Automatica 173 (2025) 112034

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Optimal consumption under a drawdown constraint over a finite horizon ☆

Xiaoshan Chen [a], Xun Li [b], Fahuai Yi [a], Xiang Yu [b,*]

[a] School of Mathematical Science, South China Normal University, Guangzhou 510631, China

[b] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 10 August 2022
Received in revised form 6 February 2024
Accepted 8 November 2024
Available online 3 December 2024

Keywords:
Optimal consumption
Drawdown constraint
Parabolic variational inequality
Gradient constraint
Free boundary

ABSTRACT

This paper studies a finite horizon utility maximization problem on excessive consumption under a drawdown constraint. Our control problem is an extension of the one considered in Angoshtari et al. (2019) to the model with a finite horizon and an extension of the one considered in Jeon and Oh (2022) to the model with zero interest rate. Contrary to Angoshtari et al. (2019), we encounter a parabolic nonlinear HJB variational inequality with a gradient constraint, in which some time-dependent free boundaries complicate the analysis significantly. Meanwhile, our methodology is built on technical PDE arguments, which differs from the martingale approach in Jeon and Oh (2022). Using the dual transform and considering the auxiliary variational inequality with gradient and function constraints, we establish the existence and uniqueness of the classical solution to the HJB variational inequality after the dimension reduction, and the associated free boundaries can be characterized in analytical form. Consequently, the piecewise optimal feedback controls and the time-dependent thresholds for the ratio of wealth and historical consumption peak can be obtained.

1. Introduction

Optimal portfolio and consumption via utility maximization has always been one of the core research topics in quantitative finance. Starting from the seminal works (Merton, 1969) and Merton (1971), a large amount of studies can be found in the literature. One notable research direction is to refine the measurement of consumption performance by encoding the impact of past consumption behavior. The so-called addictive habit formation preference recommends that the utility shall be generated by the difference between the current consumption rate and the historical weighted average of the past consumption. In addition, the infinite marginal utility mandates the addictive habit formation constraint that the consumption level needs to stay above the habit formation level, representing that the agent's standard of living can never be compromised. Along this direction, fruitful results can be found in various market models, see among Angoshtari, Bayraktar, and Young (2019, 2022), Bo, Wang, and Yu (2024), Constantinides (1990), Detemple and Zapatero (1992), Englezos and Karatzas (2009), Munk (2008), Schroder and Skiadas (2002), Yang and Yu (2022), Yu (2015, 2017) and references therein.

Another rapidly growing research stream is to investigate the impact of the past consumption maximum instead of the historical average. The pioneering work (Dybvig, 1995) examines an extension of Merton's problem under a ratcheting constraint on consumption rate such that the consumption control needs to be non-decreasing. Arun (2012) later generalizes the model in Dybvig (1995) by considering a drawdown constraint such that the consumption rate cannot fall below a proportion of the past consumption maximum. Angoshtari et al. (2019) revisit the problem by considering a drawdown constraint on the excessive dividend rate up to the bankruptcy time, which can be regarded as an extension of the problem in Arun (2012) to the model with zero interest rate. Englezos and Karatzas (2021) further extend the work in Arun (2012) by considering general utility functions using the martingale duality approach where the dual optimal stopping problem is examined therein. Tanana (2023) employs the general duality approach and establishes the existence of optimal consumption under a drawdown constraint in incomplete semimartingale market models. Jeon and Oh (2022) recently generalize the approach in Englezos and Karatzas (2021) to the model with a finite horizon and addresses the existence of a solution to the dual optimal stopping problem.

☆ X. Chen is supported by NNSF of China no.12271188. X. Li is supported by Hong Kong RGC grants under no. 15216720 and 15221621. F. Yi is supported by NNSF of China no.12271188 and 12171169. X. Yu is supported by the Hong Kong RGC General Research Fund (GRF) under Grant No. 15306523 and the Hong Kong Polytechnic University, Hong Kong research grant under no. P0039251. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Subhrakanti Dey under the direction of Editor Florian Dorfler.

* Corresponding author.
E-mail addresses: xschen@m.scnu.edu.cn (X. Chen), li.xun@polyu.edu.hk (X. Li), fhyi@scnu.edu.cn (F. Yi), xiang.yu@polyu.edu.hk (X. Yu).

https://doi.org/10.1016/j.automatica.2024.112034
0005-1098/

On the other hand, inspired by the time non-separable preference rooted in the habit formation preference, some recent studies also incorporate the past consumption maximum into the utility function as an endogenous reference level. Guasoni, Huberman, and Ren (2020) first adopt the Cobb–Douglas utility that is defined on the ratio between the consumption rate and the past consumption maximum and obtain the feedback controls in analytical form. Deng, X. Li, and Yu (2022) choose the same form of the linear habit formation preference and investigate an optimal consumption problem when the difference between the consumption rate and the past spending peak generates the utility. Later, the preference in Deng et al. (2022) is generalized to an S-shaped utility in Li, Yu, and Zhang (2024) to account for the agent's loss aversion over the relative consumption. Liang, Luo, and Yuan (2023) extend the work in Deng et al. (2022) by considering the change of risk aversion parameter concerning the reference level where an additional drawdown constraint is also required. Li, Yu, and Zhang (2022) incorporates the dynamic life insurance control and expected bequest with an additional drawdown constraint on the consumption control.

As an important add-on to the existing literature, the present paper revisits the optimal excessive consumption problem under drawdown constraint as in Angoshtari et al. (2019), however, with a finite investment horizon. We summarize the main contributions of the present paper as two-fold:

(i) From the modeling perspective, it is well documented that heterogeneous agents may have diverse choices of investment horizons in practice. In fact, as the agent's time horizon changes, the risk tolerance should be adjusted accordingly. Typically, agents seek more stable assets for short-term horizon and would call for a more aggressive strategy for the longer-term investment. Our model with a terminal horizon provides the flexibility to meet versatile needs in applications with different choices of investment horizons.

(ii) From the methodology perspective, we encounter a parabolic HJB variational inequality with gradient constraint. It is well known that the global regularity of the parabolic problem requires a more delicate analysis of the time-dependent free boundaries and the smooth fit conditions. Some previous arguments in Angoshtari et al. (2019) crucially rely on the constant free boundary points as well as some explicit expressions of the value function. The present paper contributes to some new and rigorous proofs to establish the existence and the uniqueness of the classical solution to the associated HJB variational inequality (12) (see Theorem 11). We stress that, in many existing studies using the dual transform, see Chen, Chen, and Yi. (2012), Chen and Yi (2012), Dai and Yi (2009), Guan, Yi, and Chen (2019) among others, the dual domain is often the entire $\mathbb{R}^+ \times (0, T]$, which facilitates some classical PDE arguments. In our framework, due to the drawdown constraint $c_t \geq \alpha z_t$, there exists a threshold for the ratio of the wealth level and the past consumption peak such that the minimum consumption plan needs to maintain at a subsistence level (see (9)) whenever the ratio falls below that threshold. As a result, the left boundary $\omega = 0$ is mapped to an unknown finite boundary $y_0(t) < \infty$, $t > 0$. Employing the boundary condition at $y_0(t)$ in (15), we extend the variational inequality with a gradient constraint to an auxiliary variation inequality (16) with both function and gradient constraints in the unbounded domain $\mathbb{R}^+ \times (0, T]$. We can then apply several further transformations (see Propositions 8–10) and modify some technical arguments in Chen, Li, and Yi (2019) and Chen and Yi (2012) to establish the existence and uniqueness of the solution in $C^{2,1}$ regularity to the primal parabolic HJB variational inequality (see Theorem 11). We also note a recent study (Jeon, Kim, & Yang, 2024) that deals with a parabolic variational inequality with both function and gradient constraints in a different context. In the present paper, we also provide some tailor-made arguments to characterize the associated time-dependent free boundaries in analytical form such that the smooth fit conditions hold (see Theorem 12).

The rest of the paper is organized as follows. In Section 2, we introduce the market model and the utility maximization problem under a consumption drawdown constraint. By dimension reduction using the homogeneity of power utility, we formulate the associated HJB variational inequality and its dual problem. In Section 3, we first analyze the dual linear parabolic variational inequality by considering some auxiliary problems with gradient and function constraints. By showing the existence and uniqueness of the solution to the auxiliary problems and characterizing their free boundaries, we obtain the unique classical solution to the dual HJB variational inequality. In Section 4, using the results from the dual problem, we establish the unique classical solution to the primal HJB variational inequality and verify the optimal feedback controls and all associated time-dependent thresholds in analytical form. Section 5 concludes our theoretical contributions.

2. Market model and problem formulation

2.1. Model setup

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a standard filtered probability space, where $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ satisfies the usual conditions. We consider a financial market consisting of one riskless asset and one risky asset, and the terminal time horizon is denoted by T. The riskless asset price satisfies $dB_t = rB_t dt$ where $r \geq 0$ represents the constant interest rate. The risky asset price follows the dynamics

$$dS_t = S_t(\mu + r)dt + S_t\sigma dW_t, \quad t \in [0, T],$$

where W is an $\mathbb{F}$-adapted Brownian motion and the drift $(\mu + r)$ and volatility $\sigma > 0$ are given constants. It is assumed that the excessive return $\mu > 0$, i.e., the risky asset's return is higher than the interest rate.

Let $(\pi_t)_{t\in[0,T]}$ represent the dynamic amount that the investor allocates in the risky asset and $(C_t)_{t\in[0,T]}$ denote the dynamic consumption rate by the investor. In this paper, we consider a drawdown constraint on the excess consumption rate $c_t := C_t - rX_t$ in the sense that c_t cannot go below a fraction $\alpha \in (0, 1)$ of its past maximum that

$$c_t \geq \alpha z_t, \quad t \in [0, T]. \tag{1}$$

Here, the non-decreasing reference process $(z_t)_{t\in[0,T]}$ is defined as the historical excessive spending maximum $z_t = \max\{z, \sup_{s\leq t} c_s\}$, and $z \geq 0$ is the initial reference level. The self-financing wealth process X_t is then governed by

$$dX_t = (\mu\pi_t - c_t)dt + \sigma\pi_t dW_t, \quad t \in [0, T], \tag{2}$$

with the initial wealth $X_0 = x \geq 0$. Let $\mathcal{A}(x)$ denote the set of admissible controls (π_t, c_t) if $(\pi_t)_{t\in[0,T]}$ is $\mathbb{F}$-progressively measurable, $(c_t)_{t\in[0,T]}$ is $\mathbb{F}$-predictable, the integrability condition $\mathbb{E}[\int_0^T (c_t + \pi_t^2)dt] < +\infty$ holds and the drawdown constraint $c_t \geq \alpha z_t$ is satisfied a.s. for all $t \in [0, T]$.

The goal of the agent is to maximize the expected utility of the excessive consumption rate up to time $T \wedge \tau$ under the drawdown constraint, where τ is the bankruptcy time defined by $\tau := \inf\{t \geq 0 \mid X_t \leq 0\}$. The value function of the stochastic control problem over a finite time horizon is given by

$$V(x, z, t) = \sup_{(\pi,c)\in\mathcal{A}(x)} \mathbb{E}\left[\int_t^{T\wedge\tau} e^{-\delta(s-t)}\frac{c_s^{1-p}}{1-p}ds + e^{-\delta(T\wedge\tau-t)}\frac{X_{T\wedge\tau}^{1-p}}{1-p}\middle| X_t = x, z_t = z\right], \tag{3}$$

where the constant $\delta \geq 0$ denotes the subjective time preference parameter, and $0 < p < 1$ stands for the agent's relative risk aversion coefficient.

2.2. The control problem and main result

By heuristic dynamic programming arguments and the martingale optimality condition, we note that the term $\partial_z V = 0$ holds whenever the monotone process z_t is strictly increasing, and the associated HJB variational inequality can be written as

$$\begin{cases} \max\Big\{ \sup_{\substack{\pi\in\mathbb{R},\\ \alpha z\leq c\leq z}} \Big[\partial_t V + \frac{1}{2}\sigma^2\pi^2\partial_{xx}V + (\mu\pi - c)\partial_x V \\ \qquad -\delta V + \frac{c^{1-p}}{1-p}\Big], \partial_z V \Big\} = 0, \ (x,z,t)\in Q, \\ V(0,z,t) = 0, \qquad z>0, t\in[0,T), \\ V(x,z,T) = \frac{x^{1-p}}{1-p}, \qquad x\geq 0,\ z>0, \end{cases} \tag{4}$$

where $Q := (0,+\infty)\times(0,+\infty)\times[0,T)$.

It is easy to see that $V(\beta x, \beta z, t) = \beta^{1-p}V(x,z,t)$, and we can consider

$$\omega := \frac{x}{z} \geq 0, \tag{5}$$

to reduce the dimension that $V(x,z,t) = z^{1-p}U(\omega,t)$. Moreover, let us consider the auxiliary controls $\hat{\pi}(\omega,t) = \pi(x,z,t)/z$ and $\hat{c}_t(\omega,t) = c(x,z,t)/z$. The HJB equation (4) can be rewritten as

$$\begin{cases} \max\Big\{ \partial_t U + \sup_{\hat{\pi}\in\mathbb{R}}\big\{\frac{1}{2}\sigma^2\hat{\pi}^2\partial_{\omega\omega}U + \mu\hat{\pi}\partial_\omega U\big\}, \\ \qquad + \sup_{\alpha\leq\hat{c}\leq 1}\Big\{\frac{\hat{c}^{1-p}}{1-p} - \hat{c}\partial_\omega U\Big\} - \delta U(1-p)U - \omega\partial_\omega U \Big\} = 0, \\ \qquad\qquad (\omega,t)\in\mathcal{Q}, \\ U(0,t) = 0, \quad t\in[0,T), \\ U(\omega,T) = \frac{\omega^{1-p}}{1-p}, \quad \omega\geq 0, \end{cases} \tag{6}$$

where $\mathcal{Q} := \mathbb{R}^+ \times [0,T)$.

We first present the main result of the paper, and its proof is deferred to Section 4.

Theorem 1 (*Verification Theorem*).
There exists a unique classical solution $U(\omega,t) \in C^{2,1}(\mathcal{Q})$ to problem (6), and $V(x,z,t) := z^{1-p}U(\frac{x}{z},t) \in C^{2,1}(Q)\cap C(\bar{Q})$ is the unique classical solution to problem (4). The optimal feedback controls are given by

$$\pi^*(x,z,t) = z\hat{\pi}^*\left(\frac{x}{z},t\right) = z\left[-\frac{\mu}{\sigma^2}\frac{\partial_\omega U(\frac{x}{z},t)}{\partial_{\omega\omega}U(\frac{x}{z},t)}\right], \tag{7}$$

$$c^*(x,z,t) = z\hat{c}^*\left(\frac{x}{z},t\right) \tag{8}$$

$$= \begin{cases} \alpha z, & \text{if } 0 < x \leq \omega_\alpha(t)z, \\ (\partial_\omega U)^{-\frac{1}{p}}\left(\frac{x}{z},t\right)z, & \text{if } \omega_\alpha(t)z < x < \omega_1(t)z, \\ z, & \text{if } \omega_1(t)z \leq x \leq \omega^*(t)z, \end{cases}$$

$$\mathcal{D} = \{(x,z,t)\in Q \mid \partial_z V(x,z,t) = 0\}$$
$$= \{(x,z,t)\in Q \mid x \geq \omega^*(t)z\}, \tag{9}$$
$$\mathcal{C} = \{(x,z,t)\in Q \mid \partial_z V(x,z,t) < 0\}$$
$$= \{(x,z,t)\in Q \mid x < \omega^*(t)z\}, \tag{10}$$

where $\omega_\alpha(t)$, $\omega_1(t)$ and $\omega^*(t)$ are free boundaries to problem (12), which are characterized analytically in Theorem 12.

Remark 2. Motivated by the result in Jeon and Oh (2022), we conjecture and will verify later in our model that there exists a time-dependent free boundary $\omega^*(t)$, which is the critical threshold of the wealth-to-consumption-peak ratio under the optimal control (π^*, c^*) such that if $X_t^* > \omega^*(t)z_{t-}^*$, the resulting z_t^* immediately jumps from z_{t-}^* to a new global maximum level $z_t^* = X_t^*/\omega^*(t)$. If $X_t^* \leq \omega^*(t)z_t^*$, the agent chooses the excessive consumption rate staying in $[\alpha z_t^*, z_t^*]$.

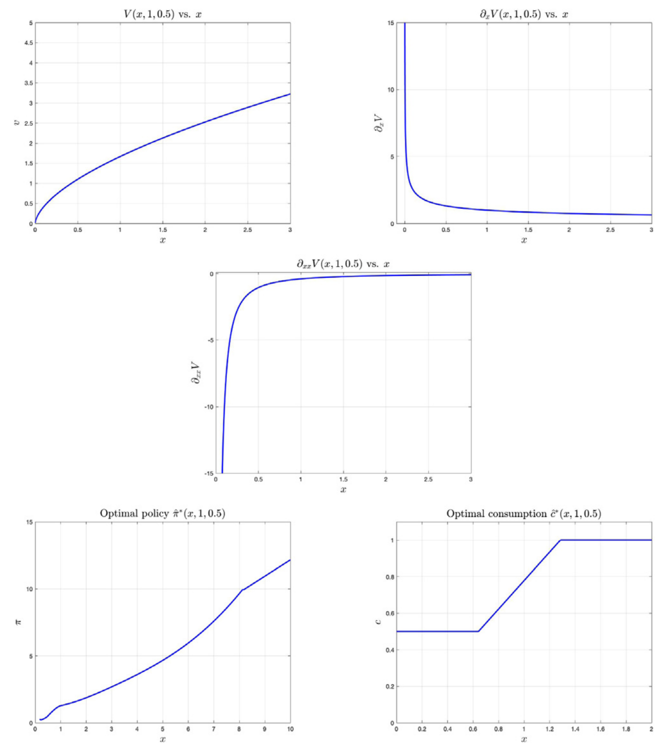

Fig. 1. The numerical illustration of the value function and its derivative with respect to x (top panel) The numerical illustration of the optimal portfolio and the optimal consumption with respect to x (bottom panel).

We plot the numerical illustrations of the value function, the optimal feedback functions of portfolio and consumption in terms of the wealth variable x while fixing $z = 1$ and $t = 0.5$ in Fig. 1. As shown in Fig. 1, the value function is strictly increasing and concave in the wealth variable x. More importantly, both optimal feedback functions of portfolio and consumption rate are increasing in x. However, comparing with the Merton's solution, the optimal portfolio is no longer a constant proportion strategy of the wealth level, in fact, it is even not simply concave or convex in the wealth variable depending on the ratio of the wealth and the past consumption peak. The right-bottom panel also illustrates the piecewise consumption behavior in Theorem 1: when $x/z \leq \omega_\alpha(t)$, the optimal consumption $\hat{c}^*(\frac{x}{z},t) = \alpha$; when $\omega_\alpha(t) < x/z < \omega_1(t)$, the optimal consumption is the first order condition $\hat{c}^*(\frac{x}{z},t) = (\partial_\omega U)^{-\frac{1}{p}}(\frac{x}{z},t)$; and when $\omega_1(t) \leq x/z \leq \omega^*(t)$, the optimal consumption is equal to the historical consumption peak.

2.3. The dual variational inequality

We first choose the candidate optimal feedback control $\hat{\pi}^*(\omega,t)$ by the first order condition that $\hat{\pi}^*(\omega,t) := -\frac{\mu}{\sigma^2}\frac{\partial_\omega U}{\partial_{\omega\omega}U} \geq 0$. Considering the constraint $\alpha \leq \hat{c} \leq 1$, we can choose the candidate optimal feedback control

$$\hat{c}^*(\omega,t) := \max\Big\{\alpha, \min\big\{1, (\partial_\omega U)^{-\frac{1}{p}}\big\}\Big\}. \tag{11}$$

Then (6) becomes

$$\begin{cases} \max\left\{ \partial_t U - \frac{1}{2}\frac{\mu^2}{\sigma^2}\frac{(\partial_\omega U)^2}{\partial_{\omega\omega} U} + \frac{(\hat{c}^*(\omega,t))^{1-p}}{1-p} - \hat{c}^*(\omega,t)\partial_\omega U \right. \\ \qquad \left. -\delta U, (1-p)U - \omega\partial_\omega U \right\} = 0, \quad (\omega,t) \in \mathcal{Q}, \\ U(0,t) = 0, \qquad t \in [0,T), \\ U(\omega,T) = \frac{\omega^{1-p}}{1-p}, \quad \omega \geq 0. \end{cases} \tag{12}$$

For the conjectured free boundary $\omega^*(t)$ in Remark 2 under the optimal control (π^*, c^*), using the relationship between $V(x,z,t)$ and $U(\omega,t)$, $\mathcal{Q}$ can also be divided into two regions, namely the continuation region and the jump region denoted by CR and JR (see Fig. 1) that

$$\begin{aligned} CR &:= \left\{(\omega,t) | (1-p)U - \omega\partial_\omega U < 0\right\} \\ &= \left\{(\omega,t) | \omega < \omega^*(t)\right\}, \end{aligned} \tag{13}$$

$$\begin{aligned} JR &:= \left\{(\omega,t) | (1-p)U - \omega\partial_\omega U = 0\right\} \\ &= \left\{(\omega,t) | \omega \geq \omega^*(t)\right\}. \end{aligned} \tag{14}$$

Later, we will rigorously characterize $\omega^*(t)$ and two regions CR and JR in analytical form in Theorem 12.

We employ the convex dual transform that $u(y,t) = \max_{\omega>0}[U(\omega,t) - \omega y]$. Let $y_0(t) := \partial_\omega U(0,t)$, it follows from the standard dual representation that $U(\omega,t) = \min_{y\in(0,y_0(t))}[u(y,t) + y\omega]$. The dual problem of (12) can be written as

$$\begin{cases} \max\left\{ \partial_t u + \frac{1}{2}\frac{\mu^2}{\sigma^2} y^2 \partial_{yy} u + \delta y \partial_y u - \delta u - f(y), \right. \\ \qquad \left. (1-p)u + py\partial_y u \right\} = 0, \\ \qquad (y,t) \in (0, y_0(t)) \times (0,T], \\ u(y_0(t),t) = \partial_y u(y_0(t),t) = 0, \\ u(y,T) = \frac{p}{1-p} y^{1-\frac{1}{p}}, \quad y \geq 0, \end{cases} \tag{15}$$

where

$$f(y) = \begin{cases} \alpha y - \frac{\alpha^{1-p}}{1-p}, & \text{if } y \geq \alpha^{-p}, \\ -\frac{p}{1-p} y^{1-\frac{1}{p}}, & \text{if } 1 < y < \alpha^{-p}, \\ y - \frac{1}{1-p}. & \text{if } y \leq 1, \end{cases}$$

Moreover, we define $y^*(t) = \partial_\omega U(\omega^*(t),t)$.

3. The solution to the dual variational inequality (15)

3.1. Auxiliary dual variational inequality

Taking advantage of the boundary condition of $u(y,t)$ on $y_0(t)$, we can expand the solution $u(y,t)$ to the problem (15) from $(0,y_0(t)) \times (0,T]$ to the enlarged domain $\mathcal{Q} = \mathbb{R}^+ \times [0,T)$. Let us consider $\hat{u}(y,t)$ that satisfies the auxiliary variational inequality on $\mathcal{Q}$ that

$$\begin{cases} \max\left\{ \partial_t \hat{u} + \frac{1}{2}\frac{\mu^2}{\sigma^2} y^2 \partial_{yy}\hat{u} + \delta y \partial_y \hat{u} - \delta\hat{u} - f(y), \right. \\ \qquad \left. (1-p)\hat{u} + py\partial_y\hat{u},\ -\hat{u} \right\} = 0, \quad (y,t) \in \mathcal{Q}, \\ \hat{u}(y,T) = \frac{p}{1-p} y^{1-\frac{1}{p}}, \quad y \geq 0. \end{cases} \tag{16}$$

For the variational inequality above, we consider the following regimes and associated free boundaries:

$\mathcal{F} := \left\{(y,t) \in \mathcal{Q} \mid \hat{u}(y,t) = 0\right\}$
(Function constraint region),

$\mathcal{G} := \left\{(y,t) \in \mathcal{Q} \mid (1-p)\hat{u}(y,t) + py\partial_y\hat{u}(y,t) = 0\right\}$
(Gradient constraint region),

$\mathcal{E} := \left\{ (y,t) \in \mathcal{Q} \mid (1-p)\hat{u}(y,t) + py\partial_y\hat{u}(y,t) < 0, \ \hat{u}(y,t) > 0 \right\}$ (Equation region).

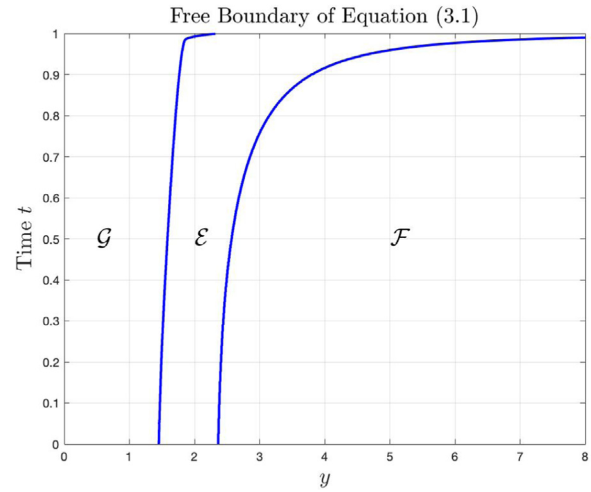

Fig. 2. The numerical illustration of free boundaries in the variational inequality (3.1).

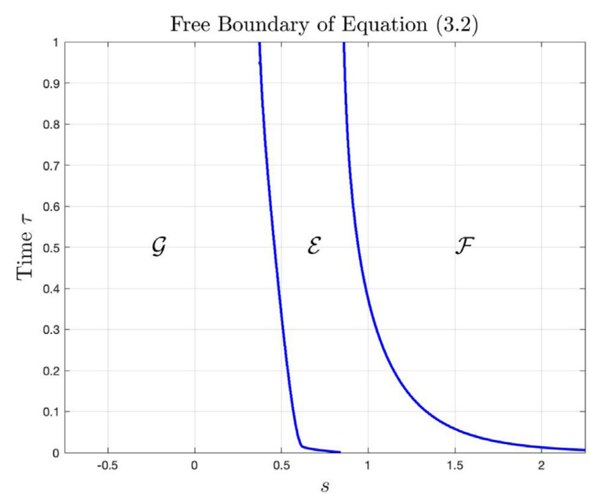

Fig. 3. The numerical illustration of free boundaries in the variational inequality (3.2).

We plot in Fig. 2 the numerical illustration of these free boundaries.

To simplify some analysis, let us further set $y = e^s$, $\tau = T - t$ and $\tilde{u}(s,\tau) = \hat{u}(y,t)$. It follows that $\tilde{u}(s,\tau)$ satisfies

$$\begin{cases} \min\left\{ \partial_\tau \tilde{u} - \frac{1}{2}\frac{\mu^2}{\sigma^2}\partial_{ss}\tilde{u} - (\delta - \frac{1}{2}\frac{\mu^2}{\sigma^2})\partial_s\tilde{u} + \delta\tilde{u} + \tilde{f}(s), \right. \\ \qquad \left. (p-1)\tilde{u} - p\partial_s\tilde{u},\ \tilde{u} \right\} = 0, \quad (s,\tau) \in \Omega, \\ \tilde{u}(s,0) = \frac{p}{1-p} e^{\frac{p-1}{p}s}, \quad s \in \mathbb{R}, \end{cases} \tag{17}$$

where $\Omega = (-\infty, +\infty) \times (0,T]$ and

$$\tilde{f}(s) = \begin{cases} \alpha e^s - \frac{\alpha^{1-p}}{1-p}, & \text{if } s \geq -p\ln\alpha, \\ -\frac{p}{1-p} e^{\frac{p-1}{p}s}, & \text{if } 0 < s < -p\ln\alpha, \\ e^s - \frac{1}{1-p}, & \text{if } s \leq 0. \end{cases}$$

The transformed regimes and the associated free boundaries are plotted in Fig. 3.

Considering the transform $v(s,\tau) := e^{\frac{1-p}{p}s}\tilde{u}(s,\tau)$, we can work with another auxiliary dual variational inequality of $v(s,\tau)$ that

$$\begin{cases} \min\left\{\partial_\tau v - \mathcal{L}_s v + e^{\frac{1-p}{p}s}\tilde{f}(s), -\partial_s v(s,\tau), v(s,\tau)\right\} = 0, \\ \qquad (s,\tau) \in \Omega, \\ v(s,0) = \frac{p}{1-p}, \quad s \in \mathbb{R}, \end{cases} \tag{18}$$

where $\mathcal{L}_s v := \frac{1}{2}\frac{\mu^2}{\sigma^2}\partial_{ss}v - \left(\frac{2-p}{2p}\frac{\mu^2}{\sigma^2} - \delta\right)\partial_s v - \left(\frac{1}{p}\delta - \frac{\mu^2(1-p)}{2p^2\sigma^2}\right)v$.

3.2. Characterization of the free boundary in (18)

Let us first analyze the auxiliary variational inequality (18), which is a parabolic variational inequality with both gradient constraint and function constraint. To this end, we first consider the following problem, for any $N > 0$, v_N satisfies

$$\begin{cases} \min\Big\{ \partial_\tau v_N - \mathcal{L}_s v_N + e^{\frac{1-p}{p}s}\tilde{f}(s), -\partial_s v_N(s,\tau) \\ \quad , v_N(s,\tau)\Big\} = 0, \quad (s,\tau) \in \Omega_N, \\ v_N(s,0) = \frac{p}{1-p}, \quad s > -N, \\ v_N(-N,\tau) = \frac{p}{1-p}, \quad \tau \in [0,T], \end{cases} \tag{19}$$

where $\Omega_N = (-N, +\infty) \times (0,T]$.

Lemma 3. *There exists a solution $v_N(s,\tau) \in W^{2,1}_{q,loc}(\Omega_N) \cap C(\overline{\Omega_N})$ to problem (19). Moreover, we have*

$$0 \le v_N(s,\tau) \le \frac{p}{1-p}, \quad (s,\tau) \in \Omega_N, \tag{20}$$

$$\partial_\tau v_N(s,\tau) \le 0, \quad (s,\tau) \in \Omega_N, \tag{21}$$

$$\lim_{s\to+\infty} v_N(s,\tau) = 0., \tag{22}$$

and the solution satisfying (20) is unique.

Proof. We can solve the problem (19) using the standard penalty approximation method. Consider v_N^ε satisfies the penalty problem

$$\begin{aligned} &\partial_\tau v_N^\varepsilon - \mathcal{L}_s v_N^\varepsilon + e^{\frac{1-p}{p}s}\tilde{f}(s) + \beta_\varepsilon(v_N^\varepsilon) \\ &+ s\beta_\varepsilon(-\partial_s v_N^\varepsilon) = 0, \quad (s,\tau) \in \Omega_N, \\ &v_N^\varepsilon(s,0) = \frac{p}{1-p}, \quad s > -N, \\ &v_N^\varepsilon(-N,\tau) = \frac{p}{1-p}, \quad \tau \in [0,T], \end{aligned} \tag{23}$$

where $\beta_\varepsilon(\lambda)$ is the penalty function satisfying

$$\beta_\varepsilon(\cdot) \in C^2(-\infty,+\infty), \quad \beta_\varepsilon(\cdot) \le 0, \quad \beta_\varepsilon'(\cdot) \ge 0,$$

$$\lim_{\varepsilon\to0}\beta_\varepsilon(\lambda) = \begin{cases} 0, & \lambda > 0, \\ -\infty, & \lambda < 0. \end{cases}$$

Using the standard fixed point theorem, we are able to show that there exists a solution $v_N^\varepsilon(s,\tau) \in W^{2,1}_{q,loc}(\Omega_N) \cap C(\overline{\Omega_N})$ to the penalty problem (23). Letting $\varepsilon \to 0$, we obtain a solution $v_N(s,\tau) \in W^{2,1}_{q,loc}(\Omega_N) \cap C(\overline{\Omega_N})$ to problem (19).

The estimate (20) follows from the facts $\partial_s v_N \le 0$, $v_N \ge 0$ and the boundary condition $v_N(-N,\tau) = \frac{p}{1-p}$. Moreover, based on the boundary condition $v_N(-N,\tau) = \frac{p}{1-p}$, we have the uniqueness of the solution to problem (19) and then the comparison principle for problem (19) holds true.

Now we will show (21). For any $0 < \delta < T$, set $v_\delta(s,\tau) = v_N(s,\tau+\delta)$, then $v_\delta(s,\tau)$ satisfies

$$\begin{cases} \min\left\{\partial_\tau v_\delta - \mathcal{L}_s v_\delta + e^{\frac{1-p}{p}s}\tilde{f}(s), -\partial_s v_\delta, v_\delta\right\} = 0, \\ \qquad (s,\tau) \in (-N,+\infty) \times (0, T-\delta], \\ v_\delta(s,0) = v_N(s,\delta) \le \frac{p}{1-p} = v_N(s,0), \quad s > -N, \\ v_\delta(-N,\tau) = \frac{p}{1-p} = v_N(-N,\tau), \quad \tau \in [0, T-\delta]. \end{cases} \tag{24}$$

Using the comparison principle between (19) and (24), we have that $v_N(s,\tau+\delta) = v_\delta(s,\tau) \le v_N(s,\tau)$, which leads to the desired result (21).

For each $\tau > 0$, define $Z_N(\tau) := \min\{s | v_N(s,\tau) = 0\}$. We next show that $Z_N(\tau)$ is finite for $\tau \in (0,T]$. Suppose that there exists a $\tau_0 > 0$ such that $Z_N(\tau_0) = +\infty$. Together with (21), it holds that $v_N(s,\tau) > 0$ for $(s,\tau) \in (-N,+\infty) \times (0,\tau_0)$, which implies

$$\begin{cases} \min\left\{\partial_\tau v_N - \mathcal{L}_s v_N + e^{\frac{1-p}{p}s}\tilde{f}(s), -\partial_s v_N(s,\tau)\right\} = 0, \\ \qquad (s,\tau) \in (-N,+\infty) \times (0,\tau_1), \\ v_N(s,0) = \frac{p}{1-p}, \quad s > -N, \\ v_N(-N,\tau) = \frac{p}{1-p}, \quad \tau \in [0,\tau_1], \end{cases} \tag{25}$$

where $0 < \tau_1 \le \tau_0$.

Note that $e^{\frac{1-p}{p}s}\tilde{f}(s) = e^{\frac{1}{p}s}\left(\alpha - \frac{\alpha^{1-p}}{1-p}e^{-s}\right) \to +\infty$ as $s \to +\infty$, it follows that there exists a constant s_0 such that $e^{\frac{1-p}{p}s}\tilde{f}(s) \ge \frac{3\alpha}{4}e^{\frac{1}{p}s}$ for $s \ge s_0$. Let $\tilde{v}(s,\tau) = \frac{p}{1-p} - \frac{\alpha}{2}(e^{\frac{1}{p}s} - e^{\frac{1}{p}s_0})\tau$. We next show that $\tilde{v}(s,\tau)$ is a super-solution to (25) on $(s_0,+\infty) \times (0,\tau_1)$.

When τ_1 is small enough, we can deduce

$$\begin{aligned} &\partial_\tau\tilde{v} - \mathcal{L}_s\tilde{v} + e^{\frac{1-p}{p}s}\tilde{f}(s) \\ &= -\frac{\alpha}{2}\left(e^{\frac{1}{p}s} - e^{\frac{1}{p}s_0}\right) + \left(\frac{\delta}{p} - \frac{\mu^2(1-p)}{2p\sigma^2}\right) \\ &\quad \times \left(\frac{p}{1-p} + \frac{\alpha}{2}e^{\frac{1}{p}s_0}\tau\right) + e^{\frac{1-p}{p}s}\tilde{f}(s) \ge 0. \end{aligned}$$

Together with $v_N(s,\tau) \le \frac{p}{1-p}$, we know that

$$\begin{cases} \min\left\{\partial_\tau\tilde{v} - \mathcal{L}_s\tilde{v} + e^{\frac{1-p}{p}s}\tilde{f}(s), -\partial_s\tilde{v}(s,\tau)\right\} = 0, \\ \qquad (s,\tau) \in (s_0,+\infty) \times (0,\tau_1), \\ \tilde{v}(s,0) = \frac{p}{1-p}, \quad s > s_0, \\ \tilde{v}(-N,\tau) = \frac{p}{1-p} \ge v_N(s_0,\tau), \quad \tau \in [0,\tau_1]. \end{cases}$$

The comparison principle implies that $v_N(s,\tau) \le \tilde{v}(s,\tau)$ for $(s,\tau) \in (s_0,+\infty) \times (0,\tau_1]$. Moreover, we have $\tilde{v}(s,\tau) < 0$ for $s > p\ln\left(e^{\frac{1}{p}s_0} + \frac{2}{\alpha}\frac{p}{(1-p)\tau_1}\right)$, and hence $v_N(s,\tau) < 0$ for $s > p\ln\left(e^{\frac{1}{p}s_0} + \frac{2}{\alpha}\frac{p}{(1-p)\tau_1}\right)$, which is a contradiction. As a result, $Z_n(\tau)$ is finite for each $\tau > 0$, and it implies that (22) holds.

Proposition 4. *There exists a solution $v(s,\tau) \in W^{2,1}_{q,loc}(\Omega) \cap C(\overline{\Omega})$ to problem (18). Moreover, we have*

$$0 \le v(s,\tau) \le \frac{p}{1-p}, \quad (s,\tau) \in \Omega, \tag{26}$$

$$\partial_\tau v(s,\tau) \le 0, \quad (s,\tau) \in \Omega, \tag{27}$$

$$\lim_{s\to+\infty} v(s,\tau) = 0. \tag{28}$$

The solution to problem (18) satisfying (26) to (28) is unique.

Proof. Let $v(s,\tau) = \lim_{N\to+\infty} v_N(s,\tau)$, then $v(s,\tau) \in W^{2,1}_{q,loc}(\Omega) \cap C(\overline{\Omega})$ is the solution of problem (18). It is easy to see that (26) and (27) can be derived from (20) and (21). Let $N_1 > N_2$, then $v_{N_1}(-N_2,\tau) \le \frac{p}{1-p} = v_{N_2}(-N_2,\tau)$. By the comparison principle, we have $v_{N_1}(s,\tau) \le v_{N_2}(s,\tau)$ for $(s,\tau) \in (-N_2,+\infty) \times (0,T]$. It hence holds that $v_{N_1}(s,\tau) = 0$ for $(s,\tau) \in (Z_{N_2},+\infty) \times (0,T]$, which implies that $Z_N(\tau)$ is decreasing in N. This, together with (22), implies (28).

We are ready to show the uniqueness of the solution to problem (18) by contradiction. Suppose v_1, v_2 are two solutions to the problem (18). Denote $\mathcal{N} = \{v_1 > v_2\} \neq \emptyset$, and let $\tau_0 = \inf\{\tau : (s,\tau) \in \mathcal{N}\}$, $A = \partial_p\mathcal{N} \cap \{(s,\tau) : \tau = \tau_0\}$. It then holds that $v_1 = v_2$ on A. Denote $\mathcal{N}_1 = \mathcal{N} \cap \{(s,\tau) : \partial_s v_1 < \partial_s v_2\}$

and $\mathcal{N}_2 = \mathcal{N} \cap \{(s,\tau) : \partial_s v_1 \geq \partial_s v_2\}$. Suppose $\mathcal{N}_1 \neq \emptyset$, by condition (28), it is easy to show that there exists a point $(s_0, \tau_0) \in \partial_p \mathcal{N}_1 \cap A$, which implies $v_1(s_0,\tau_0) = v_2(s_0,\tau_0)$, and we have

$$\begin{cases} \partial_\tau v_1 - \mathcal{L}_s v_1 + e^{\frac{1-p}{p}s}\tilde{f}(s) = 0, & (s,\tau) \in \mathcal{N}_1, \\ \partial_\tau v_2 - \mathcal{L}_s v_2 + e^{\frac{1-p}{p}s}\tilde{f}(s) \geq 0, & (s,\tau) \in \mathcal{N}_1, \\ v_1 = v_2 \text{ or } \partial_s v_1 = \partial_s v_2, & (s,\tau) \in \partial_p \mathcal{N}_1. \end{cases}$$

By condition (26) and the maximum principle, we know $v_2(s,\tau) \geq v_1(s,\tau)$ for $(s,\tau) \in \mathcal{N}_1$, which contradicts the definition of $\mathcal{N}$ and hence $\mathcal{N} \subset \mathcal{N}_2$.

We then conclude that $\partial_s v_1 \geq \partial_s v_2$ for $(s,\tau) \in \mathcal{N}$, and $v_1 = v_2$ for $(s,\tau) \in \partial_p \mathcal{N}$. This, together with the condition (28), implies that $v_1(s,\tau) \leq v_2(s,\tau)$ for $(s,\tau) \in \mathcal{N}$, which is a contradiction to the definition of $\mathcal{N}$. The proof is then complete.

As a direct result of the uniqueness of solution to problem (18), the comparison principle for problem (18) holds. Note that $v(s,\tau)$ satisfies $\partial_s v(s,\tau) \leq 0$ and $v(s,\tau) \geq 0$. For each $\tau > 0$, let us define

$$Z(\tau) := \inf\{s \mid v(s,\tau) = 0\}. \tag{29}$$

Proposition 5. *The curve $Z(\tau)$ defined in (29) satisfies $-p\ln\alpha < Z(\tau) < +\infty$ for $\tau \in (0,T]$ and*

$$\mathcal{F} = \{(s,\tau) \in \mathcal{Q} \mid s \geq Z(\tau)\}. \tag{30}$$

Moreover, $Z(\tau)$ strictly decreases in τ. In particular,

$$\lim_{\tau \to 0^+} Z(\tau) = +\infty. \tag{31}$$

Proof. The result (30) follows from definitions of $Z(\tau)$ and $\mathcal{F}$. By the variational inequality (18), we have

$$\partial_\tau v - \mathcal{L}_s v + e^{\frac{1-p}{p}s}\tilde{f}(s) \geq 0, \quad \text{if } v = 0,$$

which leads to $s \geq -p\ln\alpha - \ln(1-p)$. Hence, by the definition of $Z(\tau)$, we know $Z(\tau) > -p\ln\alpha$.

Next, we show that $Z(\tau)$ is finite for $\tau \in (0,T]$. Suppose that there exists a $\tau_0 > 0$ such that $Z(\tau_0) = +\infty$. It then holds that $v(s,\tau) > 0, (s,\tau) \in \mathbb{R} \times (0,\tau_0]$, which implies that $v(s,\tau)$ satisfies

$$\begin{cases} \min\left\{\partial_\tau v - \mathcal{L}_s v + e^{\frac{1-p}{p}s}\tilde{f}(s), -\partial_s v(s,\tau)\right\} = 0, \\ \qquad (s,\tau) \in \mathbb{R} \times (0,\tau_0], \\ v(s,0) = \frac{p}{1-p}, \quad s \in \mathbb{R}. \end{cases} \tag{32}$$

Note that $e^{\frac{1-p}{p}s}\tilde{f}(s) = \alpha e^{\frac{1}{p}s} - \frac{\alpha^{1-p}}{1-p}e^{\frac{1-p}{p}s} \to +\infty$ as $s \to +\infty$. There exists a constant s_0 such that $e^{\frac{1-p}{p}s}\tilde{f}(s) \geq \frac{\alpha}{2}e^{\frac{1}{p}s}$ for $s \geq s_0$. Let $\tilde{v}(s,\tau) := \frac{p}{1-p} - \frac{\alpha}{2}(e^{\frac{1}{p}s} - e^{\frac{1}{p}s_0})\tau$. We next show that $\tilde{v}(s,\tau)$ is a super-solution to (32) on $(s_0,+\infty) \times (0,\tau_0)$. In view that

$$\begin{aligned} &\partial_\tau \tilde{v} - \mathcal{L}_s \tilde{v} + e^{\frac{1-p}{p}s}\tilde{f}(s) \\ &\quad = -\frac{\alpha}{2}(e^{\frac{1}{p}s} - e^{\frac{1}{p}s_0}) + \left(\frac{\delta}{p} - \frac{\mu^2(1-p)}{2p^2\sigma^2}\right) \\ &\quad \times \left(\frac{p}{1-p} + \frac{\alpha}{2}e^{\frac{1}{p}s_0}\tau\right) + e^{\frac{1-p}{p}s}\tilde{f}(s) \geq 0. \end{aligned}$$

Together with $\frac{p}{1-p} \geq v(s,\tau)$, we know that $\tilde{v}(s,\tau)$ satisfies

$$\begin{cases} \min\left\{\partial_\tau \tilde{v} - \mathcal{L}_s \tilde{v} + e^{\frac{1-p}{p}s}\tilde{f}(s), -\partial_s \tilde{v}(s,\tau)\right\} \geq 0, \\ \qquad (s,\tau) \in (s_0,+\infty) \times (0,\tau_0], \\ \tilde{v}(s,0) = \frac{p}{1-p}, \quad s \in (s_0,+\infty), \\ \tilde{v}(s_0,\tau) = \frac{p}{1-p} \geq v(s_0,\tau), \quad \tau \in (0,\tau_0]. \end{cases}$$

The comparison principle implies that $v(s,\tau) \leq \tilde{v}(s,\tau)$ for $(s,\tau) \in (s_0,+\infty) \times (0,\tau_0]$. Moreover, it is easy to see that $\tilde{v}(s,\tau) < 0$ for $s > p\ln\left(e^{\frac{1}{p}s_0} + \frac{2}{\alpha}\frac{p}{(1-p)\tau_0}\right)$, which implies that $v(s,\tau) < 0$ for $s > p\ln\left(e^{\frac{1}{p}s_0} + \frac{2}{\alpha}\frac{p}{(1-p)\tau_0}\right)$, leading to a contradiction.

Finally, we show the strict monotonicity of $Z(\tau)$. For any $\tau_0 > 0$, suppose $Z(\tau_0) = s_0$, i.e. $v(s_0,\tau_0) = 0$. According to $\partial_\tau v(s,\tau) \leq 0$, we have $v(s_0,\tau) = 0$, $\tau \in (\tau_0,T]$. By the definition (29) of $Z(\tau)$, we obtain $Z(\tau) \leq s_0$ for $\tau \in (\tau_0,T]$. Suppose that $Z(\tau)$ is not strictly monotone and there exists $\tau_1 < \tau_2$, such that $Z(\tau) = s_0$ for $\tau \in [\tau_1,\tau_2]$. Denote $\Gamma := \{s_0\} \times (\tau_1,\tau_2)$. Then we have

$$\begin{cases} \partial_\tau v - \mathcal{L}_s v = -e^{\frac{1-p}{p}s}\tilde{f}(s), (s,\tau) \in (s_0-\varepsilon, s_0) \times (\tau_1,\tau_2), \\ v|_\Gamma = \partial_s v|_\Gamma = 0, \end{cases}$$

where ε is small enough. It then follows that

$$\begin{cases} \partial_\tau(\partial_\tau v) - \mathcal{L}_s(\partial_\tau v) = 0, \quad (s,\tau) \in (s_0-\varepsilon, s_0) \times (\tau_1,\tau_2), \\ \partial_\tau v|_\Gamma = \partial_{s\tau} v|_\Gamma = 0. \end{cases}$$

Together with the fact that $\partial_\tau v \leq 0$, Hopf's principle implies that

$$\partial_{s\tau} v|_\Gamma > 0 \quad \text{or} \quad \partial_\tau v \equiv 0, \quad (s,\tau) \in (s_0, s_0+\varepsilon) \times (\tau_1,\tau_2),$$

leading to a contradiction. Moreover, it follows from the monotonicity of $Z(\tau)$ and $v(s,0) = \frac{p}{1-p} > 0$ that the result (31) holds.

Let us next focus on the domain $\{s \leq Z(\tau)\}$. In view of the definition of $Z(\tau)$, $v(s,\tau)$ is a unique $W^{2,1}_{q,loc}(\Omega)$ solution to problem (18). On the domain $\{s \leq Z(\tau)\}$, $v(s,\tau)$ satisfies

$$\begin{cases} \min\left\{\partial_\tau v - \mathcal{L}_s v + e^{\frac{1-p}{p}s}\tilde{f}(s), -\partial_s v(s,\tau)\right\} = 0, \\ \qquad (s,\tau) \in \tilde{\Omega}, \\ v(s,0) = \frac{p}{1-p}, \quad s \in \mathbb{R}, \\ \partial_s v(Z(\tau),\tau) = 0, \quad \tau \in (0,T], \end{cases} \tag{33}$$

where $\tilde{\Omega} := (-\infty, Z(\tau)) \times (0,T]$. In order to analyze the free boundary arising from gradient constraint, we follow the similar idea in Chen and Yi (2012) and consider the parabolic obstacle problem

$$\begin{cases} \max\{\partial_\tau w(s,\tau) - \mathcal{L}_s w(s,\tau) - g(s), w(s,\tau)\} = 0, \\ \qquad (s,\tau) \in \tilde{\Omega}, \\ w(s,0) = 0, \quad s \in \mathbb{R}, \\ w(Z(\tau),\tau) = 0, \quad \tau \in (0,T], \end{cases} \tag{34}$$

where $g(s) = \frac{\alpha}{p}e^{\frac{1-p}{p}s}(\alpha^{-p} - e^s)I_{\{e^s \geq \alpha^{-p}\}} + \frac{1}{p}e^{\frac{1-p}{p}s}(1-e^s)I_{\{e^s \leq 1\}}$.

3.3. Characterization of the free boundary in problem (34)

Following the standard penalty approximation method as to show the existence of solution to the problem (18), it is easy to conclude the next result, and its proof is hence omitted.

Lemma 6. *There exists a unique $w(s,\tau) \in W^{2,1}_{q,loc}(\tilde{\Omega}) \cap C(\overline{\tilde{\Omega}})$ to problem (34) for any $1 < q < +\infty$.*

To study some properties of the free boundary in (34), let us first define

$$\mathcal{G}_0 := \{(s,\tau) \in \tilde{\Omega} \mid w(s,\tau) = 0\},$$
$$\mathcal{E}_0 := \{(s,\tau) \in \tilde{\Omega} \mid w(s,\tau) < 0\}.$$

Proposition 7. *There exists a function $T(s)$ such that*

$$\mathcal{G}_0 = \{(s,\tau) \mid s \in \mathbb{R}, 0 \leq \tau \leq T(s)\}, \tag{35}$$

and $T(s)$ is decreasing in s such that

$$T(s) = 0, \quad s > 0, \tag{36}$$
$$T(s) > 0, \quad s < 0. \tag{37}$$

Particularly, $T(s)$ is strictly decreasing on $\{s \mid 0 < T(s) < T\}$ and $T(s)$ is continuous.

Proof. The conjectured free boundary $\omega^*(t)$ in Remark 2 and all the transformations in Section 2 imply that $\mathcal{G}_0$ is connected in τ direction. Note that $w(s,0)=0$, let us define $T(s) := \sup\{\tau \mid w(s,\tau)=0\}$, $\forall s \in \mathbb{R}$. By the definitions of $\mathcal{G}_0$ and $T(s)$, we get the desired result (35).

We next show the monotonicity of $T(s)$. By the variational inequality (34), we have $g(s) \ge 0$ if $w(s,\tau)=0$, which implies $s \le -p\ln\alpha$. That is, $w(s,\tau) < 0$ for $s > -p\ln\alpha$, $\tau \in (0,T]$. It follows that

$$\{(s,\tau) \mid s > -p\ln\alpha,\ \tau \in (0,T]\} \subset \mathcal{E}_0.$$

For any $s_0 \le -p\ln\alpha$ such that $T(s_0) > 0$, we define an auxiliary function

$$\tilde{w}(s,\tau) := \begin{cases} 0, & \text{if } (s,\tau) \in (-\infty,s_0] \times [0,T(s_0)], \\ w(s,\tau), & \text{if } (s,\tau) \in \{(s_0,+\infty) \times [0,T(s_0)]\} \cap \tilde{\Omega}. \end{cases}$$

We show that $\tilde{w}(s,\tau)$ is the solution to problem (34) in the domain $\{\mathbb{R} \times [0,T(s_0)]\}$. By the definition of $\tilde{w}(s,\tau)$, we have that $\tilde{w}(s,0)=0$, $\tilde{w}(s,\tau) \le 0$, and

$$\begin{cases} \partial_\tau \tilde{w} - \mathcal{L}_s \tilde{w} = 0 \le g(s), \\ \qquad \text{if } (s,\tau) \in (-\infty,s_0] \times [0,T(s_0)], \\ \partial_\tau \tilde{w} - \mathcal{L}_s \tilde{w} = \partial_\tau w - \mathcal{L}_s w \le g(s), \\ \qquad \text{if } (s,\tau) \in \{(s_0,+\infty) \times [0,T(s_0)]\} \cap \tilde{\Omega}. \end{cases}$$

Moreover, if $\tilde{w}(s,\tau) < 0$, then $w(s,\tau) < 0$. Hence, we have

$$\partial_\tau \tilde{w} - \mathcal{L}_s \tilde{w} = \partial_\tau w - \mathcal{L}_s w = g(s), \quad \tilde{w}(s,\tau) < 0.$$

Thus, $\tilde{w}(s,\tau)$ is a $W^{2,1}_{q,loc}$-solution to problem (34) in the domain $\{\mathbb{R} \times [0,T(s_0)]\} \cap \tilde{\Omega}$. The uniqueness of the solution to (34) yields that

$$w(s,\tau) = \tilde{w}(s,\tau) = 0, \quad (s,\tau) \in (-\infty,s_0] \times [0,T(s_0)].$$

By the definition of $T(s)$, we obtain $T(s) \ge T(s_0)$ for $\forall s < s_0$, and $T(s)$ is decreasing in s.

Next, we show (36). Suppose that there exists $s_0 > 0$ such that $T(s_0) = \tau_0 > 0$, by the monotonicity of $T(s)$, we have

$$\begin{cases} \partial_\tau w - \mathcal{L}_s w \le g(s) \le 0, \\ \qquad (s,\tau) \in \{[0,+\infty) \times [0,\tau_0]\} \cap \tilde{\Omega}, \\ w(0,\tau) = 0, \quad w(Z(\tau),\tau) = 0, \quad \tau \in (0,\tau_0], \\ w(s,0) = 0, \quad s \in \mathbb{R}^+. \end{cases}$$

The strong maximum principle implies that $w(s,\tau) < 0$ for $(s,\tau) \in (0,+\infty) \times (0,\tau_0]$. It contradicts with $w(s_0,\tau) = 0$ for $\tau \in (0,\tau_0]$. Hence, (36) holds true.

In view of the definition of $T(s)$ and the fact $w(s,\tau) \le 0$, we have

$$\partial_\tau w(s,0) \le 0, \quad \forall s \in \mathbb{R}, \tag{38}$$

$$\partial_\tau w(s,T(s)) \le 0, \quad \forall s \in \mathbb{R}. \tag{39}$$

Suppose that there exists $s_1 < 0$ such that $T(s_1) = 0$, then we have

$$\partial_\tau w(s,0) = \mathcal{L}_s w(s,0) + g(s) > 0, \quad s \in (s_1,0),$$

yielding a contraction to (38). Hence (37) follows.

Thanks to (39), we claim the strict monotonicity of $T(s)$ in $\{s \mid 0 < T(s) < T\}$. Indeed, suppose that exists $s_1 < s_2 \le 0$ such that $0 < T(s_1) = T(s_2) < T$. Then we have $w(s,T(s_2)) = 0$, $s \in (s_1,s_2)$. Applying the equation $\partial_\tau w - \mathcal{L}_s w = g(s)$ at $(s_1,s_2) \times \{T(s_2)\}$, we have $\partial_\tau w|_{\tau=T(s_2)} = (\mathcal{L}_s w + g(s))|_{\tau=T(s_2)} > 0$, $s \in (s_1,s_2)$, which contradicts with (39). The claim therefore holds. Following the proof of the strictly monotonicity of $Z(\tau)$ in Proposition 7, we can conclude the continuity of $T(s)$.

As $T(s)$ strictly decreases and is continuous in s on $0 < T(s) < T$, there exists an inverse function of $T(s)$ denoted by $S(\tau) := T^{-1}(\tau)$, $\tau \in (0,T)$. Let us define

$$S(\tau) := \begin{cases} T^{-1}(\tau), & \text{if } \tau \in (0,T), \\ \sup\{s \mid T(s) = T\}, & \text{if } \tau = T. \end{cases} \tag{40}$$

3.4. The solution to problem (17)

In this subsection, we first use the solution to problem (34) to construct the solution to problem (33), and then obtain the solution to problem (18). Using the transform between $\tilde{u}(s,\tau)$ and $v(s,\tau)$, we can further obtain the solution to problem (17). Following the same proof of Theorem 4.6 in Chen and Yi (2012), we can get the next result.

Proposition 8. *Let $w(s,\tau)$ be the solution to problem (34) and let us define*

$$\bar{v}(s,\tau) := \int_{Z(\tau)}^{s} w(\xi,\tau)d\xi + \frac{p}{1-p}\chi\{\tau=0\}, \quad (s,\tau) \in \tilde{\Omega}. \tag{41}$$

Then $\bar{v}(s,\tau)$ is the unique solution to problem (33) satisfying $\bar{v}(s,\tau) \in C^{2,1}(\tilde{\Omega}) \cap C(\overline{\tilde{\Omega}})$ and $\partial_s \bar{v}(s,\tau) \in W^{2,1}_{q,loc}(\tilde{\Omega}) \cap C(\overline{\tilde{\Omega}})$. Moreover, if we define

$$v(s,\tau) = \begin{cases} \bar{v}(s,\tau), & \text{if } s \le Z(\tau), \\ 0, & \text{if } s > Z(\tau), \end{cases} \tag{42}$$

then $v(s,\tau) \in W^{2,1}_{q,loc}(\Omega) \cap C(\overline{\Omega})$ is the solution to problem (18). Let $S(\tau)$ be given in (40) and let $Z(\tau)$ be given in (29), $S(\tau)$ and $Z(\tau)$ are free boundaries of problem (18) such that

$$\mathcal{F} = \{v=0\} = \{(s,\tau) \mid s \ge Z(\tau), \tau \in (0,T]\}, \tag{43}$$

$$\begin{aligned} \mathcal{G} &= \{-\partial_s v = 0, v > 0\} \\ &= \{(s,\tau) \mid s \le S(\tau), \tau \in (0,T]\}, \end{aligned} \tag{44}$$

$$\begin{aligned} \mathcal{E} &= \{-\partial_s v > 0, v > 0\} \\ &= \{(s,\tau) \mid S(\tau) < s < Z(\tau), \tau \in (0,T]\}. \end{aligned} \tag{45}$$

In addition, for $(s,\tau) \in \tilde{\Omega}$., $v(s,\tau)$ satisfies

$$\partial_s v(s,\tau) - \frac{1-p}{p} v(s,\tau) < 0, \tag{46}$$

$$\partial_{ss} v(s,\tau) - \frac{2-p}{p}\partial_s v(s,\tau) + \frac{1-p}{p^2} v(s,\tau) > 0, \tag{47}$$

We then have the next result.

Proposition 9. *For $(s,\tau) \in \tilde{\Omega}$, $\tilde{u}(s,\tau) = e^{\frac{p-1}{p}s} v(s,\tau)$ is the unique solution to problem (17), where $v(s,\tau)$ is the solution to problem (18). In addition, $S(\tau)$ defined in (40) and $Z(\tau)$ defined in (29) are free boundaries of problem (17) such that*

$$\mathcal{F} = \{\tilde{u}=0\} = \{(s,\tau) \mid s \ge Z(\tau), \tau \in (0,T]\}, \tag{48}$$

$$\begin{aligned} \mathcal{G} &= \{(p-1)\tilde{u} - p\partial_s \tilde{u} = 0, \tilde{u} > 0\} \\ &= \{(s,\tau) \mid s \le S(\tau)\}, \end{aligned} \tag{49}$$

$$\begin{aligned} \mathcal{E} &= \{(p-1)\tilde{u} - p\partial_s \tilde{u} > 0, \tilde{u} > 0\} \\ &= \{(s,\tau) \mid S(\tau) < s < Z(\tau)\}. \end{aligned} \tag{50}$$

Moreover, $\tilde{u}(s,\tau) \in W^{2,1}_{q,loc}(\Omega) \cap C(\overline{\Omega})$ and $\tilde{u}(s,\tau) \in C^{2,1}(\tilde{\Omega}) \cap C(\overline{\tilde{\Omega}})$ that satisfies

$$\partial_s \tilde{u}(s,\tau) < 0, \quad (s,\tau) \in \tilde{\Omega} \tag{51}$$

$$\partial_{ss} \tilde{u}(s,\tau) - \partial_s \tilde{u}(s,\tau) > 0, \quad (s,\tau) \in \tilde{\Omega}. \tag{52}$$

Proof. It follows from transform that $\tilde{u}(s,\tau)$ is the unique solution to (17). The regularity of $\tilde{u}(s,\tau)$ can be deduced from the regularity of $v(s,\tau)$. The uniqueness of solution to problem (18) leads to the uniqueness of solution to the problem (17). We can deduce (48)–(50) from (43)–(45), and (51)–(52) from (46)–(47).

3.5. The solution to the dual variational inequality (15)

In view of the transform $y=e^s, t=T-\tau, \hat{u}(y,t)=\tilde{u}(s,\tau)$, where $\tilde{u}(s,\tau)$ is the solution to problem (17), it is standard to show that $\hat{u}(y,t)$ is the solution to problem (16) in the next result.

Proposition 10. *$\hat{u}(y,t)=\tilde{u}(s,\tau)$ is the unique solution to problem (16), In particular, let $\mathbb{Q}=(0,y_0(t))\times(0,T]$ with $y_0(t)=e^{Z(T-t)}$, and*

$$u(y,t)=\hat{u}(y,t),\quad (y,t)\in\mathbb{Q}. \tag{53}$$

Then $u(y,t)$ is the unique solution to problem (15) and $e^{S(T-t)}\in C[0,T]$ is the free boundary to problem (15) such that

$$\begin{aligned}\mathcal{G}_1&=\left\{(y,t)|(1-p)u+py\partial_y u=0\right\}\\&=\left\{(y,t)|0<y\le e^{S(T-t)},t\in[0,T)\right\},\end{aligned}\tag{54}$$

$$\begin{aligned}\mathcal{E}_1&=\left\{(y,t)|(1-p)u+py\partial_y u<0\right\}\\&=\left\{(y,t)|e^{S(T-t)}<y\le e^{Z(T-t)},t\in[0,T)\right\}.\end{aligned}\tag{55}$$

Moreover, $u(y,t)\in C^{2,1}(\mathbb{Q})\cap C(\overline{\mathbb{Q}})$ that satisfies

$$\partial_y u(y,t)<0,\quad (y,t)\in\mathbb{Q}, \tag{56}$$

$$\partial_{yy}u(y,t)>0,\quad (y,t)\in\mathbb{Q}. \tag{57}$$

4. Proof of main results

4.1. The solution to the HJB variational inequality (12)

Theorem 11. *There exists a solution $U(\omega,t)$ to the problem (12). Moreover, $U(\omega,t)\in C^{2,1}(\mathcal{Q})$ that satisfies*

$$\partial_\omega U(\omega,t)>0,\quad (\omega,t)\in\mathcal{Q}, \tag{58}$$

$$\partial_{\omega\omega}U(\omega,t)<0,\quad (\omega,t)\in\mathcal{Q}. \tag{59}$$

$$0\le U(\omega,t)\le\frac{\omega^{1-p}}{1-p},\quad (\omega,t)\in\mathcal{Q}, \tag{60}$$

Moreover, if U_1, U_2 are the solutions to problem (12) satisfying (60) and

$$\lim_{\omega\to0^+}\omega^{p-1}(U_1-U_2)=0, \tag{61}$$

then $U_1=U_2$ in $\mathcal{Q}$.

Proof. Using the fact that $u(y,t)\in C^{2,1}(\mathbb{Q})$ and (57), we can deduce the existence of a continuous inverse function of $I(y,t)=-\partial_y u(y,t)=\omega>0$ that $I^{-1}(\omega,t)\in C(\mathcal{Q})$. It follows from (56) and the boundary condition on $y_0(t)$ that we have $y_0(t)=I^{-1}(0,t)$. Set

$$U(\omega,t):=u(I^{-1}(\omega,t),t)+\omega I^{-1}(\omega,t).$$

Then $U(\omega,t)$ is continuous. Moreover, we have

$$\begin{aligned}&\partial_\omega U(\omega,t)\\&=\partial_y u\frac{1}{\partial_y I}+I^{-1}+I\frac{1}{\partial_y I}\\&=\frac{\partial_y u}{-\partial_{yy}u}+y+\frac{-\partial_y u}{-\partial_{yy}u}=y=I^{-1}(\omega,t)\in C(\mathcal{Q}),\\&\partial_{\omega\omega}U(\omega,t)=-\frac{1}{\partial_{yy}u(y,t)}\in C(\mathcal{Q}),\\&\partial_t U(\omega,t)=\partial_t u(y,t)=\partial_t u(I^{-1}(\omega,t),t)\in C(\mathcal{Q}),\\&U(0,t)=u(y_0(t),t)=0,\\&U(\omega,T)=\frac{1}{1-p}\left(I^{-1}(\omega,T)\right)^{1-\frac{1}{p}}=\frac{1}{1-p}\omega^{1-p}.\end{aligned}$$

Hence, $U(\omega,t)\in C^{2,1}(\mathcal{Q})$ is a solution to problem (12). As $\partial_\omega U(\omega,t)=y>0$, we have (58). In addition, (57) implies (59). By its definition, we know $V(x,z,t)\ge0$ and $U(\omega,t)=\frac{1}{z^{1-p}}V(x,z,t)\ge0$. Next we will show the right hand side of (60). In view of (27) and the fact $u(y,t)=y^{\frac{p-1}{p}}v(s,\tau)=y^{\frac{p-1}{p}}v(\ln y,T-t)$, then $\partial_t u(y,t)=-y^{\frac{p-1}{p}}\partial_\tau v(\ln y,T-t)\ge0$. Hence $\partial_t U(\omega,t)=\partial_t u(I^{-1}(\omega,t),t)\ge0$, together with $U(\omega,T)=\frac{\omega^{1-p}}{1-p}$, we obtain $U(\omega,t)\le\frac{\omega^{1-p}}{1-p}$, $0\le t\le T$.

In what follows, we show the uniqueness of the solution to problem (12) by the contradiction argument. Suppose U_1, U_2 are two distinct solutions to problem (12) satisfying (61) that $\mathcal{N}:=\{U_1(\omega,t)>U_2(\omega,t)\}\ne\emptyset$, where

$$\begin{aligned}\mathcal{N}_1:=\{(\omega,t)\in\mathcal{N}|(1-p)U_1(\omega,t)-\omega\partial_\omega U_1\\<(1-p)U_2(\omega,t)-\omega\partial_\omega U_2\},\\\mathcal{N}_2:=\{(\omega,t)\in\mathcal{N}|(1-p)U_1(\omega,t)-\omega\partial_\omega U_1\\\ge(1-p)U_2(\omega,t)-\omega\partial_\omega U_2\}.\end{aligned}$$

It follows from the definition of $\mathcal{N}_1$ that

$$\begin{aligned}&\partial_t U_1-\frac{1}{2}\frac{\mu^2}{\sigma^2}\frac{(\partial_\omega U_1)^2}{\partial_{\omega\omega}U_1}+\frac{(\hat{c}^*(\omega,t))^{1-p}}{1-p}\\&\qquad-\hat{c}^*(\omega,t)\partial_\omega U_1-\delta U_1=0,\\&\partial_t U_2-\frac{1}{2}\frac{\mu^2}{\sigma^2}\frac{(\partial_\omega U_2)^2}{\partial_{\omega\omega}U_2}+\frac{(\hat{c}^*(\omega,t))^{1-p}}{1-p}\\&\qquad-\hat{c}^*(\omega,t)\partial_\omega U_2-\delta U_2\le0.\end{aligned}$$

Denote $c^*(\omega,t)=g(\partial_\omega U)$ and $\tilde{u}=U_1-U_2$. It holds that

$$\begin{aligned}&\partial_t\tilde{u}+\frac{(\partial_\omega U_1)^2}{\partial_{\omega\omega}U_1\partial_{\omega\omega}U_2}\partial_{\omega\omega}\tilde{u}+\frac{\partial_\omega U_1+\partial_\omega U_2}{\partial_{\omega\omega}U_2}\partial_\omega\tilde{u}-\delta\tilde{u}\\&\ge g(\partial_\omega U_1)\partial_\omega U_1-g(\partial_\omega U_2)\partial_\omega U_2-\frac{[g(\partial_\omega U_1)]^{1-p}}{1-p}\\&\qquad+\frac{[g(\partial_\omega U_2)]^{1-p}}{1-p}.\end{aligned}\tag{62}$$

Denote

$$G(x):=g(x)x-\frac{[g(x)]^{1-p}}{1-p}.$$

It holds in $\mathcal{N}_1$ that $0<(1-p)\tilde{u}<\omega\partial_\omega\tilde{u}$, $(\omega,t)\in\mathcal{N}_1$. Hence, we obtain $\partial_\omega U_1-\partial_\omega U_2=\partial_\omega\tilde{u}>0$. In view that

$$g(\partial_\omega U)=c^*(\omega,t)=\begin{cases}1, & \partial_\omega U<1,\\(\partial_\omega U)^{-1/p}, & 1<\partial_\omega U<\alpha^{-p},\\\alpha, & \partial_\omega U>\alpha^{-p},\end{cases}$$

$$[g(x)]^{-p}=\begin{cases}1, & x<1,\\x, & 1<x<\alpha^{-p},\\\alpha^{-p}, & x>\alpha^{-p},\end{cases}$$

$$g'(x)=\begin{cases}0,\\-\frac{1}{p}x^{-\frac{1}{p}-1},\\0,\end{cases}$$

and the facts that $g'(x)\le0$ and

$$G'(x)=g'(x)[x-[g(x)]^{-p}]+g(x)=g(x)\ge0,$$

we deduce that $G(x)$ is increasing in x. It then follows that $G(\partial_\omega U_1)\ge G(\partial_\omega U_2)$, and the right hand side of (62) is nonnegative. Therefore, $\tilde{u}$ satisfies

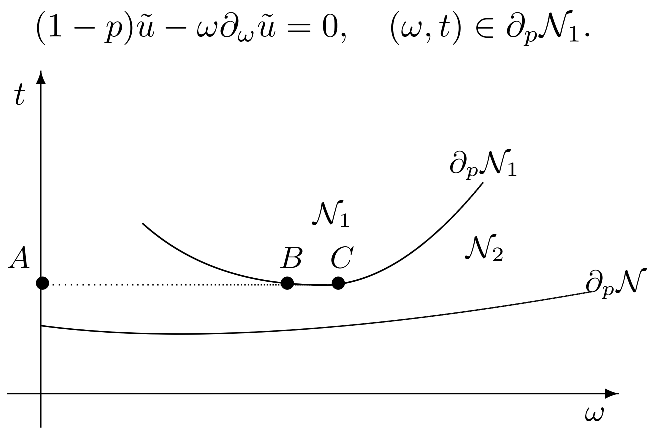

Fig. 4.

$$\partial_t\tilde{u}-\mathcal{L}^*\tilde{u}\geq 0, \quad (\omega,t)\in\mathcal{N}_1,$$

where the coefficients in the operator $\mathcal{L}^*$ are all determined. By the definition of $\mathcal{N}_1$, we have

$$(1-p)\tilde{u}-\omega\partial_\omega\tilde{u}=0, \quad (\omega,t)\in\partial_p\mathcal{N}_1.$$

Let $t_0=\inf\{t:(\omega,t)\in\mathcal{N}_1\}$ and $B(\omega_0,t_0), C(\omega_1,t_0)\in\partial_p\mathcal{N}_1$, where $\omega_0=\inf\{\omega:(\omega,t_0)\in\partial_p\mathcal{N}_1\}$, $\omega_1\geq\omega_0$ and $\overline{BC}\subset\partial_p\mathcal{N}_1$, as illustrated in the above figure. We claim that $\tilde{u}(B)=0$. If it is not true, then we consider $A(0,t_0)$. So we have $(1-p)\tilde{u}-\omega\partial_\omega\tilde{u}\geq 0$ for $(\omega,t)\in\overline{AB}$. It follows that $\partial_\omega(\omega^{p-1}\tilde{u})\leq 0$ for $(\omega,t)\in\overline{AB}$. By the condition (61), we have $\omega^{p-1}\tilde{u}(B)\leq 0$. Together with $\omega^{p-1}\tilde{u}(B)\geq 0$, we obtain $\tilde{u}(B)=0$, and hence $\tilde{u}(\omega,t)=0$ for $(\omega,t)\in\overline{BC}$. By the maximum principle, we know $U_1-U_2=\tilde{u}\leq 0$ for $(\omega,t)\in\mathcal{N}_1$, which leads to a contradiction with the definition of $\mathcal{N}_1$. Hence, $\mathcal{N}_1=\emptyset$, and $\mathcal{N}=\mathcal{N}_2$. In view that $\partial_\omega(\omega^{p-1}\tilde{u})\geq 0$ for $(\omega,t)\in\partial_p\mathcal{N}$, and $\omega^{p-1}\tilde{u}=0$ for $(\omega,t)\in\partial_p\mathcal{N}$, we have $\omega^{p-2}\tilde{u}=0$, for $(\omega,t)\in\mathcal{N}$, which contradicts the definition of $\mathcal{N}$. The uniqueness of the solution to problem (12) then follows (see Fig. 4).

4.2. Proof of Theorem 1

Theorem 12. *Let $w(s,\tau)$ be the solution to problem (34), and $S(\tau)$ be defined in (40). There exist three free boundaries $\omega^*(t)$, $\omega_1(t)$ and $\omega_\alpha(t)$ to problem (12) such that*

$$\omega_\alpha(t)<\omega_1(t)\leq\omega^*(t), \quad t\in[0,T], \tag{63}$$

$$\begin{aligned} JR&=\left\{(\omega,t)\in\overline{\mathcal{Q}}\mid(1-p)U-\omega\partial_\omega U=0\right\}\\ &=\left\{(\omega,t)\in\overline{\mathcal{Q}}\mid\omega\geq\omega^*(t),t\in[0,T)\right\}, \end{aligned}\tag{64}$$

$$\begin{aligned} CR&=\left\{(\omega,t)\in\overline{\mathcal{Q}}\mid(1-p)U-\omega\partial_\omega U<0\right\}\\ &=\left\{(\omega,t)\in\overline{\mathcal{Q}}\mid\omega<\omega^*(t),t\in[0,T)\right\}. \end{aligned}\tag{65}$$

Moreover, for $t\in[0,T)$, we have the analytical form in terms of $w(s,\tau)$ that

$$\omega^*(t)=\frac{1-p}{p}e^{-\frac{1}{p}S(T-t)}\int_{Z(T-t)}^{S(T-t)}w(\xi,T-t)d\xi, \tag{66}$$

$$\omega_1(t)=\frac{1-p}{p}\int_{Z(T-t)}^{0}w(\xi,T-t)d\xi-w(0,T-t), \tag{67}$$

$$\begin{aligned} \omega_\alpha(t)=\alpha\frac{1-p}{p}&\int_{Z(T-t)}^{-p\ln\alpha}w(\xi,T-t)d\xi\\ &-\alpha w(-p\ln\alpha,T-t). \end{aligned}\tag{68}$$

In particular, our conjectured free boundary $\omega^*(t)$ in Remark 2 is now characterized analytically by (66). In addition, the candidate optimal feedback control $\hat{c}^*(\omega,t)$ given in (11) satisfies that

$$\hat{c}^*(\omega,t)=\begin{cases}\alpha, & \text{if } 0<\omega\leq\omega_\alpha(t),\\ (\partial_\omega U)^{-\frac{1}{p}}, & \text{if } \omega_\alpha(t)<\omega<\omega_1(t),\\ 1, & \text{if } \omega_1(t)\leq\omega\leq\omega^*(t).\end{cases}\tag{69}$$

Proof. According to (57), $I(y,t)=-\partial_y u(y,t)$ is strictly decreasing in y, it follows that

$$\mathcal{G}_1=\left\{(y,t)\in\mathbb{Q}\mid I(y,t)\geq I(e^{S(T-t)},t),t\in[0,T)\right\},$$
$$\mathcal{E}_1=\left\{(y,t)\in\mathbb{Q}\mid I(y,t)<I(e^{S(T-t)},t),t\in[0,T)\right\}.$$

The above results, together with (54)–(55), imply the existence of a free boundary $\omega^*(t)=I(e^{S(T-t)},t)=-\partial_y u(e^{S(T-t)},t)$ such that (64)–(65) hold true.

Using the transform $u(y,t)=\tilde{u}(s,\tau)=e^{-\frac{1-p}{p}s}v(s,\tau)$, $y=e^s$ and $\tau=T-t$, we have that

$$\begin{aligned} &\omega^*(t)\\ &=-\partial_y u(e^{S(T-t)},t)=-e^{-S(T-t)}\partial_s\tilde{u}(S(T-t),T-t)\\ &=\frac{1-p}{p}e^{-\frac{1}{p}S(T-t)}\int_{Z(T-t)}^{S(T-t)}w(\xi,T-t)d\xi, \end{aligned}$$

where the last equality is due to the expression of $v(s,\tau)$ in (41) and the fact that $\partial_s v(S(T-t),T-t)=0$.

Next, we show (69). In the domain CR, we have

$$\partial_t U-\frac{1}{2}\frac{\mu^2}{\sigma^2}\frac{(\partial_\omega U)^2}{\partial_{\omega\omega}U}+\frac{\hat{c}^{1-p}(\omega,t)}{1-p}-\hat{c}(\omega,t)\partial_\omega U-\delta U=0,$$

where $\hat{c}(\omega,t)$ is given by (11). Then, let us consider two free boundaries $\omega_1(t)$ and $\omega_\alpha(t)$ in problem (12) satisfying that

$$\begin{cases}\partial_\omega U(\omega_1(t),t)=1,\\ \partial_\omega U(\omega_\alpha(t),t)=\alpha.\end{cases}\tag{70}$$

By the strict concavity (59) of $U(\omega,t)$, we get the desired result (63). Moreover, the strict concavity of $U(\omega,t)$ and the definition (11) of $\hat{c}(\omega,t)$ imply the expression (69).

Next, for any fixed $t\in[0,T)$, the first equation of (70) implies that $\omega_1(t)$ satisfies

$$\begin{aligned} \omega_1(t)&=(\partial_\omega U)^{-1}(1,t)=-\partial_y u(1,t)=-\partial_s\tilde{u}(0,T-t)\\ &=\frac{1-p}{p}\int_{Z(T-t)}^{0}w(\xi,\tau)d\xi-w(0,T-t). \end{aligned}$$

Similarly, it follows from the second equation of (70) that $\omega_\alpha(t)$ satisfies

$$\begin{aligned} &\omega_\alpha(t)\\ &=(\partial_\omega U)^{-1}(\alpha^{-p},t)=-\partial_y u(\alpha^{-p},t)\\ &=-\alpha^p\partial_s\tilde{u}(-p\ln\alpha,T-t)\\ &=\alpha\frac{1-p}{p}\int_{Z(T-t)}^{-p\ln\alpha}w(\xi,T-t)d\xi-\alpha w(-p\ln\alpha,T-t). \end{aligned}$$

The continuity of $\omega^*(t),\omega_1(t),\omega_\alpha(t)$ can be deduced from the expressions in (66)–(68), the continuity of $w(s,\tau)$ and $S(\tau)$ with respect to τ.

We are now ready to prove Theorem 1.

Proof (*Proof of Theorem 1*)**.** By Theorems 11 and 12, it is straightforward to check that $V(x,z,t)$ is the unique solution in $C^{2,1}(Q)\cap C(\bar{Q})$ to problem (4). In view of (64)–(65), if $x\geq\omega^*(t)z$, then

$$\partial_z V(x,z,t)=z^{-p}[(1-p)U(\omega,t)-\omega\partial_\omega U(\omega,t)]=0,$$

for $\omega=\frac{x}{z}\geq\omega^*(t)$; if $x<\omega^*(t)z$, then

$$\partial_z V(x,z,t) = z^{-p}[(1-p)U(\omega,t) - \omega\partial_\omega U(\omega,t)] < 0,$$

for $\omega = \frac{x}{z} < \omega^*(t)$. Hence, we obtain results (9)–(10).

To prove the optimality of feedback control (7) and (9), it suffices to show that the following two conditions hold for all $x, z \in \mathbb{R}^+$ and $t \le T$:

(i) The following SDE admits a unique strong solution $(X_s^*)_{s\ge t}$ that

$$\begin{cases} dX_s^* = & \left(\mu\pi^*\left(X_s^*, \frac{M_s^*}{\omega^*(s)}, s\right) - c^*\left(X_s^*, \frac{M_s^*}{\omega^*(s)}, s\right)\right) ds \\ & +\sigma\pi^*\left(X_s^*, \frac{M_s^*}{\omega^*(s)}, s\right) dW_s; \\ M_s^* = & \max\left\{\omega^*(t)z, \sup_{t\le u\le s} X_u^*\right\}; \\ X_t^* = & x. \end{cases} \tag{71}$$

(ii) $(\pi_t^*, c_t^*) := (\pi^*(X_t^*, z_t^*, t), c^*(X_t^*, z_t^*, t))$ is optimal.

For simplicity, we write $\mathbb{E}[\cdot|X_t = x, z_t = z]$ as $\mathbb{E}[\cdot]$ in the following. We first prove that condition (i) holds. By the form of π^*, c^* and Remark 2, we know that the consumption running maximum process z_s^* satisfies $z_s^* = \frac{M_s^*}{\omega^*(s)}$ for all $s \in [t, T\wedge\tau]$ with M_s^* given by (71). Hence, if the processes X_s^* and z_s^* satisfying (71) exist, then it is easy to check that $(\pi_t^*, c_t^*) = (\pi^*(X_t^*, z_t^*, t), c^*(X_t^*, z_t^*, t))$ is admissible. To show that (71) has a unique strong solution, by Theorem 7 in section 3 of Chapter 5 of Protter (2005), one needs to show that the functionals

$$G(s,X) := \pi^*(X_s, \frac{1}{\omega^*(t)}\max\{\omega^*(t)z, \sup_{t\le u\le s} X_u\}, s),$$
$$F(s,X) := c^*(X_s, \frac{1}{\omega^*(t)}\max\{\omega^*(t)z, \sup_{t\le u\le s} X_u\}, s)$$

for continuous functions $X: \mathbb{R}^+ \to \mathbb{R}^+$, are functional Lipschitz in the sense of Protter (2005). Note that one can apply the similar arguments as used in Appendix A of Angoshtari et al. (2019) to prove the Lipschitz property of the feedback functions π^* and c^*. Therefore, for any $s \in [t, T\wedge\tau]$ and continuous functions X and Y, we have

$$\begin{aligned} &|G(s,X) - G(s,Y)| \\ &\le |X_s - Y_s| \\ &+ \left|\max\left\{\omega^*(t)z, \sup_{t\le u\le s} X_u\right\} - \max\left\{\omega^*(t)z, \sup_{t\le u\le s} Y_u\right\}\right| \\ &\le 2K \sup_{t\le u\le s} |X_u - Y_u|. \end{aligned} \tag{72}$$

Hence, G is functional Lipschitz. Similarly, F is also functional Lipschitz.

To prove condition (ii), we can apply some standard localization arguments in the literature to obtain that

$$V(x,z,t) = \mathbb{E}\Bigg[\int_t^{\tau_n} e^{-\delta(s-t)}\frac{(c_s^*)^{1-p}}{1-p}ds + e^{-\delta(\tau_n - t)}V(X_{\tau_n}^*, z_{\tau_n}^*, \tau_n)\Bigg].$$

Letting $n\to\infty$ in above equation and using Monotone Convergence Theorem, we obtain

$$\begin{aligned} &\lim_{n\to\infty}\mathbb{E}\left[\int_t^{\tau_n} e^{-\delta(s-t)}\frac{(c_s^*)^{1-p}}{1-p}ds\right] \\ &= \mathbb{E}\left[\int_t^{T\wedge\tau} e^{-\delta(s-t)}\frac{(c_s^*)^{1-p}}{1-p}ds\right]. \end{aligned} \tag{73}$$

Additionally, note that the value function of the utility maximization problem on terminal wealth under a drawdown constraint with power utility function is less than the value function of Merton investment problem with power utility function ($\alpha = 0$ in our context) that

$$\sup_{(\pi,c)\in\mathcal{A}(x)}\mathbb{E}\left[\frac{1}{1-p}X_{\tau_n}^{1-p}\right] \le \sup_{(\pi,c)\in\mathcal{A}_0(x)}\mathbb{E}\left[\frac{1}{1-p}X_{\tau_n}^{1-p}\right],$$

where $\mathcal{A}_0(x)$ is the set of admissible strategies for Merton problem. Similar to Lemma A.3 of Angoshtari et al. (2019), there exists a constant $M > 0$ such that $0 \le V(x,z,t) \le \frac{x^{1-p}}{1-p}M$ for all $x \ge 0$, independent of z and t. Therefore, it holds that

$$\begin{aligned} \mathbb{E}\left[e^{-\delta(\tau_n - t)}V(X_{\tau_n}^*, z_{\tau_n}^*, \tau_n)\right] &\le M\mathbb{E}\left[e^{-\delta(\tau_n-t)}\frac{(X_{\tau_n}^*)^{1-p}}{1-p}\right] \\ &\le M\sup_{(\pi,c)\in\mathcal{A}_0(x)}\mathbb{E}\left[\frac{X_{\tau_n}^{1-p}}{1-p}\right]. \end{aligned}$$

Using the standard transversality condition in the Merton problem, the fact $V(x,z,t) \ge 0$, the terminal condition $V(x,z,T) = \frac{x^{1-p}}{1-p}$, the boundary condition $V(0,z,t) = 0$, and the Dominated Convergence Theorem, we have that

$$\begin{aligned} &\lim_{n\to\infty}\mathbb{E}\left[e^{-\delta(\tau_n-t)}V(X_{\tau_n}^*, z_{\tau_n}^*, \tau_n)\right] \\ &= \mathbb{E}\left[e^{-\delta(T\wedge\tau - t)}\frac{(X_{T\wedge\tau}^*)^{1-p}}{1-p}\right]. \end{aligned} \tag{74}$$

Combining (73) and (74), we deduce that the first equality in condition (ii) holds.

Finally, in view that the feedback controls $(\pi_t^*, c_t^*) := (\pi^*(X_t^*, z_t^*, t), c^*(X_t^*, z_t^*, t))$ are admissible, we readily have

$$\begin{aligned} &V(x,z,t) \\ &\le \sup_{(\pi,c)\in\mathcal{A}(x)}\mathbb{E}\left[\int_t^{T\wedge\tau} e^{-\delta(s-t)}\frac{c_s^{1-p}}{1-p}ds + e^{-\delta(T\wedge\tau-t)}\frac{X_{T\wedge\tau}^{1-p}}{1-p}\right]. \end{aligned}$$

For any admissible strategies $(\pi, c) \in \mathcal{A}(x)$ and corresponding state process (X, z), as $V(x,z,t)$ is the unique classical solution to problem (4), it is straightforward to derive that

$$\begin{aligned} &V(x,z,t) \\ &\ge \mathbb{E}\left[\int_t^{\tau_n} e^{-\delta(s-t)}\frac{(c_s)^{1-p}}{1-p}ds + e^{-\delta(\tau_n-t)}V(X_{\tau_n}, z_{\tau_n}, \tau_n)\right]. \end{aligned}$$

Letting $n\to\infty$, we conclude that the reverse inequality holds, which completes the proof.

5. Conclusions

We revisit the optimal consumption problem under drawdown constraint formulated in Angoshtari et al. (2019) by featuring the finite investment horizon. For this stochastic control problem under control-type constraint, we contribute to the theoretical study on the existence and uniqueness of the classical solution to the parabolic HJB variational inequality. In particular, the consumption drawdown constraint induces some time-dependent free boundaries that deserve careful investigations. Using the dual transform and considering the auxiliary variational inequality with both function and gradient constraints, we develop some technical arguments to obtain the regularity of the unique solution as well as some analytical characterization of the associated time-dependent free boundaries such that the smooth fit conditions hold. As a result, we are able to derive and verify the optimal portfolio and consumption strategies in the piecewise feedback form.

Acknowledgments

The authors sincerely thank referees for their valuable comments and suggestions.

References

Angoshtari, B., Bayraktar, E., & Young, V. R. (2019). Optimal dividend distribution under drawdown and ratcheting constraints on dividend rates. *SIAM Journal on Financial Mathematics*, *10*, 547–577.

Angoshtari, B., Bayraktar, E., & Young, V. R. (2022). Optimal investment and consumption under a habit-formation constraint. *SIAM Journal on Financial Mathematics*, *13*, 321–352.

Arun, T. (2012). The merton problem with a drawdown constraint on consumption. Preprint, available at arXiv:1210.5205.

Bo, L., Wang, S., & Yu, X. (2024). A mean field game approach to equilibrium consumption under external habit formation. *Stochastic Processes and their Applications*, *178*, Article 104461.

Chen, X., Chen, Y., & Yi., F. (2012). Parabolic variational inequality with parameter and gradient constraints. *Journal of Mathematical Analysis and Applications*, *385*(2), 928–946.

Chen, X., Li, X., & Yi, F. (2019). Optimal stopping investment with non-smooth utility over an infinite time horizon. *Journal of Industrial and Management Optimization*, *15*, 81–96.

Chen, X., & Yi, F. (2012). A problem of singular stochastic control with optimal stopping in finite horizon. *SIAM Journal on Control and Optimization*, *50*, 2151–2172.

Constantinides, G. M. (1990). Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy*, *98*(3), 519–543.

Dai, M., & Yi, F. (2009). Finite horizon optimal investment with transaction costs: A parabolic double obstacle problem. *Journal of Differential Equations*, *246*, 1445–1469.

Deng, S., X. Li, H. P., & Yu, X. (2022). Optimal consumption with reference to past spending maximum. *Finance and Stochastics*, *26*, 217–266.

Detemple, J., & Zapatero, F. (1992). Habit formation: A resolution of the equity premium puzzle. *Mathematical Finance*, *2*, 251–274.

Dybvig, P. H. (1995). Dusenberry's racheting of consumption: optimal dynamic consumption and investment given intolerance for any decline in standard of living. *Review of Economic Studies*, *62*, 287–313.

Englezos, N., & Karatzas, I. (2009). Utility maximization with habit formation: dynamic programming and stochastic PDEs. *SIAM Journal on Control and Optimization*, *48*, 481–520.

Englezos, N., & Karatzas, I. (2021). Portfolio selection with drawdown constraint on consumption: a generalization model. *Mathematical Methods of Operations Research*, *93*(2), 243–289.

Guan, C., Yi, F., & Chen, X. (2019). A fully nonlinear free boundary problem arising from optimal dividend and risk control model. *Mathematical Control and Related Fields*, *9*, 425–452.

Guasoni, P., Huberman, G., & Ren, D. (2020). Shortfall aversion. *Mathematical Finance*, *39*(2), 869–920.

Jeon, J., Kim, T., & Yang, Z. (2024). The finite-horizon retirement problem with borrowing constraint: A zero-sum stopper vs. singular-controller game. http://dx.doi.org/10.2139/ssrn.4364441, Preprint, available at SSRN.

Jeon, J., & Oh, J. (2022). Finite horizon portfolio selection problem with a drawdown constraint on consumption. *Journal of Mathematical Analysis and Applications*, *506*(1), Article 125542.

Li, X., Yu, X., & Zhang, Q. (2022). Optimal consumption and life insurance under shortfall aversion and a drawdown constraint. *Insurance: Mathematics & Economics*, *108*, 25–45.

Li, X., Yu, X., & Zhang, Q. (2024). Optimal consumption with loss aversion and reference to past spending maximum. *SIAM Journal on Financial Mathematics*, *15*(1), 121–160.

Liang, Z., Luo, X., & Yuan, F. (2023). Consumption-investment decisions with endogenous reference point and drawdown constraint. *Mathematics and Financial Economics*, *17*, 285–334.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, *51*, 247–257.

Merton, R. C. (1971). Optimal consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, *3*, 373–413.

Munk, C. (2008). Portfolio and consumption choice with stochastic investment opportunities and habit formation in preferences. *Journal of Economic Dynamics & Control*, *32*, 3560–3589.

Protter, P. E. (2005). *Stochastic integration and differential equations* (2nd ed.). Berlin: Springer-Verlag.

Schroder, M., & Skiadas, C. (2002). An isomorphism between asset pricing models with and without linear habit formation. *The Review of Financial Studies*, *15*(4), 1189–1221.

Tanana, A. (2023). Utility maximization with ratchet and drawdown constraints on consumption in incomplete semimartingale markets. *The Annals of Applied Probability*, *33*(5), 4127–4162.

Yang, Y., & Yu, X. (2022). Optimal entry and consumption under habit formation. *Advances in Applied Probability*, *54*(2), 433–459.

Yu, X. (2015). Utility maximization with addictive consumption habit formation in incomplete semimartingale markets. *The Annals of Applied Probability*, *25*(3), 1383–1419.

Yu, X. (2017). Optimal consumption under habit formation in markets with transaction costs and random endowments. *The Annals of Applied Probability*, *27*(2), 960–1002.

Xiaoshan Chen received the B.S. and Ph.D. degrees in Mathematics from South China Normal University in 2007 and 2012, respectively. She took the Postdoc position in City University of Hong Kong and Wayne State University in 2013 and 2014, respectively. She is currently working at the School of Mathematical Science, South China Normal University as an assistant professor. Her research interests include financial mathematics, partial differential equations and its application in finance, and free boundary problem.

Xun Li received the B.S. degree in 1992 from the Department of Mathematics at Shanghai University of Science and Technology, the M.S. degrees in 1995 from the Department of Mathematics at Shanghai University. He completed his Ph.D. degree in 2000 from the Department of Systems Engineering and Engineering Management at the Chinese University of Hong Kong, and stayed with the same department as a postdoctoral research fellow until 2001. From 2001 to 2003, he was a postdoctoral fellow in the Mathematical and Computational Finance Laboratory at University of Calgary. From 2003 to 2007, he was a visiting fellow in the Department of Mathematics at National University of Singapore. He joined the Department of Applied Mathematics at the Hong Kong Polytechnic University as Assistant Professor in 2007, Associate Professor in 2013, and is currently Professor. His main research areas are stochastic control and applied probability with financial applications, and he has published in journals such as SIAM Journal on Control and Optimization, Annals of Applied Probability, IEEE Transactions on Automatic Control, Automatica, Journal of Differential Equations, Mathematical Finance, Finance and Stochastics, and Quantitative Finance.

Fahuai Yi obtained his Master degree at Nankai University and doctoral degree at Zhejiang University. He is currently a retired professor from South China Normal University. His research interests include nonlinear partial differential equations, free boundary problems, stochastic control and quantitative finance.

Xiang Yu received the B.S. degree in Mathematics at Huazhong University of Science and Technology in 2007 and the Ph.D. degree in Mathematics at the University of Texas at Austin in 2012. During 2012 to 2015, he worked as the postdoc Assistant Professor in the Department of Mathematics at University of Michigan. He joined the Department of Applied Mathematics at the Hong Kong Polytechnic University as Assistant Professor in 2015 and Associate Professor in 2022. His research interests lie primarily in mathematical finance, applied probability and stochastic analysis, stochastic control and optimization. He has publications in Mathematical Finance, Finance and Stochastics, Annals of Applied Probability, Mathematics of Operations Research and SIAM Journal on Control and Optimization.