Research Paper Series

No. 140

Optimal Timing for Short Covering of an Illiquid Security

Tsz-Kin Chung[†] Keiichi Tanaka[‡]

May, 2014

[†]Graduate School of Social Sciences, Tokyo Metropolitan University

[‡]Graduate School of Social Sciences, Tokyo Metropolitan University

Optimal Timing for Short Covering of an Illiquid Security

Tsz-Kin Chung * Keiichi Tanaka †
Graduate School of Social Sciences, Tokyo Metropolitan University

May, 2014

Abstract We formulate a short-selling strategy of a stock and seek the optimal timing of short covering in the presence of a random recall and a loan fee rate in an illiquid stock loan market. The aim is to study how the optimal trading strategy of the short-seller is influenced by the relevant features of the stock loan market. We characterize the optimal timing of short covering depending on the conditions that lead to different costs and benefits of keeping the position. Depending on the parameters, not only a put-type problem but also a call-type problem emerges. The solution to the optimal stopping problem is obtained in a closed form. We present explicitly what actions the investor should take. A comparative analysis is conducted with numerical examples.

Keywords: Finance, short-selling, stock loan, recall risk, optimal stopping

1. Introduction

Short-selling is the selling of a financial security that the investor does not own. The trading provides an efficient means for investors to exploit the opportunity or hedge against downside risk when they anticipate the overpricing of a security and speculate its future decline in value. A typical situation involving short-selling transactions can be described as follows. Some institutional investors are long biased and hence only rebalance their portfolios on a quarterly or yearly basis. These institutional investors are usually mutual funds, pension funds or tracker funds. They place their stocks with the broker who acts as a custodian. The broker who holds the inventory of stocks has the discretion to lend out the stock in order to earn a loan fee income. On the other side is the short-seller (e.g., a hedge fund manager) who implements a short-selling position by borrowing the stock ("stock loan") from such a broker and selling it in the market. After that, the short-seller's objective is to "buy low, sell high" such that the stock is first sold high and purchased later at a lower price. The buying back and returning of the stock to the broker is called *short covering.* At the end, the short-seller makes a profit from a price decline or a loss from a price rise of the stock.

There are a number of real-world complications involved in the implementation of a short-selling strategy. A unique feature in a stock loan contract is that there is no guaranteed maturity and it is effectively rolled over on a daily basis as documented in D'Avolio [3]. Hence, the broker holds an option to recall the stock borrowing at any time. At a recall, the short-seller is then forced to cover the short position immediately, regardless of a profit or loss, if replacing stocks are not found and placed with the broker. The risk of such an involuntary termination of a short-selling strategy is called the recall risk. Besides the capital profit/loss associated with the recall risk, the short-seller also has to take into account the running cost of the strategy. The broker charges the short-seller a loan fee, which is

* E-mail: btkchung@gmail.com
† E-mail: tanaka-keiichi@tmu.ac.jp

calculated as the loan fee rate times the stock price times the length of the period. As noted in D'Avolio [3], the loan fee rate varies dramatically across different categories of stocks from 50 to 800 basis points. At the same time, the short-seller deposits the sale proceeds into a margin account, which generates interest income called the short interest rebate. When the interest rate is high, the income may become one of the return drivers of a short-selling strategy. Hence, the investor's decision depends on the balance between the benefit (interest income) and the cost (loan fee) of holding the short position.

The optimal trading rule of a long position has been formulated as an optimal stopping problem in [8] and extended in [7, 10], in which the investor initially holds a security and seeks the optimal timing to sell it in order to maximize the expected discounted payoff. In this paper, we formulate a short-selling strategy as an optimal stopping problem and seek the optimal timing of the short covering in the presence of a recall risk, loan fee and interest income. The feature of the random recall gives rise to an optimal stopping problem with a random time horizon. One of the interesting results is that, depending on the levels of loan fee and interest rates, the optimal stopping problem is either of a put-type problem with a down-and-out stopping rule or a call-type problem with an up-and-out stopping rule. We derive the corresponding Hamilton–Jacobi–Bellman equation governing the value function in the face of a recall risk, and find that the recall risk has a significant impact on the value function and the corresponding optimal threshold. The value function may become negative because of the possibility of a forced termination, and the short-seller is likely to stop earlier at the closer optimal threshold to the entry price as a result of the random recall (the put-type problem) or the relatively expensive net running cost of keeping the position (the call-type problem). Given the closed-form solution, we characterize explicitly the investor's active region in terms of the loan fee rate and interest rate in several ways. We show that the active region depends sensibly on the stock price volatility, expected return and recall intensity.

The paper is organized as follows. Section 2 presents the formulation of the model, the solution and some analysis on the active condition. Section 3 provides a comparative analysis with numerical examples. Section 4 concludes the paper.

2. Model

2.1. Setup

We fix the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and assume a stock price to be a one-dimensional diffusion process $X = (X_t)_{t\geq 0}$ satisfying

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \qquad X_0 = x > 0.$$

Here, $(W_t)_{t\geq 0}$ is a standard Brownian motion, μ is the expected return of the stock and σ is the volatility of the stock. The infinitesimal generator of the stock price process X is given by

$$\mathcal{L}_X = \frac{1}{2}\sigma^2 x^2 \frac{d^2}{dx^2} + \mu x \frac{d}{dx}.$$

To maintain analytical tractability, we assume that the random time of the broker's recall is an exponential random variable[1] independent of the stock price process. An exponential random variable $\tau_R \sim Exp\,(\lambda)$ with $\lambda \geq 0$ is also equipped, and is assumed to be independent of the Brownian motion $(W_t)_{t\geq 0}$. We write $\mathcal{F}^W_t = \sigma(W_s; s \in [0,t])$, $\mathbb{F}^W = \left(\mathcal{F}^W_t\right)_{t\geq 0}$ and

[1] The exponential variable is a popular choice for the modeling of random arrival times in finance and economics (see [2, 9]).

$\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$, and assume that $\mathcal{F}_t^W \vee \sigma(1_{\{\tau_R > t\}}) \subset \mathcal{F}_t$. We denote the expectation $\mathbb{E}_x[\cdot] = \mathbb{E}[\cdot | \mathcal{F}_0, X_0 = x, \tau_R > 0]$ under $\mathbb{P}$.

There is a stock loan (or securities lending) market to borrow/lend the stock against the loan fee, although the liquidity may be limited[2] in the sense that the loan contract is available only with a specific broker because of the illiquidity of the stock loan market compared with the volume of the transactions by the investor, and thus stock availability is not guaranteed. The contract may be automatically renewed instantaneously, although there is a chance that the broker will not be able to find stock to replace. Hence, we assume that the lender does not renew the contract with probability λdt over the next instant period dt (random recall). Once the loan contract is terminated, the short-seller has to cover the short position by buying stock at the market price.

The loan fee is charged instantaneously based on the current stock price, i.e., the borrower makes the loan fee payment $\delta X_t dt$ over a small time interval dt, where δ is the constant loan fee rate. The short-seller deposits the initial proceeds K from selling the stock into a margin account that pays interest continuously at a constant rate q. As a result, the net cash outflow is given by $(\delta X_t - qK)\, dt$ over a time interval dt, which can be positive or negative depending on the levels of the stock price, the loan fee rate and the interest rate. The net cash flow $\delta X_t - qK$ is sometimes referred to as the effective loan fee and can be interpreted as the net running cost of the short-selling strategy.

The short-seller is supposed to have already undertaken the short position when the stock price was equal to K and to hold the position until she buys back at the market price either at her own discretion or following a recall by the broker. She seeks the optimal timing of short covering at her own discretion. The short-seller's problem is to optimize the expected net profit discounted at her own discount rate β

$$v(x) = \sup_{\tau_E \in \mathcal{A}} \mathbb{E}_x \left[e^{-\beta(\tau_E \wedge \tau_R)} \left(K - X_{\tau_E \wedge \tau_R}\right) - \int_0^{\tau_E \wedge \tau_R} e^{-\beta s} \left(\delta X_s - qK\right) ds \right], \tag{1}$$

where $\mathcal{A}$ is the set of all $\mathbb{F}^W$-stopping times taking values in $[0, \infty]$. We assume that the subjective discount rate β is sufficiently high, compared with the growth rate of the stock price: $\beta > \max(\mu, 0)$.

The following lemma is a standard result for a stopping time and the geometric Brownian motion X_t (e.g., see Egami and Dayanik [5] for the proof).

Lemma 2.1 *Suppose that $\beta > \max(\mu, 0)$. Then for any stopping time τ, it holds that*

$$\mathbb{E}_x \left[\int_0^\tau e^{-\beta t} qK dt \right] = \frac{qK}{\beta} - \frac{qK}{\beta} \mathbb{E}_x\left[e^{-\beta\tau}\right], \quad \mathbb{E}_x \left[\int_0^\tau e^{-\beta t} \delta X_t dt \right] = \frac{\delta}{\beta - \mu} x - \frac{\delta}{\beta - \mu} \mathbb{E}_x\left[e^{-\beta\tau} X_\tau\right].$$

By Lemma 2.1, the short-seller's problem is rewritten as

$$v(x) = \sup_{\tau_E \in \mathcal{A}} \mathbb{E}_x \left[e^{-\beta(\tau_E \wedge \tau_R)} \left(\eta K - \rho X_{\tau_E \wedge \tau_R}\right)\right] + (1-\eta)K - (1-\rho)x, \tag{2}$$

with

$$\eta = 1 - \frac{q}{\beta}, \qquad \rho = 1 - \frac{\delta}{\beta - \mu}. \tag{3}$$

[2] A repo contract is similar to a stock loan contract. They differ in terms of cash flow. In a stock loan contract only the interest equivalent cash flow (loan fee) is paid to the lender without any payment of the notional amount.

Then the original problem is reduced to the auxiliary optimal stopping problem

$$u(x) = \sup_{\tau_E \in \mathcal{A}} \mathbb{E}_x \left[e^{-\beta(\tau_E \wedge \tau_R)} g\left(X_{\tau_E \wedge \tau_R} \right) \right], \tag{4}$$

where g is the gain function given by

$$g(x) = \eta K - \rho x.$$

The second term $(1-\eta)K - (1-\rho)x$ on the right-hand side of (2) represents the expected net cost of the loan fee and the interest in the case that the short position is maintained forever without recall. Therefore, $u(x)$ represents the expected revenue from the short cover $K - X_{\tau_E \wedge \tau_R}$ and the cancellation value of the ongoing stock loan that was assumed to continue forever. The coefficients ρ and η in (3) play important roles in our analysis. Apparently, the investor's optimal strategy depends on the magnitudes of the loan fee rate and the interest rate via the signs of ρ and η. A patient investor's η tends to be negative, while an impatient one's η tends to be positive. $\eta \gtreqless 0$ is equivalent to $q \lesseqgtr \beta$ and $\rho \gtreqless 0$ is equivalent to $\delta \lesseqgtr \beta - \mu$. A sufficiently low loan fee rate implies that the gain function g is a decreasing function of x, while in an expensive loan fee environment it is an increasing function. Therefore, we call the put-type problem for the first case ($\rho > 0$) and the call-type problem for the latter case ($\rho < 0$) .

2.2. Value function and optimal thresholds

The introduction of the recall of the loan contract as the result of unavailability of replacement stock creates another nonzero lower bound besides the gain function to the optimal stopping problem. There are two lower bounds of $u(x)$ for each x by definition. One is $g(x)$ corresponding to $\tau_E = 0$. The other one is the function $\phi(x)$ corresponding to $\tau_E = \infty$, given by

$$\phi(x) = \mathbb{E}_x \left[e^{-\beta \tau_R} g\left(X_{\tau_R} \right) \right] = \lambda \left[\frac{1}{\beta + \lambda} \eta K - \frac{1}{\beta + \lambda - \mu} \rho x \right]. \tag{5}$$

Note that $\phi(x) = 0$ if $\lambda = 0$ (no chance of recall), and that $\phi(x)$ takes negative values for some x with some parameter sets when $\lambda > 0$.

Depending on the signs of η and ρ, there are four cases regarding the properties of g and ϕ as shown in Figure 1.

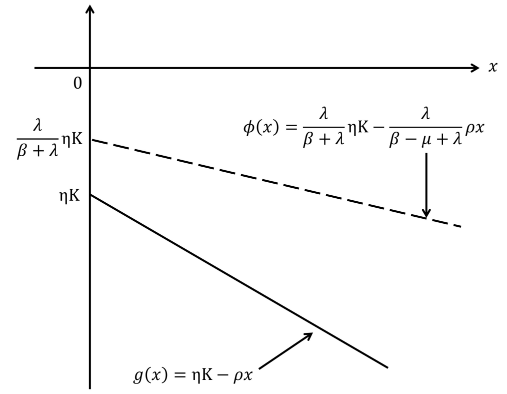
Case (a)	When $\rho \geq 0$, $\eta \leq 0$,	we see $g'(x) \leq 0$, $\phi'(x) \leq 0$, $g(x) \leq \phi(x)$ on $\mathbb{R}_+$
Case (b)	When $\rho > 0$, $\eta > 0$,	we see $g'(x) < 0$, $\phi'(x) < 0$ on $\mathbb{R}_+$, $g(x) \geq \phi(x)$ on $(0, aK]$, $g(x) < \phi(x)$ on (aK, ∞)
Case (c)	When $\rho \leq 0$, $\eta \geq 0$,	we see $g'(x) \geq 0$, $\phi'(x) \geq 0$, $g(x) \geq \phi(x)$ on $\mathbb{R}_+$
Case (d)	When $\rho < 0$, $\eta < 0$,	we see $g'(x) > 0$, $\phi'(x) > 0$ on $\mathbb{R}_+$, $g(x) < \phi(x)$ on $(0, aK)$, $g(x) \geq \phi(x)$ on $[aK, \infty)$,

where a is given by

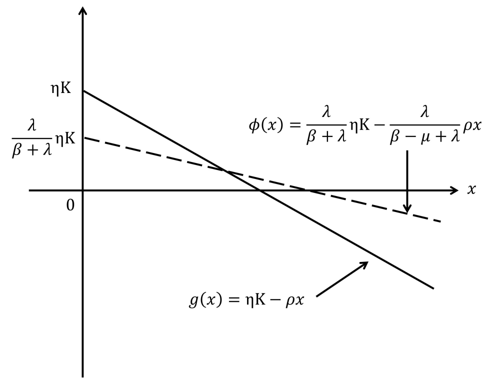
$$a = \frac{\beta}{\beta + \lambda} \frac{\beta + \lambda - \mu}{\beta - \mu} \frac{\eta}{\rho} \tag{6}$$

and satisfies $g(aK) = \phi(aK)$ because

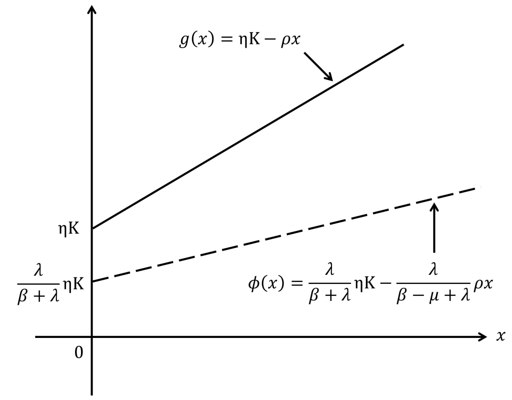
$$g(x) - \phi(x) = \frac{\beta}{\beta + \lambda} \eta K - \frac{\beta - \mu}{\beta + \lambda - \mu} \rho x. \tag{7}$$



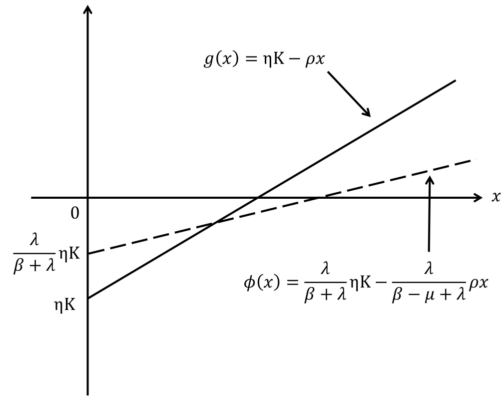
(a) When $\rho \geq 0$ and $\eta \leq 0$, the optimal strategy is to wait and see and $u(x) = \phi(x)$.



(b) When $\rho > 0$ and $\eta > 0$, the optimal strategy is determined by a put-type problem.



(c) When $\rho \leq 0$ and $\eta \geq 0$, the optimal strategy is to stop immediately and $u(x) = g(x)$.



(d) When $\rho < 0$ and $\eta < 0$, the optimal strategy is determined by a call-type problem.

Figure 1: The lower bounds g and ϕ in different cases.

This adjusted strike price aK plays a similar role to the strike price K in standard real option problems. Since it must hold that $u(x) \geq \max(g(x), \phi(x))$, one can easily conjecture that

Case (a)	$u(x) = \phi(x)$	with $\tau_E^* = \infty$
Case (b)	u is obtained as a put-type problem solution	with $\tau_E^* = \inf\{t > 0, X_t \leq b_p\}$
Case (c)	$u(x) = g(x)$	with $\tau_E^* = 0$
Case (d)	u is obtained as a call-type problem solution	with $\tau_E^* = \inf\{t > 0, X_t \geq b_c\}$

with some thresholds b_p, b_c.

The conjecture is confirmed affirmatively in the next proposition. To solve the optimal stopping problems in the presence of an exogenous Poisson signaling process, we follow a heuristic approach to derive the Hamilton–Jacobi–Bellman (HJB) equation of (4) used by Guo and Liu [6].[3] In the continuation region, we have two possible scenarios over the small time interval dt : (i) with probability λdt, the broker's recall is exercised and the short-seller is forced to cover the short position and receives $g(x) = \eta K - \rho x$; (ii) with probability $1 - \lambda dt$, the broker does not recall and the short-seller continues the strategy. The dynamic

[3]See [1, 4] for similar problems.

programming principle suggests that it holds that

$$u(x) \geq \lambda dtg(x) + (1 - \lambda dt)\, \mathbb{E}_x \left[e^{-\beta dt} u\left(X_{dt}\right) \right].$$

With application of Ito's formula to the second term, we have

$$\begin{aligned} u(x) &\geq \lambda dtg(x) + (1 - \lambda dt) \left[u(x) + \left(\mathcal{L}_X - \beta\right) u(x) dt \right] \\ &= u(x) + \left[\left(\mathcal{L}_X - \beta\right) u(x) + \lambda \left(g\left(x\right) - u(x)\right) \right] dt. \end{aligned}$$

As a result, it must hold that

$$\begin{aligned} &\widehat{\mathcal{L}}u(x) \leq 0 \quad \forall x \in \mathbb{R}_+, \\ &\widehat{\mathcal{L}}u(x) = 0 \quad \text{on the continuation region C,} \end{aligned}$$

where

$$\widehat{\mathcal{L}}f(x) = \frac{1}{2}\sigma^2 x^2 \frac{d^2 f(x)}{dx^2} + \mu x \frac{df(x)}{dx} - \beta f(x) + \lambda \left(g(x) - f(x)\right). \tag{8}$$

Hence, the auxiliary value function u must be constructed so that it satisfies the above two conditions and $u(x) \geq g(x)$ for all $x \in \mathbb{R}_+$. In the construction we denote by $m < 0$ and $n > 1$ the solutions of

$$\frac{1}{2}\sigma^2 x(x-1) + \mu x - (\beta + \lambda) = 0,$$

which are associated with the homogeneous part of the infinitesimal operator $\widehat{\mathcal{L}}$.

Proposition 2.1 *The auxiliary optimal stopping problem (4) has the solution:*[4]

Case (a) *When* $\rho \geq 0, \eta \leq 0, \quad u = \phi, \quad \tau_E^* = \infty,$

Case (b) *When* $\rho > 0, \eta > 0, \quad u = u_p, \quad \tau_E^* = \inf\left\{t \geq 0; \quad X_t \leq b_p\right\}, \quad b_p = \dfrac{m}{m-1} aK,$

Case (c) *When* $\rho \leq 0, \eta \geq 0, \quad u = g, \quad \tau_E^* = 0,$

Case (d) *When* $\rho < 0, \eta < 0, \quad u = u_c, \quad \tau_E^* = \inf\left\{t \geq 0; \quad X_t \geq b_c\right\}, \quad b_c = \dfrac{n}{n-1} aK,$

and

$$u_p(x) \triangleq \begin{cases} \left(g(b_p) - \phi(b_p)\right) \left(\dfrac{x}{b_p}\right)^m + \phi(x), & x \in \left(b_p, \infty\right), \\ g(x), & x \in \left(0, b_p\right], \end{cases} \tag{9}$$

$$u_c(x) \triangleq \begin{cases} g(x), & x \in \left[b_c, \infty\right), \\ \left(g(b_c) - \phi(b_c)\right) \left(\dfrac{x}{b_c}\right)^n + \phi(x), & x \in \left(0, b_c\right). \end{cases} \tag{10}$$

Proof. See Appendix. ■

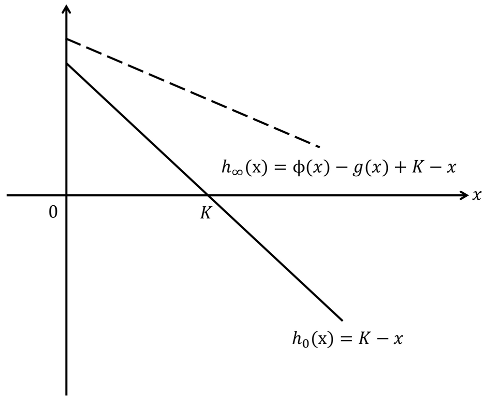
Then we are ready to obtain the value function.

Theorem 2.1 *The value function v of the optimal stopping problem (1) is given by*

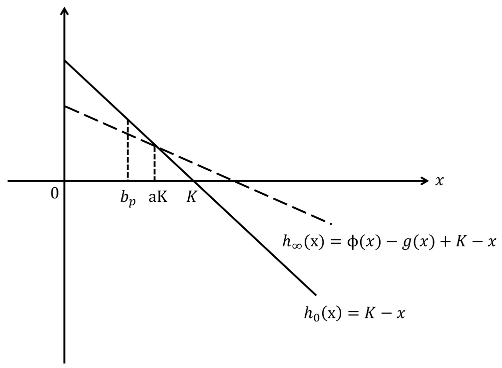
$$v(x) = u(x) - g(x) + K - x, \quad x \in \mathbb{R}_+.$$

From Proposition 2.1 and Theorem 2.1, $v(x)$ has the two lower bounds: $h_0(x) = K - x$ corresponding to $\tau_E^* = 0$ (immediate exercise) and $h_\infty(x) = \phi(x) - g(x) + K - x$ corresponding to $\tau_E^* = \infty$ (wait and see). Figure 2 illustrates what is happening in each case.

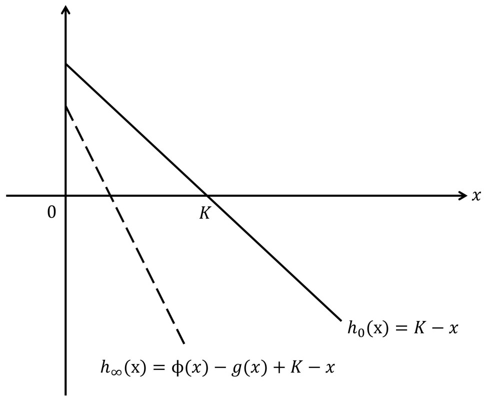
[4]When $\rho = \eta = 0$, we have $\phi = g = 0$ and it makes no difference to stop at $\tau_E^* = 0$ and $\tau_E^* = \infty$.



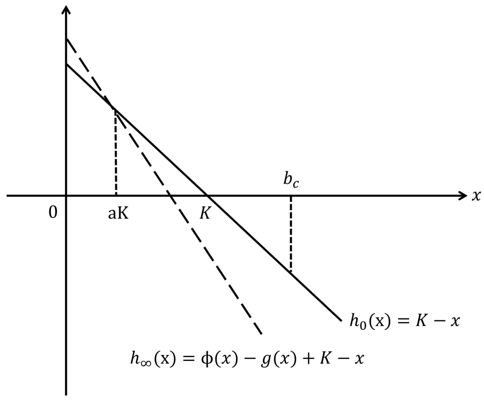
(a) When $\rho \geq 0$ and $\eta \leq 0$, the optimal strategy is $\tau_E^* = \infty$.



(b) When $\rho > 0$ and $\eta > 0$, the optimal strategy is $\tau_E^* = \inf\{t \geq 0; X_t \leq b_p\}$.



(c) When $\rho \leq 0$ and $\eta \geq 0$, the optimal strategy is $\tau_E^* = 0$.



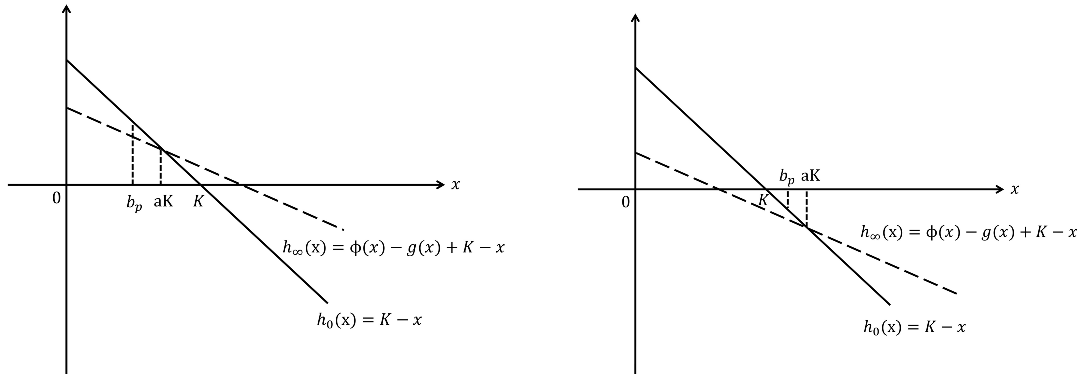
(d) When $\rho < 0$ and $\eta < 0$, the optimal strategy is $\tau_E^* = \inf\{t \geq 0; X_t \geq b_c\}$.

Figure 2: The lower bounds h_0 and h_∞ of the value function v.

- Case (a): $h_\infty(x)$ is always greater than $h_0(x)$ thanks to the cheap loan fee and the high interest income. It is optimal to wait and see until the recall time ($\tau_E^* = \infty$).
- Case (b): The net running cost of the short position is relatively cheap. It is better to stop than wait when x is sufficiently small enough to satisfy at least $h_\infty(x) < h_0(x)$. The optimal strategy is to cover the short position at the threshold b_p, i.e., $\tau_E^* = \inf\{t \geq 0; X_t \leq b_p\}$.
- Case (c): $h_0(x)$ is always greater than $h_\infty(x)$ because of the expensive loan fee and the low interest income. It is optimal to stop immediately and hence no entry occurs ($\tau_E^* = 0$).
- Case (d): The net running cost of the short position is relatively expensive. It is better to stop than wait when x is sufficiently large enough to satisfy at least $h_\infty(x) < h_0(x)$. The optimal strategy is to cover the short position at the threshold b_c, i.e., $\tau_E^* = \inf\{t \geq 0; X_t \geq b_c\}$.

In the put-type problem, the short covering at the threshold b_p involves taking profit while the short covering at b_c in the call-type problem involves a loss cut in the capital[5]. In either case, net interest income until the short covering depends on the historical path. The mandatory short covering upon recall yields another opportunity of profit taking or loss

[5]This is a voluntary loss cut in order to avoid the high running cost as the stock price increases.



(a) When $\rho > 0$, $\eta > 0$ and $a \leq 1$, then $b_p < aK \leq K$ always holds.

(b) When $\rho > 0$, $\eta > 0$ and $a > 1$. If $a < (m-1)/m$ then $b_p < K$ while $K \leq b_p$ in the case of $a \geq (m-1)/m$.

Figure 3: The active condition and optimal threshold for the put-type problem.

cut. Compared with a case without a random recall risk, the short-seller is likely to stop earlier at an optimal threshold closer to the entry price because of the random recall (an early profit taking on the put-type problem) or the relatively expensive net running cost of keeping the position (an early loss cut on the call-type problem).

Note that in Cases (b) and (d), as illustrated in Figure 2, it holds that

$$\begin{aligned} &\text{Case (b)} \qquad b_p < aK, \ \ g(b_p) - \phi(b_p) = \frac{1}{1-m}\frac{\beta}{\beta+\lambda}\eta K > 0, \\ &\text{Case (d)} \qquad aK < b_c, \ \ g(b_c) - \phi(b_c) = \frac{1}{1-n}\frac{\beta}{\beta+\lambda}\eta K > 0, \end{aligned}$$

because $m < 0$ and $n > 1$. It can be seen that

$$\begin{aligned} &\text{Case (b)} \qquad u_p(x) - \phi(x) = (g(b_p) - \phi(b_p)) \left(\frac{x}{b_p}\right)^m > 0, \quad x \in (b_p, \infty), \\ &\text{Case (d)} \qquad u_c(x) - \phi(x) = (g(b_c) - \phi(b_c)) \left(\frac{x}{b_c}\right)^n > 0, \quad x \in (0, b_c), \end{aligned}$$

represent the value of the short-seller's optionality to stop at a finite time. The quantities $(x/b_p)^m$ and $(x/b_c)^n$ are the expected present values of one dollar when the stock price hits the thresholds b_p and b_c, respectively, before the random recall. The corresponding dollar amounts at the exercises are $g(b_p) - \phi(b_p)$ and $g(b_c) - \phi(b_c)$, respectively.

2.3. The active region

We need to consider the condition that the short-seller actually enters into a short position. As Case (a) and Case (c) are trivial, let us restrict our attention to the put-type problem (Case (b)) and the call-type problem (Case (d)). In order that the investor is active in the sense of waiting for the stock price's passage over the thresholds obtained in the previous subsection until the random recall, it is clear that $K > b_p$ on the put-type problem and $K < b_c$ on the call-type problem must hold. They are rewritten in terms of a, m and n in the next proposition.

Proposition 2.2 *For the put-type problem ($\rho > 0$, $\eta > 0$), the active condition $K > b_p$ is equivalent to the condition $a < \dfrac{m-1}{m}$. For the call-type problem ($\rho < 0$, $\eta < 0$), the active* *condition $K < b_c$ is equivalent to the condition $a > \dfrac{n-1}{n}$. If ($\rho > 0$, $\eta > 0$, $a \geq \dfrac{m-1}{m}$) or ($\rho < 0$, $\eta < 0$, $a \leq \dfrac{n-1}{n}$), the investor does not enter into the short position.*

Figure 3 illustrates the active condition for the put-type problem. When $a < 1$, we see $b_p < aK < K$ and the active condition $K > b_p$ is always satisfied. When $a > 1$, however, the active condition $K > b_p$ is satisfied only when $a < (m-1)/m$.

Clearly, the active condition depends on the coefficients η, ρ and other model parameters. The active region on the (η, ρ)-plane can be characterized explicitly as

$$\rho \geq \begin{cases} \dfrac{m}{m-1}\gamma\eta, & \eta \geq 0, \\ \dfrac{n}{n-1}\gamma\eta, & \eta < 0, \end{cases} \qquad \text{where } \gamma = \frac{\beta}{\beta+\lambda}\frac{\beta+\lambda-\mu}{\beta-\mu},$$

as shown in Figure 4(a). They can be converted on the (q, δ)-plane as

$$\delta - (\beta - \mu) \leq \begin{cases} \dfrac{n}{n-1}\dfrac{\beta+\lambda-\mu}{\beta+\lambda}(q-\beta), & q > \beta, \\ \dfrac{m}{m-1}\dfrac{\beta+\lambda-\mu}{\beta+\lambda}(q-\beta), & q \leq \beta, \end{cases}$$

as in Figure 4(b). On the (q, δ)-plane, the boundaries of the two conditions in Proposition 2.2 are graphically represented by the two straight lines L_1 and L_2 in Figure 4(b). The slopes M and N of L_1 and L_2, respectively, are represented solely by the roots n and m, respectively, as in the next proposition, which also shows that a change of the model parameters leads to an expansion or a contraction of the active region as the two lines rotate around $(\beta, \beta - \mu)$.

Proposition 2.3 *(1) The slopes on the (q, δ)-plane*

$$M = \frac{m}{m-1}\frac{\beta+\lambda-\mu}{\beta+\lambda}, \quad N = \frac{n}{n-1}\frac{\beta+\lambda-\mu}{\beta+\lambda}$$

have the following properties

$$N = \frac{m-1}{m}, \quad M = \frac{n-1}{n},$$

and

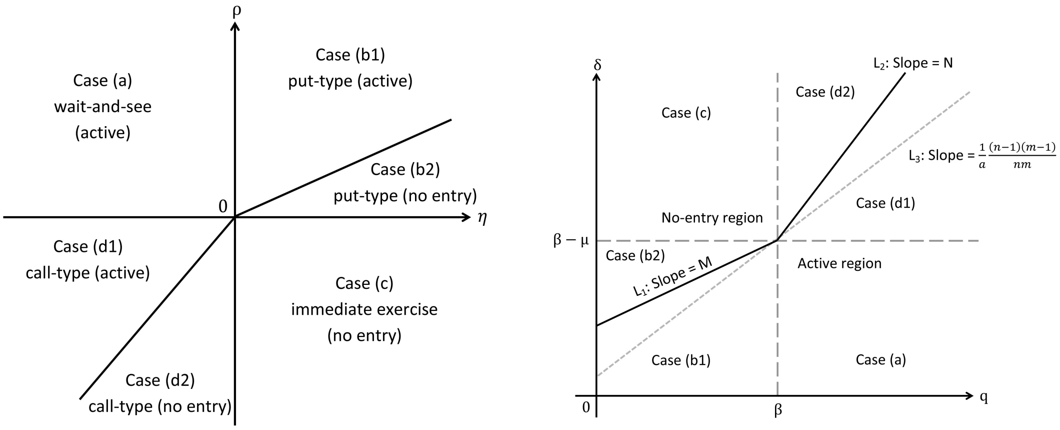
$$\frac{\partial N}{\partial \sigma^2} > 0, \quad \frac{\partial N}{\partial \mu} < 0, \quad \frac{\partial N}{\partial \lambda} < 0, \qquad \frac{\partial M}{\partial \sigma^2} < 0, \quad \frac{\partial M}{\partial \mu} < 0, \quad \frac{\partial M}{\partial \lambda} > 0.$$

(2) The thresholds b_p, b_c have the following properties

$$\frac{\partial b_p}{\partial \lambda} > 0, \qquad \frac{\partial b_c}{\partial \lambda} < 0.$$

Proof.n, m are solutions of the quadratic equation

$$\frac{1}{2}\sigma^2 x(x-1) + \mu x - (\beta+\lambda) = 0.$$



(a) Active region on the (η, ρ)-plane. (b) Active region on the (q, δ)-plane.

Figure 4: Investor's active region. The origin on the (η, ρ)-plane corresponds to the point $(\beta, \beta - \mu)$ on the (q, δ)-plane.

Then it holds that

$$\frac{\beta + \lambda - \mu}{\beta + \lambda} = \frac{(n-1)(m-1)}{nm},$$

and

$$\frac{\partial m}{\partial \sigma^2} > 0, \qquad \frac{\partial m}{\partial \mu} < 0, \qquad \frac{\partial m}{\partial \lambda} < 0, \qquad \frac{\partial n}{\partial \sigma^2} < 0, \qquad \frac{\partial n}{\partial \mu} < 0, \qquad \frac{\partial n}{\partial \lambda} > 0.$$

■

Proposition 2.3 (1) implies that the active region expands as the stock volatility σ increases (more opportunistic trade) and contracts as the recall intensity λ increases (higher recall risk) when the loan fee rate δ and the interest rate q are fixed. The separating lines rotate clockwise when μ increases. With an increase of λ, the thresholds become closer to the original sales price so that the short-seller will close the position more quickly and more conservatively.

Finally, to represent the active condition graphically in relation to L_1 and L_2, we rewrite the definition of a in (6) as

$$\delta - (\beta - \mu) = \frac{1}{a} \frac{(n-1)(m-1)}{nm} (q - \beta),$$

which is drawn as the dashed line L_3 in Figure 4(b). In Case (b) on the domain $q \leq \beta$ in the graph, the active condition $a < (m-1)/m$ is equivalent to the condition

$$\frac{1}{a} \frac{(n-1)(m-1)}{nm} > \frac{n-1}{n}, \tag{11}$$

which implies that the slope M (the RHS of (11)) of line L_1 must be smaller than the slope of L_3 (the LHS of (11)). Similarly, for Case (d), $a > (n-1)/n$ is equivalent to the condition that the slope N of line L_2 is larger than the slope of L_3. In other words, the active condition is equivalent to the condition that the line L_3 lies below the lines L_1 and L_2.

3. Comparative Analysis with Numerical Examples

In this section, we conduct a comparative analysis so as to understand the properties of the decision making using numerical examples for various levels of recall intensity, loan fee rate and interest rate. Unless otherwise stated, we assume the following parameter values: the short-seller's discount rate $\beta = 0.05$, the stock price volatility $\sigma = 0.2$, and the recall intensity $\lambda = 0.02$. The expected return will be either $\mu = -0.03$ (down-market) or $\mu = 0.03$ (up-market). The initial stock price and initial entry price are taken to be $x = K = 100$. To illustrate, we focus on the put-type problem (Case (b)) and the call-type problem (Case (d)). Recall that the thresholds are given by

$$b_p = \frac{m}{m-1} aK = \left(1 + \frac{1}{m-1}\right)\left(1 - \frac{\mu}{\beta+\lambda}\right)\frac{\eta}{\rho}K, \tag{12}$$

$$b_c = \frac{n}{n-1} aK = \left(1 + \frac{1}{n-1}\right)\left(1 - \frac{\mu}{\beta+\lambda}\right)\frac{\eta}{\rho}K, \tag{13}$$

$$\frac{\eta}{\rho} = \frac{1 - \dfrac{q}{\beta}}{1 - \dfrac{\delta}{\beta - \mu}}. \tag{14}$$

3.1. The impact of loan fee rate and interest rate

First, we study the impact of the loan fee rate and interest rate in the short-seller's problem. The basic role of the two parameters in the decision making by the short-seller is to determine $h_\infty(x)$, one of the two lower bounds of the value function

$$h_\infty(x) = \phi(x) - g(x) + K - x = \left(1 - \frac{\beta}{\beta+\lambda}\eta\right) K - \left(1 - \frac{\beta-\mu}{\beta+\lambda-\mu}\rho\right) x.$$

A different q gives a different intersection of the lower bound line via η and a different δ gives a different slope via ρ. When $q = \delta = 0$, we see $\eta = \rho = 1$, $g(x) = K - x$ and $h_\infty(x) = \phi(x)$.

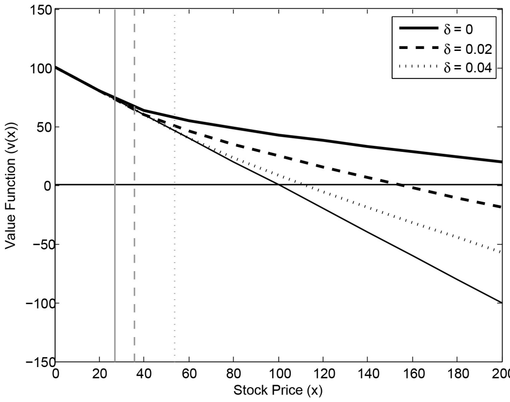
Put-type problem ($\rho > 0, \eta > 0$)

As the loan fee rate δ increases, the coefficient ρ decreases so that the slope of $h_\infty(x)$ becomes steeper while maintaining the same intersection. This generates a smaller lower bound and hence reduces the value function, as we show in Figure 5(a). On the other hand, as q increases and hence η decreases, the intersection of $h_\infty(x)$ increases while maintaining the same slope. Hence, the lower bound is increased and this produces a higher value function (Figure 5(b)). When x is close to the optimal threshold, it is less obvious by simply looking at the formula to examine the impact of the loan fee and interest income.

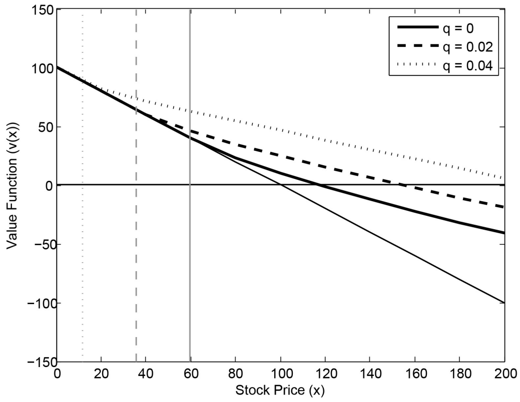
The impact of the loan fee rate and interest rate on the optimal threshold are shown in Figure 5(c) and 5(d) in a very different manner, where μ is taken as $\mu = -0.03$. The optimal threshold (12) is an increasing hyperbolic function of the loan fee rate δ and a decreasing linear function of the interest rate q. Figure 5(c) shows that the threshold is highly sensitive to the loan fee rate that is less than but close to $\beta - \mu = 0.08$. From (12) and (14), we see that as $\delta \uparrow \beta - \mu$ ($\rho \downarrow 0$), the threshold goes to infinity and we approach the immediate exercise solution (no entry). Figure 5(d) shows that the threshold depends linearly on the interest rate q. Similarly, as $q \uparrow \beta$ ($\eta \downarrow 0$), the threshold goes to zero and we recover the wait-and-see solution.

Call-type problem ($\rho < 0, \eta < 0$)

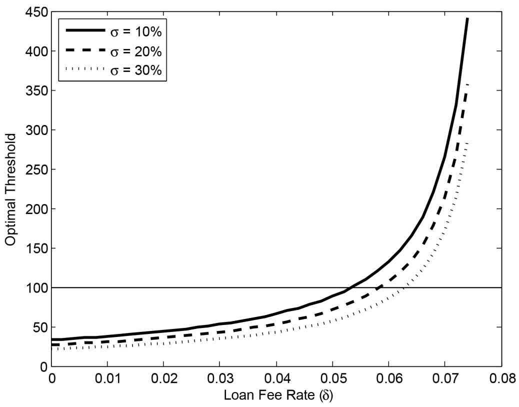
Similar analysis with varying δ and q can be applied to the call-type problem as shown in Figure 6(a) and 6(b). As the loan fee rate δ increases, the cost of holding the short position is more expensive and the investor should stop earlier, as can be seen from (13). Similar to the put-type problem, the optimal threshold is also highly sensitive to the loan fee (Figure



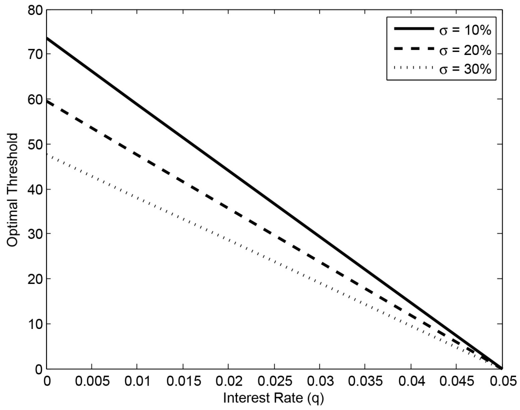
(a) Value function with loan fee rate $\delta = 0$, 0.02 and 0.04 and interest rate fixed at $q = 0.02$. The optimal thresholds are determined as 26.8, 35.7 and 53.6, respectively.



(b) Value function with interest rate $q = 0$, 0.02 and 0.04 and loan fee rate fixed at $\delta = 0.02$. The optimal thresholds are determined as 59.5, 35.7 and 11.9, respectively.



(c) Optimal threshold versus loan fee rate δ with interest rate fixed at $q = 0.02$.



(d) Optimal threshold versus interest rate q with loan fee rate fixed at $\delta = 0.02$.

Figure 5: Value function and optimal threshold for the put-type problem.

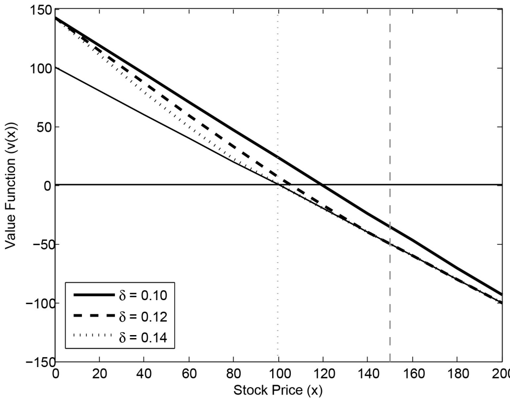
6(c)). From (13) and (14), as $\delta \downarrow \beta - \mu$ ($\rho \uparrow 0$), the threshold goes to infinity and we have the wait-and-see solution (Figure 6(d)).

For completeness, we report in Figure 7 the value function in an up-market ($\mu = 0.03$) corresponding to Figures 5(a), 5(b), 6(a) and 6(b). A similar argument holds for the case of an up-market.

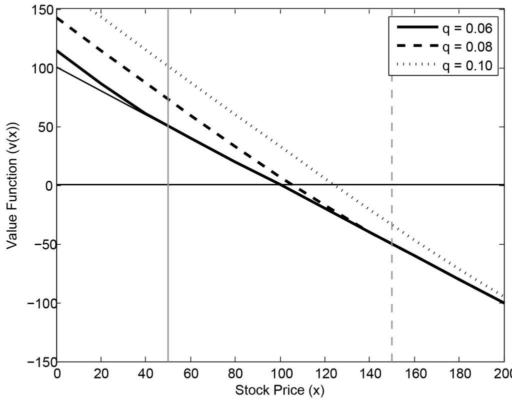
3.2. The impact of recall risk

The next objective is examining the impact of recall risk λ on the value function and the optimal threshold. For the time being, let us assume the simple case of $\delta = q = 0$ (equivalently, $\eta = \rho = 1$), which implies put-type problems and excludes the effects of the loan fee and interest income. We saw in the previous subsection that δ and q determine the location and the shape of the lower bound $h_\infty(x)$. Thus, from the results with $\delta = q = 0$, one can easily understand the corresponding results with nonzero δ and q by shifting vertically and rotating the line.

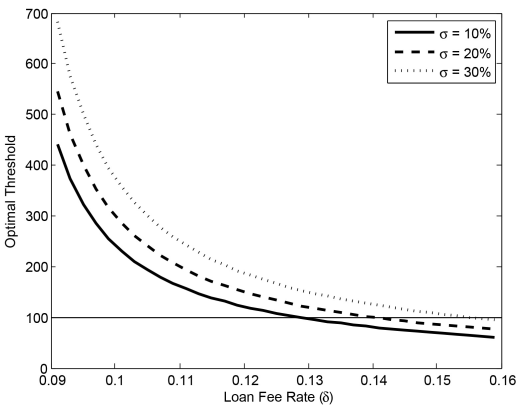
Figure 8(a) and 8(b) report the value functions $v(x)$ and the optimal thresholds b_p for different values of recall intensity λ with $\mu = -0.03$ (down-market, Figure 8(a)) and $\mu = 0.03$ (up-market, Figure 8(b)), respectively, based on Proposition 2.1 and Theorem 2.1.



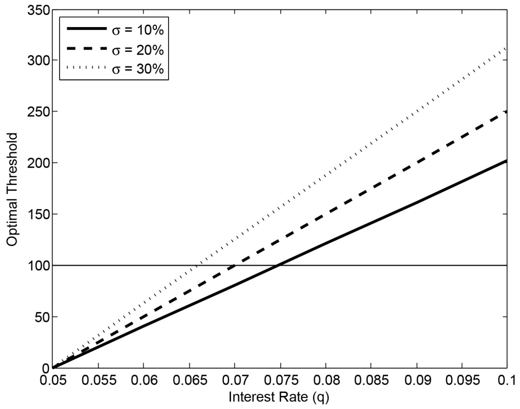
(a) Value function with loan fee rate $\delta = 0.10$, 0.12 and 0.14 and interest rate fixed at $q = 0.08$. The optimal thresholds are determined as 300, 150 and 100, respectively.



(b) Value function with interest rate $q = 0.06$, 0.08 and 0.10 and loan fee rate fixed at $\delta = 0.12$. The optimal thresholds are determined as 50, 150 and 250, respectively.



(c) Optimal threshold versus loan fee rate δ with interest rate fixed at $q = 0.08$.



(d) Optimal threshold versus interest rate q with loan fee rate fixed at $\delta = 0.12$.

Figure 6: Value function and optimal threshold for the call-type problem.

The corresponding optimal thresholds are calculated and marked for illustrative purposes.

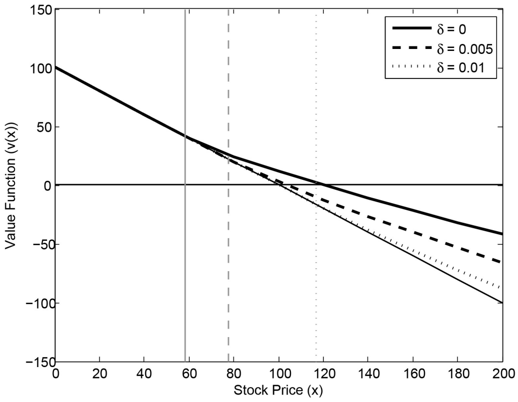
The value function $v(x)$ has two lower bounds in this case, $K - x$ and $\phi(x)$. When $\lambda = 0$, we have $\phi(x) = 0$ so that the value function is positive for all x in $\mathbb{R}_+$ (solid line).[6] As λ departs from zero, $\phi(x)$ slopes downward as in (5). Therefore, the value function can also take negative values when x is large. This is demonstrated in the cases where $\lambda = 0.02$ (dashed line) and $\lambda = 0.05$ (dotted line).

Comparing the two figures, the thresholds appear to be more sensitive to the recall risk when $\mu > 0$ (up-market) than when $\mu < 0$ (down-market). As the result the value functions do so. This phenomenon can be explained by taking a closer look at the sensitivity of b_p with respect to λ, which is positive by Proposition 2.3 (2). By differentiating (12) with respect to λ, we see

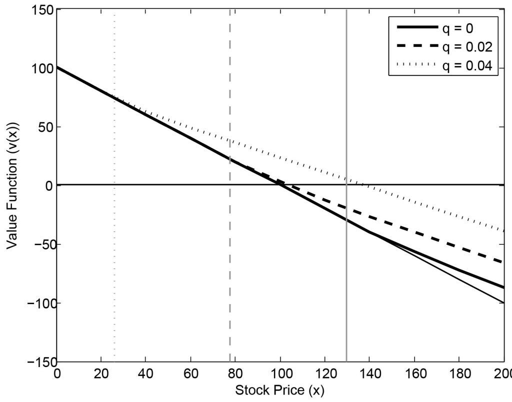
$$\frac{\partial b_p}{\partial \lambda} = \left[\frac{\partial m}{\partial \lambda} \frac{-1}{(m-1)^2} \left(1 - \frac{\mu}{\beta + \lambda} \right) + \left(1 + \frac{1}{m-1} \right) \frac{\mu}{(\beta + \lambda)^2} \right] \frac{\eta}{\rho} K.$$

As in the proof of Proposition 2.3, we have $\partial m / \partial \lambda < 0$. Hence, the first term in the

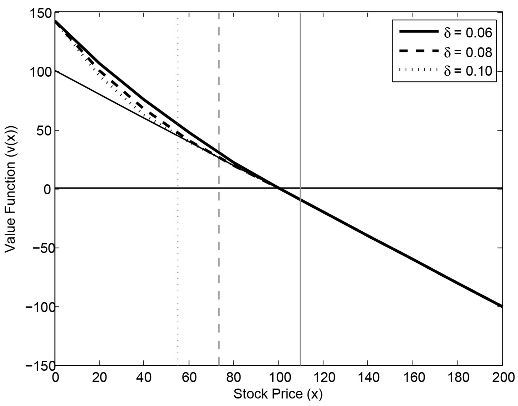
[6]This corresponds to a standard real option problem in an infinite horizon.



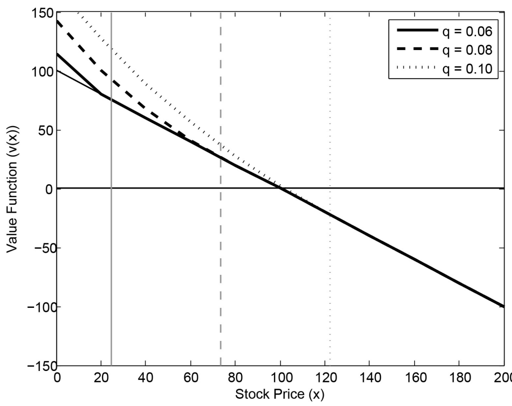
(a) Value function with loan fee rate $\delta = 0$, 0.005 and 0.01 and interest rate fixed at $q = 0.02$. The optimal thresholds are determined as 58.4, 77.9 and 116.8, respectively.



(b) Value function with interest rate $q =$ 0, 0.02 and 0.04 and loan fee rate fixed at $\delta = 0.005$. The optimal thresholds are determined as 129.8, 77.9 and 26.0, respectively.



(c) Value function with loan fee rate $\delta =$ 0.06, 0.08 and 0.10 and interest rate fixed at $q = 0.08$. The optimal thresholds are determined as 110, 73.4 and 55.0, respectively.

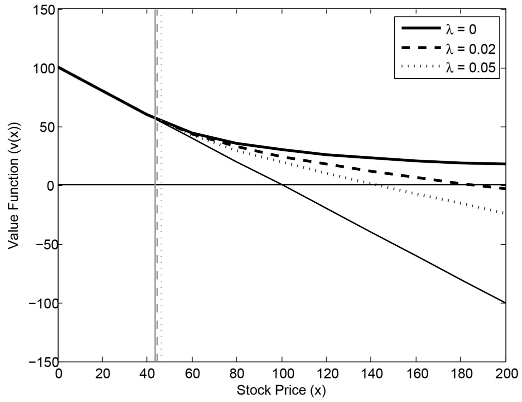


(d) Value function with interest rate $q =$ 0.06, 0.08 and 0.10 and loan fee rate fixed at $\delta = 0.08$. The optimal thresholds are determined as 24.5, 73.4 and 122.3, respectively.

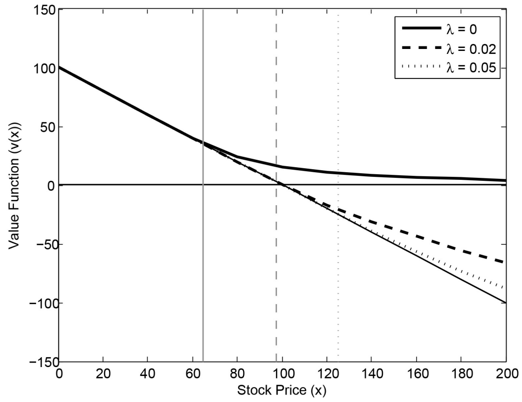
Figure 7: Value function in an up-market with $\mu = 0.03$.

brackets is always positive while the sign of the second term is the same as the sign of μ. It follows that they are cancelled out to some extent when μ is negative. This is the reason for the higher sensitivity of the thresholds for positive μ (up-market) than negative μ. Figure 8(e) confirms the claim: the optimal threshold increases strongly with the intensity when $\mu = 0.03$ (dotted line), while the corresponding impact is much weaker when $\mu = 0$ and $\mu = -0.03$ (dashed line and solid line, respectively). It implies that the short-seller has to be very nervous about the timing of taking capital profit (short covering) from a stock price movement with a positive trend as the recall risk increases, so that she will be satisfied with a smaller profit. On the other hand, in a down-trending market, the short-seller does not need to modify the target price so much in accordance with the magnitude of the recall risk.

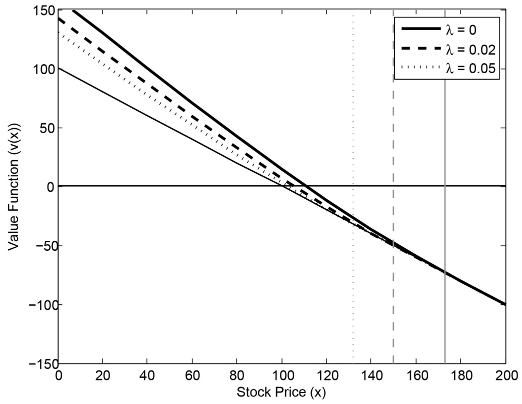
As mentioned, for general δ, q (or η, ρ) leading to a put-type problem, similar analysis and interpretations hold by shifting vertically and rotating the line of the lower bound $h_\infty(x) = \phi(x) - g(x) + K - x$. From its expression, we see that the intensity λ somewhat



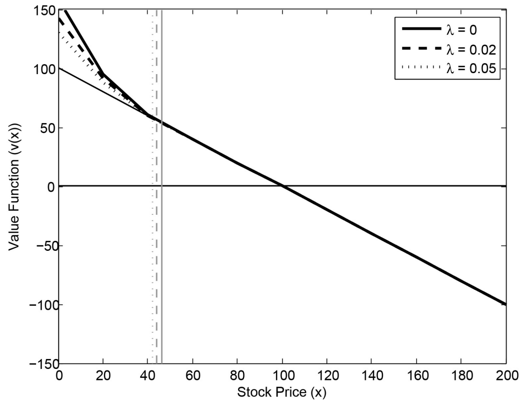
(a) Value function in a down-market ($\mu = -0.03$) with recall intensity $\lambda = 0$, 0.02 and 0.05, $\delta = q = 0$. The optimal thresholds are determined as 43.4, 44.6 and 46.1, respectively.



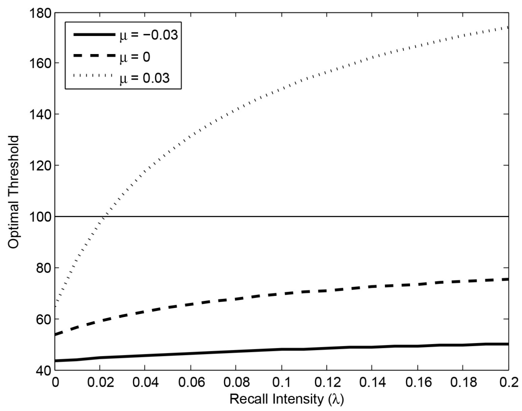
(b) Value function in an up-market ($\mu = 0.03$) with recall intensity $\lambda = 0$, 0.02 and 0.05, $\delta = q = 0$. The optimal thresholds are determined as 65.0, 97.3 and 125.0, respectively.



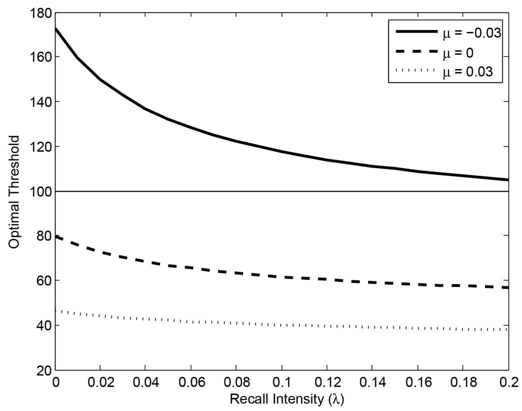
(c) Value function in a down-market ($\mu = -0.03$) with recall intensity $\lambda = 0$, 0.02 and 0.05, $\delta = 0.12$ and $q = 0.08$. The optimal thresholds are determined as 173.0, 150 and 132.2, respectively.



(d) Value function in an up-market ($\mu = 0.03$) with recall intensity $\lambda = 0$, 0.02 and 0.05, $\delta = 0.12$ and $q = 0.08$. The optimal thresholds are determined as 46.2, 44.0 and 42.0, respectively.



(e) Optimal threshold versus recall intensity λ with $\delta = q = 0$.



(f) Optimal threshold versus recall intensity λ with $\delta = 0.12$ and $q = 0.08$.

Figure 8: The impact of recall risk.

dampens the effects of η and ρ on the lower bound $h_\infty(x)$. This is because the recall risk reduces the expected holding time of the short position and hence diminishes the role of the running cost.

For the call-type problems, these impacts are opposite to the case of the put-type problems. By differentiating (13) with respect to λ, we see

$$\frac{\partial b_c}{\partial \lambda} = \left[\frac{\partial n}{\partial \lambda}\frac{-1}{(n-1)^2}\left(1-\frac{\mu}{\beta+\lambda}\right)+\left(1+\frac{1}{n-1}\right)\frac{\mu}{(\beta+\lambda)^2}\right]\frac{\eta}{\rho}K,$$

which is negative by Proposition 2.3. Moreover, we have $\partial n/\partial \lambda > 0$ such that the first term in brackets is always negative while the sign of the second term is the same as μ. Hence, the sensitivity of the call-type threshold is higher for negative μ (down-market) than positive μ (up-market). Figure 8(f) shows that the threshold decreases notably against the intensity when $\mu = -0.03$ (solid line), while the impact is as large when $\mu = 0$ and $\mu = 0.03$ (dashed and dotted lines). As a consequence, the impact of recall risk on the value function is more significant in a down-market than in an up-market for a call-type problem (Figure 8(c) and 8(d)).

3.3. The impact of volatility

We then turn our attention to how the optimal threshold depends on the stock price volatility. Proposition 2.3 together with (12) imply that the put-type optimal threshold is decreasing with the volatility because the volatility affects b_p only through the term $m/(m-1)$ in (12) and $\partial m/\partial \sigma^2 > 0$. Figure 9(a) shows that the threshold decreases gradually with the stock price volatility under the assumption $\delta = q = 0$. The intuition is that the strategy is more opportunistic given a higher probability of the stock price declining to below the entry price. In contrast, the call-type optimal threshold is increasing with the volatility because of the term $n/(n-1)$ in (13) and $\partial n/\partial \sigma^2 < 0$ (Figure 9(c)).

Note that when $\lambda = 0$ (no recall risk) and $\delta = q = 0$, we have $b_p = (m/(m-1))K < K$ such that the investor is active regardless of the market direction. In the presence of recall risk with $\delta = q = 0$, the sign of μ matters because of

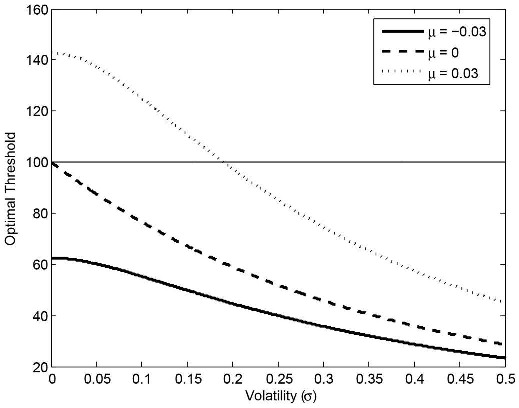
$$a = \frac{1+\dfrac{\lambda}{\beta-\mu}}{1+\dfrac{\lambda}{\beta}},$$

by (6). When $\mu \le 0$, the investor is always active (dashed and solid lines in Figure 9(a)) because in this case $a \le 1 < (m-1)/m$. In contrast, when $\mu > 0$, the investor only trades when the volatility is high enough (dotted line in Figure 9(a)) so that the derived value of $m < 0$ becomes large and satisfies $a < (m-1)/m$.

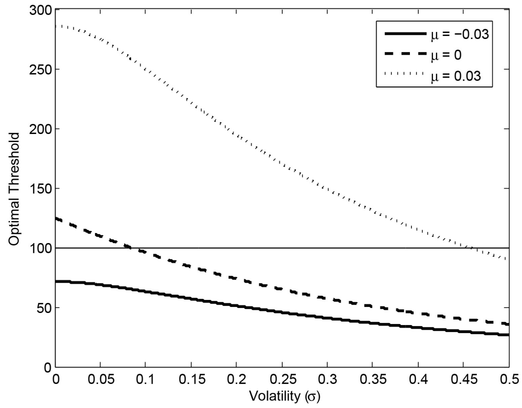
For nonzero δ, q, the condition on μ for the active region (put type or call type) can be obtained in accordance with the value of η/ρ. From (6), we see that the parameter a is proportional to the ratio η/ρ. For a put-type problem, keeping η/ρ and a fixed, the investor only trades when the volatility is high enough and $m < 0$ is large enough to satisfy $a < (m-1)/m$ (see Figure 9(a) versus 9(b)). Similarly, for a call-type problem, trade happens when the volatility is high enough such that $n > 1$ becomes small enough to satisfy the condition $a > (n-1)/n$ (see Figure 9(c) versus 9(d)).

4. Conclusion

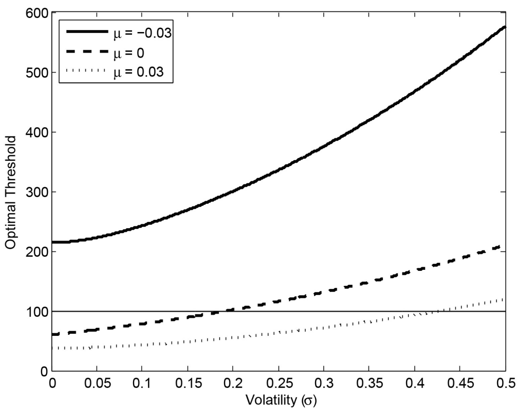
In this paper, we studied the optimal stopping problem related to a short-selling strategy in a financial market. In a short-selling transaction, the short-seller faces the possibility



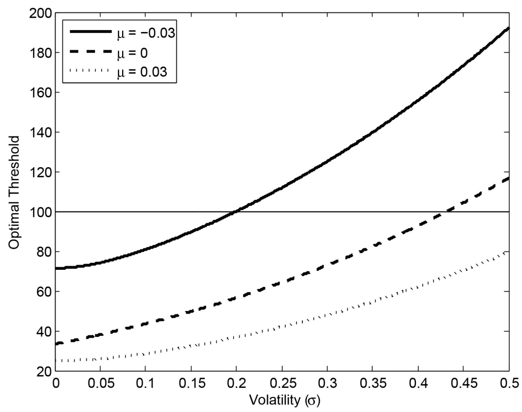
(a) Optimal threshold versus volatility σ with $\delta = q = 0$.



(b) Optimal threshold versus volatility σ with $\delta = 0.01$ and $q = 0$.



(c) Optimal threshold versus volatility σ with $\delta = 0.10$ and $q = 0.08$.



(d) Optimal threshold versus volatility σ with $\delta = 0.14$ and $q = 0.08$.

Figure 9: The impact of volatility.

of a broker recall and the short-seller might be forced to stop the strategy involuntarily and experience a loss. Our results show that, depending on the levels of the loan fee and interest rate, the optimal stopping problem is either a put-type problem with a down-and-out stopping rule or a call-type problem with an up-and-out stopping rule. The value function may become negative because of the possibility of a forced termination, and the short-seller is likely to stop earlier at the closer optimal threshold to the entry price as a result of the random recall (the put-type problem) or the relatively expensive net running cost of keeping the position (the call-type problem). The analysis in this paper will be sufficient for investors to make a short-selling decision in a simple setting. As an extension, more realistic factors such as a stop-loss limit or a nondiffusion-type stock price process may be included. These are left for future research.

References

[1] P. Carr: Randomization and the American Put. *Review of Financial Studies*, **11** (1998), 597–626.

[2] P. Carr and V. Linetsky: The Valuation of Executive Stock Options in an Intensity-Based Framework. *European Finance Review*, **4** (2000), 211–230.

[3] G. D'Avolio: The Market of Borrowing Stock. *Journal of Financial Economics*, **26**

(2002), 271–306.
[4] P. Dupuis and H. Wang: Optimal Stopping with Random Intervention Times. *Advances in Applied Probability*, **34** (2002), 141–157.
[5] M. Egami and S. Dayanik: Optimal Stopping Problems for Asset Management. *Advances in Applied Probability*, **44** (2012), 655–677.
[6] X. Guo and J. Liu: Stopping at the Maximum of Geometric Brownian Motion when Signals Are Received. *Journal of Applied Probability*, **42** (2005), 826–838.
[7] X. Guo and Q. Zhang: Optimal Selling Rules in a Regime Switching Model. *IEEE Transactions on Automatic Control*, **50** (2005), 1450–1455.
[8] G. Peskir and A. Shiryaev: *Optimal Stopping and Free-Boundary Problems* (Birkhäuser Verlag, 2005).
[9] S.R. Pliska and J. Ye: Optimal Life Insurance Purchase and Consumption/Investment under Uncertain Lifetime. *Journal of Banking and Finance*, **31** (2007), 1307–1319.
[10] Q. Zhang: Stock Trading: An Optimal Selling Rule. *SIAM Journal of Control and Optimization*, **40** (2001), 64–87.

A. Proof of Proposition 2.1

We will prove Proposition 2.1 by constructing a candidate of the value function for each case, and then verifying their optimality.

Case (a) When $\rho \geq 0$, $\eta \leq 0$, $u = \phi$, $\tau_E^* = \infty$,

Case (b) When $\rho > 0$, $\eta > 0$, $u = u_p$, $\tau_E^* = \inf\{t \geq 0; \quad X_t \leq b_p\}$,

Case (c) When $\rho \leq 0$, $\eta \geq 0$, $u = g$, $\tau_E^* = 0$,

Case (d) When $\rho < 0$, $\eta < 0$, $u = u_c$, $\tau_E^* = \inf\{t \geq 0; \quad X_t \geq b_c\}$.

A.1. Construction of candidates of the value function

The occurrence of τ_R can be regarded as a regime switch. Thus, let us define the regime $\epsilon(t) = 1_{\{\tau_R > t\}}$. A candidate for the value function before the occurrence of τ_R is denoted by $\hat{u}(\cdot, 1)$ and after τ_R by $\hat{u}(\cdot, 0)$. Obviously, $\hat{u}(x, 0) = g(x) = \eta K - \rho x$. Similarly, we denote the corresponding candidate of the optimal stopping time by $\widehat{\tau^*}$. $\hat{u}(x, 1)$ is constructed so that

$$
\begin{aligned}
&\hat{u}(x,1) \geq g(x) \quad \forall x \in \mathbb{R}_+,\\
&\widehat{\mathcal{L}}\hat{u}(x,1) \leq 0 \quad \forall x \in \mathbb{R}_+,\\
&\widehat{\mathcal{L}}\hat{u}(x,1) = 0 \quad \text{on the continuation region C},
\end{aligned}
$$

where $\widehat{\mathcal{L}}$ is defined in (8).

A.1.1. Case (a)

When $\rho \geq 0, \eta \leq 0$, $\phi(x)$ satisfies $\phi(x) \geq g(x)$ and $\widehat{\mathcal{L}}\phi(x) = 0$ for all $x \in \mathbb{R}_+$. Hence, $\hat{u}(x,1) = \phi(x)$ and the corresponding optimal stopping time is $\widehat{\tau^*} = \infty$.

A.1.2. Case (b)

When $\rho > 0$, $\eta > 0$, $g(x) \geq \phi(x)$ for sufficiently small x while $g(x) < \phi(x)$ for sufficiently large x. Therefore, we look for $\hat{u}(x,1)$ satisfying the put-type problem

$$
\begin{aligned}
&\widehat{\mathcal{L}}\hat{u}(x,1) = 0 \quad \text{for } x \in (b_p, \infty),\\
&\hat{u}(b_p,1) = g(b_p) \text{ (value matching)}, \quad \frac{d}{dx}\hat{u}(b_p,1) = \frac{d}{dx} g(b_p) \text{ (smooth pasting)}.
\end{aligned}
$$

By the standard argument in the context of a real option, the candidate value function for the put-type problem is given by

$$\hat{u}(x,1) = \begin{cases} g(x) & x \in (0, b_p], \\ Ax^m + \phi(x) & x \in (b_p, \infty), \end{cases}$$

where

$$\begin{aligned} b_p &= \frac{m}{m-1}\frac{\beta}{\beta+\lambda}\frac{\beta+\lambda-\mu}{\beta-\mu}\frac{\eta}{\rho}K = \frac{m}{m-1}aK, \\ A &= \left[\left(\frac{\beta}{\beta+\lambda}\right)\eta K - \left(\frac{\beta-\mu}{\beta+\lambda-\mu}\right)\rho b_p\right]\left(\frac{1}{b_p}\right)^m = (g(b_p) - \phi(b_p))\left(\frac{1}{b_p}\right)^m. \end{aligned}$$

It is easy to check that $0 < b_p < aK$ because $\rho > 0$, $\eta > 0$ and $m < 0$.

On $(0, b_p]$, $\hat{u}(x,1) = g(x) = \eta K - \rho x$ satisfies the condition

$$\widehat{\mathcal{L}}g(x) = (\beta - \mu)\,\rho x - \beta\eta K \le \beta\left[\left(1 - \frac{1}{n}\right)\eta K - \eta K\right] = -\beta\frac{1}{n}\eta K < 0,$$

given that $n > 1$, $\rho > 0$ and $\eta > 0$.

A.1.3. Case (c)

When $\rho \le 0$, $\eta \ge 0$, $\widehat{\mathcal{L}}g(x) \le 0$ for all $x \in \mathbb{R}_+$ and $g(x)$ is always greater than $\phi(x)$. There is no reason to wait. Thus, $\hat{u}(x,1) = g(x)$ and $\widehat{\tau^*} = 0$.

A.1.4. Case (d)

When $\rho < 0$, $\eta < 0$, $g(x) < \phi(x)$ for sufficiently small x while $g(x) \ge \phi(x)$ for sufficiently large x. Therefore, we look for $\hat{u}(x,1)$ satisfying the call-type problem

$$\begin{aligned} &\widehat{\mathcal{L}}\hat{u}(x,1) = 0 \quad \text{for } x \in (0, b_c), \\ &\hat{u}(b_c,1) = g(b_c) \text{ (value matching)}, \quad \frac{d}{dx}\hat{u}(b_c,1) = \frac{d}{dx}g(b_c) \text{ (smooth pasting)}. \end{aligned}$$

By a similar argument to Case (b), the candidate value function for the call-type problem is given by

$$\hat{u}(x,1) = \begin{cases} Bx^n + \phi(x) & x \in (0, b_c), \\ g(x) & x \in [b_c, \infty), \end{cases}$$

where

$$b_c = \frac{n}{n-1}aK, \qquad B = (g(b_c) - \phi(b_c))\left(\frac{1}{b_c}\right)^n.$$

One can check easily that $aK < b_c$ and $\widehat{\mathcal{L}}g(x) < 0$ on $[b_c, \infty)$.

A.2. Verification of the optimality

$\hat{u}(x,1)$ is constructed so that $\widehat{\mathcal{L}}\hat{u}(x,1) \le 0$ for all $x \in \mathbb{R}_+$. Therefore, $e^{-\beta t}\widehat{\mathcal{L}}\hat{u}(x,1) \le 0$ for all x. It implies that a process $\{e^{-\beta t}\hat{u}(X(t), \epsilon(t)) | 0 \le t \le \tau_R\}$ is a supermartingale

$$\hat{u}(x,1) \ge \mathbb{E}_x[e^{-\beta t}\hat{u}(X(t), \epsilon(t))], \quad \forall t \le \tau_R.$$

Then we obtain for any stopping time τ_E

$$\hat{u}(x,1) \ge \mathbb{E}_x\left[e^{-\beta(\tau_E \wedge \tau_R)}\hat{u}(X_{\tau_E \wedge \tau_R}, \epsilon(\tau_E \wedge \tau_R))\right] \ge \mathbb{E}_x\left[e^{-\beta(\tau_E \wedge \tau_R)}g(X_{\tau_E \wedge \tau_R})\right],$$

where Doob's optional sampling theorem is applied on the first inequality and $\hat{u}(x,i) \ge g(x)$ $(i = 0,1)$ on the second inequality. Hence, for each case we see $\hat{u}(x,1) \ge \mathbb{E}_x\left[e^{-\beta(\tau_E \wedge \tau_R)}g\left(X_{\tau_E \wedge \tau_R}\right)\right]$ for all τ_E. The fact that $\widehat{\tau^*}$ attains the equality or the supremum of the right-hand side will be discussed for each case below based on the construction.

A.2.1. Case (a)

Write $\tau_k = k$ and define

$$\phi_k(x) \triangleq \mathbb{E}_x\left[e^{-\beta(\tau_k\wedge\tau_R)} g\left(X_{\tau_k\wedge\tau_R}\right)\right].$$

By splitting the event $1 = 1_{\{\tau_R\leq\tau_k\}} + 1_{\{\tau_R>\tau_k\}}$, we see

$$\phi_k(x) \quad = \quad \phi(x) + \eta K \frac{\beta}{\beta+\lambda} e^{-(\beta+\lambda)\tau_k} - \rho x \frac{\beta-\mu}{\beta-\mu+\lambda} e^{-(\beta-\mu+\lambda)\tau_k}.$$

Then $\lim_{k\to\infty}\phi_k(x) = \phi(x)$ for all $x \in \mathbb{R}_+$ implies that $\hat{u}(x,1) = \phi(x)$ is equal to $\sup_{\tau_E\in\mathcal{A}} \mathbb{E}_x\left[e^{-\beta(\tau_E\wedge\tau_R)} g\left(X_{\tau_E\wedge\tau_R}\right)\right]$.

A.2.2. Case (b)

We follow the procedure used by [7]. C is the continuation region $C = \{x > b_p\}$ and D is the stopping region $D = \mathbb{R}_+ \setminus C$.

First, we consider the case of $\lambda > 0$. As $\tau_R < \infty$ a.s., $\widehat{\tau^*} \wedge \tau_R < \infty$ a.s. Furthermore, $\mathbb{E}_x[\widehat{\tau^*}\wedge\tau_R] \leq \mathbb{E}_x[\tau_R] = 1/\lambda < \infty$ so that Dynkin's formula can be applied for a stopping time $\widehat{\tau^*}\wedge\tau_R$.

$$\mathbb{E}_x\left[e^{-\beta\left(\widehat{\tau^*}\wedge\tau_R\right)}\hat{u}(X(\widehat{\tau^*}\wedge\tau_R), \epsilon(\widehat{\tau^*}\wedge\tau_R))\right] = \hat{u}(x) + \mathbb{E}_x\left[\int_0^{\widehat{\tau^*}\wedge\tau_R} e^{-\beta t}\widehat{\mathcal{L}}\hat{u}(X(t),1)dt\right] = \hat{u}(x,1).$$

$(X(\widehat{\tau^*}\wedge\tau_R), \epsilon(\widehat{\tau^*}\wedge\tau_R)) \notin D = \{(x,i)|\hat{u}(x,i) > g(x)\}$ implies

$$\hat{u}(X(\widehat{\tau^*}\wedge\tau_R), \epsilon(\widehat{\tau^*}\wedge\tau_R)) = g(X(\widehat{\tau^*}\wedge\tau_R)).$$

Hence

$$\hat{u}(x,1) \quad = \mathbb{E}_x\left[e^{-\beta\left(\widehat{\tau^*}\wedge\tau_R\right)} g(X(\widehat{\tau^*}\wedge\tau_R))\right].$$

Next, let us consider the case of $\lambda = 0$. In this case $\phi(x) \equiv 0$. We take an increasing sequence of open sets

$$C_k = C \cap \{x < k\}$$

and stopping times

$$\tau_k = \inf\{t \in [0,\infty)\,; X_t \notin C_k\}$$

for $k = 1, 2, ...$ The continuity of $X = (X_t)_{t\geq 0}$ implies that $X_{\tau_k} = b_p$ or $X_{\tau_k} = k$. When k is sufficiently large $k > b_p$, we see $C_k = (b_p, k)$ and $\mathbb{P}\{X_{\tau_k} > k\} = 0$. Then, we can express

$$\hat{u}\left(X_{\tau_k}, 1\right) = \hat{u}\left(X_{\tau_k}, 1\right)\mathbf{1}_{\left\{X_{\tau_k}=k\right\}} + \hat{u}\left(X_{\tau_k}, 1\right)\mathbf{1}_{\left\{X_{\tau_k}<k\right\}}.$$

It is noted that:

1. $\tau_k \uparrow \widehat{\tau^*}$ a.s. and this implies τ_k converges to $\widehat{\tau^*}$ in distribution;
2. $\tau_k < \infty$ a.s. and $\mathbb{E}_x[\tau_k] < \infty$ for C_k is a bounded region and X_t has a unique strong solution; and
3. on the event $\{X_{\tau_k} < k\}$, we have $\hat{u}\left(X_{\tau_k}, 1\right) = g\left(X_{\tau_k}\right)$.

Without the random recall, the candidate value function $\hat{u}(x,1)$ tends to $\phi(x) \equiv 0$ as $X_t \to \infty$. For any $\varepsilon > 0$, there exists a k_1 such that for all $k > k_1$:

$$\mathbb{E}_x \left[e^{-\beta\tau_k} \hat{u}\left(X_{\tau_k}, 1\right) \mathbf{1}_{\left\{X_{\tau_k}=k\right\}} \right] < \frac{1}{2}\varepsilon.$$

Moreover, τ_k converges to $\widehat{\tau^*}$ in distribution and there exists a k_2 such that for all $k > k_2$

$$\mathbb{E}_x \left[e^{-\beta\tau_k} g\left(X_{\tau_k}\right) \right] < \frac{1}{2}\varepsilon + \mathbb{E}_x \left[e^{-\beta\widehat{\tau^*}} g\left(X_{\widehat{\tau^*}}\right) \right].$$

Therefore, there exists $k_3 > \max\left(k_1, k_2\right)$ such that for all $k > k_3$

$$\begin{aligned}
\mathbb{E}_x \left[e^{-\beta\tau_k} \hat{u}\left(X_{\tau_k}, 1\right) \right] &= \mathbb{E}_x \left[e^{-\beta\tau_k} \hat{u}\left(X_{\tau_k}, 1\right) \mathbf{1}_{\left\{X_{\tau_k}<k\right\}} \right] + \mathbb{E}_x \left[e^{-\beta\tau_k} \hat{u}\left(X_{\tau_k}, 1\right) \mathbf{1}_{\left\{X_{\tau_k}=k\right\}} \right] \\
&= \mathbb{E}_x \left[e^{-\beta\tau_k} g\left(X_{\tau_k}\right) \mathbf{1}_{\left\{X_{\tau_k}<k\right\}} \right] + \mathbb{E}_x \left[e^{-\beta\tau_k} \hat{u}\left(X_{\tau_k}, 1\right) \mathbf{1}_{\left\{X_{\tau_k}=k\right\}} \right] \\
&< \varepsilon + \mathbb{E}_x \left[e^{-\beta\widehat{\tau^*}} g\left(X_{\widehat{\tau^*}}\right) \right].
\end{aligned}$$

By Dynkin's formula, we have

$$\mathbb{E}_x \left[e^{-\beta\tau_k} \hat{u}\left(X_{\tau_k}, 1\right) \right] = \hat{u}(x,1) + \mathbb{E}_x \left[\int_0^{\tau_k} e^{-\beta t} \widehat{\mathcal{L}} \hat{u}\left(X_t, 1\right) dt \right] = \hat{u}(x,1),$$

because $\widehat{\mathcal{L}} \hat{u}\left(X_t, 1\right) = 0$ for $x \in C_k$. As a result, we have $\hat{u}(x,1) - \varepsilon < \mathbb{E}_x \left[e^{-\beta\tau_E^*} g\left(X_{\widehat{\tau^*}}\right) \right]$ for any $\varepsilon > 0$ so that $\hat{u}(x,1) = \sup_{\tau_E} \mathbb{E}_x \left[e^{-\beta\tau_E} g\left(X_{\tau_E}\right) \right]$ in this case.

We conclude that $\hat{u}(x,1)$ is equal to the value function

$$u(x) = \sup_{\tau_E \in \mathcal{A}} \mathbb{E}_x \left[e^{-\beta\left(\tau_E \wedge \tau_R\right)} g\left(X_{\tau_E \wedge \tau_R}\right) \right]$$

and $\widehat{\tau^*}$ is the optimal stopping time τ_E^*.

A.2.3. Case (c)

It is trivial that $\widehat{\tau^*} = 0$ attains $g(x)$.

A.2.4. Case (d)

By modifying the sequence of open sets to $C_k = C \cap \{1/k < x < k\}$, the same arguments as used for Case (b) hold for Case (d).