J Optim Theory Appl (2009) 143: 455–478
DOI 10.1007/s10957-009-9576-6

Asset-Liability Management Under the Safety-First Principle

M.C. Chiu · D. Li

Published online: 23 May 2009

Abstract Under the safety-first principle (Roy in Econometrica 20:431–449, 1952), one investment goal in asset-liability (AL) management is to minimize an upper bound of the ruin probability which measures the likelihood of the final surplus being less than a given target level. We derive solutions to the safety-first AL management problem under both continuous-time and multiperiod-time settings via investigating the relationship between the safety-first AL management problem and the mean-variance AL management problem, and offer geometric interpretations. We classify investors under the safety-first principle as safety-first greedy and nongreedy investors and discuss corresponding optimal strategies for them.

Keywords Portfolio selection · Asset-liability management · Safety-first · Efficient frontier

1 Introduction

Based on a recognition that investors are often primarily concerned with avoiding loss of a significant magnitude, Roy [1] proposes the safety-first principle for portfo-

Communicated by X.Q. Yang.

This work was supported by the Research Grants Council of Hong Kong under Grants CUHK 4245/04E and N_CUHK445/05.

M.C. Chiu
Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
e-mail: mamcchiu@ust.hk

D. Li (✉)
Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong, Shatin, N.T., Hong Kong
e-mail: dli@se.cuhk.edu.hk

lio selection. Under the safety-first principle, a safety-first investor specifies a threshold level of the final wealth below which the outcome is regarded as a disaster and minimizes the ruin probability or the chance of disaster. Based on the Bienaymé-Tchebycheff inequality, Roy considers a surrogate problem that minimizes an upper bound of the ruin probability, subject to a constraint that the expected final wealth is higher than the threshold.

The Roy's safety-first principle has many far-reaching consequences. This pioneering idea initiates the investigation of shortfall constraints and downside risk, and leads to the concept of Value at Risk (VaR) in the modern risk management practice, see Jorion [2] for reference. The safety-first approach is an important complement to the Markowitz's mean-variance formulation [3–5] for portfolio selection. Moreover, shortfall constraints have been also taken into the consideration under some mean-variance portfolio selection formulations, see Korn and Trautmann [6], Korn [7], Bielecki et al. [8]. Levy and Sarnat [9] find that, when the disaster level is equal to the risk-free return, the safety-first and mean-variance criteria merge together to generate the same optimal solution in a single-period model. Arzac and Bawa [10] show that the capital asset pricing model is robust to safety-first investors under traditional distribution assumptions and a single-period setting.

Extensions of the safety-first approach are abound in the literature. For instance, Telser [11] considers a hedging problem with the safety-first criterion. Kataoka [12] develops a stochastic programming model for a portfolio selection problem under the safety-first principle. Li et al. [13] employ the embedding technique of Li and Ng [14] to build up a mathematical foundation to solve the multiperiod safety-first formulation. Specifically, the analytical trading strategy which is of a form of feedback control is derived under the condition that the disaster level is less than the return of the minimum variance portfolio. Milevsky [15] considers a market that consists of one risk-free asset and two risky assets evolving as geometric Brownian motion (GBM) with constant parameters and obtains an analytic constant-rebalanced portfolio policy for a safety-first investor. However, Milevsky's model ignores the constraint that the expected return should be larger than the disaster level as imposed by Roy [1]. This simplification essentially changes the nature of the problem under investigation from "safety-first principle" to "target reaching principle" as shown in Chiu et al. [16]. Although the literature has witnessed many works on the target reaching problems or ruin probability minimization problems, see, for example, Browne [17–19], this paper remains to consider the Roy safety-first principle with the mean constraint.

A possible excuse of removing the "mean constraint" in some extensions of the safety-first principle in the literature could be that Roy [1] does not explicitly impose any constraint on the expected final wealth. However, he actually does, although implicitly. In Roy [1, p. 434], he clearly states the following:

> "... If in default of minimizing $P(\xi \le d)$, we operate on $\frac{\sigma^2}{(m-d)^2}$, this is equivalent to maximizing $\frac{m-d}{\sigma}$ ...",

where, in terms of Roy's notation, ξ is the final wealth, σ^2 is the variance of ξ, m is the expected value of ξ, and d is the preselected disaster level. Mathematically, as σ is positive, minimizing $\frac{\sigma^2}{(m-d)^2}$ is equivalent to maximizing $\frac{m-d}{\sigma}$ if and only if $m > d$. Thus, the quoted statement indicates that the mean constraint ($m > d$) has served as an indispensable element of the original Roy's safety-first principle. Empirical analysis on the safety-first formulation includes Harlow [20] and Jansen et al. [21].

Sharpe and Tint [22] recognize that a company should make investment decisions by taking into account its liabilities. They suggest that it would be more beneficial to replace the final wealth by the final surplus in portfolio selection problems. This, however, presents a challenging asset-liability (AL) management problem to pension funds, banks, insurance companies, and academics, due to that the uncontrollable liability complicates the analysis and often requires additional technical skills. In addition, Keel and Müller [23] show that liabilities do affect the efficient frontier. Adopting the mean-variance formulation, Leippold et al. [24] derive an analytical optimal policy and obtain the efficient frontier for the multiperiod AL management problem by utilizing the embedding technique of Li and Ng [14]. Chiu and Li [25] further generalize the problem to a continuous-time setting, derive the analytical optimal trading strategy and obtain the optimal initial funding ratio. However, all these works are limited to the mean-variance criterion, and little is known about an optimal policy for the safety-first AL management. As the disaster probability of the surplus is closely related to the default probability of a company, it is both interesting and indispensable to consider the safety-first AL management problem.

This paper undertakes a comprehensive analysis of the Roy's safety-first AL management problem under both continuous time and multiperiod settings. In particular, we study the implications of the safety-first AL management on optimal portfolio policies, investigate its connection to the mean-variance criterion, and study different optimal behaviors associated with a safety-first company, a safety-first individual, a mean-variance company and a mean-variance individual. Our analysis nests the problem of Li et al. [13] as a special case by setting the level of liabilities to zero. To the best of our knowledge, ours is the first attempt to directly embed safety-first objectives into a dynamic AL management problem.

The rest of the paper is organized as follows. In Sect. 2, we formulate the continuous-time and discrete-time AL models under the safety-first criterion. Section 3 derives an optimal trading strategy and classify investors into two classes: greedy and non-greedy investors. The geometric interpretation is given in Sect. 4. We conclude the paper in Sect. 5.

2 Model

In this section, we formulate the safety-first AL management problem under two settings: continuous-time and multiperiod. The former considers the situation in which investors can rebalance their portfolios continuously and asset returns are driven by Brownian motions. Although continuous-rebalance is unrealistic in practice even when transaction costs are not considered, the continuous-time model offers the limit case in AL management as a utopian reference point. While the assumption of Brownian motions in the continuous-time model imposes a distributional assumption to the asset returns, the multiperiod model allows more flexibility. The multiperiod model only requires knowledge of the first two moments for asset returns, while it permits investors to rebalance their portfolios only at discrete time instants.

In a later section, safety-first AL problem under both models will be solved under a unified framework, which makes use of the results from the mean-variance AL problem. To pave the way for the later analysis, we summarize first the results in the mean-variance AL problem in this section.

2.1 Continuous-Time Model

Consider a financial market in which n_1+1 assets are traded continuously in the time horizon $[0,T]$. Of n_1+1 assets, the asset labelled by $i=0$ is the risk-free asset and the remaining assets, labelled by $i=1,2,\dots,n_1$, are risky assets. The process of the risk-free asset $P_0(t)$ satisfies the following differential equation:

$$\begin{aligned} dP_0 &= P_0(t)r(t)dt, \\ P_0(0) &= p_0 > 0, \end{aligned}$$

where $r(t)$ is the risk-free rate. The price processes $P_1(t),\dots,P_{n_1}(t)$ of the n_1 risky assets satisfy the following stochastic differential equations (SDEs):

$$\begin{aligned} dP_i(t) &= P_i(t)\left\{\alpha_i(t)dt + \sum_{j=1}^{n}\big(\sigma_A(t)\big)_{ij}dW^j(t)\right\}, \quad t\in[0,T], \\ P_i(0) &= p_i > 0, \quad i=1,2,\dots,n_1, \end{aligned}$$

where $W_t=(W_t^1,W_t^2,W_t^3,\dots,W_t^n)'$ is a standard $\mathcal{F}_{t\geq 0}$-adapted n-dimensional Wiener process which is defined on a fixed filtered complete probability space $(\Omega,\mathcal{F},\mathcal{P},\mathcal{F}_{t\geq 0})$ with $n\geq n_1$, W_t^i and W_t^j are mutually independent for all $i\neq j$, $\mathcal{L}^2_{\mathcal{F}_T}(\Omega,\mathbb{R}^d)$ represents the set of all $\mathbb{R}^d$-valued, $\mathcal{F}_T$-measurable stochastic processes $f(t)$, such that $\mathrm{E}[\int_0^T |f(t)|^2dt]<+\infty$, $\alpha_i(t)$ is the appreciation rate of asset i, and $\sigma_A(t)=(\sigma_A(t))_{ij}$ is the variance-covariance matrix of assets, which belongs to the Banach space of $\mathbb{R}^{n_1\times n}$-valued continuous function on $[0,T]$. As widely adopted in the literature, we assume that the non-degeneracy condition of $\sigma_A(t)\sigma_A(t)'\geq\delta_A I_{n_1}$ holds for all $t\in[0,T]$ and for some $\delta_A>0$. Besides, we assume that all the functions are measurable and uniformly bounded in $[0,T]$.

Consider that an investor with an initial wealth x_0 invests in the financial market and is subject to liabilities with an initial value l_0. The liability value process follows,

$$\begin{aligned} dl(t) &= l(t)\beta(t)dt + l(t)\sigma_L(t)dW(t), \\ l(0) &= l_0, \end{aligned} \tag{1}$$

where $\beta(t)$ is the appreciation rate of the value of liabilities and $\sigma_L(t)$ is the volatility, which belongs to $C([0,T];\mathbb{R}^{n\times 1})$, a Banach space of $\mathbb{R}^{n\times 1}$-valued continuous function on $[0,T]$, and satisfies the nondegeneracy condition. The investor is allowed to continuously trade the assets over the time period $[0,T]$.

Remark 2.1 For asset-liability management problems, the incompleteness of the market is usually induced by the nontradable liability, which may involve external risky factors which do not contribute to the risky assets. However, we will show now that this consideration is indeed nested in our model setting. To see this, consider the following liability value process:

$$dl(t) = l(t)\beta(t)dt + \sum_{j=1}^{n+k} \big(\sigma_L(t)\big)_j dW^j(t),$$

$$l(0) = l_0,$$

where $(W_t^1, \ldots, W_t^n)$ spans the market, while $(W_t^{n+1}, \ldots, W_t^{n+k})$ unspans the market. Therefore $(W_t^{n+1}, \ldots, W_t^{n+k})$ are specific risky factors behind the liability. In such a situation, define

$$\widetilde{W}_t = (W_t^1, W_t^2, \ldots, W_t^n, W_t^{n+1}, W_t^{n+2}, \ldots, W_t^{n+k})',$$

and let the variance-covariance matrix be

$$\widetilde{\sigma_A}(t) := \big(\sigma_A(t)\ \mathbf{0_{n_1\times k}}\big),$$

where $\mathbf{0_{n_1\times k}}$ is a zero matrix with n_1 rows and k columns. Then, the price process of asset i becomes

$$dP_i(t) = P_i(t)\left\{\alpha_i(t)dt + \sum_{j=1}^{n+k}\big(\widetilde{\sigma_A}(t)\big)_{ij} dW^j(t)\right\}, \quad t \in [0, T],$$

$$P_i(0) = p_i > 0, \quad i = 1, 2, \ldots, n_1.$$

Consequently, it can be transformed back to our original setting with the variance-covariance matrix $\widetilde{\sigma_A}$ and $n + k$ independent Brownian motions.

Denote $S(t) = x(t) - l(t)$ as the surplus. Then, the objective of the investor is to determine an optimal investment strategy such that the probability that her final surplus $S(T)$ is below a preselected threshold D, $\mathcal{P}(S(T) \le D)$, is minimized. The value D can be viewed as a "disaster" level from the investor's point of view. As mentioned in the introduction, Roy [1, p. 434] stated:

"we operate on $\frac{\sigma^2}{(m-d)^2}$, this is equivalent to maximizing $\frac{m-d}{\sigma}$",

which implies that the mean constraint ($m > d$) is assumed in the Roy's safety-first principle. Hence, the requirement that the nonnegative value of D be smaller than the expected final surplus $\mathrm{E}[S(T)]$ is an indispensable constraint of our problem. Applying the Bienaymé-Tchebycheff inequality, we have $\mathcal{P}(S(T) \le D) \le \frac{\mathrm{Var}[S(T)]}{(\mathrm{E}[S(T)]-D)^2}$. Thus, minimizing $\mathcal{P}(S(T) \le D)$ can be achieved by minimizing its upper bound $\frac{\mathrm{Var}[S(T)]}{(\mathrm{E}[S(T)]-D)^2}$.

In this paper, we consider liabilities to be uncontrollable, meaning that the dynamics of the value of liabilities are not affected by the trading strategy of the investor.

This consideration agrees with Sharpe and Tint [22], Keel and Müller [23], Leippold et al. [24] and Chiu and Li [25]. We also assume that the price processes for both assets and liabilities follow geometric Brownian motions (GBM) in the continuous-time setting. The use of GBM in modeling the asset price process is common in the financial market. For liabilities, Norberg [26], Josa-Fombellida and Rincón-Zapatero [27] and others have used GBM to model the liability price process.

Remark 2.2 As we work in a Brownian filtration, a Brownian stochastic exponential would be a natural choice for modeling a strictly positive liability. Other models, such as random coefficient, would introduce some major technical complications, without changing the solution methodology (completion of squares) and without adding substantial additional insights. This is why the choice of GBM appears to be justified.

Let $u_i(t)$ be the amount invested in the asset i and $N_i(t)$ be units of asset i. Then, the aggregated asset $x(t)$ takes the form $x(t) = \sum_{i=0}^{n_1} u_i(t)$, where $u_i(t) = N_i(t)P_i(t)$. By Itô's lemma, the SDE underlying the dynamics of $x(t)$ is given by

$$dx(t) = \big[r(t)x(t) + \widetilde{\alpha}(t)'u(t)\big]dt + u(t)'\sigma_A(t)dW(t), \tag{2}$$

where $\widetilde{\alpha}(t) = (\alpha_1(t) - r(t), \alpha_2(t) - r(t), \dots, \alpha_{n_1}(t) - r(t))'$ is a column vector of benchmark-asset-appreciation rates. Furthermore, the investment strategy or portfolio policy is defined by $u(t) = (u_1(t), u_2(t), \dots, u_{n_1}(t))'$. By subtracting (1) from (2), the SDE for the surplus can be derived as

$$\begin{aligned} dS(t) &= \big[r(t)S(t) + (r(t) - \beta(t))l(t) + \tilde{\alpha}(t)'u(t)\big]dt \\ &\quad + \big[u(t)'\sigma_A(t) - \sigma_L(t)l(t)\big]dW(t), \\ S(0) &= x_0 - l_0, \\ dl(t) &= \beta(t)l(t)dt + \sigma_L(t)l(t)dW(t), \\ l(0) &= l_0. \end{aligned} \tag{3}$$

Hence, the safety-first AL management is formulated as

$$\begin{aligned} (\mathrm{P}_1^c(D)) \quad \min_{u(\cdot)} \quad & \frac{\mathrm{Var}[S(T)]}{(\mathrm{E}[S(T)] - D)^2}, \\ \text{s.t.} \quad & \mathrm{E}[S(T)] > D, \\ & u(\cdot) \in \mathcal{L}_{\mathcal{F}}{}^2(0, T; \mathbb{R}^{n_1}), \\ & (3), \end{aligned}$$

where $\mathcal{L}_{\mathcal{F}}{}^2(0, T; \mathbb{R}^{n_1})$ denotes the set of all $\mathbb{R}^{n_1}$-valued, measurable stochastic processes $f(t)$ adapted to $\mathcal{F}_{t\geq 0}$ such that $\mathrm{E}\int_0^T |f(t)|^2 dt < +\infty$.

The safety-first AL management is closely related to the mean-variance AL management. Hence, we define the notion of the mean-variance efficient as follows.

Definition 2.1 The surplus process $\tilde{S}(t)$ satisfying (3) is *mean-variance efficient* if there exists no other surplus $S(t)$ satisfying (3) such that $\mathrm{E}[S(T)] \geq \mathrm{E}[\tilde{S}(T)]$ and $\mathrm{Var}[S(T)] \leq \mathrm{Var}[\tilde{S}(T)]$ with at least one strict inequality. Such a point $(\mathrm{E}[\tilde{S}(T)], \mathrm{Var}[\tilde{S}(T)])$ is called a mean-variance efficient point. The set of all efficient points forms the *mean-variance efficient frontier*.

The mean-variance portfolio optimization problem $(\mathrm{P}_2^c(\epsilon))$ can be posted as follows:

$$
\begin{aligned}
(\mathrm{P}_2^c(\epsilon)) \quad & \min_{u(\cdot)} \ \mathrm{Var}[S(T)], \\
& \text{s.t.} \quad \mathrm{E}[S(T)] \geq \epsilon, \\
& \qquad u(\cdot) \in \mathcal{L}_{\mathcal{F}}{}^2(0, T; \mathbb{R}^{n_1}), \\
& \qquad (3).
\end{aligned}
$$

According to Chiu and Li [25], the mean-variance efficient frontier of the final surplus for $(\mathrm{P}_2^C(\epsilon))$ is given by

$$
\mathrm{Var}[S(T)] = \mathfrak{A}^c\big(\mathrm{E}[S(T)] - \mathfrak{B}^c\big)^2 + \mathfrak{C}^c, \tag{4}
$$

where

$$
\begin{aligned}
\mathfrak{A}^c &= \frac{e^{-\int_0^T (\tilde{\alpha}'(\sigma_A\sigma_A{}')^{-1}\tilde{\alpha})(t)dt}}{1 - e^{-\int_0^T (\tilde{\alpha}'(\sigma_A\sigma_A{}')^{-1}\tilde{\alpha})(t)dt}}, \\
\mathfrak{B}^c &= x_0 e^{\int_0^T r(t)dt} - l_0 e^{\int_0^T (\beta - \tilde{\alpha}'(\sigma_A\sigma_A{}')^{-1}\sigma_A\sigma_L{}')(t)dt}, \\
\mathfrak{C}^c &= l_0^2 \Bigg\{ e^{\int_0^T (\sigma_L\sigma_L{}' + 2\beta)(t)dt} - e^{\int_0^T (\sigma_L\sigma_L{}')(t)dt}\varrho \\
&\qquad - \frac{e^{\int_0^T (2\beta - 2\tilde{\alpha}'(\sigma_A\sigma_A{}')^{-1}(\tilde{\alpha} + \sigma_A\sigma_L{}'))(t)dt}}{e^{-\int_0^T (\tilde{\alpha}'(\sigma_A\sigma_A{}')^{-1}\tilde{\alpha})(t)dt}} \Bigg\}, \\
\varrho &= \int_0^T (\tilde{\alpha} + \sigma_A{}'\sigma_L)'(\sigma_A\sigma_A{}')^{-1}(\tilde{\alpha} + \sigma_A\sigma_L{}') \\
&\qquad \times e^{-\int_t^T (\tilde{\alpha} + \sigma_A\sigma_L{}')'(\sigma_A\sigma_A{}')^{-1}(\tilde{\alpha} + \sigma_A\sigma_L{}')d\tau} \\
&\qquad \times e^{\int_0^t \sigma_L\sigma_L{}'d\tau + \int_t^T \sigma_L\sigma_A{}'(\sigma_A\sigma_A{}')^{-1}\sigma_A\sigma_L{}'d\tau} dt.
\end{aligned}
$$

The corresponding investment strategy $u^*(t, S, l)$ is of the following feedback form:

$$
\begin{aligned}
u^*(t, S, l) = -\big[(\sigma_A\sigma_A{}')(t)\big]^{-1} \Bigg\{ & \tilde{\alpha}(t)S - \frac{\lambda}{2\omega}\tilde{\alpha}(t)e^{\int_t^T -r(\tau)d\tau} \\
& + \big[\tilde{\alpha}(t)\big(1 - e^{\int_t^T (-r + \beta - (\sigma_A\sigma_L{}')'(\sigma_A\sigma_A{}')^{-1}\tilde{\alpha})(\tau)d\tau}\big) \\
& - (\sigma_A\sigma_L{}')(t)\big(e^{\int_t^T (-r + \beta - (\sigma_A\sigma_L{}')'(\sigma_A\sigma_A{}')^{-1}\tilde{\alpha}(\tau)d\tau}\big)\big]l \Bigg\},
\end{aligned} \tag{5}
$$

where there is a relationship between the risk-averse coefficient $\frac{\lambda}{2\omega}$ and the expected surplus,

$$\begin{aligned}\mathrm{E}[S(T)] &= e^{-\int_0^T (\tilde{\alpha}'(\sigma_A\sigma_A')^{-1}\tilde{\alpha})(t)dt}\big(x_0 e^{\int_0^T r(t)dt} - l_0 e^{\int_0^T (\beta-\tilde{\alpha}'(\sigma_A\sigma_A')^{-1}\sigma_A\sigma_L')(t)dt}\big)\\ &\quad + \frac{\lambda}{2\omega}\big(1 - e^{-\int_0^T (\tilde{\alpha}'(\sigma_A\sigma_A')^{-1}\tilde{\alpha})(t)dt}\big).\end{aligned} \tag{6}$$

The above results are derived from the following auxiliary problem in Chiu and Li [25]:

$$\begin{aligned}&\min_{u(\cdot)} \ \mathrm{E}[\omega S(T)^2 - \lambda S(T)],\\ &\text{s.t.}\quad u(\cdot) \in \mathcal{L}^2_{\mathcal{F}_T}([0,T],\mathbb{R}^{n_1}),\\ &\qquad\ (3).\end{aligned}$$

Note that the optimal policy $u^*(t,S,l)$ is actually a feedback control of both the surplus S and the liability value l. In other words, the investor adjusts her portfolio based on observed values of surplus and liability. However, the investor sometimes may like to make her decision based on wealth x and liability value l. To facilitate this possibility, the optimal feedback control can be alternatively written in terms of x and l as

$$\begin{aligned}\tilde{u}(t,x,l) &= -\big[\sigma_A(t)\sigma_A(t)'\big]^{-1}\big\{\tilde{\alpha}(t)x - \tilde{\alpha}(t)e^{\int_t^T -\alpha_0(\tau)d\tau}\\ &\quad - e^{\int_t^T (-\alpha_0+\beta-(\sigma_A\sigma_L)'(\sigma_A\sigma_A')^{-1}\tilde{\alpha})(\tau)d\tau}(\tilde{\alpha}(t)+\sigma_A(t)\sigma_L(t)')l\big\}.\end{aligned} \tag{7}$$

In (4), the values of $\mathfrak{A}^c$, $\mathfrak{B}^c$ and $\mathfrak{C}^c$ are nonnegative and $\frac{\mathfrak{C}^c}{l_0^2} = 0$ if and only if $n = n_1$, which corresponds to the situation that the number of risk factors equals to the number of risky assets. The proof can be found in Chiu and Li [25]. Therefore, when $n = n_1$, $\mathrm{Var}[S(T)]$ becomes a perfect square function of $\mathrm{E}[S(T)]$ and the value of $\frac{1}{\sqrt{\mathfrak{A}^c}}$ is called the *price of risk*. At the same time, $n = n_1$ represents a complete market. The value of $\mathfrak{B}^c$ represents the return of the minimum variance portfolio, i.e., the risk-free return minus the return of the *hedged liability*. When the investor optimally hedges the liabilities by using risky assets, the appreciation rate after hedging is the original appreciation rate, β, less the project rate $\tilde{\alpha}'(\sigma_A\sigma_A')^{-1}\sigma_A\sigma_L'$, corresponding to the hedging strategy.

2.2 Multiperiod Model

In the multiperiod setting, we assume that there are n_1 risky assets traded in the market with random rates of returns. We label these risky assets by $i = 1,2,\ldots,n_1$. Let $\alpha_{i,t}$ be the return of risky asset i on time t and β_t be the appreciation of liabilities. Denote r_t to be the risk-free rate, which is deterministic at the time t. The rates of return of the risky assets at the time period t within the planning horizon are denoted by a vector $\alpha_t = (\alpha_{1,t},\alpha_{2,t},\ldots,\alpha_{n,t})'$. It is assumed that the returns α_t and β_t

have known means $\mathrm{E}[\alpha_t] = (\mathrm{E}[\alpha_{1,t}], \ldots, \mathrm{E}[\alpha_{n,t}])'$ and $\mathrm{E}[\beta_t]$, and known covariances $\mathrm{cov}_t\big(\binom{\alpha_t}{\beta_t}\big)$. Also, we assume that the matrices

$$\mathrm{E}\left[\begin{pmatrix}\alpha_t\\ \beta_t\end{pmatrix}(\alpha_t' \quad \beta_t)\right] = \mathrm{cov}_t\left[\begin{pmatrix}\alpha_t\\ \beta_t\end{pmatrix}\right] + \mathrm{E}_t\left[\begin{pmatrix}\alpha_t\\ \beta_t\end{pmatrix}\right]\mathrm{E}_t[(\alpha_t' \quad \beta_t)]$$

of conditional second moments at all time t are positive definite. The investor is allowed to trade the assets over T consecutive transaction periods at dates $0, 1, \ldots, T-1$. At $t = 0$, she is equipped with an initial wealth x_0 and possesses an initial liability l_0. The wealth x_t can be reinvested at the beginning of each time period, but the liabilities l_t are uncontrollable. The objective of the investor is to determine an optimal investment strategy such that the probability that the final surplus $S_T := x_T - l_T$ goes below a preselected threshold D, $\mathcal{P}(S_T \le D)$, is minimized, where the nonnegative "disaster" level D is smaller than the expected final surplus $\mathrm{E}[S(T)]$. As we witnessed in the continuous-time setting, minimizing $\mathcal{P}(S_T \le D)$ can be achieved by minimizing the upper bound $\frac{\mathrm{Var}[S_T]}{(\mathrm{E}[S_T]-D)^2}$ as proposed by Roy [1].

Let π_{it} be the amount invested in the asset i at time t. Then the AL dynamics can be written as

$$\begin{aligned} x_{t+1} &= r_t x_t + \tilde{\alpha}_t' \pi_t, \\ l_{t+1} &= \beta_t l_t, \end{aligned} \tag{8}$$

where $\tilde{\alpha}_t = (\alpha_{1,t} - r_t, \alpha_{2,t} - r_t, \ldots, \alpha_{n_1,t} - r_t)'$ is the vector of benchmark-asset-appreciation rate, and $\pi_t = (\pi_{1t}, \pi_{2t}, \ldots, \pi_{n_1 t})'$ is investment strategy or portfolio policy at time t in a periodic trading market. Hence, the safety-first and mean-variance AL problems in multiperiod setting are formulated as

$$(\mathrm{P}_1^d(D)) \quad \min_{\pi} \quad \frac{\mathrm{Var}[S_T]}{(\mathrm{E}[S_T] - D)^2}, \quad \text{s.t.} \quad \mathrm{E}[S_T] > D, \quad (8),$$

and

$$(\mathrm{P}_2^d(\epsilon)) \quad \min_{\pi} \quad \mathrm{Var}[S_T], \quad \text{s.t.} \quad \mathrm{E}[S_T] \ge \epsilon, \quad (8),$$

respectively. According to Leippold et al. [24], the mean-variance efficient frontier of the final surplus takes the following form:

$$\mathrm{Var}[S(T)] = \mathfrak{A}^d\big(\mathrm{E}[S(T)] - \mathfrak{B}^d\big)^2 + \mathfrak{C}^d, \tag{9}$$

where

$$\mathfrak{A}^d = \frac{1}{\mathrm{E}[S_{T,e}]} - 1; \qquad \mathfrak{B}^d = \frac{\mathrm{E}[S_{T,0}]}{1 - \mathrm{E}[S_{T,e}]}; \qquad \mathfrak{C}^d = \mathrm{E}[S_{T,0}^2] - \frac{\mathrm{E}[S_{T,0}]^2}{1 - \mathrm{E}[S_{T,e}]},$$

$$\begin{aligned}
\mathrm{E}[S_{T,e}] &= \sum_{i=0}^{T-1} \mathrm{E}[\tilde{\alpha}'_{i+1} a_{i+2}] \mathrm{E}[\tilde{\alpha}_{i+1} a_{i+2} \tilde{\alpha}'_{i+1} a_{i+2}]^{-1} \mathrm{E}[\tilde{\alpha}_{i+1} a_{i+2}], \\
\mathrm{E}[S_{T,0}] &= x_0 \big\{ \mathrm{E}[r_0 a_1] - \mathrm{E}[\tilde{\alpha}'_0 a_1] \mathrm{E}[\tilde{\alpha}_0 a_1 \tilde{\alpha}'_0 a_1]^{-1} \mathrm{E}[\tilde{\alpha}_0 a_1 r_0 a_1] \big\} \\
&\quad - l_0 \big\{ \mathrm{E}[\beta_0 b_1] - \mathrm{E}[\tilde{\alpha}'_0 a_1] \mathrm{E}[\tilde{\alpha}_0 a_1 \tilde{\alpha}'_0 a_1]^{-1} \mathrm{E}[\tilde{\alpha}_0 \beta_0 b_1] \big\}, \\
a_T &= b_T = 1, \\
a_{T-k} &= r_{T-k} a_{T-k+1} \\
&\quad - \tilde{\alpha}'_{T-k} a_{T-k+1} \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} \tilde{\alpha}'_{T-k} a_{T-k+1}]^{-1} \\
&\quad \times \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} r_{T-k} a_{T-k+1}], \quad k = 1, 2, \ldots, T, \\
b_{T-k} &= \beta_{T-k} b_{T-k+1} \\
&\quad - \tilde{\alpha}'_{T-k} a_{T-k+1} \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} \tilde{\alpha}'_{T-k} a_{T-k+1}]^{-1} \\
&\quad \times \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} \beta_{T-k} b_{T-k+1}], \quad k = 1, 2, \ldots, T.
\end{aligned} \tag{10}$$

The corresponding optimal trading strategy is

$$\begin{aligned}
\pi^*_{T-k} = -\mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} \tilde{\alpha}'_{T-k} a_{T-k+1}]^{-1} \Big\{ & x_{T-k} \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} r_{T-k} a_{T-k+1}] \\
& - l_{T-k} \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} \beta_{T-k} b_{T-k+1}] - \frac{\lambda}{2\omega} \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1}] \Big\},
\end{aligned} \tag{11}$$

where there is a relationship between the risk-averse coefficient $\frac{\lambda}{2\omega}$ and the expected surplus,

$$\mathrm{E}[S_T] = \mathrm{E}[S_{T,0}] + \frac{\lambda}{2\omega} \mathrm{E}[S_{T,e}]. \tag{12}$$

Note that there is an essential difference between the discrete-time and continuous-time settings when deriving an optimal trading strategy. Under the continuous-time setting, the optimal policy can be expressed analytically, while, in the case of a discrete-time setting, analytical solution always can be expressed in an iterative formula, see (10) and (11). However, when returns are independent with deterministic coefficients, the closed-form solution can be obtained. Note that $\mathfrak{C}^d \geq 0$ in (9). In the next section, we will link up the safety-first criterion and the mean-variance criterion. Using the quadratic forms of the mean-variance efficient frontier, (4) and (9), we can show that both the optimal trading strategies of $(\mathrm{P}_1^c(D))$ and $(\mathrm{P}_1^d(D))$ are mean-variance efficient.

3 Solution to the Safety-First AL Management

Under both the discrete-time and continuous-time settings, the mean-variance efficient frontiers take the following form:

$$\mathrm{Var}[S_T] = \mathfrak{A}^{c,d} \big(\mathrm{E}[S_T] - \mathfrak{B}^{c,d} \big)^2 + \mathfrak{C}^{c,d}, \tag{13}$$

where the choice of superscript, c or d, depends on whether the model is continuous or discrete. In (13) $\mathfrak{B}^{c,d}$ is the return of the minimum risk portfolio, $\mathfrak{C}^{c,d}$ is the minimum variance, and $\frac{1}{\sqrt{\mathfrak{A}^{c,d}}}$ provides a rough information how much the expected return of a portfolio increases when the standard deviation increases by one unit. If $\mathfrak{C}^{c,d}$ equals zero, then $\frac{1}{\sqrt{\mathfrak{A}^{c,d}}}$ is termed the price of risk. The mean-variance AL problem can be presented in a general form as follows,

$$\begin{aligned} (\mathrm{P}_2(\epsilon)) \quad & \min_{u(\cdot)} && \mathrm{Var}[S_T], \\ & \text{s.t.} && \mathrm{E}[S_T] \geq \epsilon, \\ & && \mathfrak{W}, \end{aligned}$$

where $\mathfrak{W}$ in $(\mathrm{P}_2(\epsilon))$ represents, in the continuous-time setting, the constraints

$$\begin{aligned} & u(\cdot) \in \mathcal{L}_{\mathcal{F}}{}^2(0, T; \mathbb{R}^{n_1}), \\ & (3), \end{aligned} \tag{14}$$

or represents, in the discrete-time setting, the set of constraints

$$\begin{aligned} x_{t+1} &= r_t x_t + \tilde{\alpha}_t' u_t, \quad t = 0, 1, \ldots, T-1, \\ l_{t+1} &= \beta_t l_t, \quad t = 0, 1, \ldots, T-1. \end{aligned} \tag{15}$$

In this section, we derive the solution to the safety-first AL problem without attaching it to a specific market structure (being continuous or discrete). Using the expression of the mean-variance efficient frontier (MV-EF) in the form (13), the safety-first AL problem can be posted as follows:

$$\begin{aligned} (\mathrm{P}_1(D)) \quad & \min_{u} && \frac{\mathrm{Var}[S_T]}{(\mathrm{E}[S_T] - D)^2}, \\ & \text{s.t.} && \mathrm{E}[S_T] > D, \\ & && \mathfrak{W}, \end{aligned}$$

where the $\mathfrak{W}$ represents the set of constraints (14) or (15). Then, we have the following theorem.

Theorem 3.1 *The optimal solution of* $(\mathrm{P}_1(D))$ *with a given nonnegative* D *is an optimal solution of* $(\mathrm{P}_2(\epsilon))$ *for some* $\epsilon \geq 0$.

Proof Define Π_P to be the set of optimal solutions of problem (P). Hence,

$$\Pi_{\mathrm{P}_1(D)} = \{u \mid u \text{ is a minimizer of } \mathrm{P}_1(D)\},$$

$$\Pi_{\mathrm{P}_2(\epsilon)} = \{u \mid u \text{ is a minimizer of } \mathrm{P}_2(\epsilon)\}.$$

Suppose that $\Pi_{\mathrm{P}_1(D)} \nsubseteq \bigcup_{\epsilon \geq 0} \Pi_{\mathrm{P}_2(\epsilon)}$. It implies that there exists $u \in \Pi_{\mathrm{P}_1(D)}$ such that $u \notin \Pi_{\mathrm{P}_2(\epsilon)}$ for all $\epsilon \geq 0$. Let $V = \mathrm{Var}[S(T)]|_u$ and $E = \mathrm{E}[S(T)]|_u$.

Case 1: $D \geq \mathfrak{B}^{c,d}$. In this case, $E > \mathfrak{B}^{c,d}$. Due to the continuity of the efficient frontier, there exists $u^* \in \bigcup_{\epsilon\geq 0} \Pi_{\mathrm{P}_2(\epsilon)}$ such that $\mathrm{E}[S(T)]|_{u^*} = E$ and $\mathrm{Var}[S(T)]|_{u^*} < V$, as u is not mean-variance efficient. Consider

$$\frac{\mathrm{Var}[S(T)]|_{u^*}}{(\mathrm{E}[S(T)]|_{u^*} - D)^2} = \frac{\mathrm{Var}[S(T)]|_{u^*}}{(E-D)^2} < \frac{V}{(E-D)^2},$$

which contradicts the optimality of u in problem $(\mathrm{P}_1(D))$. Therefore, $u \in \bigcup_{\epsilon\geq 0} \Pi_{\mathrm{P}_2(\epsilon)}$.

Case 2: $D < \mathfrak{B}^{c,d}$. If $E \geq \mathfrak{B}^{c,d}$, then the proof is similar to Case 1. Otherwise, if $E < \mathfrak{B}^{c,d}$, let $u^* \in \bigcup_{\epsilon\geq 0} \Pi_{\mathrm{P}_2(\epsilon)}$ such that $\mathrm{E}[S(T)]|_{u^*} = \mathfrak{B}^{c,d} > E$. Then, $\mathrm{Var}[S(T)]|_{u^*}$ must be the minimum variance of the final surplus. Since $D < E < \mathfrak{B}^{c,d}$, we have

$$(E-D)^2 < (\mathfrak{B}^{c,d} - D)^2,$$

which leads to

$$\frac{V}{(E-D)^2} > \frac{V}{(\mathfrak{B}^{c,d} - D)^2} > \frac{\mathrm{Var}[S(T)]|_{u^*}}{(\mathrm{E}[S(T)]|_{u^*} - D)^2},$$

which contradicts the optimality of u. Hence, $u \in \bigcup_{\epsilon\geq 0} \Pi_{\mathrm{P}_2(\epsilon)}$.

We can finally conclude that $\Pi_{\mathrm{P}_1(D)} \subseteq \bigcup_{\epsilon\geq 0} \Pi_{\mathrm{P}_2(\epsilon)}$. □

Theorem 3.1 asserts that the optimal solution to $(\mathrm{P}_1(D))$ can be found from among the points on the mean-variance efficient frontier. Thus, it remains to identify the efficient solution which solves the safety-first AL problem. To do this, we have to consider several different cases.

Li et al. [13] confine their study in the case where the disaster level is less than the expected return of the minimum variance portfolio. In Milevsky [15], although the mean constraint is absent in the analysis, the disaster level is set at the maximum possible risk-free return, $x_0 e^{\alpha_0 T}$. Hence, it may be more interesting to focus on this situation first.

Let E^* be the expected final surplus under the optimal strategy for $(\mathrm{P}_1(D))$, if it exists. Then, we have the following result.

Theorem 3.2 *If* $D < \mathfrak{B}^{c,d}$, *then* $\mathfrak{B}^{c,d} < E^* = \mathfrak{B}^{c,d} + \frac{\mathfrak{C}^{c,d}}{\mathfrak{A}^{c,d}(\mathfrak{B}^{c,d} - D)} < \infty$, *and the optimal trading rule for* $(\mathrm{P}_1(D))$ *can be obtained by substituting* E^* *into* (6) *and* (5) *for the continuous-time model, and substituting* E^* *into* (12) *and* (11) *for the discrete-time model. More specifically, we have the following explicit forms of the optimal trading rule for* $(\mathrm{P}_1(D))$:

$$\begin{aligned} u^*(t,S,l) = &-\big[(\sigma_A\sigma_A{}')(t)\big]^{-1} \\ &\times \Big\{\tilde{\alpha}(t)'S + \big[\tilde{\alpha}(t)'\big(1 - e^{\int_t^T(-r+\beta-(\sigma_A\sigma_L)'(\sigma_A\sigma_A{}')^{-1}\tilde{\alpha})(\tau)d\tau}\big) \\ &- (\sigma_A\sigma_L)(t)\big(e^{\int_t^T(-r+\beta-(\sigma_A\sigma_L)'(\sigma_A\sigma_A{}')^{-1}\tilde{\alpha})(\tau)d\tau}\big)\big]l \end{aligned}$$

$$- \left(\mathfrak{B}^c + \frac{\mathfrak{C}^c}{(\mathfrak{B}^c - D)e^{-\int_0^T (\tilde{\alpha}'(\sigma_A \sigma_A')^{-1}\tilde{\alpha})(t)dt}} \right) \tilde{\alpha}(t)' e^{\int_t^T -r(\tau)d\tau} \Bigg\} \quad (16)$$

for the continuous-time setting and

$$\begin{aligned} \pi^*_{T-k} = & -\mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} \tilde{\alpha}'_{T-k} a_{T-k+1}]^{-1} \\ & \times \Big\{ x_{T-k} \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} r_{T-k} a_{T-k+1}] \\ & - l_{T-k} \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} \beta_{T-k} b_{T-k+1}] \\ & - \left(\mathfrak{B}^d + \frac{\mathfrak{C}^d}{(\mathfrak{B}^d - D)(1 - \mathrm{E}[S_{T,e}])} \right) \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1}] \Big\} \end{aligned} \quad (17)$$

for the discrete-time setting, respectively. Moreover, in the complete market, the optimal trading strategies in continuous-time and discrete-time setting become

$$\begin{aligned} u^*(t, S, l) = & -\big[(\sigma_A \sigma_A')(t)\big]^{-1} \Big\{ \tilde{\alpha}(t)' S - \mathfrak{B}^c \tilde{\alpha}(t)' e^{\int_t^T -r(\tau)d\tau} \\ & + \Big[\tilde{\alpha}(t)' \big(1 - e^{\int_t^T (-r+\beta-(\sigma_A\sigma_L)'(\sigma_A\sigma_A')^{-1}\tilde{\alpha})(\tau)d\tau}\big) \\ & - (\sigma_A \sigma_L)(t) \big(e^{\int_t^T (-r+\beta-(\sigma_A\sigma_L)'(\sigma_A\sigma_A')^{-1}\tilde{\alpha})(\tau)d\tau} \big) \Big] l \Big\} \end{aligned} \quad (18)$$

and

$$\begin{aligned} \pi^*_{T-k} = & -\mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} \tilde{\alpha}'_{T-k} a_{T-k+1}]^{-1} \\ & \times \big\{ x_{T-k} \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} r_{T-k} a_{T-k+1}] \\ & - l_{T-k} \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} \beta_{T-k} b_{T-k+1}] - \mathfrak{B}^d \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1}] \big\}, \end{aligned} \quad (19)$$

respectively.

Proof From (13), we can attack the problem ($\mathrm{P}_1(D)$) by solving the following optimization problem:

$$\begin{aligned} (\mathrm{P}_3(D)) \quad \min \quad & U(E) = \frac{\mathfrak{A}^{c,d}(E - \mathfrak{B}^{c,d})^2 + \mathfrak{C}^{c,d}}{(E - D)^2} \\ & = \mathfrak{A}^{c,d} \left(1 - \frac{\mathfrak{B}^{c,d} - D}{E - D} \right)^2 + \frac{\mathfrak{C}^{c,d}}{(E - D)^2}, \\ \text{s.t.} \quad & E > D, \end{aligned}$$

where E is the expected final surplus. By taking derivative of U with respect to E, we have

$$\frac{dU}{dE} = 2\mathfrak{A}^{c,d} \frac{\mathfrak{B}^{c,d} - D}{(E - D)^2} \left(1 - \frac{\mathfrak{B}^{c,d} - D}{E - D} \right) - 2 \frac{\mathfrak{C}^{c,d}}{(E - D)^3}. \quad (20)$$

When $\mathfrak{B}^{c,d} > D$ and $E > D$, E can be smaller than $\mathfrak{B}^{c,d}$, equal to $\mathfrak{B}^{c,d}$ or greater than $\mathfrak{B}^{c,d}$. Let $\widetilde{E}$ be the critical point. Solving

$$\left.\frac{dU}{dE}\right|_{E=\widetilde{E}} = 0$$

yields the following relation:

$$2\mathfrak{A}^{c,d}(\mathfrak{B}^{c,d} - D)\left(1 - \frac{\mathfrak{B}^{c,d} - D}{\widetilde{E} - D}\right) - \frac{2\mathfrak{C}^{c,d}}{\widetilde{E} - D} = 0,$$

i.e.,

$$\mathfrak{A}^{c,d}(\mathfrak{B}^{c,d} - D) = \frac{\mathfrak{A}^{c,d}(\mathfrak{B}^{c,d} - D)^2 + \mathfrak{C}^{c,d}}{\widetilde{E} - D}.$$

We further have

$$\widetilde{E} - D = \mathfrak{B}^{c,d} - D + \frac{\mathfrak{C}^{c,d}}{\mathfrak{A}^{c,d}(\mathfrak{B}^{c,d} - D)},$$

which gives rise to

$$\widetilde{E} = \mathfrak{B}^{c,d} + \frac{\mathfrak{C}^{c,d}}{\mathfrak{A}^{c,d}(\mathfrak{B}^{c,d} - D)} > \mathfrak{B}^{c,d} > D.$$

We also have

$$\frac{dU}{dE} = \begin{cases} < 0, & \text{if } E < \widetilde{E}, \\ = 0, & \text{if } E = \widetilde{E}, \\ > 0, & \text{if } E > \widetilde{E}. \end{cases}$$

Therefore, $\widetilde{E}$ is the minimizer of $U(E)$ and the optimal expected surplus E^* is equal to $\widetilde{E}$. The minimum value of $U(E)$ is $U(E^*) = \frac{\mathfrak{A}^{c,d}\mathfrak{C}^{c,d}}{\mathfrak{C}^{c,d} + \mathfrak{A}^{c,d}(\mathfrak{B}^{c,d} - D)^2}$. By substituting $E = \mathfrak{B}^c + \frac{\mathfrak{C}^c}{\mathfrak{A}^c(\mathfrak{B}^c - D)}$ into (6), we have

$$\begin{aligned} \frac{\lambda}{2\omega} &= -\mathfrak{B}^c\mathfrak{A}^c + \frac{\mathfrak{B}^c + \frac{\mathfrak{C}^c}{\mathfrak{A}^c(\mathfrak{B}^c - D)}}{1 - e^{-\int_0^T (\tilde{\alpha}'(\sigma_A\sigma_A')^{-1}\tilde{\alpha})(t)dt}} \\ &= \mathfrak{B}^c + \frac{\mathfrak{C}^c}{\mathfrak{A}^c(\mathfrak{B}^c - D)(1 - e^{-\int_0^T (\tilde{\alpha}'(\sigma_A\sigma_A')^{-1}\tilde{\alpha})(t)dt})} \\ &= \mathfrak{B}^c + \frac{\mathfrak{C}^c}{(\mathfrak{B}^c - D)e^{-\int_0^T (\tilde{\alpha}'(\sigma_A\sigma_A')^{-1}\tilde{\alpha})(t)dt}}. \end{aligned} \tag{21}$$

Substituting (21) further into (5) yields the continuous trading strategy in (16). For the discrete time scenario, substituting $E = \mathfrak{B}^d + \frac{\mathfrak{C}^d}{\mathfrak{A}^d(\mathfrak{B}^d - D)}$ into (12) gives rise to

$$\frac{\lambda}{2\omega} = \frac{\mathfrak{B}^d + \frac{\mathfrak{C}^d}{\mathfrak{A}^d(\mathfrak{B}^d - D)}}{\mathrm{E}[S_{T,e}]} - \frac{\mathrm{E}[S_{T,0}]}{\mathrm{E}[S_{T,e}]}$$

$$= \mathfrak{B}^d + \frac{\mathfrak{C}^d}{(\mathfrak{B}^d - D)(1 - \mathrm{E}[S_{T,e}])}. \tag{22}$$

Then, substituting (22) further into (11) yields the optimal periodic-trading strategy in (17). When $\mathfrak{C}^{c,d} \equiv 0$, the market is complete, see Chiu and Li [25]. Even though the uncontrollable liabilities exists, the investor is able to construct a portfolio to hedge the liabilities completely. Hence, by substituting $\mathfrak{C}^{c,d} \equiv 0$ into (16) and (17), the expression of the optimal trading strategy in a complete market becomes (18) for continuous-time or (19) for discrete-time, respectively. □

Theorem 3.2 reveals that the value of E^* is finite, when $D < \mathfrak{B}^{c,d}$. In such a situation, the optimal trading strategy can be uniquely determined by identifying E^*. This conclusion matches the result of Li et al. [13] and explains why existing papers only consider the disaster level to be less than the return of the minimum variance portfolio, $\mathfrak{B}^{c,d}$. Note, however, Roy [1] does not restrict his analysis only to this situation. Thus, we are now going to investigate the continuous and multiperiod trading strategies in both complete and incomplete market when the disaster level is not less than $\mathfrak{B}^{c,d}$.

Theorem 3.3 *Assume* $D \geq \mathfrak{B}^{c,d}$. *Then*:

(i) *If* $D = \mathfrak{B}^{c,d}$ *and the market is complete, the optimal solution* E^* *of* $(\mathrm{P}_3(D))$ *can be any value in the interval* (D, ∞).
(ii) *When* $D > \mathfrak{B}^{c,d}$, $(\mathrm{P}_3(D))$ *has no optimal solution.*
(iii) *Furthermore, when* $D = \mathfrak{B}^{c,d}$ *and* E *is specified at a value in* (D, ∞), *the optimal strategies for continuous and discrete setting of* $(\mathrm{P}_1(D))$ *are given as follows, respectively*:

$$\begin{aligned} u^*(t,S,l) = {} & -\big[(\sigma_A{\sigma_A}')(t)\big]^{-1}\big\{\tilde{\alpha}(t)S + \mathfrak{B}^c\mathfrak{A}^c\tilde{\alpha}(t)e^{\int_t^T -r(\tau)d\tau} \\ & + \big[\tilde{\alpha}(t)\big(1 - e^{\int_t^T(-r+\beta-(\sigma_A{\sigma_L}')'(\sigma_A{\sigma_A}')^{-1}\tilde{\alpha})(\tau)d\tau}\big) \\ & - (\sigma_A{\sigma_L}')(t)\big(e^{\int_t^T(-r+\beta-(\sigma_A{\sigma_L}')'(\sigma_A{\sigma_A}')^{-1}\tilde{\alpha})(\tau)d\tau}\big)\big]l\big\} \\ & + \frac{Ee^{\int_t^T -r(\tau)d\tau}}{1 - e^{-\int_0^T(\tilde{\alpha}'(\sigma_A{\sigma_A}')^{-1}\tilde{\alpha})(t)dt}}[(\sigma_A{\sigma_A}')(t)]^{-1}\tilde{\alpha}(t) \end{aligned} \tag{23}$$

and

$$\begin{aligned} \pi_{T-k}^* = {} & -\mathrm{E}[\tilde{\alpha}_{T-k}a_{T-k+1}\tilde{\alpha}'_{T-k}a_{T-k+1}]^{-1} \\ & \times \Big\{x_{T-k}\mathrm{E}[\tilde{\alpha}_{T-k}a_{T-k+1}r_{T-k}a_{T-k+1}] \\ & - l_{T-k}\mathrm{E}[\tilde{\alpha}_{T-k}a_{T-k+1}\beta_{T-k}b_{T-k+1}] \\ & + \frac{\mathrm{E}[S_{T,0}]}{\mathrm{E}[S_{T,e}]}\mathrm{E}[\tilde{\alpha}_{T-k}a_{T-k+1}]\Big\}, \\ & + \frac{E}{\mathrm{E}[S_{T,e}]}\mathrm{E}[\tilde{\alpha}_{T-k}a_{T-k+1}\tilde{\alpha}'_{T-k}a_{T-k+1}]^{-1}\mathrm{E}[\tilde{\alpha}_{T-k}a_{T-k+1}]. \end{aligned} \tag{24}$$

Proof When $D > \mathfrak{B}^{c,d}$, minimizing $U(E) = \frac{\mathfrak{A}^{c,d}(E-\mathfrak{B}^{c,d})^2+\mathfrak{C}^{c,d}}{(E-D)^2}$ over the entire real space yields an unconstrained global solution $\widetilde{E} = \mathfrak{B}^{c,d} + \frac{\mathfrak{C}^{c,d}}{\mathfrak{A}^{c,d}(\mathfrak{B}^{c,d}-D)}$. Clearly, $\widetilde{E} < D$. By analyzing the function $\frac{dU}{dE}$, we conclude that

$$\frac{dU}{dE} = \begin{cases} = 0, & \text{if } E = \widetilde{E}, \\ > 0, & \text{if } \widetilde{E} < E < D, \\ < 0, & \text{if } D < E. \end{cases}$$

Thus, $U(E)$ is a decreasing function on $(D, +\infty)$, which implies that E^*, the optimal solution of $\mathrm{P}_3(D)$, is at the infinity and

$$\inf U = \lim_{E\to\infty} U = \mathfrak{A}^{c,d} \geq 0.$$

When $D = \mathfrak{B}^{c,d}$ and $\mathfrak{C}^{c,d} > 0$, $U(E)$ can be simplified to $\mathfrak{A}^{c,d} + \frac{\mathfrak{C}^{c,d}}{(E-D)^2}$, which is also a decreasing function. Similarly, we have $E^* = \infty$ and

$$\inf U = \lim_{E\to\infty} U = \mathfrak{A}^{c,d} \geq 0.$$

From the above analysis, we can conclude that there is no solution for problem $(\mathrm{P}_3(D))$ when $D > \mathfrak{B}^{c,d}$ or $D = \mathfrak{B}^{c,d}$ in an incomplete market.

When $D = \mathfrak{B}^{c,d}$ and $\mathfrak{C}^{c,d} = 0$, then $U(E) \equiv \mathfrak{A}^{c,d}$ and any feasible solution is optimal for problem $(\mathrm{P}_3(D))$. This implies that, for any feasible solution E of $(\mathrm{P}_3(D))$ with $E > D$, the mean-variance efficient portfolios of continuous-time and discrete-time models given in (23) and (24), respectively, solve the safety-first criteria $(\mathrm{P}_1(D))$. □

Theorem 3.3 asserts that, when $D \geq \mathfrak{B}^{c,d}$ and $\mathfrak{C}^{c,d} > 0$, there is no optimal trading strategy due to the infinite value of E^*. It may explain why existing papers which study the Roy's safety-first model all assume that the disaster level is less than the return of the minimum variance portfolio. In other words, when $D \geq \mathfrak{B}^{c,d}$ and $\mathfrak{C}^{c,d} > 0$, the investor under the safety-first principle attempts to get an unlimited expected final surplus, even though she would face an unlimited risk. We may classify this type of investors as *greedy investors*.

3.1 Greedy and Nongreedy Investors

Theorems 3.2 and 3.3 show that the disaster level affects the existence of E^* and its corresponding optimal trading strategy. Clearly, $\mathfrak{B}^{c,d}$ is a watershed of disaster level. When $D < \mathfrak{B}^{c,d}$, E^* is finite and the disaster probability is bounded above by

$$\frac{\mathfrak{A}^{c,d}\mathfrak{C}^{c,d}}{\mathfrak{C}^{c,d} + \mathfrak{A}^{c,d}(\mathfrak{B}^{c,d} - D)^2}.$$

We classify these investors who choose the disaster level to be smaller than the return of the minimum variance portfolio *nongreedy investors* in this paper. Nongreedy investors avoid catastrophic loss in their investment by setting $D < \mathfrak{B}^{c,d}$. Accordingly, they obtain a finite expected final surplus,

$$E^* = \mathfrak{B}^{c,d} + \frac{\mathfrak{C}^{c,d}}{\mathfrak{A}^{c,d}(\mathfrak{B}^{c,d} - D)},$$

when they adopt the optimal trading strategy, either (16) or (17). Note that the expected final surplus E^* increases as the prescribed disaster level D increases and vice versa. Meanwhile, they have a corresponding finite variance

$$\mathrm{Var}[S_T]|_{E=E^*} = \frac{(\mathfrak{C}^{c,d})^2}{\mathfrak{A}^{c,d}(\mathfrak{B}^{c,d} - D)^2} + \mathfrak{C}^{c,d}.$$

In an incomplete market, if $D \geq \mathfrak{B}^{c,d}$, the expected final surplus becomes infinity according to Theorem 3.3. Also, the same consequence occurs in the complete market when $D > \mathfrak{B}^{c,d}$. However, when $D = \mathfrak{B}^{c,d}$ and $\mathfrak{C}^{c,d} = 0$, any mean-variance efficient portfolio is an optimal trading strategy in problem ($\mathrm{P}_1(D)$) if the investor pre-selects her expected final surplus to be greater than D. Even though they choose a very large value as their expected final surplus/risk averse coefficient, any mean-variance efficient portfolio is optimal in ($\mathrm{P}_1(D)$). Thus, we classify this kind investors as *greedy investors*. Although greedy investors adopt the safety-first principle as their doctrine, they actually target a mean-variance portfolio with an infinity expected surplus.

Note that Theorem 3.3 does not provide the optimal trading strategy in neither continuous-time setting nor discrete-time setting. We address now the issue of how an optimal portfolio for the greedy investors can be constructed, besides the situation where $D = \mathfrak{B}^{c,d}$ and $\mathfrak{C}^{c,d} = 0$.

Let us look at (23) and (24). For the continuous-time case, as E goes to positive infinity, the second term

$$\frac{E e^{\int_t^T -r(\tau)d\tau}}{1 - e^{-\int_0^T (\tilde{\alpha}'(\sigma_A\sigma_A')^{-1}\tilde{\alpha})(t)dt}} \big[(\sigma_A\sigma_A')(t)\big]^{-1}\tilde{\alpha}(t)$$

dominates the first term in the trading strategy (23). Because of the positivity of $\frac{E e^{\int_t^T -r(\tau)d\tau}}{1 - e^{-\int_0^T (\tilde{\alpha}'(\sigma_A\sigma_A')^{-1}\tilde{\alpha})(t)dt}}$, the optimal buy-and-sell trading strategy for the greedy investor is proportional to the vector

$$\big[(\sigma_A\sigma_A')(t)\big]^{-1}\tilde{\alpha}(t).$$

Here $(\sigma_A\sigma_A')^{-1}\tilde{\alpha}$ is related to the *Sharpe Ratio*, the excess return over variance. The greedy investor allocates her wealth according to the performance of stocks in terms of the modified Sharpe ratio, the higher the modified Sharpe ratio of a stock, the more the wealth allocated to that stock.

For example, if the jth element of $[(\sigma_A\sigma_A')(t)]^{-1}\tilde{\alpha}(t)$ is positive, the investor needs to hold the asset j long; otherwise, she needs to sell the asset j short. For the discrete-time model, the optimal buy-and-sell trading strategy for the greedy investor is proportional to the vector

$$\mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1} \tilde{\alpha}'_{T-k} a_{T-k+1}]^{-1} \mathrm{E}[\tilde{\alpha}_{T-k} a_{T-k+1}],$$

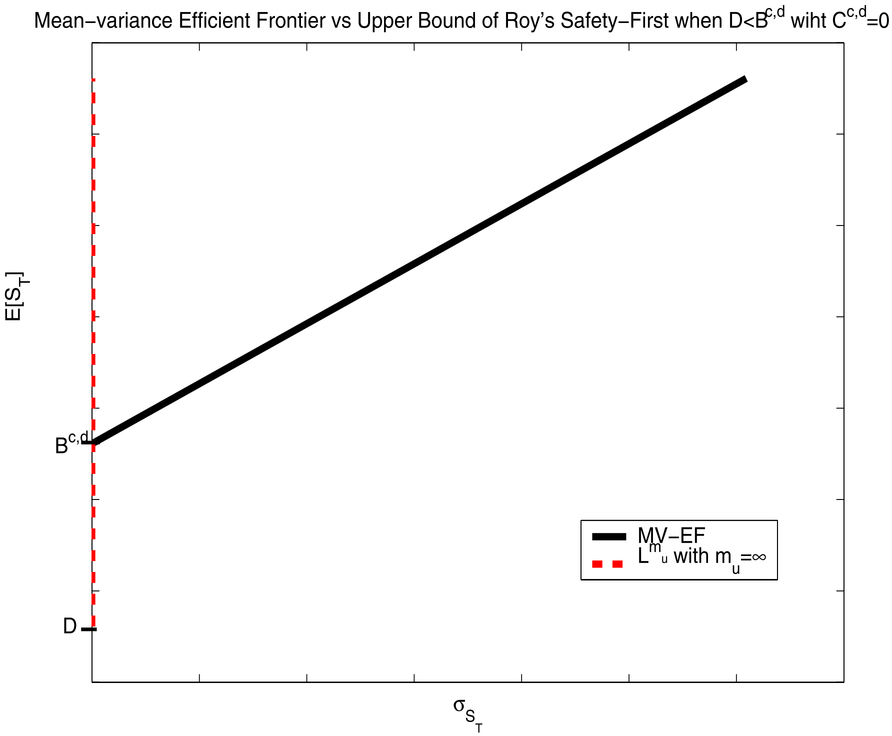

Fig. 1 When $\mathfrak{C}^{c,d}=0$ and $D<\mathfrak{B}^{c,d}$, $m_u^*=\infty$ and $E^*=\mathfrak{B}^{c,d}$

because the second term

$$\frac{E}{\mathrm{E}[S_{T,e}]}\mathrm{E}[\tilde{\alpha}_{T-k}a_{T-k+1}\tilde{\alpha}'_{T-k}a_{T-k+1}]^{-1}\mathrm{E}[\tilde{\alpha}_{T-k}a_{T-k+1}]$$

dominates the first term in the trading strategy (24).

4 Geometric Interpretation

We develop a geometric approach to further illustrate Theorems 3.2 and 3.3 in this section. There are several reasons to adopt a geometric approach. Firstly, the results derived before can be understood by readers who do not want to go through the tedious mathematical process. Secondly, the geometric approach is more intuitive and provides economic interpretations. Finally, it has been a usual practice to explain economic ideas via a graphical help of the mean-variance efficient frontier. Throughout this section, we take Theorem 3.1 for granted.

Let $m_u=\frac{\mathrm{E}[S_T]-D}{\sigma_{S_T}}|_u$ be the excess value over standard deviation for a given trading strategy u. Rearranging terms yields

$$\mathrm{E}[S_T]|_u=m_u\sigma_{S_T}|_u+D, \tag{25}$$

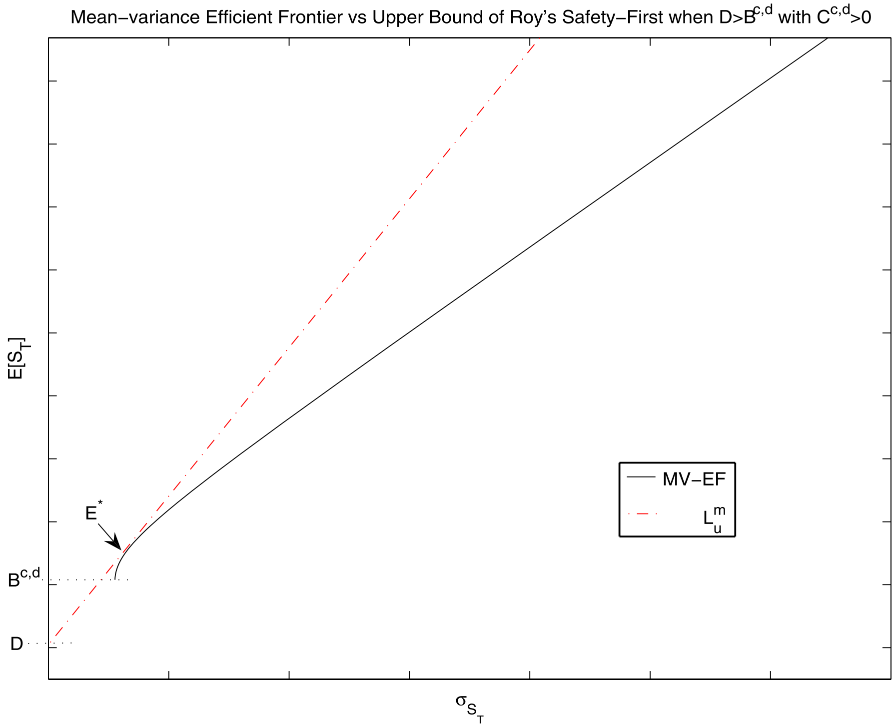

Fig. 2 When $\mathfrak{C}^{c,d} > 0$ and $D < \mathfrak{B}^{c,d}$, $E^* > \mathfrak{B}^{c,d}$ is the unique tangent point between MV-EF and L^{m_u}

which represents a straight line on the return-standard-deviation plane, with m_u being the slope and D being the intercept on the $\mathrm{E}[S_T]$-axis. We denote L^{m_u} as the set that contains all points on this straight line.

From the definition of m_u, $(\mathrm{P}_1(D))$ can be rewritten as

$$\begin{aligned}
\max_{u(\cdot)} \quad & m_u, \\
\text{s.t.} \quad & \mathrm{E}[S_T] > D, \\
& \mathfrak{W}, \\
& L^{m_u} : \mathrm{E}[S_T]|_u = m_u \sigma_{S_T}|_u + D. \qquad (26)
\end{aligned}$$

From Theorem 3.1, the optimal solution to $(\mathrm{P}_1(D))$ must be mean-variance efficient. Thus, we must have

$$(\sigma_{S_T}|_u)^2 = \mathfrak{A}^{c,d}(\mathrm{E}[S_T]|_u - \mathfrak{B}^{c,d})^2 + \mathfrak{C}^{c,d}.$$

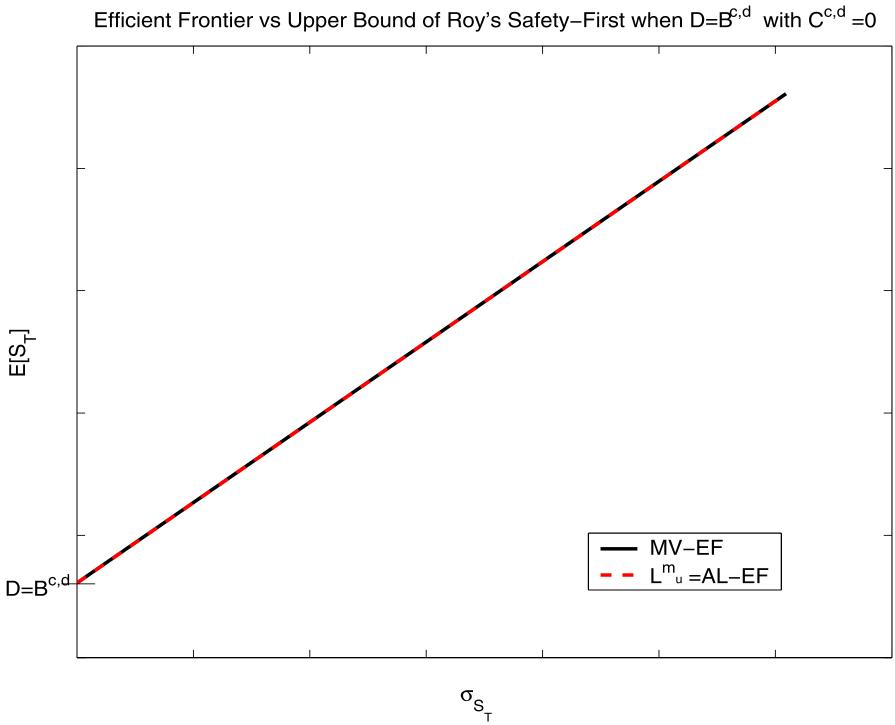

Fig. 3 When $\mathfrak{C}^{c,d}=0$ and $D=\mathfrak{B}^{c,d}$, m_u equals to the slope of MV-EF and L^{m_u} coincides with MV-EF

Thus, problem $(\mathrm{P}_1(D))$ should incorporate this constraint into its formulation, resulting in the following equivalent form:

$$
\begin{aligned}
(\mathrm{P}_4(D)) \quad & \max_{u(\cdot)} && m_u, \\
& \text{s.t.} && \mathrm{E}[S_T] > D, \\
& && L^{m_u}: \mathrm{E}[S_T]|_u = m_u \sigma_{S_T}|_u + D, \\
& && (\sigma_{S_T})^2 = \mathfrak{A}^{c,d}(\mathrm{E}[S_T]-\mathfrak{B}^{c,d})^2 + \mathfrak{C}^{c,d}.
\end{aligned}
$$

It is evident now that the optimal solution locates at the intersection of L^{m_u} and MV-EF.

We consider the situation in Theorem 3.2 with $\mathfrak{B}^{c,d} > D$. If the market is complete, i.e. $\mathfrak{C}^{c,d}=0$, then MV-EF is a straight line and $(\mathrm{P}_1(D))$ is reduced to

$$
\begin{aligned}
& \max_{u(\cdot)} && m_u, \\
& \text{s.t.} && \mathrm{E}[S_T] > D, \\
& && L^{m_u}: \mathrm{E}[S_T]|_u = m_u \sigma_{S_T}|_u + D, \\
& && \sigma_{S_T} = \sqrt{\mathfrak{A}^{c,d}}(\mathrm{E}[S_T]-\mathfrak{B}^{c,d}).
\end{aligned}
$$

Hence, $m_u=\infty$ is the maximum slope and the unique intersection point is $E=\mathfrak{B}^{c,d}$; see Fig. 1.

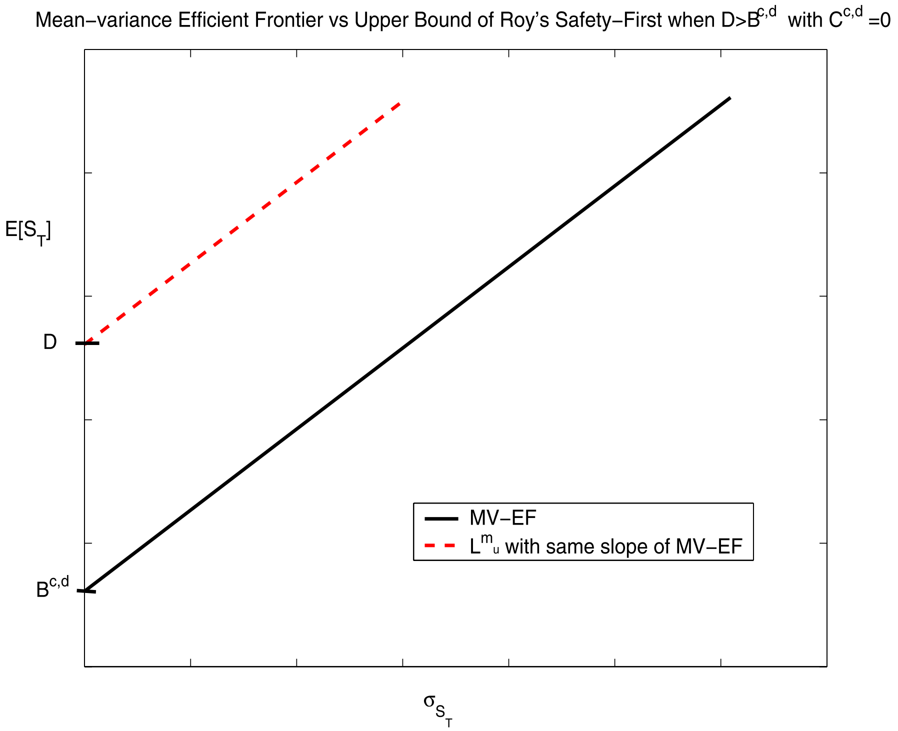

Fig. 4 When $\mathfrak{C}^{c,d} = 0$ and $D > \mathfrak{B}^{c,d}$, m_u equals to the slope of MV-EF and L^{m_u} intersects with MV-EF at infinity

For the incomplete market, i.e. $\mathfrak{C}^{c,d} > 0$, MV-EF is given by

$$(\sigma_{S_T})^2 = \mathfrak{A}^{c,d}(\mathrm{E}[S_T] - \mathfrak{B}^{c,d})^2 + \mathfrak{C}^{c,d},$$

which is a bullet-shaped curve. Obviously, the straight line

$$\mathrm{E}[S(T)] = \frac{1}{\sqrt{\mathfrak{A}^{c,d}}}\sigma_{S(T)} + \mathfrak{B}^{c,d}$$

is the asymptote of MV-EF for $\sigma_{S_T \to \infty}$.

If $D < \mathfrak{B}^{c,d}$, then E^*, which is larger than $\mathfrak{B}^{c,d}$, is the tangent point between MV-EF and L^{m_u}; see Fig. 2.

Figures 1 and 2 can be used to better our understanding of Theorem 3.2. We briefly summarize out findings again. In a complete market, $\mathfrak{C}^{c,d} = 0$ and Fig. 1 shows that $E^* = \mathfrak{B}^{c,d}$. In an incomplete market, the unique tangent point E^* is shown in Fig. 2. Besides, the value of this tangent point E^* can be easily computed as $\mathfrak{B}^{c,d} + \frac{\mathfrak{C}^{c,d}}{\mathfrak{A}^{c,d}(\mathfrak{B}^{c,d} - D)} > \mathfrak{B}^{c,d}$.

Next, we consider the situation in Theorem 3.3 with $D \geq \mathfrak{B}^{c,d}$. If the market is complete and $D = \mathfrak{B}^{c,d}$, then the maximum value of m_u equals the slope of MV-

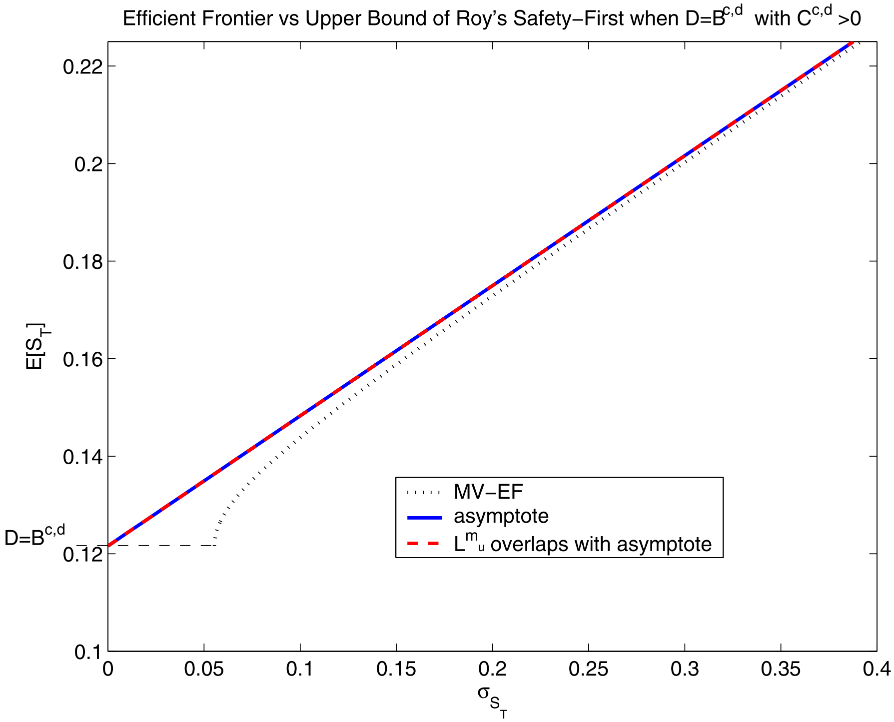

Fig. 5 When $\mathfrak{C}^{c,d} > 0$ and $D = \mathfrak{B}^{c,d}$, m_u equals to the slope of the asymptote of MV-EF, L^{m_u} merges to the asymptote, and L^{m_u} intersects with MV-EF at infinity

EF and L^{m_u} overlaps with MV-EF; see Fig. 3. There are infinitely many intersection points.

For the following cases, (i) $D > \mathfrak{B}^{c,d}$ and (ii) $D = \mathfrak{B}^{c,d}$ with $\mathfrak{C}^{c,d} > 0$, the intersection point of E^* MV-EF and L^{m_u} must be infinity; see Figs. 4, 5 and 6.

5 Conclusions

The asset-liability (AL) management problem under the Roy's safety-first principle has been studied in this paper. We have shown that the optimal trading policy locates on the mean-variance efficient frontier. When the preset disaster level is strictly less than the minimum-variance return, we have given an explicit expression of the optimal trading strategies for both continuous-time and discrete-time models. An investor who sets the disaster level less than the minimum-variance return is a genuine safety-first investor. We classify this type of investors as nongreedy investors. However, if the disaster level is set to be larger than the minimum-variance return, then the investor tends to seek for an infinite return accompanied by an infinite variance. We classify this type of investors as greedy investors. We have also provided an explicit optimal trading strategy for greedy investors. To demonstrate related economic interpretations, we have further illustrated our result by adopting a geometric approach.

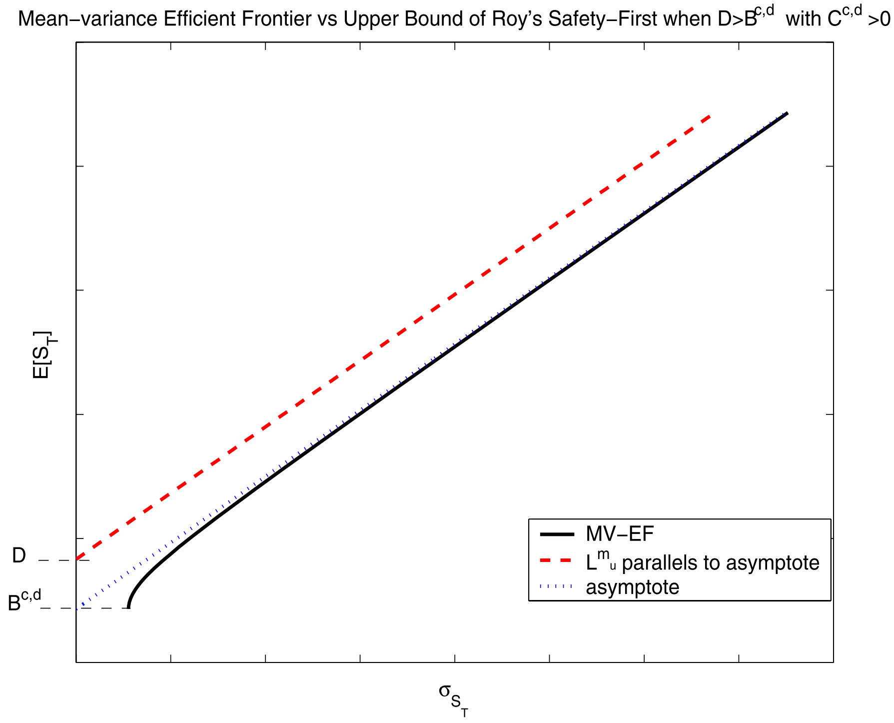

Fig. 6 When $\mathfrak{C}^{c,d} > 0$ and $D > \mathfrak{B}^{c,d}$, m_u equals to the slope of the asymptote of MV-EF, L^{m_u} is a *straight line* parallel to the asymptote, and L^{m_u} intersects with MV-EF at infinity

Acknowledgements We appreciate the careful reading and valuable comments from three anonymous referees and the Associate Editor. Special thank also goes to H.Y. Wong for insightful comments and suggestions.

References

1. Roy, A.D.: Safety first and the holding assets. Econometrica **20**, 431–449 (1952)
2. Jorion, P.: Value at Risk: The New Benchmark for Controlling Derivatives Risk. McGraw-Hill, New York (1997)
3. Markowitz, H.M.: Portfolio Selection: Efficient Diversification of Investment. Wiley, New York (1959)
4. Markowitz, H.M.: Mean-Variance Analysis in Portfolio Choice and Capital Markets. Blackwell, Cambridge (1989)
5. Markowitz, H.M.: The optimization of a quadratic function subject to linear constraints. Nav. Res. Logist. Q. **3**, 111–133 (1956)
6. Korn, R., Trautmann, S.: Continuous-time portfolio optimization under terminal wealth constraints. Math. Methods Oper. Res. **42**, 69–93 (1995)
7. Korn, R.: Optimal Portfolios—Stochastic Models for Optimal Investment and Risk Management in Continuous Time. World Scientific Publishing, Singapore (1997)
8. Bielecki, T.R., Jin, H., Pliska, S.R., Zhou, X.Y.: Continuous-time mean-variance portfolio selection with bankruptcy prohibition. Math. Financ. **15**, 213–244 (2005)
9. Levy, H., Sarnat, M.: Safety first—an expected utility principle. J. Financ. Quant. Anal. **7**, 1829–1834 (1971)

10. Arzac, E.R., Bawa, V.S.: Portfolio choice and equilibrium in capital markets with safety-first investors. J. Financ. Econ. **4**, 277–288 (1977)
11. Telser, L.G.: Safety first and hedging. Rev. Econ. Stud. **23**, 1–16 (1956)
12. Kataoka, S.: A stochastic programming model. Econometrica **31**, 181–196 (1963)
13. Li, D., Chan, T.F., Ng, W.L.: Safety-first dynamic portfolio selection. Dyn. Contin. Discrete Impuls. Syst. **4**, 585–600 (1998)
14. Li, D., Ng, W.L.: Optimal dynamic portfolio selection: multiperiod mean-variance formulation. Math. Financ. **10**, 387–406 (2000)
15. Milevsky, M.A.: Time diversification, safety-first and risk. Rev. Quant. Financ. Account. **12**, 271–281 (1999)
16. Chiu, M.C., Wong, H.Y., Li, D.: Martingale approach to Roy's safety-first principle and asset-liability management. Working Paper, Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong (2007)
17. Browne, S.: Optimal investment policies for a firm with a random risk process: exponential utility and minimizing the probability of ruin. Math. Oper. Res. **20**, 937–958 (1995)
18. Browne, S.: Survival and growth with a liability: optimal portfolio strategies in continuous time. Math. Oper. Res. **22**, 468–493 (1997)
19. Browne, S.: Reaching goals by a deadline: digital options and continuous-time active portfolio management. Adv. Appl. Probab. **31**, 551–577 (1999)
20. Harlow, W.: Asset allocation in a downside-risk framework. Financ. Anal. J. **47**, 28–40 (1991)
21. Jansen, W.D., Koedijk, K.G., Vries, C.G.: Portfolio selection with limited downside risk. J. Empir. Financ. **7**, 247–269 (2000)
22. Sharpe, W.F., Tint, L.G.: Liabilities-a new approach. J. Portf. Manag. **16**, 5–10 (1990)
23. Keel, A., Müller, H.: Efficient portfolios in the asset liability context. Astin Bull. **25**, 33–48 (1995)
24. Leippold, M., Trojani, F., Vanini, P.: A geometric approach to multiperiod mean variance optimization of assets and liabilities. J. Econ. Dyn. Control **28**, 1079–1113 (2004)
25. Chiu, M.C., Li, D.: Continuous-time mean-variance optimization of assets and liabilities. Insur. Math. Econ. **39**, 330–355 (2006)
26. Norberg, R.: Ruin problems with assets and liabilities of diffusion type. J. Stoch. Process. Their Appl. **81**, 255–269 (1999)
27. Josa-Fombellida, R., Rincón-Zapatero, J.P.: Optimal risk management in defined benefit stochastic pension funds. Insur. Math. Econ. **34**, 489–503 (2004)