https://doi.org/10.1016/j.jde.2022.08.006

A free boundary problem arising from a multi-state regime-switching stock trading model

Chonghu Guan* Jing Peng† Zuo Quan Xu‡

May 21, 2022

Abstract

In this paper, we study a free boundary problem, which arises from an optimal trading problem of a stock whose price is driven by unobservable market status and noice processes. The free boundary problem is a variational inequality system of three functions with a degenerate operator. We prove that all the four switching free boundaries are no-overlapping, monotonic and C^∞-smooth by the approximation method. We also completely determine their relative localities and provide the optimal trading strategies for the stock trading problem.

Keywords. free boundary problem; system of parabolic variational inequalities; regime-switching; stock trading

Mathematics Subject Classification. 35R35; 35K87; 91B70; 91B60.

1 Introduction

This paper considers a free boundary problem arising from a stock trading model. The stock price is driven by a two-state market status process which is unobservable to the trader. The details of the financial and stochastic background of the problem are given in Appendix A.

*School of Mathematics, Jiaying University, Meizhou 514015, Guangdong, China. This author is partially supported by NSF of Guangdong Province of China (No. 2021A1515012031 and No. 2022A1515010263), NNSF of China (No. 11901244). Email: 316346917@qq.com.

†Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China. Email: jing.peng@connect.polyu.hk. The author acknowledges financial support from NSFC (No.11971409).

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong. Email: maxu@polyu.edu.hk. The author acknowledges financial support from NSFC (No.11971409), The Hong Kong RGC (GRF No.15202421), The PolyU-SDU Joint Research Center on Financial Mathematics and the CAS AMSS-POLYU Joint Laboratory of Applied Mathematics, The Hong Kong Polytechnic University.

The model reduces to finding a triple of value functions $(v_0(p,t),\ v_1(p,t),\ v_{-1}(p,t))$ that satisfies the following variational inequality (VI) system

$$\begin{cases} \min\Big\{\partial_t v_0 - \mathcal{L}v_0,\ v_0 - v_1 + (1+K),\ v_0 - v_{-1} - (1-K)\Big\} = 0, \\ \min\Big\{\partial_t v_1 - \mathcal{L}v_1,\ v_1 - v_0 - (1-K)\Big\} = 0, \\ \min\Big\{\partial_t v_{-1} - \mathcal{L}v_{-1},\ v_{-1} - v_0 + (1+K)\Big\} = 0, \qquad (p,t)\in\Omega, \end{cases} \tag{1.1}$$

with the initial conditions

$$\begin{cases} v_0(p,0) = 0, \\ v_1(p,0) = 1-K, \\ v_{-1}(p,0) = -(1+K), \qquad 0<p<1, \end{cases} \tag{1.2}$$

where the domain is

$$\Omega = (0,1)\times(0,T],$$

and the operator $\mathcal{L}$ is defined by

$$\mathcal{L} = \frac{1}{2}\left(\frac{(\mu_1-\mu_2)p(1-p)}{\sigma}\right)^2\partial_{pp} + \big(-(\lambda_1+\lambda_2)p + \lambda_2 + (\mu_1-\mu_2)p(1-p)\big)\partial_p + (\mu_1-\mu_2)p + \mu_2 - \rho.$$

The parameters, including the percentage of transaction fee K, the expected return rates of the stock, μ_1 and μ_2, in the different market statues, the volatility rate of the stock σ, the discount rate of the investor ρ, and the coefficients of transition probabilities of the market states λ_1 and λ_2, are all constants and satisfy

$$0<K<1, \quad \mu_1>\rho>\mu_2, \quad \sigma,\ \lambda_1,\ \lambda_2>0.$$

Note that the operator $\mathcal{L}$ is degenerate on the boundaries $p=0$ and $p=1$. According to Fichera's theorem (see [10]), we must not put the boundary conditions on these two boundaries.

Partial differental equation (PDE) technologies are widely used in the economic literature to study similar regime-switching models. For instance, Yi [12] considered the pricing problem of American put option with regime-switching volatility and its related excising region; Khaliq, et al. [6] studied the numerical solution of a class of complex PDE systems about

American option with multiple states regime-switching. Dai, et al.. [3] provided theoretical analysis on the variational inequalities as well as numerical simulations.

Dai, et al. [4] considered the optimal stock trading rule with two net positions: the flat position (no stock holding) and the long position (holding one share of stock); the model can be expressed by a variational inequality system of the corresponding two value functions. The two variational inequalities are reduced to one double obstacle problem on the difference of the two value functions, and then the one-dimensional problem and the properties of its free boundary can be obtained by well-known results on the double obstacle problem in the PDE literature.

Our model is an evolutionary model of Dai, et al. [4]. We introduce the third financially meaningful position - short position. This is not a trivial extension of Dai, et al. [4] in the following sense. Because the stock are allowed to be short in our model, there is one extra (short) position than that of [4]. As a consequence, the Hamilton-Jacobi-Bellman equation (HJB) system involves three variational inequalities and three value functions, which could not be amalgamated into a single one as previous works [1, 2, 4] did, so it calls for completely new technologies to deal with.

Our model is also similar to Ngo and Pham [8]. They considered the problem of determining the optimal cut-off of the pair trading rule for a three-state regime-switching model. However their model is infinity time horizon, so there is no time variable and the system is an ordinary differential equation system for which standard smooth-fitting technique can be applied and a closed-form solution is available. By contrast, we consider a finite time model, so the HJB system becomes a PDE system for which the existence and smoothness of the solution is much harder to establish.

In this paper, we prove the VI system (1.1) with the initial condition (1.2) has four switching free boundaries which are no-overlapping, monotonic and C^∞-smooth; we also completely determine their relative localities and provide the optimal trading strategies for the stock trading problem. The main technical contribution of this paper is that, the properties of the free boundaries of the variational inequality system with coupling appearing in obstacle constraints are studied thoroughly for the first time.

The rest of the paper is arranged as follows. In Section 2, we first construct a penalty approximation system and obtain some estimations of its solution, then completely solve the VI system (1.1) by a limit argument. Section 3 is devoted to the study of the properties of the four switching free boundaries of the VI system (1.1). In Appendix A we give the financial and stochastic background of the problem; and in Appendix B we give the proof of uniqueness for the VI system (1.1).

2 Existence and uniqueness by penalty approximation method

In this section, we show the problem (1.1) admits a solution by the approximation method. The uniqueness of the solution is given in Appendix B.

For this, we first construct an approximation equation system for the VI problem (1.1). For sufficiently small $\varepsilon \in (0, 1/2)$, consider

$$\begin{cases} \partial_t v_0^\varepsilon - \mathcal{L} v_0^\varepsilon + \beta_\varepsilon(v_0^\varepsilon - v_1^\varepsilon + 1 + K) + \beta_\varepsilon(v_0^\varepsilon - v_{-1}^\varepsilon - (1-K)) = 0, \\ \partial_t v_1^\varepsilon - \mathcal{L} v_1^\varepsilon + \beta_\varepsilon(v_1^\varepsilon - v_0^\varepsilon - (1-K)) = 0, \\ \partial_t v_{-1}^\varepsilon - \mathcal{L} v_{-1}^\varepsilon + \beta_\varepsilon(v_{-1}^\varepsilon - v_0^\varepsilon + (1+K)) = 0, \qquad (p,t) \in \Omega, \end{cases} \tag{2.1}$$

with the initial condition

$$\begin{cases} v_0^\varepsilon(p, 0) = 0, \\ v_1^\varepsilon(p, 0) = 1 - K, \\ v_{-1}^\varepsilon(p, 0) = -(1+K), \qquad 0 < p < 1; \end{cases} \tag{2.2}$$

where $\beta_\varepsilon(\cdot)$ is any penalty function satisfying the following properties

$$\beta_\varepsilon(\cdot) \in C^2(-\infty, +\infty), \quad \beta_\varepsilon(0) = -c_0, \quad \beta_\varepsilon(\varepsilon) = -c_1, \quad \beta_\varepsilon(x) = 0 \ \text{ for } x \geqslant 2\varepsilon,$$

$$\beta_\varepsilon(\cdot) \leqslant 0, \quad \beta_\varepsilon'(\cdot) \geqslant 0, \quad \beta_\varepsilon''(\cdot) \leqslant 0, \quad \lim_{\varepsilon \to 0} \beta_\varepsilon(x) = \begin{cases} 0, & x > 0, \\ -\infty, & x < 0, \end{cases}$$

and

$$c_0 = 2(\mu_1 - \mu_2)(1+K) + 2, \quad c_1 = (\mu_1 - \mu_2)(1+K) + 1.$$

Clearly $c_0 > c_1 > 0$. Figure 1 demonstrates such a penalty function $\beta_\varepsilon(\cdot)$.

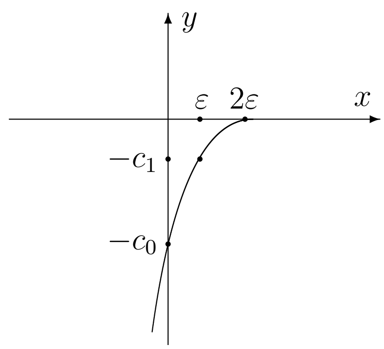

Figure 1: A penalty function $\beta_\varepsilon(\cdot)$.

We first, using the technique of Dai and Yi [2], reduce the problem (2.1) of three functions to a problem of two functions. Let

$$u_1^\varepsilon = v_0^\varepsilon - v_1^\varepsilon, \quad u_{-1}^\varepsilon = v_0^\varepsilon - v_{-1}^\varepsilon.$$

By (2.1) and (2.2), we have

$$\begin{cases} \partial_t u_1^\varepsilon - \mathcal{L}u_1^\varepsilon + \beta_\varepsilon(u_1^\varepsilon + 1 + K) - \beta_\varepsilon(-u_1^\varepsilon - (1-K)) + \beta_\varepsilon(u_{-1}^\varepsilon - (1-K)) = 0, \\ \partial_t u_{-1}^\varepsilon - \mathcal{L}u_{-1}^\varepsilon + \beta_\varepsilon(u_{-1}^\varepsilon - (1-K)) - \beta_\varepsilon(-u_{-1}^\varepsilon + (1+K)) + \beta_\varepsilon(u_1^\varepsilon + 1 + K) = 0, \end{cases} \tag{2.3}$$

and

$$\begin{cases} u_1^\varepsilon(p,0) = -(1-K), \\ u_{-1}^\varepsilon(p,0) = 1+K, \qquad 0 < p < 1. \end{cases} \tag{2.4}$$

Since the operator $\mathcal{L}$ is degenerated on the boundaries $p = 0$ and $p = 1$, we first consider the problem (2.3) in the bounded domain

$$\Omega_\varepsilon := (\varepsilon, 1-\varepsilon) \times (0, T],$$

with the initial condition (2.4) and the additional boundary condition

$$\begin{cases} \partial_p u_1^\varepsilon(\varepsilon, t) = \partial_p u_1^\varepsilon(1-\varepsilon, t) = 0, \\ \partial_p u_{-1}^\varepsilon(\varepsilon, t) = \partial_p u_{-1}^\varepsilon(1-\varepsilon, t) = 0, \qquad 0 \leqslant t < T. \end{cases} \tag{2.5}$$

We now provide the following existence result for the approximation problem (2.3).

Lemma 2.1 *The system (2.3), restricted to the domain Ω_ε, with the initial condition (2.4) and boundary condition (2.5), admits a solution $(u_1^\varepsilon, u_{-1}^\varepsilon) \in C^{2,1}(\overline{\Omega}_\varepsilon) \times C^{2,1}(\overline{\Omega}_\varepsilon)$. Moreover, the solution satisfies*

$$-(1+K)+\varepsilon \leqslant u_1^\varepsilon \leqslant -(1-K), \tag{2.6}$$

$$1-K+\varepsilon \leqslant u_{-1}^\varepsilon \leqslant 1+K, \tag{2.7}$$

in $\overline{\Omega}_\varepsilon$.

PROOF: The existence of $W_p^{2,1}(\Omega_\varepsilon)$ solution can be proved by the Schauder fixed point theorem and the comparison principle for nonlinear equation. Because the process is standard (see [12]), we omit the details. Furthermore, by the Schauder estimation, we also have $(u_1^\varepsilon, u_{-1}^\varepsilon) \in C^{2,1}(\overline{\Omega}_\varepsilon) \times C^{2,1}(\overline{\Omega}_\varepsilon)$.

We come to prove the first inequality in (2.6). Denote

$$\phi = -(1+K)+\varepsilon.$$

Then by simple calculations and recalling the definition of c_1,

$$\begin{aligned}
&\partial_t\phi - \mathcal{L}\phi + \beta_\varepsilon(\phi+1+K) - \beta_\varepsilon(-\phi-(1-K)) + \beta_\varepsilon(u_{-1}^\varepsilon - (1-K))\\
&= [(\mu_1-\mu_2)p+\mu_2-\rho][1+K-\varepsilon] + \beta_\varepsilon(\varepsilon) - \beta_\varepsilon(2K-\varepsilon) + \beta_\varepsilon(u_{-1}^\varepsilon - (1-K))\\
&= [(\mu_1-\mu_2)p+\mu_2-\rho][1+K-\varepsilon] - c_1 + \beta_\varepsilon(u_{-1}^\varepsilon - (1-K))\\
&\leqslant [(\mu_1-\mu_2)+\mu_2-\mu_2][1+K-\varepsilon] - ((\mu_1-\mu_2)(1+K)+1)\\
&< 0,
\end{aligned}$$

so by (2.3),

$$\begin{aligned}
&\partial_t\phi - \mathcal{L}\phi + \beta_\varepsilon(\phi+1+K) - \beta_\varepsilon(-\phi-(1-K))\\
&< \partial_t u_1^\varepsilon - \mathcal{L}u_1^\varepsilon + \beta_\varepsilon(u_1^\varepsilon+1+K) - \beta_\varepsilon(-u_1^\varepsilon-(1-K)).
\end{aligned}$$

By the comparison principle, we obtain $u_1^\varepsilon \geqslant \phi = -(1+K)+\varepsilon$, proving the first inequality in (2.6). Similarly, using $c_1 \geqslant (\rho-\mu_2)(1-K+\varepsilon)$, we can prove $1-K+\varepsilon \leqslant u_{-1}^\varepsilon$ in (2.7).

Now, we prove the second inequality in (2.6). Denote

$$\Phi = -(1-K).$$

Then

$$
\begin{aligned}
&\partial_t \Phi - \mathcal{L}\Phi + \beta_\varepsilon(\Phi + 1 + K) - \beta_\varepsilon(-\Phi - (1-K)) + \beta_\varepsilon(u^\varepsilon_{-1} - (1-K)) \\
&= [(\mu_1 - \mu_2)p + \mu_2 - \rho](1-K) + \beta_\varepsilon(2K) - \beta_\varepsilon(0) + \beta_\varepsilon(u^\varepsilon_{-1} - (1-K)) \\
&= [(\mu_1 - \mu_2)p + \mu_2 - \rho](1-K) + c_0 + \beta_\varepsilon(u^\varepsilon_{-1} - (1-K)) \\
&\geqslant (\mu_2 - \rho)(1-K) + c_0 + \beta_\varepsilon(1 - K + \varepsilon - (1-K)) \\
&\geqslant (\mu_2 - \mu_1)(1-K) + c_0 - c_1 \\
&> 0,
\end{aligned}
$$

where the first inequality is due to $\beta' \geqslant 0$ and the first inequality in (2.7). It follows from (2.3) that

$$
\begin{aligned}
&\partial_t \Phi - \mathcal{L}\Phi + \beta_\varepsilon(\Phi + 1 + K) - \beta_\varepsilon(-\Phi - (1-K)) \\
&> \partial_t u^\varepsilon_1 - \mathcal{L}u^\varepsilon_1 + \beta_\varepsilon(u^\varepsilon_1 + 1 + K) - \beta_\varepsilon(-u^\varepsilon_1 - (1-K)).
\end{aligned}
$$

By the comparison principle, we have $u^\varepsilon_1 \leqslant \Phi = -(1-K)$, giving the second inequality in (2.6). Similarly, using $c_0 \geqslant c_1 + (\mu_1 - \rho)(1+K)$, we can prove $u^\varepsilon_{-1} \leqslant 1 + K$ in (2.7). □

Lemma 2.2 *Let $(u^\varepsilon_1, u^\varepsilon_{-1})$ be given in Lemma 2.1, we have*

$$
\partial_p u^\varepsilon_1 \leqslant 0, \quad \partial_p u^\varepsilon_{-1} \geqslant 0. \tag{2.8}
$$

Proof: Denote $w_i = \partial_p u^\varepsilon_i$, for $i = 1, -1$. After differentiating (2.3) w.r.t. p and we get

$$
\begin{cases}
\partial_t w_1 - \mathcal{T} w_1 - (\mu_1 - \mu_2)u^\varepsilon_1 + \beta'_\varepsilon(u^\varepsilon_1 + 1 + K)w_1 \\
\qquad + \beta'_\varepsilon(-u^\varepsilon_1 - (1-K))w_1 + \beta'_\varepsilon(u^\varepsilon_{-1} - (1-K))w_{-1} = 0, \\
\partial_t w_{-1} - \mathcal{T} w_{-1} - (\mu_1 - \mu_2)u^\varepsilon_{-1} + \beta'_\varepsilon(u^\varepsilon_{-1} - (1-K))w_{-1} \\
\qquad + \beta'_\varepsilon(-u^\varepsilon_{-1} + (1+K))w_{-1} + \beta'_\varepsilon(u^\varepsilon_1 + 1 + K)w_1 = 0,
\end{cases} \tag{2.9}
$$

where the operator $\mathcal{T}$ is defined by

$$
\begin{aligned}
\mathcal{T} &:= \mathcal{L} + \frac{(\mu_1 - \mu_2)^2}{\sigma^2} p(1-p)(1-2p)\partial_p + [-(\lambda_1 + \lambda_2) + (\mu_1 - \mu_2)(1-2p)] \\
&= \frac{1}{2}\left(\frac{(\mu_1 - \mu_2)p(1-p)}{\sigma}\right)^2 \partial_{pp} \\
&\quad + \left(-(\lambda_1 + \lambda_2)p + \lambda_2 + (\mu_1 - \mu_2)p(1-p) + \frac{(\mu_1 - \mu_2)^2}{\sigma^2} p(1-p)(1-2p)\right)\partial_p \\
&\quad + \mu_2 - \rho - (\lambda_1 + \lambda_2) + (\mu_1 - \mu_2)(1-p).
\end{aligned}
$$

Define

$$W_i = e^{-\lambda t} w_i, \quad i = 1, -1,$$

where λ is a constant that will be determined latter. To prove (2.8), it suffices to prove

$$W_1 \leqslant 0, \quad W_{-1} \geqslant 0. \tag{2.10}$$

Using $u_1^\varepsilon < 0$, $u_{-1}^\varepsilon > 0$ and $\mu_1 > \mu_2$, we get from (2.9) that

$$\begin{cases} \partial_t W_1 - \mathcal{T} W_1 + \lambda W_1 + \beta_\varepsilon'(u_1^\varepsilon + 1 + K) W_1 \\ \qquad + \beta_\varepsilon'(-u_1^\varepsilon - (1-K)) W_1 + \beta_\varepsilon'(u_{-1}^\varepsilon - (1-K)) W_{-1} < 0, \\ \partial_t W_{-1} - \mathcal{T} W_{-1} + \lambda W_{-1} + \beta_\varepsilon'(u_{-1}^\varepsilon - (1-K)) W_{-1} \\ \qquad + \beta_\varepsilon'(-u_{-1}^\varepsilon + (1+K)) W_{-1} + \beta_\varepsilon'(u_1^\varepsilon + 1 + K) W_1 > 0. \end{cases} \tag{2.11}$$

Clearly, there exist (p_1, t_1), $(p_{-1}, t_{-1}) \in \overline{\Omega}_\varepsilon$ such that

$$W_1(p_1, t_1) = \max_{(p,t)\in\overline{\Omega}_\varepsilon} W_1(p,t), \quad W_{-1}(p_{-1}, t_{-1}) = \min_{(p,t)\in\overline{\Omega}_\varepsilon} W_{-1}(p,t).$$

Suppose (2.10) was not true, then we would have

$$\max\{W_1(p_1, t_1), -W_{-1}(p_{-1}, t_{-1})\} > 0.$$

Without loss of generality, we may assume that

$$W_1(p_1, t_1) \geqslant -W_{-1}(p_{-1}, t_{-1}) \quad \text{and} \quad W_1(p_1, t_1) > 0.$$

Notice $W_1 = 0$ on $\partial_p \Omega_\varepsilon$ (the parabolic boundary of Ω_ε), so (p_1, t_1) is inside the domain Ω_ε or at the upper boundary of Ω_ε, and thus,

$$\partial_p W_1(p_1, t_1) = 0, \quad \partial_{pp} W_1(p_1, t_1) \leqslant 0, \quad \partial_t W_1(p_1, t_1) \geqslant 0.$$

Choosing $\lambda > \mu_2 - \rho - (\lambda_1 + \lambda_2) + (\mu_1 - \mu_2)(1-p) + 1$, the above, after simple calculation, implies that

$$\partial_t W_1 - \mathcal{T} W_1 + (\lambda - \beta_\varepsilon'(0))\, W_1 \Big|_{(p_1, t_1)} > 0. \tag{2.12}$$

One the other hand, together with $W_1(p_1,t_1) > 0$ and $\beta' \geqslant 0$, the first inequality in (2.11) leads to

$$\begin{aligned}
&\partial_t W_1 - \mathcal{T} W_1 + \lambda W_1\Big|_{(p_1,t_1)}\\
&< -\beta_\varepsilon'(u_1^\varepsilon + 1 + K)W_1 - \beta_\varepsilon'(-u_1^\varepsilon - (1-K))W_1 - \beta_\varepsilon'(u_{-1}^\varepsilon - (1-K))W_{-1}\Big|_{(p_1,t_1)}\\
&\leqslant -\beta_\varepsilon'(u_{-1}^\varepsilon - (1-K))W_{-1}\Big|_{(p_1,t_1)}.
\end{aligned}$$

Because $-W_{-1}(p_1,t_1) \leqslant -W_{-1}(p_{-1},t_{-1}) \leqslant W_1(p_1,t_1)$, $\beta' \geqslant 0$, $\beta'' \leqslant 0$, and $u_{-1}^\varepsilon > 1-K$, the above is

$$\leqslant \beta_\varepsilon'(u_{-1}^\varepsilon - (1-K))W_1\Big|_{(p_1,t_1)} \leqslant \beta_\varepsilon'(0)W_1\Big|_{(p_1,t_1)},$$

contradicting to (2.12). Therefore, (2.10) holds true and the claim is proved. □

Lemma 2.3 *Let $(u_1^\varepsilon, u_{-1}^\varepsilon)$ be given in Lemma 2.1, we have*

$$u_1^\varepsilon + u_{-1}^\varepsilon \geqslant 0. \tag{2.13}$$

PROOF: Denote $U = u_1^\varepsilon + u_{-1}^\varepsilon$. Adding up the two equations in (2.3), we have

$$\begin{aligned}
&\partial_t U - \mathcal{L}U + \left[\beta_\varepsilon(u_1^\varepsilon + 1 + K) - \beta_\varepsilon(-u_{-1}^\varepsilon + (1+K))\right]\\
&+\left[\beta_\varepsilon(u_{-1}^\varepsilon - (1-K)) - \beta_\varepsilon(-u_1^\varepsilon - (1-K))\right] + \beta_\varepsilon(u_1^\varepsilon + 1 + K) + \beta_\varepsilon(u_{-1}^\varepsilon - (1-K)) = 0.
\end{aligned}$$

Using the mean value theorem, there exit $\xi(p,t)$ between $u_1^\varepsilon + 1 + K$ and $-u_{-1}^\varepsilon + 1 + K$, $\eta(p,t)$ between $u_{-1}^\varepsilon - (1-K)$ and $-u_{-1}^\varepsilon + (1+K)$, such that

$$\begin{aligned}
\beta_\varepsilon(u_1^\varepsilon + 1 + K) - \beta_\varepsilon(-u_{-1}^\varepsilon + (1+K)) &= \beta_\varepsilon'(\xi)U,\\
\beta_\varepsilon(u_{-1}^\varepsilon - (1-K)) - \beta_\varepsilon(-u_1^\varepsilon - (1-K)) &= \beta_\varepsilon'(\eta)U.
\end{aligned}$$

It follows that

$$\partial_t U - \mathcal{L}U + \beta_\varepsilon'(\xi)U + \beta_\varepsilon'(\eta)U = -\beta_\varepsilon(u_1^\varepsilon + 1 + K) - \beta_\varepsilon(u_{-1}^\varepsilon - (1-K)) \geqslant 0,$$

Using the maximum principle we get $U \geqslant 0$, that is (2.13). □

Proposition 2.4 *The system* (2.1) *with the initial condition* (2.2) *and the following boundary condition*

$$\begin{cases} \partial_p v_0^\varepsilon(\varepsilon,t) = \partial_p v_0^\varepsilon(1-\varepsilon,t) = 0, \\ \partial_p v_1^\varepsilon(\varepsilon,t) = \partial_p v_1^\varepsilon(1-\varepsilon,t) = 0, \\ \partial_p v_{-1}^\varepsilon(\varepsilon,t) = \partial_p v_{-1}^\varepsilon(1-\varepsilon,t) = 0, \end{cases} \tag{2.14}$$

has a solution $(v_0^\varepsilon, v_1^\varepsilon, v_{-1}^\varepsilon) \in C^{2,1}(\overline{\Omega}_\varepsilon) \times C^{2,1}(\overline{\Omega}_\varepsilon) \times C^{2,1}(\overline{\Omega}_\varepsilon)$, *and it is bounded by some constant independent of* ε.

PROOF: Let $(u_1^\varepsilon, u_{-1}^\varepsilon)$ be given in Lemma 2.1. Then the following linear problem of v_0^ε,

$$\begin{cases} \partial_t v_0^\varepsilon - \mathcal{L} v_0^\varepsilon + \beta_\varepsilon(u_1^\varepsilon + 1 + K) + \beta_\varepsilon(u_{-1}^\varepsilon - (1-K)) = 0, \\ \partial_p v_0^\varepsilon(\varepsilon,t) = 0, \quad \partial_p v_0^\varepsilon(1-\varepsilon,t) = 0, \quad v_0^\varepsilon(p,0) = 0, \qquad (p,t) \in \Omega_\varepsilon, \end{cases}$$

has a unique solution $v_0^\varepsilon \in C^{2,1}(\overline{\Omega}_\varepsilon)$. Let $v_1^\varepsilon = v_0^\varepsilon - u_1^\varepsilon$ and $v_{-1}^\varepsilon = v_0^\varepsilon - u_{-1}^\varepsilon$, then it is easy to verify that $(v_0^\varepsilon, v_1^\varepsilon, v_{-1}^\varepsilon)$ is a solution to the system (2.1) with the initial condition (2.2) and the boundary condition (2.14).

Moreover, using the first inequality in (2.6) and the first inequality in (2.7), it is easy to prove

$$0 \leqslant v_0^\varepsilon \leqslant 2c_1 e^{\mu_1 t} \tag{2.15}$$

by the comparison principle. Together with (2.6) and (2.7), we conclude $|v_i^\varepsilon|$, $i = 0, 1, -1$ are bounded by some constant independent of ε. □

Our first main result, which completely characterizes the solution of the system (1.1), is given below.

Theorem 2.5 *The variational inequality system* (1.1) *with the initial condition* (1.2) *has a unique bounded solution* (v_0, v_1, v_{-1}) *such that* $v_i \in C^{1+\alpha, \frac{1+\alpha}{2}}(\Omega \bigcup \{t=0\}) \bigcap W^{2,1}_{q,\mathrm{loc}}(\Omega \bigcup \{t = 0\})$, *for any* $q > 3$, $i = 0, 1, -1$, *where* $\alpha = 1 - 3/q$. *Furthermore,*

$$-(1+K) \leqslant v_0 - v_1 \leqslant -(1-K), \tag{2.16}$$

$$1 - K \leqslant v_0 - v_{-1} \leqslant 1 + K, \tag{2.17}$$

$$\partial_p(v_0 - v_1) \leqslant 0, \quad \partial_p(v_0 - v_{-1}) \geqslant 0. \tag{2.18}$$

$$v_0 - v_1 + v_0 - v_{-1} \geqslant 0. \tag{2.19}$$

Proof: Let $(v_0^\varepsilon, v_1^\varepsilon, v_{-1}^\varepsilon)$ be given in Proposition 2.4. Fix $r \in (0, 1/2)$, apply $W_p^{2,1}$ interior estimate (see [7]) to (2.1), we have for any $q > 3$,

$$|v_0^\varepsilon|_{W_q^{2,1}(\Omega_r)},\ |v_1^\varepsilon|_{W_q^{2,1}(\Omega_r)},\ |v_{-1}^\varepsilon|_{W_q^{2,1}(\Omega_r)} \leqslant C_r,$$

where C_r is independent of ε. For any $r \in (0, 1/2)$, $q > 3$ and $i = -1, 0, 1$, by the Sobolev Embedding Theorem, i.e., $W_q^{2,1}(\Omega_r) \subseteq\subseteq C^{1+\alpha, \frac{1+\alpha}{2}}(\overline{\Omega}_r)$, with $\alpha = 1 - 3/q$, there exists a subsequence of $v_i^\varepsilon \in C^{1+\alpha, \frac{1+\alpha}{2}}(\Omega \bigcup \{t = 0\}) \bigcap W_{q,\text{loc}}^{2,1}(\Omega \bigcup \{t = 0\})$, which we still denote by v_i^ε, such that

$$v_i^\varepsilon \longrightarrow v_i \quad \text{uniformly in } C^{1+\alpha, \frac{1+\alpha}{2}}(\overline{\Omega}_r), \quad \text{weekly in } W_q^{2,1}(\Omega_r).$$

Since $|v_i^\varepsilon|$, $i = 0, 1, -1$ are bounded by some constant independent of ε, so are $|v_i|$, $i = 0, 1, -1$.

By (2.1), we have

$$\partial_t v_i^\varepsilon - \mathcal{L} v_i^\varepsilon \geqslant 0 \quad \text{in } \Omega.$$

Letting $\varepsilon \to 0$, we have

$$\partial_t v_i - \mathcal{L} v_i \geqslant 0 \quad \text{in } \Omega. \tag{2.20}$$

Moreover, from (2.6), (2.7), (2.8) and (2.13), we obtain (2.16), (2.17), (2.18) and (2.19), respectively.

Now, we prove the first variational inequality in (1.1). From (2.20), (2.16) and (2.17) we obtain

$$\min \Big\{ \partial_t v_0 - \mathcal{L} v_0,\ v_0 - v_1 + (1 + K),\ v_0 - v_{-1} - (1 - K) \Big\} \geqslant 0.$$

In the following, we come to prove the equality holds.

Suppose $v_0 - v_1 + (1 + K) > 2\varepsilon_0$ and $v_0 - v_{-1} - (1 - K) > 2\varepsilon_0$ hold at some point $(p, t) \in \Omega \bigcup \{t = 0\}$ for some $\varepsilon_0 > 0$. By the continuity of v_i and the uniformly convergence of v_i^ε in $C^{1+\alpha, \frac{1+\alpha}{2}}(\Omega \bigcup \{t = 0\})$, we have

$$(v_0^\varepsilon - v_1^\varepsilon)(p, t) + (1 + K) > \varepsilon_0, \quad (v_0^\varepsilon - v_{-1}^\varepsilon)(p, t) - (1 - K) > \varepsilon_0,$$

for small enough $\varepsilon > 0$. Thus by the first equation in (2.1),

$$(\partial_t v_0^\varepsilon - \mathcal{L} v_0^\varepsilon)(p, t) = 0.$$

Let $\varepsilon \to 0$ we get $(\partial_t v_0 - \mathcal{L} v_0)(p, t) = 0$. Therefore, we proved (v_0, v_1, v_{-1}) satisfies the first variational inequality in (1.1). The other two variational inequalities in (1.1) can be proved similarly.

The proof of uniqueness is a slight modification of the standard proof, so we put it in Appendix B. □

From now on, let (v_0, v_1, v_{-1}) denote the unique solution of the VI system (1.1), given in Theorem 2.5.

3 Free boundaries and optimal trading strategies

In this section, we study the free boundaries for the VI system (1.1), and provide the optimal trading strategies for the stock trading problem given in Appendix A.

Define four trading regions $\mathcal{S}_{i,j}$, switching from position i to position j, $i \neq j \in \{0, -1, 1\}$, as

$$\mathcal{S}_{0,1} = \{(p,t) \in \Omega \mid v_0 - v_1 + 1 + K = 0\},$$

$$\mathcal{S}_{0,-1} = \{(p,t) \in \Omega \mid v_0 - v_{-1} - (1 - K) = 0\},$$

$$\mathcal{S}_{1,0} = \{(p,t) \in \Omega \mid v_1 - v_0 - (1 - K) = 0\},$$

$$\mathcal{S}_{-1,0} = \{(p,t) \in \Omega \mid v_{-1} - v_0 + 1 + K = 0\}.$$

Define the free trading boundaries for $0 \leqslant t < T$ as

$$p_{0,1}(t) = \inf\{p \mid (p,t) \in \mathcal{S}_{0,1}\},$$

$$p_{-1,0}(t) = \inf\{p \mid (p,t) \in \mathcal{S}_{-1,0}\},$$

$$p_{0,-1}(t) = \sup\{p \mid (p,t) \in \mathcal{S}_{0,-1}\},$$

$$p_{1,0}(t) = \sup\{p \mid (p,t) \in \mathcal{S}_{1,0}\}.$$

Since $\partial_p(v_0 - v_1) \leqslant 0$, $\partial_p(v_0 - v_{-1}) \geqslant 0$ in Ω, we see that

$$\mathcal{S}_{0,1} = \{(p,t) \in \Omega \mid p \geqslant p_{0,1}(t)\},$$

$$\mathcal{S}_{-1,0} = \{(p,t) \in \Omega \mid p \geqslant p_{-1,0}(t)\},$$

$$\mathcal{S}_{0,-1} = \{(p,t) \in \Omega \mid p \leqslant p_{0,-1}(t)\},$$

$$\mathcal{S}_{1,0} = \{(p,t) \in \Omega \mid p \leqslant p_{1,0}(t)\}.$$

Define a critical threshold

$$p_0 := \frac{\rho - \mu_2}{\mu_1 - \mu_2} \in (0,1).$$

The second main result of this paper is the completely characterization of the free boundaries defined above.

Theorem 3.1 *The free boundaries $p_{0,-1}(t)$ and $p_{-1,0}(t)$ are strictly increasing, and $p_{0,1}(t)$ and $p_{1,0}(t)$ are strictly decreasing; and they have no-overlapping,*

$$0 \leqslant p_{0,-1}(t) < p_{1,0}(t) < p_0 < p_{-1,0}(t) < p_{0,1}(t) \leqslant 1,$$

with

$$p_{1,0}(0+) = p_0, \quad p_{-1,0}(0+) = p_0.$$

Also, they are C^∞ smooth, i.e.,

$$p_{0,-1}(t), \quad p_{1,0}(t), \quad p_{-1,0}(t), \quad p_{0,1}(t) \in C^\infty((0,T]).$$

Moreover, there exists $t_1 \geqslant \frac{1}{\mu_1-\rho} \log \frac{1+K}{1-K}$ such that $p_{0,1}(t) = 1$ if $t \leqslant t_1$; and $t_0 \geqslant \frac{1}{\rho-\mu_2} \log \frac{1+K}{1-K}$, such that $p_{0,-}(t) = 0$ if $t \leqslant t_0$.

Theorem 3.1 is an immediately consequence of Propositions 3.2 - 3.9 in the rest part of this section. Figure 2 demonstrates the shapes and locations of these free boundaries.

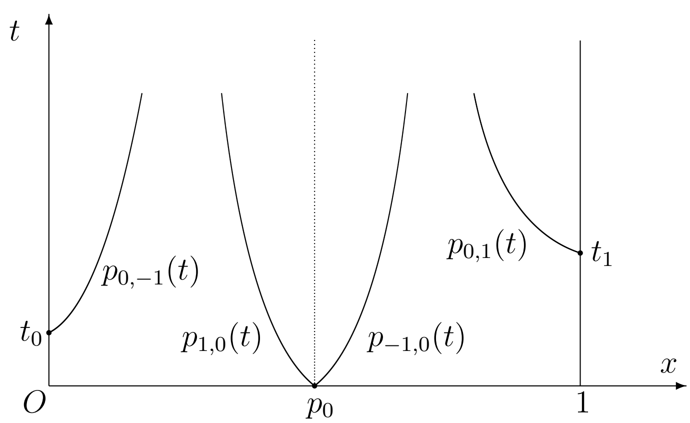

Figure 2: The free boundaries.

We summarize the optimal trading strategies in all the positions and status in Table 1.

status	in short position	in flat position	in long position
$0 < p \leqslant p_{0,-1}(t)$	do nothing	sell 1 share	sell 2 shares
$p_{0,-1}(t) < p \leqslant p_{1,0}(t)$	do nothing	do nothing	sell 1 share
$p_{1,0}(t) < p < p_{-1,0}(t)$	do nothing	do nothing	do nothing
$p_{-1,0}(t) \leqslant p < p_{0,1}(t)$	buy 1 share	do nothing	do nothing
$p_{0,1}(t) \leqslant p < 1$	buy 2 shares	buy 1 share	do nothing

Table 1: The optimal strategy

Proposition 3.2 *We have*

$$\mathcal{S}_{0,1} \subseteq \mathcal{S}_{-1,0} \subseteq \big[p_0, 1\big) \times (0, T], \tag{3.1}$$

and

$$\mathcal{S}_{0,-1} \subseteq \mathcal{S}_{1,0} \subseteq \big(0, p_0\big] \times (0, T]. \tag{3.2}$$

As a consequence,

$$0 \leqslant p_{0,-1}(t) \leqslant p_{1,0}(t) \leqslant p_0 \leqslant p_{-1,0}(t) \leqslant p_{0,1}(t) \leqslant 1, \quad 0 \leqslant t < T. \tag{3.3}$$

PROOF: We only prove (3.1) as the proof of (3.2) is similar.

Due to (2.19) and (1.1), we have

$$v_0 - v_1 + 1 + K \geqslant v_{-1} - v_0 + 1 + K \geqslant 0.$$

This implies the first part of (3.1), namely, $\mathcal{S}_{0,1} \subseteq \mathcal{S}_{-1,0}$.

Now, we prove the second part of (3.1), i.e., $\mathcal{S}_{-1,0} \subseteq \big[p_0, 1\big) \times (0, T]$. For any $(p, t) \in \mathcal{S}_{-1,0}$, we have $v_{-1} - v_0 + (1 + K) = 0$. Thus by (1.1),

$$\partial_t v_{-1} - \mathcal{L} v_{-1} \geqslant 0.$$

Suppose $(p, t) \in \mathcal{S}_{-1,0} \setminus \mathcal{S}_{0,1}$, i.e. $v_0 - v_1 + 1 + K > 0$. Noticing $v_0 - v_{-1} - (1 - K) = 2K > 0$, we have by (1.1),

$$\partial_t v_0 - \mathcal{L} v_0 = 0.$$

So

$$0 \leqslant \Big(\partial_t - \mathcal{L}\Big)\Big(v_{-1} - v_0\Big) = \Big(\partial_t - \mathcal{L}\Big)\Big(-(1 + K)\Big) = [(\mu_1 - \mu_2)p + \mu_2 - \rho](1 + K)$$

leading to

$$p \geqslant \frac{\rho - \mu_2}{\mu_1 - \mu_2} = p_0. \tag{3.4}$$

Otherwise, $(p, t) \in \mathcal{S}_{0,1}$, i.e. $v_0 - v_1 + 1 + K = 0$, then $v_1 - v_0 - (1 - K) = 2K > 0$. By (1.1),

$$\partial_t v_1 - \mathcal{L} v_1 = 0.$$

Note that $v_{-1} - v_1 + 2(1 + K) = 0$, thus

$$0 \leqslant \Big(\partial_t - \mathcal{L}\Big)\Big(v_{-1} - v_1\Big) = \Big(\partial_t - \mathcal{L}\Big)\Big(-2(1 + K)\Big) = 2[(\mu_1 - \mu_2)p + \mu_2 - \rho](1 + K)$$

which also implies (3.4). So we proved the second part of (3.1). $\square$

The financial significance of this proposition is obvious. For instance, the first part of (3.1) means that if the trader wants to transfer from the present short position to the long position, he must first enter the flat position before going to the long position. The second part of (3.1) means that the trader should not buy the stock when the (estimated) chance of the market status being bull, p, is too small (that is, less than the constant p_0). The others can be explained similarly. In one sentence, the higher the chance to be a bull market, the more the stock holding.

Lemma 3.3 *We have*

$$\partial_t(v_0 - v_1) \leqslant 0, \tag{3.5}$$

$$\partial_t(v_0 - v_{-1}) \leqslant 0. \tag{3.6}$$

PROOF: Come back to $(u_1^\varepsilon, u_{-1}^\varepsilon)$, the solution of (2.3). It only needs to prove that, for $i = 1$, -1, and any fixed $\Delta t \in (0, T)$,

$$\widehat{u}_i^\varepsilon(p,t) := u_i^\varepsilon(p, t+\Delta t) \leqslant u_i^\varepsilon(p,t) + [3\varepsilon - \varepsilon^2 p(1-p)]e^{\lambda T} \quad \text{in } Q := [\varepsilon, 1-\varepsilon] \times [0, T-\Delta t],$$

where $\lambda = \mu_1 + \mu_2$ is a constant. Define

$$U_i = e^{-\lambda t} u_i^\varepsilon.$$

Our problem reduces to proving $\widehat{U}_i(p,t) := U_i(p, t+\Delta t)$ satisfies

$$e^{\lambda \Delta t}\widehat{U}_i(p,t) \leqslant U_i(p,t) + [3\varepsilon - \varepsilon^2 p(1-p)]e^{\lambda(T-t)} \quad \text{in } Q. \tag{3.7}$$

Suppose (3.7) was not true when $i = 1$. Denote by (p^*, t^*) a maximum point of $e^{\lambda\Delta t}\widehat{U}_1 - U_1 - [3\varepsilon - \varepsilon^2 p(1-p)]$ in Q, then we have

$$\Big(e^{\lambda\Delta t}\widehat{U}_1 - U_1 - 2\varepsilon\Big)(p^*, t^*) > 0,$$

which implies

$$\Big(\widehat{u}_1^\varepsilon - u_1^\varepsilon - 2\varepsilon\Big)(p^*, t^*) > 0. \tag{3.8}$$

Since $\partial_p\Big(e^{\lambda\Delta t}\widehat{U}_1 - U_1 - [3\varepsilon - \varepsilon^2 p(1-p)]\Big) = \varepsilon^2(1-2\varepsilon) > 0$ on $\{p = \varepsilon\}$ and $\partial_p\Big(e^{\lambda\Delta t}\widehat{U}_1 - U_1 - [3\varepsilon - \varepsilon^2 p(1-p)]\Big) = \varepsilon^2(2\varepsilon - 1) < 0$ on $\{p = 1-\varepsilon\}$, (p^*, t^*) does not lie on the two boundaries. Also since

$$\Big(e^{\lambda\Delta t}\widehat{U}_1 - U_1 - [3\varepsilon - \varepsilon^2 p(1-p)]\Big)(p, 0) = u_1^\varepsilon(p, \Delta t) - [-(1-K)] - [3\varepsilon - \varepsilon^2 p(1-p)] < 0,$$

we conclude that (p^*,t^*) is inside the domain Q or at the upper boundary of Q so that

$$\partial_p\Big(e^{\lambda\Delta t}\widehat{U}_1 - U_1 - [3\varepsilon - \varepsilon^2 p(1-p)]\Big)(p^*,t^*) = 0,$$

$$\partial_{pp}\Big(e^{\lambda\Delta t}\widehat{U}_1 - U_1 - [3\varepsilon - \varepsilon^2 p(1-p)]\Big)(p^*,t^*) \leqslant 0,$$

$$\partial_t\Big(e^{\lambda\Delta t}\widehat{U}_1 - U_1 - [3\varepsilon - \varepsilon^2 p(1-p)]\Big)(p^*,t^*) \geqslant 0.$$

The above implies that if ε is small enough,

$$\Big(\partial_t - \mathcal{L}\Big)\Big(\widehat{u}_1^\varepsilon - u_1^\varepsilon\Big)\Big|_{(p^*,t^*)} = e^{\lambda t}\Big(\partial_t - \mathcal{L} + \lambda\Big)\Big(e^{\lambda\Delta t}\widehat{U}_1^\varepsilon - U_1^\varepsilon\Big)\Big|_{(p^*,t^*)} > 0. \tag{3.9}$$

On the other hand, since both u_1^ε and $\widehat{u}_1^\varepsilon$ satisfy the first equation in (2.3), we have

$$\begin{aligned}\Big(\partial_t - \mathcal{L}\Big)(\widehat{u}_1^\varepsilon - u_1^\varepsilon) = &-\beta_\varepsilon(\widehat{u}_1^\varepsilon + 1 + K) + \beta_\varepsilon(-\widehat{u}_1^\varepsilon - (1-K)) - \beta_\varepsilon(\widehat{u}_{-1}^\varepsilon - (1-K)) \\ &+ \beta_\varepsilon(u_1^\varepsilon + 1 + K) - \beta_\varepsilon(-u_1^\varepsilon - (1-K)) + \beta_\varepsilon(u_{-1}^\varepsilon - (1-K)). \end{aligned} \tag{3.10}$$

By (3.8) and (2.6), we get

$$-u_1^\varepsilon(p^*,t^*) - (1-K) > -\widehat{u}_1^\varepsilon(p^*,t^*) - (1-K) + 2\varepsilon \geqslant 2\varepsilon.$$

And by (2.13),

$$u_{-1}^\varepsilon(p^*,t^*) - (1-K) \geqslant -u_1^\varepsilon(p^*,t^*) - (1-K) > 2\varepsilon.$$

Thus, by the definition of β_ε,

$$\beta_\varepsilon(u_{-1}^\varepsilon - (1-K))\Big|_{(p^*,t^*)} = \beta_\varepsilon(-u_1^\varepsilon - (1-K))\Big|_{(p^*,t^*)} = 0.$$

Moreover, due to $\beta_\varepsilon' \geqslant 0$, $-\widehat{u}_1^\varepsilon \leqslant \widehat{u}_{-1}^\varepsilon$, and $u_1^\varepsilon(p^*,t^*) < \widehat{u}_1^\varepsilon(p^*,t^*)$, we have

$$\beta_\varepsilon(-\widehat{u}_1^\varepsilon - (1-K)) - \beta_\varepsilon(\widehat{u}_{-1}^\varepsilon - (1-K))\Big|_{(p^*,t^*)} \leqslant 0, \quad \beta_\varepsilon(u_1^\varepsilon + 1 + K) - \beta_\varepsilon(\widehat{u}_1^\varepsilon + 1 + K)\Big|_{(p^*,t^*)} \leqslant 0.$$

Therefore, by (3.10) we have

$$\Big(\partial_t - \mathcal{L}\Big)\Big(\widehat{u}_1^\varepsilon - u_1^\varepsilon\Big)\Big|_{(p^*,t^*)} \leqslant 0,$$

which contradicts (3.9). Therefore, (3.7) is true.

Letting $\varepsilon \to 0$ in (3.7), we get

$$\Big(v_0 - v_1\Big)(p, t+\Delta t) \leqslant \Big(v_0 - v_1\Big)(p,t),$$

which implies (3.5). The proof of (3.6) is similar. □

By Lemma 3.3, we obtain

Proposition 3.4 *The free boundaries $p_{0,-1}(t)$ and $p_{-1,0}(t)$ are increasing, and $p_{0,1}(t)$ and $p_{1,0}(t)$ are decreasing.*

The financial significance of this proposition is clear. Because the investor must take the flat position at maturity, the investor is getting less likely to transfer from the flat position to the short or long position, and more likely to move from the short and long positions to the flat position.

When t is very small, we know the exact values of $p_{0,1}(t)$ and $p_{0,-1}(t)$.

Proposition 3.5 *There exists $t_1 \geqslant \frac{1}{\mu_1-\rho}\log\frac{1+K}{1-K}$ such that $p_{0,1}(t) = 1$ if $t \leqslant t_1$; and, there exists $t_0 \geqslant \frac{1}{\rho-\mu_2}\log\frac{1+K}{1-K}$ such that $p_{0,-1}(t) = 0$ if $t \leqslant t_0$.*

PROOF: We will only prove the first assertion as the proof of the other one is similar. Note that $u_1 = v_0 - v_1$ satisfies

$$\min\Big\{\partial_t u_1 - \mathcal{L}u_1,\ u_1 + (1+K)\Big\} = 0 \quad \text{in} \quad [p_0, 1] \times [0, T]. \tag{3.11}$$

Especially, at the right boundary $p = 1$, we have

$$\begin{cases} \partial_t u_1(1,t) - (\mu_1 - \rho)u_1(1,t) = -\lambda_1 \partial_p u_1(1,t), & \text{if } u_1(1,t) > -(1+K); \\ \partial_t u_1(1,t) - (\mu_1 - \rho)u_1(1,t) \geqslant -\lambda_1 \partial_p u_1(1,t), & \text{if } u_1(1,t) = -(1+K); \\ u_1(1,0) = -(1-K). \end{cases} \tag{3.12}$$

Define

$$\begin{cases} Z(t) = -(1-K)e^{(\mu_1-\rho)t}, \quad \text{if} \quad t < \frac{1}{\mu_1-\rho}\log\frac{1+K}{1-K}; \\ Z(t) = -(1+K), \quad \text{if} \quad t \geqslant \frac{1}{\mu_1-\rho}\log\frac{1+K}{1-K}. \end{cases}$$

Then

$$\begin{cases} \partial_t Z(t) - (\mu_1 - \rho)Z(t) = 0, \quad \text{if} \quad Z(t) > -(1+K); \\ \partial_t Z(t) - (\mu_1 - \rho)Z(t) \geqslant 0, \quad \text{if} \quad Z(t) = -(1+K); \\ Z(0) = -(1-K). \end{cases}$$

Since $\partial_p u_1(1,t) \leqslant 0$, we see that $Z(t)$ is a sub-solution of (3.12). Therefore, $u_1(1,t) \geqslant Z(t)$. In particular, when $t < \frac{1}{\mu_1-\rho}\log\frac{1+K}{1-K}$, we have $u_1(1,t) \geqslant Z(t) > -(1+K)$, which implies $p_{0,1}(t) = 1$. □

Proposition 3.6 *The free boundaries $p_{1,0}(t)$ and $p_{-1,0}(t)$ are continuous. Moreover, both their initial points are p_0, i.e.,*

$$p_{1,0}(0+) = p_0, \quad p_{-1,0}(0+) = p_0.$$

PROOF: We first prove the continuity property of $p_{1,0}(t)$. By contrary, suppose $p_{1,0}(t)$ is discontinuous at a point $t_0 \in (0, T]$, then by the decreasing property of $p_{1,0}(t)$, we have

$$p_{1,0}(t_0+) < p_{1,0}(t_0-) \leqslant p_0.$$

Then $u_1 = v_0 - v_1$ satisfies

$$u_1(p, t_0) = -(1-K), \quad \forall\, p \in (p_{1,0}(t_0+), p_{1,0}(t_0-)).$$

Using (1.1), we have

$$\Big(\partial_t u_1 - \mathcal{L} u_1\Big)(p, t_0) = 0, \quad p \in (p_{1,0}(t_0+), p_{1,0}(t_0-)).$$

So $\partial_t u_1(p, t_0) = \mathcal{L} u_1(p, t_0) = [(\mu_1 - \mu_2)p + \mu_2 - \rho][-(1-K)] > 0$ for $p \in (p_{1,0}(t_0+), p_{1,0}(t_0-)) \subset [0, p_0]$, which contradicts (3.5). Similarly, we can prove $p_{1,0}(0+) = p_0$ and the corresponding properties of $p_{-1,0}(t)$. □

This result is some surprising to us. It tells us when the time is close to the maturity (that is, $t \leqslant t_1$), the investor should *not* transfer from the flat position to the other positions at all. Such phenomenon has also been observed in Dai, et al. [1]. Both of them, we think, are due to the presence of transaction costs.

Proposition 3.7 *The free boundaries $p_{0,1}(t)$ and $p_{0,-1}(t)$ are continuous.*

PROOF: We just prove the continuity property of $p_{0,1}(t)$. If not, suppose t_0 was the discontinuous point, by the decreasing property of $p_{0,1}(t)$, we would have

$$p_{0,1}(t_0+) < p_{0,1}(t_0-),$$

thus, $u_1 = v_0 - v_1$ satisfies

$$u_1(p, t_0) = -(1+K), \quad \forall p \in (p_{0,1}(t_0+), p_{0,1}(t_0-)). \tag{3.13}$$

Let $\mathcal{D} := [p_{0,1}(t_0+), p_{0,1}(t_0-)] \times [t_0 - \varepsilon, t_0]$. Using (1.1), we have

$$\big(\partial_t u_1 - \mathcal{L} u_1\big)(p, t) = 0, \quad (p, t) \in \mathcal{D}.$$

Since $(p_{0,1}(t_0+), t_0)$ is the minimal point of u_1 in $\mathcal{D}$, by the strong maximum principle, (3.13) is impossible. □

Based on (3.3) and the monotonicity of the four free boundaries, we further have the no-overlapping property.

Proposition 3.8 *For the four free boundaries, we have,*

$$0 \leqslant p_{0,-1}(t) < p_{1,0}(t) < p_0 < p_{-1,0}(t) < p_{0,1}(t) \leqslant 1, \quad t \in (0,T].$$

Moreover, they are strictly monotone.

PROOF: We just prove the first strict inequality. Suppose there exits $t_0 \in (0,T]$ such that $p_{0,-1}(t_0) = p_{1,0}(t_0)$. Denote $c = p_{0,-1}(t_0) = p_{1,0}(t_0)$. Since $p_{0,-1}(t)$ is increasing, $p_{1,0}(t)$ is decreasing and $p_{0,-1}(t) \leqslant p_{1,0}(t)$, we have

$$p_{0,-1}(t) = p_{1,0}(t) = c, \quad \forall\, t \in [t_0, T].$$

Thus $u_1 = v_0 - v_1$ satisfies

$$u_1(c,t) = -(1-K), \quad \partial_p u_1(c,t) = 0, \quad \partial_t u_1(c,t) = 0, \quad \partial_{tp} u_1(c,t) = 0, \quad \forall\, t \in [t_0, T]. \tag{3.14}$$

By (3.5), $\partial_t u_1 \leqslant 0$, the above means $\partial_t u_1$ gets its maximum value 0 at (c,t) for any $t \in [t_0, T]$.

On the other hand, note that

$$\Big(\partial_t - \mathcal{L}\Big) u_1 = 0, \quad \forall\, (p,t) \in [c, p_0] \times [t_0, T],$$

differential it w.r.t. t we have

$$\Big(\partial_t - \mathcal{L}\Big) \partial_t u_1 = 0, \quad \forall\, (p,t) \in [c, p_0] \times [t_0, T].$$

Since $\partial_t u_1$ gets its maximum value at (c,t) for any $t \in [t_0, T]$, by the Hopf lemma, we have

$$\partial_{tp} u_1(c,t) < 0,$$

which contradicts (3.14). The remaining conclusions can be proved similarly. □

Proposition 3.9 *We have*

$$p_{0,-1}(t), \quad p_{1,0}(t), \quad p_{-1,0}(t), \quad p_{0,1}(t) \in C^\infty((0,T]).$$

PROOF: The proof of the smoothness is fairly long and technical, we refer interested reads to [4, 5]. □

Appendix A Financial background

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ be a fixed filtered complete probability space on which is defined a standard one-dimensional Brownian motion W. It represents the financial market. A stock is given in the market and its price process $S = (S_t)_{t\geqslant 0}$ satisfies the following stochastic differential equation (SDE):

$$\mathrm{d}S_t = \mu(I_t)S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t,$$

where the volatility $\sigma > 0$ is a known constant, but the market status process I_t and the noice process W_t are unobservable to an investor. The investor can only observe the stock price process, so his information $\mathcal{F}_t$ is equal to $\sigma\{S(s) : 0 \leqslant s \leqslant t\}$ augmented by all the $\mathbb{P}$-null sets and $\mathcal{F}_T \subseteq \mathcal{F}$.

To model the drift uncertainty, we assume $I_s \in \{1, 2\}$ is a two-state Markov chain. At each time s, $I_s = 1$ indicates a bull market status and $I_s = 2$ a bear market status. Suppose the transition probabilities are given by

$$\mathbb{P}(I_{s+h} = j \mid I_s = i) = \begin{cases} \lambda_i h + o(h), & j \neq i, \\ \\ 1 - \lambda_i h + o(h), & j = i, \end{cases} \qquad i, j = 1, 2,$$

as $h \to 0$. We assume I and W are independent processes and $\mu_1 > \mu_2$ where $\mu_i = \mu(i)$, $i = 1, 2$. If $\mu_1 = \mu_2$, then the market status uncertainty disappears and our financial market becomes the classical Black-Scholes market.

The investor is allowed to trade the stock at any time. But, at each time, the investor can only take one of the three positions: $\{-1, 0, 1\}$, where $i = 0$ corresponds to a flat position (no stock holding), $i = 1$ to a long position (holding one share of the stock), while $i = -1$ to a short position (short sale one share of the stock).

The investor's trading strategy beginning at time t is modeled by a sequence $(\tau_n, \iota_n)_{n\in\mathbb{Z}_+}$, where $t = \tau_0 < \tau_1 < \cdots$ is a strictly increasing sequence of $(\mathcal{F}_s)_{s\geqslant t}$-stopping times, and ι_n valued in $\{-1, 0, 1\}$, required to be $\mathcal{F}_{\tau_n}$ measurable, represents the position decided at τ_n until the next trading time. By misuse of notation, we denote by α^t_s the value of the position at any time s (begin with time t), namely,

$$\alpha^t_s = \iota_0 1_{\{t\leqslant s<\tau_1\}} + \sum_{n=1}^{\infty} \iota_n 1_{\{\tau_n\leqslant s<\tau_{n+1}\}}, \quad s \geqslant t.$$

Since $(\tau_n, \iota_n)_{n\in\mathbb{Z}_+}$ and α^t are one-to-one, we will not distinguish them and call them switching controls starting from time t.

Let $0 < K < 1$ denote the percentage of transaction fee. We denote by $g_{i,j}(S)$ the trading gain when switching from a position i to j, $i, j \in \{-1, 0, 1\}$, $j \neq i$, for a current stock price S. The switching gain functions are given by:

$$\begin{aligned} g_{0,1}(S) &= g_{-1,0}(S) = -S(1+K), \\ g_{0,-1}(S) &= g_{1,0}(S) = S(1-K). \end{aligned}$$

This means the investor needs to pay a proportional transaction costs when purchasing or selling the stock. For integrity, set

$$\begin{aligned} g_{0,0}(S) &= g_{1,1}(S) = g_{-1,-1}(S) = 0, \\ g_{-1,1}(S) &= g_{-1,0}(S) + g_{0,1}(S) = -2S(1+K), \\ g_{1,-1}(S) &= g_{1,0}(S) + g_{0,-1}(S) = 2S(1-K). \end{aligned}$$

By misuse of notation, we also set $g(S, i, j) = g_{i,j}(S)$.

At initial time t, given the stock price $S_t = S$, the reward functions of the decision sequences for a switching control α starting from time t, is given by

$$J(S, t, \alpha) = \mathbb{E}\left[\sum_{n=1}^{\infty} e^{-\rho(\tau_n - t)} g(S_{\tau_n}, \alpha_{\tau_n^-}, \alpha_{\tau_n}) 1_{\{\tau_n \leqslant T\}} + e^{-\rho(T-t)} g(S_T, \alpha_T, 0) \middle| S_t = S \right],$$

where $\rho > 0$ is a constant discount factor satisfying

$$\mu_2 < \rho < \mu_1.$$

If ρ is out of the above range, the problem admits trivial solutions only. The last term in the reword function means that at the maturity time T the net position must be flat.

Note that only the stock price S_s is observable to the investor at time s. The market status I_s is not directly observable. Thus, it is necessary to convert the problem into a completely observable one. One way to accomplish this is to use the Wonham filter [11]. Let $p_s = \mathbb{P}(I_s = 1 \mid \mathcal{F}_s)$. Then we can show (see [11]) that p_s satisfies the following SDE:

$$\mathrm{d}p_s = [-(\lambda_1 + \lambda_2)p_s + \lambda_2]\mathrm{d}s + (\mu_1 - \mu_2)p_s(1 - p_s)\sigma \mathrm{d}\nu_s,$$

where ν_s, given by

$$\mathrm{d}\nu_s = \frac{\mathrm{d}\log(S_s) - [(\mu_1 - \mu_2)p_s + \mu_2 - \sigma^2/2]\mathrm{d}s}{\sigma},$$

is called the innovation process and is a standard Brownian motion under the filtration $(\mathcal{F}_t)_{t \geqslant 0}$ (see, e.g., [9]).

Given $S_t = S$ and $p_t = p$, the investor's problem is to choose a switching control $(\alpha_s^t)_{s\geqslant t}$ starting from t, to maximize the reward function

$$J(S, p, t, \alpha^t)$$

subject to

$$\begin{cases} \mathrm{d}S_s = S_s[(\mu_1 - \mu_2)p_s + \mu_2]\mathrm{d}s + S_s\sigma\mathrm{d}\nu_s, \quad S_t = S, \\ \\ \mathrm{d}p_s = [-(\lambda_1 + \lambda_2)p_s + \lambda_2]\mathrm{d}s + (\mu_1 - \mu_2)p_s(1 - p_s)\sigma\mathrm{d}\nu_s, \quad p_t = p. \end{cases}$$

Indeed, this new problem is a completely observable one, because the conditional probabilities can be obtained using the stock price up to time s.

Let $V_i(S, p, t)$ denote the value function with the state (S, p, t) and net position $i \in \{-1, 0, 1\}$ at time t. That is,

$$V_i(S, p, t) = \sup_{\alpha\in\mathcal{A}_t^i} J(S, p, t, \alpha^t),$$

where $\mathcal{A}_t^i$ denotes the set of admissible switching controls α^t starting from t with the initial position $\iota_0 = i$.

Define an operator

$$\begin{aligned} \mathcal{T} = &\frac{1}{2}\Big(\frac{(\mu_1 - \mu_2)p(1 - p)}{\sigma}\Big)^2\partial_{pp} + \frac{1}{2}\sigma^2 S^2\partial_{SS} + S(\mu_1 - \mu_2)p(1 - p)\partial_{Sp} \\ &+ [-(\lambda_1 + \lambda_2)p + \lambda_2]\partial_p + S[(\mu_1 - \mu_2)p + \mu_2]\partial_S - \rho. \end{aligned}$$

The HJB equation associated with our optimal stopping time problem can be given formally as follows

$$\begin{cases} \min\Big\{ -\partial_t V_0 - \mathcal{T}V_0,\ V_0 - V_1 + S(1 + K),\ V_0 - V_{-1} - S(1 - K)\Big\} = 0, & \\ & 0 < p < 1, \\ \min\Big\{ -\partial_t V_1 - \mathcal{T}V_1,\ V_1 - V_0 - S(1 - K)\Big\} = 0, & \\ & 0 \leqslant t < T, \\ \min\Big\{ -\partial_t V_{-1} - \mathcal{T}V_{-1},\ V_{-1} - V_0 + S(1 + K)\Big\} = 0, & \end{cases}$$

with the terminal condition

$$\begin{cases} V_0(S, p, T) = 0, & \\ \\ V_1(S, p, T) = S(1 - K), & \qquad 0 < p < 1. \\ \\ V_{-1}(S, p, T) = -S(1 + K), & \end{cases}$$

It is easy to see from the model that the value functions $V_i(S,p,t),\ i=-1,0,1,$ are affine in S. This motivates us to adopt the following transformation $\widehat{V}_i(p,t)=V_i(S,p,t)/S$. Further, for convenience of discussion, we transform the terminal value problem to an initial value problem. Let $v_i(p,t)=\widehat{V}_i(p,T-t)$, then v_i satisfies the following variational inequality system

$$\begin{cases} \min\Big\{\partial_t v_0-\mathcal{L}v_0,\ v_0-v_1+(1+K),\ v_0-v_{-1}-(1-K)\Big\}=0, \\ \min\Big\{\partial_t v_1-\mathcal{L}v_1,\ v_1-v_0-(1-K)\Big\}=0, \qquad (p,t)\in\Omega=(0,1)\times[0,T), \\ \min\Big\{\partial_t v_{-1}-\mathcal{L}v_{-1},\ v_{-1}-v_0+(1+K)\Big\}=0, \end{cases}$$

with the initial condition

$$\begin{cases} v_0(p,0)=0, \\ v_1(p,0)=1-K, \qquad 0<p<1, \\ v_{-1}(p,0)=-(1+K), \end{cases}$$

where

$$\mathcal{L}=\frac{1}{2}\Big(\frac{(\mu_1-\mu_2)p(1-p)}{\sigma}\Big)^2\partial_{pp} \\ +[-(\lambda_1+\lambda_2)p+\lambda_2+(\mu_1-\mu_2)p(1-p)]\partial_p+(\mu_1-\mu_2)p+\mu_2-\rho.$$

This is the variational inequality system (1.1) investigated in this paper.

Now, we would like to provide an upper and a lower bounds for the value function. For the flat position, if we choose the strategy of not trading, the final reward is 0, which gives $V_0(S,p,t)\geqslant 0$. On the other hand, if we set $\mu_2=\mu_1$ and $\rho=0,\ K=0$, namely there is no difference in stock returns, no transaction cost or discount, then obviously, the optimal strategy for the investor in the flat position is to buy the stork immediately and hold it to maturity and then sell it, the corresponding reward is $S(e^{\mu_1(T-t)}-1)$, which gives an upper bound for $V_0(S,p,t)$. Moreover, according to the trading rules, we have

$$S(1-K)\leqslant V_1-V_0\leqslant S(1+K),\quad S(1-K)\leqslant V_0-V_{-1}\ \leqslant S(1+K),$$

so $|v_i(p,t)|=|V_i(S,p,t)|/S,\ i=0,1,-1,$ are bounded.

Appendix B The proof of the uniqueness

Suppose (v_0, v_1, v_{-1}) and $(\widehat{v}_0, \widehat{v}_1, \widehat{v}_{-1})$ are two bounded solutions of (1.1) with the initial condition (1.2). Let

$$\nu_i = e^{-\lambda t} v_i, \quad \widehat{\nu}_i = e^{-\lambda t}\widehat{v}_i, \quad i = 0, 1, -1,$$

with $\lambda = \mu_1 + \mu_2$. Then (ν_0, ν_1, ν_{-1}) and $(\widehat{\nu}_0, \widehat{\nu}_1, \widehat{\nu}_{-1})$ are solutions of

$$\min\Big\{(\partial_t - \mathcal{L} + \lambda)\nu_i,\ \nu_i - \max_{j\neq i}(\nu_j + h_{i,j})\Big\} = 0, \quad (p,t) \in \Omega, \quad i = -1, 0, 1,$$

with the initial condition (1.2), where

$$\begin{aligned}
h_{0,1} &= h_{-1,0} = -e^{-\lambda t}(1+K),\\
h_{1,0} &= h_{0,-1} = e^{-\lambda t}(1-K),\\
h_{-1,1} &= h_{-1,0} + h_{0,1} = -2e^{-\lambda t}(1+K),\\
h_{1,-1} &= h_{1,0} + h_{0,-1} = 2e^{-\lambda t}(1-K).
\end{aligned}$$

For any sufficiently small $\varepsilon > 0$, we come to prove

$$\nu_i \leqslant \widehat{\nu}_i + \phi_\varepsilon,\ i = 0, 1, -1, \tag{B.1}$$

where

$$\phi_\varepsilon(p) = \varepsilon - \varepsilon^2(\log p + \log(1-p)).$$

We argue by contradiction. Suppose (B.1) was not true, denote by (p^*, t^*) a maximum point of

$$\Phi := \max_{i\in\{0,1,-1\}}\{\nu_i - \widehat{\nu}_i - \phi_\varepsilon\},$$

then we have $\Phi(p^*, t^*) > 0$. Since $\lim\limits_{p\to 0+}\phi_\varepsilon(p) = \lim\limits_{p\to 1-}\phi_\varepsilon(p) = +\infty$ and $\Phi(p, 0) = -\phi_\varepsilon(p) < 0$, (p^*, t^*) is inside the domain Ω or at the upper boundary of Ω.

Note that

$$\Big|\frac{1}{2}\Big(\frac{(\mu_1-\mu_2)p(1-p)}{\sigma}\Big)^2\partial_{pp}\phi_\varepsilon\Big| \leqslant \varepsilon^2\Big(\frac{\mu_1-\mu_2}{\sigma}\Big)^2$$

and

$$\begin{aligned}
\lim_{p\to 0+}[-(\lambda_1+\lambda_2)p + \lambda_2 + (\mu_1-\mu_2)p(1-p)]\partial_p\phi_\varepsilon = -\infty,\\
\lim_{p\to 1-}[-(\lambda_1+\lambda_2)p + \lambda_2 + (\mu_1-\mu_2)p(1-p)]\partial_p\phi_\varepsilon = -\infty,
\end{aligned}$$

we have

$$(\partial_t - \mathcal{L} + \lambda)\phi_\varepsilon \geqslant 0.$$

Now, suppose $\Phi(p^*, t^*) = (\nu_i - \widehat{\nu}_i - \phi_\varepsilon)(p^*, t^*)$ for some $i \in \{0, 1, -1\}$. If $(\nu_i - \max_{j \neq i}(\nu_j + h_{i,j}))(p^*, t^*) > 0$, by the VI system of ν_i, we have $(\partial_t - \mathcal{L} + \lambda)\nu_i(p^*, t^*) = 0$ and thus $(\partial_t - \mathcal{L} + \lambda)(\nu_i - \widehat{\nu}_i - \phi_\varepsilon)(p^*, t^*) \leqslant 0$. On the other hand, since $\nu_i - \widehat{\nu}_i - \phi_\varepsilon$ takes its positive maximum value at the point (p^*, t^*), we have $(\partial_t - \mathcal{L} + \lambda)(\nu_i - \widehat{\nu}_i - \phi_\varepsilon)(p^*, t^*) > 0$, a contradiction. So there exists a $j \neq i$ such that $(\nu_i - \nu_j - h_{i,j})(p^*, t^*) = 0$. Since $\nu_i - \nu_j \geqslant \widehat{\nu}_i - \widehat{\nu}_j \geqslant h_{i,j}$, we conclude $(\widehat{\nu}_i - \widehat{\nu}_j - h_{i,j})(p^*, t^*) = 0$. Consequently, $(\nu_j - \widehat{\nu}_j - \phi_\varepsilon)(p^*, t^*) = (\nu_i - \widehat{\nu}_i - \phi_\varepsilon)(p^*, t^*) = \Phi(p^*, t^*)$, namely (p^*, t^*) is a maximum point of $\nu_j - \widehat{\nu}_j - \phi_\varepsilon$ as well. Again, if $(\nu_j - \max_{k \neq j}(\nu_k + h_{j,k}))(p^*, t^*) > 0$, by the VI system we have $(\partial_t - \mathcal{L} + \lambda)(\nu_j - \widehat{\nu}_j - \phi_\varepsilon)(p^*, t^*) \leqslant 0$ which contradicts that (p^*, t^*) is a maximum point of $\nu_j - \widehat{\nu}_j - \phi_\varepsilon$. So there exists a $k \neq j$ such that $(\nu_j - \nu_k - h_{j,k})(p^*, t^*) = 0$. But $h_{j,i} + h_{i,j} < 0$, so we conclude $k \neq i$. Repeating the above argument, (p^*, t^*) is a maximum point of $\nu_k - \widehat{\nu}_k - \phi_\varepsilon$. Note that $(\nu_k - \nu_j - h_{k,j})(p^*, t^*) = -(h_{j,k} + h_{k,j})(p^*, t^*) > 0$ and $(\nu_k - \nu_i - h_{k,i})(p^*, t^*) = -(h_{i,j} + h_{j,k} + h_{k,i})(p^*, t^*) > 0$, so by the VI system we have $(\partial_t - \mathcal{L} + \lambda)(\nu_i - \widehat{\nu}_i - \phi_\varepsilon)(p^*, t^*) \leqslant 0$, which contradicts that $\nu_k - \widehat{\nu}_k - \phi_\varepsilon$ takes its positive maximum value at (p^*, t^*). Hence we prove (B.1).

Letting $\varepsilon \to 0$ in (B.1), we get $\nu_i \leqslant \widehat{\nu}_i$, $i = 0, 1, -1$, which implies the uniqueness.

References

[1] M. Dai, Z. Xu, and X. Zhou. *Continuous-time markowitz's model with transaction costs.* SIAM J. Financ. Math., 1:96-125, 2010.

[2] M. Dai and F. Yi. *Finite horizon optimal investment with transaction costs: A parabolic double obstacle problem.* J. Differ. Equ., 246:1445-1469, 2009.

[3] M. Dai, Q. Zhang, and Zhu Q.J. *Optimal trend following trading rules.* Social Science Electronic Pulishing, 41(2), 2011.

[4] M. Dai, Q. Zhang, and Q.J. Zhu. *Trend following trading under a regime switching model.* SIAM J. Financ. Math., 1:780-810, 2010.

[5] A. Friedman. *Parabolic variational inequalities in one space dimension and smoothness of the free boundary.* J. Funct. Anal., 18:151-176, 1975.

[6] A.Q.M. Khaliq, B. Kleefeld, and R.H. Liu. *Solving complex pde systems for pricing american options with regime-switching by efficient exponential time differencing schemes.* Numer. Meth. Part Differ. Equ., 29(1):320-336, 2013.

[7] G.M. Lieberman. *Second Order Parabolic Differential Equations.* World Scientific, 1996.

[8] M.M. Ngo and H. Pham. *Optimal switching for the pairs trading rule: A viscosity solutions approach.* J. Math. Anal. Appl., 441:403-425, 2016.

[9] B. Øksendal. *Stochastic differential equations, 6td ed.* Springer-Verlag, Berlin, New York, 2003.

[10] O.A. Oleinik and E.V. Radkevic. *Second-order equations with nonnegative characteristic form.* American Mathematical Society Rhode Island and Plenum Press, New York, 1973.

[11] W.M. Wonham. *Some applications of stochastic differential equations to optimal nonlinear filtering.* SIAM J. Control Optim., 2:347-369, 1965.

[12] F. Yi. *American put option with regime-switching volatility (finite time horizon) variational inequality approach.* Math. Meth. Appl. Sci., 31:1461-1477, 2008.