A simple integral equation approach for optimal investment stopping problems with partial information

Jie Xing*, Jingtang Ma† and Harry Zheng ‡

July 8, 2024

Abstract

In this paper we study a finite horizon optimal investment stopping problem with unobservable random variable for the return of risky asset. Using the Bayesian filter and the dual control approach, we transform the original primal problem into a dual finite horizon optimal stopping problem, which results in the dual value function satisfying a variational inequality with two state variables. For a class of utility functions that include power utility and non-HARA utility, we show that the free boundary satisfies a Volterra type nonlinear integral equation with expectation over the joint distribution of the dual state process and the filtered probability process and we simplify and solve the integral equation with the dimension reduction and backward recursive methods. We also construct two simple closed form approximations for the free boundary using its asymptotic properties and show their accuracy and efficiency with numerical examples. Furthermore, we demonstrate that different model parameters may lead to one, or two, or no free boundaries with a simple example.

Mathematical subject classification 2020: 49L20, 90C46, 93E20

OR/MS subject classification: Finance/portfolio, finance/investment

Keywords: Optimal investment stopping problem, partial information, dual control, free boundary, integral equations, closed form approximation

1 Introduction

There has been extensive research in utility maximization of wealth and consumption, see Pham (2009) for excellent exposition in theory, methodology, and applications. Among many different model formulations, the mixed optimal control and optimal stopping problem (see Karatzas and Wang (2000)) is one of most difficult to solve and highly useful models with many applications, see for example Henderson and Hobson (2008) on asset sale, Yang and Koo (2018) on early retirement option. If the underlying state process follows a Markov process, the value function satisfies a

*School of Big Data Application and Economics, Guizhou University of Finance and Economics, Guiyang, 550025, P.R. China (Email: xj3553@mail.gufe.edu.cn).

†Co-corresponding author. School of Mathematics, Southwestern University of Finance and Economics, Chengdu, 611130, China (Email: mjt@swufe.edu.cn).

‡Co-corresponding author. Department of Mathematics, Imperial College, London SW7 2BZ, UK (Email: h.zheng@imperial.ac.uk).

variational inequality that is in general difficult to solve as one has to find the boundary of the stopping region, called the free boundary. The notion of viscosity solution, first defined by Crandall et al. (1992), provides a framework to study the existence and uniqueness of the continuous solution to the variational inequality, see Fleming and Soner (1993), Shreve and Soner (1994), Soner and Touzi (2002) and Zariphopoulou (1992). However, the viscosity solution approach in general does not provide information on the free boundary such as the smooth pasting condition that is critical in analyzing the obstacle problem. In a Black-Scholes market model with known deterministic coefficients and infinite investment horizon, the free boundary is a single point that may be determined with the continuity and smooth pasting properties of the value function, and the convex dual approach may be used to simplify the dynamic programming equation in the continuation region, see Choi and Shim (2006), Bensoussan et al. (2016), Koo et al. (2021). If the investment horizon is finite, the free boundary is an unknown function of time and it is much more difficult to determine, one may only derive some quantitative properties of the free boundary and the value function (continuity, monotonicity, etc.), see Jang et al. (2024), Jian et al. (2014), and Ma et al. (2019).

In practice, model coefficients are not deterministic nor observable. Ekström and Vaicenavicius (2016) point out that the incompleteness of information is inevitable as one needs a very long time series to estimate the drift which is rarely available for public offering stock. To circumvent this difficulty, one may use the filtering theory to estimate model coefficients based on observable market information. There are three special but important filters, they are Kalman-Bucy filter for linear diffusion, Wonham filter for finite state Markov chain, and Bayesian filter for random variable. All of them have been extensively studied in portfolio optimization, see, for example, Lakner (1998) and Papanicolaou (2019) for the linear diffusion model, Sass and Haussmann (2004) and Eksi and Ku (2017) for the continuous-time finite state Markov chain model, Bismuth et al. (2019) and Bodnar et al. (2017) for the random variable model. For optimal stopping with partial information in finance, we refer to Ekström and Vaicenavicius (2016) for optimal liquidation, De Angelis (2020) for optimal dividends, Xu and Yi (2019) for optimal strategy of stock loan, Décamps et al. (2005) and Klein (2009) for optimal investment timing, and Chen et al. (2021) for optimal retirement with investment and consumption.

This paper generalizes the results in Ma et al. (2019) for a one-dimensional complete market model to the framework of partial information setting, specifically, we study a general optimal investment and stopping problem in finite time horizon with unobservable random drift of risky asset. Using the Bayesian filter and the dual approach, we convert the primal mixed optimal control and stopping problem into a two-dimensional dual optimal stopping problem with the dual process having stochastic volatility, which leads to a variational inequality with two state variables as well as time variable. For a class of dual utility functions (see (3.1)), which include power utilities, and under some assumption (see Assumption 3.1), we prove the existence of a unique continuous two-dimensional free boundary that separates the whole space into two parts: the continuation region and the stopping region (see Propositions 3.2 and 3.4) and we show that there may be two or no free boundaries if the required assumption is not satisfied (see Example 4.2). The free boundary satisfies a Volterra type nonlinear integral equation (see (3.9)) that is highly difficult to solve as it involves the expectation over some joint distribution of the dual process and the filtered probability process, which is in general unknown and the techniques in Ma et al. (2019) are not applicable to the analysis of the free boundary. Thanks to the dimension reduction method, first introduced in Klein (2009), we simplify the nonlinear integral equation with expectation (see (3.9)) further to a nonlinear integral equation without expectation (see (3.19)) that can be solved easily using the iteration method in Detemple (2005). Furthermore, using the asymptotic property of the free boundary and linear interpolation, we construct two global closed-form approximations (see (3.27) and (3.28)) which are shown to be accurate by numerical examples.

We next compare our work with three papers in the literature that are closely related to ours in model formulation and methodology. The first one is Chen et al. (2021) that also uses the dual approach to solving the primal optimal investment, consumption and retirement problem with partial information. Some differences are minor, for example, the drift term in Chen et al. (2021) is a hidden continuous-time two state Markov chain process and estimated with the Wonham filter whereas in this work an unobservable two-state random variable and estimated with the Bayesian filter, and the properties of the dual value function and the dual free boundary in Chen et al. (2021) are derived for power utilities whereas in this work for a more general class of utility functions (see (3.1)). The key differences are that we characterize the free boundary with a numerically solvable nonlinear integral equation (see (3.19)) and also propose two global closed-form approximations (see (3.27) and (3.28)), none of which are discussed in Chen et al. (2021). On the other hand, Chen et al. (2021) have some detailed small-scale asymptotic analysis of the dual value function whereas in this work we do not have any asymptotic analysis as our focus is on the dual free boundary, not the dual value function that can be determined once the dual free boundary is known (see Proposition 3.9).

The second one is Klein (2009) that first introduces the dimension reduction method which we also use in this paper. The key differences are the following: Firstly, Klein (2009) solves an optimal stopping problem with drift uncertainty, whereas we have a mixed optimal investment and stopping problem with drift uncertainty, so Klein (2009) can not be applied to our problem directly. To tackle this difficulty, we introduce a dual process Y and convert the original problem into a dual optimal stopping problem with stochastic volatility (see (2.11)). Secondly, Klein (2009) uses the measure change to make the original state process to have complete information with constant drift, whereas under the same measure change, the dual process Y has incomplete information with unknown drift and volatility that are updated continuously, which is due to the stochastic volatility feature of the dual process Y. To overcome this technical difficulty, we combine the measure change and the dual control to introduce a new dual process $\mathcal{Y}$ that has complete information with constant drift and volatility (see (3.13)) and has a relation with the original dual process Y. Thirdly, Klein (2009) studies an infinite horizon optimal stopping problem, so the free boundary is a single point that can be determined in closed form with the smooth pasting condition, whereas our dual problem has a finite horizon which makes essentially impossible to find the free boundary in a closed form, we use the integral equation and the iteration method to determine the free boundary and also construct two global closed-form approximations, which is completely different from Klein (2009) in methodology.

The third one is Ma et al. (2019) that studies the optimal investment stopping problem in a Black-Scholes market model. The primal problem in that paper is converted into an equivalent dual optimal stopping problem with its free boundary satisfying a Volterra type nonlinear integral equation that can be calculated easily due to the geometric Brownian motion structure and a global closed-form formula of the approximate free boundary can also be derived. In sharp contrast to Ma et al. (2019), the unobservable drift coefficient in this paper requires different mathematical machinery due to the unknown free boundary in a three-dimensional space (see Proposition 3.2). We use the dual control approach to characterizing the equivalent dual value function and deriving the nonlinear integral equation for the free boundary (see Proposition 3.6). However, the expectation in the integral equation (3.9) can not be computed directly due to the unknown joint transition density of the two-dimensional dual state process. Therefore, the asymptotic analysis technique in Ma et al. (2019) is not applicable. Instead, we adopt the methodology proposed by Klein (2009) and modify it to reduce the dimension of our dual problem. The numerical examples show that our proposed method is accurate and efficient.

The rest of the paper is organized as follows. In Section 2, the model is formulated, and the primal problem is converted into a dual optimal stopping problem and the verification theorem is given. In Section 3, the main results of this paper are presented. In Section 4, the examples for power and non-HARA utility are studied. In Section 5, the numerical results are presented. Section 6 concludes. In Appendix A, the proofs of the main results and examples are collected.

2 Model formulation and dual approach

In this paper, we consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Here the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ satisfies the usual conditions and $\mathbb{P}$ denotes the probability measure. We assume μ is an $\mathcal{F}_0-$measurable random variable and $\{B_t\}_{t\geq 0}$ an $\mathcal{F}_t-$adapted standard Brownian motion, independent of μ. The market consists of one riskless saving account with constant interest rate $r > 0$ and one risky asset, whose price evolves according to

$$\mathrm{d}S_t = \mu S_t \mathrm{d}t + \sigma S_t \mathrm{d}B_t, \quad 0 \leq t \leq T,$$

where σ is a constant volatility rate. We assume that one can observe $\mathbb{F}^S = \{\mathcal{F}_t^S\}_{t\geq 0}$, the natural filtration generated by S, augmented with all $\mathbb{P}$-null sets, but can not observe directly μ and $\{B_t\}_{t\geq 0}$. Moreover, we assume that μ takes value μ_l or μ_h with $r < \mu_l < \mu_h$ and the agent's initial estimate of the probability of the event $\{\mu = \mu_h\}$ is a constant $p \in (0, 1)$.

Let $\{X_t\}_{t\geq 0}$ denote the wealth process and π_t the amount of wealth an investor holds in risky asset at time t. By self-financing condition, the investor's wealth X, starting with initial capital x, satisfies

$$\mathrm{d}X_t = rX_t \mathrm{d}t + \pi_t\big((\mu - r)\mathrm{d}t + \sigma \mathrm{d}B_t\big), \quad 0 \leq t \leq T.$$

We assume the portfolio process $\{\pi_t\}_{t\geq 0}$ is $\mathcal{F}_t^S-$progressively measurable and satisfies $\mathbb{E}\Big[\int_0^T |\pi_t|^2 \mathrm{d}t\Big] < \infty$. Denote the set of all $\mathcal{F}_t^S-$adapted stopping time taking values in $[t, T]$ by $\mathcal{T}_{t,T}$. The optimal investment stopping problem is given by

$$\sup_{\pi,\tau} \mathbb{E}\big[e^{-\beta\tau} U(X_\tau - K)\big], \tag{2.1}$$

where $\tau \in \mathcal{T}_{0,T}$, $\beta > 0$ is a discount factor, $K \geq 0$ the minimum wealth threshold value, and $U : (0, +\infty) \to \mathbb{R}$ a utility function that is twice continuously differentiable, increasing, strictly concave, satisfying $U(0) = 0$, $U(\infty) = \infty$, $U'(0) = \infty$, $U'(\infty) = 0$, $U(x) < C(1 + x^\gamma)$ for $x > 0$ and some constants $C > 0$ and $0 < \gamma < 1$.

The following notations are used throughout this paper: $Q_x := (0, T) \times (K, \infty) \times (0, 1)$, $Q_y := (0, T) \times \mathbb{R}_+ \times (0, 1)$, $Q_z := (0, T) \times \mathbb{R}_+ \times \mathbb{R}_+$, these are the domains of variables for primal and dual problems, $a \wedge b := \min\{a, b\}$, $\partial_x V, \partial_{xx} V$ are the first and second order partial derivatives of V with respect to x, similar notations are used for other partial derivatives.

Since one can only observe the asset price process S, but not its drift coefficient μ nor its driving Brownian motion process B, we first define an observable drift coefficient process $\{\mu_t\}_{t\geq 0}$ by the conditional expectation of μ, given the filtration $\mathbb{F}^S$, as

$$\mu_t = \mathbb{E}[\mu | \mathcal{F}_t^S],$$

and an observable probability process $\{P_t\}_{t\geq 0}$ by the conditional probability of $\mu = \mu_h$, given $\mathbb{F}^S = \{\mathcal{F}_t^S\}_{t\geq 0}$, as

$$P_t = \mathbb{P}(\mu = \mu_h | \mathcal{F}_t^S). \tag{2.2}$$

Clearly, $\mu_t = \mu_h P_t + \mu_l(1-P_t)$. Finally, define the innovation process $\{\tilde{B}_t\}_{t\geq 0}$ by

$$\mathrm{d}\tilde{B}_t = \sigma^{-1}(\mu - \mu_t)\mathrm{d}t + \mathrm{d}B_t.$$

Then $\{\tilde{B}_t\}_{t\geq 0}$ is an $\mathcal{F}_t^S$ standard Brownian motion, see Bain and Crisan (2009, Chapter 3). The process $\{P_t\}_{t\geq 0}$ satisfies the following SDE

$$\mathrm{d}P_t = \Theta P_t(1-P_t)\mathrm{d}\tilde{B}_t, \quad \text{with} \quad P_0 = p, \tag{2.3}$$

where $\Theta = \frac{\mu_h - \mu_l}{\sigma}$, see Décamps et al. (2005). We can rewrite the asset price process S as

$$\mathrm{d}S_t = \big[(\mu_h - \mu_l)P_t + \mu_l\big]S_t\mathrm{d}t + \sigma S_t \mathrm{d}\tilde{B}_t,$$

with observable drift coefficient process and Brownian motion process. The wealth process X then satisfies

$$\mathrm{d}X_t \quad = \quad \big(\pi_t\vartheta(P_t)\sigma + rX_t\big)\mathrm{d}t + \sigma\pi_t\mathrm{d}\tilde{B}_t, \tag{2.4}$$

where $\vartheta(p) := p\theta_h + (1-p)\theta_l$ and $\theta_h := \frac{\mu_h - r}{\sigma}$ and $\theta_l := \frac{\mu_l - r}{\sigma}$.

Solving problem (2.1), regarded as an optimal investment stopping problem of the Markovian processes (2.3) and (2.4), reduces to determining

$$V(t,x,p) = \sup_{\pi,\tau}\mathbb{E}\big[e^{-\beta(\tau-t)}U(X_\tau^{t,x,p} - K)\big]. \tag{2.5}$$

Applying the dynamic programming principle, in the region Q_x the value function satisfies

$$\min\Big\{\big(-\partial_t V - \sup_{\pi}\mathcal{L}_{X,P}^{\pi}V + \beta V\big)(t,x,p), V(t,x,p) - U(x-K)\Big\} = 0 \tag{2.6}$$

with terminal condition $V(T,x,p) = U(x-K)$, where $\mathcal{L}_{X,P}^{\pi}$ is the infinitesimal generator of (X,P), given by

$$\mathcal{L}_{X,P}^{\pi}V = (\pi\vartheta(p)\sigma + rx)\partial_x V + \frac{1}{2}\sigma^2\pi^2\partial xxV + \frac{1}{2}\Theta^2 p^2(1-p)^2\partial_{pp}V + \Theta p(1-p)\sigma\pi\partial_{xp}V.$$

It is in general highly difficult to solve the variational equation (2.6) as one needs to find the free boundary as well as to solve a nonlinear PDE with two state variables.

We now use the duality approach to studying the primal problem (2.5). We first introduce the state price density process $\mathcal{H}_s := e^{-r(s-t)}M_s$ with

$$M_s := \exp\Big\{-\int_t^s \vartheta(P_u)\mathrm{d}\tilde{B}_u - \frac{1}{2}\int_t^s \vartheta^2(P_u)\mathrm{d}u\Big\}, \quad s \geq t.$$

Using Itô's lemma and the fact that $\tilde{B}$ is a standard Brownian motion under measure $\mathbb{P}$, we get

$$\mathrm{d}(\mathcal{H}_s X_s) = e^{-r(s-t)}M_s\,(\sigma\pi_s - \vartheta(P_s)X_s)\,\mathrm{d}\tilde{B}_s, \quad s \geq t,$$

which means the process $\{\mathcal{H}_s X_s\}_{s\geq t}$ is a positive local martingale, hence a super-martingale. By the optional sampling theorem, we have the following budget constraint

$$\mathbb{E}[\mathcal{H}_\tau X_\tau] \leq x, \tag{2.7}$$

for any bounded stopping time $\tau \geq t$. The dual function of $U(\cdot - K)$ is given by, for $y > 0$,

$$\tilde{U}_K(y) := \sup_{x>K}[U(x-K) - xy] = \sup_{x>0}[U(x) - xy] - Ky.$$

It is easy to check that

$$-Ky \leq \tilde{U}_K(y) \leq C + Cy^{\frac{\gamma}{\gamma-1}} - Ky \tag{2.8}$$

for some positive constant C. Thanks to (2.7), we have for any Lagrange multiplier $y > 0$,

$$\begin{aligned} V(t,x,p) &\leq \sup_{\pi,\tau} \Big\{ \mathbb{E}[e^{-\beta(\tau-t)}U(X_\tau - K)] + y(x - \mathbb{E}[H_\tau X_\tau]) \Big\} \\ &\leq \sup_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}[e^{-\beta(\tau-t)}\tilde{U}_K(Y_\tau^{t,y,p})] + xy, \end{aligned} \tag{2.9}$$

where $Y_s := ye^{\beta(\tau-t)}\mathcal{H}_\tau$. Itô's lemma implies that the dual process Y satisfies the SDE:

$$\mathrm{d}Y_s = (\beta - r)Y_s\mathrm{d}t - \vartheta(P_s)Y_s\mathrm{d}\tilde{B}_s, \quad t \leq s \leq T, \quad \text{with} \quad Y_t = y. \tag{2.10}$$

The dual problem is given by the following optimal stopping problem:

$$\tilde{V}(t,y,p) = \sup_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}[e^{-\beta(\tau-t)}\tilde{U}_K(Y_\tau^{t,y,p})]. \tag{2.11}$$

The dynamic programming principle gives that the dual value function satisfies

$$\min\big\{ -\partial_t\tilde{V} - \mathcal{L}_{Y,P}\tilde{V} + \beta\tilde{V}, \tilde{V} - \tilde{U}_K \big\} = 0, \quad (t,y,p) \in Q_y, \tag{2.12}$$

$$\tilde{V}(T,y,p) = \tilde{U}_K(y), \quad (y,p) \in \mathbb{R}_+ \times (0,1), \tag{2.13}$$

where $\mathcal{L}_{Y,P}$ is the infinitesimal generator of (Y,P) given by

$$\mathcal{L}_{Y,P}\tilde{V} = \frac{1}{2}\vartheta^2(p)y^2\partial_{yy}\tilde{V} + \frac{1}{2}\Theta^2p^2(1-p)^2\partial_{pp}\tilde{V} - \Theta p(1-p)\vartheta(p)y\partial_{yp}\tilde{V} + (\beta - r)y\partial_y\tilde{V}.$$

Compared with (2.6), the variational equation (2.12) is simpler as one only needs to solve a linear PDE, not a nonlinear PDE as in (2.6), in the continuous region, but one still needs to find the free boundary with two state variables, a highly difficult problem which we will address in the next section.

We introduce the continuation region and stopping region of the dual problem as follows

$$\mathcal{C}_y = \big\{(t,y,p) \in Q_y : \tilde{V}(t,y,p) > \tilde{U}_K(y)\big\},$$
$$\mathcal{S}_y = \big\{(t,y,p) \in Q_y : \tilde{V}(t,y,p) = \tilde{U}_K(y)\big\}.$$

Since $\mathcal{C}_y$ is determined once $\mathcal{S}_y$ is known, we will focus on finding the stopping region $\mathcal{S}_y$ and will not write $\mathcal{C}_y$ and continuous regions for other related problems in the rest of the paper.

Since $\vartheta(p)$ is bounded, we find that for any $\bar{\alpha} \in \mathbb{R}$

$$\mathbb{E}\Big[\sup_{0\leq s\leq T} \big(Y_s^{y,p}\big)^{\bar{\alpha}}\Big] \leq C, \tag{2.14}$$

where C is a constant depending on $y, T, \bar{\alpha}$ and independent of p (e.g. see Jacka and Ocejo (2018, Appendix A)). By (2.8) and (2.14), the standard optimal stopping theory (see Peskir and Shiryaev (2006, Corollary 2.9)) yields that the optimal stopping time is given by

$$\tau^*(t,y,p) := \inf\big\{s \geq t : (s, Y_s^{t,y,p}, P_s^{t,p}) \in \mathcal{S}_y\big\} \wedge T. \tag{2.15}$$

In the end of this section, we give a verification theorem which shows that the primal value function V, the optimal stopping time τ^*, the optimal control π^*, and the optimal wealth X^* can all be recovered from those of the dual counterpart and that we only need to focus on solving the dual problem, which is what we will do in the rest of the paper.

Proposition 2.1. *Let $\tilde{V}$ be a function satisfying the variational inequality* (2.12) *with terminal condition* (2.13) *such that*

(i) $\tilde{V} \in C^1(Q_y) \cap C(\bar{Q}_y)$*;*

(ii) $\tilde{V} \in C^{1,2,2}(Q_y \backslash \partial\mathcal{C}_y)$ *with locally bounded derivatives near* $\partial\mathcal{C}_y$*;*

(iii) $\left|\tilde{V}(t,y,p)\right| \leq C\Big(y^q + 1\Big)$ *for some constant* $q < 0$*;*

(iv) $\tilde{V}(t,\cdot,p)$ *is strictly convex;*

(v) $-\partial_y\tilde{V}(t,y,p) \to +\infty$ *as* $y \to 0$ *and* $-\partial_y\tilde{V}(t,y,p) \to \hat{K} \leq K$ *for some positive constant* $\hat{K}$*.*

Then the value function for problem (2.5) *is given by*

$$V(t,x,p) = \inf_{y>0}\left[\tilde{V}(t,y,p) + xy\right], \quad (t,x,p) \in Q_x.$$

Moreover, for $s \geq t$, the optimal stopping time, the optimal portfolio strategy and optimal wealth are respectively given by

$$\begin{aligned}
\tau^* &= \inf\left\{s \geq t : (s, Y_s, P_s) \in \mathcal{S}_y\right\} \wedge T, \\
\pi_s^* &= \frac{\vartheta(P_s)Y_s\partial_{yy}\tilde{V}(s,Y_s,P_s) - \Theta P_s(1-P_s)\partial_{yp}\tilde{V}(s,Y_s,P_s)}{\sigma}\mathbb{1}_{\{t\leq s\leq \tau^*\}}, \\
X_s^* &= -\partial_y\tilde{V}(s,Y_s,P_s),
\end{aligned}$$

with $Y_t = y_*$ and $P_t = p$, where y_* is the unique solution to the equation $\partial_y\tilde{V}(t,y,p) + x = 0$ with $X_t^* = x$ and $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function of some set.

Proof. The proof of the verification theorem is standard. For details, see e.g. Jang et al. (2024, Theorem 14). □

3 Main results

In the rest of this paper, analogous to Ma et al. (2019), we consider the dual utility function of the following form

$$\tilde{U}_K(y) = \sum_{j=1}^{J} -\frac{1}{q_j}y^{q_j} - Ky, \tag{3.1}$$

where $q_1 < q_2 < \cdots < q_J < 0$. Note that $\tilde{U}_0(y)$ is the dual function of the power utility $U(x) = \frac{1}{\gamma}x^\gamma$, $\gamma \in (0,1)$, with $J = 1$, $q_1 = \frac{\gamma}{\gamma-1}$, and that of the non-HARA utility

$$U(x) = \frac{1}{3}H^{-3}(x) + H^{-1}(x) + xH(x), \tag{3.2}$$

where $H(x) = (\frac{2}{-1+\sqrt{1+4x}})^{1/2}$, with $J = 2$, $q_1 = -3$, $q_2 = -1$. Therefore, (3.1) covers a broad class of utility functions. We introduce an important function by

$$\phi(y,p) := \sum_{j=1}^{J} A_j(p) y^{q_j} - rKy, \tag{3.3}$$

where

$$A_j(p) := \frac{1}{2}\vartheta^2(p)(q_j - 1) + \Big(1 - \frac{1}{q_j}\Big)\beta - r, \quad j = 1, 2, \ldots, J.$$

Note that $A_1 < A_2 < \cdots < A_J$. We assume the following assumption holds in the rest of this section.

Assumption 3.1. *The parameters of the model satisfy $K > 0$ and $A_1(1) > 0$.*

3.1 The free boundary

To study the optimal stopping problem, the following monotonicity of $\tilde{V}$ is critical for subsequent analysis.

Proposition 3.1. *The function $y \mapsto \tilde{V}(t,y,p)$ is strictly convex. The functions $t \mapsto \tilde{V}(t,y,p)$ and $y \mapsto \tilde{V}(t,y,p)$ are decreasing, and $p \mapsto \tilde{V}(t,y,p)$ is increasing. Furthermore, if Assumption 3.1 holds, then the function $y \mapsto \tilde{V}(t,y,p) - \tilde{U}_K(y)$ is increasing.*

Proof. See Appendix A.1. □

For the shape of the stopping region $\mathcal{S}_y$, we have the following result.

Proposition 3.2. *There exists a unique free boundary b defined by*

$$b(t,p) := \sup\big\{y \in \mathbb{R}_+ : \tilde{V}(t,y,p) = \tilde{U}_K(y)\big\} \tag{3.4}$$

such that

$$\mathcal{S}_y = \big\{(t,y,p) \in Q_y : y \le b(t,p)\big\}. \tag{3.5}$$

Moreover, the functions $t \mapsto b(t,p)$ is increasing and $p \mapsto b(t,p)$ is decreasing.

Proof. See Appendix A.2. □

We next show that $\tilde{V}$ is continuous differentiable across the free boundary b.

Proposition 3.3. *For any $(t,p) \in (0,T) \times (0,1)$, we have*

$$\begin{aligned}
&\partial_y \tilde{V}\big(t, b(t,p), p\big) = \tilde{U}_K'\big(b(t,p)\big),\\
&\partial_p \tilde{V}\big(t, b(t,p), p\big) = 0,\\
&\partial_t \tilde{V}\big(t, b(t,p), p\big) = 0.
\end{aligned}$$

Furthermore, the function $\tilde{V}$ is C^1 in Q_y and the following limits hold

$$\lim_{y\to 0} -\partial_y \tilde{V}(t,y,p) = +\infty, \quad \lim_{y\to\infty} -\partial_y \tilde{V}(t,y,p) := \hat{K} \le K \tag{3.6}$$

for any fixed $(t,p) \in [0,T] \times (0,1)$, where $\hat{K}$ is a non-negative constant.

Proof. The C^1 regularity of $\tilde{V}$ can be derived by penalty method, we refer to Chen et al. (2021, Theorem 2) for details. (3.6) follows from the same proof of Ma et al. (2019, Corollary 2.4). $\square$

Proposition 3.4. *The free boundary $b(t,p)$ defined by (3.4) is continuous for any $(t,p) \in (0,T) \times (0,1)$.*

Proof. See Appendix A.3. $\square$

To calculate the free boundary b, it is important to find the value of b near the terminal time T. The next proposition shows such a result.

Proposition 3.5. *The free boundary b defined by (3.4) satisfies*

$$\lim_{t \uparrow T} b(t,p) = \ell(p), \tag{3.7}$$

where $\ell(p)$ is the unique solution to the equation $\phi(y,p) = 0$ for any $p \in (0,1)$.

Proof. See Appendix A.4. $\square$

The free boundary can be characterized by a Volterra type nonlinear integral equation, see Detemple (2005, Theorem 70).

Proposition 3.6. *The value function $\tilde{V}$ has the following representation*

$$\tilde{V}(t,y,p) = -\mathbb{E}\Big[\int_t^T e^{-\beta(s-t)}\phi(Y_s^{t,y,p}, P_s^{t,p})\mathbb{1}_{\{Y_s^{t,y,p} > b(s,P_s^{t,p})\}}\mathrm{d}s\Big] + \tilde{U}_K(y). \tag{3.8}$$

Moreover, the free boundary $b(t,p)$ satisfies the following Volterra type non-linear integral equation:

$$\mathbb{E}\Big[\int_t^T e^{-\beta(s-t)}\phi(Y_s^{t,b(t,p),p}, P_s^{t,p})\mathbb{1}_{\{Y_s^{t,b(t,p),p} > b(s,P_s^{t,p})\}}\mathrm{d}s\Big] = 0 \tag{3.9}$$

with terminal condition given by Proposition 3.5.

3.2 Dimension reduction

As the law of the process (Y,P) is not available in general, we are not able to calculate the expectation in (3.9) directly. It is difficult to solve the integral equation (3.9) with the traditional backward recursion method proposed in Detemple (2005). Following Klein (2009), we first reduce the dimension of equation (3.9) and then apply a simple backward recursive method to solve it. The critical technique used in this procedure is measure change. To this end, we define a likelihood ratio process $\{\Phi_t\}_{t\geq 0}$ by

$$\Phi_t := \frac{P_t}{1-P_t}, \quad t \geq 0. \tag{3.10}$$

Furthermore, introduce a new process W by

$$\mathrm{d}W_t := \Theta P_t \mathrm{d}t + \mathrm{d}\tilde{B}_t,$$

and a new probability measure by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t^S} := \exp\Big(-\frac{1}{2}\int_0^t \Theta^2 P_u^2 \mathrm{d}u - \int_0^t \Theta P_u \mathrm{d}\tilde{B}_u\Big)$$

$$= \quad \exp\Big(\frac{1}{2}\int_0^t \Theta^2 P_u^2 \mathrm{d}u - \int_0^t \Theta P_u \mathrm{d}W_u\Big). \tag{3.11}$$

Girsanov's theorem gives that W is a standard Brownian motion under $\mathbb{Q}$. Then the likelihood ratio process satisfies

$$\mathrm{d}\Phi_t = \Theta\Phi_t \mathrm{d}W_t, \quad \text{with} \quad \Phi_0 = \varphi := \frac{p}{1-p}.$$

If we let $F_t := \frac{1+\Phi_t}{1+\varphi}$, then Itô's lemma yields

$$\mathrm{d}F_t = \Theta P_t F_t \mathrm{d}W_t, \quad \text{with} \quad F_0 = 1.$$

By (3.11), it turns out that

$$F_t = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\Big|_{\mathcal{F}_t^S}. \tag{3.12}$$

Under the measure $\mathbb{Q}$, the wealth process X satisfies

$$\mathrm{d}X_s = (rX_s + \pi_s\theta_l\sigma)\mathrm{d}s + \sigma\pi_s \mathrm{d}W_s, \quad s \geq t,$$

which implies that the corresponding dual process, denoted by $\mathcal{Y}$, should satisfy

$$\mathrm{d}\mathcal{Y}_s = (\beta - r)\mathcal{Y}_s \mathrm{d}s - \theta_l \mathcal{Y}_s \mathrm{d}W_s, \quad s \geq t \tag{3.13}$$

and the corresponding budget constraint becomes

$$\mathbb{E}^{\mathbb{Q}}[\tilde{\mathcal{H}}_\tau X_\tau] \leq x,$$

where $\tilde{\mathcal{H}}_s := e^{-\beta(s-t)}\mathcal{Y}_s^{t,1}$. For any $y > 0$, we have

$$\begin{aligned} V(t,x,p) &= \sup_{\pi,\tau} \mathbb{E}^{\mathbb{Q}}[e^{-\beta(\tau-t)}F_\tau U(X_\tau - K)] \\ &\leq \sup_{\pi,\tau}\Big\{\mathbb{E}^{\mathbb{Q}}[e^{-\beta(\tau-t)}F_\tau U(X_\tau - K)] + y(x - \mathbb{E}^{\mathbb{Q}}[\tilde{\mathcal{H}}_\tau X_\tau])\Big\} \\ &\leq \sup_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}[e^{-\beta(\tau-t)}F_\tau \tilde{U}_K(\mathcal{Y}_\tau/F_\tau)] + xy. \end{aligned} \tag{3.14}$$

Inspired by (2.9) and (3.14), we apply Itô's lemma to FY and deduce that

$$\mathrm{d}(F_sY_s) = (\beta - r)F_sY_s\mathrm{d}s - \theta_l F_sY_s\mathrm{d}W_s, \quad F_tY_t = y.$$

Comparing with (3.13), we derive the relation between Y under $\mathbb{P}$ and $\mathcal{Y}$ under $\mathbb{Q}$ as $\mathcal{Y}_t = F_tY_t$. Using the observation above, we can rewrite the dual problem (2.11) under measure $\mathbb{Q}$ as follows.

Proposition 3.7. *Let $\mathbb{Q}$ be a probability measure given by* (3.11)*. Then the dual value function $\tilde{V}$ has the following representation*

$$\tilde{V}(t,y,p) = \frac{1}{1+\varphi}\sup_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}\Big[\int_t^\tau e^{-\beta(s-t)}\psi\Big(s,z,\Phi_s^{t,\varphi}\Big)\mathrm{d}s\Big] + \tilde{U}_K(y),$$

where

$$\varphi = \frac{p}{1-p}, \quad z = \varphi^{-\varpi}(1+\varphi)ye^{-\varsigma t}, \quad \psi(s,z,\Phi) = -(1+\Phi)\phi\Big(\frac{ze^{\varsigma s}\Phi^{\varpi}}{1+\Phi}, \frac{\Phi}{1+\Phi}\Big)$$

and $\varsigma = \beta - r - \frac{1}{2}\theta_l^2 - \frac{1}{2}\Theta\theta_l$ *and* $\varpi = -\frac{\theta_l}{\Theta} < 0$.

Proof. See Appendix A.5. □

Let $\tilde{\mathcal{V}}(t,z,\varphi) := (1+\varphi)(\tilde{V}(t,y,p) - \tilde{U}_K(y))$. Thanks to Proposition 3.7, we study the following optimal stopping problem

$$\tilde{\mathcal{V}}(t,z,\varphi) \quad = \quad \sup_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}\Big[\int_t^\tau e^{-\beta(s-t)}\psi\Big(s,z,\Phi_s\Big)\mathrm{d}s\Big], \tag{3.15}$$

By dynamic programming principle, $\tilde{\mathcal{V}}$ satisfies

$$\min\big\{(-\partial_t\tilde{\mathcal{V}} - \frac{1}{2}\Theta^2\varphi^2\partial_{\varphi\varphi}\tilde{\mathcal{V}} + \beta\tilde{\mathcal{V}})(t,z,\varphi) - \psi(t,z,\varphi), \tilde{\mathcal{V}}(t,z,\varphi)\big\} = 0 \quad \text{in} \quad Q_z, \tag{3.16}$$

with terminal condition $\tilde{\mathcal{V}}(T,z,\varphi) = 0$. Note that z can be viewed as a parameter other than a variable. In this sense we have reduced the dimension of the original dual problem (2.11). We continue the study of problem (3.15) by introducing the stopping region as follows

$$\mathcal{S}_z := \{(t,z,\varphi) \in Q_z : \tilde{\mathcal{V}}(t,z,\varphi) = 0\}.$$

In order to study the shape of $\mathcal{S}_z$, we state the following proposition.

Proposition 3.8. *There exists a positive function $(t,z) \mapsto \hat{b}(t,z)$ such that the stopping region with respect to problem* (3.15) *can be represented as*

$$\mathcal{S}_z = \{(t,z,\varphi) \in Q_z : \varphi \geq \hat{b}(t,z)\}, \tag{3.17}$$

with the function $z \mapsto \hat{b}(t,z)$ being strictly increasing. Moreover,

$$b(t,p) = \frac{p^{\varpi}}{(1-p)^{\varpi-1}} e^{\varsigma t}\hat{b}^{-1}\Big(t, \frac{p}{1-p}\Big), \tag{3.18}$$

where $\hat{b}^{-1}(t,\varphi)$ denotes the inverse function of $\hat{b}(t,\cdot)$ for any fixed $t \in (0,T)$.

Proof. See Appendix A.6. □

Using the shape of the continuous region and measure change, we can derive the nonlinear Volterra type integral equation with respect to the free boundary $\hat{b}(t,z)$.

Proposition 3.9. *The function $\tilde{\mathcal{V}}$ has the following representation*

$$\tilde{\mathcal{V}}(t,z,\varphi) = \int_t^T e^{-\beta(s-t)} \int_{-\infty}^{d(s-t,\hat{b}(s,z),\varphi)} \psi(s,z,\varphi e^{-\frac{1}{2}\Theta^2(s-t)+\Theta\sqrt{s-t}\eta})n(\eta)\mathrm{d}\eta\mathrm{d}s.$$

Moreover, the free boundary $\hat{b}$ satisfies the following nonlinear integral equation:

$$\int_t^T e^{-\beta(s-t)} \int_{-\infty}^{d(s-t,\hat{b}(s,z),\hat{b}(t,z))} \psi(s,z,\hat{b}(t,z)e^{-\frac{1}{2}\Theta^2(s-t)+\Theta\sqrt{s-t}\eta})n(\eta)\mathrm{d}\eta\mathrm{d}s = 0 \tag{3.19}$$

where $n(\eta) := \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\eta^2}$ and $d(s,\hat{\varphi},\varphi) := \frac{\log(\hat{\varphi}/\varphi)+\frac{1}{2}\Theta^2 s}{\Theta\sqrt{s}}$. The terminal condition is given by

$$\lim_{t\uparrow T}\hat{b}(t,z_0) = \frac{p_0}{1-p_0} \tag{3.20}$$

for any fixed $z_0 > 0$, where p_0 is the unique solution to

$$z_0 = \ell(p)\frac{(1-p)^{\varpi-1}}{p^{\varpi}}e^{-\varsigma T}.$$

Proof. See Appendix A.7. □

3.3 Global closed form approximation

Using the dimension reduction method, the free boundary $\hat{b}$ is characterized by the integral equation (3.19) that can be solved numerically with the backward recursive method, which is accurate but time consuming. We now establish a simple global closed-form approximation of the free boundary and will compare their performance with some numerical tests in Section 5. To this end, for any $p \in [0,1]$, we consider the following optimal stopping problem

$$\tilde{\mathscr{V}}(t,y;p) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\big[e^{-\beta(\tau-t)}\tilde{U}_K(\tilde{Y}_\tau^{t,y})\big], \tag{3.21}$$

where the dynamic of $\tilde{Y}$ follows

$$\mathrm{d}\tilde{Y}_s = (\beta - r)\tilde{Y}_s \mathrm{d}s - \vartheta(p)\tilde{Y}_s \mathrm{d}\tilde{B}_s, \quad s \geq t, \quad \tilde{Y}_t = y,$$

with $\vartheta(p) = \Theta p + \theta_l$.

For any fixed $p \in [0,1]$, define the stopping region of problem (3.21) by

$$\tilde{\mathscr{S}}_y^p = \big\{(t,y) \in [0,T) \times (0,+\infty) : \tilde{\mathscr{V}}(t,y;p) = \tilde{U}_K(y)\big\}.$$

The properties of the free boundary associated with the optimal stopping problem (3.21) are summarized in the next result, its proof is similar to Ma et al. (2019, Theorem 3.7, Theorem 3.9) and therefore omitted.

Proposition 3.10. *Let $\kappa(p) := (2r - 2\beta + \vartheta^2(p))/\vartheta^2(p)$, $\lambda(p) := 2\beta/\vartheta^2(p)$. There exists a positive function $\tilde{b}$ such that*

$$\tilde{\mathscr{S}}_y^p = \big\{(t,y) \in [0,T) \times (0,+\infty) : y \leq \tilde{b}(t;p)\big\}. \tag{3.22}$$

The free boundary $\tilde{b}$ is increasing in t with limits

$$\lim_{t \uparrow T} \tilde{b}(t;p) = \ell(p), \tag{3.23}$$

and

$$\lim_{T-t \to \infty} \tilde{b}(t;p) = \hat{\ell}(p), \tag{3.24}$$

where $\hat{\ell}(p)$ is the unique solution to

$$\sum_{j=1}^{J} -\frac{1}{q_j}\Big[q_j - \frac{1}{2}\big(\kappa(p) - \sqrt{\kappa^2(p) + 4\lambda(p)}\big)\Big] y^{q_j - 1} - K\Big[1 - \frac{1}{2}\big(\kappa(p) - \sqrt{\kappa^2(p) + 4\lambda(p)}\big)\Big] = 0.$$

Furthermore, as $t \to T$, the free boundary $\tilde{b}(t;p)$ satisfies

$$\lim_{t \uparrow T} \frac{\ell(p) - \tilde{b}(t;p)}{\vartheta(p)\ell(p)\sqrt{2(T-t)}} = A, \tag{3.25}$$

where A is the unique solution to

$$\frac{1}{2}e^{-x^2} - \frac{\sqrt{\pi}}{2}x + x^2 \int_0^1 e^{-x^2\eta^2} \frac{3\eta^2 + \eta^4}{(1+\eta^2)^2} \mathrm{d}\eta = 0.$$

Using Proposition 3.10, we have the following asymptotic result of b at $p=0$ and $p=1$.

Proposition 3.11. *The free boundary b defined by* (3.4) *satisfies*

$$\lim_{p\downarrow 0} b(t,p) = \tilde{b}(t;0), \quad \lim_{p\uparrow 1} b(t,p) = \tilde{b}(t;1), \tag{3.26}$$

where $\tilde{b}$ is defined by (3.22).

Proof. See Appendix A.8. □

Propositions 3.5, 3.10 and 3.11 show that free boundaries b and $\tilde{b}$ have the same limits at $t=T$ and $p=0,1$. We know, for fixed p, there is a simple and accurate global closed-form approximation for $\tilde{b}$ (see Ma et al. (2019)), we therefore suggest to use that approximation for the free boundary b too, albeit P is now a process, not a constant.

Inspired by (3.7), (3.23) - (3.26), we construct an approximation of the form

$$b(t,p) \approx b_I(t,p) := \ell(p) - (\ell(p) - \hat{\ell}(p))\sqrt{1-e^{-2\alpha_*(T-t)}} \tag{3.27}$$

with $\alpha_* = \left(\frac{A\vartheta(p)\ell(p)}{\ell(p)-\hat{\ell}(p)}\right)^2$.

We may also use the previous approximation and linear interpolation to produce a new approximation as follows

$$b(t,p) \approx b_{II}(t,p) := b_I(t,0)(1-p) + b_I(t,1)p \tag{3.28}$$

with $b_I(t,0)$ and $b_I(t,1)$ given by (3.27). We shall give some numerical examples to verify the validity of the previous two formulas in Section 5.

4 Examples

In this section, we consider some special cases of utility functions, including power and non-HARA utility functions, and investigate the shape of investment stopping region. We especially pay attention to the situation when Assumption 3.1 does not hold. We first present an example for power utility, that is, we assume $J=1$, $q_1 = \frac{\gamma}{\gamma-1}$ with $0<\gamma<1$. Then $\tilde{U}_0(y)$ is the dual function of the power utility $U(x) = \frac{1}{\gamma}x^\gamma$. In this case the function ϕ defined in (3.3) is given by

$$\phi(y,p) = A_1(p)y^{q_1} - rKy,$$

with $A_1(p) = \frac{1}{2}\vartheta^2(p)(q_1-1) + (1-\frac{1}{q_1})\beta - r$.

Example 4.1. *Assume $J=1$, $q_1 = \frac{\gamma}{\gamma-1}$ with $0<\gamma<1$. We have the following cases:*

(i) $A_1(1)>0$ (equivalently $\beta > r\gamma - \frac{q_1}{2}\theta_h^2$). If $K>0$, there exists a unique free boundary given by (3.4)*; if $K=0$, it is optimal to stop the investment immediately.*

(ii) $A_1(1)\le 0$, $A_1(0)>0$ (equivalently $r\gamma - \frac{q_1}{2}\theta_l^2 < \beta \le r\gamma - \frac{q_1}{2}\theta_h^2$) and $K\ge 0$. There exists a unique free boundary $g(t,y)\le p_0$ with p_0 being the unique solution to $A_1(p)=0$ such that

$$\mathcal{S}_y = \big\{(t,y,p)\in Q_y : p \le g(t,y)\big\}. \tag{4.1}$$

(iii) $A_1(0)\le 0$ (equivalently $\beta \le r\gamma - \frac{q_1}{2}\theta_l^2$) and $K\ge 0$. There is no free boundary and it is not optimal to stop before the maturity.

Proof. See Appendix A.9. □

Remark 4.1. *Example 4.1 implies that, for power utility agent, if the discount factor β is large and $K > 0$, the agent should sell all risky assets once the wealth process reaches the high threshold value; if β is large and $K = 0$, investing in risky assets is never optimal and the agent should put all his money in the bank account immediately; if β is small, the agent should invest in risky assets until the maturity.*

We next present two examples for non-HARA utility, that is, we assume $J = 2$, $q_1 = -3$, $q_2 = -1$. Then $\tilde{U}_0(y)$ is the dual function of the non-HARA utility function U in (3.2). In this case the function ϕ defined in (3.3) is given by

$$\phi(y,p) = A_1(p)y^{-3} + A_2(p)y^{-1} - rKy,$$

where $A_1(p) = -2\vartheta^2(p) + \frac{4}{3}\beta - r$, $A_2(p) = -\vartheta^2(p) + 2\beta - r$. Note that A_1 and A_2 are decreasing functions of p. Lemma A.3 in Appendix A.10 gives the full characterization of all cases and free boundaries, which depends on the combination of signs of $A_i(p)$ for $i = h, l$ and $p = 0, 1$ and some other measures. To appreciate its practical meanings and implications, we use it to identify the relation of the range of the discount factor β and its associated free boundary in the next example.

Example 4.2. *Assume $J = 2$, $q_1 = -3$, $q_2 = -1$ and $K > 0$. Denote by, for $i = h, l$,*

$$\beta_{1,i} := \frac{3}{2}\theta_i^2 + \frac{3}{4}r, \quad \beta_{2,i} := \frac{1}{2}\theta_i^2 + \frac{1}{2}r, \quad \beta_{3,i} := \beta_{2,i} + \sqrt{rK(\frac{4}{3}\theta_i^2 + \frac{1}{3}r) + \frac{4}{9}r^2K^2} - \frac{2}{3}rK.$$

Note that $\beta_{2,i} < \beta_{3,i} < \beta_{1,i}$ for $i = h, l$. Assume further $\beta_{1,l} > \beta_{2,h}$. We have the following results:

(i) If $\beta > \beta_{1,h}$, there exists one free boundary $b(t,p)$;

(ii) If $\beta_{3,l} < \beta \le \beta_{1,h}$, there exist two free boundaries $b_1(t,p)$ and $b_2(t,p)$;

(iii) If $\beta \le \beta_{3,l}$, there is no free boundary.

Proof. See Appendix A.10. □

Remark 4.2. *Example 4.2 shows that, for non-HARA utility agent, if the discount factor β is large, one should sell the risky asset once the wealth process is above the threshold wealth level, if β is small, then one should hold the risky asset until the maturity, and for β in the mid range, one should stop once the wealth process reaches the lower or upper threshold wealth level. These optimal decisions are based on the assumption that $\beta_{1,l} > \beta_{2,h}$. If $\beta_{1,l} \le \beta_{2,h}$, one can use Lemma A.3 to derive a similar relation of the discount factor and the free boundary. Specifically, β can be classified into five intervals with one free boundary if $\beta \in (\beta_{1,l}, \beta_{2,h}] \cup (\beta_{1,h}, \infty)$, two free boundaries if $\beta \in (\beta_{3,l}, \beta_{1,l}] \cup (\beta_{2,h}, \beta_{1,h}]$, and no free boundary if $\beta \in (0, \beta_{3,l}]$.*

For $K = 0$ and non-HARA utility, we have the following results that can be similarly proved as those in Example 4.2 by studying the set $\{(t,y,p) : \phi(y,p) \ge 0\}$, so omitted here.

Example 4.3. *Assume $J = 2$, $q_1 = -3$, $q_2 = -1$ and $K = 0$. We have the following cases.*

(i) $A_1(1) < 0$, $A_1(0) \le 0$ (equivalently $\beta \le \beta_{1,l}$). There exists a function $\bar{b}(t,p)$ such that

$$\mathcal{S}_y = \big\{(t,y,p) \in Q_y : y \ge \bar{b}(t,p)\big\}.$$

(ii) $A_1(1) \ge 0$ (equivalently $\beta \ge \beta_{1,h}$). It is optimal to stop the investment immediately.

(iii) $A_1(1) < 0$, $A_1(0) > 0$, $A_2(1) > 0$ *(equivalently* $\beta_{1,l} < \beta < \beta_{1,h}$, $\beta > \beta_{2,h}$*). There exists a positive function* $g(t, y)$ *such that*

$$\mathcal{S}_y = [0, T) \times (0, \tilde{y}_1] \times (0, g]$$

with $\tilde{y}_1$ *being the unique solution to* $\phi(y, 1) = 0$.

(iv) $A_1(1) < 0$, $A_1(0) > 0$, $A_2(1) \le 0$ *(equivalently* $\beta_{1,l} < \beta \le \beta_{2,h}$*). There exists a positive function* $g(t, y)$ *such that*

$$\mathcal{S}_y = [0, T) \times (0, \infty) \times (0, g].$$

Remark 4.3. *In Example 4.3, we consider the non-HARA utility function without any constraints on the initial wealth. Compared with Example 4.2, Example 4.3 also shows that if* β *is small enough, there is no stopping boundary for* $K > 0$ *whereas there is a unique stopping boundary for* $K = 0$ *and the agent should sell the stocks once the wealth is sufficiently low; if* β *is sufficiently large, there is a unique stopping boundary for* $K > 0$ *whereas there is no stopping boundary for* $K = 0$ *and the agent should stop investing in the risky assets at once; if* $\beta \in (\beta_{1,l}, \beta_{1,h}) \cap (\beta_{2,h}, \infty)$*, there exists two stopping boundaries for* $K > 0$ *whereas there exists only one stopping boundary for* $K = 0$ *and the agent should stop the investment when the posterior probability process reaches sufficiently high level; if* $\beta \in (\beta_{1,l}, \beta_{2,h})$*, there is a unique stopping boundary no matter* $K > 0$ *or* $K = 0$*, which further illustrates that the portfolio insurance* K *plays a key role in choosing the optimal trading strategy and we can not set* $K = 0$ *to simplify our problem.*

5 Numerical examples

In this section, we present the numerical results computed by the integral equation method (IEM) and global closed form approximation methods (GCA) . For IEM, we adopt the backward recursion method in Detemple (2005) to compute the boundary $\hat{b}$ by solving the integral equation (3.19). Denote by

$$\mathcal{J}(s, t, \hat{\varphi}, \varphi, z) := e^{-\beta s} \int_{-\infty}^{d(s,\hat{\varphi},\varphi)} \psi(t + s, z, \varphi e^{-\frac{1}{2}\Theta^2 s + \Theta\sqrt{s}\eta}) n(\eta) \mathrm{d}\eta.$$

Then we need to solve

$$\int_0^{T-t} \mathcal{J}(s, t, \hat{b}(t + s, z), \hat{b}(t, z), z) \mathrm{d}s = 0$$

with terminal condition (3.20). Set the discretization mesh $t_i = T - i\Delta$ for $i = 0, \dots, N$ with $\Delta = T/N$ and denote by $\hat{b}_i$ the approximation of $\hat{b}(t_i, z)$. Using the trapezoidal rule to discretize the integral equation, we get

$$0 = \sum_{j=1}^{N-i} \left(\mathcal{J}(j\Delta, t_i, \hat{b}_{i+j}, \hat{b}_i, z) + \mathcal{J}((j-1)\Delta, t_i, \hat{b}_{i+j-1}, \hat{b}_i, z) \right) \tag{5.1}$$

for $i = 0, 1, \dots, N - 1$. There are N equations and $N + 1$ unknowns $\hat{b}_0, \dots \hat{b}_N$. Using (3.7), we may set $\hat{b}_N = \ell(p)$ and then find $\hat{b}_i$, $i = N - 1, \dots, 0$, by solving equations (5.1) in a backward way and with the Newton iteration method. After getting $\hat{b}_i$ for $i = 0, 1, \dots, N$, we can obtain the corresponding value for free boundary b by the relation (3.18). Finally, the free boundary of problem (2.5) can be calculated by the dual relation

$$B(t, p) = -\tilde{U}'_K(b(t, p)).$$

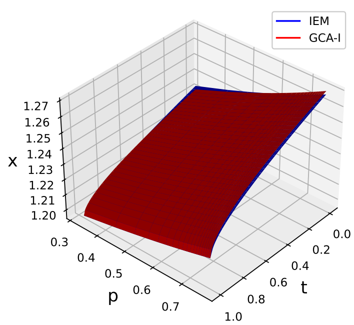
Besides, we have constructed two global closed-form approximations (GCAs) in (3.27) and (3.28).

Example 5.1. *In this example, we plot the free boundary and the optimal strategy of the optimal investment stopping problem* (2.5)*. We consider the dual utility function of the form*

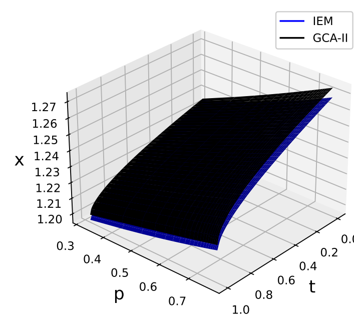
$$\tilde{U}_K(y) = -\frac{1}{q}y^q - Ky$$

with $q = -3$ *and* $K = 1$*. In this case,* $\tilde{U}_0(y)$ *is the dual function of the power utility* $U(x) = \frac{4}{3}x^{\frac{3}{4}}$*. The other parameters used in this example are* $\beta = 0.05$*,* $r = 0.01$*,* $\sigma = 0.25$*,* $\mu_h = 0.03$*,* $\mu_l = 0.02$*,* $T = 1$*. The initial wealth is* $x_0 = 1.2$ *and the initial belief* $p_0 = 0.5$*.*

We notice that the parameters above satisfy the Assumption 3.1. Thus, there exists a unique free boundary b for the dual problem (2.11) given by (3.4). In Figure 1, we plot the optimal exercise boundary by the integral equation method (IEM), using (3.27) and (3.28) and we view the solution derived by IEM as a benchmark. It is shown that the errors of the two approximation formulas are small and GCA-I is more accurate than GCA-II. In Figure 2(a) and Figure 2(b), the sample path of optimal wealth and optimal trading strategy are depicted. To find the optimal stopping time, we solve $\partial_y \tilde{V}(0, y, p_0) + x_0 = 0$ firstly to obtain $y = y_*$. The optimal stopping time is the first time that the process $\{\Phi\}_{t\geq 0}$ hits the free boundary $\hat{b}(t, z)$ starting from $\Phi_0 = p_0/(1 - p_0)$ with $z = \frac{y_*(1-p_0)^{\varpi-1}}{p_0^{\varpi}}$, which is depicted as τ_0 in Figure 2(c). Figure 2 reveals that it is optimal to stop investing in the risky assets at $t = \tau_0$ when the corresponding path of Φ hits the free boundary $\hat{b}$ for the first time, and when the path of Φ does not hit the free boundary $\hat{b}$ before terminal date T the individual should hold the risky asset until the terminal time.



(a)



(b)

Figure 1: (a) The optimal stopping boundary compared with GCA-I; (b) The optimal stopping boundary compared with GCA-II.

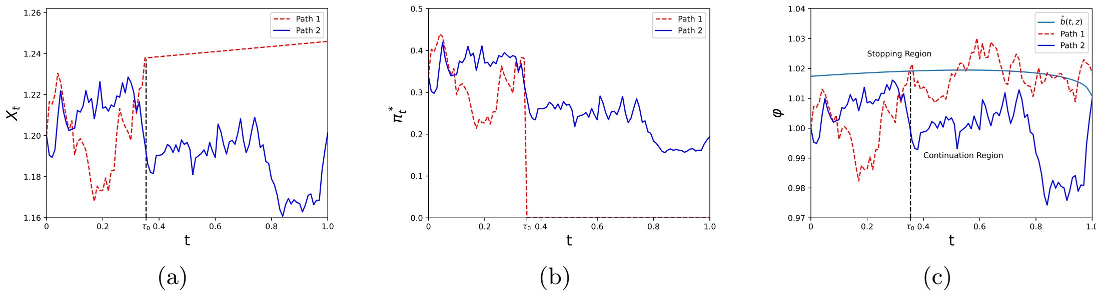


Figure 2: (a) Two different sample paths of wealth with initial wealth $x_0 = 1.2$; (b) Two different sample paths of optimal investment strategy with initial wealth $x_0 = 1.2$; (c) Two different sample paths of likelihood ratio process with initial ratio $\varphi_0 = 1$ and optimal stopping time.

Example 5.2. *In this example, we compare the optimal values and the optimal investment strategies obtained by the two GCAs and the IEM at the time $t = 0$ for utility function used in Example 5.1.*

(i) We compare the numerical results between the two GCAs and IEM. The parameters used are the same as Example 5.1. The numerical result is shown in Table 1.

(ii) Table 2 presents the mean and standard deviation of the relative difference between IEM and the GCAs. We fix $K = 1$, $T = 1$, initial wealth $x_0 = 1.5$ and initial belief $p_0 = 0.5$. The rest parameters are selected randomly: 10 *samples of μ_h from the uniform distribution on interval* $[0.03, 0.08]$, *μ_l on* $[0.02, 0.06]$, *r on* $[0.01, 0.05]$, *β on* $[0.04, 0.08]$, *σ on* $[0.2, 0.5]$, *q on* $[-4, -1]$. *We also require the parameters satisfy Assumption 3.1.*

Table 1: Comparison between the GCAs and IEM for Example 5.2 (i).

Method	For prime value			For optimal strategy		
	Value	Diff.	Time (s)	Strategy	Diff.	Time (s)
GCA-I	0.399607	6.75e-5	18.2	0.338528	7.68e-5	3.1
GCA-II	0.399515	1.63e-4	17.4	0.333473	1.50e-2	2.8
IEM	0.399580	–	5637.2	0.338554	–	1842.3

Table 2: Comparison between the GCAs and IEM for Example 5.2 (ii).

Method	For prime value			For optimal strategy		
	Avg. diff.	Std. diff.	Avg. time (s)	Avg. diff.	Std. diff.	Avg. time (s)
GCA-I	6.73e-5	1.41e-4	18.5	3.59e-3	6.23e-3	3.1
GCA-II	1.43e-4	3.24e-4	18.6	5.34e-3	6.87e-3	3.0
IEM	–	–	5637.2	–	–	1842.3

From the numerics in Tables 1 and 2, we observe that the difference between the GCAs and IEM optimal values is very small, whereas the computational time for GCAs is much less than that for IEM. The GCA-I is much more accurate than GCA-II. Compared to the optimal values, the error for computing the optimal strategies using both the IEM and GCAs is larger. This is not surprising, as the optimal strategies are involved with the second derivatives of the dual value functions.

6 Conclusions

In this paper we have given rigorous analysis on the free boundary for a class of two-dimensional mixed optimal investment and stopping problems with unobservable random drift of risky asset. The problem is degenerate and the free boundary is a three-variable function of time, wealth and initial belief. We have characterized the properties of the free boundary and found it with the integral equation method and the global closed form approximation method. There remain some open problems, for example, the current approach can not be extended if the drift coefficient is a hidden two-state Markov chain. Detailed comments are given as follows. Assume that the drift coefficient μ is a hidden two-state Markov chain, that is, $\mu = \mu(\alpha_t) \in \{\mu_h, \mu_l\}$, where the Markov chain α is characterized by the generator of the form

$$\begin{pmatrix} -\lambda_h & \lambda_h \\ \lambda_l & -\lambda_l \end{pmatrix}, \quad \lambda_h,\, \lambda_l > 0. \tag{6.1}$$

As usual, we define the conditional probability process P as (2.2). By Chen et al. (2021), the Wonham filter P satisfies

$$\mathrm{d}P_t = (\lambda_l - (\lambda_h + \lambda_l)P_t)\,\mathrm{d}t + \Theta P_t(1 - P_t)\mathrm{d}\tilde{B}_t$$

with innovation process $\tilde{B}$ given by

$$\mathrm{d}\tilde{B}_s = \frac{1}{\sigma}\frac{\mathrm{d}S_t}{S_t} - \mathbb{E}[\mu(\alpha_t)|\mathcal{F}_t^S]\mathrm{d}t.$$

It turns out the likelihood ratio process Φ, defined in (3.10), satisfies

$$\mathrm{d}\Phi_t = (\lambda_l - \lambda_h\Phi_t)(1 + \Phi_t)\mathrm{d}t + \Theta\Phi_t\mathrm{d}W_t$$

under the measure Q defined by (3.11). Furthermore, if we define $F_t := \frac{1+\Phi_t}{1+\varphi}$, then (3.12) does not hold in this case. Therefore, the dimension reduction technique is not applicable when we consider the Wonham filter and the nonlinear integral equation (3.9) can not be simplified further. We leave this and other questions for future research.

Acknowledgment: The authors are very grateful to two anonymous reviewers whose constructive comments and suggestions have helped to improve the paper of the previous version. Jie Xing was supported by National Natural Science Foundation of China (Grant No. 12101151) and Youth Foundation of Guizhou University of Finance and Economics (Grant No. 2022KYQN10). Jingtang Ma was supported by National Natural Science Foundation of China (Grant No. 12071373). Harry Zheng was supported by Engineering and Physical Sciences Research Council of UK (Grant No. EP/V008331/1).

Appendix A Proofs

In this appendix, we shall give the proof of the main results and examples. For convenience, since X and P are time-homogeneous, we rewrite (2.11) and (2.15) as

$$\tilde{V}(t, y, p) = \sup_{\tau \in \mathcal{T}_{0,T-t}} \mathbb{E}[e^{-\beta\tau}\tilde{U}_K(Y_\tau^{y,p})],$$

and

$$\tau^*(t, y, p) := \inf\left\{s \geq 0 : (t + s, Y_s^{y,p}, P_s^p) \in \mathcal{S}_y\right\} \wedge (T - t),$$

respectively. Denote by $L^{\tilde{p}}(\mathcal{F}_T)\,(\tilde{p}\geq 1)$ the space of $\mathcal{F}_T$-measurable random variables ξ with norm $\{\mathbb{E}[\xi^{\tilde{p}}]\}^{1/\tilde{p}}<\infty$, $\mathscr{L}_t^{\tilde{p}}$ the space of continuous $\mathcal{F}_t$-measurable stochastic process X with norm $\Big\{\mathbb{E}\big[\int_0^t |X_s|^{\tilde{p}}ds\big]\Big\}^{1/\tilde{p}}<\infty$, $W_{\tilde{p}}^{1,2,2}(Q)\,(\tilde{p}\geq 1)$ the Sobolev space and $W_{\tilde{p},loc}^{1,2,2}(Q)\,(\tilde{p}\geq 1)$ the local Sobolev space, see Friedman (1982) for details of these spaces used extensively in the nonlinear PDE theory.

A.1 Proof of Proposition 3.1

The main difficulty of the proof lies in the monotonicity of $\tilde{V}(t,y,\cdot)$ because the operator $\mathcal{L}_{Y,P}$ is degenerate and the traditional method in Friedman (1982) can not be applied directly. To overcome it, we choose a Brownian motion $\hat{B}$ independent of $\tilde{B}$ and let $\bar{B}_t := \rho\tilde{B}_t+\sqrt{1-\rho^2}\hat{B}_t$ with correlation coefficient $-1<\rho<1$. We consider the following auxiliary optimal stopping problem for the moment

$$\bar{V}(t,y,p;\rho)=\sup_{\tau\in\mathcal{T}_{0,T-t}}\mathbb{E}\big[e^{-\beta\tau}\tilde{U}_K(\bar{Y}_\tau^{y,p})\big], \tag{A.1}$$

where

$$\begin{aligned}
\mathrm{d}\bar{Y}_t&=(\beta-r)\bar{Y}_t\mathrm{d}t-\vartheta(P_t)\bar{Y}_t\mathrm{d}\bar{B}_t,\\
\mathrm{d}P_t&=\Theta P_t(1-P_t)\mathrm{d}\tilde{B}_t.
\end{aligned}$$

The properties of $\bar{V}$ are summarized in the following lemma.

Lemma A.1. *The value function of problem* (A.1) *satisfies* $\bar{V}\in W_{\tilde{p},loc}^{1,2,2}(Q_y)\cap C(\bar{Q}_y)$ *for any* $\tilde{p}>2$ *and* $\bar{V}(t,\cdot,p;\rho)$ *is strictly convex. Moreover,*

$$\begin{aligned}
&\partial_t\bar{V}(t,y,p;\rho)\leq 0,\quad \partial_y\bar{V}(t,y,p;\rho)\leq 0,\quad \partial_p\bar{V}(t,y,p;\rho)\geq 0,\\
&\lim_{\rho\to 1}\bar{V}(t,y,p;\rho)=\tilde{V}(t,y,p).
\end{aligned}$$

Proof of Lemma A.1. Let

$$\bar{\mathcal{S}}_y:=\big\{(t,y,p)\in Q_y:\bar{V}(t,y,p;\rho)=\tilde{U}_K(y)\big\}.$$

We split the proof into four steps.

Step 1. We prove that $\bar{V}\in W_{\tilde{p},loc}^{1,2,2}(Q_y)\cap C(\bar{Q}_y)$. By dynamic programming principle, $\bar{V}$ satisfies

$$\begin{aligned}
&\min\big\{-\bar{V}_t-\mathcal{L}_{\bar{Y},P}\bar{V}+\beta\bar{V},\bar{V}-\tilde{U}_K\big\}=0,\ (t,y,p)\in Q_y,\\
&\bar{V}(T,y,p)=\tilde{U}_K(y),\ y>0,
\end{aligned}$$

where $\mathcal{L}_{\bar{Y},P}$ is the infinitesimal generator of $(\bar{Y},P)$, given by

$$\mathcal{L}_{\bar{Y},P}\bar{V}=\frac{1}{2}\vartheta^2(p)y^2\partial_{yy}\bar{V}+\frac{1}{2}\Theta^2p^2(1-p)^2\partial_{pp}\bar{V}-\rho\Theta p(1-p)\vartheta(p)y\partial_{yp}\bar{V}+(\beta-r)y\partial_y\bar{V}.$$

We notice that the operator $\mathcal{L}_{\bar{Y},P}$ is non-degenerate. According to standard penalty method (see Friedman (1982, Theorem 8.2)), it follows that $\bar{V}\in W_{\tilde{p},loc}^{1,2,2}(Q_y)\cap C(\bar{Q}_y)$.

Step 2. It is not difficult to verify that

$$\mathbb{E}\Big[\sup_{0\leq t\leq T}\big|\tilde{U}_K(\bar{Y}_t)\big|\Big]<\infty.$$

Standard optimal stopping theory (see Peskir and Shiryaev (2006, Corollary 2.9)) yields that the optimal stopping time is given by

$$\bar{\tau}^*(t,y,p) := \inf\big\{s \geq 0 : (s+t, \bar{Y}_s^{y,p}, P_s^p) \in \bar{\mathcal{S}}_y\big\} \wedge (T-t).$$

Consequently, the convexity of $\bar{V}(t,\cdot,p)$ follows from the fact that $\tilde{U}_K(\cdot)$ is strictly convex and the affine property of $\bar{Y}_t^{y,p}$ in y.

Step 3. We prove the monotonicity of $\bar{V}$. Indeed, $\partial_t\bar{V} \leq 0$ and $\partial_y\bar{V} \leq 0$ follow easily by the definition of $\bar{V}$ and $\tilde{U}_K$ is non-increasing .

It remains to show $\partial_p\bar{V} \geq 0$. Obviously, we find that $\partial_p\bar{V} = 0$ if $(t,y,p) \in \bar{\mathcal{S}}_y$. Assume that $(t,y,p) \in \mathcal{C}_y$. Noting that in $\mathcal{C}_y$, $\bar{V}$ satisfies

$$-\mathcal{L}_{\bar{Y},P}\bar{V} + \beta\bar{V} = \partial_t\bar{V}$$

According to Friedman (1964, Theorem 10, p. 72), $\partial_t\bar{V}$ is C^1 with Hölder continuous derivatives with respect to the state variables. By Friedman (1964, Theorem 20, p. 87), we find that the function $(y,p) \mapsto \bar{V}(t,y,p)$ is C^3 in the domain $\mathcal{C}_y$. Differentiating the previous PDE with respect to p, we obtain $\partial_p\bar{V}$ satisfies

$$-\partial_t(\partial_p\bar{V}) - \mathcal{L}^p_{\bar{Y},P}\partial_p\bar{V} + \beta\partial_p\bar{V} = \Theta\vartheta(p)y^2\partial_{yy}\bar{V} \geq 0,$$

where

$$\mathcal{L}^p_{\bar{Y},P}\bar{V} := \mathcal{L}_{\bar{Y},P}\bar{V} + \Theta^2p(1-p)(1-2p)\partial_p\bar{V} - \rho\big[\Theta(1-2p)\vartheta(p) + \Theta^2p(1-p)\big]y\partial_y\bar{V}.$$

Moreover, Step 1 gives that $\partial_p\bar{V} = 0$ on $\partial\bar{\mathcal{C}}_y$ because the Sobolev embedding implies $\bar{V} \in C^1$ and $\partial_p\bar{V} = 0$ in $\mathcal{S}_y$. We have $\partial_p\bar{V} \geq 0$ by maximum principle.

Step 4. We show that $\lim\limits_{\rho\to 1}\bar{V}(t,y,p;\rho) = \tilde{V}(t,y,p)$. Choosing $\bar{\tau}^*$ as an optimal stopping time for $\bar{V}(t,y,p;\rho)$, we deduce that

$$\bar{V}(t,y,p;\rho) - \tilde{V}(t,y,p) \leq -\sum_{j=1}^{J}\frac{1}{q_j}\mathbb{E}\Big[\big|(\bar{Y}_{\bar{\tau}^*})^{q_j} - (Y_{\bar{\tau}^*})^{q_j}\big|\Big] + K\mathbb{E}\Big[\big|\bar{Y}_{\bar{\tau}^*} - Y_{\bar{\tau}^*}\big|\Big]. \tag{A.2}$$

Furthermore, for any $q \in \mathbb{R}$ we have

$$\begin{aligned}
&\mathbb{E}\Big[\big|(\bar{Y}_{\bar{\tau}^*})^q - (Y_{\bar{\tau}^*})^q\big|\Big] \\
\leq\ &\mathbb{E}\Big[\big((\bar{Y}_{\bar{\tau}^*})^q + (Y_{\bar{\tau}^*})^q\big)\Big|(1-\rho)\int_0^{\bar{\tau}^*}q\vartheta(P_s)\mathrm{d}\tilde{B}_s - \sqrt{1-\rho^2}\int_0^{\bar{\tau}^*}q\vartheta(P_s)\mathrm{d}\hat{B}_s\Big|\Big] \\
\leq\ &\Big(\big\|(\bar{Y}_{\bar{\tau}^*})^q\big\|_{L^2(\mathcal{F}_T)} + \big\|(Y_{\bar{\tau}^*})^q\big\|_{L^2(\mathcal{F}_T)}\Big)\Big((1-\rho)\Big\|\int_0^{\bar{\tau}^*}q\vartheta(P_s)\mathrm{d}\tilde{B}_s\Big\|_{L^2(\mathcal{F}_T)} \\
&+ \sqrt{1-\rho^2}\Big\|\int_0^{\bar{\tau}^*}q\vartheta(P_s)\mathrm{d}\hat{B}_s\Big\|_{L^2(\mathcal{F}_T)}\Big) \\
\leq\ &-Cq\big(1-\rho+\sqrt{1-\rho^2}\big)\|\vartheta(P)\|_{\mathscr{L}^2_T},
\end{aligned} \tag{A.3}$$

for some appropriate constant $C > 0$, where the first inequality follows from the inequality $|e^x - e^y| \leq (e^x + e^y)|x-y|$, the second inequality follows from Hölder's inequality and Minkowski inequality and the third inequality follows from (2.14). Thus, combining (A.2) and (A.3), we have

$$\overline{\lim_{\rho\to 1}}\,\bar{V}(t,y,p;\rho) - \tilde{V}(t,y,p) \leq 0. \tag{A.4}$$

Similarly, choosing τ^* as optimal stopping time for $\tilde{V}(t,y,p)$, we have

$$\begin{aligned}\bar{V}(t,y,p;\rho)-\tilde{V}(t,y,p) &\geq -\mathbb{E}\Big[\big|\tilde{U}_K(\bar{Y}_{\tau^*})-\tilde{U}_K(Y_{\tau^*})\big|\Big]\\ &\geq \sum_{j=1}^{J}\frac{1}{q_j}\mathbb{E}\Big[\big|(\bar{Y}_{\tau^*})^{q_j}-(Y_{\tau^*})^{q_j}\big|\Big]-K\mathbb{E}\Big[\big|\bar{Y}_{\tau^*}-Y_{\tau^*}\big|\Big]\to 0,\end{aligned}$$

as $\rho\to 1$, which gives that

$$\varliminf_{\rho\to 1}\bar{V}(t,y,p;\rho)-\tilde{V}(t,y,p)\geq 0. \tag{A.5}$$

Combining (A.4) and (A.5), we conclude that $\lim_{\rho\to 1}\bar{V}(t,y,p;\rho)=\tilde{V}(t,y,p)$. □

Proof of Proposition 3.1. A similar argument as in the proof of Lemma A.1 shows that $\tilde{V}(t,\cdot,p)$ is strictly convex, $\tilde{V}$ is non-increasing both in t and y. To show that $\tilde{V}(t,y,\cdot)$ is non-decreasing, choosing $0<p_1<p_2<1$, thanks to Lemma A.1, we find

$$\tilde{V}(t,y,p_1)=\lim_{\rho\to 1}\bar{V}(t,y,p_1;\rho)\leq\lim_{\rho\to 1}\bar{V}(t,y,p_2;\rho)=\tilde{V}(t,y,p_2).$$

In the end, we assume that Assumption 3.1 holds and show that $\tilde{V}(t,\cdot,p)-\tilde{U}_K(\cdot)$ is non-decreasing. Thanks to Dynkin's formula, for any $\tau\in\mathcal{T}_{t,T}$, we deduce that

$$\mathbb{E}[e^{-\beta\tau}\tilde{U}_K(Y_\tau)]-\tilde{U}_K(y)=-\mathbb{E}\Big[\int_0^\tau e^{-\beta s}\phi(Y_s,P_s)\mathrm{d}s\Big]. \tag{A.6}$$

Thus, we have

$$\tilde{V}(t,y,p)-\tilde{U}_K(y)=\sup_{\tau\in\mathcal{T}_{t,T}}-\mathbb{E}\Big[\int_0^\tau e^{-\beta s}\phi(Y_s,P_s)\mathrm{d}s\Big].$$

By Assumption 3.1 we deduce that $\phi(\cdot,p)$ is non-increasing. Since $Y_s^{y_1,p}<Y_s^{y_2,p}$ for any $s\geq 0$ with $y_1<y_2$, we conclude that $\tilde{V}(t,\cdot,p)-\tilde{U}_K(\cdot)$ is non-decreasing as desired. □

A.2 Proof of Proposition 3.2

Proof of Proposition 3.2. Under Assumption 3.1, due to the variational inequality (2.12), we know that

$$\mathcal{S}_y\subset\Big\{(t,y,p)\in Q_y : y\leq\ell(p)\Big\},$$

where $\ell(p)$ is the unique solution to $\phi(y,p)=0$ for any $p\in(0,1)$. By Proposition 3.1, the stopping region is downward connected and the continuation region is upward connected. Hence, there exists a function b defined by (3.4) such that (3.5) holds.

To show the monotonicity of b, we denote the $t-$section of the stopping region by

$$\mathcal{S}_y^t=\big\{(y,p) : y\leq b(t,p)\big\}.$$

Since $\tilde{V}(\cdot,y,p)$ is non-increasing, the $t-$section of $\mathcal{S}_y$ satisfies $\mathcal{S}_y^{t_1}\subset\mathcal{S}_y^{t_2}$ for $t_1<t_2$. That is, $b(\cdot,y)$ is non-decreasing. Similarly, define the $p-$section of $\mathcal{S}_y$ by

$$\mathcal{S}_y^p=\big\{(t,y) : y\leq b(t,p)\big\}.$$

Then $\mathcal{S}_y^{p_2}\subset\mathcal{S}_y^{p_1}$ for $p_1<p_2$ as $\tilde{V}(t,y,\cdot)$ is non-decreasing, i.e. $b(t,\cdot)$ is non-increasing. □

A.3 Proof of Proposition 3.4

Proof of Proposition 3.4. We borrow some ideas from De Angelis (2020, Proposition 5.2) and divide the proof into four steps.

Step 1. We prove that $b(\cdot,p)$ is right continuous and $b(t,\cdot)$ is left continuous. Let $\{t_n\}_{n\geq 1}$ be a sequence such that $t_n \downarrow t_0$ as $n\to\infty$. Then the sequence $\{b(t_n,p)\}_{n\geq 1}$ is non-increasing. So the $\lim_{n\to\infty} b(t_n,p)$ exists and $\lim_{n\to\infty} b(t_n,p) \geq b(t_0,p)$. On the other hand, since

$$\tilde V\big(t_n, b(t_n,p), p\big) = \tilde U_K\big(b(t_n,p)\big), \quad n\geq 1,$$

the continuity of $\tilde V$ and $\tilde U_K$ implies

$$\tilde V\big(t_0, \lim_{n\to\infty} b(t_n,p), p\big) = \tilde U_K\big(\lim_{n\to\infty} b(t_n,p)\big).$$

The definition of $b(t_0,p)$ yields $\lim_{n\to\infty} b(t_n,p) \leq b(t_0,p)$, which means that $b(\cdot,p)$ is right continuous. Analogously, a symmetric argument also shows that $b(t,\cdot)$ is left continuous.

Step 2. We prove that $b(t,\cdot)$ is right continuous. We pick $t_0\in(0,T)$ and $p_0\in(0,1)$. As $b(t_0,\cdot)$ is non-increasing, it follows that $b(t_0,p_0)\geq b(t_0,p_0+)$. Indeed, we claim that the equality holds and argue by contradiction. Suppose that $b(t_0,p_0) > b(t_0,p_0+)$, then we can pick $y_1,y_2>0$ such that $b(t_0,p_0) > y_2 > y_1 > b(t_0,p_0+)$. Denote $\mathcal{N} := (y_1,y_2)\times(p_0,1)$. It follows that for any $(y,p)\in\mathcal{N}$,

$$y > b(t_0,p_0+) \geq b(t_0,p),$$

where the second inequality follows from the fact that $b(t_0,\cdot)$ is non-increasing. That is,

$$(t_0,y,p)\in\mathcal{C}_y \quad \text{for} \quad (y,p)\in\mathcal{N}. \tag{A.7}$$

Moreover, from the definition of b, we know that

$$(t_0,y,p_0)\in\mathcal{S}_y \quad \text{for} \quad y\in(y_1,y_2). \tag{A.8}$$

Letting

$$v(t,y,p) := \tilde V(t,y,p) - \tilde U_K(y), \tag{A.9}$$

then in the region $\mathcal{C}_y$, we have

$$(\partial_t + \mathcal{L}_{Y,P})v - \beta v = \phi(y,p). \tag{A.10}$$

By (A.7), we see that in region $\mathcal{N}$, $v(t_0,\cdot,\cdot)$ satisfies

$$\mathcal{L}_{Y,P} v(t_0,y,p) - \beta v(t_0,y,p) = \phi(y,p) - \partial_t v(t_0,y,p).$$

Letting $u=\partial_y v$, we find that $u(t_0,\cdot,\cdot)$ satisfies

$$\mathscr{L} u(t_0,y,p) - \beta u(t_0,y,p) = -\partial_t u(t_0,y,p) + \partial_y\phi(y,p) \quad \text{in} \quad \mathcal{N}, \tag{A.11}$$

where

$$\mathscr{L} = \mathcal{L}_{Y,P} + y\vartheta^2(p)\partial_y - \Theta p(1-p)\vartheta(p)\partial_p + \beta - r.$$

We pick a positive function $\bar\psi(y)\in C_c^\infty(y_1,y_2)$ and define

$$\mathcal{G}(p) := \int_{y_1}^{y_2} \partial_{pp} u(t_0,y,p)\bar\psi(y)\mathrm{d}y. \tag{A.12}$$

By (A.11), we deduce that

$$\begin{aligned}\frac{1}{2}\Theta^2 p^2(1-p)^2\mathcal{G}(p) &= -\int_{y_1}^{y_2}\partial_t u(t_0,y,p)\bar{\psi}(y)dy+\int_{y_1}^{y_2}\partial_y\phi(y,p)\bar{\psi}(y)\mathrm{d}y\\ &\quad -\int_{y_1}^{y_2}\big[\mathscr{L}u-\beta u-\frac{1}{2}\Theta^2p^2(1-p)^2\partial_{pp}u\big](t_0,y,p)\bar{\psi}(y)\mathrm{d}y.\end{aligned}$$

Since $\partial_y\phi<0$ by the Assumption 3.1, integrating by parts and letting $p\to p_0$ in the previous equation, then using the C^1 continuity of $\tilde{V}$ (see Proposition 3.3) and (A.8), we have

$$\begin{aligned}\lim_{p\to p_0}\frac{1}{2}\Theta^2p^2(1-p)^2\mathcal{G}(p) &= -\int_{y_1}^{y_2}(u\cdot\mathscr{L}_1\bar{\psi})(t_0,y,p_0)\mathrm{d}y-\int_{y_1}^{y_2}(\partial_p v\cdot\mathscr{L}_2\bar{\psi})(t_0,y,p_0)\mathrm{d}y\\ &\quad -\lim_{p\to p_0}\int_{y_1}^{y_2}\partial_t u(t_0,y,p)\bar{\psi}(y)\mathrm{d}y+\int_{y_1}^{y_2}\partial_y\phi(y,p_0)\bar{\psi}(y)\mathrm{d}y\\ &= \lim_{p\to p_0}\int_{y_1}^{y_2}\partial_t v(t_0,y,p)\bar{\psi}'(y)dy+\int_{y_1}^{y_2}\partial_y\phi(y,p_0)\bar{\psi}(y)\mathrm{d}y\\ &= \int_{y_1}^{y_2}\partial_y\phi(y,p_0)\bar{\psi}(y)\mathrm{d}y\\ &< 0,\end{aligned}$$

where the last inequality follows from $\phi(\cdot,p)$ is strictly decreasing, and

$$\begin{aligned}\mathscr{L}_1 &= \frac{1}{2}\vartheta^2(p)y^2\partial_{yy}+(\vartheta^2(p)-\beta+r)y\partial_y-\beta,\\ \mathscr{L}_2 &= -\Theta p(1-p)\vartheta(p)y\partial_{yy}-\Theta p(1-p)\vartheta(p)\partial_y.\end{aligned}$$

This implies

$$\lim_{p\to p_0}\mathcal{G}(p)<0.$$

Thus, there exists small $\epsilon>0$ such that

$$\begin{aligned}0 &> \int_{p_0}^{p_0+\epsilon}\int_{p_0}^{z}\mathcal{G}(p)\mathrm{d}p\mathrm{d}z\\ &= -\int_{p_0}^{p_0+\epsilon}\int_{p_0}^{z}\int_{y_1}^{y_2}\partial_{pp}v(t_0,y,p)\bar{\psi}'(y)\mathrm{d}y\mathrm{d}p\mathrm{d}z\\ &= -\int_{y_1}^{y_2}\int_{p_0}^{p_0+\epsilon}\big[\partial_p v(t_0,y,z)-\partial_p v(t_0,y,p_0)\big]\bar{\psi}'(y)\mathrm{d}z\mathrm{d}y\\ &= -\int_{y_1}^{y_2}\big[v(t_0,y,p_0+\epsilon)-v(t_0,y,p_0)\big]\bar{\psi}'(y)\mathrm{d}y\\ &= \int_{y_1}^{y_2}\partial_y v(t_0,y,p_0+\epsilon)\bar{\psi}(y)\mathrm{d}y,\end{aligned}$$

where the first equality follows from integrating by parts and (A.12), the second equality follows from Fubini's theorem, the third and the final equality follow from that $v\in C^1$, (A.8), and integrating by parts. However, $\partial_y v=\partial_y\tilde{V}-\tilde{U}_K'\geq 0$ by Proposition 3.1. This forces the integration above to be non-negative and provides a contradiction. Hence, $b(t_0,\cdot)$ is right continuous as claimed.

Step 3. We prove that $b(\cdot,p)$ is left continuous. As in step 2, we pick $t_0\in(0,T)$ and $p_0\in(0,1)$. Since $b(\cdot,p)$ is non-decreasing, we have $\lim_{t\uparrow t_0}b(t,p_0)\leq b(t_0,p_0)$. Suppose that $\lim_{t\uparrow t_0}b(t,p_0)<b(t_0,p_0)$. We pick $y_1,y_2>0$ such that

$$\lim_{t\uparrow t_0}b(t,p_0)<y_1<y_2<b(t_0,p_0). \tag{A.13}$$

Since $b(\cdot,p)$ is non-decreasing and $b(t,\cdot)$ is non-increasing, we have

$$\lim_{p\downarrow p_0}\lim_{t\uparrow t_0} b(t,p) = \lim_{t\uparrow t_0}\lim_{p\downarrow p_0} b(t,p) = \lim_{t\uparrow t_0} b(t,p_0). \tag{A.14}$$

Combining (A.13) and (A.14), we have there exists some $\delta > 0$ such that

$$\lim_{t\uparrow t_0} b(t,p) < y_1 < y_2 < b(t_0,p) \tag{A.15}$$

for any $p \in (p_0, p_0+\delta)$ as we have already shown $b(t,\cdot)$ is right continuous in step 2. Denote $\mathcal{U} := (0,t_0)\times(y_1,y_2)\times(p_0,p_0+\delta)$. Then for any $(t,y,p)\in\mathcal{U}$, since $b(\cdot,p)$ is non-decreasing, we deduce that

$$y > y_1 > \lim_{t\uparrow t_0} b(t,p) \geq b(t,p),$$

which implies $(t,y,p)\in\mathcal{C}_y$. We must have $\mathcal{U}\subset\mathcal{C}_y$ and u satisfies (A.11) with t_0 replaced by t. Moreover, (A.15) means that

$$v(t_0,y,p) = 0, \quad (y,p)\in(y_1,y_2)\times(p_0,p_0+\delta). \tag{A.16}$$

We pick positive functions $\bar\psi(y)\in C_c^\infty(y_1,y_2)$ and $\chi(p)\in C_c^\infty(p_0,p_0+\delta)$ with $\int_{y_1}^{y_2}\bar\psi(y)\mathrm{d}y = \int_{p_0}^{p_0+\delta}\chi(p)\mathrm{d}p = 1$, and denote

$$G(t) := \int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2}\partial_t u(t,y,p)\bar\psi(y)\chi(p)\mathrm{d}y\mathrm{d}p.$$

By (A.11) , we deduce that

$$G(t) = -\int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2}(\mathscr{L}u-\beta u)(t,y,p)\bar\psi(y)\chi(p)\mathrm{d}y\mathrm{d}p + \int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2}\partial_y\phi(y,p)\bar\psi(y)\chi(p)\mathrm{d}y\mathrm{d}p.$$

Integrating by parts, we derive

$$\begin{aligned}
G(t) &= -\int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2} u(t,y,p)(\mathscr{L}^*-r)(\bar\psi\cdot\chi)(y,p)\mathrm{d}y\mathrm{d}p \\
&\quad + \int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2}\partial_y\phi(y,p)\bar\psi(y)\chi(p)\mathrm{d}y\mathrm{d}p \\
&= \int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2} v(t,y,p)\frac{\partial}{\partial y}\big((\mathscr{L}^*-r)(\bar\psi\cdot\chi)\big)(y,p)\mathrm{d}y\mathrm{d}p \\
&\quad + \int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2}\partial_y\phi(y,p)\bar\psi(y)\chi(p)\mathrm{d}y\mathrm{d}p \\
&\leq \int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2} v(t,y,p)\frac{\partial}{\partial y}\big((\mathscr{L}^*-r)(\bar\psi\cdot\chi)\big)(y,p)\mathrm{d}y\mathrm{d}p - rK,
\end{aligned} \tag{A.17}$$

where $\mathscr{L}^*$ is the adjoint of the operator $\mathscr{L}-\beta+r$, and the last inequality holds as Assumption 3.1 implies $\partial_y\phi \leq -rK$. Letting $t\to t_0$ in (A.17), (A.16) gives

$$\lim_{t\to t_0} G(t) \leq -rK < -\frac{rK}{2}.$$

So there exists some $0 < \epsilon_1 < \epsilon_2$ such that

$$
\begin{aligned}
-\frac{rK}{2}(\epsilon_2-\epsilon_1) &> \int_{\epsilon_1}^{\epsilon_2} G(t_0+s)\mathrm{d}s \\
&= \int_{\epsilon_1}^{\epsilon_2}\int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2} \partial_t u(t_0+s,y,p)\bar{\psi}(y)\chi(p)\mathrm{d}y\mathrm{d}p\mathrm{d}s \\
&= \int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2} \big(u(t_0+\epsilon_2,y,p)-u(t_0+\epsilon_1,y,p)\big)\bar{\psi}(y)\chi(p)\mathrm{d}y\mathrm{d}p \\
&= \int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2} u(t_0+\epsilon_2,y,p)\bar{\psi}(y)\chi(p)\mathrm{d}y\mathrm{d}p \\
&\quad + \int_{p_0}^{p_0+\delta}\int_{y_1}^{y_2} v(t_0+\epsilon_1,y,p)\bar{\psi}'(y)\chi(p)\mathrm{d}y\mathrm{d}p, \qquad \text{(A.18)}
\end{aligned}
$$

where the second equality follows from Fubini's theorem and the last equality follows from integrating by parts. Using the fact that $u = \partial_y\tilde{V} - \tilde{U}_K' \geq 0$ (see Proposition 3.1) and (A.16), letting $\epsilon_1 \to 0$ in (A.18), we have

$$0 > -\frac{rK}{2}\epsilon_2 \geq \int_{\epsilon_1}^{\epsilon_2} G(t_0+s)\mathrm{d}s \geq 0,$$

which provides a contradiction. The claim follows.

Step 4. Since $b(\cdot,p)$ is non-decreasing for any fixed $0 < p < 1$, the continuity of b in t and p separately implies that b is continuous in $(0,T)\times(0,1)$ (see Kruse and Deely (1969)). □

A.4 Proof of Proposition 3.5

Proof of Proposition 3.5. Since $b(\cdot,p)$ is non-decreasing and

$$\mathcal{S}_y \subset \big\{(t,y,p)\in Q_y : y \leq \ell(p)\big\}$$

implied by the variational inequality (2.12), we deduce that

$$\lim_{t\uparrow T} b(t,p) \leq \ell(p).$$

We fix $p = p_0$ and argue by contradiction. Assume $\lim_{t\uparrow T} b(t,p_0) < \ell(p_0)$. Then we can pick y_1, y_2 such that $\lim_{t\uparrow T} b(t,p_0) < y_1 < y_2 < \ell(p_0)$. Thus, there exists some $t' < T$ and $\delta' > 0$ such that

$$b(t,p_0) < y_1, \quad y_2 < \ell(p) \quad \text{for} \quad (t,p)\in(t',T)\times(p_0,p_0+\delta'). \tag{A.19}$$

Denote $\mathcal{D} := (t',T)\times(y_1,y_2)\times(p_0,p_0+\delta')$. Since $b(t,\cdot)$ is non-increasing, for any $(t,y,p)\in\mathcal{D}$, we have

$$y > y_1 > b(t,p_0) \geq b(t,p), \tag{A.20}$$

which means that $\mathcal{D}\subset\mathcal{C}_y$.

Let $\tilde{\psi}(y)\in C_c^\infty(y_1,y_2)$ and $\tilde{\chi}(p)\in C_c^\infty(p_0,p_0+\delta')$ with $\tilde{\psi}(y)\geq 0$ and $\tilde{\chi}(p)\geq 0$. We introduce

$$\tilde{\mathcal{G}}(t) := \int_{p_0}^{p_0+\delta'}\int_{y_1}^{y_2} \partial_t v(t,y,p)\tilde{\psi}(y)\tilde{\chi}(p)\mathrm{d}y\mathrm{d}p,$$

where v is defined in (A.9). Recall that v satisfies (A.10). Then integrating by parts, we deduce that

$$\begin{aligned}
\lim_{t\uparrow T}\tilde{\mathcal{G}}(t) &= \lim_{t\uparrow T}\int_{p_0}^{p_0+\delta'}\int_{y_1}^{y_2}\big(-\mathcal{L}_{Y,P}v+\beta v+\phi\big)(t,y,p)\tilde{\psi}(y)\tilde{\chi}(p)\mathrm{d}y\mathrm{d}p \\
&= \lim_{t\uparrow T}\int_{p_0}^{p_0+\delta'}\int_{y_1}^{y_2} v(t,y,p)\big(-\mathcal{L}^*_{Y,P}+\beta\big)\big(\tilde{\psi}(y)\tilde{\chi}(p)\big)+\phi(y,p)\tilde{\psi}(y)\tilde{\chi}(p)\mathrm{d}y\mathrm{d}p \\
&= \int_{p_0}^{p_0+\delta'}\int_{y_1}^{y_2}\phi(y,p)\tilde{\psi}(y)\tilde{\chi}(p)\mathrm{d}y\mathrm{d}p, \qquad (A.21)
\end{aligned}$$

where $\mathcal{L}^*_{Y,P}$ is the adjoint of the operator $\mathcal{L}_{Y,P}$ and we have used the terminal condition $v(T,y,p)=0$ in the last equality.

For any $y\in(y_1,y_2)$, since $y<y_2<\ell(p)$ for $p\in(p_0,p_0+\delta')$ (see (A.19)) , we deduce that $\phi(y,p)>0$ for $(y,p)\in(y_1,y_2)\times(p_0,p_0+\delta')$. Hence, we must have $\lim_{t\uparrow T}\tilde{\mathcal{G}}(t)>0$ by (A.21), which further gives there exists some $0<\bar{\delta}<T-t'$ such that

$$\begin{aligned}
0<\int_{T-\bar{\delta}}^{T}\tilde{\mathcal{G}}(t)\mathrm{d}t &= \int_{T-\bar{\delta}}^{T}\int_{p_0}^{p_0+\delta'}\int_{y_1}^{y_2}\partial_t v(t,y,p)\tilde{\psi}(y)\tilde{\chi}(p)\mathrm{d}y\mathrm{d}p\mathrm{d}t \\
&= \int_{p_0}^{p_0+\delta'}\int_{y_1}^{y_2}\big[v(T,y,p)-v(T-\bar{\delta},y,p)\big]\tilde{\psi}(y)\tilde{\chi}(p)\mathrm{d}y\mathrm{d}p \\
&= -\int_{p_0}^{p_0+\delta'}\int_{y_1}^{y_2}v(T-\bar{\delta},y,p)\tilde{\psi}(y)\tilde{\chi}(p)\mathrm{d}y\mathrm{d}p<0,
\end{aligned}$$

where the last inequality follows from $(T-\bar{\delta},y,p)\in\mathcal{D}\subset\mathcal{C}_y$ by (A.20). This is a contradiction. The proof is complete. □

A.5 Proof of Proposition 3.7

Proof of Proposition 3.7. Letting $\mathcal{Y}_t:=F_tY_t$ and applying Itô's lemma to $\mathcal{Y}$, we deduce that

$$\mathrm{d}\mathcal{Y}_t=(\beta-r)\mathcal{Y}_t\mathrm{d}t-\theta_l\mathcal{Y}_t\mathrm{d}W_t$$

with $\mathcal{Y}_0=y$. Now (A.6) yields

$$\begin{aligned}
\tilde{V}(t,y,p)-\tilde{U}_K(y) &= \sup_{\tau\in\mathcal{T}_{0,T-t}}\mathbb{E}\Big[-\int_0^{\tau}e^{-\beta s}\phi(Y_s,P_s)\mathrm{d}s\Big] \\
&= \frac{1}{1+\varphi}\sup_{\tau\in\mathcal{T}_{0,T-t}}\mathbb{E}^{\mathbb{Q}}\Big[-\int_0^{\tau}e^{-\beta s}(1+\Phi_s)\phi\Big(\frac{\mathcal{Y}_s(1+\varphi)}{1+\Phi_s},\frac{\Phi_s}{1+\Phi_s}\Big)\mathrm{d}s\Big],
\end{aligned}$$

where the second equality follows from measure change and $\mathcal{Y}_s=Y_sF_s$, $P_s=\Phi_s/(1+\Phi_s)$. Since $\mathcal{Y}$ and Φ are driven by the same Brownian motion W, we have

$$\mathcal{Y}_t=ye^{\varsigma t}\Big(\frac{\Phi_t}{\varphi}\Big)^{\varpi}$$

with $\varsigma=\beta-r-\frac{1}{2}\theta_l^2-\frac{1}{2}\Theta\theta_l$ and $\varpi=-\frac{\theta_l}{\Theta}<0$. The proposition follows from the transformation $z=\frac{y(1+\varphi)}{\varphi^{\varpi}}e^{-\varsigma t}$ and the fact that Φ is time-homogeneous.

□

A.6 Proof of Proposition 3.8

To prove the proposition, we need the following lemma.

Lemma A.2. *The function $\tilde{\mathcal{V}}(t,z,\varphi)$ is non-increasing in φ and non-decreasing in z.*

Proof of Lemma A.2. Firstly, we claim that $\psi(t,z,\cdot)$ is non-increasing. Recalling the relation that $y=\frac{ze^{\varsigma t}\varphi^{\varpi}}{1+\varphi}$ and $p=\frac{\varphi}{1+\varphi}$, we deduce that

$$\frac{\partial\psi}{\partial\varphi}=\sum_{j=1}^{J}\Big(-A_j-A_jq_j\frac{\varpi}{p}+A_jq_j+\vartheta(p)\Theta(1-p)(1-q_j)\Big)y^{q_j}+\frac{rK\varpi}{p}y.$$

Consider the auxiliary function

$$\bar{F}_j(p):=-A_j(p)-A_j(p)q_j\frac{\varpi}{p}+A_j(p)q_j+\vartheta(p)\Theta(1-p)(1-q_j),\quad 0<p<1.$$

Under Assumption 3.1, we have $A_j>0$. It follows that $\lim_{p\downarrow 0}\bar{F}_j(p)=-\infty$ and $\lim_{p\uparrow 1}\bar{F}_j(p)<0$. Furthermore,

$$\begin{aligned}\bar{F}_j'(p) &= \vartheta(p)\Theta(1-q_j)q_j\frac{\varpi}{p}+\frac{A_jq_j\varpi}{p^2}-\vartheta(p)\Theta(1-q_j)q_j+\Theta^2(1-p)(1-q_j)\\ &> 0,\end{aligned}$$

which gives $\bar{F}_j(p)<0$. Noting that $\varpi<0$, we have $\frac{\partial\psi}{\partial\varphi}<0$. The claim now follows.

Using $\psi(t,z,\cdot)$ is strictly decreasing, we have for any $\varphi_1<\varphi_2$,

$$\begin{aligned}\tilde{\mathcal{V}}(t,z,\varphi_1) &= \sup_{\tau\in\mathcal{T}_{0,T-t}}\mathbb{E}^{\mathbb{Q}}\Big[\int_0^{\tau}e^{-\beta s}\psi\Big(t+s,z,\Phi_s^{\varphi_1}\Big)\mathrm{d}s\Big]\\ &\geq \sup_{\tau\in\mathcal{T}_{0,T-t}}\mathbb{E}^{\mathbb{Q}}\Big[\int_0^{\tau}e^{-\beta s}\psi\Big(t+s,z,\Phi_s^{\varphi_2}\Big)\mathrm{d}s\Big]=\tilde{\mathcal{V}}(t,z,\varphi_2).\end{aligned}$$

That is, $\tilde{\mathcal{V}}(t,z,\cdot)$ is non-increasing. Similarly, the Assumption 3.1 means that $\psi(t,\cdot,\varphi)$ is strictly increasing, which further gives that $\tilde{\mathcal{V}}(t,\cdot,\varphi)$ is non-decreasing. □

Proof of Proposition 3.8. (3.16) implies that

$$\mathcal{S}_z\subset\{(t,z,\varphi)\in Q_z:\psi(t,z,\varphi)\leq 0\}.$$

Since $\lim_{\varphi\to 0}\psi(t,z,\varphi)=+\infty$, $\lim_{\varphi\to+\infty}\psi(t,z,\varphi)=-\infty$, $\partial_\varphi\psi(t,z,\varphi)<0$ (see Lemma A.2), we deduce that there exists a function $(t,z)\mapsto\Upsilon(t,z)$ such that

$$\mathcal{S}_z\subset\{(t,z,\varphi)\in Q_z:\varphi\geq\Upsilon(t,z)\}.$$

We pick some $\varphi_0>0$ such that $(t,z,\varphi_0)\in\mathcal{S}_z$. As $\tilde{\mathcal{V}}(t,z,\cdot)$ is non-increasing (see Lemma A.2), it follows that $(t,z,\varphi)\in\mathcal{S}_z$ for any $\varphi>\varphi_0$. Thus, the stopping region is upward connected, i.e. there exists a unique free boundary $\hat{b}$ such that (3.17) holds.

Since $\tilde{\mathcal{V}}(t,\cdot,\varphi)$ is non-decreasing (see Lemma A.2), we have $\hat{b}(t,\cdot)$ is non-decreasing. Thus, we can define the generalized inverse function of $\hat{b}(t,\cdot)$ as follows

$$\hat{b}^{-1}(t,\varphi):=\inf\{z\in\mathbb{R}_+:\ \hat{b}(t,z)>\varphi\}$$

for any $\varphi > 0$. Then the stopping region defined in the $(t, y, p)-$coordinate can be rewritten as

$$\mathcal{S}_y = \Big\{(t, y, p) \in Q_y : \, y \leq \frac{p^{\varpi}}{(1-p)^{\varpi-1}} e^{\varsigma t} \hat{b}^{-1}\Big(t, \frac{p}{1-p}\Big)\Big\}.$$

This implies that (3.18) holds. Furthermore, the relation (3.18) and Proposition 3.4 imply that $\hat{b}^{-1}(t, \varphi)$ is a continuous function. Consequently, $\hat{b}(t, \cdot)$ is strictly increasing. The proof is completed. □

A.7 Proof of Proposition 3.9

Proof of Proposition 3.9. Recalling that

$$\Phi_t = \frac{P_t}{1-P_t}, \quad \mathcal{Y}_t = ye^{\varsigma t}\Big(\frac{\Phi_t}{\varphi}\Big)^{\varpi}, \quad \mathcal{Y}_t = Y_t F_t,$$
$$\tilde{\mathcal{V}}(t, z, \varphi) = (1+\varphi)(\tilde{V}(t, y, p) - \tilde{U}_K(y)), \quad z = \frac{(1-p)^{\varpi-1}}{p^{\varpi}} ye^{-\varsigma t},$$
$$\{Y_s^{t,y,p} > b(s, P_s^{t,p})\} = \{\Phi_s < \hat{b}(t+s, z)\},$$

(3.8) yields

$$\begin{aligned}\tilde{\mathcal{V}}(t, z, \varphi) &= \mathbb{E}^{\mathbb{Q}}\Big[\int_0^{T-t} e^{-\beta s}\psi(t+s, z, \Phi_s)\mathbb{1}_{\{\Phi_s < \hat{b}(t+s,z)\}} \mathrm{d}s\Big] \\ &= \int_0^{T-t} e^{-\beta s} \int_{-\infty}^{d(s,\hat{b}(t+s,z),\varphi)} \psi(t+s, z, \varphi e^{-\frac{1}{2}\Theta^2 s + \Theta\sqrt{s}\eta}) n(\eta) \mathrm{d}\eta \mathrm{d}s\end{aligned}$$

where $n(\eta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}}$, and $d(s, \hat{b}(t+s, z), \varphi) = \frac{\log(\hat{b}(t+s,z)/\varphi) + \frac{1}{2}\Theta^2 s}{\Theta\sqrt{s}}$. Since $\tilde{\mathcal{V}}$ is continuous, the free boundary $\hat{b}(t, z)$ satisfies (3.19).

Finally, we show that the terminal condition (3.20) holds. For this, let $\mathcal{K}(p) := \ell(p)\frac{(1-p)^{\varpi-1}}{p^{\varpi}} e^{-\varsigma T}$. By Proposition 3.5, recalling that $b(t, p) = \frac{p^{\varpi}}{(1-p)^{\varpi-1}} e^{\varsigma t} \hat{b}^{-1}\Big(t, \frac{p}{1-p}\Big)$ and $\hat{b}(t, \cdot)$ is strictly increasing (Proposition 3.8), we deduce that

$$\hat{b}(T, \mathcal{K}(p)) = \frac{p}{1-p}.$$

It remains to show that $\mathcal{K}(p) = z_0$ has a unique solution for any fixed $z_0 > 0$. The existence follows from $\lim_{p\to 0} \mathcal{K}(p) = 0$ and $\lim_{p\to 1} \mathcal{K}(p) = +\infty$. To prove the uniqueness, we argue by contradiction. If there exist $p_1 < p_2$ such that $\mathcal{K}(p_1) = z_0 = \mathcal{K}(p_2)$, then it follows that

$$\hat{b}(T, \mathcal{K}(p_1)) = \frac{p_1}{1-p_1} = \hat{b}(T, \mathcal{K}(p_2)) = \frac{p_2}{1-p_2}.$$

This provides a contradiction. □

A.8 Proof of Proposition 3.11

Proof of Proposition 3.11. We only need to prove the first limit in (3.26). The other limit can be proved in a similar way. Firstly, under Assumption 3.1, we claim that

$$\lim_{p\downarrow 0} \tilde{V}(t, y, p) = \tilde{\mathscr{V}}(t, y; 0). \tag{A.22}$$

Choosing $\tau_*^l(t,y)$ as optimal stopping time for $\tilde{\mathscr{V}}(t,y;0)$, then we deduce

$$\begin{aligned}\tilde{V}(t,y,p)-\tilde{\mathscr{V}}(t,y;0) &\geq \mathbb{E}\Big[e^{-\beta\tau_*^l}\big(\tilde{U}_K(Y_{\tau_*^l})-\tilde{U}_K(\tilde{Y}_{\tau_*^l})\big)\Big]\\ &\geq \sum_{j=1}^J \frac{1}{q_j}\mathbb{E}\Big[\big|(Y_{\tau_*^l})^{q_j}-(\tilde{Y}_{\tau_*^l})^{q_j}\big|\Big]-K\mathbb{E}\Big[\big|Y_{\tau_*^l}-\tilde{Y}_{\tau_*^l}\big|\Big]. \qquad \text{(A.23)}\end{aligned}$$

Since $\lim_{p\to 0} P_t^p = 0$ (see Décamps et al. (2005, Lemma 4.3)), it follows that $\vartheta(P_t^p)\to\theta_l=\vartheta(0)$ as $p\to 0$. Also, it is not difficult to see that for any $q\in\mathbb{R}$

$$\mathbb{E}\big[\sup_{0\leq s\leq T}(\tilde{Y}_s)^q\big]\leq C,$$

for some appropriate constant C. Now, letting $p\to 0$ in (A.23), in exactly the same way as in (A.2) - (A.4), we deduce that

$$\varliminf_{p\downarrow 0}\tilde{V}(t,y,p)-\tilde{\mathscr{V}}(t,y;0)\geq 0.$$

Moreover, a symmetric argument by choosing τ^* as optimal stopping time for $\tilde{V}(t,y,p)$ also shows that

$$\varlimsup_{p\downarrow 0}\tilde{V}(t,y,p)-\tilde{\mathscr{V}}(t,y;0)\leq 0.$$

Thus, the claim now follows.

Furthermore, applying dynamic programming principle, $\tilde{\mathscr{V}}(t,y;0)$ satisfies the following variational inequality

$$\min\big\{-\partial_t\tilde{\mathscr{V}}-\mathcal{L}_{\tilde{\mathcal{Y}}}\tilde{\mathscr{V}}+\beta\tilde{\mathscr{V}},\tilde{\mathscr{V}}-\tilde{U}_K\big\}=0, \qquad \text{(A.24)}$$

where $\tilde{\mathcal{Y}}$ satisfies

$$\mathrm{d}\tilde{\mathcal{Y}}_s=(\beta-r)\tilde{\mathcal{Y}}_s\mathrm{d}s-\theta_l\tilde{\mathcal{Y}}_s\mathrm{d}\tilde{B}_s, \quad \tilde{\mathcal{Y}}_t=y$$

with $\mathcal{L}_{\tilde{\mathcal{Y}}}$ being the infinitesimal generator of $\tilde{\mathcal{Y}}$.

Suppose that $\lim_{p\downarrow 0} b(t,p):=\tilde{g}(t)$. Taking limit $p\to 0$ in (3.8), noting that $\mathbb{P}(\tilde{\mathcal{Y}}_s^{t,y}=\tilde{g}(s))=0$, (A.22) yields

$$\tilde{\mathscr{V}}(t,y;0)-\tilde{U}_K(y)=-\mathbb{E}\Big[\int_t^T e^{-\beta(s-t)}\phi(\tilde{\mathcal{Y}}_s^{t,y},0)\mathbb{1}_{\{\tilde{\mathcal{Y}}_s^{t,y}>\tilde{g}(s)\}}\mathrm{d}s\Big]$$

Applying Feynman-Kac formula, $\tilde{\mathscr{V}}(t,y;0)-\tilde{U}_K(y)$ satisfies

$$(\partial_t+\mathcal{L}_{\tilde{\mathcal{Y}}}-\beta)(\tilde{\mathscr{V}}(t,y;0)-\tilde{U}_K(y))=\phi(y,0)\mathbb{1}_{\{y>\tilde{g}(t)\}}.$$

It follows that if $y>\tilde{g}(t)$, then $\tilde{\mathscr{V}}$ satisfies

$$(\partial_t+\mathcal{L}_{\tilde{\mathcal{Y}}}-\beta)\tilde{\mathscr{V}}(t,y;0)=0;$$

if $y<\tilde{g}(t)$, then $\mathcal{V}$ satisfies

$$(\partial_t+\mathcal{L}_{\tilde{\mathcal{Y}}}-\beta)\tilde{\mathscr{V}}(t,y;0)=-\phi(y,0)<0$$

as $y<\tilde{g}(t)\leq\ell(0)$. By the variational inequality (A.24), we find that $(t,y)\in\tilde{\mathscr{S}}_y^{0c}$ for $y>\tilde{g}(t)$ and $(t,y)\in\tilde{\mathscr{S}}_y^{0}$ for $y<\tilde{g}(t)$, i.e. $\tilde{g}(t)=\tilde{b}(t;0)$ is the free boundary that split the whole region into two parts. The proof is completed.

□

A.9 Proof of Example 4.1

Proof of Example 4.1. Firstly, note that

$$\mathcal{S}_y \subset \{(t,y,p):\ \phi(y,p) \geq 0\}. \tag{A.25}$$

(i) If $K > 0$, then (i) follows from Proposition 3.2. Suppose that $K = 0$, then we have $\phi(y,p) \geq 0$ in $\mathbb{R}_+ \times (0,1)$. Using (A.6) and choosing τ^* as the optimal stopping time for $\tilde{V}(t,y,p)$, we have

$$0 \leq \tilde{V}(t,y,p) - \tilde{U}_K(y) = -\mathbb{E}\Big[\int_0^{\tau^*} e^{-\beta s}\phi(Y_s,P_s)\mathrm{d}s\Big] \leq 0,$$

which means that the optimal stopping time $\tau^* = 0$.

(ii) The assumption implies that $A_1(1) \leq 0$, $A_1(0) > 0$ and $A_1(\cdot)$ is strictly decreasing, from which we deduce that

$$\begin{aligned}\mathcal{S}_y \subset \{(t,y,p):\ \phi(y,p) \geq 0\} &= \{(t,y,p):\ A_1(p)y^{q_1-1} \geq rK\}\\ &= \{(t,y,p):\ p \leq A_1^{-1}\Big(\frac{rK}{y^{q_1-1}}\Big)\}\\ &\subset \{(t,y,p):\ 0 < p \leq p_0\}\end{aligned}$$

with $p_0 \in (0,1]$ being the unique solution to $A_1(p) = 0$. Since $\tilde{V}(t,y,\cdot)$ is non-decreasing (see Proposition 3.1), it follows that $\mathcal{S}_y$ is downward connected, i.e. there exists a positive function $g(t,y) \leq p_0$ such that (4.1) holds.

(iii) Notice that $\phi(y,p) \leq 0$ in $\mathbb{R}_+ \times (0,1)$ in this case. Using (A.27), we have $\mathcal{S}_y = \emptyset$, which means that the optimal stopping time $\tau^* = T$. □

A.10 Proof of Example 4.2

We first state a lemma that would help us connect β with the free boundary.

Lemma A.3. *For non-HARA utility function (3.2), let $\tilde{\Delta}(p) := A_2^2(p) + 4rKA_1(p)$ and assume $K > 0$. We have the following cases and results.*

(i) If $A_1(1) > 0$, then there exists a unique continuous free boundary $b(t,p)$ defined by (3.4).

(ii) If $A_1(0) \leq 0$, then

(a) If $A_2(1) \geq 0$ and $\tilde{\Delta}(0) > 0$, then there exist two positive functions $b_1(t,p)$ and $b_2(t,p)$ such that

$$\mathcal{S}_y = [t^*,T) \times [b_1,b_2] \times (0,\tilde{p}_0], \tag{A.26}$$

where $t^* \in [0,T)$ and $\tilde{p}_0 = 1$ if $\tilde{\Delta}(1) \geq 0$ and $\tilde{p}_0$ is the unique solution to $\tilde{\Delta}(p) = 0$ if $\tilde{\Delta}(1) < 0$.

(b) If $A_2(1) \geq 0$ and $\tilde{\Delta}(0) \leq 0$, then it is not optimal to stop the investment before the maturity T.

(c) If $A_2(1) < 0$, $A_2(0) > 0$, $\tilde{\Delta}(0) > 0$, then there exist two positive functions $b_1(t,p)$ and $b_2(t,p)$ such that

$$\mathcal{S}_y = [t^*,T) \times [b_1,b_2] \times (0,\tilde{p}_0],$$

where $t^* \in [0,T)$ and $\tilde{p}_0$ is the unique solution to $\tilde{\Delta}(p) = 0$ on $(0,\tilde{p}_2)$ with $\tilde{p}_2$ being the unique solution to $A_2(p) = 0$.

(d) If $A_2(1) < 0$, $A_2(0) \le 0$ or $A_2(1) < 0$, $A_2(0) > 0$, $\tilde{\Delta}(0) \le 0$, then it is not optimal to stop the investment before the maturity T.

(iii) If $A_1(1) \le 0 < A_1(0)$, then

(a) If $A_2(1) < 0$, then there exists a positive function $g(t, y)$ such that

$$\mathcal{S}_y = [0, T) \times (0, \tilde{y}_0] \times (0, g],$$

with $\tilde{y}_0$ being the unique positive solution to $\phi(y, 0) = 0$.

(b) If $A_2(1) \ge 0$, then there exist two positive function $b_1(t, p)$ and $b_2(t, p)$ such that

$$\mathcal{S}_y = [0, T) \times (0, b_2] \times (0, \tilde{p}_1] \cup [t^*, T) \times [b_1, b_2] \times (\tilde{p}_1, \tilde{p}_0),$$

where $t^* \in [0, T)$, $\tilde{p}_1$ is the unique solution to $A_1(p) = 0$, and $\tilde{p}_0 = 1$ if $\tilde{\Delta}(1) \ge 0$ and $\tilde{p}_0$ is the unique solution to $\tilde{\Delta}(p) = 0$ if $\tilde{\Delta}(1) < 0$.

Proof of Example 4.2. We now use Lemma A.3 to prove the results of Example 4.2. Denote by

$$\beta_{4,i} := \beta_{2,i} - \sqrt{rK(\frac{4}{3}\theta_i^2 + \frac{1}{3}r) + \frac{4}{9}r^2K^2} - \frac{2}{3}rK, \quad i = h, l.$$

We have $\beta_{4,i} < \beta_{2,i} < \beta_{3,i} < \beta_{1,i}$ for $i = h, l$, where $\beta_{1,i}, \beta_{2,i}, \beta_{3,i}$ are defined in Example 4.2. Using the notations above, we find that $A_i(0) > 0$ is equivalent to $\beta > \beta_{i,l}$ for $i = 1, 2$; $A_i(1) > 0$ is equivalent to $\beta > \beta_{i,h}$ for $i = 1, 2$; $\tilde{\Delta}(0) > 0$ is equivalent to $\beta < \beta_{4,l}$ or $\beta > \beta_{3,l}$; $\tilde{\Delta}(1) > 0$ is equivalent to $\beta < \beta_{4,h}$ or $\beta > \beta_{3,h}$.

We can now translate the results of Lemma A.3 in terms of β. To determine specific intervals for β, we need to discuss $\beta_{2,h} > \beta_{3,l}$ and $\beta_{2,h} \le \beta_{3,l}$ separately and use the assumption $\beta_{2,h} < \beta_{1,l}$. We only prove the results for $\beta_{2,h} > \beta_{3,l}$, the results for $\beta_{2,h} \le \beta_{3,l}$ can be proved similarly and omitted here.

(i) $A_1(1) > 0$ or, equivalently, $\beta > \beta_{1,h}$, there exists one free boundary.

(ii)(a) $A_1(0) \le 0, A_2(1) \ge 0, \tilde{\Delta}(0) > 0$ or, equivalently, $\beta \le \beta_{1,l}, \beta \ge \beta_{2,h}, \beta > \beta_{3,l}$ (the case $\beta < \beta_{4,l}$ can not happen as it contradicts to other conditions on β), which gives $\beta_{2,h} \le \beta \le \beta_{1,l}$, there exist two free boundaries.

(ii)(b) $A_1(0) \le 0, A_2(1) \ge 0, \tilde{\Delta}(0) \le 0$ or, equivalently, $\beta \le \beta_{1,l}, \beta \ge \beta_{2,h}, \beta \le \beta_{3,l}$, which is impossible as $\beta_{2,h} > \beta_{3,l}$.

(ii)(c) $A_1(0) \le 0, A_2(1) < 0, A_2(0) > 0, \tilde{\Delta}(0) > 0$ or, equivalently, $\beta \le \beta_{1,l}, \beta < \beta_{2,h}, \beta > \beta_{2,l}, \beta > \beta_{3,l}$ (the case $\beta < \beta_{4,l}$ can not happen as it contradicts to other conditions on β), which gives $\beta_{3,l} < \beta < \beta_{2,h}$, there exist two free boundaries.

(ii)(d) Case 1: $A_1(0) \le 0, A_2(1) < 0, A_2(0) \le 0$ or, equivalently, $\beta \le \beta_{1,l}, \beta < \beta_{2,h}, \beta \le \beta_{2,l}$, which gives $\beta \le \beta_{2,l}$, there is no free boundary. Case 2: $A_1(0) \le 0, A_2(1) < 0, A_2(0) > 0, \tilde{\Delta}(0) \le 0$ or, equivalently, $\beta \le \beta_{1,l}, \beta < \beta_{2,h}, \beta > \beta_{2,l}, \beta \le \beta_{3,l}$, which gives $\beta_{2,l} < \beta \le \beta_{3,l}$, there is no free boundary.

(iii)(a) $A_1(1) \le 0 < A_1(0), A_2(1) < 0$ or, equivalently, $\beta \le \beta_{1,h}, \beta > \beta_{1,l}, \beta < \beta_{2,h}$, which is impossible as $\beta_{2,h} < \beta_{1,l}$.

(iii)(b) $A_1(1) \le 0 < A_1(0), A_2(1) \ge 0$ or, equivalently, $\beta \le \beta_{1,h}, \beta > \beta_{1,l}, \beta \ge \beta_{2,h}$, which gives $\beta_{1,l} < \beta \le \beta_{1,h}$, there exist two free boundaries.

Combining (ii)(a), (ii)(c), (iii)(b) shows that for $\beta_{3,l} < \beta \le \beta_{1,h}$ there exist two free boundaries. Combining (ii)(d) Cases 1 and 2 shows that for $\beta \le \beta_{3,l}$ there is no free boundary. This completes the proof of Example 4.2. □

Proof of Lemma A.3. To study the shape of the stopping region $\mathcal{S}_y$, differentiating $\tilde{\Delta}$, we have

$$\tilde{\Delta}'(p) = -4\Theta\vartheta(p)A_2(p) - 16\Theta rK\vartheta(p) = 4\Theta\vartheta(p)(-A_2(p) - 4rK).$$

(i) Under Assumption 3.1, the shape of the stopping region follows from Proposition 3.2.

(ii) (a) Firstly, we assume that $\tilde{\Delta}(1) \ge 0$. In this case, we have $\tilde{\Delta}(p) > 0$. It follows that there exists two positive functions $\ell_1(p)$ and $\ell_2(p)$ such that

$$\big\{(t, y, p) \in Q_y : \phi(y, p) \ge 0\big\} = \big\{(t, y, p) \in Q_y : \ell_1(p) \le y \le \ell_2(p)\big\},$$

with ℓ_1 and ℓ_2 defined by

$$\ell_1(p) := \Big(\frac{-A_2(p) - \sqrt{A_2^2(p) + 4rKA_1(p)}}{2A_1(p)}\Big)^{-1/2}, \tag{A.27}$$

$$\ell_2(p) := \Big(\frac{-A_2(p) + \sqrt{A_2^2(p) + 4rKA_1(p)}}{2A_1(p)}\Big)^{-1/2}. \tag{A.28}$$

We now prove (A.26) holds. Note that (A.27) implies

$$\mathcal{S}_y \subset \mathcal{M} := \big\{(t, y, p) \in Q_y : \ell_1(p) \le y \le \ell_2(p)\big\}. \tag{A.29}$$

$\mathcal{S}_y$ is not empty. Otherwise, in Q_y,

$$-\partial_t\tilde{V} - \mathcal{L}_{Y,P}\tilde{V} + \beta\tilde{V} = 0,$$

from which we deduce that $\partial_t\tilde{V}(T, y, p) = \phi(y, p) > 0$ for $\ell_1(p) < y < \ell_2(p)$. This is contradict with $\partial_t\tilde{V} \le 0$ by Proposition 3.1.

Also, the monotonicity of $\tilde{V}(\cdot, y, p)$ yields that if there exists some $t_0 \in (0, T)$ such that $\{t_0\} \times (0, +\infty) \times (0, 1) \subset \mathcal{C}_y$, then $[0, t_0) \times (0, +\infty) \times (0, 1) \subset \mathcal{C}_y$. Thus, thanks to (A.29), there exists some $t^* \in [0, T)$ such that we can define

$$b_1(t, p) := \inf\big\{y \in \mathbb{R}_+ : \tilde{V}(t, y, p) = \tilde{U}_K(y)\big\},$$
$$b_2(t, p) := \sup\big\{y \in \mathbb{R}_+ : \tilde{V}(t, y, p) = \tilde{U}_K(y)\big\},$$

for $t \in [t^*, T)$. Moreover, $(t, y, p) \in (0, t^*) \times (0, +\infty) \times (0, 1) \subset \mathcal{C}_y$ if $t^* > 0$.

Denote $\Lambda := \big\{(t, y, p) \in [t^*, T) \times (0, +\infty) \times (0, 1) : b_1(t, p) \le y \le b_2(t, p)\big\}$ and recall that $v = \tilde{V} - \tilde{U}_K$ in (A.9). Obviously, we see that $\mathcal{S}_y \subset \Lambda$. It suffices to show $\Lambda \subset \mathcal{S}_y$. Otherwise, the region $\tilde{\mathcal{M}} := \mathcal{C}_y \cap \Lambda$ is nonempty and in this region,

$$\partial_t v + \mathcal{L}_{Y,P}v - \beta v = \phi(y, p) \ge 0,$$

since $\tilde{\mathcal{M}} \subset \Lambda \subset \{(t, y, p) \in Q_y : \phi(y, p) \ge 0\}$. Moreover, $v = 0$ on $\partial\tilde{\mathcal{M}}$. The maximum principle gives $v \le 0$ in $\tilde{\mathcal{M}}$. This provides a contradiction as $v > 0$ in $\tilde{\mathcal{M}}$. Hence, $\Lambda = \mathcal{S}_y$, i.e. (A.26) holds.

If $\tilde{\Delta}(1) < 0$, noting that

$$\{(t, y, p) : \phi(y, p) \geq 0\} = \{(t, y, p) : \ell_1(p) \leq y \leq \ell_2(p), p \leq \tilde{p}_0\}$$

where ℓ_i $(i = 1, 2)$ is defined in (A.27) and (A.28) and $\tilde{p}_0$ is the unique solution to $\tilde{\Delta}(p) = 0$, the rest of the proof can be carried out in exactly the same way as the case $\tilde{\Delta}(1) \geq 0$.

(ii)(c) The assumption implies that $A_1(p) < 0$ on $(0, 1)$, $A_2(p) > 0$ on $(0, \tilde{p}_2)$ and $A_2(p) < 0$ on $(\tilde{p}_2, 1)$ for some $\tilde{p}_2 \in (0, 1)$. Moreover, $\tilde{\Delta}(\tilde{p}_2) < 0$, $\tilde{\Delta}(0) > 0$, $\tilde{\Delta}'(p) < 0$ on $(0, \tilde{p}_2)$. Then $\tilde{\Delta}(p) > 0$ on $(0, \tilde{p}_0)$ and $\tilde{\Delta}(p) < 0$ on $(\tilde{p}_0, \tilde{p}_2)$. Thus,

$$\{(t, y, p) : \phi(y, p) \geq 0\} = \{(t, y, p) : \ell_1(p) \leq y \leq \ell_2(p), 0 < p \leq \tilde{p}_0\}$$

for some positive functions ℓ_i $(i = 1, 2)$ defined in (A.27) and (A.28), the rest of the proof can be carried out in exactly the same way as (ii)(a).

In the cases (ii)(b) and (ii)(d), we immediately have $\phi < 0$, which gives $\mathcal{S}_y = \emptyset$ by (A.27), i.e. it is optimal to hold the risky asset until terminal time T.

(iii)(a) Suppose $A_1(0) > 0$, $A_1(1) \leq 0$, $A_2(1) < 0$. Let $\tilde{y}_0$ be the unique solution to $\phi(y, 0) = 0$. It follows that $\phi(y, 1) < 0$ for $y > 0$, $\phi(y, 0) < 0$ for $y > \tilde{y}_0$ and $\phi(y, 0) > 0$ for $y < \tilde{y}_0$. Since $\partial_p \phi(y, p) < 0$, there exists a unique solution $h_1(y)$ such that $\phi(y, p) > 0$ if $p < h_1(y)$ and $\phi(y, p) < 0$ if $p > h_1(y)$ for $y < \tilde{y}_0$, i.e.

$$\{(t, y, p) : \phi \geq 0\} = \{(t, y, p) : p \leq h_1(y)\}.$$

Since $\tilde{V}$ is non-decreasing in p, it follows that $\mathcal{S}_y$ is downward connected. Consequently, (a) in (iii) holds.

(iii)(b) Suppose $A_1(0) > 0$, $A_1(1) \leq 0$, $A_2(1) \geq 0$. Consider the following function defined by

$$\tilde{\ell}_2(p) := \begin{cases} (\frac{rK}{A_2(p)})^{-1/2}, & p = \tilde{p}_1, \\ \ell_2(p), & p \neq \tilde{p}_1. \end{cases}$$

Let $\tilde{p}_1$ be the unique solution to $A_1(p) = 0$. Clearly, $A_1(p) \geq 0$ on $(0, \tilde{p}_1]$, $A_1(p) < 0$ on $(\tilde{p}_1, 1)$ and $A_2(p) > 0$ on $(0, 1)$. Thus, $\phi(y, p) \geq 0$ is equivalent to $y \leq \tilde{\ell}_2(p)$ on $(0, \tilde{p}_1]$.

On the other hand, we notice that $\tilde{\Delta}(\tilde{p}_1) = A_2^2(\tilde{p}_1) > 0$, and $\tilde{\Delta}'(p) < 0$. Thus, if $\tilde{\Delta}(1) < 0$, then there exists some $\tilde{p}_0 \in (\tilde{p}_1, 1)$ such that $\tilde{\Delta}(p) > 0$ for $p \in (\tilde{p}_1, \tilde{p}_0)$. Then it follows that $\phi(y, p) \geq 0$ is equivalent to $\ell_1(p) \leq y \leq \ell_2(p)$ on $(\tilde{p}_1, \tilde{p}_0)$. In conclusion,

$$\{(t, y, p) : \phi \geq 0\} = \{(t, y, p) : 0 < p \leq \tilde{p}_1, y \leq \tilde{\ell}_2(p) \text{ or } \tilde{p}_1 < p < \tilde{p}_0, \ell_1(p) \leq y \leq \ell_2(p)\}. \quad \text{(A.30)}$$

Similarly, we can show that if $\tilde{\Delta}(1) \geq 0$, then (A.30) holds with $\tilde{p}_0 = 1$. The rest of the proof follows from almost the same argument as (ii)(a). □

References

Bain, A. and Crisan, D. (2009). Fundamentals of Stochastic Filtering, Springer.

Bensoussan, A., Jang, B.G., and Park, S. (2016). Unemployment risks and optimal retirement in an incomplete market, Operations Research, 64, 1015–1032.

Bismuth, A., Gueant, O. and Pu, J. (2019). Portfolio choice, portfolio liquidation, and portfolio transition under drift uncertainty. Mathematics and Financial Economics, 13, 661–719.

Bodnar, T., Mazur, S. and Okhrin, Y. (2017). Bayesian estimation of the global minimum variance portfolio. European Journal of Operational Research, 256, 292–307.

Crandall, M.G., Ishii, H. and Lions, P.L. (1992). User's guide to viscosity solutions of second order partial differential equations, Bulletin of The American Mathematical Society, 27, 1–67.

Choi, K.J. and Shim, G. (2006). Disutility, optimal retirement, and portfolio selection, Mathematical Finance, 16, 443–467.

Chen, K.X., Jeon, J. and Wong, H.Y. (2021). Optimal retirement under partial information, Mathematics of Operations Research, 47, 1802–1832.

Décamps, J.P., Mariotti, T. and Villeneuve, S. (2005). Investment timing under incomplete information, Mathematics of Operations Research, 30, 472–500.

Detemple, J. (2005). American-Style Derivatives: Valuation and Computation, Chapman & Hall/CRC Financial Mathematics Series.

De Angelis, T. (2020). Optimal dividends with partial information and stopping of a degenerate reflecting diffusion, Finance and Stochastics, 24, 71–123.

Eksi, Z. and Ku, H. (2017). Portfolio optimization for a large investor under partial information and price impact, Mathematical Methods of Operations Research, 86, 601–623.

Ekström, E. and Vaicenavicius, J. (2016). Optimal liquidation of an asset under drift uncertainty, SIAM Journal on Financial Mathematics, 7, 357–381.

Fleming, W. and Soner, S. (1993). Controlled Markov Processes and Viscosity Solutions, Springer.

Friedman, A. (1964). Partial Differential Equations of Parabolic Type, Prentice-Hall.

Friedman, A. (1982). Variational Principles and Free Boundary Problems, John Wiley & Sons.

Henderson, V. and Hobson, D. (2008). An explicit solution for an optimal stopping/optimal control problem which models an asset sale, The Annals of Applied Probability, 18, 1681–1705.

Jian, X., Li, X. and Yi, F.H. (2014). Optimal investment with stopping in finite horizon, Journal of Inequalities and Applications, 432, 1–14.

Jacka, S.D. and Ocejo, A. (2018). On the regularity of American options with regime-switching uncertainty, Stochastic Processes & Their Applications, 128, 803–818.

Jang, H.J., Xu, Z.Q. and Zheng, H. (2024). Optimal investment, heterogeneous consumption and best time for retirement, Operations Research, 72, 832–847.

Kruse, R.L. and Deely, J.J. (1969). Joint continuity of monotonic functions, The American Mathematical Monthly, 76, 74–76.

Karatzas, I. and Wang, H. (2000). Utility maximization with discretionary stopping, SIAM Journal on Control and Optimization, 30, 306–329.

Klein, M. (2009). Comment on "Investment Timing Under Incomplete Information", Mathematics of Operations Research, 34, 249–254.

Koo, H.K., Roh, K.H. and Shin, Y.H. (2021). Optimal consumption/investment and retirement with necessities and luxuries, Mathematical Methods of Operations Research, 94, 281–317.

Lakner, P. (1998). Optimal trading strategy for an investor: the case of partial information, Stochastic Processes and their Applications, 76, 77–97.

Ma, J.T., Xing, J. and Zheng, H. (2019). Global closed-form approximation of free boundary for optimal investment stopping problems, SIAM Journal on Control and Optimization, 57, 2092–2121.

Papanicolaou, A. (2019). Backward SDEs for control with partial information, Mathematical Finance, 29, 208–248.

Peskir, G. and Shiryaev, A. (2006). Optimal Stopping and Free-Boundary Problems, Birkhäuser.

Pham, H. (2009). Continuous-time Stochastic Control and Optimization with Financial Applications, Springer.

Shreve, S.E. and Soner, H.M. (1994). Optimal investment and consumption with transaction costs, The Annals of Applied Probability, 4, 609–692.

Soner, H.M. and Touzi, N. (2002). Stochastic target problems, dynamic programming, and viscosity solutions, SIAM Journal on Control and Optimization, 41, 404–424.

Sass, J. and Haussmann, U.G. (2004). Optimizing the terminal wealth under partial information: The drift process as a continuous time Markov chain, Finance and Stochastics, 8, 553–577.

Xu, Z.Q. and Yi, F.H. (2019). Optimal redeeming strategy of stock loans under drift uncertainty, Mathematics of Operations Research, 45, 384–401.

Yang, Z. and Koo, H.K. (2018). Optimal consumption and portfolio selection with early retirement option, Mathematics of Operations Research, 43, 1378–1404.

Zariphopoulou, T. (1992). Investment-consumption models with transaction fees and Markov-chain parameters, SIAM Journal on Control and Optimization, 30, 613–636.