(c^*)})$ solves (3.1).*

Proof. For any pair (c, ϑ) , we have

$$E \left[(H - c - G_T(\vartheta))^2 \right] \geq E \left[(H - c - G_T(\xi^{(c)}))^2 \right] \geq E \left[(H - c^* - G_T(\xi^{(c^*)}))^2 \right]$$

by the definitions of $\xi^{(c)}$ and c^* , respectively.

q.e.d.

Corollary 3.2. *The solution of (3.1) is given by the pair $(V_0, \xi^{(V_0)})$ with*

$$(3.2) \quad V_0 = \frac{E[H\widetilde{Z}^0]}{E[\widetilde{Z}^0]} = \widetilde{E}[H]$$

(the second equality only holds if $E[\widetilde{Z}^0] > 0$).

Proof. By Corollary 2.5, the function $c \mapsto E \left[(H - c - G_T(\xi^{(c)}))^2 \right]$ is minimized by c^* with $c^* E[\widetilde{Z}^0] = E[H\widetilde{Z}^0]$, i.e., by V_0 ; note that V_0 is well-defined due to (2.8). Hence the assertion follows from Theorem 2.2 and Lemma 3.1.

q.e.d.

Remarks. 1) Corollary 3.2 shows a feature common to many optimization problems in financial mathematics: the optimal initial capital is the expectation of H under a suitable martingale measure for X . It is tempting (and was in fact suggested by Schäl (1994)) to interpret V_0 as a *fair hedging price* for H . However, this is not always appropriate; an example in section 5 illustrates the problem which may arise from such a definition.

2) If $E[\tilde{Z}^0] = 0$, then Corollary 2.5 shows that $E\left[(H - c - G_T(\xi^{(c)}))^2\right]$ does not depend on c . Hence every pair $(V_0, \xi^{(V_0)})$ with $V_0 \in \mathbb{R}$ solves (3.1), and (3.2) implies by our conventions that we choose $V_0 = 0$. An analogous comment also applies to Corollary 3.4 below.

3) By Corollary 2.5, the optimal pair $(V_0, \xi^{(V_0)})$ also satisfies

$$E\left[H - V_0 - G_T(\xi^{(V_0)})\right] = 0.$$

As in Lemma 3.8 of Schäl (1994), this implies that $(V_0, \xi^{(V_0)})$ can be extended to a mean-self-financing strategy. We refer to Schäl (1994) and Schweizer (1988) for precise definitions and more details on this question.

3.2. The variance-optimal strategy

As a second problem, consider

$$(3.3) \quad \text{Minimize } \text{Var}[H - G_T(\vartheta)] \text{ over all } \vartheta \in \Theta.$$

The next result is again valid for any set $\Theta \neq \emptyset$ and any mapping $G_T : \Theta \rightarrow \mathcal{L}^2(P)$.

Lemma 3.3. *If (1.1) has a solution $\xi^{(c)} \in \Theta$ for every $c \in \mathbb{R}$, and if c^* minimizes the function $c \mapsto E\left[(H - c - G_T(\xi^{(c)}))^2\right]$, then $\xi^{(c^*)}$ solves (3.3).*

Proof. For every $\vartheta \in \Theta$, we have

$$\begin{aligned} \text{Var}[H - G_T(\vartheta)] &= E\left[(H - E[H - G_T(\vartheta)] - G_T(\vartheta))^2\right] \\ &\geq E\left[\left(H - E[H - G_T(\vartheta)] - G_T(\xi^{(E[H - G_T(\vartheta)])})\right)^2\right] \\ &\geq E\left[\left(H - c^* - G_T(\xi^{(c^*)})\right)^2\right] \\ &\geq \text{Var}\left[H - c^* - G_T(\xi^{(c^*)})\right] \\ &= \text{Var}\left[H - G_T(\xi^{(c^*)})\right], \end{aligned}$$

where the first inequality uses the definition of $\xi^{(c)}$ with $c := E[H - G_T(\vartheta)]$ and the second the definition of c^* .

q.e.d.

Corollary 3.4. *The solution of (3.3) is given by $\xi^{(V_0)}$ with V_0 as in (3.2).*

Proof. This follows immediately from Lemma 3.3 and the proof of Corollary 3.2.

q.e.d.

3.3. The mean-variance frontier

The third problem we consider is

$$(3.4) \quad \begin{aligned} &\text{Given } m \in \mathbb{R}, \text{ minimize } \text{Var}[H - G_T(\vartheta)] \text{ over all } \vartheta \in \Theta \\ &\text{satisfying the constraint } E[H - G_T(\vartheta)] = m. \end{aligned}$$

Lemma 3.5. *For every $c \in \mathbb{R}$, $\xi^{(c)}$ is H -mean-variance efficient in the sense that*

$$\text{Var} \left[H - G_T(\xi^{(c)}) \right] \leq \text{Var}[H - G_T(\vartheta)]$$

for every $\vartheta \in \Theta$ such that

$$E[H - G_T(\vartheta)] = E \left[H - G_T(\xi^{(c)}) \right].$$

Proof. Let $m = E \left[H - G_T(\xi^{(c)}) \right]$ and take any $\vartheta \in \Theta$ with $E[H - G_T(\vartheta)] = m$. Then the definition of $\xi^{(c)}$ implies

$$\begin{aligned} \text{Var}[H - G_T(\vartheta)] &= \text{Var}[H - c - G_T(\vartheta)] \\ &= E \left[(H - c - G_T(\vartheta))^2 \right] - (m - c)^2 \\ &\geq E \left[(H - c - G_T(\xi^{(c)}))^2 \right] - \left(E \left[H - c - G_T(\xi^{(c)}) \right] \right)^2 \\ &= \text{Var} \left[H - c - G_T(\xi^{(c)}) \right] \\ &= \text{Var} \left[H - G_T(\xi^{(c)}) \right]. \end{aligned}$$

q.e.d.

Corollary 3.6. *Suppose that X is not a martingale. For every $m \in \mathbb{R}$, the solution of (3.4) is then given by $\xi^{(c_m)}$ with*

$$(3.5) \quad c_m = \frac{m - E[H\tilde{Z}^0]}{1 - E[\tilde{Z}^0]}.$$

Proof. Fix $m \in \mathbb{R}$. By Lemma 3.5, it is enough to show that there exists $c \in \mathbb{R}$ with $E \left[H - G_T(\xi^{(c)}) \right] = m$, since the corresponding strategy $\xi^{(c)}$ will then solve (3.4). But Corollary 2.5 implies that for every $c \in \mathbb{R}$

$$E \left[H - G_T(\xi^{(c)}) \right] = E[H\tilde{Z}^0] + c \left(1 - E[\tilde{Z}^0] \right),$$

and this equals m if c is given by c_m in (3.5). Because X is not a martingale, $E[\tilde{Z}^0] \neq 1$ by Theorem 5 of Schweizer (1993a), and so c_m is well-defined.

q.e.d.

Remark. If X is a martingale, then $E[G_T(\vartheta)] = 0$ for every $\vartheta \in \Theta$; hence (3.4) only makes sense for $m = E[H]$. In that case, the solution is given by $\xi^{(E[H])}$, since

$$\begin{aligned} \text{Var}[H - G_T(\vartheta)] &= E \left[(H - E[H] - G_T(\vartheta))^2 \right] \\ &\geq E \left[\left(H - E[H] - G_T(\xi^{(E[H])}) \right)^2 \right] \\ &= \text{Var} \left[H - G_T(\xi^{(E[H])}) \right] \end{aligned}$$

for all $\vartheta \in \Theta$. As a matter of fact, the strategy $\xi^{(c)}$ does not depend on c in the martingale case, but only on H ; see subsection 4.1.

4. Special cases

In this section, we return to the basic problem (1.1) and indicate the simplifications arising in several special cases.

4.1. The martingale case

If X is a (P, \mathbb{F}) -martingale (and as usual square-integrable), (1.1) becomes very simple. First of all, the process A is identically 0 so that (ND) is trivially satisfied. Furthermore, it is clear that $G_T(\Theta)$ is closed in $\mathcal{L}^2(P)$ since X as a martingale has pairwise orthogonal increments. The adjustment process β is identically 0; it coincides with $\tilde{\lambda}$, and the measures \tilde{P} , \hat{P} and P all coincide. By Theorem 2.4, the optimal strategy for fixed c and H is

$$\xi_k^{(c)} = \varrho_k = \frac{E[H \Delta X_k | \mathcal{F}_{k-1}]}{E[\Delta X_k^2 | \mathcal{F}_{k-1}]};$$

note that this does not depend on c , which justifies the remark after Corollary 3.6. The decomposition (2.20) is the well-known *Kunita-Watanabe decomposition* of H with respect to the martingale X ; see for instance Dellacherie/Meyer (1982), Theorem VIII.51. In particular, we obtain

$$(4.1) \quad \xi^{(c)} = \varrho = \xi^H \quad \text{for every } c \in \mathbb{R}.$$

Lemma 2.7 yields

$$H_0 = V_0 = E[H],$$

and

$$E \left[\left(H - c - G_T(\xi^{(c)}) \right)^2 \right] = (E[H] - c)^2 + E \left[(L_T^H)^2 \right]$$

since

$$H - c - G_T(\xi^{(c)}) = E[H] - c + L_T^H$$

by (2.20) and (4.1). In particular, the *minimal expected net quadratic loss* or (in the terminology of Schäl (1992)) *minimal total risk*

$$J_0 := \min_{\substack{\vartheta \in \Theta, \\ c \in \mathbb{R}}} E \left[(H - c - G_T(\vartheta))^2 \right]$$

is given by

$$(4.2) \quad J_0 = E \left[(L_T^H)^2 \right].$$

4.2. The case where H is attainable

In this subsection, we assume that the contingent claim H is *attainable* in the sense that $L_T^H = 0$ P -a.s. in the decomposition (2.20). This means that H can be represented as

$$(4.3) \quad H = H_0 + \sum_{j=1}^T \xi_j^H \Delta X_j \quad P\text{-a.s.},$$

i.e., as the sum of a constant and a (discrete-time) stochastic integral with respect to X . We shall impose no special conditions on X , except as usual that (ND) is satisfied. Assumption (4.3) implies that

$$\widehat{V} = H_0 + G(\xi^H)$$

and $\gamma \equiv 0$ by (2.24). Hence Theorem 2.8 yields for each k

$$\xi_k^{(c)} - \xi_k^H = \beta_k \left(H_0 - c - \sum_{j=1}^{k-1} (\xi_j^{(c)} - \xi_j^H) \Delta X_j \right) \quad P\text{-a.s.}$$

and therefore by an induction argument

$$(4.4) \quad \xi_k^{(c)} = \xi_k^H + (H_0 - c) \beta_k \prod_{\ell=1}^{k-1} (1 - \beta_\ell \Delta X_\ell) \quad P\text{-a.s. for } k = 1, \dots, T.$$

This implies for each k that

$$\widehat{V}_k - c - G_k(\xi^{(c)}) = (H_0 - c) \prod_{j=1}^k (1 - \beta_j \Delta X_j) \quad P\text{-a.s.}$$

In particular,

$$H - c - G_T(\xi^{(c)}) = (H_0 - c) \prod_{j=1}^T (1 - \beta_j \Delta X_j) = (H_0 - c) \widetilde{Z}^0$$

which implies by (2.8) that

$$\min_{\vartheta \in \Theta} E \left[(H - c - G_T(\vartheta))^2 \right] = E \left[(H - c - G_T(\xi^{(c)}))^2 \right] = (H_0 - c)^2 E[\widetilde{Z}^0].$$

The solution of (3.1) is therefore given by the pair (H_0, ξ^H) , since $\xi^{(H_0)} = \xi^H$ by (4.4). Alternatively, this can be deduced from Corollary 3.2, since $V_0 = H_0$ by Lemma 2.7 and the assumption (4.3). In particular, we see that the minimal total risk J_0 becomes 0; this is of course obvious from (4.3).

4.3. The case where X has a deterministic mean-variance tradeoff

In this subsection, we consider the special case where X has a *deterministic mean-variance tradeoff* in the sense that

$$(4.5) \quad \text{the process } \widehat{K} \text{ is deterministic.}$$

Under this assumption, (1.1) was solved by Schäl (1994). Note that (4.5) is equivalent to saying that the process $(\widetilde{\lambda}_k \Delta A_k)_{k=1, \dots, T}$ is deterministic, since

$$(4.6) \quad 1 - \widetilde{\lambda}_k \Delta A_k = 1 - \frac{(E[\Delta X_k | \mathcal{F}_{k-1}])^2}{E[\Delta X_k^2 | \mathcal{F}_{k-1}]} = \frac{\text{Var}[\Delta X_k | \mathcal{F}_{k-1}]}{E[\Delta X_k^2 | \mathcal{F}_{k-1}]} = \left(1 + \frac{(E[\Delta X_k | \mathcal{F}_{k-1}])^2}{\text{Var}[\Delta X_k | \mathcal{F}_{k-1}]}\right)^{-1}.$$

Remark. Assumption (4.5) implies that for each k , there is a constant $\delta_k \in [0, 1]$ such that $\widetilde{\lambda}_k \Delta A_k = \delta_k$ P -a.s.; furthermore, $\delta_k = 1$ if and only if $\text{Var}[\Delta X_k | \mathcal{F}_{k-1}] = 0$ P -a.s. Thus (4.5) implies (ND) under the additional nondegeneracy condition that for each k ,

$$\text{Var}[\Delta X_k | \mathcal{F}_{k-1}] > 0 \quad \text{with positive probability;}$$

see also Example 1.

In the remainder of this subsection, we shall assume that X satisfies (ND), and we consider a fixed contingent claim $H \in \mathcal{L}^2(P)$. The basic result underlying the subsequent simplifications is then

Lemma 4.1. *If X has a deterministic mean-variance tradeoff, then*

$$E \left[\prod_{j=k}^T (1 - \widetilde{\lambda}_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right] = \prod_{j=k}^T (1 - \widetilde{\lambda}_j \Delta A_j) \quad P\text{-a.s. for } k = 1, \dots, T.$$

Proof. For fixed k ,

$$E \left[\prod_{j=k}^T (1 - \widetilde{\lambda}_j \Delta X_j) \middle| \mathcal{F}_{T-1} \right] = (1 - \widetilde{\lambda}_T \Delta A_T) \prod_{j=k}^{T-1} (1 - \widetilde{\lambda}_j \Delta X_j)$$

by the definition of A . By (4.5), the first term is deterministic, and so the assertion follows by successive conditioning on $\mathcal{F}_{T-2}, \dots, \mathcal{F}_{k-1}$.

q.e.d.

Corollary 4.2. *If X has a deterministic mean-variance tradeoff, the adjustment process β coincides with $\widetilde{\lambda}$, and the measures \widetilde{P} and \widehat{P} coincide.*

Proof. We first show by backward induction that $\beta \equiv \widetilde{\lambda}$. By (1.2) and (2.2), $\beta_T = \widetilde{\lambda}_T$. For

$k < T$,

$$\begin{aligned}
\beta_k &= \frac{E \left[\Delta X_k \prod_{j=k+1}^T (1 - \beta_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right]}{E \left[\Delta X_k^2 \prod_{j=k+1}^T (1 - \beta_j \Delta X_j)^2 \middle| \mathcal{F}_{k-1} \right]} \\
&= \frac{E \left[\Delta X_k \prod_{j=k+1}^T (1 - \tilde{\lambda}_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right]}{E \left[\Delta X_k^2 \prod_{j=k+1}^T (1 - \tilde{\lambda}_j \Delta X_j) \middle| \mathcal{F}_{k-1} \right]} \\
&= \frac{E[\Delta X_k | \mathcal{F}_{k-1}] \prod_{j=k+1}^T (1 - \tilde{\lambda}_j \Delta A_j)}{E[\Delta X_k^2 | \mathcal{F}_{k-1}] \prod_{j=k+1}^T (1 - \tilde{\lambda}_j \Delta A_j)} \\
&= \tilde{\lambda}_k
\end{aligned}$$

by the induction hypothesis, (2.6) and Lemma 4.1 (after conditioning on \mathcal{F}_k) and (4.5). This implies by Lemma 4.1

$$E[\tilde{Z}^0] = E \left[\prod_{j=1}^T (1 - \tilde{\lambda}_j \Delta X_j) \right] = \prod_{j=1}^T (1 - \tilde{\lambda}_j \Delta A_j) > 0$$

by (1.3). Thus \tilde{Z} is well-defined and equals \hat{Z}_T by (2.23) and (2.21), so that $\tilde{P} = \hat{P}$.

q.e.d.

Proposition 4.3. *If X has a deterministic mean-variance tradeoff, the solution $\xi^{(c)}$ of (1.1) satisfies*

$$(4.7) \quad \xi_k^{(c)} = \xi_k^H + \tilde{\lambda}_k \left(\hat{V}_{k-1} - c - G_{k-1}(\xi^{(c)}) \right) \quad P\text{-a.s. for } k = 1, \dots, T.$$

Proof. By Theorem 2.8 and Corollary 4.2, it is enough to show that (4.5) implies

$$(4.8) \quad \gamma_k = 0 \quad P\text{-a.s. for } k = 1, \dots, T.$$

But for every fixed k , we have for $j > k$

$$\begin{aligned}
&E \left[\Delta L_j^H \Delta X_k \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{j-1} \right] \\
&= E \left[\Delta L_j^H (1 - \tilde{\lambda}_j \Delta X_j) E \left[\prod_{\ell=j+1}^T (1 - \tilde{\lambda}_\ell \Delta X_\ell) \middle| \mathcal{F}_j \right] \middle| \mathcal{F}_{j-1} \right] \Delta X_k \prod_{\ell=k+1}^{j-1} (1 - \tilde{\lambda}_\ell \Delta X_\ell) \\
&= E \left[\Delta L_j^H (1 - \tilde{\lambda}_j \Delta X_j) \middle| \mathcal{F}_{j-1} \right] \prod_{\ell=j+1}^T (1 - \tilde{\lambda}_\ell \Delta A_\ell) \Delta X_k \prod_{\ell=k+1}^{j-1} (1 - \tilde{\lambda}_\ell \Delta X_\ell) \\
&= 0 \quad P\text{-a.s.},
\end{aligned}$$

where the first step uses Corollary 4.2, the second Lemma 4.1 and (4.5), and the third the fact that L^H is a martingale and strongly orthogonal to M . In the same way, we obtain

$$E \left[\Delta L_k^H \Delta X_k \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{k-1} \right] = E \left[\Delta L_k^H \Delta X_k \middle| \mathcal{F}_{k-1} \right] \prod_{\ell=k+1}^T (1 - \tilde{\lambda}_\ell \Delta A_\ell) = 0$$

P -a.s. and therefore by summing over $j \geq k$ and conditioning on \mathcal{F}_{k-1}

$$E \left[(L_T^H - L_{k-1}^H) \Delta X_k \prod_{\ell=k+1}^T (1 - \beta_\ell \Delta X_\ell) \middle| \mathcal{F}_{k-1} \right] = 0 \quad P\text{-a.s.},$$

hence (4.8) by the definition of γ .

q.e.d.

The result in Proposition 4.3 was previously obtained by Schäl (1994). However, it should be emphasized that his method of proof is completely different from the approach taken here. He starts by defining a predictable process $\psi^{(c)} = \left(\psi_k^{(c)} \right)_{k=1, \dots, T}$ recursively by

$$(4.9) \quad \psi_k^{(c)} = \xi_k^H + \tilde{\lambda}_k \left(\widehat{V}_{k-1} - c - G_{k-1}(\psi^{(c)}) \right)$$

and then shows that $\psi^{(c)}$ is in Θ and solves (1.1). Both these arguments rely on the condition that X has a deterministic mean-variance tradeoff; see also the remark following Corollary 2.5. Note that (4.9) has exactly the same structure as (4.7); we could therefore recover the results of Schäl (1994) by showing that $\psi^{(c)}$ is in Θ , since this implies by (4.9) and (4.7) that $\psi^{(c)}$ and $\xi^{(c)}$ coincide, and in particular by Proposition 4.3 that $\psi^{(c)}$ solves (1.1). We prefer to give an alternative proof which also works in a continuous-time framework and which seems a bit more elegant than the proof in Schäl (1994). The inspiration for this argument comes from Duffie/Richardson (1991); see also Schweizer (1992, 1993c).

Theorem 4.4. *Suppose that X has a deterministic mean-variance tradeoff. For every fixed $c \in \mathbb{R}$, the process $\psi^{(c)}$ defined by (4.9) is in Θ and solves (1.1). Furthermore,*

$$(4.10) \quad \min_{\vartheta \in \Theta} E \left[(H - c - G_T(\vartheta))^2 \right] = E \left[(H - c - G_T(\psi^{(c)}))^2 \right] \\ = \left((H_0 - c)^2 + E \left[(L_0^H)^2 \right] \right) \prod_{j=1}^T (1 - \tilde{\lambda}_j \Delta A_j) + \sum_{k=1}^T E \left[(\Delta L_k^H)^2 \right] \prod_{j=k+1}^T (1 - \tilde{\lambda}_j \Delta A_j).$$

The solution of (3.1) is given by the pair $(H_0, \psi^{(H_0)})$, and the minimal total risk is

$$(4.11) \quad J_0 = E \left[(L_0^H)^2 \right] \prod_{j=1}^T (1 - \tilde{\lambda}_j \Delta A_j) + \sum_{k=1}^T E \left[(\Delta L_k^H)^2 \right] \prod_{j=k+1}^T (1 - \tilde{\lambda}_j \Delta A_j).$$

Finally, $\psi^{(c)}$ and $\xi^{(c)}$ coincide.

Proof. 1) We first show by induction that $\psi^{(c)} \in \Theta$. For every k ,

$$\tilde{\lambda}_k^2 E \left[\Delta X_k^2 \middle| \mathcal{F}_{k-1} \right] = \tilde{\lambda}_k \Delta A_k$$

is deterministic. This implies by (4.9) that

$$\begin{aligned} E \left[\left(\psi_1^{(c)} \Delta X_1 \right)^2 \right] &= E \left[\left(\xi_1^H \Delta X_1 + \tilde{\lambda}_1 \Delta X_1 (\widehat{V}_0 - c) \right)^2 \right] \\ &\leq 2E \left[(\xi_1^H \Delta X_1)^2 \right] + 2E \left[(\widehat{V}_0 - c)^2 \tilde{\lambda}_1^2 E \left[\Delta X_1^2 | \mathcal{F}_0 \right] \right] < \infty, \end{aligned}$$

since $\xi^H \in \Theta$ and \widehat{V} is square-integrable. Now suppose that $\psi_j^{(c)} \Delta X_j \in \mathcal{L}^2(P)$ for $j = 1, \dots, k-1$. Then $\widehat{V}_{k-1} - c - G_{k-1}(\psi^{(c)}) \in \mathcal{L}^2(P)$ and an analogous argument as above shows that $\psi_k^{(c)} \Delta X_k \in \mathcal{L}^2(P)$, thus completing the induction.

2) To show that $\psi^{(c)}$ solves (1.1), we fix $\vartheta \in \Theta$ and define the function

$$f(k) := E \left[\left(\widehat{V}_k - c - G_k(\psi^{(c)}) \right) G_k(\vartheta) \prod_{j=k+1}^T (1 - \tilde{\lambda}_j \Delta X_j) \right] \quad \text{for } k = 0, 1, \dots, T.$$

Since (4.9) yields by (2.24)

$$\begin{aligned} &\left(\widehat{V}_k - c - G_k(\psi^{(c)}) \right) G_k(\vartheta) \\ &= \left(\widehat{V}_{k-1} - c - G_{k-1}(\psi^{(c)}) + (\xi_k^H - \psi_k^{(c)}) \Delta X_k + \Delta L_k^H \right) (G_{k-1}(\vartheta) + \vartheta_k \Delta X_k) \\ &= \left(\left(\widehat{V}_{k-1} - c - G_{k-1}(\psi^{(c)}) \right) (1 - \tilde{\lambda}_k \Delta X_k) + \Delta L_k^H \right) (G_{k-1}(\vartheta) + \vartheta_k \Delta X_k), \end{aligned}$$

Lemma 4.1, (2.10) and Corollary 4.2 imply that

$$f(k) = f(k-1),$$

because L^H is a martingale and strongly orthogonal to M . Hence

$$E \left[\left(H - c - G_T(\psi^{(c)}) \right) G_T(\vartheta) \right] = f(T) = f(0) = 0$$

for every $\vartheta \in \Theta$, so that $\psi^{(c)}$ solves (1.1).

3) It is clear from (4.9) and (4.7) that $\psi^{(c)}$ and $\xi^{(c)}$ must coincide. Furthermore, the solution of (3.1) is obviously obtained by minimizing $E \left[\left(H - c - G_T(\psi^{(c)}) \right)^2 \right]$ over c , and thus it only remains to prove (4.10). To that end, we define the function

$$g(k) := E \left[\left(\widehat{V}_k - c - G_k(\psi^{(c)}) \right)^2 \prod_{j=k+1}^T (1 - \tilde{\lambda}_j \Delta X_j) \right] \quad \text{for } k = 0, 1, \dots, T.$$

Then (4.9) implies as before that

$$\widehat{V}_k - c - G_k(\psi^{(c)}) = \left(\widehat{V}_{k-1} - c - G_{k-1}(\psi^{(c)}) \right) (1 - \tilde{\lambda}_k \Delta X_k) + \Delta L_k^H$$

and therefore by Lemma 4.1, since L^H is a martingale and strongly orthogonal to M ,

$$\begin{aligned} g(k) &= E \left[\left(\widehat{V}_{k-1} - c - G_{k-1}(\psi^{(c)}) \right)^2 (1 - \widetilde{\lambda}_k \Delta X_k)^2 \prod_{j=k+1}^T (1 - \widetilde{\lambda}_j \Delta X_j) \right] \\ &\quad + E \left[(\Delta L_k^H)^2 \prod_{j=k+1}^T (1 - \widetilde{\lambda}_j \Delta X_j) \right] \\ &= g(k-1) + E [(\Delta L_k^H)^2] \prod_{j=k+1}^T (1 - \widetilde{\lambda}_j \Delta A_j). \end{aligned}$$

Here, we have used Corollary 4.2 and (2.6) to simplify the first term and Lemma 4.1 for the second one. Hence we obtain

$$E \left[\left(H - c - G_T(\psi^{(c)}) \right)^2 \right] = g(T) = g(0) + \sum_{k=1}^T E [(\Delta L_k^H)^2] \prod_{j=k+1}^T (1 - \widetilde{\lambda}_j \Delta A_j),$$

and since

$$g(0) = E \left[(\widehat{V}_0 - c)^2 \prod_{j=1}^T (1 - \widetilde{\lambda}_j \Delta X_j) \right] = E [(H_0 - c + L_0^H)^2] \prod_{j=1}^T (1 - \widetilde{\lambda}_j \Delta A_j)$$

by Lemma 4.1, we obtain (4.10) after noting that $E[L_0^H] = 0$.

q.e.d.

Remarks. 1) Theorem 4.4 contains the main results of Schäl (1994); note that due to (4.6), his formula for J_0 agrees with ours. The additional term involving $E[(L_0^H)^2]$ in (4.10) and (4.11) is due to the fact that we have not assumed \mathcal{F}_0 to be trivial.

2) To obtain the solution of (3.1), we could also have used Corollary 3.2 and the fact that $H_0 = V_0$ by Corollary 4.2 and Lemma 2.7.

3) If X is a martingale, then the expression (4.11) for J_0 reduces to (4.2), since $\widetilde{\lambda} \equiv 0$ and L^H as a martingale has pairwise orthogonal increments. If X does not have a deterministic mean-variance tradeoff, an explicit formula like (4.10) does not seem to be available.

5. Explicit examples

The purpose of this section is to illustrate the previously developed concepts by means of several examples where explicit computations are possible.

Example 2. Suppose that $X_0 = 0$ and that ΔX_1 takes the values $+1, 0, -1$ with probability $\frac{1}{3}$ each. Given that $X_1 \neq +1$, ΔX_2 takes the values ± 1 with probability $\frac{1}{2}$ each. The conditional distribution of ΔX_2 given $X_1 = +1$ is denoted by ν , and we shall assume that

$$(5.1) \quad \int_{-\infty}^{\infty} x^2 \nu(dx) < \infty$$

and

$$(5.2) \quad \nu(\{0\}) < 1.$$

The filtration \mathbb{F} will be that generated by X . See Figure 1 for a graphical illustration of X .

(Insert Figure 1 here)

To simplify the notation, we shall denote the value of any \mathcal{F}_1 -measurable random variable Y on the sets $\{X_1 = +1\}$, $\{X_1 = 0\}$, $\{X_1 = -1\}$ by $Y^{(+)}$, $Y^{(0)}$ and $Y^{(-)}$, respectively. Thus we have for instance

$$\Delta A_2^{(-)} = E[\Delta X_2 | X_1 = -1] = 0.$$

It is then easy to check that

$$\tilde{\lambda}_1 = \tilde{\lambda}_2^{(-)} = \tilde{\lambda}_2^{(0)} = 0$$

and

$$(5.3) \quad \tilde{\lambda}_2^{(+)} = \frac{E[\Delta X_2 | X_1 = +1]}{E[\Delta X_2^2 | X_1 = +1]} = \frac{\int_{-\infty}^{\infty} x \nu(dx)}{\int_{-\infty}^{\infty} x^2 \nu(dx)};$$

this is well-defined by (5.1) and Jensen's inequality, and

$$(5.4) \quad \tilde{\lambda}_2^{(+)} \Delta A_2^{(+)} = \frac{\left(\int_{-\infty}^{\infty} x \nu(dx) \right)^2}{\int_{-\infty}^{\infty} x^2 \nu(dx)} < 1$$

by (5.2) and Jensen's inequality. In particular, we see that (ND) is satisfied.

Next we compute the adjustment process β . By its definition, $\beta_2 = \tilde{\lambda}_2$ and

$$(5.5) \quad \begin{aligned} \beta_1 &= \frac{E[\Delta X_1(1 - \beta_2 \Delta X_2)]}{E[\Delta X_1^2(1 - \beta_2 \Delta X_2)^2]} \\ &= \frac{E[\Delta X_1(1 - \tilde{\lambda}_2 \Delta A_2)]}{E[\Delta X_1^2(1 - \tilde{\lambda}_2 \Delta A_2)]} \\ &= \frac{-\left(\int_{-\infty}^{\infty} x \nu(dx) \right)^2}{2 \int_{-\infty}^{\infty} x^2 \nu(dx) - \left(\int_{-\infty}^{\infty} x \nu(dx) \right)^2}, \end{aligned}$$

where the first step uses $\beta_2 = \tilde{\lambda}_2$, conditioning on \mathcal{F}_1 and (2.6), and the second uses (5.4) and the structure of X . This shows that the processes β and $\tilde{\lambda}$ are different as soon as

$$\int_{-\infty}^{\infty} x \nu(dx) \neq 0,$$

which in the present setting is equivalent to saying that X is not a martingale. Furthermore, it is clear from (5.3) and (5.4) that both

$$\tilde{Z}^0 = (1 - \beta_1 \Delta X_1)(1 - \beta_2 \Delta X_2)$$

and

$$\hat{Z}_2 = \frac{(1 - \tilde{\lambda}_1 \Delta X_1)(1 - \tilde{\lambda}_2 \Delta X_2)}{(1 - \tilde{\lambda}_1 \Delta A_1)(1 - \tilde{\lambda}_2 \Delta A_2)}$$

will become negative with positive probability if $\text{supp } \nu$ is unbounded. This shows that both \tilde{P} and \hat{P} will in general not be measures, but only signed measures.

Example 3. Now consider the special case of Example 2 where

$$\nu = \frac{1}{3}\delta_{\{+2\}} + \frac{1}{2}\delta_{\{+1\}} + \frac{1}{6}\delta_{\{-1\}},$$

where $\delta_{\{x\}}$ denotes a unit mass at the point x ; see Figure 2.

(Insert Figure 2 here)

Then (5.3) – (5.5) simplify to

$$\begin{aligned}\tilde{\lambda}_2^{(+)} &= \frac{1}{2}, \\ \tilde{\lambda}_2^{(+)} \Delta A_2^{(+)} &= \frac{1}{2}, \\ \beta_1 &= -\frac{1}{3}.\end{aligned}$$

If we number the trajectories ω as ω_1 to ω_7 , starting with $\omega_1 = \{\Delta X_1 = +1, \Delta X_2 = +2\}$, $\omega_2 = \{\Delta X_1 = +1, \Delta X_2 = +1\}$ and so on, we can write the random variable \tilde{Z}^0 as a vector,

$$\tilde{Z}^0 = \left(0, \frac{2}{3}, 2, 1, 1, \frac{2}{3}, \frac{2}{3}\right).$$

Hence

$$E[\tilde{Z}^0] = \frac{7}{9}$$

and

$$(5.6) \quad \tilde{Z} = \frac{\tilde{Z}^0}{E[\tilde{Z}^0]} = \left(0, \frac{6}{7}, \frac{18}{7}, \frac{9}{7}, \frac{9}{7}, \frac{6}{7}, \frac{6}{7}\right).$$

Similarly, we obtain

$$(5.7) \quad \hat{Z}_2 = (0, 1, 3, 1, 1, 1, 1)$$

which clearly shows that \tilde{Z} and \hat{Z}_2 , hence also the measures \tilde{P} and \hat{P} , do not agree. The explicit expressions for \tilde{P} and \hat{P} in terms of their transition probabilities are given in Figures 3 and 4, respectively. Note that both \tilde{P} and \hat{P} are martingale measures for X , but not equivalent to P .

(Insert Figure 3 here)

(Insert Figure 4 here)

This example also illustrates another point. Since

$$\tilde{Z}(\omega) = \frac{\tilde{P}[\{\omega\}]}{P[\{\omega\}]} = \frac{E[\tilde{Z}^0 I_{\{\omega\}}]}{E[\tilde{Z}^0]P[\{\omega\}]}$$

for every ω , Corollary 3.2 shows that $P[\{\omega\}]\tilde{Z}(\omega)$ gives the optimal initial capital V_0 corresponding to the contingent claim $H^\omega := I_{\{\omega\}}$. Similarly, Lemma 2.7 shows that $P[\{\omega\}]\hat{Z}_2(\omega)$ equals the constant H_0 in the decomposition (2.20) of H^ω . A comparison of (5.6) and (5.7) thus reveals that H_0 and V_0 will not agree in general. Moreover, there will be no general ordering between H_0 and V_0 ; both $H_0 > V_0$ and $H_0 < V_0$ can occur, as well as $H_0 = V_0$. \tilde{Z} and \hat{Z}_2 are often called *state prices* or *state price densities* with respect to \tilde{P} and \hat{P} , respectively; see for instance Back (1991).

Another interesting feature of Example 3 is the fact that it provides us with an example of a contingent claim, namely

$$H^{\omega_1} = I_{\{\Delta X_1=+1, \Delta X_2=+2\}},$$

which is bounded, nonnegative and positive with positive probability, and yet has both $H_0 = 0$ and $V_0 = 0$. This shows that an interpretation of either H_0 or V_0 as a fair price of H does not always make sense from an economic point of view. The terminology “fair hedging price” suggested by Schäl (1994) should therefore not be used carelessly. Note that the problem here is not due to any inherent pathology of X ; it is obvious that X admits an equivalent martingale measure and therefore (see Harrison/Kreps (1979)) provides no arbitrage opportunities.

To round off the example, we now compute the optimal strategy $\xi^{(c)}$ and the mean and second moment of the net loss $H - c - G_T(\xi^{(c)})$ for the contingent claim $H^{\omega_1} = I_{\{\omega_1\}}$. First of all, ϱ in (2.11) is computed to be given by

$$\varrho_1 = \varrho_2^{(-)} = \varrho_2^{(0)} = 0 \quad , \quad \varrho_2^{(+)} = \frac{1}{3}.$$

Using Theorem 2.4, we then obtain $\xi^{(c)}$ as

$$\begin{aligned} \xi_1^{(c)} &= \frac{c}{3}, \\ \left(\xi_2^{(c)}\right)^{(-)} &= \left(\xi_2^{(c)}\right)^{(0)} = 0, \\ \left(\xi_2^{(c)}\right)^{(+)} &= \frac{1-2c}{3} = \frac{1}{3} + c - \frac{c}{3}. \end{aligned}$$

This tells us that at time 0, we should buy $\frac{c}{3}$ shares. If the stock goes up, we should buy another $\frac{1}{3} + c$ shares at time 1; otherwise, we simply sell our $\frac{c}{3}$ shares at time 1. At time 2, we liquidate whatever we have. We remark that $\xi^{(c)}$ could of course also be obtained by parametrizing ξ as $(\xi_1, \xi_2^{(-)}, \xi_2^{(0)}, \xi_2^{(+)})$ and then minimizing the quadratic function

$$E\left[(H - c - G_2(\xi))^2\right] = h\left(\xi_1, \xi_2^{(-)}, \xi_2^{(0)}, \xi_2^{(+)}\right).$$

For the total net loss, Corollary 2.5 yields

$$E \left[H - c - G_2(\xi^{(c)}) \right] = -\frac{7c}{9}$$

and

$$E \left[\left(H - c - G_2(\xi^{(c)}) \right)^2 \right] = \frac{1}{27} + \frac{7c^2}{9}$$

with an obvious minimum at $c = 0 = V_0$.

Example 4. Our final example shows that if X does not satisfy (ND), then $G_T(\Theta)$ need not be closed in $\mathcal{L}^2(P)$ and (1.1) may fail to have a solution. This counterexample is due to Walter Schachermayer (private communication).

Let $\Omega = [0, 1] \times \{-1, +1\}$ with its Borel σ -algebra \mathcal{F} ; elements of Ω will be denoted by $\omega = (u, v)$ with $u \in [0, 1]$, $v \in \{-1, +1\}$, and we denote by $U(\omega) := u$ the first and by $V(\omega) := v$ the second coordinate. Let $\mathcal{F}_0 = \mathcal{F}_1 = \sigma(U)$, $\mathcal{F}_2 = \mathcal{F}$, and let P be the measure on (Ω, \mathcal{F}) such that U is distributed uniformly on $[0, 1]$ and the conditional distribution of V given U is $U^2 \delta_{\{+1\}} + (1 - U^2) \delta_{\{-1\}}$. Finally, let $X_0 = 0$, $\Delta X_1 = 1$ and

$$\Delta X_2 = V^+(1 + U) - 1 = V^+U - V^-,$$

so that

$$\Delta X_2(u, v) = uI_{\{v=+1\}} - I_{\{v=-1\}}.$$

This model can be interpreted as follows. At time 0, we observe the value of a random variable U distributed uniformly on $[0, 1]$. Whatever the value of U , $X_0 = 0$ and $X_1 = 1$. At time 2, we toss a coin with (random) probability U^2 of getting heads. If the coin turns up heads, $\Delta X_2 = U$; otherwise, $\Delta X_2 = -1$.

Consider now the contingent claim

$$H = \left(\frac{1}{U} + 1\right)V^+ = \frac{1}{U}V^+(1 + U).$$

Then $H \in \mathcal{L}^2(P)$, since

$$E [H^2] = E \left[\left(\frac{1}{U} + 1\right)^2 E [(V^+)^2 | U] \right] = E [(1 + U)^2] < \infty.$$

If ξ is a predictable process with $G_2(\xi) = H$ P -a.s., then

$$(5.8) \quad \frac{1}{U}V^+(1 + U) = H = \xi_1 \Delta X_1 + \xi_2 \Delta X_2 = \xi_1 + \xi_2 (V^+(1 + U) - 1)$$

implies that

$$\xi_1 = \xi_2 = \frac{1}{U} \quad P\text{-a.s.}$$

by considering (5.8) separately on $\{V = +1\}$ and $\{V = -1\}$. However,

$$\xi_1 \Delta X_1 = \frac{1}{U} \notin \mathcal{L}^2(P)$$

shows that $\xi \notin \Theta$, and since there is no other predictable process ϑ with $G_2(\vartheta) = H$ P -a.s., we conclude that

$$H \notin G_2(\Theta).$$

But if we set

$$\xi^n := \xi I_{\{U \geq \frac{1}{n}\}} = \frac{1}{U} I_{\{U \geq \frac{1}{n}\}},$$

then

$$\xi_1^n \Delta X_1 = \frac{1}{U} I_{\{U \geq \frac{1}{n}\}} \in \mathcal{L}^2(P)$$

and

$$\xi_2^n \Delta X_2 = \frac{1}{U} (V^+(1+U) - 1) I_{\{U \geq \frac{1}{n}\}} \in \mathcal{L}^2(P),$$

hence $\xi^n \in \Theta$ for every $n \in \mathbb{N}$, and

$$G_2(\xi^n) = \frac{1}{U} V^+(1+U) I_{\{U \geq \frac{1}{n}\}} = H I_{\{U \geq \frac{1}{n}\}}$$

converges to H in $\mathcal{L}^2(P)$. This shows that $G_2(\Theta)$ is not closed in $\mathcal{L}^2(P)$ and that (1.1) does not have a solution for this contingent claim H and for $c = 0$.

To conclude the example, we show that X violates condition (ND). In fact,

$$E[\Delta X_2 | \mathcal{F}_1] = E[V^+(1+U) - 1 | U] = U^3 + U^2 - 1$$

and

$$E[\Delta X_2^2 | \mathcal{F}_1] = E[(V^+U - V^-)^2 | U] = U^4 - U^2 + 1,$$

and therefore the ratio

$$\frac{(E[\Delta X_2 | \mathcal{F}_1])^2}{E[\Delta X_2^2 | \mathcal{F}_1]} = \frac{(U^3 + U^2 - 1)^2}{U^4 - U^2 + 1}$$

is not uniformly bounded away from 1, since the right-hand side tends to 1 as U approaches 0.

Acknowledgement. This work is based on a part of my Habilitationsschrift at the University of Göttingen. Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged.

References

- K. Back (1991), “Asset Pricing for General Processes”, *Journal of Mathematical Economics* 20, 371–395
- C. Dellacherie and P.-A. Meyer (1982), “Probabilities and Potential B”, *North-Holland*
- D. Duffie (1988), “Security Markets. Stochastic Models”, *Academic Press*
- D. Duffie and H. R. Richardson (1991), “Mean-Variance Hedging in Continuous Time”, *Annals of Applied Probability* 1, 1–15
- H. Föllmer and M. Schweizer (1991), “Hedging of Contingent Claims under Incomplete Information”, in: *M. H. A. Davis and R. J. Elliott (eds.), “Applied Stochastic Analysis”, Stochastics Monographs, Vol. 5, Gordon and Breach, London/New York, 389–414*

- J. M. Harrison and D. M. Kreps (1979), “Martingales and Arbitrage in Multiperiod Securities Markets”, *Journal of Economic Theory* 20, 381–408
- J. M. Harrison and S. R. Pliska (1981), “Martingales and Stochastic Integrals in the Theory of Continuous Trading”, *Stochastic Processes and their Applications* 11, 215–260
- C. Hipp (1993), “Hedging General Claims”, *Proceedings of the 3rd AFIR Colloquium, Rome, Vol. 2*, 603–613
- N. Hofmann, E. Platen and M. Schweizer (1992), “Option Pricing under Incompleteness and Stochastic Volatility”, *Mathematical Finance* 2, 153–187
- D. Luenberger (1969), “Optimization by Vector Space Methods”, *Wiley*
- S. M. Müller (1985), “Arbitrage Pricing of Contingent Claims”, *Lecture Notes in Economics and Mathematical Systems* 254, *Springer*
- M. Schäl (1994), “On Quadratic Cost Criteria for Option Hedging”, *Mathematics of Operations Research* 19, 121–131
- M. Schweizer (1988), “Hedging of Options in a General Semimartingale Model”, *Diss. ETHZ No. 8615*
- M. Schweizer (1991), “Option Hedging for Semimartingales”, *Stochastic Processes and their Applications* 37, 339–363
- M. Schweizer (1992), “Mean-Variance Hedging for General Claims”, *Annals of Applied Probability* 2, 171–179
- M. Schweizer (1993a), “A New Characterization of the Martingale Property”, *SFB 303 discussion paper No. B-246, University of Bonn*
- M. Schweizer (1993b), “Approximating Random Variables by Stochastic Integrals, and Applications in Financial Mathematics”, *Habilitationschrift, University of Göttingen*
- M. Schweizer (1993c), “Approximating Random Variables by Stochastic Integrals”, *preprint, University of Göttingen, to appear in Annals of Probability*