Belak, Christoph; Mich, Lukas; Seifried, Frank T.

Optimal investment for retail investors

Mathematical Finance

Provided in Cooperation with:
John Wiley & Sons

Suggested Citation: Belak, Christoph; Mich, Lukas; Seifried, Frank T. (2021) : Optimal investment for retail investors, Mathematical Finance, ISSN 1467-9965, Wiley Periodicals, Inc., Hoboken, USA, Vol. 32, Iss. 2, pp. 555-594, https://doi.org/10.1111/mafi.12336

Received: 3 September 2019 | Accepted: 23 July 2021

DOI: 10.1111/mafi.12336

ORIGINAL ARTICLE

MATHEMATICAL FINANCE An International Journal of Mathematics, Statistics and Financial Economics WILEY

Optimal investment for retail investors

Christoph Belak[1] | **Lukas Mich**[2] | **Frank T. Seifried**[2]

[1] Technische Universität Berlin, Institute of Mathematics, Germany

[2] University of Trier, Department IV – Mathematics, Germany

Correspondence
Christoph Belak, Technische Universität Berlin, Institute of Mathematics, Straße des 17. Juni 136, 10623 Berlin, Germany.
Email: belak@math.tu-berlin.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: GRK 2126: Algorithmic Optimization (ALOP); Deutsche Forschungsgemeinschaft, Grant/Award Number: GRK 2544: Stochastic Analysis in Interaction.

Abstract

We study optimal portfolio decisions for a retail investor that faces a strictly positive transaction cost in a classical Black-Scholes market. We provide a construction of optimal trading strategies and characterize the value function as the unique viscosity solution of the associated quasi-variational inequalities. Moreover, we numerically investigate the optimal trading regions for a variety of real-world cost structures faced by retail investors. We find that the cost structure has a strong effect on the qualitative shape of the no-trading region and optimal strategies.

KEYWORDS
portfolio optimization, transaction costs, retail investor

JEL CLASSIFICATION
G11, C61

1 | INTRODUCTION

Classical transaction cost models typically assume that costs are affine functions of the trading volume,[1] that is, either costs which are proportional to the trading volume, (Czichowsky & Schachermayer, 2017; Davis & Norman, 1990; Kallsen & Muhle-Karbe, 2010; Shreve & Soner, 1994), purely fixed costs, (Altarovici et al., 2015, 2017; Feodoria, 2016; Liu, 2004), or a mix of the two, (Altarovici et al., 2017; Belak & Christensen, 2019; Eastham & Hastings, 1988; Korn, 1998; Liu, 2004; Øksendal & Sulem, 2002). In real-world markets, however, retail investors can also face different cost structures which are, for example, proportional to the trading volume, but

additionally floored and capped at certain minimal and maximal cost levels. Thus, one example of a retail investor's transaction cost as a function of the trading volume is

$$C : \mathbb{R} \setminus \{0\} \to \mathbb{R}, \quad \Delta \mapsto C(\Delta) \triangleq \min\{\max\{C_{\min}, c|\Delta|\}, C_{\max}\}, \tag{1}$$

where $0 < C_{\min} < C_{\max}$ and $c \in (0, 1)$. Here, $C_{\min}$ represents the minimal cost due per trade, c is the proportional cost per unit trading volume, and $C_{\max}$ is the maximal trading cost.

In the literature, rather little is known about optimal investment decisions in the presence of transaction costs such as (1), except for general results that are agnostic toward the specific cost structure. Thus in Bielecki and Pliska (2000) a risk-sensitive growth rate criterion is considered for general cost functions, and the authors provide a verification theorem for the value function which allows to construct an optimal investment strategy under the assumption of the existence of a sufficiently smooth solution of the associated Bellman equation.[2] Moreover, in Palczewski and Stettner (2007), an iterated optimal stopping approach is employed to construct optimal trading strategies for a lifetime consumption-portfolio problem with general cost functions, in which consumption is only allowed to take place at trading dates. While the focus of Palczewski and Stettner (2007) is on the theoretical study of existence of optimal strategies in their general setting, this paper provides a detailed investigation, including a qualitative analysis of optimal strategies, for real-world cost structures including (1).

Thus in this paper we consider a retail investor in a Black-Scholes market that faces strictly positive transaction costs and who wishes to maximize expected utility from terminal wealth. Being a retail investor, it is natural to assume that short sales and leverage, that is, short positions in either the money market account or the stock, are prohibited.[3] In this setting, we use arguments based on the stochastic Perron's method (see Bayraktar & Sîrbu, 2012, 2013, 2014, for early developments) to characterize the value function as the unique viscosity solution of the associated Bellman equation, which in this setting is represented by a system of quasi-variational inequalities (QVIs). Given this characterization of the value function, we employ the superharmonic function technique introduced in Christensen (2014) and further refined in Belak & Christensen (2019) and Belak et al. (2017) to provide an explicit construction of optimal investment strategies in terms of a trading region and post-trade target positions. On the basis of our theoretical results, we provide a detailed numerical investigation of the structure and shape of the trading regions and post-trade target positions leading to the optimal investment strategy for a range of cost structures faced by retail investors: fixed costs, fixed plus proportional costs, piecewise constant costs, and the floored and capped proportional cost structure given by (1).

Our numerical results indicate a strong dependence of the optimal trading strategies on the cost structure: For fixed, fixed plus proportional, and piecewise constant costs, the no-trading region is V-shaped. In the case of fixed costs, optimal trades shift the portfolio onto the optimal position in the absence of transaction costs, whereas for fixed plus proportional costs optimal post-trade target portfolios form a wedge around the frictionless optimizer. For piecewise constant costs, we observe optimality of multiple small trades performed back to back instead of single large trades. For the floored and capped proportional cost (1), the no-trading region is best described as being vVv-shaped. In particular, there exist portfolio positions for which it is not optimal to trade, even though there are both portfolios closer to and further away from the frictionless optimizer for which it *is* optimal to make a transaction. Second, we identify some novel boundary effects for short time horizons. Third and most importantly, we find that optimal transactions can be characterized via distinct wealth regimes: For moderate amounts of wealth, the retail investor optimally trades only at the floored cost, with target portfolios distinct from the frictionless optimizer. By

contrast, investors with large levels of wealth trade at the capped cost and onto the frictionless optimal position, thus effectively facing fixed transaction costs.

The remainder of this article is structured as follows. In Section 2 we introduce the model and provide the mathematical formulation of the retail investor's portfolio optimization problem. In Section 3 we state the main mathematical results of this paper. Section 4 contains our main qualitative findings via a detailed analysis of optimal trading strategies. Finally, Section 5 provides the proof of the viscosity characterization and the construction of optimal strategies.

2 | RETAIL INVESTOR PORTFOLIO PROBLEM

In all that follows, we fix a filtered probability space $(\Omega, \mathfrak{A}, \mathfrak{F}, \mathbb{P})$ with $\mathfrak{F} = \{\mathfrak{F}_t\}_{t\in[0,T]}$ that supports a standard $\mathfrak{F}$-Wiener process $W = \{W_t\}_{t\in[0,T]}$ and satisfies the usual conditions.

2.1 | Financial market and transaction costs

We consider a retail investor that has access to a classical Black-Scholes market $P = (P^0, P^1)$ consisting of a money market account $P^0 = \{P^0_t\}_{t\in[0,T]}$ with risk-free rate $r \in \mathbb{R}$ and a stock (or stock index) $P^1 = \{P^1_t\}_{t\in[0,T]}$ with drift $\mu \in \mathbb{R}$ and volatility $\sigma > 0$. The dynamics of P^0 and P^1 are thus given by

$$\mathrm{d}P^0_t = rP^0_t\,\mathrm{d}t \quad \text{and} \quad \mathrm{d}P^1_t = \mu P^1_t\,\mathrm{d}t + \sigma P^1_t\,\mathrm{d}W_t, \quad t \in [0,T].$$

The investor faces strictly positive transaction costs that are increasing in the trading volume. More precisely, we take as given a continuous *transaction cost function*

$$\mathrm{C} : \mathbb{R} \to \mathbb{R}_+, \qquad \Delta \mapsto \mathrm{C}(\Delta) \tag{2}$$

and assume $|\Delta| \mapsto \mathrm{C}(|\Delta|)$ is increasing and[4] $\mathrm{C}_{\min} \triangleq \mathrm{C}(0) > 0$. Moreover, we suppose that

$$\Delta \mapsto \Delta + \mathrm{C}(\Delta) \qquad \text{is strictly increasing on } \mathbb{R}. \tag{3}$$

Since the range of this map contains $\mathbb{R}_+$, there exists a continuous and strictly increasing inverse map denoted by $\chi : \mathbb{R}_+ \to \mathbb{R}$. Note that (3) only has to be checked on $(-\infty, 0]$ since C is increasing on $\mathbb{R}_+$. Strict monotonicity is satisfied in the common situation in which C is piecewise linear on $(-\infty, 0]$ with slope in $(-1, 0]$, which, in particular, is the case for the following examples.

Examples. In Section 4 we investigate optimal trading strategies for the following real-world cost functions encountered by retail investors:

- ▷ *Fixed Costs*: The investor pays a fixed cost $\mathrm{C}_{\min} > 0$ on each trade, that is,

$$\mathrm{C}(\Delta) \triangleq \mathrm{C}_{\min}, \qquad \Delta \in \mathbb{R}.$$

- ▷ *Fixed plus Proportional Costs*: The investor pays a fixed cost $\mathrm{C}_{\min} > 0$ plus a proportional cost $\mathrm{c} \in (0,1)$ on each trade, that is,

$$\mathrm{C}(\Delta) \triangleq \mathrm{C}_{\min} + \mathrm{c}|\Delta|, \qquad \Delta \in \mathbb{R}.$$

This corresponds to the cost function studied in Belak and Christensen (2019).

▷ *Floored and Capped Proportional Costs*: The investor pays proportional costs $c \in (0,1)$ floored and capped at levels $0 < C_{min} < C_{max}$, that is,

$$C(\Delta) \triangleq \min\{\max\{C_{min}, c|\Delta|\}, C_{max}\}, \qquad \Delta \in \mathbb{R}.$$

▷ *Piecewise Fixed Costs*: The investor pays a fixed cost $C_{min} > 0$ for trades not exceeding a threshold $D > 0$, and a cost of $C_{max} > C_{min}$ otherwise, that is,[5]

$$C(\Delta) \triangleq C_{min}\mathbb{1}_{[0,D]}(|\Delta|) + C_{max}\mathbb{1}_{(D,\infty)}(|\Delta|), \qquad \Delta \in \mathbb{R}.$$

Portfolio positions are specified as pairs $x = (x_0, x_1) \in \mathbb{R}^2$ where x_0 and x_1 represent the dollar amounts invested in the money market and in the stock, respectively. The sets of portfolios without leveraged or short positions in the stock (equivalently, without short positions in either the money market account or the stock) and the set of non-zero portfolios without leverage or shorting are defined by

$$\overline{\mathcal{S}} \triangleq \mathbb{R}^2_+ \qquad \text{and} \qquad \mathcal{S} \triangleq \mathbb{R}^2_+ \setminus \{0\}.$$

Given the cost function (2), the transaction $\Delta \in \mathbb{R}$ shifts a portfolio $x = (x_0, x_1) \in \overline{\mathcal{S}}$ to the new position $\Gamma(x, \Delta)$, where the *rebalancing function* Γ is given by

$$\Gamma : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2, \qquad (x, \Delta) \mapsto \Gamma(x, \Delta) \triangleq (x_0 - \Delta - C(\Delta), x_1 + \Delta).$$

A transaction Δ is called *feasible* for the portfolio $x \in \overline{\mathcal{S}}$ if it does not result in a short position in either asset, and we denote the set of all feasible transactions by

$$\mathcal{D}(x) \triangleq \{\Delta \in \mathbb{R} : \Gamma(x, \Delta) \in \overline{\mathcal{S}}\}.$$

Recall that χ denotes the strictly increasing and continuous inverse of $\Delta \mapsto \Delta + C(\Delta)$. Since

$$\Gamma(x, \Delta) \in \overline{\mathcal{S}} \quad \text{if and only if} \quad x_0 - \Delta - C(\Delta) \geq 0 \quad \text{and} \quad x_1 + \Delta \geq 0 \tag{4}$$

$$\text{if and only if} \quad \Delta \in (-\infty, \chi(x_0)] \quad \text{and} \quad \Delta \in [-x_1, \infty), \tag{5}$$

we see that the set of feasible transactions can be represented as

$$\mathcal{D}(x) = [-x_1, \chi(x_0)], \qquad x \in \overline{\mathcal{S}}. \tag{6}$$

Next, let $\mathcal{S}_\emptyset$ denote the set of portfolio positions for which no feasible transaction exists, that is,

$$\mathcal{S}_\emptyset \triangleq \{x \in \overline{\mathcal{S}} : \mathcal{D}(x) = \emptyset\} = \{x \in \overline{\mathcal{S}} : \chi(x_0) < -x_1\},$$

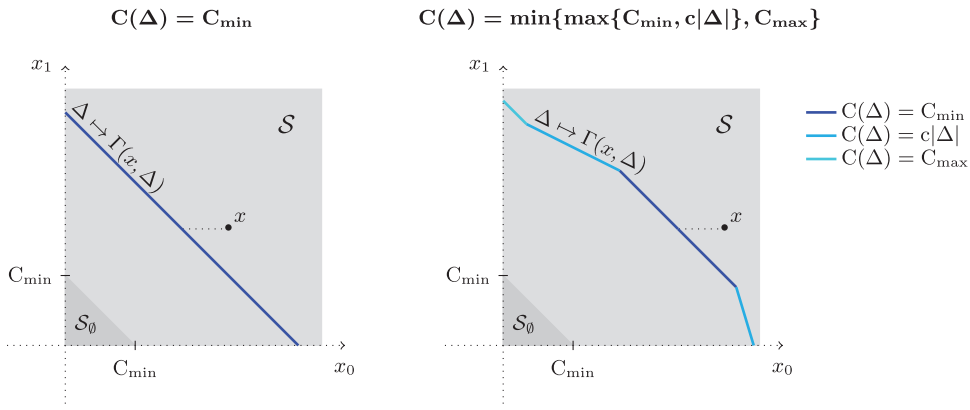

FIGURE 1 Illustration of the solvency region $\mathcal{S}$, the set $\mathcal{S}_\emptyset$ of portfolios for which no feasible transactions exist, and the set $\{\Gamma(x,\Delta) : \Delta \in \mathcal{D}(x)\}$ of portfolios which can be reached by a feasible transaction from x for fixed costs (left) and floored and capped proportional costs (right) [Color figure can be viewed at wileyonlinelibrary.com]

and $\overline{\mathcal{S}}_\emptyset$ and $\partial\mathcal{S}_\emptyset$ denote the closure and the $\overline{\mathcal{S}}$-relative boundary of $\mathcal{S}_\emptyset$, respectively. We note that (6) implies

$$\mathcal{D}(x) \neq \emptyset \quad \text{if and only if} \quad -x_1 \in \mathcal{D}(x) \quad \text{if and only if} \quad x_0 + x_1 - \mathrm{C}(-x_1) \geq 0.$$

But then it follows that

$$\mathcal{S}_\emptyset = \left\{x \in \overline{\mathcal{S}} : x_0 + x_1 < \mathrm{C}(-x_1)\right\} \supseteq \left\{x \in \overline{\mathcal{S}} : x_0 + x_1 < \mathrm{C}_{\min}\right\},$$

from which we see that $\mathcal{S}_\emptyset$ is open relative to $\overline{\mathcal{S}}$; also, it is clear that

$$\partial\mathcal{S}_\emptyset = \{x \in \overline{\mathcal{S}} : x_0 + x_1 = \mathrm{C}(-x_1)\} \qquad \text{and} \qquad \overline{\mathcal{S}}_\emptyset = \{x \in \overline{\mathcal{S}} : x_0 + x_1 \leq \mathrm{C}(-x_1)\}.$$

Figure 1 illustrates $\mathcal{S}_\emptyset$ and the set $\{\Gamma(x,\Delta) : \Delta \in \mathcal{D}(x)\}$ of portfolios which can be reached by a transaction from x for fixed costs and floored and capped proportional costs.

2.2 | Trading strategies and portfolio dynamics

It is well-known that, in the presence of transaction costs that are bounded from below, trading with infinite activity leads to immediate bankruptcy. Hence a *trading strategy* is specified by a sequence $\Lambda = \{(\tau_k, \Delta_k)\}_{k\in\mathbb{N}}$, where $\{\tau_k\}_{k\in\mathbb{N}}$ is an increasing sequence of $\mathfrak{F}$-stopping times representing the trading dates, and each Δ_k, $k \in \mathbb{N}$, is an $\mathbb{R}$-valued $\mathfrak{F}_{\tau_k}$-measurable random variable specifying the volume of the k^{th} trade. Starting from an initial portfolio position $x \in \overline{\mathcal{S}}$ at time $t \in [0,T]$, the dynamics of the retail investor's portfolio $X = X^{t,x,\Lambda} = \{X_s^{t,x,\Lambda}\}_{s\in[t,T]}$ are given by

$$X_s^0 = x_0 + \int_t^s rX_u^0 \,\mathrm{d}u - \sum_{k=1}^{\infty} [\Delta_k + \mathrm{C}(\Delta_k)]\mathbb{1}_{\{\tau_k \leq s\}}, \quad s \in [t,T], \tag{7}$$

$$X_s^1 = x_1 + \int_t^s \mu X_u^1 \, \mathrm{d}u + \int_t^s \sigma X_u^1 \, \mathrm{d}W_u + \sum_{k=1}^{\infty} \Delta_k \mathbb{1}_{\{\tau_k \leq s\}}, \quad s \in [t,T]. \tag{8}$$

We furthermore set $X_{t-}^{t,x,\Lambda} \triangleq x$ to account for the possibility of a trade at time t. A trading strategy Λ is called *admissible* for the initial portfolio position (t,x) if it does not involve leverage or borrowing, that is,

$$\tau_1 \geq t \qquad \text{and} \qquad X_s^{t,x,\Lambda} \in \overline{S}, \qquad s \in [t,T].$$

The set of admissible trading strategies for the initial position (t,x) is denoted by $\mathcal{A}(t,x)$.

Remark. Since transaction costs are bounded from below, admissibility and absence of arbitrage imply in particular that the investor trades only finitely many times a.s., that is, we have

$$\mathbb{P}\left[\lim_{k\to\infty} \tau_k > T\right] = 1 \quad \text{for all } \Lambda = \{(\tau_k, \Delta_k)\}_{k\in\mathbb{N}} \in \mathcal{A}(t,x); \tag{9}$$

see Belak & Christensen, 2019, Lemma A.4, for a formal argument. Moreover, since leveraged positions are ruled out, a standard moments estimate for SDEs, see Belak & Christensen, 2019, Lemma A.5, yields a constant $M > 0$ such that

$$\sup_{\Lambda\in\mathcal{A}(t,x)} \mathbb{E}\left[\sup_{s\in[t,T]} \left|X_s^{t,x,\Lambda}\right|^2\right] \leq M(1+|x|^2), \qquad (t,x) \in [0,T]\times\overline{S}. \tag{10}$$

2.3 | Portfolio problem

The retail investor aims to maximize expected utility from liquid wealth at terminal time T. Her risk preferences are captured by a power utility function with relative risk aversion parameter $1-p$ where $p \in (0,1)$, so the investor's utility function for liquid wealth is given by

$$U : \mathbb{R}_+ \to \mathbb{R}_+, \qquad \ell \mapsto U(\ell) \triangleq \frac{1}{p}\ell^p. \tag{11}$$

We denote by $\mathrm{L}(x)$ the liquidation value of a portfolio $x \in \overline{S}$, where

$$\mathrm{L} : \overline{S} \to \mathbb{R}_+, \quad x \mapsto \mathrm{L}(x) \triangleq x_0 + (x_1 - \mathrm{C}(-x_1))^+.$$

This definition of L guarantees that the investor liquidates her stock position only in case this does not induce a net loss, that is, the revenue from selling is at least as big as the trading cost. Conversely, stocks being limited liability securities, the investor cannot be forced to sell them, and she will thus not do so if she were to incur a loss. Setting $U_{\mathrm{L}} \triangleq U \circ \mathrm{L}$, the retail investor's portfolio problem reads

$$\mathcal{V}(t,x) \triangleq \sup_{\Lambda\in\mathcal{A}(t,x)} \mathbb{E}\left[U_{\mathrm{L}}\left(X_T^{t,x,\Lambda}\right)\right], \qquad (t,x) \in [0,T]\times\overline{S}.$$

3 | MATHEMATICAL RESULTS

In this section, we state and discuss the main mathematical results of this article; their proofs are deferred to Section 5 .

3.1 | Characterization of the value function

Our first main result identifies the value function

$$\mathcal{V} : [0,T]\times\overline{\mathcal{S}} \to \mathbb{R}_+, \qquad (t,x) \mapsto \mathcal{V}(t,x) \triangleq \sup_{\Lambda\in\mathcal{A}(t,x)} \mathbb{E}\Big[U_{\mathrm{L}}\Big(X_T^{t,x,\Lambda}\Big)\Big] \tag{12}$$

as the unique continuous viscosity solution of the dynamic programing equation associated with the retail investor's portfolio optimization problem. In order to state this result, we need to introduce some notation. First, we denote the *infinitesimal generator* of the uncontrolled state process by

$$\mathcal{L}[\varphi](t,x) \triangleq -\frac{\partial\varphi}{\partial t}(t,x) - rx_0\frac{\partial\varphi}{\partial x_0}(t,x) - \mu x_1\frac{\partial\varphi}{\partial x_1}(t,x) - \frac{1}{2}\sigma^2x_1^2\frac{\partial^2\varphi}{\partial x_1^2}(t,x)$$

for all $(t,x)\in[0,T]\times\overline{\mathcal{S}}$ and every sufficiently smooth function $\varphi : [0,T]\times\overline{\mathcal{S}}\to\mathbb{R}$. Second, for every locally bounded function $\varphi : [0,T]\times\overline{\mathcal{S}}\to\mathbb{R}$ we define[6] the *maximum operator*

$$\mathcal{M}[\varphi](t,x) \triangleq \begin{cases} \sup_{\Delta\in\mathcal{D}(x)}\varphi(t,\Gamma(x,\Delta)) & \text{if } \mathcal{D}(x)\neq\emptyset \\ \inf_{(\bar t,\bar x)\in[0,T]\times\overline{\mathcal{S}}_\emptyset}[\varphi(\bar t,\bar x)-1] & \text{if } \mathcal{D}(x)=\emptyset \end{cases}$$

for all $(t,x)\in[0,T]\times\overline{\mathcal{S}}$. Finally, for $n\in\mathbb{N}_0$, we write $\mathcal{M}^n$ for the n-fold concatenation of the maximum operator, that is,

$$\mathcal{M}^n[\varphi] \triangleq \underbrace{\mathcal{M}\circ\cdots\circ\mathcal{M}}_{n\text{-times}}[\varphi],$$

where $\mathcal{M}^0[\varphi]=\varphi$ is taken to be the identity. With this notation in place, the first main result of this article can be stated as follows.

Main Result 1 (Viscosity Characterization). *The value function $\mathcal{V}$ defined in (12) satisfies*

$$0\le\mathcal{V}(t,x)\le K(1+|x|^p), \qquad (t,x)\in[0,T]\times\overline{\mathcal{S}} \tag{13}$$

for $p\in(0,1)$ from (11) and some $K>0$, and it is a continuous viscosity solution of the quasi-variational inequalities (QVIs)

$$\min\{\mathcal{L}[\mathcal{V}](t,x), \mathcal{V}(t,x)-\mathcal{M}[\mathcal{V}](t,x)\}=0, \qquad (t,x)\in[0,T)\times\mathcal{S} \tag{14}$$

with boundary/terminal conditions

$$\mathcal{V}(t,x) = \max_{n\in\mathbb{N}_0} \mathcal{M}^n[U_{\mathrm{L}}](x), \qquad (t,x) \in ([0,T]\times\{0\}) \cup (\{T\}\times\overline{\mathcal{S}}).$$

Moreover, $\mathcal{V}$ is unique within the class of functions satisfying (13).

Proof. This follows from Theorems 5.4, Theorem 5.11 and Theorem 5.12 in Section 5. □

The boundary/terminal conditions in Main Result 1 arise from the observation that for certain cost structures it might be less costly to liquidate a portfolio through multiple small trades instead of one large trade. Main Result 1 not only provides a characterization of the value function $\mathcal{V}$ for the retail investor's portfolio problem, but simultaneously demonstrates that $\mathcal{V}$ is continuous. This is the key ingredient required to explicitly construct optimal trading strategies; we elaborate on this in the following.

3.2 | Construction of optimal trading strategies

We first define a candidate optimal strategy in terms of the continuation region $\mathcal{C}$ and the intervention region $\mathcal{I}$ induced by the value function $\mathcal{V}$, that is,[7]

$$\mathcal{C} \triangleq \{(t,x) \in [0,T]\times\overline{\mathcal{S}} : \mathcal{V}(t,x) > \mathcal{M}[\mathcal{V}](t,x)\}, \tag{15}$$

$$\mathcal{I} \triangleq \{(t,x) \in [0,T]\times\overline{\mathcal{S}} : \mathcal{V}(t,x) = \mathcal{M}[\mathcal{V}](t,x)\}. \tag{16}$$

Main Result 1 guarantees that $\mathcal{V} \geq \mathcal{M}[\mathcal{V}]$ and hence the sets $\mathcal{C}$ and $\mathcal{I}$ partition the state space. The candidate optimal strategy is intuitively described as follows: Do nothing as long as the portfolio remains inside the continuation region $\mathcal{C}$; if the intervention region $\mathcal{I}$ is hit, trade a volume that corresponds to a maximizer of $\mathcal{M}[\mathcal{V}]$. Therefore, we subsequently also refer to $\mathcal{C}$ as the no-trading region and $\mathcal{I}$ as the trading region.

To make this precise, note that since $\mathcal{V}$ is continuous and each of the sets $\mathcal{D}(x)$ is compact, the measurable selection result in Schäl (1974) implies that there exists a Borel measurable function

$$\delta : [0,T]\times(\overline{\mathcal{S}} \setminus \mathcal{S}_\emptyset) \to \mathbb{R}, \qquad (t,x) \mapsto \delta(t,x),$$

that satisfies

$$\delta(t,x) \in \mathcal{D}(x) \quad \text{and} \quad \mathcal{M}[\mathcal{V}](t,x) = \mathcal{V}(t,\Gamma(x,\delta(t,x))), \quad (t,x) \in [0,T]\times(\overline{\mathcal{S}} \setminus \mathcal{S}_\emptyset).$$

Given an initial time $t \in [0,T]$ and an initial portfolio $x \in \overline{\mathcal{S}}$, we set $\tau_0^* \triangleq t$ and define the candidate optimal trading strategy $\Lambda^* = \{(\tau_k^*, \Delta_k^*)\}_{k\in\mathbb{N}}$ iteratively by setting[8]

$$\tau_k^* \triangleq \inf\left\{u \in [\tau_{k-1}^*, T] : \left(u, X_{u-}^{t,x,\Lambda^*}\right) \in \mathcal{I}\right\} \quad \text{and} \quad \Delta_k^* \triangleq \delta\left(\tau_k^*, X_{\tau_k^*-}^{t,x,\Lambda^*}\right)\mathbb{1}_{\{\tau_k^* \leq T\}} \tag{17}$$

TABLE 1 Model parameters for numerical simulations

r	μ	σ	p	T
3.0%	10.2%	40.0%	0.1	1

TABLE 2 Parameters for the various cost functions

Cost Structure	C_{min}	C_{max}	c	D
$C(\Delta) = C_{min}$	8.9	–	–	–
$C(\Delta) = C_{min} + c\|\Delta\|$	8.9	–	0.25%	–
$C(\Delta) = C_{min}\mathbb{1}_{[0,D]}(\|\Delta\|) + C_{max}\mathbb{1}_{(D,\infty)}(\|\Delta\|)$	8.9	58.9	–	10,000
$C(\Delta) = \min\{\max\{C_{min}, c\|\Delta\|\}, C_{max}\}$	8.9	58.9	0.25%	–

for each $k \in \mathbb{N}$. Our second main result demonstrates that this iteration is well-defined, and that Λ^* is optimal for the retail investor's portfolio problem. See also Figure 4 for a sample path of an optimal portfolio process.

Main Result 2 (Optimal Strategy). *Let $(t,x) \in [0,T] \times \overline{\mathcal{S}}$. Then $\Lambda^* = \{(\tau_k^*, \Delta_k^*)\}_{k\in\mathbb{N}}$ defined in terms of (15) and (16) is well-defined and optimal for the retail investor's portfolio problem, that is,*

$$\Lambda^* \in \mathcal{A}(t,x) \quad \text{and} \quad \mathcal{V}(t,x) = \mathbb{E}\left[U_L\left(X_T^{t,x,\Lambda^*}\right)\right].$$

Proof. This follows immediately from Theorem 5.11 and Theorem 5.12 in Section 5. □

Together, Main Results 1 and 2 provide a complete solution of the retail investor's portfolio problem. In particular, the retail investor's optimal trading strategy is fully described by the no-trading region $\mathcal{C}$ and the target positions on its boundary; these can be identified numerically by solving the QVIs (14).

4 | ANALYSIS OF OPTIMAL TRADING STRATEGIES

In this section, we analyze, illustrate and discuss the structure of optimal trading strategies for the retail investor's portfolio problem in detail. Unless stated otherwise, quantitative results are based on the market parameters in Table 1. The coefficients are such that, in the absence of transaction costs, the optimal fraction of wealth invested in the stock is given by

$$\pi^* \triangleq \frac{\mu - r}{(1-p)\sigma^2} = \frac{1}{2},$$

that is, the investor optimally holds equal amounts of money in the money market account and the stock at all times. Finally, we denote by $\tau \triangleq T - t$ the remaining investment horizon.

In what follows, we present numerical results for the optimal trading regions for the cost functions introduced in Section 2, that is, for (i) fixed costs, (ii) fixed plus proportional costs, (iii) piecewise fixed costs, and (iv) floored and capped proportional costs. The parameters for these cost structures are summarized in Table 2.

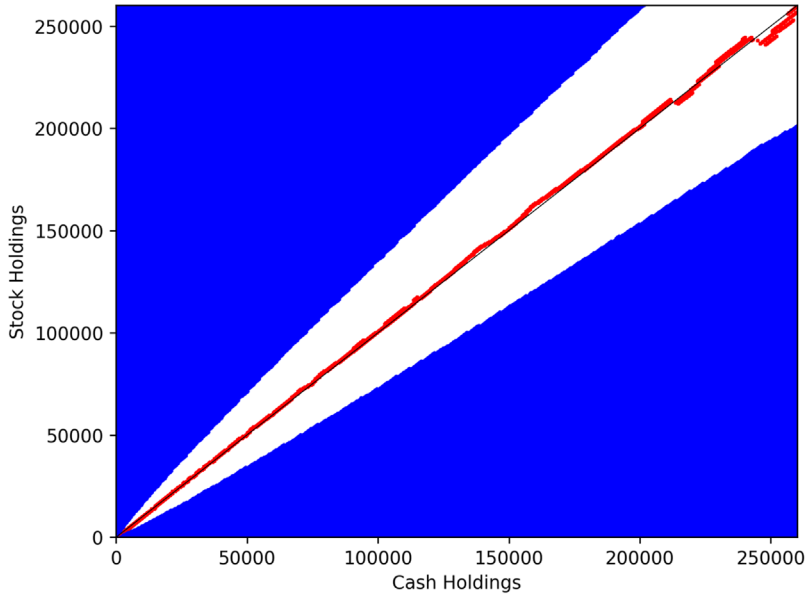

FIGURE 2 Trading regions for $\tau = 1$ with fixed costs [Color figure can be viewed at wileyonlinelibrary.com]

Our numerical results are obtained by solving the QVIs (14) using a finite difference scheme based on penalization of the non-local term, followed by a policy iteration; see Azimzadeh and Forsyth (2016) for a description of this scheme. In contrast to Azimzadeh and Forsyth (2016), we perform computations on a *triangular grid* such that $x_0 + x_1 \leq 600,000$ to avoid truncation in the computation of the maximum operator, which we found to significantly improve the stability of the algorithm. Trading regions are reported on a rectangular grid with $x_0, x_1 \leq 260,000$.

In all subsequent plots, the frictionless optimal positions are indicated by a solid black line, which we refer to as the *Merton line*. *Pre-trade portfolio positions*, that is, portfolios in the trading region $\mathcal{I} = \{\mathcal{V} = \mathcal{M}[\mathcal{V}]\}$, are colored in shades of blue; the associated *target positions*, that is, the portfolio positions resulting after optimal trades, are colored in shades of red.

4.1 | Optimal trading regions and target portfolios

Figures 2 to 7 depict the optimal trading regions of the various cost structures for $\tau = 1$; for larger time horizons, the optimal trading regions become stationary and hardly differ from the trading regions displayed here. Thus the following discussion applies as long as the outstanding investment horizon is not too small; boundary effects as terminal time approaches are discussed separately in Subsection 4.2. In general, as expected, the investor trades whenever the portfolio is sufficiently far away from the Merton line; optimal transactions always move the portfolio position towards it.

4.1.1 | Fixed costs and fixed plus proportional costs

Figures 2 and 3 depict the optimal trading regions for the fixed cost structure $C(\Delta) = C_{\min}$ and the fixed plus proportional cost structure $C(\Delta) = C_{\min} + c|\Delta|$, respectively. Portfolio optimization

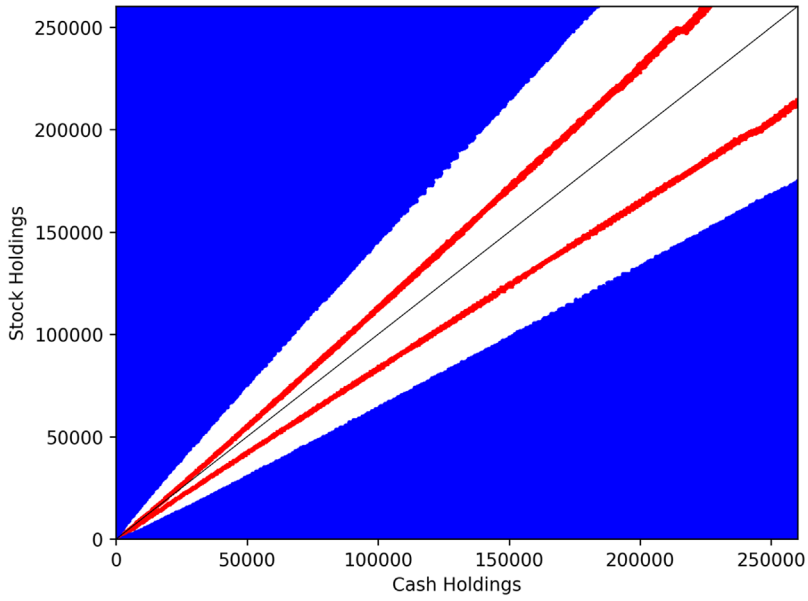

FIGURE 3 Trading regions for $\tau = 1$ with fixed plus proportional costs [Color figure can be viewed at wileyonlinelibrary.com]

problems with these cost functions are well-studied in the literature; see, for example, Altarovici et al., 2017; Belak & Christensen, 2019; Korn, 1998; Liu, 2004; Øksendal & Sulem, 2002.

In both cases, the no-trading region is approximately cone- or V-shaped and situated around the Merton line. Under the optimal strategy, the portfolio evolves uncontrolled inside the white no-trading region, that is, grows with an interest rate of r in the cash position and fluctuates with drift μ and volatility σ in the stock position; as soon as the portfolio process hits the blue trading region, the trader sells (when above the Merton line) or buys (when below the Merton line) shares of the stock to shift the portfolio onto a point on the red target portfolio set; see Figure 4 for a sample path of the optimal portfolio in the case of fixed plus proportional cost.[9]

In the case of purely fixed costs, only the frequency of trades but not the trading volume is penalized and hence the target portfolios coincide approximately[10] with the Merton line. In contrast, the fixed plus proportional cost structure additionally penalizes the size of a trade. Thus, in this case it is optimal to perform smaller trades towards the Merton line but not directly onto it.

4.1.2 | Piecewise constant costs

Since the piecewise constant cost structure

$$C(\Delta) = C_{\min} \mathbb{1}_{[0,D]}(|\Delta|) + C_{\max} \mathbb{1}_{(D,\infty)}(|\Delta|), \qquad \Delta \in \mathbb{R}$$

is discontinuous, it does not fit directly into the theoretical setup of this article. Nevertheless, our results are applicable if we consider a continuous piecewise linear approximation of this cost function in the interval $[D - \varepsilon, D + \varepsilon]$ for $\varepsilon > 0$ small. The resulting trading regions arising from

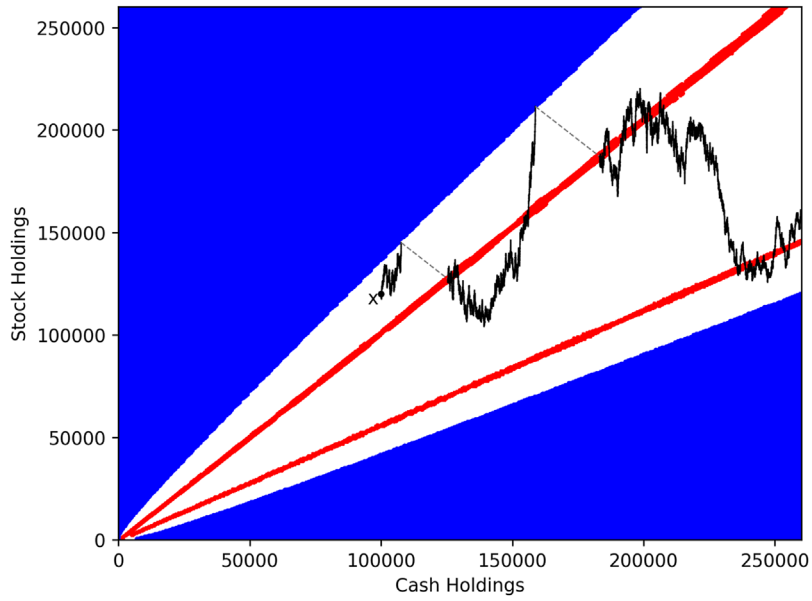

FIGURE 4 Sample path of an optimal portfolio process in the presence of fixed plus proportional costs. The portfolio starts in x and hits the trading region twice. When the trading region is hit, stocks are sold so that the portfolio is shifted onto the red target portfolio line. We use the parameters in Table 1 with interest rate and volatility replaced by 8.4% and 20%, respectively [Color figure can be viewed at wileyonlinelibrary.com]

this approximation are depicted in Figures 5 and 6. To distinguish trade sizes, we use light shades of blue/red for optimal trades Δ incurring the small cost, that is, $C(\Delta) = C_{min}$, and dark shades of blue/red for optimal trades incurring the large cost, that is, $C(\Delta) = C_{max}$.

A new feature arising with this cost structure is that target portfolios are not necessarily located in the no-trading region. For this reason, Figure 5 plots the blue trading region on top of the red target portfolios, whereas Figure 6 shows the red target portfolios plotted on top of the blue trading region. As can be seen, there is a significant overlap of the trading region and the target portfolios. Each portfolio in this intersection corresponds to a position for which it is optimal to perform at least two small trades back to back. This is a consequence of the observation that for large trades $|\Delta| > D$ it is cheaper to split the trade into n smaller trades of size $|\Delta|/n$ provided that $nC_{min} < C_{max}$. Given the parameters in Table 2, we find that splitting a trade of size $|\Delta|$ into n trades of size $|\Delta|/n$ is optimal if and only if

$$(n-1)\cdot 10,000 < |\Delta| \leq n \cdot 10,000, \qquad n = 1, \dots, 6.$$

For trade sizes exceeding 60,000$ it is never optimal to split the trade. We note, however, that such behavior can only be observed at the first trading date or possibly during terminal liquidation. That is, as soon as the portfolio has reached the no-trading region once and ignoring terminal liquidation, the investor will perform only a single trade at a time. Since the portfolios on the boundary of the trading region are shaded in light blue, we find that these single trades incur the small cost C_{min}. This implies that large trades with $C(\Delta) = C_{max}$, which are located at the top left and bottom right in Figures 5 and 6 and take the portfolio directly onto the Merton line, only occur at initial time.

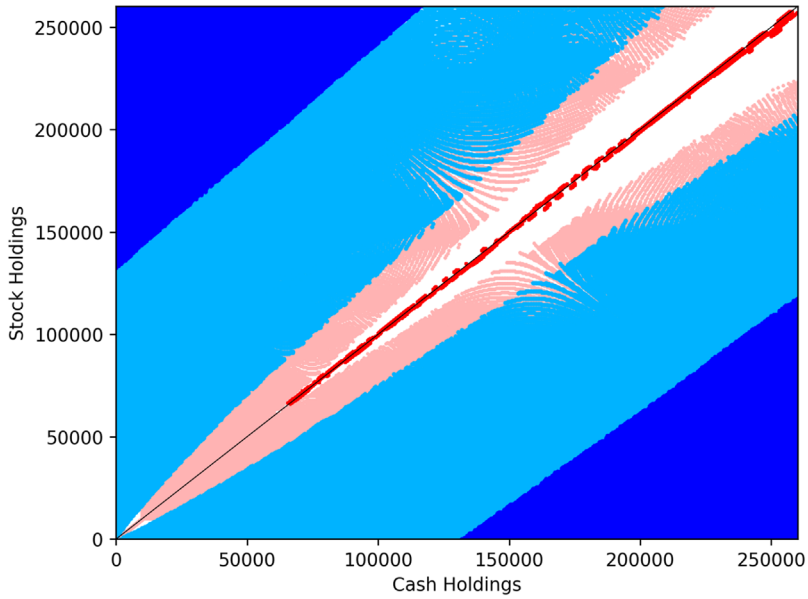

FIGURE 5 Trading regions for $\tau = 1$ with piecewise fixed costs; trading regions plotted *on top* of target portfolios [Color figure can be viewed at wileyonlinelibrary.com]

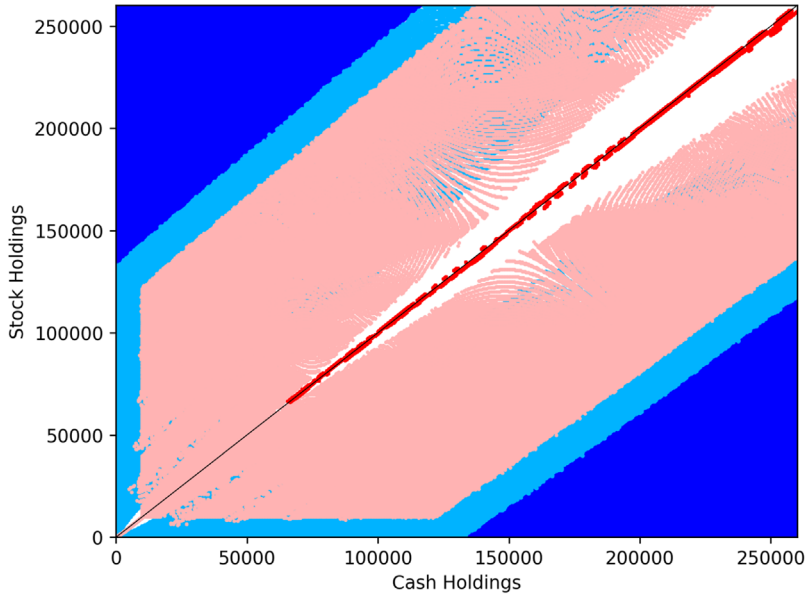

FIGURE 6 Trading regions for $\tau = 1$ with piecewise fixed costs; trading regions plotted *below* target portfolios [Color figure can be viewed at wileyonlinelibrary.com]

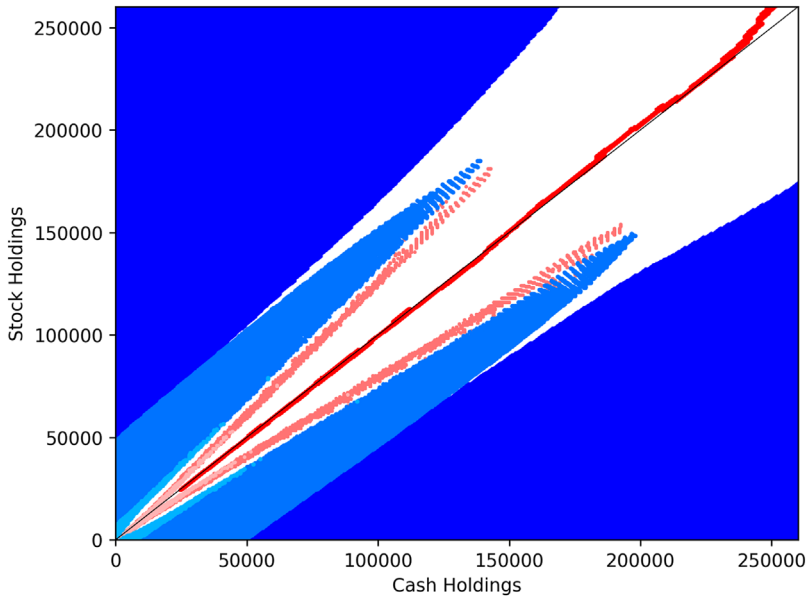

FIGURE 7 Trading regions for $\tau = 1$ with floored and capped proportional costs [Color figure can be viewed at wileyonlinelibrary.com]

4.1.3 | Floored and capped proportional costs

Let us now turn to floored and capped proportional costs $C(\Delta) = \min\{\max\{C_{min}, c|\Delta|\}, C_{max}\}$. Note that, given the cost parameters in Table 2, we have

$$C(\Delta) = C_{min} \quad \text{for } |\Delta| \leq 3,560, \tag{18}$$

$$C(\Delta) = c|\Delta| \quad \text{for } 3,560 \leq |\Delta| \leq 23,560, \tag{19}$$

$$C(\Delta) = C_{max} \quad \text{for } 23,560 \leq |\Delta|. \tag{20}$$

The resulting trading region for $\tau = 1$ is depicted in Figure 7. To distinguish the different cost regimes, we use three different shades of blue and red: We use light shades for the *floored cost region*, where optimal trades incur the minimal transaction cost C_{min}; medium shades for the *proportional cost region*, where optimal trades incur transaction costs in the interval (C_{min}, C_{max}); and dark shades for the *capped cost region* with optimal transaction cost C_{max}.

A surprising feature of the trading regions in Figure 7 is the emergence of the two white v-shaped areas splitting the intervention region with proportional cost trades (medium blue) from the intervention region with capped cost trades (dark blue), resulting in the white vVv-shaped no-trading region. Note that portfolios inside the two outer v-shaped wedges (the two small white wedges bounded between the dark and medium blue trading regions) are no-trade positions, that is, the investor takes no action here. If the portfolio moves sufficiently far away from the Merton line, a capped cost trade onto the Merton line is performed[11], whereas, if her portfolio moves closer

to the Merton line, a proportional cost trade onto the medium red wedge is performed. Note that, once the optimal portfolio position leaves the outer two v-shaped wedges, it never returns there since the interest rate r is positive, so this can occur only for the first transaction.

The possibly most important insight from Figure 7 is that we are able to identify distinct wealth-dependent regimes of optimal transactions (disregarding the first trade, that is, disregarding the two outer v-shaped areas of the no-trading region, see above):

- In the *moderate wealth regime*, all target portfolios are located on a wedge around the Merton line, all optimal trades incur (approximately) the floored cost, and feature a transaction size of C_{min}/c (the largest trade size that still incurs the floored cost).[12] For our specific choice of model parameters, a portfolio is in the moderate wealth regime if the cash position is below 135,000$ for sell orders and below 190,000$ for buy orders.
- In the *large wealth regime*, all optimal transactions involve the capped cost and the target portfolios are on the Merton line. In particular, a retail investor with a large amount of wealth acts exactly as though she faced fixed transaction costs of size C_{max}. For sell orders (resp. buy orders), the large wealth regime consists of those portfolios with a cash position exceeding 135,000$ (resp. 190,000$).

The distinction between moderate and large wealth regimes serves as a possible explanation for the existence of the small outer v-shaped wedges of the no-trading region. In the moderate wealth regime, wealth levels are too small to come anywhere near trade sizes involving the capped cost C_{max} and hence the investor effectively faces floored proportional costs. Since such costs penalize trading volumes, the target portfolios are not located on the frictionless optimizer but form a wedge around it. Conversely, for large wealth levels, the investor effectively faces fixed costs of size C_{max} which do not penalize trade sizes and hence she only performs large trades onto the Merton line. Thus, as the wealth regime transitions from moderate to large and the wedge shaped target portfolios of the moderate wealth regime move further away from the Merton line, portfolio positions arise for which it is optimal to wait and let the portfolio evolve uncontrolled at first. Now either the portfolio moves further away from the Merton line, making a large trade with capped cost more favorable, or the portfolio moves closer to the Merton line, hence making a small trade onto the target portfolio wedge more attractive. One may therefore interpret the small v-shaped areas of the no-trading region as a transition region between the moderate and large wealth regime.

We note that the change of the optimal trade size from the moderate to the large wealth regime is discontinuous as it switches from $|\Delta| \approx C_{min}/c$ (the largest trade size incurring the floored cost C_{min}) to trade sizes exceeding C_{max}/c (the smallest trade size incurring the capped cost C_{max}). The transition between the wealth regimes hence corresponds to a paradigm shift: In the large wealth regime, the investor acts as if she only faced fixed costs of size C_{max} (see Figure 2), whereas in the moderate wealth regime the investor acts as if she only faced floored proportional costs. To substantiate the claim for the moderate wealth regime, we depict in Figure 8 the optimal trading regions in the absence of a cap, that is, for $C_{max} = \infty$. We observe that the lines making up the target portfolio wedge appear to run parallel to the boundary of the trading region and hence all optimal trade sizes are indeed approximately equal (ignoring once again any initial trade which might be larger). We also observe that the behavior here is different from the case of fixed plus proportional costs, where the target portfolios are not parallel to the boundary of the trading region; see Figure 3.

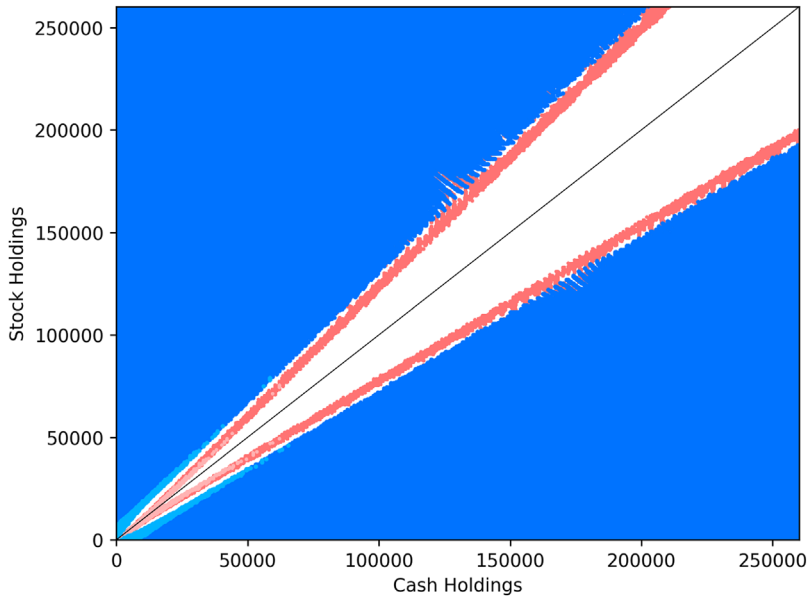

FIGURE 8 Trading regions for $\tau = 1$ with floored proportional costs without a cap [Color figure can be viewed at wileyonlinelibrary.com]

4.2 | Short investment time horizons

In this subsection we analyze the rich structure of the trading regions as the investment period tends to zero. We focus the discussion on floored and capped proportional costs.

4.2.1 | Horizons $\tau = 0.25$ and $\tau = 0.15$

Figures 9 and 10 display the optimal trading regions for time to maturity $\tau = 0.25$ and $\tau = 0.15$, respectively. In both cases, the qualitative structure is analogous to the case $\tau = 1$.

The main difference emerges in the moderate wealth regime. The target portfolios no longer form a wedge around the Merton line, but develop a kink on both the selling side (above the Merton line) and buying side (below the Merton line). This is due to the fact that the investor anticipates the end of the investment period, where the entire risky position is to be liquidated. Notice that both kinks of the restarting positions are in a vicinity of stock holdings of around 23,560$, which is exactly the threshold between proportional and capped cost trades; see (19). On the buying side, it becomes less and less attractive to trade towards this level, as any risky assets bought would have to be liquidated within a short time frame, incurring transaction costs twice. By contrast, for larger stock holdings, the liquidation at terminal time is expected to be in the capped cost region, thus bounding the liquidation cost. This causes the continuation region to widen faster for moderate wealth levels than for large wealth levels, producing the kink on the buying side.

The kink on the selling side emerges for a similar reason: As noted above, in the capped cost region it becomes less attractive to sell shares shortly before the end of the investment horizon.

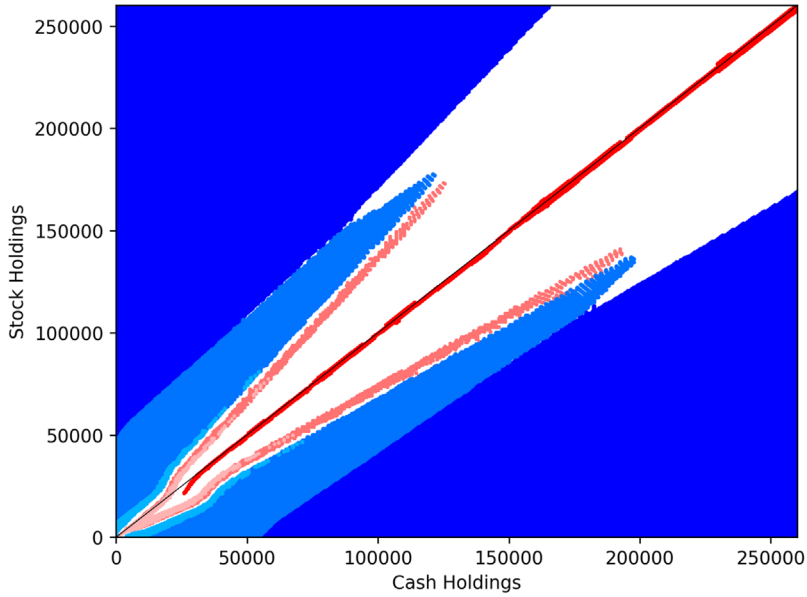

FIGURE 9 Trading regions for $\tau = 0.25$ with floored and capped proportional costs [Color figure can be viewed at wileyonlinelibrary.com]

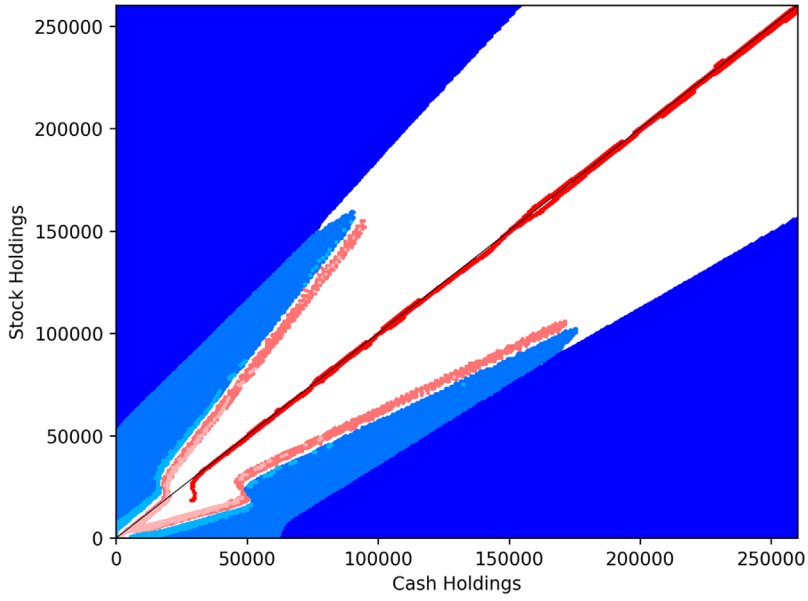

FIGURE 10 Trading regions for $\tau = 0.15$ with floored and capped proportional costs [Color figure can be viewed at wileyonlinelibrary.com]

With proportional costs, however, this makes almost no difference, as selling a part of the stock holdings before maturity and liquidating the rest at terminal time incurs approximately the same cost as liquidating the entire position at once. For this reason, the continuation region widens in the capped cost region (it is preferable to keep the portfolio until the end), whereas it shrinks in

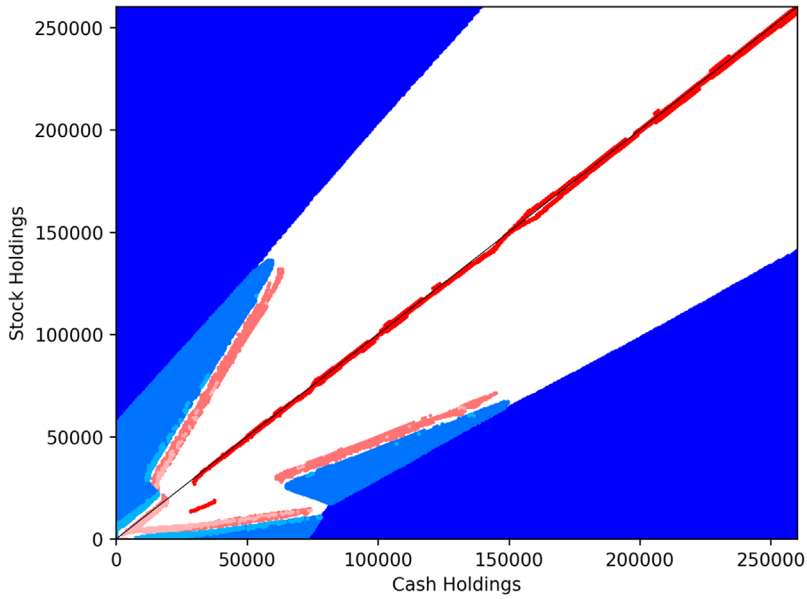

FIGURE 11 Trading regions for $\tau = 0.10$ with floored and capped proportional costs [Color figure can be viewed at wileyonlinelibrary.com]

the proportional cost region (the investor begins to liquidate the portfolio early, with the added benefit of taking it closer to the Merton line), thus causing the appearance of the kink on the selling side.

Finally, for $\tau = 0.15$, we observe that the target positions associated with capped cost trades for moderate wealth levels are located below the Merton line. This may be explained by the fact that these positions are further away from the intervention region than the Merton line, hence increasing the probability that no further trade is necessary before terminal time.

4.2.2 | Horizons $\tau = 0.10$, $\tau = 0.06$, $\tau = 0.03$ and $\tau = 0.01$

Figures 11 to 14 illustrate the evolution of the optimal trading regions as the investment horizon tends to zero.

Several effects emerge, most of which are explained by the difference in speed by which the continuation region widens in the different cost regions. In Figure 11, that is, for $\tau = 0.10$, we see that the target positions for floored/proportional cost transactions are no longer connected. On the buying side, for cash holdings below 25,000$ proportional trades disappear and are replaced by floored cost trades; in Figure 12, that is, for $\tau = 0.06$, proportional trades on the buying side are eliminated entirely. The novel feature in Figure 13, that is, for $\tau = 0.03$, is that the trading region is given by three connected regions instead of two, as the sell-side trade region with capped costs splits from the sell-side trade region with floored and proportional costs. Finally, in Figure 14, that is, for $\tau = 0.01$, it is no longer optimal to make any trades unless the portfolio is on the selling side and in the proportional cost region (see the discussion above).

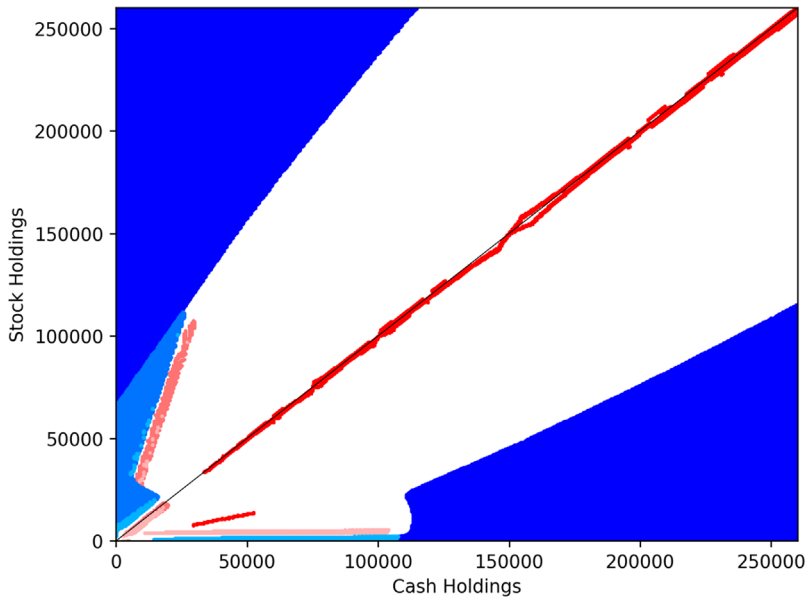

FIGURE 12 Trading regions for $\tau = 0.06$ with floored and capped proportional costs [Color figure can be viewed at wileyonlinelibrary.com]

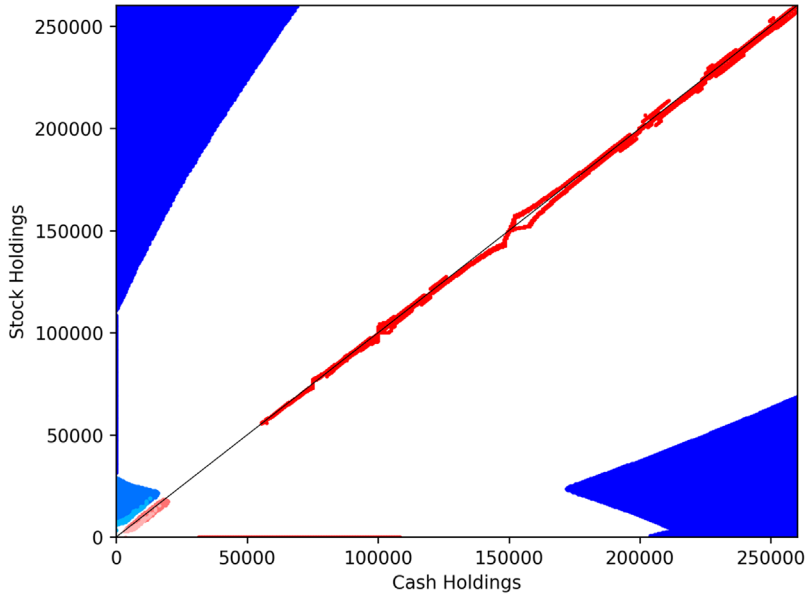

FIGURE 13 Trading regions for $\tau = 0.03$ with floored and capped proportional costs [Color figure can be viewed at wileyonlinelibrary.com]

5 | VISCOSITY CHARACTERIZATION AND OPTIMAL STRATEGIES

In this section we prove the two main results announced in Section 3: The viscosity characterization of the value function and the optimality of the candidate trading strategy.

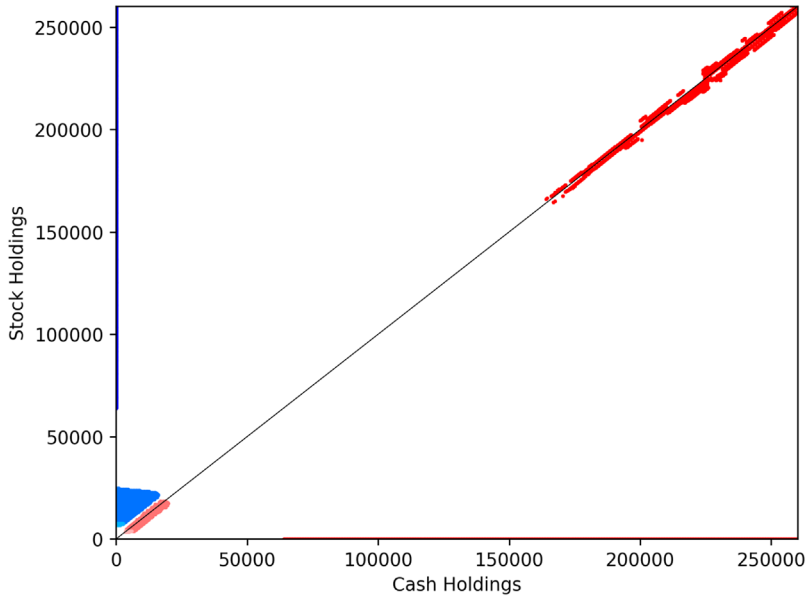

FIGURE 14 Trading regions for $\tau = 0.01$ with floored and capped proportional costs [Color figure can be viewed at wileyonlinelibrary.com]

Our mathematical approach is based on the stochastic Perron's method and the superharmonic function technique, similarly as in the analysis of portfolio problems with fixed plus proportional costs in Belak and Christensen (2019). In the present setting, however, some key technical arguments can be sharpened and streamlined. Thus we directly characterize the smallest stochastic supersolution $\mathbb{V}$ as the unique viscosity solution of the Bellman equation (Theorems 5.4 and 5.11); then we define a candidate optimal strategy in terms of $\mathbb{V}$ and provide a verification theorem (Theorem 5.12) that simultaneously establishes optimality of the candidate strategy and the fact that $\mathbb{V}$ coincides with the value function.[13] The main advantage of this direct approach is that it is significantly easier to verify the viscosity supersolution property, as we can avoid the iterated optimal stopping approximation of the value function used in Belak and Christensen (2019).

5.1 | Preliminary results

We subsequently denote by LSC and USC the sets of lower and upper semicontinuous functions $h : [0,T] \times \overline{S} \to \mathbb{R}$, respectively. If $h : [0,T] \times \overline{S} \to \mathbb{R}$ is locally bounded, we denote its lower semicontinuous envelope by h_* and its upper semicontinuous envelope by h^*.

Lemma 5.1 (Semicontinuity of $\mathcal{M}$). *For any function $h : [0,T] \times \overline{S} \to \mathbb{R}_+$, the following holds:*

1. *If $h \in$ USC, then $\mathcal{M}[h]^*(t,x) = \mathcal{M}[h](t,x)$ for all $(t,x) \in [0,T] \times \overline{S}$.*
2. *If $h \in$ LSC, then $\mathcal{M}[h]_*(t,x) = \mathcal{M}[h](t,x)$ for all $(t,x) \in [0,T] \times (\overline{S} \setminus \overline{S}_\emptyset)$.*

Proof. ad 1. Let $h \in$ USC. To show that $\mathcal{M}[h]^* = \mathcal{M}[h]$, it obviously suffices to show that $\mathcal{M}[h]$ is upper semicontinuous. For this, let $(t,x) \in [0,T] \times \overline{S}$ and choose a sequence $\{(t_k, x_k)\}_{k \in \mathbb{N}} \subset [0,T] \times \overline{S}$ converging to (t,x). Since $\mathcal{M}[h]$ is constant on $[0,T] \times S_\emptyset$ and $S_\emptyset$ is open relative to $\overline{S}$,

we may assume that $x \in \overline{S} \setminus S_\emptyset$. Moreover, by dropping to a subsequence, we may assume that either

$$x_k \in \overline{S} \setminus S_\emptyset \quad \text{for all } k \in \mathbb{N} \quad \text{or} \quad x_k \in S_\emptyset \quad \text{for all } k \in \mathbb{N}.$$

In the latter case, we have $x \in \partial S_\emptyset$ and $x_0 + x_1 - \mathrm{C}(-x_1) = 0$, that is, $\chi(x_0) = -x_1$. Thus, we see that $\mathcal{D}(x) = \{-x_1\}$ and $\Gamma(x, -x_1) = 0$; but this and the definition of $\mathcal{M}[h]$ on $[0,T] \times S_\emptyset$ imply that

$$\limsup_{k\to\infty} \mathcal{M}[h](t_k, x_k) \le \limsup_{k\to\infty} h(t_k, 0) \le h(t, 0) = h(t, \Gamma(x, -x_1)) = \mathcal{M}[h](t, x),$$

thus giving upper semicontinuity. Hence in the following we assume that $x_k \in \overline{S} \setminus S_\emptyset$ for all $k \in \mathbb{N}$. We drop to a subsequence if necessary to ensure that

$$\limsup_{k\to\infty} \mathcal{M}[h](t_k, x_k) = \lim_{k\to\infty} \mathcal{M}[h](t_k, x_k).$$

For each $k \in \mathbb{N}$, the set $\mathcal{D}(x_k)$ is non-empty and compact. By upper semicontinuity of h, we therefore find $\Delta_k \in \mathcal{D}(x_k)$ such that

$$\mathcal{M}[h](t_k, x_k) = h(t_k, \Gamma(x_k, \Delta_k)).$$

Dropping to yet another subsequence if necessary, the representation (6) of $\mathcal{D}(x_k)$ and continuity of χ show that $\{\Delta_k\}_{k\in\mathbb{N}}$ converges to some $\Delta \in \mathcal{D}(x)$. But then upper semicontinuity of h yields

$$\begin{aligned} \limsup_{k\to\infty} \mathcal{M}[h](t_k, x_k) &= \lim_{k\to\infty} \mathcal{M}[h](t_k, x_k) \\ &= \lim_{k\to\infty} h(t_k, \Gamma(x_k, \Delta_k)) \le h(t, \Gamma(x, \Delta)) \le \mathcal{M}[h](t, x), \end{aligned} \tag{21}$$

which concludes the first part of the proof.

ad 2. Now suppose that $h \in \mathrm{LSC}$. We fix $(t, x) \in [0,T] \times (\overline{S} \setminus \overline{S}_\emptyset)$ and choose an arbitrary sequence $\{(t_k, x_k)\}_{k\in\mathbb{N}} \subset [0,T] \times \overline{S}$ converging to (t, x). Since $x \notin \overline{S}_\emptyset$, it follows that $x_k \notin \overline{S}_\emptyset$ eventually and hence, without loss of generality, $\mathcal{D}(x_k) \neq \emptyset$ for all $k \in \mathbb{N}$. Now take as given some $\Delta \in \mathcal{D}(x)$. By (6) and continuity of χ, for each $k \in \mathbb{N}$, we find $\Delta_k \in \mathcal{D}(x_k)$ such that $\Delta_k \to \Delta$ as $k \to \infty$. But then

$$\liminf_{k\to\infty} \mathcal{M}[h](t_k, x_k) \ge \liminf_{k\to\infty} h\big(t_k, \Gamma(x_k, \Delta_k)\big) \ge h\big(t, \Gamma(x, \Delta)\big).$$

Since $\Delta \in \mathcal{D}(x)$ was chosen arbitrarily, this implies

$$\liminf_{k\to\infty} \mathcal{M}[h](t_k, x_k) \ge \mathcal{M}[h](t, x),$$

that is, $\mathcal{M}[h]$ is lower semicontinuous on $[0,T] \times (\overline{S} \setminus \overline{S}_\emptyset)$ and thus equal to $\mathcal{M}[h]_*$. □

We close this subsection by introducing a suitable notion of viscosity solutions of the QVIs (14). Since (14) are the only quasi-variational inequalities in this paper, we henceforth briefly refer to (14) as *the QVIs*.

Definition 5.2 (Viscosity Solutions of QVIs). Let $h : [0,T] \times \overline{S} \to \mathbb{R}$ be locally bounded.

(i) We say that h is a *viscosity subsolution* of the QVIs if, for all $(t,x) \in [0,T) \times S$ and all $\varphi \in C^2([0,T) \times S)$ with $\varphi \geq h^*$ and $\varphi(t,x) = h^*(t,x)$, we have

$$\min\{\mathcal{L}[\varphi](t,x), h^*(t,x) - \mathcal{M}[h^*]^*(t,x)\} \leq 0. \tag{22}$$

(ii) We say that h is a *viscosity supersolution* of the QVIs if, for all $(t,x) \in [0,T) \times S$ and all $\varphi \in C^2([0,T) \times S)$ with $\varphi \leq h_*$ and $\varphi(t,x) = h_*(t,x)$, we have

$$\min\{\mathcal{L}[\varphi](t,x), h_*(t,x) - \mathcal{M}[h_*]_*(t,x)\} \geq 0. \tag{23}$$

(iii) h is called a *viscosity solution* of the QVIs if it is both a viscosity sub- and supersolution.

Remark. If h is continuous, we see that the left-hand side of (22) (resp. (23)) corresponds to the upper (resp. lower) semicontinuous envelope of

$$\min\{\mathcal{L}[\varphi](t,x), h(t,x) - \mathcal{M}[h](t,x)\},$$

which is common practice when defining viscosity solutions of discontinuous differential equations. For the supersolution property, we have $\mathcal{M}[h^*]^* = \mathcal{M}[h^*]$, so we could technically refrain from passing to the upper semicontinuous envelope here. However, in general it is not true that $\mathcal{M}[h_*]_* = \mathcal{M}[h_*]$, as the equality may fail on $[0,T] \times \partial S_\emptyset$.

5.2 | A comparison principle for the QVIs

The aim of this subsection is to establish a comparison principle that is sufficiently strong to establish uniqueness and continuity for viscosity solutions of the QVIs. The comparison principle is obtained by perturbing viscosity solutions with a (strict) classical supersolution, an idea which goes back to Ishii (1993). The supersolution we use is given by the following result.

Lemma 5.3 (Classical Supersolution). *Let $\varepsilon \in \{0,1\}$, $q \in [p,1)$, $\lambda > q \max\{r,\mu,0\}$, and $C > 0$ and define*

$$\Psi_\varepsilon^q : [0,T] \times \overline{S} \to \mathbb{R}_+, \qquad (t,x) \mapsto \Psi_\varepsilon^q(t,x) \triangleq C(\varepsilon + x_0 + x_1)^q e^{\lambda(T-t)}. \tag{24}$$

Then there exists a continuous function $\kappa : \overline{S} \to \mathbb{R}_+$ that is strictly positive on S such that

$$\min\left\{\mathcal{L}\left[\Psi_\varepsilon^q\right](t,x), \Psi_\varepsilon^q(t,x) - \mathcal{M}\left[\Psi_\varepsilon^q\right](t,x)\right\} \geq \kappa(x) > 0, \qquad (t,x) \in [0,T) \times S.$$

Proof. Fix $(t,x) \in [0,T) \times \mathcal{S}$. An explicit computation shows that

$$\mathcal{L}\left[\Psi_\varepsilon^q\right](t,x) = Ce^{\lambda(T-t)}(\varepsilon + x_0 + x_1)^{q-1}\Bigg[\lambda\varepsilon + (\lambda - qr)x_0 + (\lambda - q\mu)x_1 \tag{25}$$

$$+ \frac{1}{2}(1-q)q\sigma^2\frac{x_1^2}{\varepsilon + x_0 + x_1}\Bigg] \tag{26}$$

$$\geq (\lambda - q\max\{r,\mu,0\})Ce^{\lambda(T-t)}(\varepsilon + x_0 + x_1)^q > 0. \tag{27}$$

Moreover, whenever $x \notin \mathcal{S}_\emptyset$, we have $x_0 + x_1 \geq \mathrm{C}_{\min}$ and

$$\Psi_\varepsilon^q(t,x) - \mathcal{M}\left[\Psi_\varepsilon^q\right](t,x) = Ce^{\lambda(T-t)}\inf_{\Delta \in D(x)}\left[(\varepsilon + x_0 + x_1)^q - (\varepsilon + x_0 + x_1 - \mathrm{C}(\Delta))^q\right] \tag{28}$$

$$= Ce^{\lambda(T-t)}\left[(\varepsilon + x_0 + x_1)^q - (\varepsilon + x_0 + x_1 - \mathrm{C}_{\min})^q\right] > 0. \tag{29}$$

Since $\Psi_\varepsilon^q(t,x) - \mathcal{M}[\Psi_\varepsilon^q](t,x) \geq \Psi_\varepsilon^q(t,x) - \Psi_\varepsilon^q(t,x) + 1 = 1$ if $x \in \mathcal{S}_\emptyset$, this completes the proof. □

Before we state the comparison principle, we introduce some short-hand notation by defining[14]

$$F_{\mathcal{L}} : \overline{\mathcal{S}} \times \mathbb{R} \times \mathbb{R}^2 \times \mathbb{S}^2 \to \mathbb{R}$$

via

$$F_{\mathcal{L}}(x,a,b,M) \triangleq -a - rx_0b_0 - \mu x_1 b_1 - \frac{1}{2}\sigma^2 x_1^2 M_{22}$$

for all $x = (x_0,x_1) \in \overline{\mathcal{S}}$, $a \in \mathbb{R}$, $b = (b_0,b_1) \in \mathbb{R}^2$, and $M = (M_{ij})_{i=1,2}^{j=1,2} \in \mathbb{S}^2$. Note that

$$\mathcal{L}[\varphi](t,x) = F_{\mathcal{L}}\left(x, \frac{\partial\varphi}{\partial t}(t,x), \frac{\partial\varphi}{\partial x}(t,x), \frac{\partial^2\varphi}{\partial x^2}(t,x)\right), \qquad (t,x) \in [0,T) \times \mathcal{S},$$

for every $\varphi \in \mathrm{C}^2([0,T) \times \mathcal{S})$.

Theorem 5.4 (Comparison Principle). *Let $u \in$ USC and $v \in$ LSC be a viscosity subsolution of the QVIs and a viscosity supersolution of the QVIs, respectively. Suppose that*

$$u(t,0) = 0 \quad \textit{for all } t \in [0,T] \quad \textit{and} \quad u(T,x) \leq v(T,x) \quad \textit{for all } x \in \overline{\mathcal{S}}, \tag{30}$$

and that there exists a constant $K > 0$ such that

$$0 \leq u(t,x), v(t,x) \leq K(1 + |x|^p), \qquad (t,x) \in [0,T] \times \overline{\mathcal{S}}. \tag{31}$$

Then v dominates u everywhere, that is,

$$u(t,x) \le v(t,x), \qquad (t,x) \in [0,T] \times \overline{S}.$$

Proof. Fix $q \in (p,1)$ and choose $C > 0$ sufficiently large such that $u, v \le \Psi_1^q$ on $[0,T] \times \overline{S}$, where Ψ_1^q is given by (24); this is possible by (31). For any $\eta > 1$, we define $u_\eta \in \mathrm{USC}$ and $v_\eta \in \mathrm{LSC}$ by

$$u_\eta \triangleq \frac{\eta+1}{\eta} u - \frac{1}{\eta}\Psi_1^q \quad \text{and} \quad v_\eta \triangleq \frac{\eta-1}{\eta} v + \frac{1}{\eta}\Psi_1^q.$$

We proceed to show that $u_\eta \le v_\eta$ on $[0,T] \times \overline{S}$, which implies the result once we send $\eta \to \infty$. We argue by contradiction and suppose that

$$u_\eta(t^*,x^*) > v_\eta(t^*,x^*) \quad \text{for some } (t^*,x^*) \in [0,T] \times \overline{S}.$$

Step 1. For each $k \in \mathbb{N}_0$, we define $\phi_k : ([0,T] \times \overline{S})^2 \to \mathbb{R}$ by

$$\phi_k(t,x,\hat{t},\hat{x}) \triangleq u_\eta(t,x) - v_\eta(\hat{t},\hat{x}) - \frac{k}{2}[|t-\hat{t}|^2 + |x-\hat{x}|^2], \qquad (t,x),(\hat{t},\hat{x}) \in [0,T] \times \overline{S},$$

and set

$$\Theta_k \triangleq \sup_{(t,x),(\hat{t},\hat{x}) \in [0,T] \times \overline{S}} \phi_k(t,x,\hat{t},\hat{x}) \quad \text{and} \quad \Theta \triangleq \sup_{(t,x) \in [0,T] \times \overline{S}} \phi_0(t,x,t,x).$$

It is immediately seen that

$$0 < u_\eta(t^*,x^*) - v_\eta(t^*,x^*) \le \Theta \le \Theta_{k+1} \le \Theta_k \le \Theta_0, \qquad k \in \mathbb{N}.$$

This implies that every maximizing sequence for some Θ_k, $k \in \mathbb{N}_0$, must eventually be contained in the set

$$F \triangleq \{(t,x,\hat{t},\hat{x}) \in ([0,T] \times \overline{S})^2 : u_\eta(t,x) - v_\eta(\hat{t},\hat{x}) \ge 0\}.$$

Since u_η and $-v_\eta$ are upper semicontinuous, F is closed. Moreover, by (31) and the fact that $q > p$, F is bounded and hence compact. But then, for all $k \in \mathbb{N}_0$,

$$\Theta_k = \phi_k(t_k,x_k,\hat{t}_k,\hat{x}_k) < \infty \quad \text{for some } (t_k,x_k,\hat{t}_k,\hat{x}_k) \in F,$$

and after dropping to a subsequence we may assume that $\{(t_k,x_k,\hat{t}_k,\hat{x}_k)\}_{k \in \mathbb{N}}$ is convergent. Since $\Theta_k > 0$ and $u_\eta, -v_\eta \in \mathrm{USC}$, we have

$$\frac{k}{2}\left[|t_k - \hat{t}_k|^2 + |x_k - \hat{x}_k|^2\right] \le \sup_{(t,x,\hat{t},\hat{x}) \in F} [u_\eta(t,x) - v_\eta(\hat{t},\hat{x})] < \infty,$$

so we must have

$$(\bar{t},\bar{x}) \triangleq \lim_{k\to\infty}(t_k,x_k) = \lim_{k\to\infty}(\hat{t}_k,\hat{x}_k).$$

But then, since $\Theta \le \Theta_k$ and $u_\eta, -v_\eta \in \mathrm{USC}$, we obtain

$$\begin{aligned} 0 &\le \limsup_{k\to\infty} \frac{k}{2}\left[|t_k-\hat{t}_k|^2 + |x_k-\hat{x}_k|^2\right] \\ &= \limsup_{k\to\infty}[u_\eta(t_k,x_k) - v_\eta(\hat{t}_k,\hat{x}_k) - \Theta_k] \le u_\eta(\bar{t},\bar{x}) - v_\eta(\bar{t},\bar{x}) - \Theta \le 0. \end{aligned} \tag{32}$$

We have thus shown that

$$(\bar{t},\bar{x}) = \lim_{k\to\infty}(t_k,x_k) = \lim_{k\to\infty}(\hat{t}_k,\hat{x}_k) \quad \text{and} \quad \lim_{k\to\infty}\frac{k}{2}\left[|t_k-\hat{t}_k|^2 + |x_k-\hat{x}_k|^2\right] = 0, \tag{33}$$

$$\lim_{k\to\infty} u_\eta(t_k,x_k) = u_\eta(\bar{t},\bar{x}) \quad \text{and} \quad \lim_{k\to\infty} v_\eta(\hat{t}_k,\hat{x}_k) = v_\eta(\bar{t},\bar{x}), \tag{34}$$

and that

$$\lim_{k\to\infty}\Theta_k = \Theta = \phi_0(\bar{t},\bar{x},\bar{t},\bar{x}) = u_\eta(\bar{t},\bar{x}) - v_\eta(\bar{t},\bar{x}). \tag{35}$$

Note that (35) implies in particular that $\bar{t} < T$ and hence, without loss of generality, $t_k, \hat{t}_k < T$ for every $k \in \mathbb{N}$. Indeed, if this were not the case, (30) and the estimate $u, v \le \Psi_1^q$ would yield the contradiction

$$\Theta = u_\eta(\bar{t},\bar{x}) - v_\eta(\bar{t},\bar{x}) = u(T,\bar{x}) - v(T,\bar{x}) + \frac{1}{\eta}\left[u(T,\bar{x}) + v(T,\bar{x}) - 2\Psi_1^q(T,\bar{x})\right] \le 0.$$

Similarly, we cannot have $\bar{x} = 0$ since otherwise (30) and non-negativity of v_η and Ψ_1^q imply

$$\Theta = u_\eta(\bar{t},\bar{x}) - v_\eta(\bar{t},\bar{x}) \le u_\eta(\bar{t},0) = -\frac{1}{\eta}\Psi_1^q(\bar{t},0) \le 0.$$

We may therefore also assume that $x_k, \hat{x}_k \in S$ for every $k \in \mathbb{N}$.

Step 2. Since $(t_k,x_k),(\hat{t}_k,\hat{x}_k) \in [0,T)\times S$ for all $k \in \mathbb{N}$ by Step 1, we can apply Ishii's lemma, see Crandall et al., 1992, Theorem 3.2, to obtain $M_k, N_k \in \mathbb{S}^2$ with[15]

$$\begin{pmatrix} M_k & 0 \\ 0 & -N_k \end{pmatrix} \le 3k \begin{pmatrix} \mathrm{I} & -\mathrm{I} \\ -\mathrm{I} & \mathrm{I} \end{pmatrix} \tag{36}$$

such that[16]

$$(k(t_k-\hat{t}_k), k(x_k-\hat{x}_k), M_k) \in \overline{\mathcal{J}}^{2,+} u_\eta(t_k,x_k), \tag{37}$$

$$(k(t_k - \hat{t}_k), k(x_k - \hat{x}_k), N_k) \in \overline{\mathcal{J}}^{2,-} v_\eta(\hat{t}_k, \hat{x}_k). \tag{38}$$

Since u and v are, respectively, viscosity sub- and supersolutions and Ψ_1^q is a strict classical supersolution, the same argument as in Belak & Christensen, 2019, Proposition 4.2, shows that

$$-\frac{\bar{\kappa}}{\eta} \geq \min\{F_{\mathcal{L}}(x_k, k(t_k - \hat{t}_k), k(x_k - \hat{x}_k), M_k), u_\eta(t_k, x_k) - \mathcal{M}[u_\eta]^*(t_k, x_k)\}, \tag{39}$$

$$\frac{\bar{\kappa}}{\eta} \leq \min\{F_{\mathcal{L}}(\hat{x}_k, k(t_k - \hat{t}_k), k(x_k - \hat{x}_k), N_k), v_\eta(t_k, x_k) - \mathcal{M}[v_\eta]_*(t_k, x_k)\} \tag{40}$$

$$\leq F_{\mathcal{L}}(\hat{x}_k, k(t_k - \hat{t}_k), k(x_k - \hat{x}_k), N_k), \tag{41}$$

where $\bar{\kappa} \triangleq \inf_{(t,x,\hat{t},\hat{x}) \in F} \min\{\kappa(x), \kappa(\hat{x})\} > 0$ and κ is the continuous function provided by Lemma 5.3.

Step 3. Let us now argue that in (39), after dropping to a subsequence, we may assume that

$$F_{\mathcal{L}}(x_k, k(t_k - \hat{t}_k), k(x_k - \hat{x}_k), M_k) \leq -\frac{\bar{\kappa}}{\eta}, \qquad k \in \mathbb{N}. \tag{42}$$

We argue by contradiction and assume that this is not the case, that is, the latter inequality is only valid for at most finitely many $k \in \mathbb{N}$. By (39), this means that there exists $K \in \mathbb{N}$ with

$$u_\eta(t_k, x_k) \leq \mathcal{M}[u_\eta]^*(t_k, x_k) - \frac{\bar{\kappa}}{\eta}, \qquad k \geq K. \tag{43}$$

Note that this is only possible if $x_k \notin S_\emptyset$ for all $k \geq K$, and hence we see that $\bar{x} \notin S_\emptyset$. Upon making K larger, using (35) and the convergence in (34), we furthermore find that

$$\Theta = u_\eta(\bar{t}, \bar{x}) - v_\eta(\bar{t}, \bar{x}) \leq u_\eta(t_k, x_k) - v_\eta(\hat{t}_k, \hat{x}_k) + \frac{\bar{\kappa}}{4\eta}, \qquad k \geq K. \tag{44}$$

Similarly, making K even larger if necessary, upper semicontinuity of $\mathcal{M}[u_\eta]^*$ yields

$$\mathcal{M}[u_\eta]^*(t_k, x_k) \leq \mathcal{M}[u_\eta]^*(\bar{t}, \bar{x}) + \frac{\bar{\kappa}}{4\eta}, \qquad k \geq K. \tag{45}$$

Since u_η is upper semicontinuous, we have $\mathcal{M}[u_\eta]^* = \mathcal{M}[u_\eta]$ by Lemma 5.1. But then due to compactness of $\mathcal{D}(\bar{x})$ and upper semicontinuity of u_η there exists $\Delta \in \mathcal{D}(\bar{x})$ such that

$$\mathcal{M}[u_\eta]^*(\bar{t}, \bar{x}) = \mathcal{M}[u_\eta](\bar{t}, \bar{x}) = u_\eta(\bar{t}, \Gamma(\bar{x}, \Delta)). \tag{46}$$

If we now successively plug (43), (45), and then (46) into (44), we arrive at

$$\Theta \leq u_\eta(\bar{t}, \Gamma(\bar{x}, \Delta)) - v_\eta(\hat{t}_k, \hat{x}_k) - \frac{\bar{\kappa}}{2\eta}, \qquad k \geq K. \tag{47}$$

If $\Gamma(\bar{x},\Delta)=0$, then (30) gives $u_\eta(\bar{t},\Gamma(\bar{x},\Delta))=0$ and hence we obtain the contradiction

$$\Theta \le -v_\eta(\hat{t}_k,\hat{x}_k)-\frac{\bar{\kappa}}{2\eta}<0.$$

We must therefore have $\Gamma(\bar{x},\Delta)\neq 0$; but since $\mathcal{D}(x)=\{-x_1\}$ for all $x\in\partial S_\emptyset$, this is only possible if $\bar{x}\notin \overline{S}_\emptyset$. Now (40) gives

$$v_\eta(\hat{t}_k,\hat{x}_k)\ge \mathcal{M}[v_\eta]_*(\hat{t}_k,\hat{x}_k)+\frac{\bar{\kappa}}{\eta},\qquad k\in\mathbb{N},$$

and by lower semicontinuity of $\mathcal{M}[v_\eta]_*$ we can assume that

$$\mathcal{M}[v_\eta]_*(\hat{t}_k,\hat{x}_k)\ge \mathcal{M}[v_\eta]_*(\bar{t},\bar{x})-\frac{\bar{\kappa}}{2\eta},\qquad k\ge K.$$

Since $\bar{x}\notin\overline{S}_\emptyset$ and v_η is lower semicontinuous, Lemma 5.1 and $\Delta\in\mathcal{D}(\bar{x})$ imply that

$$\mathcal{M}[v_\eta]_*(\bar{t},\bar{x})=\mathcal{M}[v_\eta](\bar{t},\bar{x})\ge v_\eta(\bar{t},\Gamma(\bar{x},\Delta)).$$

Plugging the latter three inequalities into (47) thus gives

$$\Theta\le u_\eta(\bar{t},\Gamma(\bar{x},\Delta))-v_\eta(\bar{t},\Gamma(\bar{x},\Delta))-\frac{\bar{\kappa}}{\eta}\le\Theta-\frac{\bar{\kappa}}{\eta}.$$

Since this is a contradiction, it follows that we may assume that (42) holds.

Step 4. Combining (41) from Step 2 and (42) from Step 3 shows that, for all $k\in\mathbb{N}$,

$$\frac{2\bar{\kappa}}{\eta}\le F_{\mathcal{L}}(\hat{x}_k,k(t_k-\hat{t}_k),k(x_k-\hat{x}_k),N_k)-F_{\mathcal{L}}(x_k,k(t_k-\hat{t}_k),k(x_k-\hat{x}_k),M_k).$$

Using (36), it is readily confirmed that there exists a constant $L>0$ such that

$$\frac{2\bar{\kappa}}{\eta}\le F_{\mathcal{L}}(\hat{x}_k,k(t_k-\hat{t}_k),k(x_k-\hat{x}_k),N_k)-F_{\mathcal{L}}(x_k,k(t_k-\hat{t}_k),k(x_k-\hat{x}_k),M_k)\le kL|x_k-\hat{x}_k|^2$$

for all $k\in\mathbb{N}$. Now send $k\to\infty$ and use (33) to obtain the final contradiction $2\bar{\kappa}/\eta\le 0$. □

In the same spirit, we have a comparison result at terminal time.

Corollary 5.5 (Terminal Comparison). *Let $u\in$ USC and $v\in$ LSC be such that*

$$\min\{u(T,x)-U_{\mathrm{L}}(x),u(T,x)-\mathcal{M}[u]^*(T,x)\}\le 0, x\in S, \tag{48}$$

$$\min\{v(T,x)-U_{\mathrm{L}}(x),v(T,x)-\mathcal{M}[v]_*(T,x)\}\ge 0, x\in S. \tag{49}$$

Suppose moreover that $u(T,0)=0$ *and that there exists a constant* $K>0$ *such that*

$$0 \le u(T,x), v(T,x) \le K(1+|x|^p), \qquad x \in \overline{S}. \tag{50}$$

Then v *dominates* u *everywhere on* $\overline{S}$*, that is,* $u(T,\cdot) \le v(T,\cdot)$.

Proof. This follows by repeating the proof of Theorem 5.4 with the time index set $[0,T]$ replaced by the singleton $\{T\}$. Indeed, upon assuming that there exist $\eta>0$ and $x^* \in \overline{S}$ with

$$u_\eta(T,x^*) - v_\eta(T,x^*) > 0,$$

we can follow the proof of Theorem 5.4 to obtain sequences $\{x_k\}_{k\in\mathbb{N}}, \{\hat{x}_k\}_{k\in\mathbb{N}} \subset \overline{S}$ converging to some $\bar{x} \in \overline{S}$ such that

$$\lim_{k\to\infty} u_\eta(T,x_k) = u_\eta(T,\bar{x}), \qquad \lim_{k\to\infty} v_\eta(T,\hat{x}_k) = v_\eta(T,\bar{x}), \qquad u_\eta(T,\bar{x}) - v_\eta(T,\bar{x}) > 0,$$

and such that

$$u_\eta(T,x_k) - U_{\mathrm{L}}(x_k) \le -\frac{\bar{\kappa}}{\eta} < 0 < \frac{\bar{\kappa}}{\eta} \le v_\eta(T,\hat{x}_k) - U_{\mathrm{L}}(\hat{x}_k)$$

for a strictly positive constant $\bar{\kappa}>0$ which does not depend on $k \in \mathbb{N}$. We note here that the latter chain of inequalities requires to choose the constants in Ψ_1^q such that $\Psi_1^q(T,\cdot) > U_{\mathrm{L}}$, which is clearly possible. Passing to the limit implies

$$u_\eta(T,\bar{x}) - U_{\mathrm{L}}(\bar{x}) < v_\eta(T,\bar{x}) - U_{\mathrm{L}}(\bar{x})$$

which contradicts $u_\eta(T,\bar{x}) - v_\eta(T,\bar{x}) > 0$ and hence completes the proof. □

5.3 | Stochastic supersolutions and the viscosity property

We next demonstrate that there exists a viscosity solution $\mathbb{V}$ of the QVIs. We use a variant of the stochastic Perron's method, in which it is shown that $\mathbb{V}$ can be constructed as the pointwise minimum of the set of stochastic supersolutions of the QVIs.

Definition 5.6 (Stochastic Supersolutions). We denote by $\mathbb{H}$ the set of *stochastic supersolutions* of the QVIs, that is, the set of all functions $h : [0,T] \times \overline{S} \to \mathbb{R}$ such that

($\mathbb{H}_1$) h is upper semicontinuous;
($\mathbb{H}_2$) There exists a constant $K>0$ such that

$$h(t,x) \le K(1+|x|^p), \qquad (t,x) \in [0,T] \times \overline{S};$$

($\mathbb{H}_3$) h satisfies the terminal condition

$$h(T,x) \ge U_{\mathrm{L}}(x), \qquad x \in \overline{S};$$

($\mathbb{H}_4$) h is decreasing in the direction of transactions, that is,

$$h(t,x) \geq \mathcal{M}[h](t,x), \qquad (t,x) \in [0,T] \times \overline{\mathcal{S}};$$

($\mathbb{H}_5$) For any pair of $\mathfrak{F}$-stopping times θ, ρ with $0 \leq \theta \leq \rho \leq T$ and any $\mathfrak{F}_\theta$-measurable random vector $\xi = (\xi_0, \xi_1)$ taking values in $\overline{\mathcal{S}}$ with $\mathbb{E}[|\xi|^2] < \infty$, we have

$$h(\theta,\xi) \geq \mathbb{E}\left[h\left(\rho, \bar{X}_\rho^{\theta,\xi}\right) \middle| \mathfrak{F}_\theta \right],$$

where $\bar{X}^{\theta,\xi} = \{\bar{X}_t^{\theta,\xi}\}_{t\in[\theta,T]}$ denotes the uncontrolled portfolio process with $\bar{X}_\theta^{\theta,\xi} = \xi$.

Let us first argue that the set of stochastic supersolutions is not empty.

Lemma 5.7 (Stochastic Supersolution). *Provided that $C > 1/p$, the function Ψ_0^p defined in (24) is a stochastic supersolution of the QVIs, that is, $\Psi_0^p \in \mathbb{H}$.*

Proof. Being continuous, Ψ_0^p evidently satisfies ($\mathbb{H}_1$). The growth condition ($\mathbb{H}_2$) is immediate from the definition of Ψ_0^p, and the terminal condition ($\mathbb{H}_3$) follows from the fact that

$$\Psi_0^p(T,x) = C(x_0 + x_1)^p \geq \frac{1}{p}(\mathrm{L}(x))^p = U_{\mathrm{L}}(x), \qquad x \in \overline{\mathcal{S}}.$$

The property ($\mathbb{H}_4$), that is, $\Psi_0^p - \mathcal{M}[\Psi_0^p] \geq 0$, has already been established in Lemma 5.3. Regarding ($\mathbb{H}_5$), we fix two $\mathfrak{F}$-stopping times θ, ρ with $0 \leq \theta \leq \rho \leq T$ and an $\mathfrak{F}_\theta$-measurable and $\overline{\mathcal{S}}$-valued random vector ξ with $\mathbb{E}[|\xi|^2] < \infty$. Denote by $\{\rho_k\}_{k\in\mathbb{N}}$ a localizing sequence of the local martingale

$$\int_\theta^\cdot \sigma \bar{X}_u^{\theta,\xi} \frac{\partial \Psi_0^p}{\partial x_1}\left(u, \bar{X}_u^{\theta,\xi}\right) \mathrm{d}W_u.$$

Then Itō's formula, the supersolution property of Ψ_0^p established in Lemma 5.3, and Fatou's lemma show that

$$\Psi_0^p(\theta,\xi) \geq \liminf_{k\to\infty} \mathbb{E}\left[\Psi_0^p\left(\rho_k \wedge \rho, \bar{X}_{\rho_k\wedge\rho}^{\theta,\xi}\right) \middle| \mathfrak{F}_\theta \right] \geq \mathbb{E}\left[\Psi_0^p\left(\rho, \bar{X}_\rho^{\theta,\xi}\right) \middle| \mathfrak{F}_\theta \right].$$

Thus Ψ_0^p satisfies ($\mathbb{H}_5$), and the proof is complete. □

For each $h \in \mathbb{H}$, we note that Fatou's lemma, (9) and (10) imply that

$h(\,\cdot\,, X^{t,x,\Lambda}_{\cdot})$ is a strong supermartingale for all $\Lambda \in \mathcal{A}(t,x)$ and $(t,x) \in [0,T] \times \overline{\mathcal{S}}$;

see, for example, Belak et al., 2017, Lemma 3.4, or Belak & Christensen, 2019, Lemma 5.2, for a detailed argument. Using $(\mathbb{H}_3)$, it follows in particular that

$$h(t,x) \geq \mathbb{E}\left[h\left(T, X_T^{t,x,\Lambda}\right)\right] \geq \mathbb{E}\left[U_{\mathrm{L}}\left(X_T^{t,x,\Lambda}\right)\right] \quad \text{for all } \Lambda \in \mathcal{A}(t,x) \text{ and } (t,x) \in [0,T] \times \overline{\mathcal{S}}$$

and thus $h \geq \mathcal{V} \geq 0$, where $\mathcal{V}$ is the value function of the retail investor's portfolio problem, see (12). Thus we have $\mathbb{V} \geq \mathcal{V} \geq 0$ where the function $\mathbb{V}$ is defined by

$$\mathbb{V} : [0,T] \times \overline{\mathcal{S}} \to \mathbb{R}_+, \qquad (t,x) \mapsto \mathbb{V}(t,x) \triangleq \inf_{h \in \mathbb{H}} h(t,x). \tag{51}$$

By Bayraktar & Sîrbu, 2012, Proposition 4.1, the infimum in (51) can be restricted to a countable subset of $\mathbb{H}$, which implies that $\mathbb{V} \in \mathbb{H}$. As a consequence, $\mathbb{V}$ is the pointwise minimum of the members of $\mathbb{H}$.

In the following, we demonstrate that $\mathbb{V}$ is a viscosity solution of the QVIs. We begin with the subsolution property.

Proposition 5.8 (Viscosity Subsolution). *The function $\mathbb{V}$ defined in (51) is a viscosity subsolution of the QVIs.*

Proof. Being a member of $\mathbb{H}$, the function $\mathbb{V}$ is upper semicontinuous; hence we have $\mathbb{V} = \mathbb{V}^*$ and $\mathcal{M}[\mathbb{V}] = \mathcal{M}[\mathbb{V}]^*$ by Lemma 5.1. Assume by contradiction that there exist $(t^*, x^*) \in [0,T) \times \mathcal{S}$ and a test function $\varphi \in \mathrm{C}^2([0,T) \times \mathcal{S})$ with $\varphi \geq \mathbb{V}$, $\varphi(t^*, x^*) = \mathbb{V}(t^*, x^*)$, and

$$\min\{\mathcal{L}[\varphi](t^*, x^*), \mathbb{V}(t^*, x^*) - \mathcal{M}[\mathbb{V}](t^*, x^*)\} = 2\kappa > 0 \tag{52}$$

for some $\kappa > 0$. We can assume without loss that the maximum of $\mathbb{V} - \varphi$ at (t^*, x^*) is global (as only the behavior of φ in a neighborhood of (t^*, x^*) is relevant) and strict (consider $\tilde{\varphi}(t,x) \triangleq \varphi(t,x) + |(t,x) - (t^*, x^*)|^4$ instead). Using $\varphi(t^*, x^*) = \mathbb{V}(t^*, x^*)$ in (52), continuity of φ and $\mathcal{L}[\varphi]$, and lower semicontinuity of $-\mathcal{M}[\mathbb{V}]$ it follows that there exists $\varepsilon > 0$ such that

$$\min\{\mathcal{L}[\varphi](t,x), \varphi(t,x) - \mathcal{M}[\mathbb{V}](t,x)\} \geq \kappa > 0, \qquad (t,x) \in \overline{\mathcal{B}}_\varepsilon(t^*, x^*), \tag{53}$$

where we set

$$\overline{\mathcal{B}}_\varepsilon(t^*, x^*) \triangleq \{(t,x) \in [0,T] \times \overline{\mathcal{S}} : |(t,x) - (t^*, x^*)| \leq \varepsilon\}, \tag{54}$$

$$\mathcal{B}_\varepsilon(t^*, x^*) \triangleq \{(t,x) \in [0,T] \times \overline{\mathcal{S}} : |(t,x) - (t^*, x^*)| < \varepsilon\}. \tag{55}$$

Upon making ε smaller if necessary, we may in addition assume that

$$\overline{\mathcal{B}}_\varepsilon(t^*, x^*) \cap ([0,T] \times \{0\}) = \emptyset = \overline{\mathcal{B}}_\varepsilon(t^*, x^*) \cap (\{T\} \times \overline{\mathcal{S}}). \tag{56}$$

Now define

$$\mathcal{D} \triangleq \overline{\mathcal{B}}_\varepsilon(t^*, x^*) \setminus \mathcal{B}_{\varepsilon/2}(t^*, x^*).$$

Since $\mathcal{D}$ is compact and the global maximum of $\mathbb{V} - \varphi \in \mathrm{USC}$ at (t^*, x^*) is strict, there exists some $\delta \in (0, \kappa)$ such that

$$\mathbb{V}(t,x) + \delta \leq \varphi(t,x), \qquad (t,x) \in \mathcal{D}. \tag{57}$$

Fixing $\eta \in (0, \delta)$, we define

$$\varphi^\eta : [0,T) \times \mathcal{S} \to \mathbb{R}, \qquad (t,x) \mapsto \varphi^\eta(t,x) \triangleq \varphi(t,x) - \eta,$$

and

$$h^\eta : [0,T] \times \overline{\mathcal{S}} \to \mathbb{R}, \quad (t,x) \mapsto h^\eta(t,x) \triangleq \begin{cases} \min\{\mathbb{V}(t,x), \varphi^\eta(t,x)\} & \text{if } (t,x) \in \overline{\mathcal{B}}_\varepsilon(t^*,x^*), \\ \mathbb{V}(t,x) & \text{otherwise.} \end{cases}$$

Since the partial derivatives of φ^η and φ coincide, it follows from (53) that

$$\mathcal{L}[\varphi^\eta](t,x) = \mathcal{L}[\varphi](t,x) \geq \kappa > 0, \qquad (t,x) \in \overline{\mathcal{B}}_\varepsilon(t^*,x^*). \tag{58}$$

Moreover, we clearly have $\varphi^\eta(t^*,x^*) = \varphi(t^*,x^*) - \eta = \mathbb{V}(t^*,x^*) - \eta < \mathbb{V}(t^*,x^*)$ and thus

$$h^\eta(t^*,x^*) = \varphi^\eta(t^*,x^*) < \mathbb{V}(t^*,x^*). \tag{59}$$

By (53), (56), (57), and (58) and a standard argument as in Belak et al., 2017, Theorem 4.1, it follows that $h^\eta \in \mathbb{H}$. But in view of (59) this is incompatible with the definition of $\mathbb{V}$ in (51), and we conclude that $\mathbb{V}$ is a viscosity subsolution of the QVIs. □

The following result characterizes the behavior of $\mathbb{V}$ on the boundary of the state space, that is, on the sets $\{T\} \times \mathcal{S}$ and $[0,T] \times \{0\}$.

Proposition 5.9 (Terminal Inequalities). *The function $\mathbb{V}$ defined in (51) satisfies*

$$\min\{\mathbb{V}(T,x) - U_{\mathrm{L}}(x), \mathbb{V}(T,x) - \mathcal{M}[\mathbb{V}](T,x)\} \leq 0, \qquad x \in \mathcal{S}. \tag{60}$$

Moreover, it holds that

$$\mathbb{V}(t,0) = 0 = U_{\mathrm{L}}(0) = \max_{n \in \mathbb{N}} \mathcal{M}^n[U_{\mathrm{L}}](0), \qquad t \in [0,T].$$

Proof. We first recall that $\Psi_0^p \in \mathbb{H}$ if $C > 1/p$ by Lemma 5.7 and hence

$$0 \leq \mathbb{V}(t,0) \leq \Psi_0^p(t,0) = 0, \qquad t \in [0,T].$$

Thus, it only remains to verify (60). We argue by contradiction and suppose that there exists $x^* \in \mathcal{S}$ with

$$\min\{\mathbb{V}(T,x^*) - U_{\mathrm{L}}(x^*), \mathbb{V}(T,x^*) - \mathcal{M}[\mathbb{V}](T,x^*)\} \triangleq \kappa > 0.$$

For each $\varepsilon, \delta > 0$, we define the sets

$$\mathcal{B}(\delta, \varepsilon) \triangleq (T - \delta, T] \times \{x \in \overline{S} : |x - x^*| < \varepsilon\}, \tag{61}$$

$$\overline{\mathcal{B}}(\delta, \varepsilon) \triangleq [T - \delta, T] \times \{x \in \overline{S} : |x - x^*| \leq \varepsilon\}, \tag{62}$$

$$\mathcal{D}(\delta, \varepsilon) \triangleq \overline{\mathcal{B}}(\delta, \varepsilon) \setminus \mathcal{B}(\delta/2, \varepsilon/2) = [T - \delta, T - \delta/2] \times \{x \in \overline{S} : \varepsilon/2 \leq |x - x^*| \leq \varepsilon\}. \tag{63}$$

Since U_{L} is continuous and $\mathcal{M}[\mathbb{V}]$ is upper semicontinuous by Lemma 5.1, we can choose $\varepsilon \in (0, \kappa)$ such that $\varepsilon < \min\{|x^*|, T\}$ and

$$\min\{\mathbb{V}(T, x^*) - U_{\mathrm{L}}(x), \mathbb{V}(T, x^*) - \mathcal{M}[\mathbb{V}](t, x)\} \geq \varepsilon, \qquad (t, x) \in \overline{\mathcal{B}}(\varepsilon, \varepsilon). \tag{64}$$

Since $\mathbb{V}$ is locally bounded, there exists $\beta > 0$ sufficiently small such that

$$\mathbb{V}(T, x^*) + \frac{\varepsilon^2}{4\beta} \geq \varepsilon + \sup_{(t,x) \in \mathcal{D}(\delta, \varepsilon)} \mathbb{V}(t, x), \qquad \delta \in (0, \varepsilon]. \tag{65}$$

With a fixed constant $L > 0$ to be specified below, we consider the function

$$\varphi : [0, T] \times \overline{S} \to \mathbb{R}, \qquad (t, x) \mapsto \varphi(t, x) \triangleq \mathbb{V}(T, x^*) + \frac{1}{\beta}|x^* - x|^2 + L(T - t).$$

Since the spatial partial derivatives of φ are independent of t and bounded on $\overline{\mathcal{B}}(\varepsilon, \varepsilon)$, and since $(\partial/\partial t)\varphi(t, x) = -L$, we can choose L sufficiently large to ensure that

$$\mathcal{L}[\varphi](t, x) \geq 0, \qquad (t, x) \in \overline{\mathcal{B}}(\varepsilon, \varepsilon). \tag{66}$$

Having fixed L in this way, we choose $\delta < \min\{\varepsilon/(2L), \varepsilon\}$. By (65) and the fact that $|x - x^*| \geq \varepsilon/2$ and $T - t \geq -\delta$ for all $(t, x) \in \mathcal{D}(\delta, \varepsilon)$, we have

$$\varphi(t, x) \geq \mathbb{V}(t, x) + \frac{\varepsilon}{2}, \qquad (t, x) \in \mathcal{D}(\delta, \varepsilon). \tag{67}$$

Moreover, since $T - t \leq -\delta \leq -\varepsilon/(2L)$, it follows from (64) that

$$\varphi(t, x) \geq \mathbb{V}(T, x^*) - L\delta \geq U_{\mathrm{L}}(x), \qquad (t, x) \in \overline{\mathcal{B}}(\delta, \varepsilon). \tag{68}$$

Fixing $\eta \in (0, \varepsilon/2)$, we define

$$\varphi^\eta : [0, T] \times \overline{S} \to \mathbb{R}, \qquad (t, x) \mapsto \varphi^\eta(t, x) \triangleq \varphi(t, x) - \eta,$$

as well as

$$h^\eta : [0,T]\times\overline{S}\to\mathbb{R},\quad (t,x)\mapsto h^\eta(t,x)\triangleq\begin{cases}\min\{\mathbb{V}(t,x),\varphi^\eta(t,x)\} & \text{if } (t,x)\in\overline{\mathcal{B}}(\delta,\varepsilon),\\ \mathbb{V}(t,x) & \text{otherwise.}\end{cases}$$

Then $\varphi^\eta(T,x^*)=\varphi(T,x^*)-\eta=\mathbb{V}(T,x^*)-\eta<\mathbb{V}(T,x^*)$ and hence

$$h^\eta(T,x^*)=\varphi^\eta(T,x^*)<\mathbb{V}(T,x^*).$$

Using (66), (67) and (68), one can check as in Belak et al., 2017, Proposition 4.2, that $h^\eta\in\mathbb{H}$, contradicting the minimality of $\mathbb{V}$. □

Finally, we establish the supersolution property of $\mathbb{V}$. This is simpler because it follows quite directly from the properties of the members in $\mathbb{H}$.

Proposition 5.10 (Viscosity Supersolutions). *Each Borel measurable function $h:[0,T]\times\overline{S}\to\mathbb{R}_+$ satisfying $(\mathbb{H}_2)$ to $(\mathbb{H}_5)$ is a viscosity supersolution of the QVIs with*

$$\min\{h_*(T,x)-U_{\mathrm{L}}(x),h_*(T,x)-\mathcal{M}[h_*]_*(T,x)\}\geq 0,\qquad x\in\overline{S}. \tag{69}$$

In particular, $\mathbb{V}$ is a viscosity subsolution of the QVIs.

Proof. By $(\mathbb{H}_4)$, h satisfies

$$h(t,x)\geq\mathcal{M}[h](t,x)\geq\mathcal{M}[h_*](t,x)\geq\mathcal{M}[h_*]_*(t,x),\qquad (t,x)\in[0,T]\times\overline{S}.$$

But then, since $\mathcal{M}[h_*]_*$ is lower semicontinuous, it must be dominated by the lower semicontinuous envelope of h, that is,

$$h_*(t,x)-\mathcal{M}[h_*]_*(t,x)\geq 0,\qquad (t,x)\in[0,T]\times\overline{S}. \tag{70}$$

Now fix $(\bar t,\bar x)\in[0,T)\times S$ and $\varphi\in C^2([0,T)\times S)$ with $\varphi\leq h_*$ and $\varphi(\bar t,\bar x)=h_*(\bar t,\bar x)$. We choose a sequence $\{(t_k,x_k)\}_{k\in\mathbb{N}}\subset[0,T)\times S$ converging to $(\bar t,\bar x)$ such that

$$h_*(\bar t,\bar x)=\lim_{k\to\infty}h(t_k,x_k).$$

Since φ is continuous and $\varphi\leq h_*\leq h$, we see that

$$0\leq\gamma_k\triangleq h(t_k,x_k)-\varphi(t_k,x_k)\to 0\quad\text{as } k\to\infty.$$

Now fix a sequence $\{\delta_k\}_{k\in\mathbb{N}}$ of strictly positive real numbers with

$$\lim_{k\to\infty}\delta_k=\lim_{k\to\infty}\frac{\gamma_k}{\delta_k}=0.$$

Moreover, let $\varepsilon > 0$ and define

$$\rho_k \triangleq \inf\{t \in [t_k, T] : |\bar{X}^k_u - x_k| \geq \varepsilon\} \wedge (t_k + \delta_k) \wedge T, \qquad k \in \mathbb{N},$$

where $\bar{X}^k \triangleq \bar{X}^{t_k, x_k}$. Using $(\mathbb{H}_5)$, the inequality $h \geq h_* \geq \varphi$, and Itō's formula yield

$$h(t_k, x_k) \geq \mathbb{E}\big[h\big(\rho_k, \bar{X}^k_{\rho_k}\big)\big] \geq \mathbb{E}\big[\varphi\big(\rho_k, \bar{X}^k_{\rho_k}\big)\big] = \varphi(t_k, x_k) - \mathbb{E}\left[\int_{t_k}^{\rho_k} \mathcal{L}[\varphi]\big(u, \bar{X}^k_u\big)\, \mathrm{d}u\right].$$

Upon rearranging and dividing by δ_k, it follows that

$$\frac{\gamma_k}{\delta_k} + \mathbb{E}\left[\frac{1}{\delta_k}\int_{t_k}^{\rho_k} \mathcal{L}[\varphi]\big(u, \bar{X}^k_u\big)\, \mathrm{d}u\right] \geq 0.$$

Now $\rho_k(\omega) = t_k + \delta_k$ for eventually all $k \in \mathbb{N}$ and $\mathbb{P}$-almost every $\omega \in \Omega$. Thus, upon sending $k \to \infty$, the mean value theorem and dominated convergence imply that

$$\mathcal{L}[\varphi](\bar{t}, \bar{x}) \geq 0.$$

In combination with (70), this means that h is a viscosity supersolution of the QVIs. We now establish $h_*(T, \cdot) \geq U_{\mathrm{L}}$ to obtain (69). For this, fix $x \in \overline{S}$ and choose a sequence $\{(t_k, x_k)\}_{k\in\mathbb{N}} \subset [0, T] \times \overline{S}$ converging to (T, x) such that

$$\lim_{k\to\infty} h(t_k, x_k) = h_*(T, x).$$

Then $(\mathbb{H}_5)$, $(\mathbb{H}_3)$, and Fatou's lemma yield

$$h_*(T, x) = \lim_{k\to\infty} h(t_k, x_k) \geq \liminf_{k\to\infty} \mathbb{E}\left[h\left(T, \bar{X}_T^{t_k, x_k}\right)\right] \geq \liminf_{k\to\infty} \mathbb{E}\left[U_{\mathrm{L}}\left(\bar{X}_T^{t_k, x_k}\right)\right] \geq U_{\mathrm{L}}(x),$$

and thus (69) is verified. This completes the proof. □

Combining the viscosity sub- and supersolution properties of $\mathbb{V}$ with the comparison principle characterizes $\mathbb{V}$ as the unique continuous viscosity solution of the QVIs.

Theorem 5.11 (Viscosity Characterization of $\mathbb{V}$). *The function $\mathbb{V}$ defined in (51) is a continuous viscosity solution of the QVIs. It is unique in the class of nonnegative functions satisfying the growth condition $(\mathbb{H}_2)$ and the boundary/terminal conditions*

$$\mathbb{V}(t, x) = \max_{n\in\mathbb{N}_0} \mathcal{M}^n[U_{\mathrm{L}}](x), \qquad (t, x) \in ([0, T] \times \{0\}) \cup (\{T\} \times \overline{S}).$$

Proof. By Propositions 5.8 and 5.10, $\mathbb{V}$ is a viscosity solution of the QVIs. Moreover, by Propositions 5.9 and 5.10, $\mathbb{V}(\cdot, 0) = 0$ and $\mathbb{V}(T, \cdot)$ satisfies the terminal inequalities (60) and (69). Applying the comparison principle for the terminal inequalities, see Corollary 5.5, therefore shows that

$\mathbb{V}(T,\cdot) = \mathbb{V}_*(T,\cdot)$ is continuous; since $\mathbb{V} \geq \mathcal{M}[\mathbb{V}]$ everywhere we even have

$$\min\{\mathbb{V}(T,x) - U_{\mathrm{L}}(x), \mathbb{V}(T,x) - \mathcal{M}[\mathbb{V}](T,x)\} = 0, \qquad x \in \overline{\mathcal{S}}.$$

But this already implies

$$\mathbb{V}(T,x) = \max_{n\in\mathbb{N}_0} \mathcal{M}^n[U_{\mathrm{L}}](x), \qquad x \in \overline{\mathcal{S}}.$$

Since $\mathbb{V}(T,\cdot)$ is continuous, we can apply the comparison principle for the QVIs in Theorem 5.4 to obtain continuity of $\mathbb{V} = \mathbb{V}^* \leq \mathbb{V}_*$ as well as uniqueness. □

5.4 | Construction of optimal strategies

In this final subsection we show that the value function $\mathcal{V} = \mathbb{V}$, and we provide an explicit construction of optimal trading strategies for the retail investor's portfolio problem. For this, we define the continuation and intervention regions defined in terms of $\mathbb{V}$ via[17]

$$\mathcal{C} \triangleq \{(t,x) \in [0,T] \times \overline{\mathcal{S}} : \mathbb{V}(t,x) > \mathcal{M}[\mathbb{V}](t,x)\}, \tag{71}$$

$$\mathcal{I} \triangleq \{(t,x) \in [0,T] \times \overline{\mathcal{S}} : \mathbb{V}(t,x) = \mathcal{M}[\mathbb{V}](t,x)\}. \tag{72}$$

Since $\mathbb{V}$ is continuous and $\mathcal{D}$ is compact-valued, a classical measurable selection argument, see Schäl (1974), yields a Borel measurable function

$$\delta : [0,T] \times (\overline{\mathcal{S}} \setminus \mathcal{S}_\emptyset) \to \mathbb{R}, \qquad (t,x) \mapsto \delta(t,x),$$

such that

$$\delta(t,x) \in \mathcal{D}(x) \quad \text{and} \quad \mathcal{M}[\mathbb{V}](t,x) = \mathbb{V}(t, \Gamma(x, \delta(t,x))), \quad (t,x) \in [0,T] \times (\overline{\mathcal{S}} \setminus \mathcal{S}_\emptyset).$$

For any fixed $(t,x) \in [0,T] \times \overline{\mathcal{S}}$, we define a candidate optimal strategy $\Lambda^* = \{(\tau_k^*, \Delta_k^*)\}_{k\in\mathbb{N}}$ as follows: Set $(\tau_0^*, \xi_0^*) \triangleq (t,x)$ and, iteratively for all $k \in \mathbb{N}$,

$$\bar{X}^k \triangleq \bar{X}^{\tau_{k-1}^*, \xi_{k-1}^*}, \quad \tau_k^* \triangleq \inf\left\{u \in [\tau_{k-1}^*, T] : \left(u, \bar{X}_u^k\right) \in \mathcal{I}\right\}, \tag{73}$$

$$\Delta_k^* \triangleq \delta\left(\tau_k^*, \bar{X}_{\tau_k^*}^k\right) \mathbb{1}_{\{\tau_k^* \leq T\}}, \quad \xi_k^* \triangleq \Gamma\left(\bar{X}_{\tau_k^*}^k, \Delta_k^*\right), \tag{74}$$

where we recall that for any $[0,T]$-valued stopping time τ and $\overline{\mathcal{S}}$-valued random variable ξ, we write $\bar{X}^{\tau,\xi} = \{\bar{X}_t^{\tau,\xi}\}_{t\in[\tau,T]}$ for the uncontrolled portfolio process with $\bar{X}_\tau^{\tau,\xi} = \xi$. From the above construction, it follows immediately that $\Lambda^* \in \mathcal{A}(t,x)$. The following verification result demonstrates rigorously that Λ^* is optimal and $\mathbb{V} = \mathcal{V}$; its proof is based on the superharmonic function technique in Belak & Christensen, 2019, Belak et al., 2017, Christensen, 2014.

Theorem 5.12 (Verification Theorem). *For every* $(t,x) \in [0,T] \times \overline{S}$ *we have*

$$\mathbb{V}(t,x) = \mathcal{V}(t,x) = \mathbb{E}\left[U_{\mathrm{L}}\left(X_T^{t,x,\Lambda^*}\right)\right]$$

where $\Lambda^* = \{(\tau_k^*, \Delta_k^*)\}_{k\in\mathbb{N}}$ *is the trading strategy defined via (73) and (74).*

Proof. We fix $(t,x) \in [0,T] \times \overline{S}$. Since Λ^* is admissible, we have $\mathbb{E}[U_{\mathrm{L}}(X_T^{t,x,\Lambda^*})] \le \mathcal{V}(t,x)$. As we have already shown that $\mathbb{V} \ge \mathcal{V}$, it suffices to demonstrate that

$$\mathbb{V}(t,x) = \mathbb{E}\left[U_{\mathrm{L}}\left(X_T^{t,x,\Lambda^*}\right)\right].$$

We set $X^* \triangleq X^{t,x,\Lambda^*}$ for ease of notation.

Step 1. For every $\lambda \in (0,1)$, we define

$$\mathcal{C}_\lambda \triangleq \{(t,x) \in [0,T] \times \overline{S} : \lambda \mathbb{V}(t,x) > \mathcal{M}[\mathbb{V}](t,x)\}, \tag{75}$$

$$\mathcal{I}_\lambda \triangleq \{(t,x) \in [0,T] \times \overline{S} : \lambda \mathbb{V}(t,x) \le \mathcal{M}[\mathbb{V}](t,x)\}. \tag{76}$$

Since $\lambda\mathbb{V} - \mathcal{M}[\mathbb{V}]$ is lower semicontinuous, $\mathcal{I}_\lambda$ is closed and hence $\mathcal{C}_\lambda$ is open. Moreover, $\mathcal{I}_\lambda$ is decreasing in λ with $\mathcal{I} = \bigcap_{\lambda\in(0,1)} \mathcal{I}_\lambda$. For each $\lambda \in (0,1)$, we construct a family of stopping times via

$$\vartheta_{\bar{t},\bar{x}}^\lambda \triangleq \inf\left\{u \in [\bar{t},T] : \left(u, \bar{X}_u^{\bar{t},\bar{x}}\right) \in \mathcal{I}_\lambda\right\} \wedge T, \qquad (\bar{t},\bar{x}) \in [0,T] \times \overline{S}.$$

With this, we define two functions

$$h : [0,T] \times \overline{S} \to \mathbb{R}_+, \qquad (\bar{t},\bar{x}) \mapsto h(\bar{t},\bar{x}) \triangleq \mathbb{E}\left[\mathbb{V}\left(\vartheta_{\bar{t},\bar{x}}^\lambda, \bar{X}_{\vartheta_{\bar{t},\bar{x}}^\lambda}^{\bar{t},\bar{x}}\right)\right],$$

and

$$h_\lambda : [0,T] \times \overline{S} \to \mathbb{R}_+, \qquad (\bar{t},\bar{x}) \mapsto h_\lambda(\bar{t},\bar{x}) \triangleq \lambda \mathbb{V}(\bar{t},\bar{x}) + (1-\lambda)h(\bar{t},\bar{x}).$$

Step 2. We show that $h_\lambda \ge \mathbb{V}$. For this, using that h_λ is clearly Borel measurable, it suffices to show that h_λ satisfies $(\mathbb{H}_2)$ to $(\mathbb{H}_5)$; indeed, in that case Proposition 5.10 implies that h_λ is a viscosity supersolution of the QVIs, so the comparison principle in Theorem 5.4 implies that $h_\lambda \ge (h_\lambda)_* \ge \mathbb{V}$. We first observe that since $\mathbb{V}$ satisfies $(\mathbb{H}_5)$, we have

$$h(\bar{t},\bar{x}) = \mathbb{E}\left[\mathbb{V}\left(\vartheta_{\bar{t},\bar{x}}^\lambda, \bar{X}_{\vartheta_{\bar{t},\bar{x}}^\lambda}^{\bar{t},\bar{x}}\right)\right] \le \mathbb{V}(\bar{t},\bar{x}), \qquad (\bar{t},\bar{x}) \in [0,T] \times \overline{S}$$

and hence $h_\lambda \le \mathbb{V}$, so h_λ satisfies the growth condition $(\mathbb{H}_2)$ because $\mathbb{V}$ does. The terminal condition $(\mathbb{H}_3)$ for h_λ holds because $h(T,\cdot) = \mathbb{V}(T,\cdot)$, whence $h_\lambda(T,\cdot) = \mathbb{V}(T,\cdot) \ge U_{\mathrm{L}}$. To establish

($\mathbb{H}_4$) for h_λ, we fix $(\bar{t},\bar{x}) \in [0,T]\times\overline{S}$. Since $h \le \mathbb{V}$, we have

$$\begin{aligned}\mathcal{M}[h_\lambda](\bar{t},\bar{x}) &\le \lambda\mathcal{M}[\mathbb{V}](\bar{t},\bar{x}) + (1-\lambda)\mathcal{M}[h](\bar{t},\bar{x})\\ &\le \lambda\mathcal{M}[\mathbb{V}](\bar{t},\bar{x}) + (1-\lambda)\mathcal{M}[\mathbb{V}](\bar{t},\bar{x}) = \mathcal{M}[\mathbb{V}](\bar{t},\bar{x}).\end{aligned} \tag{77}$$

If $(\bar{t},\bar{x}) \in \mathcal{I}_\lambda$, then $\theta^\lambda_{\bar{t},\bar{x}} = \bar{t}$. Thus $h(\bar{t},\bar{x}) = \mathbb{V}(\bar{t},\bar{x})$ and therefore also $h_\lambda(\bar{t},\bar{x}) = \mathbb{V}(\bar{t},\bar{x})$; since $\mathbb{V}$ satisfies ($\mathbb{H}_4$), it follows that

$$\mathcal{M}[h_\lambda](\bar{t},\bar{x}) \le \mathcal{M}[\mathbb{V}](\bar{t},\bar{x}) \le \mathbb{V}(\bar{t},\bar{x}) = h_\lambda(\bar{t},\bar{x}), \qquad (\bar{t},\bar{x}) \in \mathcal{I}_\lambda.$$

If, on the other hand, $(\bar{t},\bar{x}) \in \mathcal{C}_\lambda$, then $\mathcal{M}[\mathbb{V}](\bar{t},\bar{x}) < \lambda\mathbb{V}(\bar{t},\bar{x}) \le \mathbb{V}(\bar{t},\bar{x})$ and thus

$$\mathcal{M}[h_\lambda](\bar{t},\bar{x}) \le \mathcal{M}[\mathbb{V}](\bar{t},\bar{x}) < \mathbb{V}(\bar{t},\bar{x}), \qquad (\bar{t},\bar{x}) \in \mathcal{C}_\lambda.$$

In summary, we have demonstrated that h_λ satisfies ($\mathbb{H}_4$). It remains to verify ($\mathbb{H}_5$). Since $\mathbb{V}$ satisfies ($\mathbb{H}_5$), by linearity it is clearly sufficient to show that h satisfies ($\mathbb{H}_5$). But this property is inherited from $\mathbb{V}$ by pathwise uniqueness and the strong Markov property of $\bar{X}$. We therefore conclude that $h_\lambda \ge \mathbb{V}$.

Step 3. For $k \in \mathbb{N}_0$, let us set

$$(\tau,\xi) \triangleq (\tau^*_k,\xi^*_k), \qquad \bar{X} \triangleq \bar{X}^{\tau,\xi}, \quad \text{and} \quad \vartheta^\lambda \triangleq \vartheta^\lambda_{\tau,\xi}.$$

By definition of h and the strong Markov property, we have

$$h(\tau,\xi) = \mathbb{E}\left[\mathbb{V}\left(\vartheta^\lambda_{\bar{t},\bar{x}}, \bar{X}^{\bar{t},\bar{x}}_{\vartheta^\lambda_{\bar{t},\bar{x}}}\right)\right]\Big|_{(\bar{t},\bar{x})=(\tau,\xi)} = \mathbb{E}\Big[\mathbb{V}\big(\vartheta^\lambda, \bar{X}_{\vartheta^\lambda}\big)\Big|\mathfrak{F}_\tau\Big] \qquad \text{on } \{\tau \le T\}.$$

Since $h_\lambda \ge \mathbb{V}$, it follows that

$$\mathbb{V}(\tau,\xi) \le h_\lambda(\tau,\xi) = \lambda\mathbb{V}(\tau,\xi) + (1-\lambda)\mathbb{E}\Big[\mathbb{V}\big(\vartheta^\lambda, \bar{X}_{\vartheta^\lambda}\big)\Big|\mathfrak{F}_\tau\Big] \qquad \text{on } \{\tau \le T\}.$$

Upon rearranging, dividing by $(1-\lambda)$, and using property ($\mathbb{H}_5$) of $\mathbb{V}$, we obtain

$$\mathbb{V}(\tau,\xi) \le \mathbb{E}\Big[\mathbb{V}\big(\vartheta^\lambda, \bar{X}_{\vartheta^\lambda}\big)\Big|\mathfrak{F}_\tau\Big] \le \mathbb{V}(\tau,\xi) \qquad \text{on } \{\tau \le T\}. \tag{78}$$

Step 4. Since $\vartheta^\lambda \le \tau^*_{k+1} \wedge T$ and the mapping $\lambda \mapsto \vartheta^\lambda$ is increasing, it follows that $\vartheta \triangleq \lim_{\lambda\uparrow 1}\vartheta^\lambda$ exists and satisfies $\vartheta \le \tau^*_{k+1} \wedge T$. On the other hand, since $\mathbb{V}$ is continuous and satisfies ($\mathbb{H}_4$), $(\vartheta^\lambda, \bar{X}_{\vartheta^\lambda}) \in \mathcal{I}_\lambda$, and $\mathcal{M}[\mathbb{V}]$ is upper semicontinuous, we have

$$\mathcal{M}[\mathbb{V}](\vartheta,\bar{X}_\vartheta) \le \mathbb{V}(\vartheta,\bar{X}_\vartheta) = \lim_{\lambda\uparrow 1}\mathbb{V}(\vartheta^\lambda,\bar{X}_{\vartheta^\lambda}) \tag{79}$$

$$\le \limsup_{\lambda\uparrow 1}\frac{1}{\lambda}\mathcal{M}[\mathbb{V}](\vartheta^\lambda,\bar{X}_{\vartheta^\lambda}) \le \mathcal{M}[\mathbb{V}](\vartheta,\bar{X}_\vartheta) \quad \text{on } \{\tau \le T\}, \tag{80}$$

which is only possible if $\mathcal{M}[\mathbb{V}](\vartheta, \bar{X}_\vartheta) = \mathbb{V}(\vartheta, \bar{X}_\vartheta)$, that is, $\vartheta = \tau^*_{k+1}$ on $\{\tau^*_{k+1} \le T\}$. As a consequence, using (78) and the fact that $(\vartheta^\lambda, \bar{X}_{\vartheta^\lambda}) \in \mathcal{I}_\lambda$, dominated convergence and upper semicontinuity of $\mathcal{M}[\mathbb{V}]$, and finally $(\mathbb{H}_4)$ and $(\mathbb{H}_5)$, it follows that

$$\mathbb{V}(\tau, \xi) = \lim_{\lambda \uparrow 1} \mathbb{E}\Big[\mathbb{V}\big(\vartheta^\lambda, \bar{X}_{\vartheta^\lambda}\big)\Big|\mathfrak{F}_\tau\Big] \tag{81}$$

$$\le \limsup_{\lambda \uparrow 1} \frac{1}{\lambda} \mathbb{E}\Big[\mathcal{M}[\mathbb{V}]\big(\vartheta^\lambda, \bar{X}_{\vartheta^\lambda}\big)\Big|\mathfrak{F}_\tau\Big] \tag{82}$$

$$\le \mathbb{E}\Big[\mathcal{M}[\mathbb{V}]\Big(\tau^*_{k+1}, \bar{X}_{\tau^*_{k+1}}\Big)\Big|\mathfrak{F}_\tau\Big] \le \mathbb{E}\Big[\mathbb{V}\Big(\tau^*_{k+1}, \bar{X}_{\tau^*_{k+1}}\Big)\Big|\mathfrak{F}_\tau\Big] \le \mathbb{V}(\tau, \xi) \tag{83}$$

on $\{\tau^*_{k+1} \le T\}$, where in fact we have equality everywhere. Now by definition of Λ^* and using $(\tau^*_k, \xi^*_k) = (\tau, \xi)$, we obtain

$$\mathbb{V}(\tau^*_k, \xi^*_k) = \mathbb{E}\Big[\mathcal{M}[\mathbb{V}]\Big(\tau^*_{k+1}, \bar{X}_{\tau^*_{k+1}}\Big)\Big|\mathfrak{F}_{\tau^*_k}\Big] = \mathbb{E}\Big[\mathbb{V}\Big(\tau^*_{k+1}, \xi^*_{k+1}\Big)\Big|\mathfrak{F}_{\tau^*_k}\Big] \quad \text{on } \{\tau^*_{k+1} \le T\}.$$

Iteratively applying this equality, using the definition of X^*, the fact $\mathbb{P}[\lim_{k\to\infty} \tau^*_k > T] = 1$ and dominated convergence, and finally the terminal condition $\mathbb{V}(T, \cdot) = \max_{n\in\mathbb{N}} \mathcal{M}^n[U_{\mathrm{L}}]$, it follows that

$$\mathbb{V}(t, x) = \lim_{k\to\infty} \mathbb{E}\Big[\mathbb{V}(\tau^*_k, \xi^*_k)\mathbb{1}_{\{\tau^*_k \le T\}} + U_{\mathrm{L}}(X^*_T)\mathbb{1}_{\{\tau^*_k > T\}}\Big] = \mathbb{E}[U_{\mathrm{L}}(X^*_T)], \tag{84}$$

and the proof is complete. □

ACKNOWLEDGMENTS

Lukas Mich and Frank Seifried gratefully acknowledge financial support from Deutsche Forschungsgemeinschaft (DFG) within the Research Training Group 2126. Christoph Belak gratefully acknowledges financial support from Deutsche Forschungsgemeinschaft (DFG) within the International Research Training Group 2544.

Open access funding enabled and organized by Projekt DEAL.

ENDNOTES

[1] Here, we think of transaction costs as brokerage fees and do not consider implicit costs caused by frictions such as price impact as, for example, in Garleanu and Pedersen (2016). Costs that are proportional to the investor's wealth have also been studied in the literature, see, for example, Morton and Pliska (1995).

[2] Note, however, that since the cost in (1) is bounded above, one expects this to lead to a degenerate solution for a growth rate criterion. In particular, it seems to be difficult to verify the assumptions of the verification theorem in Bielecki and Pliska (2000) for the costs in (1).

[3] Borrowing from their cash accounts to attain leverage on their stock positions is difficult for small retail investors; while it is possible for retail investors that are able to pledge sufficient additional assets, even in that case the borrowing rate is typically significantly higher than the rate earned on cash deposits.

[4] Note that $\Delta = 0$ is allowed, that is, a degenerate transaction of size zero leads to a strictly positive cost. Such trades are admissible in the real world, and it is convenient to include them mathematically to ensure compactness of the set of feasible transactions.

[5] Since this cost function is discontinuous, we consider a suitable continuous approximation instead.

[6] The definition of $\mathcal{M}[\varphi](t,x)$ in the case $\mathcal{D}(x)=\emptyset$ is mainly a technical convention. It is chosen to guarantee that $\varphi(t,x) > \mathcal{M}[\varphi](t,x)$ on $S_\emptyset$ and that $\mathcal{M}$ preserves upper semicontinuity; see Lemma 5.1 .

[7] In the proofs in Section 5, it is mathematically more convenient to use a slightly different line of argument: We first construct a viscosity solution $\mathbb{V}$ of the QVIs (Theorem 5.11) and define the candidate optimal strategy in terms of $\mathbb{V}$; then we establish a verification theorem (Theorem 5.12) and apply it to show simultaneously that (i) $\mathbb{V}=\mathcal{V}$, that is, $\mathbb{V}$ coincides with the value function; and (ii) the candidate strategy is optimal. The conclusions stated in Main Results 1 and 2 are, of course, the same.

[8] The definition is with a slight abuse of notation if multiple trades occur simultaneously. See Section 5 for the rigorous definition.

[9] In Figure 4, we choose a larger interest rate r of 8.4%, a smaller volatility σ of 20% and keep the other parameters as in Table 1. The reasons for this change are purely cosmetic: If the interest rate is too small, the cash position (which behaves like $t \mapsto x_0 e^{rt}$ in the absence of trading) barely moves, making it hard to visualize the sample path appropriately. The volatility, on the other hand, is decreased to 20% to keep the Merton ratio π^* at 50%.

[10] Minor deviations of target portfolios from the Merton line in the top right corner are due to errors propagating from an artificial boundary condition at $x_0 + x_1 = 600{,}000$, which is required by the numerical scheme.

[11] Again, deviations of target portfolios from the Merton line in the top right corner are due to the influence of an artificial boundary condition.

[12] Numerically, it is difficult to distinguish trades which are just below and above the threshold C_{min}/c separating floored cost trades and proportional cost trades. We observe that the lower left part of the target portfolio wedge is light red (hence corresponding to floored cost trades), whereas the remainder of the wedge is mostly made up of medium red portfolios (corresponding to proportional cost trades). Since the lines making up the target portfolio wedge are parallel to the boundary of the trading region, all trades occur at approximately the same cost. Thus, all trade sizes in this regime are approximately equal to C_{min}/c at a cost of approximately C_{min}.

[13] By contrast, in Belak and Christensen (2019) the approach is to first characterize the value function as the unique viscosity solution of the Bellman equation and show that it coincides with the smallest stochastic supersolution; then define a candidate optimal strategy in terms of the value function, and finally verify its optimality.

[14] Here, $\mathbb{S}^2 \subset \mathbb{R}^{2\times 2}$ denotes the set of symmetric 2×2 matrices.

[15] Here, I denotes the 2×2 identity matrix. Note that, formally, the version of Ishii's lemma applied here yields matrices $\tilde{M}_k, \tilde{N}_k \in \mathbb{S}^3$ and we obtain M_k, N_k by removing the first column and row. This is justified since there are no second-order derivatives with respect to the time variable in the QVIs.

[16] Here, $\overline{\mathcal{J}}^{2,+} u_\eta(t_k,x_k)$ and $\overline{\mathcal{J}}^{2,-} v_\eta(\hat{t}_k,\hat{x}_k)$ denote the closures of the second-order super- and subjects of u_η and v_η at (t_k,x_k) and $(\hat{t}_k,\hat{x}_k)$, respectively.

[17] Note that we use $\mathbb{V}$ as defined in (51), *not* the value function $\mathcal{V}$. Our Verification Theorem 5.12 below shows that, in fact, $\mathcal{V}=\mathbb{V}$.

REFERENCES

Altarovici, A., Muhle-Karbe, J., & Soner, H. M. (2015). Asymptotics for fixed transaction costs. *Finance and Stochastics*, *19*(2), 363–414.

Altarovici, A., Reppen, M., & Soner, H. M. (2017). Optimal consumption and investment with fixed and proportional transaction costs. *SIAM Journal on Control and Optimization*, *55*(3), 1673–1710.

Azimzadeh, P., & Forsyth, P. A. (2016). Weakly chained matrices, policy iteration, and impulse control. *SIAM Journal on Numerical Analysis*, *54*(3), 1341–1364.

Bayraktar, E., & Sîrbu, M. (2012). Stochastic Perron's method and verification without smoothness using viscosity comparison: The linear case. *Proceedings of the American Mathematical Society*, *140*(10), 3645–3654.

Bayraktar, E., & Sîrbu, M. (2013). Stochastic Perron's method for Hamilton-Jacobi-Bellman equations. *SIAM Journal on Control and Optimization*, *51*(6), 4274–4294.

Bayraktar, E., & Sîrbu, M. (2014). Stochastic Perron's method and verification without smoothness using viscosity comparison: Obstacle problems and Dynkin games. *Proceedings of the American Mathematical Society*, *142*(4), 1399–1412.

Belak, C., & Christensen, S. (2019). Utility maximisation in a factor model with constant and proportional transaction costs. *Finance and Stochastics*, *23*(1), 29–96.

Belak, C., Christensen, S., & Seifried, F. T. (2017). A general verification result for stochastic impulse control problems. *SIAM Journal on Control and Optimization*, *55*(2), 627–649.

Bielecki, T. R., & Pliska, S. R. (2000). Risk sensitive asset management with transaction costs. *Finance and Stochastics*, *4*(1), 1–33.

Christensen, S. (2014). On the solution of general impulse control problems using superharmonic functions. *Stochastic Processes and their Applications*, *124*(1), 709–729.

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

Czichowsky, C., & Schachermayer, W. (2017). Portfolio optimisation beyond semimartingales: Shadow prices and fractional Brownian motion. *The Annals of Applied Probability*, *27*(3), 1414–1451.

Davis, M. H. A., & Norman, A. R. (1990). Portfolio selection with transaction costs. *Mathematics of Operations Research*, *15*(4), 676–713.

Eastham, J. F., & Hastings, K. J. (1988). Optimal impulse control of portfolios. *Mathematics of Operations Research*, *13*(4), 588–605.

Feodoria, M.-R. (2016). *Optimal investment and utility indifference pricing in the presence of small fixed transaction costs*. PhD thesis, Christian-Albrechts-Universität Kiel.

Garleanu, N., & Pedersen, L. H. (2016). Dynamic portfolio choice with frictions. *Journal of Economic Theory*, *165*, 487–516.

Ishii, K. (1993). Viscosity solutions of nonlinear second order elliptic PDEs associated with impulse control problems. *Funkcialaj Ekvacioj*, *36*(1), 123–141.

Kallsen, J., & Muhle-Karbe, J. (2010). On using shadow prices in portfolio optimization with transaction costs. *The Annals of Applied Probability*, *20*(4), 1341–1358.

Korn, R. (1998). Portfolio optimisation with strictly positive transaction costs and impulse control. *Finance and Stochastics*, *2*(2), 85–114.

Liu, H. (2004). Optimal consumption and investment with transaction costs and multiple risky assets. *The Journal of Finance*, *59*(1), 289–338.

Morton, A. J., & Pliska, S. R. (1995). Optimal portfolio management with fixed transaction costs. *Mathematical Finance*, *5*(4), 337–356.

Øksendal, B., & Sulem, A. (2002). Optimal consumption and portfolio with both fixed and proportional transaction costs. *SIAM Journal on Control and Optimization*, *40*(6), 1765–1790.

Palczewski, J., & Stettner, L. (2007). Impulsive control of portfolios. *Applied Mathematics and Optimization*, *56*(1), 67–103.

Schäl, M. (1974). A selection theorem for optimization problems. *Archiv der Mathematik (Basel)*, *25*, 219–224.

Shreve, S. E., & Soner, H. M. (1994). Optimal investment and consumption with transaction costs. *The Annals of Applied Probability*, *4*(3), 609–692.

How to cite this article: Belak C, Mich L, Seifried FT. Optimal investment for retail investors. *Mathematical Finance*. 2022;*32*:555–594. https://doi.org/10.1111/mafi.12336