Continuous-Time Markowitz's Model with Transaction Costs*

Min Dai† Zuo Quan Xu‡ Xun Yu Zhou§

October 7, 2009

Abstract

A continuous-time Markowitz's mean-variance portfolio selection problem is studied in a market with one stock, one bond, and proportional transaction costs. This is a singular stochastic control problem, inherently with a finite time horizon. Via a series of transformations, the problem is turned into a so-called double obstacle problem, a well studied problem in physics and partial differential equation literature, featuring two time-varying free boundaries. The two boundaries, which define the buy, sell, and no-trade regions, are proved to be smooth in time. This in turn characterizes the optimal strategy, via a Skorokhod problem, as one that tries to keep a certain adjusted bond–stock position within the no-trade region. Several features of the optimal strategy are revealed that are remarkably different from its no-transaction-cost counterpart. It is shown that there exists a critical length in time, which is dependent on the stock excess return as well as the transaction fees but *independent* of the investment target and the stock volatility, so that an expected terminal return may not be achievable if the planning horizon is shorter than that critical length (while in the absence of transaction costs any expected return can be reached in an arbitrary period of time). It is further demonstrated that anyone following the optimal strategy should not buy the stock beyond the point when the time to maturity is shorter than the aforementioned critical length. Moreover, the investor would be less likely to buy the stock and more likely to sell the stock when the maturity date is getting closer. These features, while consistent with the widely accepted investment wisdom, suggest that the planning horizon is an integral part of the investment opportunities.

Key Words. continuous time, mean-variance, transaction costs, singular stochastic control, planning horizon, Lagrange multiplier, double-obstacle problem, Skorokhod problem

*Dai is partially supported by Singapore MOE AcRF grant (No. R-146-000-096-112) and NUS RMI grant (No. R-146-000-117-720/646). Zhou acknowledges financial support from Nomura Centre for Mathematical Finance and a start-up fund of the University of Oxford, and both Xu and Zhou acknowledge research grants from the Oxford–Man Institute of Quantitative Finance. The authors thank the referees and the AE for their constructive comments which have led to a much improved version of the paper.

†Department of Mathematics, National University of Singapore (NUS), Singapore. Also an affiliated member of Risk Management Institute of NUS.

‡Mathematical Institute and Nomura Centre for Mathematical Finance, University of Oxford, 24–29 St Giles, Oxford OX1 3LB, UK.

§Mathematical Institute and Nomura Centre for Mathematical Finance, and Oxford–Man Institute of Quantitative Finance, The University of Oxford, 24–29 St Giles, Oxford OX1 3LB, UK, and Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong. Email: <zhouxy@maths.ox.ac.uk>.

1 Introduction

Markowitz's (single-period) mean–variance (MV) portfolio selection model [Markowitz (1952)] marked the start of the modern quantitative finance theory. Extensions to the dynamic – especially continuous-time – setting in the asset allocation literature have, however, been dominated by the expected utility maximization (EUM) models, which take a considerable departure from the MV model. While the utility approach was theoretically justified by von Neumann and Morgenstern (1947), in practice "few if any investors know their utility functions; nor do the functions which financial engineers and financial economists find analytically convenient necessarily represent a particular investor's attitude towards risk and return" [Markowitz and Zhou (2004)]. Meanwhile, there are technical and conceptual difficulties in studying a dynamic MV model. In particular, an optimal trading strategy generated initially may no longer be optimal half way through. This so-called *time inconsistency* means that dynamic programming – which is the main tool for solving dynamic optimization problems – is not directly applicable. Furthermore, one could argue that it would be hard for an investor to follow a time-inconsistent strategy. Kydland and Prescott (1977) instead argue that time-inconsistent solutions are economically meaningful if the investor can commit at the initial time to follow a strategy (called a *pre-committed strategy*).

Time inconsistent control problems have recently attracted some interest; see Björk and Murgoci (2008) and Ekeland and Lazrak (2007). We note also that there are problems other than dynamic MV analysis that are inherently time inconsistent. For example, a dynamic behavioral portfolio selection problem is time inconsistent due to the distortions in probabilities [Jin and Zhou (2008)].

Basak and Chabakauri (2008) specifically address the time inconsistency problem in MV analysis by proposing the construction of a trading strategy that is locally optimal in an MV sense and time-consistent, although it is not globally optimal in the sense of Problem 2.1 (to be formulated in Section 2). Basak and Chabakauri (2008) further show that their strategy solves a global optimization problem with a state-dependent CARA utility function.

In this paper, we solve the global Problem 2.1, and do so when trading is subject to transaction costs. The solution obtained is pre-committed, instead of time-consistent. We solve the problem by reformulating it in a way that makes it amenable to dynamic programming.

Richardson (1989) is probably the earliest paper that studies a faithful extension of the MV model to the continuous-time setting (albeit in the context of a single stock with a constant risk-free rate), followed by Bajeux-Besnainou and Portait (1998). Li and Ng (2000), in a discrete-time setting, developed an embedding technique to change the originally time-inconsistent MV problem into a stochastic LQ control problem. This technique was extended by Zhou and Li (2000), along with a stochastic linear–quadratic control approach, to the continuous-time case.[1] Further extensions and improvements are carried out in, among many others, Lim and Zhou (2002), Lim (2004), Bielecki *et al.* (2005), and Xia (2005).

All the existing works on continuous-time MV models have assumed that there is no transaction cost, leading to results that are analytically elegant, and sometimes truly surprising [for example, it is shown in Li and Zhou (2006) that any efficient strategy realizes its goal – no matter how high it is – with a probability of at least 80%]. However, elegant they may be, certain investment behaviors derived from the results simply contradict the conventional wisdom, which in turn hints that the models may not have been properly formulated. For instance, the results

[1] One should note that the trading strategies derived in Li and Ng (2000) and Zhou and Li (2000) are not time-consistent.

dictate that an optimal strategy must trade all the time; moreover, there must be risky exposures at any time [see Chiu and Zhou (2009)]. These are certainly not consistent with the common investment advice. Indeed, the assumption that there is no transaction cost is flawed, which misleadingly allows an investor to continuously trade without any penalty.

Portfolio selection subject to transaction costs has been studied extensively, albeit in the realm of utility maximization. Mathematically such a problem is a singular stochastic control problem. Two different types of models must be distinguished: one in an infinite planning horizon and the other in a finite horizon. See Magill and Constantinides (1976), Davis and Norman (1990), and Shreve and Soner (1994) for the former, and Davis, Panas and Zariphopoulou (1993), Cvitanic and Karatzas (1996), and Gennotte and Jung (1994) for the latter. Technically, the latter is substantially more difficult than the former, since in the finite horizon case there is an additional time variable in the related Hamilton-Jacobi-Bellman (HJB) equation or variational inequality (VI). This is why the research on finite-horizon problems had been predominantly on qualitative and numerical solutions until Liu and Loewenstein (2002) devised an analytical approach based on an approximation of the finite horizon by a sequence of Erlang distributed random horizons. Dai and Yi (2009) subsequently employed a different analytical approach – a PDE one – to study the same problem.

This paper aims to analytically solve the MV model with transaction costs. Note that such a problem is inherently one in a finite time horizon, because the very nature of the Markowitz problem is about striking a balance between the risk and return of the wealth at a finite, *terminal* time. Compared with its EUM counterpart, there is a feasibility issue that must be addressed before an optimal solution is sought. Precisely speaking, the MV model is to minimize the variance of the terminal wealth subject to the constraint that an investment target – certain expected net terminal wealth – is achieved. The feasibility is about whether such a target is achievable by at least one admissible investment strategy. For a Black–Scholes market without transaction costs, it has been shown [Lim and Zhou (2002)] that *any* target can be reached in an arbitrary length of time (so long as the risk involved is not a concern, that is). For a more complicated model with random investment opportunities and no-bankruptcy constraint, the feasibility is painstakingly investigated in Bielecki *et al* (2005). In this paper we show that the length of the planning horizon is a determinant of this issue. In fact, there exists a critical length of time, which is dependent *only* on the stock excess return and the transaction fees, so that a sufficiently high target is not achievable if the planning horizon is shorter than that critical length. This certainly makes good sense intuitively.

To obtain an optimal strategy, technically we follow the idea of Dai and Yi (2009) of eventually turning the associated VI into a double-obstacle problem, a problem that has been well studied in physics and PDE theory. That said, there are indeed intriguing subtleties when actually carrying it out. In particular, this paper is the first to prove (to the best of the authors' knowledge) that the two free boundaries that define the buy, sell and no-trade regions are smooth. This smoothness is critical in deriving the optimal strategy via a Skorokhod problem.[2] The optimal strategy is rather simple in implementation; it is to keep a certain adjusted bond–stock position within the no-trade region. Several features of the optimal strategy are revealed that are remarkably different from its no-transaction-cost counterpart. Among them it is notable that one should no longer buy stock beyond the point when the time to maturity is shorter than the aforementioned critical length associated with the feasibility. Moreover,

[2]This smoothness also plays a crucial role in studying the finite horizon optimal investment and consumption with transaction costs under utility framework, see Section 3 of Dai et al. (2009) where an integral over one free boundary is used.

one is less likely to buy the stock and more likely to sell the stock when the maturity date is getting closer. These are consistent with the widely accepted financial advice, and suggest that the planning horizon should be regarded as a part of the investment opportunity set when it comes to continuous time portfolio selection.

The remainder of the paper is organized as follows. The model under consideration is formulated in section 2, and the feasibility issue is addressed in section 3. The optimal strategy is derived in sections 4–6 via several steps, including Lagrange relaxation, transformation of the HJB equation to a double obstacle problem, and the Skorokhod problem. Finally, the paper is concluded with remarks in section 7. Some technical proofs are relegated to an appendix.

2 Problem Formulation

We consider a continuous-time market where there are only two investment instruments: a bond and a stock with price dynamics given respectively by

$$\begin{aligned} \mathrm{d}R(t) &= rR(t)\,\mathrm{d}t, \\ \mathrm{d}S(t) &= \alpha S(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}B(t). \end{aligned}$$

Here $r > 0$, $\alpha > r$ and $\sigma > 0$ are constants, and the process $\{B(t)\}_{t\in[0,T]}$ is a standard one-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbf{P})$ with $B(0) = 0$ almost surely. We assume that the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ is generated by the Brownian motion, is right continuous, and each $\mathcal{F}_t$ contains all the $\mathbf{P}$-null sets of $\mathcal{F}$. We denote by $L^2_{\mathcal{F}}$ the set of square integrable $\{\mathcal{F}_t\}_{t\in[0,T]}$-adapted processes,

$$L^2_{\mathcal{F}} \stackrel{\text{def}}{=\!=} \left\{ X \;\middle|\; \begin{array}{l} \text{The process } X = \{X(t)\}_{t\in[0,T]} \text{ is an } \{\mathcal{F}_t\}_{t\in[0,T]}\text{-} \\ \text{adapted process such that } \int_0^T \mathbf{E}\left[X^2(t)\right]\mathrm{d}t < \infty \end{array} \right\},$$

and by $L^2_{\mathcal{F}_T}$ the set of square integrable $\mathcal{F}_T$-measurable random variables,

$$L^2_{\mathcal{F}_T} \stackrel{\text{def}}{=\!=} \left\{ X \;\middle|\; \text{X is an } \mathcal{F}_T\text{-measurable random variable such that } \mathbf{E}\left[X^2\right] < \infty \right\}.$$

There is a self-financing investor with a finite investment horizon $[0, T]$ who invests $X(t)$ dollars in the bond and $Y(t)$ dollars in the stock at time t. Any stock transaction incurs a *proportional transaction fee*, with $\lambda \in [0, +\infty)$ and $\mu \in [0, 1)$ being the proportions paid when buying and selling the stock, respectively. Throughout this paper, we assume that $\lambda + \mu > 0$, which means transaction costs must be involved. The bond–stock value process, starting from (x, y) at $t = 0$, evolves according to the equations:

$$X^{x,M,N}(t) = x + r\int_0^t X^{x,M,N}(s)\,\mathrm{d}s - (1+\lambda)M(t) + (1-\mu)N(t), \tag{2.1}$$

$$Y^{y,M,N}(t) = y + \alpha\int_0^t Y^{y,M,N}(s)\,\mathrm{d}s + \sigma\int_0^t Y^{y,M,N}(s)\,\mathrm{d}B(s) + M(t) - N(t), \tag{2.2}$$

where $M(t)$ and $N(t)$ denote respectively the cumulative stock purchase and sell up to time t. Sometimes we simply use X, Y or $X^{M,N}$, $Y^{M,N}$ instead of $X^{x,M,N}$, $Y^{y,M,N}$ if there is no ambiguity.

The *admissible strategy set* $\mathcal{A}$ of the investor is defined as follows:

$$\mathcal{A} \stackrel{\text{def}}{=\!=} \left\{ (M,N) \;\middle|\; \begin{array}{l} \text{The processes } M = \{M(t)\}_{t\in[0,T]} \text{ and } N = \{N(t)\}_{t\in[0,T]} \text{ are} \\ \{\mathcal{F}_t\}_{t\in[0,T]}\text{-adapted, RCLL, nonnegative and nondecreasing, and the} \\ \text{processes } X^{x,M,N} \text{ and } Y^{y,M,N} \text{ are both in } L^2_{\mathcal{F}}, \text{ for any } (x,y) \in \mathbb{R}^2 \end{array} \right\}.$$

(M, N) is called *an admissible strategy* if $(M, N) \in \mathcal{A}$. Correspondingly, $(X^{x,M,N}, Y^{y,M,N})$ is called an *admissible (bond–stock) process* if $(x, y) \in \mathbb{R}^2$ and $(M, N) \in \mathcal{A}$.

For an admissible process $(X^{x,M,N}, Y^{y,M,N})$, we define the investor's *net wealth process* by

$$W^{X,Y}(t) \stackrel{\text{def}}{=\!=} X(t) + (1-\mu)Y(t)^+ - (1+\lambda)Y(t)^-, \quad t \in [0, T].$$

Namely, $W^{X,Y}(t)$ is the net worth of the investor's portfolio at t after the transaction cost is deducted. The investor's *attainable net wealth set* at the maturity time T is defined as

$$\mathcal{W}_0^{x,y} \stackrel{\text{def}}{=\!=} \left\{ W^{X,Y}(T) \,\middle|\, \begin{array}{l} W^{X,Y}(T) \text{ is the net wealth at } T \text{ of an ad-} \\ \text{missible process } (X, Y) \text{ with } X(0-) = x, \\ Y(0-) = y. \end{array} \right\}.$$

In the spirit of the original Markowitz's MV portfolio theory, an *efficient strategy* is a trading strategy for which there does not exist another strategy that has higher mean and no higher variance, and/or has less variance and no less mean at the terminal time T. In other words, an efficient strategy is one that is *Pareto optimal*. Clearly, there could be many efficient strategies, and the terminal means and variances corresponding to all the efficient strategies form an *efficient frontier*. The positioning on the efficient frontier of a specific investor is dictated by his/her risk preference.

It is now well known that the efficient frontier can be obtained from solving the following *variance minimizing* problem:

Problem 2.1.

$$\begin{array}{ll} \text{Minimize} & \mathbf{Var}(W), \\ \text{subject to} & \mathbf{E}[W] = z, \;\; W \in \mathcal{W}_0^{x,y}. \end{array}$$

Here z is a parameter satisfying

$$z > e^{rT}x + (1-\mu)e^{rT}y^+ - (1+\lambda)e^{rT}y^-,$$

which means that the target expected terminal wealth is higher than that of the simple "all-bond" strategy (i.e. initially liquidating the stock investment and putting all the money in the bond account). The optimal solutions to the above problem with varying values of z will trace out the efficient frontier we are looking for. For this reason, although Problem 2.1 is indeed an auxilary mathematical problem introduced to help solve the original mean–variance problem, it is sometimes (as in this paper) itself called the mean–variance problem.

It is immediate to see that Problem 2.1 is equivalent to the following problem.

Problem 2.2.

$$\begin{array}{ll} \text{Minimize} & \mathbf{E}[W^2], \\ \text{subject to} & \mathbf{E}[W] = z, \;\; W \in \mathcal{W}_0^{x,y}. \end{array}$$

3 Feasibility

In contrast with the EUM problem, the MV model, Problem 2.2, has an inherent constraint $\mathbf{E}[W] = z$. Is there always an admissible strategy to meet this constraint no matter how aggressive the target z is? This is the so-called

feasibility issue. The issue is important and unique to the MV problem, and will be addressed fully in this section. To begin with, we introduce two lemmas.

Lemma 3.1. *If $W_1 \in \mathcal{W}_0^{x,y}$, $W_2 \in L^2_{\mathcal{F}_T}$ and $W_2 \leqslant W_1$, then $W_2 \in \mathcal{W}_0^{x,y}$.*

Proof. By the definition of $\mathcal{W}_0^{x,y}$, there exists $(M,N) \in \mathcal{A}$ such that $X^{M,N}(0-) = x$, $Y^{M,N}(0-) = y$ and $W^{X^{M,N},Y^{M,N}}(T) = W_1$. We define

$$\overline{M}(t) = \begin{cases} M(t), & \text{if } t < T, \\ M(T) + \dfrac{W_1 - W_2}{\lambda + \mu}, & \text{if } t = T, \end{cases} \qquad \overline{N}(t) = \begin{cases} N(t), & \text{if } t < T, \\ N(T) + \dfrac{W_1 - W_2}{\lambda + \mu}, & \text{if } t = T. \end{cases}$$

Then $(\overline{M}, \overline{N}) \in \mathcal{A}$ and

$$X^{\overline{M},\overline{N}}(t) = \begin{cases} X^{M,N}(t), & \text{if } t < T, \\ X^{M,N}(T) - W_1 + W_2, & \text{if } t = T, \end{cases} \qquad Y^{\overline{M},\overline{N}}(t) = Y^{M,N}(t), \quad t \in [0,T].$$

Therefore $W_2 = W^{X^{\overline{M},\overline{N}},Y^{\overline{M},\overline{N}}}(T) \in \mathcal{W}_0^{x,y}$. □

The proof above is very intuitive. If a higher terminal wealth is achievable by an admissible strategy, then so is a lower one, by simply "wasting money", i.e., buying and selling the same amount of the stock at T, thanks to the presence of the transaction costs. This is *not* necessarily true when there is no transaction cost.

Lemma 3.2. *For any $(x,y) \in \mathbb{R}^2$, we have (1) the set $\mathcal{W}_0^{x,y}$ is convex; (2) If $(x_i,y_i) \in \mathbb{R}^2$, $W_i \in \mathcal{W}_0^{x_i,y_i}$, $i = 1,2$, then $W_1 + W_2 \in \mathcal{W}_0^{x_1+x_2,y_1+y_2}$; (3) if $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$, then $\mathcal{W}_0^{x_1,y_1} \subseteq \mathcal{W}_0^{x_2,y_2}$; (4) $\mathcal{W}_0^{x-(1+\lambda)\rho,y+\rho} \subseteq \mathcal{W}_0^{x,y}$ and $\mathcal{W}_0^{x+(1-\mu)\rho,y-\rho} \subseteq \mathcal{W}_0^{x,y}$ for any $\rho > 0$; (5) $\mathcal{W}_0^{\rho x,\rho y} = \rho \mathcal{W}_0^{x,y}$ for any $\rho > 0$; (6) if $x + (1-\mu)y^+ - (1+\lambda)y^- \geqslant 0$, then $0 \in \mathcal{W}_0^{x,y}$.*

Proof. (1) For any W_1, $W_2 \in \mathcal{W}_0^{x,y}$, assume that $W_i = W^{X_i,Y_i}(T)$, where $(X_i,Y_i) = (X^{x,M_i,N_i}, Y^{y,M_i,N_i})$, $(M_i,N_i) \in \mathcal{A}$, $i = 1,2$. For any $k \in (0,1)$, let $M = kM_1 + (1-k)M_2$, $N = kN_1 + (1-k)N_2$. Then $(M,N) \in \mathcal{A}$, and

$$X^{x,M,N} = kX_1 + (1-k)X_2,$$
$$Y^{y,M,N} = kY_1 + (1-k)Y_2.$$

Thus

$$\begin{aligned} W^{X,Y}(T) &= X(T) + (1-\mu)Y(T)^+ - (1+\lambda)Y(T)^- = X(T) + (1-\mu)Y(T) - (\mu+\lambda)Y(T)^- \\ &\geqslant kX_1(T) + (1-k)X_2(T) + (1-\mu)(kY_1(T) + (1-k)Y_2(T)) \\ &\quad - (\mu+\lambda)(kY_1(T)^- + (1-k)Y_2(T)^-) \\ &= k(X_1(T) + (1-\mu)Y_1(T)^+ - (1+\lambda)Y_1(T)^-) \\ &\quad + (1-k)(X_2(T) + (1-\mu)Y_2(T)^+ - (1+\lambda)Y_2(T)^-) \\ &= kW_1 + (1-k)W_2. \end{aligned}$$

Since $W^{X,Y}(T) \in \mathcal{W}_0^{x,y}$, $kW_1 + (1-k)W_2 \in L^2_{\mathcal{F}_T}$, it follows from Lemma 3.1 that $kW_1 + (1-k)W_2 \in \mathcal{W}_0^{x,y}$.

(2) This can be proved by the same argument as above.

(3) If $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$, then for any $(M,N) \in \mathcal{A}$, we have $X^{x_1,M,N} \leqslant X^{x_2,M,N}$, $Y^{y_1,M,N} \leqslant Y^{y_2,M,N}$. So $W^{X^{x_1,M,N},Y^{y_1,M,N}}(T) \leqslant W^{X^{x_2,M,N},Y^{y_2,M,N}}(T) \in \mathcal{W}_0^{x_2,y_2}$. Therefore by Lemma 3.1, we have $W^{X^{x_1,M,N},Y^{y_1,M,N}}(T) \in \mathcal{W}_0^{x_2,y_2}$. This shows $\mathcal{W}_0^{x_1,y_1} \subseteq \mathcal{W}_0^{x_2,y_2}$.

(4) For any $\rho > 0$, $(M,N) \in \mathcal{A}$, we define

$$\overline{M}(s) = M(s) + \rho, \quad \forall s \in [0,T].$$

Then $(X^{x-(1+\lambda)\rho,M,N}, Y^{y+\rho,M,N}) = (X^{x,\overline{M},N}, Y^{y,\overline{M},N})$. So $W^{X^{x-(1+\lambda)\rho,M,N},Y^{y+\rho,M,N}} = W^{X^{x,\overline{M},N},Y^{y,\overline{M},N}} \in \mathcal{W}_0^{x,y}$. Hence $\mathcal{W}_0^{x-(1+\lambda)\rho,y+\rho} \subseteq \mathcal{W}_0^{x,y}$. Similarly, we can prove that $\mathcal{W}_0^{x+(1-\mu)\rho,y-\rho} \subseteq \mathcal{W}_0^{x,y}$.

(5) Noting that $W^{X^{\rho x,\rho M,\rho N},Y^{\rho y,\rho M,\rho N}} = \rho W^{X^{x,M,N},Y^{y,M,N}}$ $\forall \rho > 0$, we have immediately $\mathcal{W}_0^{\rho x,\rho y} = \rho \mathcal{W}_0^{x,y}$.

(6) If $y \geqslant 0$ and $x + (1-\mu)y \geqslant 0$, then obviously $0 \in \mathcal{W}_0^{x+(1-\mu)y,0}$ by Lemma 3.1. Noting (4) proved above, we conclude that $0 \in \mathcal{W}_0^{x,y}$. Similarly we can prove the case of $y < 0$ and $x + (1+\lambda)y \geqslant 0$. □

Denote

$$\hat{z} \overset{\text{def}}{=\!=} \sup\left\{\mathbf{E}[W] \,\middle|\, W \in \mathcal{W}_0^{x,y}\right\}. \tag{3.1}$$

In view of Lemma 3.1, Problem 2.2 is feasible when

$$z \in \mathcal{D} \overset{\text{def}}{=\!=} \left(e^{rT}x + (1-\mu)e^{rT}y^+ - (1+\lambda)e^{rT}y^-,\ \hat{z}\right). \tag{3.2}$$

It is clear that Problem 2.2 is not feasible when $z > \hat{z}$. So, it remains to investigate whether Problem 2.2 admits a feasible solution when $z = \hat{z}$.

It is well known that in the absence of transaction costs (i.e. $\lambda = \mu = 0$), we have $\hat{z} = +\infty$ and thus Problem 2.2 is always feasible for any $z \geqslant e^{rT}x + (1-\mu)e^{rT}y^+ - (1+\lambda)e^{rT}y^-$ [see Lim and Zhou (2002)]. In other words, no matter how *small* the investor's initial wealth is, the investor can always arrive at an *arbitrarily* large expected return in a *split second*, by taking a huge leverage on the stock. The following theorem indicates that things become very different when the transaction costs get involved.

Theorem 3.3. *Assume* $T \leqslant T^* \overset{\text{def}}{=\!=} \frac{1}{\alpha - r}\ln\left(\frac{1+\lambda}{1-\mu}\right)$. *Then*

$$\hat{z} = \begin{cases} e^{rT}x + (1-\mu)e^{\alpha T}y, & \text{if } y > 0, \\ e^{rT}x + (1+\lambda)e^{rT}y, & \text{if } y \leqslant 0. \end{cases}$$

Moreover, if $y > 0$ *and* $z = \hat{z}$, *then Problem 2.1 admits a unique feasible (thus optimal) solution and the optimal strategy is* $(M,N) \equiv (0,0)$. *If* $y \leqslant 0$, *then* $\mathcal{D} = \emptyset$.

Proof. For any $(M, N) \in \mathcal{A}$, due to (2.1), (2.2) and Itô's formula, we have

$$\begin{aligned}
X(T) + (1-\mu)Y(T) &= e^{rT}x + (1-\mu)e^{\alpha T}y + \int_0^T e^{r(T-t)}\left(\mathrm{d}X(t) - rX(t)\,\mathrm{d}t\right) \\
&\quad + \int_0^T (1-\mu)e^{\alpha(T-t)}\left(\mathrm{d}Y(t) - \alpha Y(t)\,\mathrm{d}t\right) \\
&= e^{rT}x + (1-\mu)e^{\alpha T}y + \int_0^T \left((1-\mu)e^{\alpha(T-t)} - (1+\lambda)e^{r(T-t)}\right)\mathrm{d}M^c(t) \\
&\quad + \int_0^T (1-\mu)\left(e^{r(T-t)} - e^{\alpha(T-t)}\right)\mathrm{d}N^c(t) + \int_0^T (1-\mu)e^{\alpha(T-t)}\sigma Y(t)\,\mathrm{d}B(t) \\
&\quad + \sum_{0\leqslant t\leqslant T}\left((1-\mu)e^{\alpha(T-t)} - (1+\lambda)e^{r(T-t)}\right)\left(M(t) - M(t-)\right) \\
&\quad + \sum_{0\leqslant t\leqslant T}(1-\mu)\left(e^{r(T-t)} - e^{\alpha(T-t)}\right)\left(N(t) - N(t-)\right) \\
&\leqslant e^{rT}x + (1-\mu)e^{\alpha T}y + \int_0^T (1-\mu)e^{\alpha(T-t)}\sigma Y(t)\,\mathrm{d}B(t),
\end{aligned}$$

where we have noted $T \leqslant \frac{1}{\alpha - r}\ln\left(\frac{1+\lambda}{1-\mu}\right)$. So

$$\mathbf{E}[X(T) + (1-\mu)Y(T)] \leqslant e^{rT}x + (1-\mu)e^{\alpha T}y.$$

It follows

$$\mathbf{E}\left[W^{X,Y}\right] = \mathbf{E}[X(T) + (1-\mu)Y(T)^+ - (1+\lambda)Y(T)^-] \leqslant \mathbf{E}[X(T) + (1-\mu)Y(T)] \leqslant e^{rT}x + (1-\mu)e^{\alpha T}y.$$

When $y \geqslant 0$, we have $\mathbf{E}[W^{X,Y}] = e^{rT}x + (1-\mu)e^{\alpha T}y$ if and only if $(M, N) \equiv (0, 0)$. Thus, if $y \geqslant 0$, then $\hat{z} = e^{rT}x + (1-\mu)e^{\alpha T}y$, and $(0, 0)$ is the unique feasible strategy when $z = \hat{z}$.

Now we turn to the case of $y < 0$. Define $\tau = \inf\{t \in [0, T) : Y(t) > 0\} \wedge T$. Then τ is a stopping time [cf. Theorem 2.33, Klebaner (2004)]. On the set $\{Y(\tau) \geqslant 0\}$, by the same argument as above, we have

$$W_T^{X,Y} \leqslant e^{r(T-\tau)}X(\tau) + (1-\mu)e^{\alpha(T-\tau)}Y(\tau) \leqslant e^{r(T-\tau)}X(\tau) + (1+\lambda)e^{r(T-\tau)}Y(\tau).$$

On the set $\{Y(\tau) < 0\}$, we have $\tau = T$,

$$W_T^{X,Y} = X(T) + (1+\lambda)Y(T) = e^{r(T-\tau)}X(\tau) + (1+\lambda)e^{r(T-\tau)}Y(\tau).$$

On the other hand, noting that $Y(t) \leqslant 0$, $t \in [0, \tau)$, we have

$$\begin{aligned}
&e^{r(T-\tau)}X(\tau) + (1+\lambda)e^{r(T-\tau)}Y(\tau) \\
&= e^{rT}x + (1+\lambda)e^{rT}y - \int_0^\tau (\lambda+\mu)e^{r(T-t)}\,\mathrm{d}N^c(t) + \int_0^\tau (\alpha - r)e^{r(T-t)}Y(t)\,\mathrm{d}t \\
&\quad + \int_0^\tau e^{r(T-t)}(1-\mu)\sigma Y(t)\,\mathrm{d}B(t) - \sum_{0\leqslant t\leqslant\tau}(\lambda+\mu)e^{r(T-t)}(N(t) - N(t-)) \\
&\leqslant e^{rT}x + (1+\lambda)e^{rT}y + \int_0^\tau e^{r(T-t)}(1-\mu)\sigma Y(t)\,\mathrm{d}B(t).
\end{aligned}$$

It follows

$$\mathbf{E}[e^{r(T-\tau)}X(\tau) + (1+\lambda)e^{r(T-\tau)}Y(\tau)] \leqslant e^{rT}x + (1+\lambda)e^{rT}y.$$

Therefore,

$$\mathbf{E}[W_T^{X,Y}] \leqslant \mathbf{E}[e^{r(T-\tau)}X(\tau) + (1+\lambda)e^{r(T-\tau)}Y(\tau)] \leqslant e^{rT}x + (1+\lambda)e^{rT}y.$$

This indicates that $\mathbf{E}[W_T^{X,Y}] = e^{rT}x + (1+\lambda)e^{rT}y$ if and only if the investor puts all of his wealth in the bond at time 0. Thus $\hat{z} = e^{rT}x + (1+\lambda)e^{rT}y$ if $y < 0$. □

This result demonstrates the importance of the length of the investment planning horizon, T, by examining the situation when T is not long enough. In this "short horizon" case, if the investor starts with a short position in stock, then the only sensible strategy is the all-bond one, since any *other* strategy will just be worse off in *both* mean and variance. On the other hand, if one starts with a long stock position, then the highest expected terminal net wealth (without considering the variance) is achieved by the "stay-put" strategy, one that does not switch at all between bond and stock from the very beginning. Therefore, any efficient strategy is between the two extreme strategies, those of "all-bond" and "stay-put", according to an individual investor's risk preference.

More significantly, Theorem 3.3 specifies explicitly this critical length of horizon, $T^* = \frac{1}{\alpha - r}\ln\left(\frac{1+\lambda}{1-\mu}\right)$. It is intriguing that T^* depends only on the excess return, $\alpha - r$, and the transaction fees λ, μ, *not* on the individual target z or the stock volatility σ. Later we will show that, indeed, $\hat{z} = +\infty$ when $T > T^*$ in Corollary 6.3. Therefore T^* is such a critical value in time that divides between "global feasibility" and "limited feasibility" of the underlying MV portfolio selection problem. It signifies the opportunity that a longer time horizon would provide in achieving a higher potential gain. In this sense, the length of the planning horizon should be really included in the set of the investment opportunities, as opposed to the hitherto widely accepted notion that the investment opportunity set consists of only the probabilistic characteristics of the returns. Moreover, it follows from the expression of T^* that the less transaction cost and/or the higher excess return of the stock the shorter time it requires to attain the global feasibility. These, of course, all make perfect sense economically.

In the remaining part of this paper, we only consider the case when $\mathcal{D} \neq \emptyset$ and $z \in \mathcal{D}$.

4 Unconstrained Problem and Double-Obstacle Problem

4.1 Lagrangian Relaxation and HJB equation

By virtue of Lemma 3.2, Problem 2.2 is a convex constrained optimization problem. We shall utilize the well-known Lagrange multiplier method to remove the constraint.

Let us introduce the following unconstrained problem.

Problem 4.1 (Unconstrained Problem).

$$\begin{aligned} &\text{Minimize} && \mathbf{E}[W^2] - 2\ell(\mathbf{E}[W] - z) \\ &\text{subject to} && W \in \mathcal{W}_0^{x,y}, \end{aligned}$$

or equivalently,

Problem 4.2.

$$\begin{aligned} &\text{Minimize} && \mathbf{E}[(W-\ell)^2] \\ &\text{subject to} && W \in \mathcal{W}_0^{x,y}. \end{aligned}$$

Define the value function of Problem 2.2 as follows:

$$V_1(x,y;z) \stackrel{\text{def}}{=\!=} \inf_{\substack{W\in\mathcal{W}_0^{x,y}\\ \mathbf{E}[W]=z}} \mathbf{E}[W^2], \quad z\in\mathcal{D}.$$

The following result, showing the connection between Problem 2.2 and Problem 4.2, can be proved by a standard convex analysis argument.

Proposition 4.1. *Problem 2.2 and Problem 4.2 have the following relations.*

(1) If W_z^* solves Problem 2.2 with parameter $z\in\mathcal{D}$, then there exits $\ell\in\mathbb{R}$ such that W_z^* also solves Problem 4.2 with parameter ℓ.

(2) Conversely, if W_ℓ solves Problem 4.2 with parameter $\ell\in\mathbb{R}$, then it must also solve Problem 2.2 with parameter $z=\mathbf{E}[W_\ell]$.

It is easy to see that $\mathcal{W}_0^{x,y}-\ell=\mathcal{W}_0^{x-\ell e^{-rT},y}$. As a consequence, we consider the following problem instead of Problem 4.2:

Problem 4.3.

$$\begin{array}{ll}\text{Minimize} & \mathbf{E}[W^2]\\ \text{subject to} & W\in\mathcal{W}_0^{x-\ell e^{-rT},y}\end{array}$$

To solve the above problem, we use dynamic programming. In doing so we need to parameterize the initial time. Consider the dynamics (2.1)–(2.2) where the initial time 0 is revised to some $s\in[0,T)$, and define $\mathcal{W}_s^{x,y}$ as the counterpart of $\mathcal{W}_0^{x,y}$ where the initial time is s and initial bond–stock position is (x,y). We then define the value function of Problem 4.3 as

$$V(t,x,y)\stackrel{\text{def}}{=\!=}\inf_{W\in\mathcal{W}_t^{x,y}}\mathbf{E}[W^2],\quad (t,x,y)\in[0,T)\times\mathbb{R}^2. \tag{4.1}$$

The following proposition establishes a link between Problem 4.3 and Problem 2.2.

Proposition 4.2. *If $z\in\mathcal{D}$, then*

$$\sup_{\ell\in\mathbb{R}}(V(0,x-\ell e^{-rT},y)-(\ell-z)^2)=V_1(x,y;z)-z^2.$$

Proof. Note that

$$\begin{aligned}&\sup_{\ell\in\mathbb{R}}(V(0,x-\ell e^{-rT},y)-(\ell-z)^2)=\sup_{\ell\in\mathbb{R}}\inf_{W\in\mathcal{W}_0^{x-\ell e^{-rT},y}}\mathbf{E}[W^2-(\ell-z)^2]\\ =&\sup_{\ell\in\mathbb{R}}\inf_{W\in\mathcal{W}_0^{x,y}}\mathbf{E}[(W-\ell)^2-(\ell-z)^2]\leqslant\sup_{\ell\in\mathbb{R}}\inf_{\substack{W\in\mathcal{W}_0^{x,y}\\ \mathbf{E}[W]=z}}\mathbf{E}[(W-\ell)^2-(\ell-z)^2]\\ =&\sup_{\ell\in\mathbb{R}}\inf_{\substack{W\in\mathcal{W}_0^{x,y}\\ \mathbf{E}[W]=z}}(\mathbf{E}[W^2]-z^2)=\inf_{\substack{W\in\mathcal{W}_0^{x,y}\\ \mathbf{E}[W]=z}}(\mathbf{E}[W^2]-z^2)=V_1(x,y;z)-z^2.\end{aligned}$$

Therefore

$$\sup_{\ell\in\mathbb{R}}(V(0,x-\ell e^{-rT},y)-(\ell-z)^2)\leqslant V_1(x,y;z)-z^2.$$

Since V_1 is convex and z is an interior point of $\mathcal{D}$, by convex analysis, there exists $\ell^* \in \mathbb{R}$ such that

$$V_1(x,y;z) - 2\ell^* z \leqslant V_1(x,y;\tilde{z}) - 2\ell^* \tilde{z}, \quad \forall\, \tilde{z} \in \mathcal{D}.$$

For any $W \in \mathcal{W}_0^{x,y}$, by the definition of V_1, we have

$$\begin{aligned}\mathbf{E}[(W-\ell^*)^2 - (\ell^*-z)^2] = \mathbf{E}[W^2] - 2\ell^*(\mathbf{E}[W]-z) - z^2 &\geqslant V_1(x,y;\mathbf{E}[W]) - 2\ell^*(\mathbf{E}[W]-z) - z^2 \\ &\geqslant V_1(x,y;z) - z^2.\end{aligned}$$

It follows

$$\begin{aligned}&\sup_{\ell\in\mathbb{R}}(V(0,x-\ell e^{-rT},y) - (\ell-z)^2) = \sup_{\ell\in\mathbb{R}}\inf_{W\in\mathcal{W}_0^{x,y}} \mathbf{E}[(W-\ell)^2 - (\ell-z)^2] \\ \geqslant &\inf_{W\in\mathcal{W}_0^{x,y}} \mathbf{E}[(W-\ell^*)^2 - (\ell^*-z)^2] \geqslant V_1(x,y;z) - z^2,\end{aligned}$$

which yields the desired result. □

Therefore, we need only to study the value function $V(t,x,y)$, which we set out to do now.

Lemma 4.3. *The value function V defined in (4.1) has the following properties.*

(1) For any $t \in [0,T)$, $V(t,\cdot,\cdot)$ is convex and continuous in $\mathbb{R}^2$.

(2) For any $t \in [0,T)$, $V(t,x,y)$ is nonincreasing in x and y.

(3) For any $\rho > 0$, $t \in [0,T)$, we have $V(t,x+(1-\mu)\rho, y-\rho) \geqslant V(t,x,y)$, $V(t,x-(1+\lambda)\rho, y+\rho) \geqslant V(t,x,y)$.

(4) For any $\rho > 0$, $t \in [0,T)$, we have $V(t,\rho x,\rho y) = \rho^2 V(t,x,y)$.

(5) If $x + (1-\mu)y^+ - (1+\lambda)y^- \geqslant 0$, then $V(t,x,y) = 0$.

Proof. All the results can be easily proved in term of the definition of V and Lemma 3.2. □

Due to part (5) of the above lemma, we only need to consider Problem 4.3 in the *insolvency region*

$$\mathscr{S} \overset{\text{def}}{=\!=} \left\{(x,y) \in \mathbb{R}^2 \middle|\, x + (1-\mu)y^+ - (1+\lambda)y^- < 0\right\}.$$

It is well known that the value function V is a viscosity solution to the following Hamilton-Jacobi-Bellman (HJB) equation or variational inequality (VI) with terminal condition:

$$\begin{cases} \min\ \left\{\varphi_t + \mathcal{L}_0\varphi, (1-\mu)\varphi_x - \varphi_y, \varphi_y - (1+\lambda)\varphi_x\right\} = 0, & \forall\, (t,x,y) \in [0,T)\times\mathscr{S}, \\ \varphi(T,x,y) = (x + (1-\mu)y^+ - (1+\lambda)y^-)^2, & \forall\, (x,y) \in \mathscr{S}, \end{cases} \tag{4.2}$$

where

$$\mathcal{L}_0\varphi \overset{\text{def}}{=\!=} \frac{1}{2}\sigma^2 y^2 \varphi_{yy} + \alpha y \varphi_y + rx\varphi_x.$$

The idea of the subsequent analysis is to construct a particular solution to the HJB equation, and then employ the verification theorem to obtain an optimal strategy. The construction of the solution is built upon a series of

transformations on equation (4.2) until we reach an equation related to the so-called *double-obstacle* problem in physics which has been well studied in the partial differential equation literature.

We will show in Proposition 4.4 below that the constructed solution φ satisfies $\varphi_y - (1+\lambda)\,\varphi_x = 0$ when $y < 0$. Hence, we only need to focus on $y > 0$. A substantial technical difficulty arises with the HJB equation (4.2) in that the spatial variable (x, y) is two dimensional. However, the homogeneity of Lemma 4.3-(4) motivates us to make the transformation

$$\varphi(t,x,y) = y^2 \overline{V}(t, \frac{x}{y}), \text{ for } y > 0,$$

so as to reduce the dimension by one. Accordingly, (4.2) is turned to

$$\begin{cases} \min\left\{\overline{V}_t + \mathcal{L}_1 \overline{V}, (x+1-\mu)\overline{V}_x - 2\overline{V}, -(x+1+\lambda)\overline{V}_x + 2\overline{V}\right\} = 0, & \forall\ (t,x) \in [0,T) \times \mathscr{X}, \\ \overline{V}(T,x) = (x+1-\mu)^2, & \forall\, x \in \mathscr{X}, \end{cases}$$

where $\mathscr{X} \overset{\text{def}}{=\!=} (-\infty, -(1-\mu))$, and

$$\mathcal{L}_1 \overline{V} = \frac{1}{2}\sigma^2 x^2 \overline{V}_{xx} - \left(\alpha - r + \sigma^2\right) x \overline{V}_x + \left(2\alpha + \sigma^2\right) \overline{V}.$$

Further, let

$$w(t,x) \overset{\text{def}}{=\!=} \frac{1}{2} \ln \overline{V}(t,x), \quad (t,x) \in [0,T) \times \mathscr{X}.$$

It is not hard to show that $w\,(t,x)$ is governed by

$$\begin{cases} \min\ \left\{w_t + \mathcal{L}_2 w, \frac{1}{w_x} - (x+1-\mu), (x+1+\lambda) - \frac{1}{w_x}\right\} = 0, & \forall\ (t,x) \in [0,T) \times \mathscr{X}, \\ w(t,x) = \ln(-x-(1-\mu)), & \forall\, x \in \mathscr{X}, \end{cases} \tag{4.3}$$

where

$$\mathcal{L}_2 w \overset{\text{def}}{=\!=} \frac{1}{2}\sigma^2 x^2 (w_{xx} + 2w_x^2) - (\alpha - r + \sigma^2) x w_x + \alpha + \frac{1}{2}\sigma^2.$$

4.2 A Related Double-Obstacle Problem

Equation (4.3) is a variational inequality with gradient constraints, which is hard to study. As in Dai and Yi (2009), we will relate it to a double obstacle problem that is tractable. We refer interested readers to Friedman (1988) for obstacle problems.

Let

$$v(t,x) \overset{\text{def}}{=\!=} \frac{1}{w_x(t,x)}, \quad (t,x) \in [0,T) \times \mathscr{X}. \tag{4.4}$$

Notice that

$$\frac{\partial}{\partial x}\mathcal{L}_2 w = -\frac{1}{v^2}\left[\frac{1}{2}\sigma^2 x^2 v_{xx} - (\alpha - r)\, x v_x + \left(\alpha - r + \sigma^2\right) v + \sigma^2 \left(\frac{2x^2 v_x - x^2 v_x^2}{v} - 2x\right)\right].$$

This *inspires* us to consider the following double-obstacle problem:

$$\begin{cases} \max\left\{\min\left\{-v_t - \mathcal{L}v, v - (x+1-\mu)\right\}, v - (x+1+\lambda)\right\} = 0, & \forall\ (t,x) \in [0,T) \times \mathscr{X} \\ v(T,x) = x + 1 - \mu, & \forall\, x \in \mathscr{X}, \end{cases} \tag{4.5}$$

where

$$\mathcal{L}v \overset{\text{def}}{=\!=} \frac{1}{2}\sigma^2 x^2 v_{xx} - (\alpha - r)xv_x + (\alpha - r + \sigma^2)v + \sigma^2\left(\frac{2x^2 v_x - x^2 v_x^2}{v} - 2x\right). \tag{4.6}$$

It should be emphasized that at this stage we have yet to know if equation (4.5) is mathematically equivalent to equation (4.3) via the transformation (4.4). However, the following propositions show that (4.5) is solvable, and the solution to (4.3) can be constructed through the solutions of (4.5).

Proposition 4.4. *Equation (4.5) has a solution $v \in W_p^{1,2}([0,T)\times(-N,-(1-\mu)))$, for any $N > -(1-\mu)$, $p \in (1,\infty)$. Moreover,*

$$v_t \leqslant 0, \tag{4.7}$$

$$0 \leqslant v_x \leqslant 1, \tag{4.8}$$

and there exist two decreasing functions $x_s^*(\cdot) \in C^\infty[0,T)$ and $x_b^*(\cdot) \in C^\infty[0,T_0)$ such that

$$\{(t,x) \in [0,T)\times\mathscr{X} :\ v(t,x) = x + 1 - \mu\} = \{(t,x) \in [0,T)\times\mathscr{X} :\ x \geqslant x_s^*(t)\} \tag{4.9}$$

and

$$\{(t,x) \in [0,T)\times\mathscr{X} :\ v(t,x) = x + 1 + \lambda\} = \{(t,x) \in [0,T_0)\times\mathscr{X} :\ x \leqslant x_b^*(t)\} \tag{4.10}$$

where

$$T_0 = \max\left\{T - \frac{1}{\alpha - r}\ln\left(\frac{1+\lambda}{1-\mu}\right), 0\right\}. \tag{4.11}$$

Further, we have

$$\lim_{t\uparrow T} x_s^*(t) = (1-\mu)x_M, \quad \lim_{T\uparrow\infty} x_s^*(0) = x_{s,\infty}^*, \quad \lim_{t\uparrow T_0} x_b^*(t) = -\infty, \quad \lim_{T\uparrow\infty} x_b^*(0) = x_{b,\infty}^*, \tag{4.12}$$

where $x_M = -\frac{\alpha - r + \sigma^2}{\alpha - r}$, and $x_{s,\infty}^*$ and $x_{b,\infty}^*$ are defined in (A.1) and (A.2).

Proposition 4.5. *Define*

$$w(t,x) \overset{\text{def}}{=\!=} \mathfrak{A}(t) + \ln(-x_s^*(t) - (1-\mu)) + \int_{x_s^*(t)}^x \frac{1}{v(t,y)}\,\mathrm{d}y, \tag{4.13}$$

where

$$\mathfrak{A}(t) \overset{\text{def}}{=\!=} \int_t^T \frac{r x_s^{*2}(\tau) + (\alpha + r)(1-\mu)x_s^*(\tau) + (\alpha + \frac{1}{2}\sigma^2)(1-\mu)^2}{(x_s^*(\tau) + 1 - \mu)^2}\,\mathrm{d}\tau. \tag{4.14}$$

Then $w \in C^{1,2}([0,T)\times\mathscr{X})$ is a solution to equation (4.3). Moreover, for any $(t,x,y) \in [0,T)\times\mathscr{S}$, we define

$$\varphi(t,x,y) \overset{\text{def}}{=\!=} \begin{cases} y^2 e^{2w(t,\frac{x}{y})}, & \text{if } y > 0, \\ e^{2\mathfrak{B}(t)}(x + (1+\lambda)y)^2, & \text{if } y \leqslant 0 \end{cases}, \tag{4.15}$$

where $w(t,x)$ is given in (4.13) and

$$\mathfrak{B}(t) \overset{\text{def}}{=\!=} \int_t^T \frac{r x_b^{*2}(\tau) + (\alpha + r)(1+\lambda)x_b^*(\tau) + (\alpha + \frac{1}{2}\sigma^2)(1+\lambda)^2}{(x_b^*(\tau) + 1 + \lambda)^2}\,\mathrm{d}\tau.$$

Then $\varphi \in C^{1,2,2}([0,T)\times\mathscr{S}\setminus\{y=0\})$ is a solution to the HJB equation (4.2).

Due to their considerable technicality, the proofs of the preceding two propositions are placed in Appendix A. Most of the above results are similar to those obtained by Dai and Yi (2009) where they considered the expected utility portfolio selection with transaction costs. Nonetheless, there is one breakthrough made by the present paper: both $x_s^*(\cdot)$ and $x_b^*(\cdot)$ are proven to be C^∞, whereas Dai and Yi (2009) only obtained the smoothness of $x_s^*(\cdot)$. In fact, Dai and Yi (2009) essentially considered a double obstacle problem for w_x. In contrast, the present paper takes the double obstacle problem for $1/w_x$ into consideration. This seemly innocent modification in fact simplifies the proof greatly. More importantly, it allows us to provide a unified framework to obtain the smoothness of $x_s^*(\cdot)$ and $x_b^*(\cdot)$. Later we will see that the smoothness of $x_s^*(\cdot)$ and $x_b^*(\cdot)$ plays a critical role in the proof of the existence of an optimal strategy.

Note that (4.8) is important in the study of the double obstacle problem (4.5). In essence, the result is based on parts (1) and (2) of Lemma 4.3.

In the subsequent section, we plan to show that $\varphi(t,x,y)$ is nothing but the value function through the verification theorem and a Skorokhod problem. At the same time, an optimal strategy will be constructed.

5 Skorokhod Problem and Optimal Strategy

Due to (4.9)–(4.10) and Proposition 4.5, we define

$$\begin{aligned}
\mathcal{SR} &= \{(t,x,y) \in [0,T) \times \mathscr{S} \mid y > 0,\ \ x \geqslant x_s^*(t)y\}\,, \\
\mathcal{BR} &= \{(t,x,y) \in [0,T) \times \mathscr{S} \mid y > 0,\ \ x \leqslant x_b^*(t)y,\ \text{or } y \leqslant 0\}\,, \\
\mathcal{NT} &= \{(t,x,y) \in [0,T) \times \mathscr{S} \mid y > 0,\ x_b^*(t)y < x < x_s^*(t)y\}\,,
\end{aligned} \tag{5.1}$$

which stand for the sell region, buy region and no trade region, respectively. Here we set $x_b^*(t) = -\infty$ when $t \in [T_0, T]$ in view of the fact that $\lim\limits_{t \uparrow T_0} x_b^*(t) = -\infty$; hence $\mathcal{BR} = \emptyset$ when $t \in [T_0, T]$. Notice that these regions do not depend on the target z.

5.1 Skorokhod Problem and Verification Theorem

In order to find the optimal solution to the MV problem, we need to study the so-called Skorokhod problem.

Problem 5.1 (Skorokhod Problem). Given $(0, X(0), Y(0)) \in \overline{\mathcal{NT}}$, find an admissible strategy (M, N) such that the corresponding bond–stock value process (X, Y) is continuous in $[0, T]$, and $(t, X(t), Y(t)) \in \overline{\mathcal{NT}}$, for any $t \in [0, T]$.

In other words, a solution to the Skorokhod problem is an investment strategy with which the trading only takes place on the boundary of the no trade region. It turns out the solution can be constructed via solving the following more specific problem.

Problem 5.2. Given $(0, X(0), Y(0)) \in \overline{\mathcal{NT}}$, find a process of bounded total variation k and a continuous process

(X, Y) such that for any $t \in [0, T]$,

$$\begin{aligned}
(t, X(t), Y(t)) &\in \overline{\mathcal{NT}}, \\
\mathrm{d}X(t) &= rX(t)\,\mathrm{d}t + \gamma_1(X(t), Y(t))\,\mathrm{d}|k|(t), \\
\mathrm{d}Y(t) &= \alpha Y(t)\,\mathrm{d}t + \sigma Y(t)\,\mathrm{d}B(t) + \gamma_2(X(t), Y(t))\,\mathrm{d}|k|(t), \\
|k|(t) &= \int_0^t \mathbf{1}_{\{(s,X(s),Y(s))\in\partial\mathcal{NT}\}}\,\mathrm{d}|k|(s),
\end{aligned}$$

where $|k|(t)$ stands for the total variation of k on $[0, t]$,

$$(\gamma_1(x, y), \gamma_2(x, y)) \overset{\text{def}}{=\!=} \begin{cases} \frac{1}{\sqrt{(1+\lambda)^2+1}}(-(1+\lambda), 1), & \text{if } (t, x, y) \in \partial_1\mathcal{NT}, \\ \frac{1}{\sqrt{(1-\mu)^2+1}}(1-\mu, -1), & \text{if } (t, x, y) \in \partial_2\mathcal{NT}, \end{cases}$$

and

$$\begin{aligned}
\partial_1\mathcal{NT} &\overset{\text{def}}{=\!=} \{(t, x, y) \in [0, T) \times \mathscr{S} \mid y > 0,\ \ x = x_b^*(t)y\}. \\
\partial_2\mathcal{NT} &\overset{\text{def}}{=\!=} \{(t, x, y) \in [0, T) \times \mathscr{S} \mid y > 0,\ \ x = x_s^*(t)y\}.
\end{aligned}$$

Letting a triplet (k, X, Y) solves Problem 5.2, define

$$M(t) \overset{\text{def}}{=\!=} \frac{1}{\sqrt{(1+\lambda)^2+1}} \int_0^t \mathbf{1}_{\{(s,X(s),Y(s))\in\partial_1\mathcal{NT}\}}\,\mathrm{d}|k|(s), \tag{5.2}$$

$$N(t) \overset{\text{def}}{=\!=} \frac{1}{\sqrt{(1-\mu)^2+1}} \int_0^t \mathbf{1}_{\{(s,Y(s),Y(s))\in\partial_2\mathcal{NT}\}}\,\mathrm{d}|k|(s). \tag{5.3}$$

Then (M, N) is a solution to the Skorokhod problem. Moreover, since (M, N, X, Y) satisfies equations (2.1) and (2.2), we can prove that the corresponding terminal net wealth $W^{X,Y}(T)$ is the optimal solution to Problem 4.3. To prove that we need the verification theorem.

Note that one can extend naturally the definition of the Skorokhod problem to the time horizon $[s, T]$, for any $s \in [0, T)$.

Theorem 5.1 (Verification Theorem). *Let φ be defined in (4.15) and V be the value function defined in (4.1). If the Skorokhod problem admits a solution in $[s, T]$, where $s \in [0, T)$, then*

$$V(t, x, y) = \varphi(t, x, y),\ \forall\ (t, x, y) \in [s, T] \times \mathbb{R}^2.$$

Proof. The proof is rather standard in the singular control literature. We only give a sketch and refer interested readers to Karatzas and Shreve (1998). Similar to Karatzas and Shreve (1998), we can show that the function $\varphi \leqslant V$ in $\overline{\mathcal{NT}}$. Moreover, if τ is a stopping time valued in $[s, T]$, where $s \in [0, T)$, and (X, Y) is a solution to Problem 5.1 in $[s, \tau]$, then

$$\varphi(s, X(s), Y(s)) = \mathbf{E}[\varphi(\tau, X(\tau), Y(\tau))].$$

Particularly, if $\tau = T$, then $\varphi = V$ in $\overline{\mathcal{NT}} \cap [s, T] \times \mathbb{R}^2$, and

$$V(s, X(s), Y(s)) = \mathbf{E}\left[\left(W^{X,Y}(T)\right)^2\right]. \tag{5.4}$$

The verification theorem in the $\mathcal{NT}$ follows. By Lemma 4.3, we know that $V(t,\cdot,\cdot)$ is convex in $\mathbb{R}^2$. So we can define its *subdifferential* as

$$\partial V(t,x,y) \stackrel{\text{def}}{=} \left\{(\delta_x,\delta_y)\middle| V(t,\bar{x},\bar{y}) \geqslant V(t,x,y) + \delta_x \cdot (\bar{x}-x) \ + \delta_y \cdot (\bar{y}-y), \forall (\bar{x},\bar{y}) \in \mathbb{R}^2\right\}.$$

Then, we are able to utilize the convex analysis as in Shreve and Soner (1994) to obtain the verification theorem in $\mathcal{BR}$ and $\mathcal{SR}$. □

5.2 Solution to Skorokhod Problem

Note that, in the Skorokhod problem, Problem 5.1, the reflection boundary depends on time t. This is very different from the standard Skorokhod problem in the literature; see, e.g., Lions and Sznitman (1984). To remove the dependence of the reflection boundary on time, we introduce a new state variable $Z(t)$ and instead consider an equivalent problem.

Problem 5.3. Given $(0, X(0), Y(0)) \in \overline{\mathcal{NT}}$, find a process of bounded total variation k and a continuous process (Z, X, Y) such that, for any $t \in [0,T]$,

$$\begin{aligned}
(Z(t), X(t), Y(t)) &\in \overline{\mathcal{NT}},\\
\mathrm{d}X(t) &= rX(t)\,\mathrm{d}t + \gamma_1(X(t),Y(t))\,\mathrm{d}|k|(t),\\
\mathrm{d}Y(t) &= \alpha Y(t)\,\mathrm{d}t + \sigma Y(t)\,\mathrm{d}B(t) + \gamma_2(X(t),Y(t))\,\mathrm{d}|k|(t),\\
\mathrm{d}Z(t) &= \mathrm{d}t + \gamma_3(Z(t),X(t),Y(t))\,\mathrm{d}|k|(t),\\
|k|(t) &= \int_0^t \mathbf{1}_{\{(Z(s),X(s),Y(s))\in\partial\mathcal{NT}\}}\,\mathrm{d}|k|(s),
\end{aligned}$$

where $Z(0) = 0$, $\gamma_3 \equiv 0$.

Clearly $Z(t) \equiv t$ because $\gamma_3(z,x,y) \equiv 0$. Therefore if (k,Z,X,Y) solves Problem 5.3, then (k,X,Y) solves Problem 5.2. It is worthwhile pointing out that the reflection boundary of Problem 5.3 becomes time independent.

Now let us consider Problem 5.3.

Theorem 5.2. *There exists a unique solution to Problem 5.3 in* $[0,T]$.

Proof. Both Lions and Sznitman (1984) and Dupus and Ishii (1993) have established existence and uniqueness for the Skorokhod problem on a domain with sufficiently smooth boundary. Since the C^∞ smoothness of $x_s^*(\cdot)$ and $x_b^*(\cdot)$ is in place, the proof is similar to that of Lemma 9.3 of Shreve and Soner (1994). It is worthwhile pointing out that Shreve and Soner (1994) did not concern the smoothness of x_s^* and x_b^* because they took into consideration a stationary problem which leads to time-independent policies (free boundaries). □

Thanks to Theorem 5.1 and Theorem 5.2, we have

Corollary 5.3. $V(t,x,y) = \varphi(t,x,y),\ \forall\ (t,x,y) \in [0,T]\times\mathbb{R}^2$.

6 Main Results

Theorem 6.1. *For any initial position $(x_0, y_0) \in \mathscr{S}$, define*

$$(X(0), Y(0)) \overset{\text{def}}{=\!=} \begin{cases} \left(\frac{(x_0+(1-\mu)y_0)x_s^*(0)}{x_s^*(0)+1-\mu}, \frac{x_0+(1-\mu)y_0}{x_s^*(0)+1-\mu} \right), & \text{if } (0, x_0, y_0) \in \mathcal{SR}, \\ (x_0, y_0), & \text{if } (0, x_0, y_0) \in \overline{\mathcal{NT}}, \\ \left(\frac{(x_0+(1+\lambda)y_0)x_b^*(0)}{x_b^*(0)+1+\lambda}, \frac{x_0+(1+\lambda)y_0}{x_b^*(0)+1+\lambda} \right), & \text{if } (0, x_0, y_0) \in \mathcal{BR}. \end{cases}$$

Let (k, Z, X, Y) be the solution to Problem 5.3 as stipulated in Theorem 5.2. Then (X, Y) is the unique solution to the Skorokhod problem, Problem 5.1, and

$$V(0, x_0, y_0) = \mathbf{E}\left[\left(W^{X,Y}(T) \right)^2 \right].$$

Moreover, the optimal strategy (M, N) is defined by (5.2) and (5.3).

Proof. Noting that $(0, X(0), Y(0)) \in \overline{\mathcal{NT}}$, by (5.4),

$$V(0, X(0), Y(0)) = \mathbf{E}\left[\left(W^{X,Y}(T) \right)^2 \right].$$

Since $V = \varphi$, it is not hard to check that

$$V(0, x_0, y_0) = V(0, X(0), Y(0)).$$

The proof is complete. □

As a final task before reaching the main result, we prove the existence of the Lagrange multiplier.

Proposition 6.2. *For any $(x, y, z) \in \mathbb{R}^2 \times \mathcal{D}$, there exists a unique $\ell^* \in \mathbb{R}$ such that*

$$V(0, x - \ell^* e^{-rT}, y) - (\ell^* - z)^2 = \sup_{\ell \in \mathbb{R}} (V(0, x - \ell e^{-rT}, y) - (\ell - z)^2).$$

Moreover, ℓ^* is determined by equation

$$e^{-rT} V_x(0, x - \ell^* e^{-rT}, y) + 2\ell^* = 2z. \tag{6.1}$$

The proof is placed in Appendix B.

Corollary 6.3. *If $T > \frac{1}{\alpha - r} \ln\left(\frac{1+\lambda}{1-\mu}\right)$, then*

$$\hat{z} = +\infty, \quad \mathcal{D} = (e^{rT}x + (1-\mu)e^{rT}y^+ - (1+\lambda)e^{rT}y^-, +\infty).$$

Proof. Suppose $\hat{z} < +\infty$. Then by the definition of $\hat{z}$, for any $W \in \mathcal{W}_0^{x,y}$, we have $\mathbf{E}[W] \leqslant \hat{z}$. So for any $z > \hat{z}$, $\ell \geqslant 0$,

$$\begin{aligned} V(0, x - \ell e^{-rT}, y) - (\ell - z)^2 &= \inf_{W \in \mathcal{W}_0^{x - \ell e^{-rT}, y}} \mathbf{E}[W^2 - (\ell - z)^2] = \inf_{W \in \mathcal{W}_0^{x,y}} \mathbf{E}[(W - \ell)^2 - (\ell - z)^2] \\ &= \inf_{W \in \mathcal{W}_0^{x,y}} (\mathbf{E}[W^2] + 2\ell(z - \mathbf{E}[W]) - z^2) \\ &\geqslant 2\ell(z - \hat{z}) - z^2. \end{aligned}$$

Consequently,

$$\sup_{\ell\in\mathbb{R}}(V(0,x-\ell e^{-rT},y)-(\ell-z)^2)=+\infty.$$

However, the proof of Proposition 6.2 (Appendix B) shows that the above supremum is finite under the condition $T>\frac{1}{\alpha-r}\ln\left(\frac{1+\lambda}{1-\mu}\right)$; see (B.1). The proof is complete. □

Now we arrive at the complete solution to the MV problem, Problem 2.1.

Theorem 6.4. *Problem 2.1 admits an optimal solution if and only if $z\in\widetilde{\mathcal{D}}$, where*

$$\widetilde{\mathcal{D}}\overset{\text{def}}{=\!=}\begin{cases}(e^{rT}x+(1-\mu)e^{rT}y^+-(1+\lambda)y^-,+\infty), & \text{if } T>\frac{1}{\alpha-r}\ln\left(\frac{1+\lambda}{1-\mu}\right),\\ (e^{rT}x+(1-\mu)e^{rT}y,e^{rT}x+(1-\mu)e^{\alpha T}y], & \text{if } T\leqslant\frac{1}{\alpha-r}\ln\left(\frac{1+\lambda}{1-\mu}\right),\ y>0,\\ \emptyset, & \text{if } T\leqslant\frac{1}{\alpha-r}\ln\left(\frac{1+\lambda}{1-\mu}\right),\ y\leqslant 0.\end{cases}$$

Moreover, if z is on the boundary of $\widetilde{\mathcal{D}}$, i.e., $T\leqslant\frac{1}{\alpha-r}\ln\left(\frac{1+\lambda}{1-\mu}\right)$, $y>0$, and $z=e^{rT}x+(1-\mu)e^{\alpha T}y$, then the optimal strategy is $(M,N)\equiv(0,0)$; otherwise, the value function and the optimal solution are given by Theorem 6.1, in which the initial position $(x_0,y_0)=(x-\ell^*e^{-rT},y)$ where ℓ^* is determined by equation (6.1).

Proof. If $z\notin\widetilde{\mathcal{D}}$, then there is no feasible solution by Theorem 3.3; so Problem 2.1 admits no optimal solution. If $z=e^{rT}x+(1-\mu)e^{\alpha T}y$ while $T\leqslant\frac{1}{\alpha-r}\ln\left(\frac{1+\lambda}{1-\mu}\right)$ and $y>0$, then Theorem 3.3 again shows that the optimal strategy is $(M,N)\equiv(0,0)$.

In all the other cases, it follows from Proposition 6.2 that there exists a unique Lagrange multiplier ℓ^* such that

$$V(0,x-\ell^*e^{-rT},y)-(\ell^*-z)^2=\sup_{\ell\in\mathbb{R}}(V(0,x-\ell e^{-rT},y)-(\ell-z)^2).$$

Appealing to Proposition 4.2, we have

$$V(0,x-\ell^*e^{-rT},y)-(\ell^*-z)^2=V_1(x,y;z)-z^2.$$

Theorem 6.1 then dictates that there exists an admissible strategy $(M^*,N^*)\in\mathcal{A}$ such that

$$V(0,x-\ell^*e^{-rT},y)=\mathbf{E}\left[\left(W^{X^{x-\ell^*e^{-rT},M^*,N^*},Y^{y,M^*,N^*}}(T)\right)^2\right].$$

Noting that for any $(M,N)\in\mathcal{A}$, we have

$$\begin{aligned}X^{x-\ell^*e^{-rT},M,N}(T)&=X^{x,M,N}(T)-\ell^*,\\ Y^{y,M,N}(T)&=Y^{y,M,N}(T),\\ W^{X^{x-\ell^*e^{-rT},M,N},Y^{y,M,N}}(T)&=W^{X^{x,M,N},Y^{y,M,N}}(T)-\ell^*;\end{aligned}$$

so

$$V(0,x-\ell^*e^{-rT},y)=\mathbf{E}\left[\left(W^{X^{x,M^*,N^*},Y^{y,M^*,N^*}}(T)-\ell^*\right)^2\right].$$

By the definition of V, for any $(M, N) \in \mathcal{A}$,

$$V(0, x - \ell^* e^{-rT}, y) \leqslant \mathbf{E}\left[\left(W^{X^{x-\ell^* e^{-rT}, M, N}, Y^{y, M, N}}(T)\right)^2\right].$$
$$= \mathbf{E}\left[\left(W^{X^{x, M, N}, Y^{y, M, N}}(T) - \ell^*\right)^2\right].$$

Therefore W^* is optimal to Problem 4.2 with parameter ℓ^*, where W^* is defined by

$$W^* = W^{X^{x, M^*, N^*}, Y^{y, M^*, N^*}}(T).$$

Owing to Proposition 4.1, W^* is optimal to Problem 2.2 with parameter $\mathbf{E}[W^*]$. By the uniqueness of ℓ^*, we have $\mathbf{E}[W^*] = z$. Thus W^* is the optimal solution to both Problems 2.2 and 2.1, and (M^*, N^*) is the optimal strategy. □

The preceding theorem fully describes the behavior of an optimal MV investor under transaction costs. If the planning horizon is not long enough (the precise critical length depends only on the stock excess return and the transaction fees), then what could be achieved at the terminal time (in terms of the expected wealth) is rather limited. Otherwise, any terminal target is achievable by an investment strategy, while an optimal (efficient) strategy is to minimize the corresponding risk (represented by the variance). The optimal strategy is characterized by three regions (those of sell, buy, and no trade) defined by (5.1). The implementation of the strategy is very simple: a transaction takes place only when the "adjusted" bond–stock process, $(X - \ell^* e^{-rT}, Y)$, reaches the boundary of the no-trade zone so as for the process to stay within the zone (if initially the process is outside of the no-trade zone then a transaction is carried out as in Theorem 6.1 to move it instantaneously into the no-trade zone).

The optimal strategy presented here is markedly different from its no-transaction counterpart [see, e.g., Zhou and Li (2000)]. With transaction costs, an investor tries *not* to trade unless absolutely necessary, so as to keep the "adjusted" bond–stock ratio, $\frac{x-\ell^* e^{-rT}}{y}$, between the two barriers, $x_b^*(t)$ and $x_s^*(t)$, at any given time t. When there is no transactions cost, however, the two barriers coincide; so the optimal strategy is to keep the above ratio exactly *at* the barrier[3]. This, in turn, requires the optimal strategy to trade *all the time*. Clearly, the strategy presented here is more consistent with the actual investors' behaviors.

Let us examine more closely the trade zone consisting of the sell and buy regions, defined in (5.1). By and large, when the adjusted bond–stock ratio, $\frac{x-\ell^* e^{-rT}}{y}$, starts to be greater than a critical barrier (namely $x_s^*(t)$, which is time-varying), then one needs to reduce the stock holdings. When the ratio starts to be smaller than *another* barrier ($x_b^*(t)$, again time-varying), then one must accumulate the stock. It is interesting to see that $y \leqslant 0$ always triggers buying; in other words shorting the stock is never favored, and any short position must be covered immediately. The essential reason behind this is the standing assumption that $\alpha > r$; so there is no good reason to short the stock.

Another not-so-obvious yet extremely intriguing behavior of the optimal strategy is that when the time to maturity is short enough (precisely, when the remaining time is less than $T^* = \frac{1}{\alpha - r} \ln\left(\frac{1+\lambda}{1-\mu}\right)$), then one should not buy stock any longer (unless to cover a possible short position). This is seen from the fact that $\lim_{t \uparrow T_0} x_b^*(t) = -\infty$ stated in Proposition 4.4 along with the definition of the buy region. Moreover, since both the two barriers are

[3] In an EUM model–the Merton problem for example–the optimal solution is to keep the bond–stock ratio exactly at a certain value. In the MV model the ratio must be "adjusted" in order to account for the constraint of meeting the terminal target.

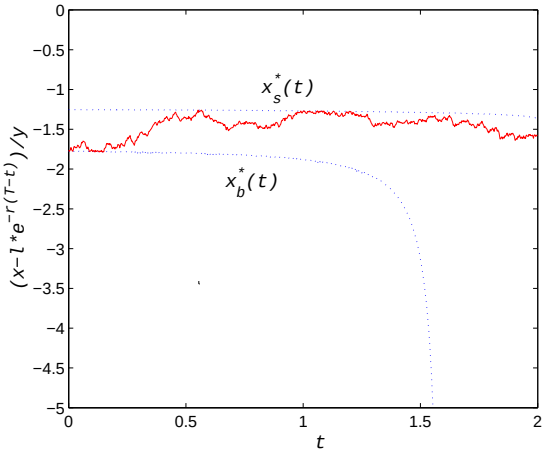

Figure 1: A Sample Path Corresponding to Optimal Strategy

decreasing in time, the buy region gets smaller and the sell region gets bigger as time passes along. This suggests that the investor would be less likely to buy the stock and more likely to sell the stock when the maturity date is getting closer. These phenomena, again, are in line with what prevail in practice.

We end this section by a numerical example. Consider a market with the following parameters:

$$(\alpha,\ r,\ \sigma,\ \lambda,\ \mu,\ T) = (0.15,\ 0.05,\ 0.2,\ 0.02,\ 0.02,\ 2).$$

In this case, $T > \frac{1}{\alpha - r} \ln\left(\frac{1+\lambda}{1-\mu}\right)$. In terms of a penalty method developed by Dai and Zhong (2009), we numerically solve equation (4.5) and then construct the two free boundaries by Proposition 4.4. Consider an investor having the initial position $(x,\ y) = (-1,\ 1)$ with an expected return $z = 1.1$ at time T. Based on (6.1) we can calculate that $\ell^* = 4.5069$; so the adjusted initial position is $(x - \ell^* e^{-rT}, y) = (-5.078, 1)$. The optimal strategy is the following. At time 0, applying Theorem 6.1, the investor carries out a transaction so as to move his adjusted position to the boundary of $\mathcal{NT}$. This is realized by buying 4.3395 units (in terms of the dollar amount) of the stock, with the new adjusted position to be $(-9.5043, 5.3395)$. After the initial time, the investor trades only on the boundaries of the $\mathcal{NT}$ region just to keep his adjusted position within the $\mathcal{NT}$ region.

Next consider another investor with an initial position $(x,\ y) = (1, 0)$ and expected return $z = 1.2$. In this case $\ell^* = 2.3690$ and $(x - \ell^* e^{-rT}, y) = (-1.1436, 0)$. So initially the investor buys 1.5047 worth of the stock moving the adjusted position to (-2.6784,1.5047) which is on the boundary of $\mathcal{NT}$. Afterwards, the trading strategy is simply to keep the adjusted position within $\mathcal{NT}$. The two boundaries are depicted in the Figure 1, where the horizontal axis is time t and the vertical one is the ratio between $x - \ell^* e^{-r(T-t)}$ and y. A sample path corresponding to the optimal strategy is also illustrated.

Finally we compare the MV efficiency with and without transaction costs, by plotting the respective efficient frontiers. We use the same model parameters as above and take the initial position $(x,\ y) = (1, 0)$. Figure 2 depicts the efficient frontiers with different transaction costs.[4] It is worth pointing out that the frontiers in the presence of transaction costs are still straight lines due to the availability of a risk-free asset. The figure shows clearly that the

[4]The no-transaction cost frontier follows from the analytical solutions in Zhou and Li (2000).

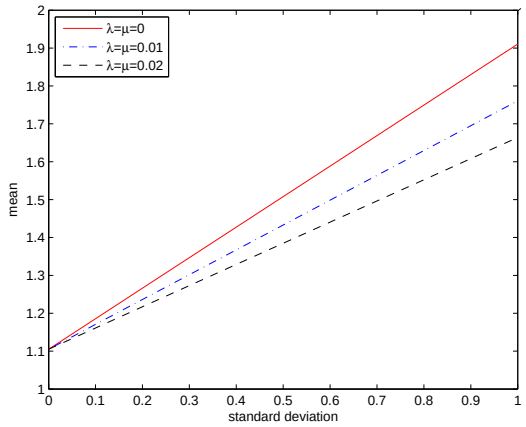

Figure 2: Efficient Frontiers with Different Transaction Costs

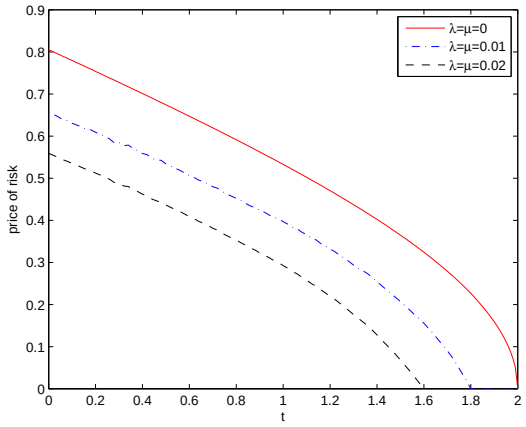

Figure 3: Prices of Risk in Time

MV efficiency declines as the transaction costs increase. Indeed, one could easily derive numerically the rate of the efficiency decline with respect to the transaction costs.

Now that the frontiers are straight lines, we plot in Figure 3 the slopes of the lines (known as the prices of risk) against different times (the expiration date T=2 is fixed). When transaction costs are incurred, we have shown that there is a threshold time value after which one never buys stock and hence the corresponding price of risk is 0. The threshold values $T - \frac{1}{\alpha - r}\log((1+\lambda)/(1-\mu)) = 1.6$ or 1.8 for $\lambda = \mu = 0.02$ or 0.01 respectively. These are verified by Figure 3.

7 Concluding Remarks

This paper investigates a continuous-time Markowitz's mean–variance portfolio selection model with proportional transaction costs. In the terminology of stochastic control theory, this is a singular control problem. We use the Lagrangian multiplier and partial differential equation to approach the problem. The problem has been completely

solved in the following sense. First, the feasibility of the model has been fully characterized by certain relationship among the parameters. Second, the value function is given via a PDE, which is analytically proven to be uniquely solvable and numerically tractable, whereas the Lagrange multiplier is determined by an algebraic equation. Third, the optimal strategy is expressed in terms of the free boundaries of the PDE. Economically, the results in the paper have revealed three critical differences arising from the presence of transaction costs. First, the expected return on the portfolio may not be achievable if the time to maturity is not long enough, while without transaction costs, any expected return can be achieved in an arbitrarily short time. Second, instead of trading all the time so as to keep a constant adjusted ratio between the stock and bond, there exist time-dependent upper and lower boundaries so that transaction is only carried out when the ratio is on the boundaries. Third, there is a critical time which depends only on the stock excess return and the transaction fees, such that beyond that time it is optimal not to buy stock at all. Finally, although shorting is allowed in our model, it is never favored by an optimal strategy. Our results are closer to real investment practice where people tend not to invest more in risky assets towards the end of the investment horizon.

Methodologically this paper employs the PDE approach of Dai and Yi (2009) developed for EUM (CRRA utility). In both MV (this paper) and CRRA [Liu and Loewenstein (2002) and Dai and Yi (2009)] cases it is shown that if the investor is holding the stock then he should stop buying more shares if the time left is too short. This is an interesting and important feature of the finite horizon problem with transaction costs. The intuition behind this is that the investor should not purchase any additional shares if the remaining investment period is not long enough to offset at least the transaction costs. However, there are significant differences between the MV and EUM models.[5] First of all, the issue of feasibility is unique to the MV model, which itself is interesting both mathematically and economically. Second, the present paper has shown that whenever the investor is shorting the stock the MV problem (with transaction costs) has no feasible solution if the remaining time is short enough, or otherwise one should immediately buy shares. There are no corresponding results in the EUM setting.[6] Last but not least, the smoothness of the switching boundaries is proved for the first time in this paper, which is instrumental in deriving rigorously the optimal trading strategies.

References

[1] BAJEUX-BESNAINOU, I., AND R. PORTAIT, Dynamic Asset Allocation in a Mean-Variance Framework, *Management Science*, Vol. 44, No. 11, S79-S95, 1998

[2] BASAK, S., AND G. CHABAKAURI, Dynamic Mean–Variance Asset Allocation, Working paper, London Business School, 2008

[3] BIELECKI, T. R., H. JIN, S. R. PLISKA, AND X. Y. ZHOU, Continuous-Time Mean-Variance Portfolio Selection with Bankruptcy Prohibition, *Mathematical Finance*, Vol. 15, No. 2, 213-244, 2005

[5] Strictly speaking, it is not entirely "fair" to compare the two models, because the MV model studied here allows bankruptcy whereas the EUM model with the usual Inada condition inherently leads to a bankruptcy-prohibited solution. It would be more meaningful to develop an MV model with no-bankruptcy and transaction costs and then to compare the solution with its EUM counterpart.

[6] In Liu and Loewenstein (2002) and Dai and Yi (2009) only the case when $y > 0$ is discussed, although we suspect that the methodology in our paper could be adapted to deal with the $y < 0$ case in the EUM setting.

[4] BJÖRK, T., AND MURGOCI, A., A General Theory of Markovian Time Inconsistent Stochastic Control Problems, Working paper, Stockholm School of Economics, 2008

[5] BLACK, F., AND M. SCHOLES, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, Vol. 81, 635-654, 1973

[6] CHIU, C. H., AND X. ZHOU, The Premium of Dynamic Trading, *Quantitative Finance*, forthcoming.

[7] CVITANIC, J., AND I. KARATZAS, Hedging and Portfolio Optimization under Transaction Costs: A Martingale Approach, *Mathematical Finance*, Vol. 6, No. 2, 133-165, 1996

[8] DAI, M., L. S. JIANG, P. F. LI, AND F. H. YI, Finite-Horizon Optimal Investment and Consumption with Transaction Costs, *SIAM Journal on Control and Optimization*, Vol. 48, No. 2, 1134-1154, 2009

[9] DAI, M., AND F. H. YI, Finite-Horizon Optimal Investment with Transaction Costs: A Parabolic Double Obstacle Problem, *Journal of Differential Equations*, Vol. 246, 1445-1469, 2009

[10] DAI, M., AND Y.F. ZHONG, Penalty Methods for Continuous-time Portfolio Selection with Proportional Transaction Costs, forthcoming in *Journal of Computational Finance*, 2009

[11] DAVIS, M. H. A., AND A. R. NORMAN, Portfolio Selection with Transaction Costs, *Mathematics of Operations Research*, Vol. 15, No. 4, 676-713, 1990

[12] DAVIS, M. H. A., V. G. PANAS, AND T. ZARIPHOPOULOU, European Option Pricing with Transaction Costs, *SIAM Journal of Control and Optimization*, Vol. 31, No. 2, 470-493, 1993

[13] DUFFIE, D., AND H. RICHARDSON, Mean-Variance Hedging in Continuous Time, *Annals of Applied Probability*, Vol. 1, 1-15, 1991

[14] DUPUS, P., AND H. ISHII, Stochastic Differential Equations with Oblique Reflections on Nonsmooth Domains, *Annals of Probability*, Vol. 21, 554-580, 1993

[15] EKELAND, I., AND A. LAZRAK, Being Serious about Non-Commitment: Subgame Perfect Equilibrium in Continuous Time, Working paper, University of British Columbia, 2007

[16] FLEMING, W. H., AND H. M. SONER, *Controlled Markov Processes and Viscosity Solutions*, New York, Springer, 2006

[17] FRIEDMAN, A., Parabolic Variational Inequalities in One Space Dimension and Smoothness of the Free Boundary, *Journal of Functional Analysis*, Vol. 18, No. 2, 115-212, 1975

[18] FRIEDMAN, A., *Variational Principles and Free-Boundary Problems*, Pobert. Erieger Publishing Company, INC., Florida, 1988

[19] GENNOTTE, G., AND A. JUNG, Investment Strategies under Transaction Costs: The Finite Horizon Case, *Management Science*, Vol. 40, 385-404, 1994

[20] JIN, H., AND X. Y. ZHOU, Behavioral Portfolio Selection in Continuous Time, *Mathematical Finance*, Vol. 18, 385-426, 2008

[21] KARATZAS, I., AND S. E. SHREVE, *Brownian Motion and Stochastic Calculus*, New York, Hong Kong, Springer-Verlag, 1991

[22] KARATZAS, I., AND S. E. SHREVE, *Methods of Mathematical Finance*, New York, Springer, 1998

[23] KLEBANER, F. C., *Introduction to Stochastic Calculus with Applications*, World Scientific Publishing Co. Pte. Ltd., Beijing, 2004

[24] KOHLMANN, M., AND X. Y. ZHOU, Relationship between Backward Stochastic Differential Equations and Stochastic Controls: A Linear-Quadratic Approach, *SIAM Journal on Control and Optimization*, Vol. 38, 1392-1407, 2000

[25] KORN, R., Optimal Portfolio, *World Scientific Publishing*, Co. Pte. Ltd, Singapore, 1997

[26] KYDLAND, F. E., AND E. C. PRESCOTT, Rules Rather Than Discretion: The Inconsistency of Optimal Plans, *The Journal of Political Economy*, Vol. 85, 473–492, 1977

[27] LI, D., AND W. L. NG, Optimal Dynamic Portfolio Selection: Multi-Period Mean-Variance Formulation, *Mathematical Finance*, Vol. 10, 387-406, 2000

[28] LI, X, AND X. Y. ZHOU, Continuous-Time Mean-Variance Efficiency: The 80% rule, *Annals of Applied Probability*, Vol. 16, No. 4, 1751–1763, 2006

[29] LI, X, X. Y. ZHOU, AND A. E. B. LIM, Dynamic Mean-Variance Portfolio Selection with No-Shorting Constraints, *SIAM Journal on Control and Optimization*, Vol. 40, 1540-1555, 2001

[30] LIM, A. E. B., Quadratic Hedging and Mean-Variance Portfolio Selection with Random Parameters in an Incomplete Market. *Mathematics of Operations Research* Vol. 29, 132-161, 2004

[31] LIM, A. E. B., AND X. Y. ZHOU, Mean-Variance Portfolio Selection with Random Parameters in a Complete Market, *Mathematics of Operations Research*, Vol. 27, No. 1, 101-120, 2002

[32] LIONS, P. L., AND A. S. SZNITMAN, Stochastic Differential Equations with Reflecting Boundary Conditions, *Communications on Pure and Applied Mathematics*, Vol. 37, 511-537, 1984

[33] LIU, H., AND M. LOEWENSTEIN, Optimal Portfolio Selection with Transaction Costs and Finite Horizons, *Review of Financial Studies*, Vol. 15, No. 3, 805-835, 2002

[34] MAGILL, M. J. P., AND G. M. CONSTANTINIDES, Portfolio Selection with Transaction Costs, *Journal of Economic Theory*, Vol. 13, 245-263, 1976

[35] MARKOWITZ, H. M., Portfolio Selection, *Journal of Finance*, Vol. 7, No. 1, 77-91, 1952

[36] MARKOWITZ, H. M., AND X. Y. ZHOU, Private Communication, 2004

[37] MERTON, R. C., Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case, *Review of Economics and Statistics*, Vol. 51, No. 3, 247-57, 1969

[38] MERTON, R. C., An Analytic Derivation of the Efficient Portfolio Frontier, *Journal of Financial and Quantitative Analysis*, Vol. 7, 1851-1872, 1972

[39] VON NEUMANN J., AND O. MORGENSTERN, *Theory of Games and Economic Behavior*, Princeton University Press, 1947

[40] PROTTER, M. H., AND H. F. WEINBERGER, *Maximum Principles in Differential Equations*, Prentice Hall, Englewood Cliffs, N.J., 1967

[41] RICHARDSON, H. R., A Minimum Variance Result in Continuous Trading Portfolio Optimization, *Management Science*, Vol. 35, No. 9, 1045-1055, 1989

[42] SHREVE, S. E., AND H. M. SONER, Optimal Investment and Consumption with Transaction Costs, *The Annals of Applied Probability*, Vol. 4, No. 3, 609-692, 1994

[43] SONER, H. M., AND S. E. SHREVE, A Free Boundary Problem Related to Singular Stochastic Control: The Parabolic Case, *Communication in Partial Differential Equations*, Vol. 16, No. 2&3, 373-424, 1991

[44] XIA, J. M., Mean-Variance Portfolio Choice: Quadratic Partial Hedging, *Mathematical Finance*, Vol. 15, No. 3, 533-538, 2005

[45] XU, Z. Q., *Continuous-Time Mean-Variance Portfolio Selection with Proportional Transaction Costs*, Thesis (Ph.D.), Chinese University of Hong Kong, 2006

[46] YONG, J. M., AND X. Y. ZHOU, *Stochastic Controls: Hamilton Systems and HJB equation*, New York, Springer, 1999

[47] ZHOU, X. Y., AND D. LI, Continuous-Time Mean-Variance Portfolio Selection: A Stochastic LQ Framework, *Applied Mathematics and Optimization*, Vol. 42, 19-33, 2000

[48] ZHOU, X. Y., AND G. YIN, Markowitz's Mean-Variance Portfolio Selection with Regime Switching: A Continuous-Time Model. *SIAM Journal on Control and Optimization* Vol. 42, No. 4, 2003

A Proofs of Propositions 4.4 and 4.5

Proposition 4.5 is straightforward once Proposition 4.4 is proved. So we prove Proposition 4.4 only. Note that equation (4.5) could have a singularity if $v = 0$. To remove the possible singularity, let us begin with the stationary counterpart of the problem. As in Theorem 6.1 of Dai and Yi (2009), we are able to show that the semi-explicit stationary solution is available through a Riccati equation

$$\begin{cases} \mathcal{L}v_\infty = 0, \text{ if } x \in (x^*_{b,\infty}, x^*_{s,\infty}), \\ v_\infty(x^*_{b,\infty}) = x^*_{b,\infty} + 1 + \lambda, \quad v'_\infty(x^*_{b,\infty}) = 1, \\ v_\infty(x^*_{s,\infty}) = x^*_{s,\infty} + 1 - \mu, \quad v'_\infty(x^*_{s,\infty}) = 1, \end{cases}$$

with

$$x^*_{s,\infty} \overset{\text{def}}{=\!=} -\frac{a}{a+k^*}(1-\mu), \tag{A.1}$$

$$x^*_{b,\infty} \overset{\text{def}}{=\!=} -\frac{a}{a+\frac{k^*}{k^*-1}}(1+\lambda), \tag{A.2}$$

where $\mathcal{L}$ is defined in (4.6),

$$a \overset{\text{def}}{=\!=} -\frac{2(\alpha - r + \sigma^2)}{\sigma^2} \in (-\infty, -2),$$

$k^* \in (1,2)$ is the solution to

$$F(k) = \frac{1+\lambda}{1-\mu},$$

$$F(k) \overset{\text{def}}{=\!=} \begin{cases} \dfrac{a+\frac{k}{k-1}}{a+k}\left(\dfrac{(c_1+\frac{k-1}{k}a)(c_2+\frac{1}{k}a)}{(c_2+\frac{k-1}{k}a)(c_1+\frac{1}{k}a)}\right)^{\frac{1}{2(c_2-c_1)}}, & \text{if } \Delta_k < 0, \\ \dfrac{a+\frac{k}{k-1}}{a+k}\exp\left(\dfrac{1}{2}\left(\dfrac{1}{\frac{1}{k}a+\frac{1-a}{4}} - \dfrac{1}{\frac{k-1}{k}a+\frac{1-a}{4}}\right)\right), & \text{if } \Delta_k = 0, \\ \dfrac{a+\frac{k}{k-1}}{a+k}\exp\left(\dfrac{1}{\sqrt{2\Delta_k}}\left(\arctan\dfrac{k(a-1)-4a}{2k\sqrt{2\Delta_k}} - \arctan\dfrac{4a-k(3a+1)}{2k\sqrt{2\Delta_k}}\right)\right), & \text{if } \Delta_k > 0, \end{cases}$$

c_1, c_2 are the two roots of $2c^2 + (a-1)c + \frac{k-1}{k^2}a^2 = 0$, and

$$\Delta_k \overset{\text{def}}{=\!=} \frac{k-1}{k^2}a^2 - \frac{1}{8}(a-1)^2.$$

Let $v(t,x)$ be a solution to equation (4.5) restricted to the region $[0,T) \times \left(-\infty, x^*_{s,\infty}\right)$ with a boundary condition $v(t, x^*_{s,\infty}) = x^*_{s,\infty} + 1 - \mu$. Apparently, v_∞ is a super-solution to equation (4.5) in the region, i.e., $v(t,x) \le v_\infty(x)$ for all $x < -(1-\mu)$, $t \in [0,T)$. It is easy to show that $v_\infty(x)$ is increasing in x. We then deduce

$$v(t,x) \le v_\infty(x^*_{s,\infty}) = x^*_{s,\infty} + 1 - \mu \overset{\text{def}}{=\!=} -C_0 < 0 \ \text{ for } \ x < x^*_{s,\infty}. \tag{A.3}$$

In what follows, we will confine equation (4.5) to the restricted region $[0,T) \times \left(-\infty, x^*_{s,\infty}\right)$ in which, due to (A.3), the equation has no singularity. It is worthwhile pointing out that $v(t,x)$ can be trivially extended to the original region by letting $v(t,x) = x + 1 - \mu$ for $x \ge x^*_{s,\infty}$.

In terms of a penalized approximation (see, for example, Friedman (1988)), it is not hard to show that $v(t,x) \in W_p^{1,2}\left([0,T) \times (-N, x^*_{s,\infty})\right)$ for any $-N < x^*_{s,\infty}$, $p > 1$. By the maximum principle, (4.7) and (4.8) follow. Then we have

$$\frac{\partial}{\partial x}\left[v - (x+1-\mu)\right] = \frac{\partial}{\partial x}\left[v - (x+1+\lambda)\right] = v_x - 1 \leqslant 0,$$

which implies the existence of $x_s(\cdot)$ (or $x_b(\cdot)$) as a single-value function. The monotonicity of $x_s(\cdot)$ and $x_b(\cdot)$ is due to

$$\frac{\partial}{\partial t}\left[v - (x+1-\mu)\right] = \frac{\partial}{\partial t}\left[v - (x+1+\lambda)\right] = v_t \leqslant 0.$$

The proof of (4.12) is the same as that of Theorem 4.5 and 4.7 of Dai and Yi (2009).

It remains to show the smoothness of $x^*_b(\cdot)$ and $x^*_s(\cdot)$. To begin with, let us make a transformation and introduce two lemmas.

Let $z = \log(-x)$, $u(t,z) = v(t,x)$. Then

$$\begin{cases} \max\left\{\min\left\{-u_t - \mathcal{L}_1 u, u - (-e^z + 1 - \mu)\right\}, u - (-e^z + 1 + \lambda)\right\} = 0, \\ u(T,z) = -e^z + 1 - \mu, \qquad\qquad (t,z) \in [0,T) \times \mathscr{Z}, \\ u\left(t, \log\left(-x^*_{s,\infty}\right)\right) = x^*_{s,\infty} + 1 - \mu, \end{cases}$$

where $\mathscr{Z} = \left(\log\left(-x^*_{s,\infty}\right), +\infty\right)$, and

$$\mathcal{L}_1 u = \frac{\sigma^2}{2} u_{zz} - \left(\alpha - r + \frac{\sigma^2}{2}\right) u_z + \left(\alpha - r + \sigma^2\right) u \; - \sigma^2 \left[\frac{u_z^2}{u} + 2e^z \frac{u_z}{u} - 2e^z\right].$$

Lemma A.1. *For any $(t,z) \in [0,T) \times \mathscr{Z}$, we have (1) $u \leqslant -C_0$; (2) $u_z = x v_x \geqslant x = -e^z$, i.e., $u_z + e^z \geqslant 0$; (3) $u - u_z \geqslant 0$. Moreover, there is a constant $C_1 > 0$ such that $u - u_z \geqslant C_1$.*

Proof. Part (1) and (2) are immediate from (A.3) and (4.8). Let us prove Part (3). Denote $w = u_z$. So,

$$\frac{\partial}{\partial z}\left(-u_t - \mathcal{L}_1 u\right) = -w_t - \mathcal{L}_2 w + 2e^z \sigma^2 \left(\frac{u_z}{u} - 1\right)$$

where

$$\mathcal{L}_2 w = \frac{\sigma^2}{2} w_{zz} - \left(\alpha - r + \frac{\sigma^2}{2}\right) w_z + \left(\alpha - r + \sigma^2\right) w - \sigma^2 \left(\frac{2\left(u_z + e^z\right)}{u} w_z - \frac{u_z\left(u_z + 2e^z\right)}{u^2} w\right).$$

We define

$$\begin{aligned} \text{SR} &= \left\{(t,z) \in [0,T) \times \mathscr{Z} \;\middle|\; u = -e^z + 1 - \mu\right\}, \\ \text{BR} &= \left\{(t,z) \in [0,T) \times \mathscr{Z} \;\middle|\; u = -e^z + 1 + \lambda\right\}, \\ \text{NT} &= \left\{(t,z) \in [0,T) \times \mathscr{Z} \;\middle|\; -e^z + 1 - \mu < u < -e^z + 1 + \lambda\right\}. \end{aligned}$$

Notice that we can rewrite

$$-u_t - \mathcal{L}_1 u = -u_t - \mathcal{L}_2 u + 2e^z \sigma^2 \left(\frac{u_z}{u} - 1\right).$$

Denote $H = u - u_z$. Then

$$-H_t - \mathcal{L}_2 H = 0 \text{ in NT}$$

Clearly $H = 1 - \mu$ in SR and at $t = T$, and $H = 1 + \lambda$ in BR. Hence, applying the maximum principle yields $H \geq 0$. Moreover, it is not hard to verify that the coefficients in $\mathcal{L}_2 H$ are bounded. We then infer that there is a constant $C_1 > 0$, such that $H \geqslant C_1$.

Lemma A.2. *There is a constant $C_2 > 0$, such that $u_t \geqslant -C_2$.*

Proof. Let $z_s(t) = \log(-x^*_s(t))$ be corresponding the selling boundary. For $z > z_s(t)$

$$\begin{aligned} u_t|_{t=T} &= -\mathcal{L}_1\left(-e^z + 1 - \mu\right) = -\left(\alpha - r\right)\left(1 - \mu\right) - \frac{\left(1-\mu\right)^2}{-e^z + 1 - \mu} \\ &\geqslant -\left(\alpha - r\right)\left(1 - \mu\right). \end{aligned}$$

Applying the maximum principle gives the desired result.

We are now to prove that both $z_s(\cdot)$ and $z_b(\cdot)$ are C^∞, where $z_s(t) = \log(-x_s^*(t))$ and $z_b(t) = \log(-x_b^*(t))$. Thanks to the bootstrap technique, we only need to show that they are Lipschitz-continuous. Hence, it suffices to prove the cone property, namely, for any $(t, z_0) \in [0, T) \times \mathscr{Z}$, there exists a constant $C > 0$ such that

$$\begin{aligned}(T-t)u_t + C\frac{\partial}{\partial z}\Big(u - (-e^z + 1 - \mu)\Big)\Big|_{(t,z_0)} &\geqslant 0,\\ (T-t)u_t + C\frac{\partial}{\partial z}\Big(u - (-e^z + 1 + \lambda)\Big)\Big|_{(t,z_0)} &\geqslant 0,\end{aligned}$$

which is equivalent to

$$(T-t)u_t + C\left(u_z + e^z\right)\big|_{(t,z_0)} \geqslant 0. \tag{A.4}$$

Now let us prove (A.4). We can only focus on the NT region. Note that

$$\frac{\partial}{\partial t}\Big(- u_t - \mathcal{L}_1 u\Big) = \left(-\frac{\partial}{\partial t} - \mathcal{L}_2\right) u_t.$$

It follows

$$\left(-\frac{\partial}{\partial t} - \mathcal{L}_2\right)\left[(T-t)\, u_t\right] = u_t \text{ in NT.}$$

On the other hand, it is not hard to check

$$\begin{aligned}\left(-\frac{\partial}{\partial t} - \mathcal{L}_2\right)(u_z + e^z) &= \sigma^2 e^z \frac{u - u_z}{u^2}\left(u + u_z + 2e^z\right)\\ &\geqslant \sigma^2 e^z \frac{u - u_z}{u^2}\left(u + u_z - 2u_z\right)\\ &= \sigma^2 e^z \frac{(u - u_z)^2}{u^2} \geqslant \frac{C_1^2\sigma^2 e^z}{u^2}, \text{ in NT}\end{aligned}$$

where $u - u_z \geqslant C_1$ is used in the last inequality. Thus,

$$\begin{aligned}&\left(-\frac{\partial}{\partial t} - \mathcal{L}_2\right)\left[(T-t)u_t + C\left(u_z + e^z\right)\right]\\ \geqslant\ & u_t + C\frac{C_1^2\sigma^2 e^z}{u^2} \geqslant -C_2 + C\frac{C_1^2\sigma^2 e^z}{u^2}, \text{ in NT.}\end{aligned}$$

Since NT is unbounded, we can follow Soner and Shreve (1991) to introduce an auxiliary function $\psi(t, z; z_0) = e^{a(T-t)}\left(z - z_0\right)^2$ with a constant $a > 0$. We can choose a big enough so that

$$\left(-\frac{\partial}{\partial t} - \mathcal{L}_2\right)\psi\left(t, z; z_0\right) \geqslant C_3\left(z - z_0\right)^2 - C_4,$$

where C_3 and C_4 are positive constants independent of (t, z). It follows

$$\begin{aligned}&\left(-\frac{\partial}{\partial t} - \mathcal{L}_2\right)\left[(T-t)u_t + C\left(u_z + e^z\right) + \psi\left(t, z; z_0\right)\right]\\ \geqslant\ & -C_2 + C\frac{C_1^2\sigma^2 e^z}{u^2} + C_3\left(z - z_0\right)^2 - C_4.\end{aligned}$$

Then we can choose $r > 0$ such that

$$C_3 r^2 - C_2 - C_4 \geqslant 0$$

and choose $C > 0$ big enough such that

$$C\frac{C_1^2\sigma^2 e^z}{u^2} - C_2 - C_4 \geqslant 0 \text{ for } |z - z_0| \leqslant r.$$

It then follows

$$\left(-\frac{\partial}{\partial t}-\mathcal{L}_2\right)\left[(T-t)u_t+C\left(u_z+e^z\right)+\psi\left(t,z;z_0\right)\right]\geqslant 0,\ \text{in NT.}$$

Applying the maximum principle and penalty approximation, we conclude

$$(T-t)u_t+C\left(u_z+e^z\right)+\psi\left(t,z;z_0\right)\geqslant 0,\ (t,z)\in[0,T)\times\mathscr{Z}.$$

Letting $z=z_0$, we get the desired result.

B Proof of Proposition 6.2

Proof. From

$$\begin{aligned}V(0,x-\ell e^{-rT},y)-(\ell-z)^2&=\inf_{W\in\mathcal{W}_0^{x-\ell e^{-rT},y}}\mathbf{E}[W^2-(\ell-z)^2]=\inf_{W\in\mathcal{W}_0^{x,y}}\mathbf{E}[(W-\ell)^2-(\ell-z)^2]\\&=\inf_{W\in\mathcal{W}_0^{x,y}}(\mathbf{E}[W^2]-2\ell(\mathbf{E}[W]-z)),\end{aligned}$$

it follows that $V(0,x-\ell e^{-rT},y)-(\ell-z)^2$ is concave in ℓ. So its maximum attains at point ℓ^* which satisfies

$$\frac{\partial}{\partial\ell}\Big(V(0,x-\ell e^{-rT},y)-(\ell-z)^2\Big)\Big|_{\ell=\ell^*}=0,$$

i.e.,

$$e^{-rT}V_x(0,x-\ell^*e^{-rT},y)+2\ell^*=2z.$$

Define

$$f(\ell)\overset{\text{def}}{=\!=}e^{-rT}V_x(0,x-\ell e^{-rT},y)+2\ell.$$

Then by the convexity of $V(0,x-\ell e^{-rT},y)-(\ell-z)^2$ in ℓ, we have that f is increasing. Since $V_x\leqslant 0$, we have $f(z)\leqslant 2z$. By the monotonicity of f, the existence of ℓ^* depends on $\lim_{\ell\to+\infty}f(\ell)$.

- We first consider the case when $T_0>0$. In this case $x_b^*(0)\in(-\infty,0)$. If $y\leqslant 0$, then

  $$(0,x-\ell e^{-rT},y)\in\mathcal{BR},\quad\forall\,\ell\geqslant z.$$

  If $y>0$, then

  $$(0,x-\ell e^{-rT},y)\in\mathcal{BR},\quad\forall\,\ell\geqslant e^{rT}(x-x_b^*(0)y).$$

  Therefore,

  $$\begin{aligned}\lim_{\ell\to+\infty}f(\ell)&=\lim_{\ell\to+\infty}\left(e^{-rT}V_x\left(0,x-\ell e^{-rT},y\right)+2\ell\right)\\&=\lim_{\ell\to+\infty}\left(2e^{-rT}e^{2\mathfrak{B}(0)}\left(x-\ell e^{-rT}+(1+\lambda)y\right)+2\ell\right)\\&=\lim_{\ell\to+\infty}2\left(1-e^{-2rT+2\mathfrak{B}(0)}\right)\ell+2e^{-rT}(x+(1+\lambda)y)\\&=+\infty,\end{aligned}$$

where we have used the fact that $\mathfrak{B}(0) < rT$ when $T_0 > 0$. Therefore, for any

$$z \in (e^{rT}x + (1-\mu)e^{rT}y^+ - (1+\lambda)e^{rT}y^-, +\infty)$$

there exists ℓ^* such that

$$\begin{aligned} &e^{-rT}V_x(0, x-\ell^* e^{-rT}, y) + 2\ell^* = 2z, \\ &V\left(0, x-\ell^* e^{-rT}, y\right) - (\ell^* - z)^2 = \sup_{\ell\in\mathbb{R}}(V(0, x-\ell e^{-rT}, y) - (\ell - z)^2). \end{aligned} \tag{B.1}$$

Now we prove the uniqueness. For $T_0 > 0$, we have

$$\mathfrak{A}(0) < rT, \quad \mathfrak{B}(0) < rT.$$

If $(0, x-\ell e^{-rT}, y) \in \mathcal{SR}$, then

$$f'(\ell) = -e^{-2rT}V_{xx}(0, x-\ell e^{-rT}, y) + 2 = -2e^{-2rT+2\mathfrak{A}(0)} + 2 > 0.$$

Similarly, if $(0, x-\ell e^{-rT}, y) \in \mathcal{BR}$, then

$$f'(\ell) = -e^{-2rT}V_{xx}(0, x-\ell e^{-rT}, y) + 2 = -2e^{-2rT+2\mathfrak{B}(0)} + 2 > 0.$$

By the maximum principle, we have

$$f'(\ell) > 0, \quad \text{for } (0, x-\ell e^{-rT}, y) \in \mathcal{NT}.$$

This implies the uniqueness of ℓ^*.

- Now, we move to the case when $T_0 = 0$. According to Theorem 3.3, we have

$$\mathcal{D} = \begin{cases} (e^{rT}x + (1-\mu)e^{rT}y, e^{rT}x + (1-\mu)e^{\alpha T}y), & \text{if } y > 0, \\ \emptyset, & \text{if } y \leqslant 0. \end{cases}$$

We only need to consider the case of $y > 0$. Note that in this case,

$$(0, x-\ell e^{-rT}, y) \in \mathcal{NT}, \quad \forall\, \ell \geqslant e^{rT}(x - x_s^*(0)y).$$

By the homogeneity property, we have

$$V_x(t, \rho x, \rho y) = \rho V_x(t, x, y), \quad \forall\, (t, x, y, \rho) \in [0, T) \times \mathbb{R}^2 \times \mathbb{R}_+.$$

So we can make the following transformation in $\mathcal{NT}$,

$$z = -\frac{y}{x} \in \left(0, \tfrac{-1}{x_s^*(0)}\right), \quad \bar{v}(t, z) = -\frac{1}{x}V_x(t, x, y).$$

Then

$$\begin{cases} \bar{v}_t + \frac{1}{2}\sigma^2 z^2 \bar{v}_{zz} + (\alpha - r)z\bar{v}_z + 2r\bar{v} = 0, \ \ (t, z) \in [0, T) \times \left(0, \tfrac{-1}{x_s^*(0)}\right). \\ \bar{v}(T, z) = 2(-1 + (1-\mu)z). \end{cases}$$

Therefore

$$\bar{v}(t, 0) = -2e^{2r(T-t)}.$$

Let $\tilde{v}(t,z) = \bar{v}_z(t,z)$, which satisfies

$$\begin{cases} \tilde{v}_t + \frac{1}{2}\sigma^2 z^2 \tilde{v}_{zz} + (\alpha - r + \sigma^2) z\tilde{v}_z + (\alpha + r)\tilde{v} = 0, \quad (t,z) \in [0,T) \times \left(0, \frac{-1}{x_s^*(0)}\right), \\ \tilde{v}(T,z) = 2(1-\mu). \end{cases}$$

Therefore

$$\tilde{v}(t,0) = 2(1-\mu)e^{(\alpha+r)(T-t)}.$$

It follows

$$\begin{aligned} V_x(0,x,y) &= -x\bar{v}\left(0, -\tfrac{y}{x}\right) = -x\left(\bar{v}(0,0) - \tfrac{y}{x}\bar{v}_z(0,0) + O\left(\tfrac{y^2}{x^2}\right)\right) \\ &= 2xe^{2rT} + 2y(1-\mu)e^{(\alpha+r)T} + O\left(\tfrac{y^2}{|x|}\right). \end{aligned}$$

So

$$\begin{aligned} \lim_{\ell\to+\infty} f(\ell) &= \lim_{\ell\to+\infty} \left(e^{-rT}V_x(0, x-\ell e^{-rT}, y) + 2\ell\right) \\ &= \lim_{\ell\to+\infty} \left(e^{-rT}\left(2(x-\ell e^{-rT})e^{2rT} + 2y(1-\mu)e^{(\alpha+r)T}\right) + O\left(\frac{y^2}{|x-\ell e^{-rT}|}\right) + 2\ell\right) \\ &= 2(e^{rT}x + (1-\mu)e^{\alpha T}y). \end{aligned}$$

The monotonicity of f ensures the existence of ℓ^*. The proof for the uniqueness is similar as above.

The proof is complete. □