Adv. Appl. Prob. **48,** 392–405 (2016)
*doi:*10.1017/apr.2016.6
© *Applied Probability Trust* 2016

MARKOV DECISION PROCESS ALGORITHMS FOR WEALTH ALLOCATION PROBLEMS WITH DEFAULTABLE BONDS

IKER PEREZ,*
DAVID HODGE ** AND
HUILING LE,** *The University of Nottingham*

Abstract

In this paper we are concerned with analysing optimal wealth allocation techniques within a defaultable financial market similar to Bielecki and Jang (2007). We study a portfolio optimization problem combining a continuous-time jump market and a defaultable security; and present numerical solutions through the conversion into a Markov decision process and characterization of its value function as a unique fixed point to a contracting operator. In this paper we analyse allocation strategies under several families of utility functions, and highlight significant portfolio selection differences with previously reported results.

Keywords: Portfolio optimization; defaultable bond; Markov decision process

2010 Mathematics Subject Classification: Primary 91G10
Secondary 90C40

1. Introduction

Let T be a finite time horizon and denote by $X = (X_t)_{t\geq 0}$ a continuous-time stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. Assume that X describes the evolution of a wealth process dependent on an allocation *strategy* or *policy*, taking values on a set Π. In this paper we are concerned with the study of a variation of a *portfolio optimization* problem of the form

$$V(t,x) = \sup_{\pi\in\Pi} \mathbb{E}[U(X_T^\pi) \mid X_t^\pi = x] \quad \text{for all } (t,x) \in [0,T] \times \mathbb{R}_+. \tag{1}$$

Here, the supremum is taken over all admissible policies in Π and function U is the *utility* determining a certain performance criterion.

Research within the field of portfolio optimization was triggered during the late 1960s with the work of Merton [16], who made use of stochastic control techniques to maximize expected discounted utilities of consumption. Later, his work was extended to different default-free frameworks where market uncertainty was mainly modelled by continuous processes with Brownian components; such work includes those of Fleming and Pang [11], Karatzas and Shreve [13], and Pham [17], among others. In the last decade, however, it is the optimal investment linked to defaultable claims that has attracted major attention. High-yield corporate

Received 16 April 2015; revision received 6 August 2015.
* Current address: Horizon Digital Economy Research, The University of Nottingham, Geospatial Building, Triumph Road, Nottingham NG7 2TU, UK Email address: iker.perez@nottingham.ac.uk
** Postal address: School of Mathematical Sciences, The University of Nottingham, University Park, Nottingham NG7 2RD, UK.

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

bonds offer attractive risk-return profiles and have become popular in comparison to stocks or default-free bonds; recent work in this area includes those of Bielecki and Jang [6], Bo *et al.* [8], Lakner and Liang [15], and Capponi and Figueroa-Lopez [9].

Bielecki and Jang [6] first considered a market including a defaultable bond, a risk-free account and a stock driven by Brownian dynamics, and analysed optimal asset allocations for a variation of problem (1) with a constant relative risk aversion (CRRA) utility, given by

$$V(t,x,h) = \sup_{\pi\in\Pi} \mathbb{E}\left[\frac{(X_T^\pi)^\gamma}{\gamma} \;\middle|\; X_t^\pi = x,\ H_t = h \right], \quad \text{with } 0 < \gamma < 1,$$

for all $(t,x,h) \in [0,T] \times \mathbb{R}_+ \times \{0,1\}$; here h denotes the current value of a *default process* $H = (H_t)_{t\geq 0}$ that models the state of the defaultable bond under the intensity-based approach to credit risk; see Bielecki and Rutkowski [7]. For this matter, the authors assumed constant parameters governing the system and default intensity, and derived closed-form solutions for the optimal allocations, pointing out that investments on the defaultable security are only justified under the presence of reasonable interest premiums. In addition, the results allocated a constant fraction of wealth in the Brownian asset, in a similar fashion to Merton [16].

Bo *et al.* [8] approached a perpetual allocation problem for an investor with logarithmic utility, considering a defaultable perpetual bond along with a traditional stock and a risk-free account in a similar manner to Bielecki and Jang [6]. Their work modelled stochastically the intensities and premium process and made use of heuristic arguments in order to postulate the price dynamics of the defaultable bond. Their results established monotonicity conditions on the optimal investment on defaultable bonds with respect to the risk premium and recovery of wealth at default. On the other hand, Lakner and Liang [15] employed duality theory to obtain similar optimal allocation strategies in a 2-way market, including a continuous-time money market account and a defaultable bond whose prices can jump; and Capponi and Figueroa-Lopez [9] extended the work in Bielecki and Jang [6] to a defaultable market with different economical regimes, where all assets are dependent on a finite state continuous-time Markov process $Y = (Y_t)_{t\geq 0}$; in their work they obtained the explicit solution to the optimization problem

$$V(t,x,h;y) = \sup_{\pi\in\Pi} \mathbb{E}[U(X_T^\pi) \mid X_t^\pi = x,\ H_t = h,\ Y_t = y]$$

with logarithmic and risk averse CRRA utilities, for all $(t,x,h) \in [0,T] \times \mathbb{R}_+ \times \{0,1\}$ and market regimes $y \in \{y_1, \ldots, y_N\}$, with $N > 0$. Their numerical economic analysis highlighted the preference of investors to buy defaultable bonds when the macroeconomic regimes yield high expected returns and the planning horizon is large.

Results in the literature do however primarily relate to markets incorporating Brownian assets and are limited with regards to the choices of utility functions that they provide solutions for. In this paper we incorporate the presence of a defaultable bond in a finite horizon market with a bank account and a continuous-time jump asset driven by a *piecewise deterministic Markov process* as introduced in Almudevar [1]. In this circumstance, it is possible to build a bridge between a problem formulated in continuous-time and the theory of discrete-time Markov decision processes (MDPs), reducing the optimization problem to a discrete-time model by considering an embedded state process. Similar financial markets, in the absence of the defaultable claim, have previously been explored by Bäuerle and Rieder [2] and Kirch and Runggaldier [14]. Kirch and Runggaldier [14] presented an algorithm for the evaluation of hedging strategies for European claims, addressing an optimization problem aiming to minimize the expected value

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

of a convex loss function of the hedging error; in their work, stock dynamics are driven by a geometric Poisson process.

On the other hand, Bäuerle and Rieder [2] considered the general portfolio utility maximization problem (1). In their case, the wealth process X reflects the evolution of wealth in a portfolio mixing a bank account and a generalized family of pure jump models; in addition, utility U is any increasing concave function. The authors make use of the embedding procedure previously explored by Almudevar [1] in order to convert the problem into a discrete-time MDP, and offer a proof for the validity of value iteration and policy improvement algorithms to approximate optimal allocation policies.

In this paper we make use of the results on credit risk presented in [7] along with the theory for MDPs reviewed in [4] and [18]; and extend the work in [2] and [3] to the context of defaultable markets explored in [6], [8], [9], and [15] and references therein. By means of a conversion of the optimization problem into a MDP, its value function is characterized as the unique fixed point to a dynamic programming operator and optimal wealth allocations are numerically approximated through value iteration. Thus we overcome the need to assume any particular form for the utility function and provide means of analyzing portfolio strategies incorporating illiquid markets. This allows us to undertake a numerical analysis exploring the dependence of portfolio selections on the *risk premium* and different parameters describing the system. In doing so, we are able to examine extensions of the work in [6], [8], [9], and [15] to more general families of logarithmic and exponential utility functions. The results highlight the nature of the significantly different allocation procedures under an exponential family of utilities, and the existence of a dependency on optimal stock allocation to default event, in a model with short-selling restrictions. In order for the presented procedure to hold, default intensities and interest rates are assumed constant in a similar manner to that in [6].

The paper is organised as follows. In Section 2 we provide an introduction to the market analysed and present the problem of interest. In Section 3 we derive the infinitesimal dynamics of the evolution of a joint wealth process within the optimization problem. Sections 4 and 5 are concerned with validating a procedure in order to introduce an equivalent MDP to our optimization problem, and we present the main technical results of this paper. Finally, in Sections 6 and 7 we present a numerical analysis and make comments on optimal portfolio strategies, drawing comparisons with previous results that lead to the key contributions of this paper. In addition, possible extensions of the model and drawbacks of this approach are discussed.

2. Introduction to the market and formulation of the problem

Let $(\Omega, \mathcal{G}, \mathbb{P})$ denote a complete probability space equipped with a filtration $\{\mathcal{G}_t\}_{t\geq 0}$. Here $\mathbb{P}$ refers to the real-world (also called historical) probability measure and $\{\mathcal{G}_t\}_{t\geq 0}$ is the *enlargement* of a *reference filtration* $\{\mathcal{F}_t\}_{t\geq 0}$ denoted $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t$ and satisfying the usual assumptions of completeness and right continuity; $\mathcal{H}_t$ will be introduced later. We consider a frictionless financial market consisting of a risk-free bank account $B = (B_t)_{0\leq t\leq T}$, a pure-jump asset $S = (S_t)_{0\leq t\leq T}$ and a defaultable bond $P = (P_t)_{0\leq t\leq T}$. The dynamics of each of the components of the market are given as follows.

Risk free bank account. Let $B_0 = 1$ and $r > 0$ denote the market fixed-interest rate. The deterministic dynamics of B are given by $\mathrm{d}B_t = rB_t\mathrm{d}t$.

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

Pure jump asset. Let $C = (C_t)_{0\le t\le T}$ be a compound Poisson process defined on the space $(\Omega, \mathcal{G}, \{\mathcal{F}_t\}_{t\ge 0}, \mathbb{P})$, given by

$$C_t = \sum_{n=1}^{N_t} Y_n,$$

where $N = (N_t)_{0\le t\le T}$ denotes a Poisson process with intensity $\nu > 0$ and $(Y_n)_{n\in\mathbb{N}}$ is a sequence of independent and identically distributed random variables, with $\mathbb{E}[Y_n] < \infty$, $Y_n \ge -1$, and distribution $\gamma(\mathrm{d}y)$. Here $\{\mathcal{F}_t\}_{t\ge 0}$ is a suitable complete and right-continuous filtration.

Asset S is a piecewise deterministic Markov process (cf. Almudevar [1]) adapted to $\mathcal{F}_t$ and is given by

$$\mathrm{d}S_t = S_{t^-}(\mu \mathrm{d}t + \mathrm{d}C_t),$$

where μ is the constant *appreciation rate* of the asset and $S_0 > 1$.

Defaultable bond. We consider a tradeable zero-coupon bond with face value of one unit and recovery at default. Let $\tau > 0$ be an exponentially distributed random variable defined on $(\Omega, \mathcal{G}, \{\mathcal{H}_t\}_{t\ge 0}, \mathbb{P})$ with intensity $\lambda_{\mathbb{P}}$; we make use of the intensity-based approach for modelling credit risk as introduced in Bielecki and Rutkowski [7] and let τ model the default time of the bond P. Here $\mathcal{H}_t = \sigma(H_s : s \le t)$ is the filtration generated by the one-jump process $H_t = \mathbf{1}_{\{\tau\le t\}}$, after completion and regularization on the right; C_t and H_t, as well as $\mathcal{F}_t$ and $\mathcal{H}_t$, are assumed to be independent and $\lambda_{\mathbb{P}}$ denotes the *hazard rate* of τ, so that the compensated process

$$\mathrm{d}M_t = \mathrm{d}H_t - \lambda_{\mathbb{P}}\mathrm{d}(t\wedge\tau) \tag{2}$$

with $M_0 = 0$ is a $(\mathcal{G}_t, \mathbb{P})$-martingale, with $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t$. Lastly, we denote by $Z = (Z_t)_{0\le t\le T}$ the $\mathcal{F}_t$-adapted *recovery process* of P, i.e. the process determining the wealth recovery upon default.

Then, the time-t price of this defaultable bond P with maturity at T is given by

$$P_t = B_t \mathbb{E}_{\mathbb{Q}}\left[B_T^{-1}(1 - H_T) + \int_t^T B_u^{-1} Z_u \mathrm{d}H_u \;\middle|\; \mathcal{G}_t \right], \tag{3}$$

where $\mathbb{Q}$ is a martingale measure equivalent to $\mathbb{P}$. Intuitively, P_t models the discounted $\mathbb{Q}$-expected value of the pay-off $(1 - H_T) + H_T Z_\tau$. The existence of such an equivalent measure on $(\Omega, \mathcal{G})$ follows from the results on change of measures presented in Bielecki and Rutkowski [7, Chapter 4].

Consider now an investor wishing to invest in this market. Denote by π_t^B the percentage of total wealth at time t invested on the risk-less bond; analogously π_t^S and π_t^P denote the time-t proportions on the asset and defaultable bond. The *portfolio process* $\pi = (\pi_t^B, \pi_t^S, \pi_t^P)_{0\le t\le T}$ is a $\mathcal{G}_t$-predictable process taking values in

$$\mathcal{U} = \left\{ (u_1, u_2, u_3) \in \mathbb{R}_+^3 : \sum_{i=1}^3 u_i = 1 \right\}, \tag{4}$$

so that short-selling is not allowed and wealth is fully invested at all times and remains positive; in addition, note that $\pi_t^P = 0$ for $t > \tau$ is a must.

Denote by $X^\pi = (X_t^\pi)_{0\le t\le T}$ the wealth process associated to a strategy $\pi \in \Pi$; its infinitesimal dynamics and explicit form are derived later. Here, Π denotes the family of

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

all measurable portfolio processes π taking values in $\mathcal{U}$. In view of (1), for a given increasing and concave utility function $U : (0, \infty) \to \mathbb{R}^+$, let

$$V_\pi(t, x, h) = \mathbb{E}_{t,x,h}[U(X_T^\pi)] \tag{5}$$

denote the expected terminal reward associated to a portfolio strategy $\pi \in \Pi$, at time t and with values $X_t^\pi = x$ and $H_t = h$. Here, $\mathbb{E}_{t,x,h}$ denotes the expectation under the conditional probability measure $\mathbb{P}|_{(X_t^\pi = x, H_t = h)}$. In this paper we are concerned with identifying the optimal policy $\pi^* \in \Pi$ maximizing rewards (5), so that

$$V_{\pi^*}(t, x, h) = \sup_{\pi \in \Pi} V_\pi(t, x, h) \quad \text{for all } (t, x, h) \in [0, T] \times \mathbb{R}^+ \times \{0, 1\}. \tag{6}$$

Note that $V_{\pi^*}(T, x, h) = U(x)$ for all $(x, h) \in \mathbb{R}^+ \times \{0, 1\}$ and problem V_{π^*} is tractable since $\mathbb{E}[Y_n] < \infty$.

3. The $\mathbb{P}$-dynamics of the defaultable bond and wealth evolution

Following the results in Bielecki and Rutkowski [7, Section 4.4] and Jeanblanc *et al.* [12, Section 8.6], let $\eta = \eta(\tau) = \phi e^{-\lambda_{\mathbb{P}}(\phi - 1)\tau}$ be a random variable satisfying $\eta > 0$ and $\mathbb{E}_{\mathbb{P}}[\eta] = 1$, where ϕ is a strictly positive constant. Then, the change of measure with *Radon–Nikodým density process*

$$\eta_t = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{G}_t} = \mathbb{E}_{\mathbb{P}}[\eta(\tau) \mid \mathcal{G}_t] = \mathbb{E}_{\mathbb{P}}[\eta(\tau) \mid \mathcal{H}_t] \tag{7}$$

is such that τ is an exponentially distributed random variable under $\mathbb{Q}$, with intensity $\lambda_{\mathbb{Q}} = \phi\lambda_{\mathbb{P}}$. We know (see Jeanblanc *et al.* [12]) that the stochastic process η_t defined by (7) is a $(\mathcal{G}_t, \mathbb{P})$-martingale with $\eta_0 = 1$ and

$$\mathrm{d}\eta_t = \eta_{t^-}(\phi - 1)\mathrm{d}M_t,$$

where M_t is defined by (2). In practice, default intensities are independently estimated, using credit ratings and company data for the real-world intensity $\lambda_{\mathbb{P}}$ and derivatives prices (including credit default swap and options) for $\lambda_{\mathbb{Q}}$; their underlying ratio ϕ is named the 'risk premium' and represents the reward investors claim for bearing the risk of default in P.

In order to obtain the $\mathbb{P}$-dynamics of P defined by (3) we make use of the models for valuation of contingent claims subject to default risk in Duffie and Singleton [10]. In addition, we make the following assumption.

Assumption 1. (Recovery of market.) *The wealth recovery upon default in P is given by a fraction of its current market value, i.e. $Z_t = (1 - L)P_{t^-}$ for all $t < T$, with $0 \le L \le 1$ constant.*

The application of Duffie and Singleton [10, Theorem 1] shows that under Assumption 1 and real-world probability measure $\mathbb{P}$, we have

$$\mathrm{d}P_t = \begin{cases} P_{t^-}[(r + \phi\lambda_{\mathbb{P}}L)\mathrm{d}t - \mathrm{d}H_t] & \text{if } t \le T \wedge \tau, \\ 0 & \text{if } \tau < t \le T, \end{cases} \tag{8}$$

with $P_0 = e^{-(r + \phi\lambda_{\mathbb{P}}L)T}$. We note that the price of P drops to zero at default; however, for portfolio optimization purposes we must account for the gain derived from its recovery value, so that we consider the $\mathbb{P}$-dynamics of a *gain process* for this purpose. We denote

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

by $G = (G_t)_{0\le t\le T}$ the wealth gain process resulting from holding one defaultable bond P, given by

$$\mathrm{d}G_t = \mathrm{d}P_t + Z_t \mathrm{d}H_t \tag{9}$$

with $G_0 = P_0$. Note that P and G differ in the sense that G incorporates the wealth recovered in case of default in P, so that $G_t = Z_\tau$ for $t \ge \tau$. Also, we observe in (8) that its dynamics are determined by

$$\mathrm{d}G_t = \begin{cases} G_{t^-}[(r + \phi\lambda_{\mathbb{P}} L)\mathrm{d}t - L\mathrm{d}H_t] & \text{for } t \le T \wedge \tau, \\ 0 & \text{for } \tau < t \le T \end{cases}$$

with $G_0 = P_0$. Thus, the time-t infinitesimal gain of a wealth process associated to a strategy $\pi \in \Pi$, denoted by $X^\pi = (X_t^\pi)_{0\le t\le T}$ in (5), is given by

$$\mathrm{d}X_t^\pi = X_{t^-}^\pi \left[(1 - \pi_t^P - \pi_t^S)\frac{\mathrm{d}B_t}{B_t} + \pi_t^S \frac{\mathrm{d}S_t}{S_{t^-}} + \pi_t^P \frac{\mathrm{d}G_t}{G_{t^-}} \right].$$

The explicit form of X is derived using Itô calculus and is given by

$$X_t^\pi = X_0 \exp\left\{ \int_0^t (r + \pi_s^S(\mu - r) + \pi_s^P \phi\lambda_{\mathbb{P}} L)\mathrm{d}s \right\} (1 - \pi_\tau^P L)^{H_t} \prod_{n=1}^{N_t} (1 + \pi_{T_n}^S Y_n), \tag{10}$$

where X_0 stands for the initial wealth.

4. A discrete-time MDP

Let $\Psi = (\Psi_n)_{n\ge 0}$ denote the increasing sequence of joint jump times in N and H, given by

$$\Psi_n = T_n \mathbf{1}_{\{T_n < \tau\}} + \tau \mathbf{1}_{\{T_{n-1} < \tau < T_n\}} + T_{n-1} \mathbf{1}_{\{\tau < T_{n-1}\}} \tag{11}$$

with $\Psi_0 = 0$. Intuitively, Ψ represents an ordered discrete counting process incorporating default time τ to jump times $(T_n)_{n\ge 0}$ in asset S. In addition, we refer to the counting steps $n \ge 0$ of Ψ as *decision epochs*. We define the MDP composed by the following 4-tuple $(E, \mathcal{A}, Q, R)$.

The *state space* E is given by $E = [0, T] \times \mathbb{R}^+ \times \{0, 1\}$ and supports times Ψ_n, with associated wealth X_{Ψ_n} and states of default process H_{Ψ_n}, immediately after each jump. We use the notation Ξ_n to denote the nth *state of the system*, given by

$$\Xi_n = \begin{cases} (\Psi_n, X_{\Psi_n}, H_{\Psi_n}) \in E & \text{if } \Psi_n \le T, \\ \Delta & \text{otherwise,} \end{cases}$$

for $n \ge 0$. We have $\Delta \notin E$ as an external *absorption state*, allowing us to set up an *infinite-horizon* optimization problem as described in [4] and [18].

The *action space* $\mathcal{A}$ stands for the set of deterministic *control* actions

$$\mathcal{A} = \{\alpha : \mathbb{R}^+ \to \mathcal{U} \text{ measurable}\}, \tag{12}$$

where $\mathcal{U}$ is given by (4). A control $\alpha \in \mathcal{A}$ is a function of time and $\alpha(t) \in \mathcal{U}$ determines the allocation of wealth at time t after a jump in Ψ. We note that for a given state $\Xi_n \in E \cup \{\Delta\}$ only a subclass of actions $D_n(\Xi_n) \subseteq \mathcal{A}$ may be admissible (for example, if bond P defaulted).

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

In addition to $\mathcal{A}$, we denote by F the set of all deterministic *policies* or *decision rules* given by

$$F = \{f : E \cup \{\Delta\} \to \mathcal{A} \text{ measurable}\}. \tag{13}$$

At any decision epoch n, a policy $f_n \in F$ maps a state Ξ_n to an admissible control action in $D_n(\Xi_n)$; we denote the resulting control by $f_n^{\Xi_n}$. The policy determines, as a function of the system state, the control chosen at epoch n. This, therefore results in a function $f_n^{\Xi_n} : \mathbb{R}^+ \to \mathcal{U}$ that models the time evolving allocation of wealth in our portfolio π, so that

$$\pi_t = f_n^{\Xi_n}(t - \Psi_n) \quad \text{for } t \in [\Psi_n, \Psi_{n+1}). \tag{14}$$

A portfolio process $\pi \in \Pi$ is called a *Markov portfolio strategy* if it is defined by a *Markov policy*, i.e. a sequence of functions $(f_n)_{n\geq 0}$ with $f_n \in F$. If policies $f_n \equiv f$ for all $n \geq 0$, the Markov policy is called *stationary*, implying that decisions are independent of the epoch number and only dependent on the system state. It is key to note that for a specified Markov policy, the controls to take at each epoch are random, since they depend on the system states to be observed.

We now define the *transition probability* Q. For current state $\Xi_n \in E$ and control $f_n^{\Xi_n} \in D_n(\Xi_n)$, the transition probability describes the probability for the system to adopt a specific state in epoch $n+1$ (or time Ψ_{n+1}). Let $f_n^{\Xi_n}(t) = (\alpha_t^B, \alpha_t^S, \alpha_t^P) \in \mathcal{U}$ denote the proportions of wealth allocated to each financial instrument at t time units after jump time Ψ_n, according to control $f_n^{\Xi_n}$; we note from (14) that this is equivalent to the global portfolio wealth allocation $\pi_{t+\Psi_n}$ at time $t + \Psi_n$. Analogously, let $\Gamma_t^{f_n^{\Xi_n}}$ denote the associated wealth at t time units after Ψ_n; this is equivalent to the global wealth $X_{t+\Psi_n}^{\pi}$ at time $t + \Psi_n$. Note from (10) that $\Gamma_t^{f_n^{\Xi_n}}$ is a deterministic function of the last system state, given by

$$\Gamma_t^{f_n^{\Xi_n}}(X_{\Psi_n}, H_{\Psi_n}) = X_{\Psi_n} \mathrm{e}^{\int_0^t (r+\alpha_s^S(\mu-r))\mathrm{d}s} \Big[H_{\Psi_n} + (1 - H_{\Psi_n}) \mathrm{e}^{\int_0^t \alpha_s^P \lambda_{\mathbb{P}} L \phi \mathrm{d}s} \Big]. \tag{15}$$

For an arbitrary $\Xi_n = (t', x, h)$, the transition probability Q is given by

$$\begin{aligned} Q(B \mid \Xi_n, f_n^{\Xi_n}) &= \mathbb{P}(\Xi_{n+1} \in B \mid \mathcal{G}_{\Psi_n}, f_n^{\Xi_n}) \\ &= \nu \int_0^{T-t'} \mathrm{e}^{-(\nu+(1-h)\lambda_{\mathbb{P}})s} \int_{-1}^{\infty} \mathbf{1}_B(t'+s, \Gamma_s^{f_n^{\Xi_n}}(x,h)(1+\alpha_s^S y), h) \gamma(\mathrm{d}y)\mathrm{d}s \\ &\quad + (1-h)\lambda_{\mathbb{P}} \int_0^{T-t'} \mathrm{e}^{-(\nu+\lambda_{\mathbb{P}})s} \mathbf{1}_B(t'+s, \Gamma_s^{f_n^{\Xi_n}}(x,0)(1-\alpha_s^P L), 1)\mathrm{d}s \end{aligned} \tag{16}$$

for $B \subseteq E$; in addition $Q(\{\Delta\} \mid \Xi_n, f_n^{\Xi_n}) = 1 - Q(E \mid \Xi_n, f_n^{\Xi_n})$. Since Δ is an absorbing state, we define $Q(\{\Delta\} \mid \Delta, \alpha) = 1$ for all controls $\alpha \in \mathcal{A}$. Intuitively, (16) gives the probability for the system state at epoch $n+1$ to fall within a subset B of the state space, given all information in $\mathcal{G}_{\Psi_n}$.

The *reward function* R is a function $R : E \times \mathcal{A} \to \mathbb{R}$ given by

$$R(t, x, h, \alpha) = \mathrm{e}^{-(\nu+(1-h)\lambda_{\mathbb{P}})(T-t)} U(\Gamma_{T-t}^{\alpha}(x, h)). \tag{17}$$

The adoption of such a nonnegative reward function ensures the reducibility of optimization problem (6) to an infinite-horizon discrete-time MDP, as will be shown in Lemma 1 below.

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

We note that the term $\exp(-(\nu + (1-h)\lambda_{\mathbb{P}})(T-t))$ defines the likelihood of no jumps in a Poisson process with rate $\nu+(1-h)\lambda_{\mathbb{P}}$ over a period of time $T-t$; this will be a key observation in the proof of Lemma 1. In addition, we define $R(\Delta, \alpha) = 0$ for all $\alpha \in \mathcal{A}$.

For an arbitrary state $(t, x, h) \in E$, we let $v(t, x, h)$ denote the *optimal total expected reward* over all Markov policies $(f_n)_{n\geq 0}$ with $f_n \in F$, given by

$$v(t, x, h) = \sup_{(f_n)} \mathbb{E}_{t,x,h}\left[\sum_{k=0}^{\infty} R(\Xi_k, f_k^{\Xi_k})\right]. \tag{18}$$

5. Main results

We now present the result on the equivalence between the portfolio optimization problem (6) and the MDP $(E, \mathcal{A}, Q, R)$.

Lemma 1. *For any $(t, x, h) \in E$, we have $V_{\pi^*}(t, x, h) = v(t, x, h)$, where $v(t, x, h)$ is defined by (18).*

Proof. We treat the $t = 0$ case. The result at arbitrary time points can be proved similarly upon redefinition of terminal time $T' = T - t$ and adjustment of notation as pointed out in [4, Chapter 8]. Denote by Π_{M} the set of all Markovian portfolio strategies and note that $\Pi_{\mathrm{M}} \subseteq \Pi$. Due to the Markovian structure of the state process the optimal strategy in (6) must be Markovian (cf. Bertsekas and Shreve [5]), so that

$$V_{\pi^*}(0, x, h) = \sup_{\pi\in\Pi} V_{\pi}(0, x, h) = \sup_{\pi\in\Pi_{\mathrm{M}}} \mathbb{E}_{x,h}[U(X_T^{\pi})],$$

i.e. the supremum is attained in the set Π_{M}. Any $\pi \in \Pi_{\mathrm{M}}$ is defined by a sequence of decision rules $f_n \in F$ forming a Markov policy $(f_n)_{n\geq 0}$ as described in (14). Therefore, for such a policy we need to show that $\mathbb{E}_{x,h}[U(X_T^{\pi})] = \mathbb{E}_{x,h}[\sum_{k=0}^{\infty} R(\Xi_k, f_k^{\Xi_k})]$. For this, we note that

$$\begin{aligned}\mathbb{E}_{x,h}[U(X_T^{\pi})] &= \mathbb{E}_{x,h}\left[\sum_{k=0}^{\infty} U(X_T^{\pi})\,\mathbf{1}_{\{\Psi_k\leq T<\Psi_{k+1}\}}\right]\\ &= \sum_{k=0}^{\infty} \mathbb{E}_{x,h}[\mathbb{E}_{x,h}[U(X_T^{\pi})\,\mathbf{1}_{\{\Psi_k\leq T<\Psi_{k+1}\}} \mid \mathcal{G}_{\Psi_k}]],\end{aligned}$$

where Ψ is the nondecreasing counting process in (11) incorporating default time in H_t to jump times in N_t; we recall that these are $\mathcal{G}_t$-adapted processes with exponentially distributed jumps of intensities $\lambda_{\mathbb{P}}$ and ν. In view of (14) and (15) we note that wealth X^{π} can be expressed as a deterministic function of the previous system state, i.e.

$$X_t^{\pi} = \Gamma_{t-\Psi_k}^{f_k^{\Xi_k}}(X_{\Psi_k}, H_{\Psi_k}) \quad \text{for } t \in [\Psi_k, \Psi_{k+1}),$$

with $X_0^{\pi} = x$. Therefore,

$$\begin{aligned}\mathbb{E}_{x,h}[U(X_T^{\pi})] &= \sum_{k=0}^{\infty} \mathbb{E}_{x,h}\left[\mathbb{E}_{x,h}\left[U\left(\Gamma_{T-\Psi_k}^{f_k^{\Xi_k}}(X_{\Psi_k}, H_{\Psi_k})\right)\mathbf{1}_{\{\Psi_k\leq T<\Psi_{k+1}\}} \,\middle|\, \mathcal{G}_{\Psi_k}\right]\right]\\ &= \sum_{k=0}^{\infty} \mathbb{E}_{x,h}\left[U\left(\Gamma_{T-\Psi_k}^{f_k^{\Xi_k}}(X_{\Psi_k}, H_{\Psi_k})\right)\mathbb{P}(\Psi_{k+1} > T \geq \Psi_k \mid \mathcal{G}_{\Psi_k})\right].\end{aligned} \tag{19}$$

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

In addition, we note that

$$\begin{aligned}\mathbb{P}(\Psi_{k+1} > T \geq \Psi_k \mid \mathcal{G}_{\Psi_k}) &= \mathbf{1}_{\{T \geq \Psi_k\}} \mathbb{P}(\Psi_{k+1} > T \mid \mathcal{G}_{\Psi_k}) \\ &= \mathbf{1}_{\{T \geq \Psi_k\}} \mathrm{e}^{-(\nu + (1 - H_{\Psi_k})\lambda_{\mathbb{P}})(T - \Psi_k)}.\end{aligned}$$

Thus, the right-hand side of (19) can be expressed as

$$\sum_{k=0}^{\infty} \mathbb{E}_{x,h}\Big[\mathbf{1}_{\{T \geq \Psi_k\}} \mathrm{e}^{-(\nu + (1 - H_{\Psi_k})\lambda_{\mathbb{P}})(T - \Psi_k)} U\Big(\Gamma_{T-\Psi_k}^{f_k^{\Xi_k}}(X_{\Psi_k}, H_{\Psi_k})\Big)\Big] = \sum_{k=0}^{\infty} \mathbb{E}_{x,h}[R(\Xi_k, f_k^{\Xi_k})].$$

This completes the proof. □

It has been shown that value function V_{π^*} in (6) can be derived as the sum of expected rewards v in (18). Thus, the theory of MDPs exposed in Putterman [18] and Bäuerle and Rieder [4] confirms the usefulness of iterative methods in order to approximate optimal portfolio strategies for our problem. Concretely, it is possible to construct a complete metric space with a reward operator in a similar manner to Bäuerle and Rieder [2], [3], where V_{π^*} is identified as its fixed point.

Let $\mathbb{M}(E)$ define the set of measurable functions mapping the state space E into the positive subset of the real line, i.e.

$$\mathbb{M}(E) = \{g \colon E \to \mathbb{R}^+ \colon g \text{ measurable}\}.$$

We note that the *maximal reward operator* $\mathcal{T}$ for the MDP $(E, \mathcal{A}, Q, R)$ is a *dynamic programming* operator acting on $\mathbb{M}(E)$, such that

$$(\mathcal{T}g)(t, x, h) = \sup_{\alpha \in \mathcal{A}} \left\{ R(t, x, h, \alpha) + \sum_k \int g(s, y, k) Q(\mathrm{d}s, \mathrm{d}y, k \mid t, x, h, \alpha) \right\}$$

for all $g \in \mathbb{M}(E)$ and $(t, x, h) \in E$. Additionally, we denote by $(\mathcal{L}g)(t, x, h|\alpha)$ the term within brackets, i.e.

$$(\mathcal{L}g)(t, x, h \mid \alpha) = R(t, x, h, \alpha) + \sum_k \int g(s, y, k) Q(\mathrm{d}s, \mathrm{d}y, k \mid t, x, h, \alpha),$$

and refer to it as the *reward operator*, so that

$$(\mathcal{T}g)(t, x, h) = \sup_{\alpha \in \mathcal{A}} (\mathcal{L}g)(t, x, h \mid \alpha).$$

Now, let $\mathbb{C}_\vartheta(E)$ be the function space defined by

$$\mathbb{C}_\vartheta(E) = \{g \in \mathbb{M}(E) \colon g \text{ continuous and concave in } x \text{ and } \|g\|_\vartheta < \infty\},$$

where

$$\|g\|_\vartheta = \sup_{(t,x,h) \in E} \frac{g(t, x, h)}{(1 + x)e^{\vartheta(T-t)}}$$

for some fixed and finite $\vartheta \geq 0$.

Theorem 1. *Operator $\mathcal{T}$ is a contraction mapping on the metric space $(\mathbb{C}_\vartheta(E), \|\cdot\|_\vartheta)$ and there exists an optimal stationary portfolio strategy $\pi^* \in \Pi$, defined by a Markov policy $(f)_{n \geq 0}$ with $f \in F$ as shown in (14), so that V_{π^*} in (6) is the unique fixed point of $\mathcal{T}$ in $\mathbb{C}_\vartheta(E)$.*

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

We refer the reader to [2] and [3], and more generally [4] for an overview on the techniques useful to prove the above Theorem; here, we have omitted it. We note that Theorem 1 implies that a single decision rule $f\colon \Xi_n \to \mathcal{A}$ is optimal for all epochs $n \geq 0$, and the control chosen after each jump in Ψ is only dependent on the state of the system Ξ.

6. Numerical analysis

We present and analyse computational results to our discrete-time infinite-horizon optimization problem $(E, \mathcal{A}, Q, R)$ defined in (11)–(18) for different measures of risk aversion. Numerical approximations to allocation strategies $\pi^* \in \Pi$, along with optimal values V_{π^*}, are obtained through the method of value iteration, using an homogeneous space discretization as introduced in [2, Section 5.3]

The equivalence result of Lemma 1 warrants the optimality of these strategies in the original portfolio optimization problem (6). Thus, we take advantage of the flexibility of the method regarding the choice of utility function and, in view of the original problem, determine characteristics of optimal wealth allocation strategies under different families of utilities, as well as the impact of generalizing utilities towards risky investments. Additionally, we assess the influence on allocation strategies of the different parameters defining the model and, more importantly, the effect of the short-selling restriction imposed on the original definition of the problem.

Numerical calculations in this section are undertaken with a set interest rate of $r = 0.05$. In addition, values such as jump intensities λ and ν, risk premium ϕ, loss at default L, and appreciation rate of the stock μ are, unless otherwise stated, fixed to sensible positive values within a financial context. This is done using parameter choices for numerical simulations in [6] and [9] as a reference, therefore allowing for direct comparisons of our results with recent work on portfolio management with defaultable bonds, and establishing general properties on optimal strategies with respect to variations on utility functions and time, wealth, and default state values.

The focus is on popular power, logarithmic, and exponential utility measures of risk aversion. In Figure 1 we present pre-default value functions under different choices of measures. We note that these are increasing in wealth and decreasing in time. In these cases, the optimal allocation strategies correspond to varying fractional distributions of wealth between the defaultable bond and the bank account; and the convergences in the grid have been in all cases achieved under 10 iterations, using an initial candidate V according to the strategy of investing all wealth in bond B.

6.1. Performance analysis of utility functions

Optimal allocations under different utilities vary on time, wealth values, and level of aversion towards risky investments. Under an exponential measure of constant absolute risk aversion, the level of optimal risky investments is highly dependent on wealth values; in this case, both π^P and π^S are decreasing functions of wealth for $x > \kappa$, with $\kappa \in \mathbb{R}_+$ small as observed in the case of a defaultable bond in Figure 2. In addition, the optimal allocation in P slightly increases as $t \to T$; this is opposed to previously reported optimal strategies under power and logarithmic utilities, where a mild increase of aversion towards the exposure to risky bonds is observed as time approaches deadline. Here, we also observe such aversion at times close to the deadline under power and logarithmic utilities, while remaining nearly time-invariant when the planning horizon is large. Certainly, as time approaches the deadline (and maturity in P under definition (3)) there exists an increase on the value of P and a decrease on the likelihood

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

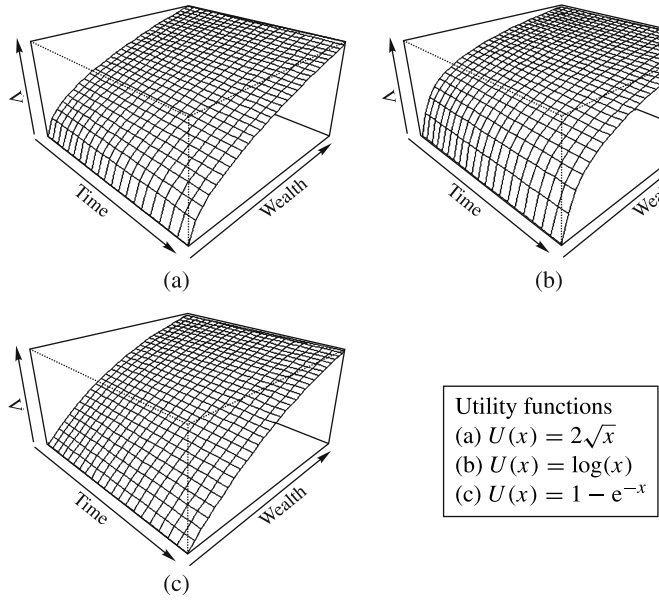


FIGURE 1: Approximation to pre-default V for different utility functions U. Results obtained through the method of value iteration with convergence in 10 iterations. $T = 1$, $r = \mu = 0.05$, $\lambda = 0.25$ $\phi = 1.3$, $L = 0.5$, and $\nu = 10$.

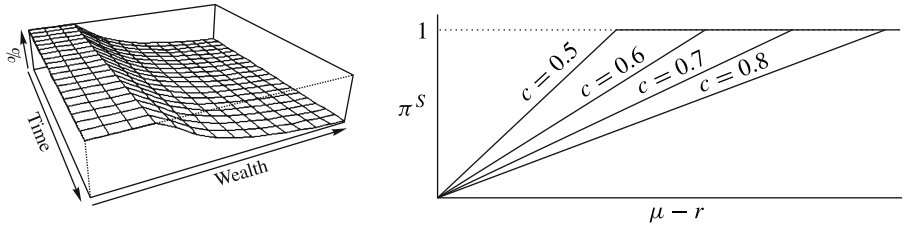


FIGURE 2: On the left, optimal π^P, for $U(x) = 1 - e^{-x}$ and varying values of $t \in [0, T]$ and $x \geq 0$ ($r = 0.05$). On the right, optimal π^S after default, as a function of the distance between the appreciation and interest rates and for power utility measures $U(x) = (x^{1-c})/(1-c)$. Maximum allocation equals 1, since no short-selling is allowed. On both, $\lambda = 0.25$ and $\nu = 10$.

of default, implying that the defaultable bond becomes relatively cheap only when the planning horizon is large.

Stock investments remain time-invariant under both power and logarithmic measures, consistent with the previous result. However, the short-selling restriction imposed on the portfolio optimization problem causes allocations π^S to remain invariant to a default event only if pre-default bond allocations π^B are strictly positive; if $\pi^B = 0$ at default time, both bond and stock percentage investments may increase following a default event in P. In Figure 2 we present varying levels of the optimal percentage allocation π^S for varying values of the difference between the appreciation rate of the stock μ and the interest rate r under power utility functions $U(x) = (x^{1-c})/(1-c)$, showing that this is a linearly increasing function on $\mu - r$ and a decreasing function on the level of constant relative risk aversion $R(x) = c$.

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

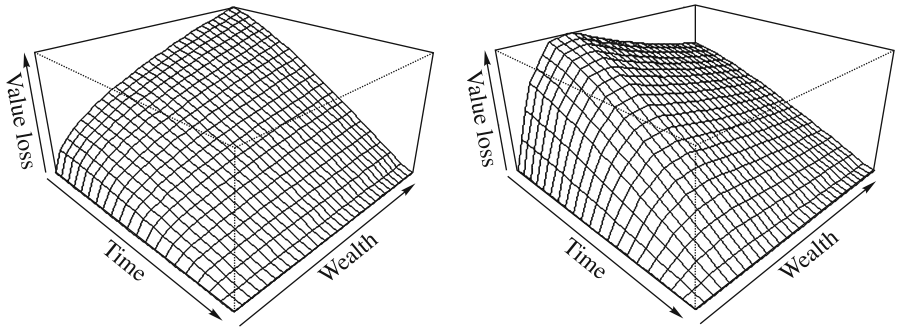


FIGURE 3: Approximation of the loss in V at default. Here $T = 1$, $r = \mu = 0.05$, $\lambda = 0.25$ $\phi = 1.3$, $L = 0.5$, and $\nu = 10$. On the left-hand side $U(x) = \sqrt{x}/2$, on the right-hand side $U(x) = 1 - \mathrm{e}^{-x}$.

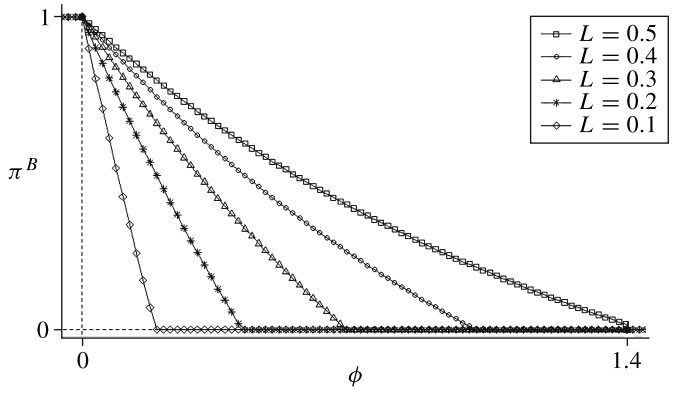


FIGURE 4: Approximation of pre-default π^B in a two-bond market, for different risk premium ϕ and loss on default L. Parameters $r = 0.05$, $\nu = 10$, $\lambda = 0.25$, and utility $U(x) = 2\sqrt{x}$.

Moreover, we note in Figure 3 that for fixed $t \in [0, T]$ and wealth $x \in \mathbb{R}_+$, the value function V is such that $V(t, x, 0) \geq V(t, x, 1)$ for all $(t, x) \in [0, t] \times \mathbb{R}_+$. In addition, $V(t, x, 0) - V(t, x, 1)$ is decreasing in time and equal to 0 at $t = T$, a common feature under all utilities. Certainly, a default event decreases the dimensionality of the problem through a reduction in the choices of investment opportunities. Under exponential utilities and for $x > \kappa$, the losses in value are decreasing functions on wealth.

Finally, the utilities analysed present common properties with regards to alterations on the values of several parameters defining the model. Optimal allocations π^P are increasing functions of the risk premium ϕ and decreasing functions of the loss value L at default, as illustrated in Figure 4 for a given pre-default state $(t, x, 0) \in E$ and utility $U(x) = 2\sqrt{x}$ in a two-bond market.

A higher incentive for bearing risk in P motivates a higher investment; on the contrary, the opposite effect is caused by decreasing the return on recovery, despite the fact that it increases the yield on the bond. It is also never optimal to invest in a defaultable bond provided $\phi \leq 1$. In addition, optimal risky investments present a similar dependency on the level of aversion under different utilities; these are decreasing functions of the level of relative/absolute risk aversion, as observed in Figure 5 for a defaultable bond under power and exponential utilities.

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

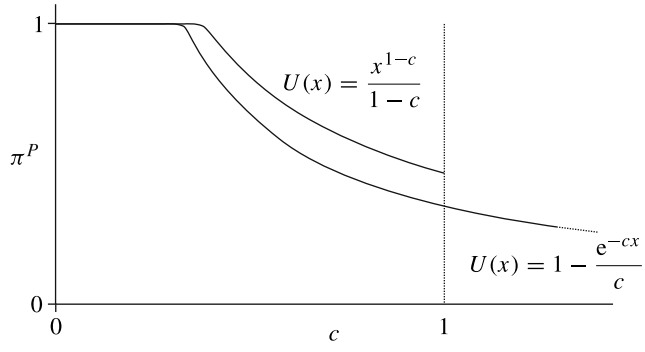


FIGURE 5: Optimal allocation π^P for utilities $U(x) = (x^{1-c})/(1-c)$, $U(x) = 1-(e^{-cx})/c$ and varying values of $c \geq 0$ in a two-bond market with fixed $(x, t, 0) \in E$. Parameters $r = 0.05$, $\nu = 10$, and $\lambda = 0.25$

7. Discussion

We have presented an extension of results in Bäuerle and Rieder [2], [3] to the context of a defaultable market, in order to study optimal wealth allocation strategies for risk adverse investors, allowing for the use of broad families of utility functions. The original continuous-time portfolio optimization problem has been transformed into a discrete-time MDP and its value function has been characterized as the unique fixed point to a dynamic programming operator, justifying the use of value iteration algorithms to provide the approximations of results of our interest.

The numerical analysis has been focused on the dependence of optimal portfolio selections on the risk premium, recovery of market value and several other parameters defining the model, and it has extended the scope of the results in [6], [8], [9], and [15] to broader families of utility functions, highlighting relevant divergences on optimal strategies with respect to variations and generalizations in choices of utilities. In addition, in this paper we have examined the impact of a short-selling restriction within the market, identifying a dependency on optimal stock allocations with respect to default event on a corporate bond.

The analysis in Section 6 suggests that, similarly to [6], [8], and [9], investments on defaultable bonds are only justified when the associated risk is correctly priced, measured in terms of risk premium coefficients ϕ. Also, similar monotonicity properties on optimal defaultable bond allocations have been identified in comparison to those presented in Bielecki and Jang [6] and Capponi and Figueroa-Lopez [9], under power and logarithmic utilities, so that these are decreasing on ϕ, increasing on L, and there exists a reduction of the risk aversion as time approaches maturity; this work suggests that such properties extend to generalizations of logarithmic utility functions. On the contrary, under exponential measures, there exists a slight increase in the risk aversion towards P in time, and optimal defaultable bond allocations are highly dependent on the wealth value and decreasing for $x > \kappa$, for some small $\kappa \in \mathbb{R}_+$. Additionally, we observed that in this case $V(t, x, 0) - V(t, x, 1)$ is decreasing on x for $x \geq \kappa$. We appreciate that many of these monotonicity observations are merely empirical at this stage, and definite scope for further work exists in trying to establish these properties in the generality observed here.

Furthermore, we have shown that the investment in the risky bond and stock is always prioritized as the levels of constant relative or absolute risk aversion are diminished. Also, optimal stock investments have been identified as linear functions of the appreciation rate of

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press

the stock and interest rate, similarly to Merton [16]. However, unlike results reported in Bielecki and Jang [6] and Capponi and Figueroa-Lopez [9], a short-selling restriction has been identified to trigger a dependency on the allocation with respect to default event in P.

Finally, we note that the problem of considering a diversified portfolio involving multiple assets and defaultable bonds is a natural extension to this work, but it is not addressed here in order to avoid the technicalities of extensive models. Other natural extensions of the model under the reduction to an MDP approach were pointed out in Bäuerle and Rieder [2]. These include the introduction of regime switching markets, where the different economical regimes are modelled by a continuous-time Markov chain $(I_t)_{t\geq 0}$ in a similar manner to Capponi and Figueroa-Lopez [9], so that parameters and coefficients defining the bank account, asset, and defaultable bond vary according to the different states of I. In this scenario, the state space within the formulation of the MDP gains a degree of dimensionality, but the embedding procedure remains similar and overcomes making strong assumptions regarding parameters defining the model.

References

[1] Almudevar, A. (2001). A dynamic programming algorithm for the optimal control of piecewise deterministic Markov processes. *SIAM J. Control Optimization.* **40,** 525–539.

[2] Bäuerle, N. and Rieder, U. (2009). MDP algorithms for portfolio optimization problems in pure jump markets. *Finance Stoch.* **13,** 591–611.

[3] Bäuerle, N. and Rieder, U. (2010). Optimal control of piecewise deterministic Markov processes with finite time horizon. In *Modern Trends in Controlled Stochastic Processes: Theory and Applications,* Luniver, Frome, pp. 123–143.

[4] Bäuerle, N. and Rieder, U. (2011). *Markov Decision Processes with Applications to Finance*. Springer, Heidelberg.

[5] Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic Optimal Control*. Academic Press, New York.

[6] Bielecki, T. R. and Jang, I. (2006). Portfolio optimization with a defaultable security. *Asia–Pacific Financial Markets* **13,** 113–127.

[7] Bielecki, T. R. and Rutkowski, M. (2002). *Credit Risk: Modelling, Valuation and Hedging*. Springer, Berlin.

[8] Bo, L., Wang, Y. and Yang, X. (2010). An optimal portfolio problem in a defaultable market. *Adv. Appl. Prob.* **42,** 689–705.

[9] Capponi, A. and Figueroa-López, J. E. (2014). Dynamic portfolio optimization with a defaultable security and regime-switching. *Math. Finance* **24,** 207–249.

[10] Duffie, D. and Singleton, K. J. (1999). Modelling term structures of defaultable bonds. *Rev. Financial Studies* **12,** 687–720.

[11] Fleming, W. H. and Pang, T. (2004). An application of stochastic control theory to financial economics. *SIAM J. Control Optimization* **43,** 502–531.

[12] Jeanblanc, M., Yor, M. and Chesney, M. (2009). *Mathematical Methods for Financial Markets*. Springer, London.

[13] Karatzas, I. and Shreve, S. E. (1998). *Methods of Mathematical Finance*. Springer, New York.

[14] Kirch, M. and Runggaldier, W. J. (2004). Efficient hedging when asset prices follow a geometric Poisson process with unknown intensities. *SIAM J. Control Optimization* **43,** 1174–1195.

[15] Lakner, P. and Liang, W. (2007). Optimal investment in a defaultable bond. *Math. Financial Econom.* **1,** 283–310.

[16] Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: the continuous-time case. *Rev. Econom. Statist.* **51,** 247–257.

[17] Pham, H. (2002). Smooth solutions to optimal investment models with stochastic volatilities and portfolio constraints. *Appl. Math. Optimization* **46,** 55–78.

[18] Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York.

https://doi.org/10.1017/apr.2016.6 Published online by Cambridge University Press