
Volume 146 Number 4 July 2011 ISSN 0022-0531
ELSEVIER
JOURNAL OF
Economic
Theory
Editors: Christian Hellwig and Karl Shell
Associate Editors
Marco Battaglini
Dirk Bergemann
Alberto Bisin
Yeon-Koo Che
In-Koo Cho
Olivier Compte
Philip H. Dybvig
Larry G. Epstein
Yossi Feinberg
Christian Ghiglino
Jacob Goeree
Piero Gottardi
Jean-Michel Grandmont
Hugo Hopenhayn
Johannes Horner
Vijay Krishna
Felix Kubler
Ricardo Lagos
Massimo Marinacci
Thomas Mariotti
David Martimort
Kiminori Matsuyama
Richard McLean
Giuseppe Moscarini
Efe Ok
Alessandro Pavan
James Peck
Lasse Heje Pedersen
Fabrizio Perri
Benjamin Polak
Stephen A. Ross
Hamid Sabourian
William H. Sandholm
Manuel S. Santos
Chris Shannon
Marciano Siniscalchi
Lones Smith
William L. Thomson
Juuso Välimäki
Randall Wright
Administrative Editors: Edward McDevitt and Susan Witherow Schulze
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Journal of Economic Theory 146 (2011) 1598–1630

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Illiquidity, position limits, and optimal investment for mutual funds ☆

Min Dai [a], Hanqing Jin [b], Hong Liu [c,*]

[a] *Department of Mathematics, National University of Singapore, Singapore*
[b] *Mathematical Institute and Oxford-Man Institute of Quantitative Finance, University of Oxford, UK*
[c] *Olin Business School, Washington University in St. Louis, United States*

Received 8 November 2009; final version received 18 October 2010; accepted 28 January 2011

Available online 8 March 2011

Abstract

We study the optimal trading strategy of mutual funds that face both position limits and differential illiquidity. We provide explicit characterization of the optimal trading strategy and conduct an extensive analytical and numerical analysis of the optimal trading strategy. We show that the optimal trading boundaries are increasing in both the lower and the upper position limits. We find that position limits can affect current trading strategy even when they are not currently binding and other seemingly intuitive trading strategies can be costly. We also examine the optimal choice of position limits.

JEL classification: D11; D91; G11; C61

Keywords: Illiquidity; Portfolio constraints; Position limits; Transaction costs; Mutual funds; Optimal investment

☆ We are indebted to the Associate Editor of the Journal of Economic Theory for constructive suggestions. We thank Phil Dybvig, Bill Marshall, Guofu Zhou, two anonymous referees, and the seminar participants at the 2008 Bachelier Finance Society Conference, the First Singapore Conference on Quantitative Finance, and Washington University for helpful comments. We thank Yifei Zhong for excellent research assistance. Dai is also an affiliated member of Risk Management Institute (NUS) and is partially supported by the NUS RMI grant (No. R-146-000-004/124-720/646) and Singapore MOE AcRF grant (No. R-146-000-096-112).

* Corresponding author. Fax: +1 314 935 6359.
E-mail address: liuh@wustl.edu (H. Liu).

0022-0531/$ – see front matter
doi:10.1016/j.jet.2011.03.014

1. Introduction

Mutual funds are often restricted to allocating certain percentages of fund assets to certain securities that have different degrees of illiquidity. These restrictions are often specified in a fund's prospectus and differ across investment styles and country-specific regulations. For example, a small cap fund may set a lower bound on its holdings of small cap stocks. In Switzerland, regulations require that at least two thirds of a fund's assets be invested in the relevant geographical sectors (e.g., Switzerland, Europe) or asset classes depending on the fund's category. In France, regulations prevent bond and money market funds from investing more than 10% in stocks. Mutual funds can also face significant transaction costs in trading securities in some asset classes. Wermers [33] concludes that transaction costs drag down net mutual fund returns by as much as 0.8%, about the same impact as fund expenses. Consistent with this finding, Chalmers et al. [8] conclude that annual trading costs for equity mutual funds are large and exhibit substantial cross-sectional variation, averaging 0.78% of fund assets per year and having an interquartile range of 0.59%. Karceski et al. [24] find that about 46% of all small cap mutual funds have trading costs that are higher than the annual fees investors pay. The prevalence of turnover constraints also suggests the importance of transaction costs (e.g., Clarke et al. [9]).

There is a large literature on the optimal trading strategy of a mutual fund.[1] However, most of this literature does not consider the significant trading costs faced by funds or the widespread position limits imposed upon mutual funds. As is well known, the presence of transaction costs and position limits can have a drastic impact on the optimal trading strategy and portfolio performance.[2] The coexistence of the position limits and asset illiquidity and the interactions among them can thus be important for the optimal trading strategy of a mutual fund. However, the existing literature ignores this coexistence and the interactions. Therefore, the optimal trading strategy of a fund that is subject to position limits and asset illiquidity is still unknown.

In this paper, we study the optimal investment problem of a mutual fund that faces position limits and trades a risk-free asset, a liquid stock, and an illiquid stock that is subject to proportional transaction costs. Because the implied Hamilton–Jacobi–Bellman equation is highly nonlinear and difficult to analyze, we convert the original problem into a double obstacle problem that is much easier to analyze. Using this alternative approach, we are able to characterize the value function and to provide many analytical comparative statics on the optimal trading strategy. We show that there exists a unique optimal trading strategy and the value function is smooth except on a measure zero set. The optimal trading strategy for the illiquid stock is determined by a time-varying buy boundary and a time-varying sell boundary between which no transaction occurs.

In addition, we establish some important monotonicity properties analytically for the trading boundaries, which are also useful for improving the precision and robustness of the numerical procedure. For example, both the buy boundary and the sell boundary (in terms of the fraction of assets under management (AUM) invested in the illiquid stock) are monotonically increasing in the position limits. In addition, in most cases the optimal buy (sell) boundary is monotonically decreasing (increasing) in calendar time when time to horizon is short.

We also conduct an extensive numerical analysis on optimal trading strategies. Our numerical analysis shows that in the presence of transaction costs, even for log preferences, the optimal

1 See, for example, Carpenter [7], Basak et al. [6], Cuoco and Kaniel [13].

2 See, for example, Davis and Norman [21], Cuoco [12], Cuoco and Liu [15], Balduzzi and Lynch [5], Liu and Loewenstein [28], Liu [27].

trading strategy is nonmyopic with respect to position limits, in the sense that position limits can affect current trading strategy even when they are not currently binding. Intuitively, even though the position limits are not currently binding, they will for sure bind when time to horizon becomes short enough. In anticipation of this future binding of the position limits, the fund changes its current trading boundaries due to transaction cost concerns. The correlation coefficient between the two stocks affects the efficiency of diversification and thus can significantly alter the optimal trading strategy in both the liquid stock and the illiquid stock.

One alternative intuitive trading strategy in the presence of position limits is to take the optimal trading strategy for the unconstrained case and modify it myopically, i.e., set a trading boundary to the position upper bound or lower bound if and only if the unconstrained boundaries violate the bounds. We show that this myopically modified strategy can be quite costly. The main reason for this large cost is that with the myopic strategy, the no-transaction region may be too narrow and thus an investor can incur large transaction costs. This result shows the importance of adopting the optimal trading strategy that simultaneously takes into account both position limits and transaction costs.

To partly address the endogeneity issue of the position limits, we also examine the optimal choice of the position limits by investors who have different risk preferences from the fund managers.[3] Our analysis shows that the optimal upper (lower) bound is increasing (decreasing) in transaction costs. This is because loosening the limits reduces transaction frequency and hence transaction costs. In addition, the optimal position limits can be sensitive to the level of transaction costs, which suggests that transaction cost can be a significant factor in determining the optimal position limits.

As far as we know, this is the first paper to characterize and compute the optimal trading strategies for mutual funds that face both asset illiquidity and position limits, which we view as the most important contribution of this paper.[4] Although the qualitative impact of position limits on trading strategies may be clear even without explicitly computing them, what is more important for mutual funds (or any investment institution) is to quantitatively determine the optimal trading strategies. For example, as we show in the paper, adopting some other seemingly reasonable trading strategies can severely worsen the fund performance. Given the large size of a typical fund and the prevalence of asset illiquidity and position limits for most mutual funds, even a small improvement in the trading strategy is likely to have a significant impact on fund performance. With the enormous amount of assets under management in the mutual fund industry, a better understanding of the optimal trading strategy for a typical fund is necessary for a better understanding of the impact of mutual funds on asset pricing. The analytical and numerical results on the optimal trading strategies also provide an important foundation for research on fund performance evaluation and optimal contracting.[5]

[3] There are obviously other reasons for imposing position limits, e.g., different investment horizons, asymmetric information, etc.

[4] Admittedly, our model has left out some other factors of mutual funds (e.g., tax management, fund flows) that affect fund trading to some extent and only focuses on the joint impact of transaction costs and position limits on the optimal trading strategy. However, we believe this is a significant step toward a better understanding of how mutual funds should trade.

[5] In contrast to Constantinides [10] and consistent with empirical evidence (e.g., Amihud, Mendelson, and Pedersen [2], Acharya and Pedersen [1]), we also find that, as expected, position limits can make transaction costs have a first-order effect on the liquidity premium, similar to the impact of a time-varying investment opportunity set (e.g., Jang, Koo, Liu and Loewenstein [26], Lynch and Tan [29]). Surprisingly, however, we show that the liquidity premium can be higher when position limits are *less* binding. The details on these results are not presented in the text to save space.

We also make significant contributions on solution methods for problems with transaction costs and portfolio constraints. First, as far as we know, this is the first paper that applies the techniques for double obstacle problems to solve and analyze a portfolio selection problem with both transaction costs and position limits. As our analysis indicates, this alternative method can be used to solve a large class of portfolio selection problems with transaction costs, portfolio constraints, and finite investment horizons (e.g., optimal investment problem of an investor who is subject to margin requirement and transaction costs, with or without labor income).

Second, our numerical procedure produces economically sensible and numerically stable results that satisfy all the analytical properties we establish. As shown in the vast literature on portfolio selection with transaction costs (e.g., Constantinides [10]), the computation of the optimal trading boundaries even in the absence of portfolio constraints and the time-to-horizon effect is challenging, because one needs to solve the HJB equation with two free boundaries. The simultaneous presence of portfolio constraints and the time-to-horizon effect makes it much more difficult. Most of the existing literature uses smooth pasting conditions on the trading boundaries to determine the optimal trading strategy. In contrast, we use a finite difference scheme to directly discretize the HJB equation into a system of nonlinear algebraic equations and then use a projected successive over-relaxation method to solve this system. This alternative method can readily deal with both portfolio constraints and the time-to-horizon effect.

This paper is closely related to two strands of literature: one on portfolio selection with transaction costs and the other on portfolio selection with portfolio constraints. The first strand of literature (e.g., Constantinides [10], Davis and Norman [21], Gârleanu and Pedersen [25], Lynch and Tan [30]) finds that the presence of transaction costs can dramatically change the optimal trading strategy. For example, Gârleanu and Pedersen [25] derive in closed form the optimal dynamic portfolio policy when trading is costly and security returns are predictable by signals with different mean-reversion speeds. They show that the optimal updated portfolio is a linear combination of the existing portfolio, the optimal portfolio absent trading costs, and the optimal portfolio based on future expected returns and transaction costs. The second strand of literature (e.g., Cvitanić and Karatzas [16], Cuoco [12], Cuoco and Liu [14]) shows that portfolio constraints can also have a large impact on the optimal trading strategy. However, as far as we know, this is the first paper to consider the joint impact of transaction costs and portfolio constraints on the optimal trading strategy, and our results show that this joint consideration is important both quantitatively and qualitatively.

As shown in the literature (e.g., Almazan et al. [4]), managers may choose not to adopt certain investment strategies even when they are not restricted from doing so. However, this does not necessarily imply that position limits are unimportant for mutual fund investment strategies. As Almazan et al. [4] explain, it is possible that "circumstances requiring the use of certain investment practices might not arise in a given reporting period" and "alternatively, it is possible for a portfolio manager to adopt a constraint on a purely voluntary basis." Studies on whether position limits can be (economically) significantly binding for mutual funds are scarce, possibly due to measurement difficulty and data availability. However, some existing literature and indirect evidence suggest position limits can indeed be (economically) significantly binding for mutual funds. For example, Clarke et al. [9] find that "constraints on short positions and turnover, for example, are fairly common and materially restrictive. Other constraints, such as market capitalization and value/growth neutrality with respect to the benchmark or economic sectors constraints, can further restrict an active portfolio's composition." Consistent with this finding, many prospectuses of mutual funds list investment-style risk, sector risk, industry concentration risk, and country risk as primary risks of the corresponding funds. This suggests that position limits required for concentrations on certain styles, sectors, industries, and geographical regions can significantly bind in some states and in some time periods and can significantly affect fund performance.

The remainder of the paper is organized as follows. In Section 2, we describe the model. We solve the first benchmark case without transaction costs in Section 3. We solve the second benchmark case with transaction costs but without position limits in Section 4. Section 5 provides a verification theorem and an analytical analysis of the main problem with both transaction costs and position limits. In Section 6, we conduct an extensive numerical analysis. Section 7 concludes. All the proofs are provided in Appendix A.

2. The model

We consider a fund manager who has a finite horizon $T \in (0, \infty)$ and maximizes the expectation of his constant relative risk averse (CRRA) utility from terminal wealth.[6] The fund can invest in three assets. The first asset ("the bond") is a money market account growing at a continuously compounded, constant rate r.[7] The second asset is a liquid risky asset ("the liquid stock," e.g., a large cap stock, S&P index) whose price process S_{Lt} evolves as

$$\frac{dS_{Lt}}{S_{Lt}} = \mu_L\, dt + \sigma_L\, dB_{Lt}, \tag{1}$$

where both μ_L and $\sigma_L > 0$ are constants and B_{Lt} is a one-dimensional standard Brownian motion. The third asset is an illiquid risky asset ("the illiquid stock," e.g., a small cap stock, an emerging market portfolio). The investor can buy the illiquid stock at the ask price $S^A_{It} = (1+\theta)S_{It}$ and sell the stock at the bid price $S^B_{It} = (1-\alpha)S_{It}$, where $\theta \geqslant 0$ and $0 \leqslant \alpha < 1$ represent the proportional transaction cost rates and S_{It} follows the process

$$\frac{dS_{It}}{S_{It}} = \mu_I\, dt + \sigma_I\, dB_{It}, \tag{2}$$

where μ_I and $\sigma_I > 0$ are both constants and B_{It} is another one-dimensional standard Brownian motion that has a correlation of ρ with B_{Lt} with $|\rho| < 1$.[8]

When $\alpha + \theta > 0$, the above model gives rise to equations governing the evolution of the dollar amount invested in the liquid assets (i.e., the bond and the liquid stock), x_t, and the dollar amount invested in the illiquid stock, y_t:

$$dx_t = rx_t\, dt + \xi_t(\mu_L - r)\, dt + \xi_t\sigma_L\, dB_{Lt} - (1+\theta)\, dI_t + (1-\alpha)\, dD_t, \tag{3}$$

$$dy_t = \mu_I y_t\, dt + \sigma_I y_t\, dB_{It} + dI_t - dD_t, \tag{4}$$

where stochastic process ξ denotes the dollar amount invested in the liquid stock and the processes D and I represent the cumulative dollar amount of sales and purchases of the illiquid stock, respectively. D and I are nondecreasing and right continuous adapted processes with $D_0 = I_0 = 0$.

---

6 This form of utility function is consistent with a linear fee structure predominantly adopted by mutual fund companies (e.g., Das and Sundaram [19], Elton, Gruber, and Blake [20]) and is also commonly used in the literature (e.g., Carpenter [7], Basak, Pavlova, and Shapiro [6], Cuoco and Kaniel [13]).

7 Although significant cash position is rare for most mutual funds, some funds do hold some cash. Later, we analyze the case where cash position is restricted.

8 The case with perfect correlation is straightforward to analyze, but needs a separate treatment. We thus omit it to save space.

Let $A_t = x_t + y_t > 0$ be the fund's assets under management (AUM) (on paper) at time t. The fund is subject to the following exogenously given constraints on its trading strategy[9]:

$$\underline{b} \leqslant \frac{y_t}{A_t} \leqslant \bar{b}, \quad \forall t \geqslant 0, \tag{5}$$

where $-\frac{1}{\theta} \leqslant \underline{b} < \bar{b} \leqslant \frac{1}{\alpha}$ are constants.[10] These constraints restrict the fraction of AUM (on paper) that must be invested in the illiquid asset and imply that the fund is always solvent after liquidation, i.e., the liquidation value[11]

$$\hat{A}_t \geqslant 0, \quad \forall t \geqslant 0, \tag{6}$$

where

$$\hat{A}_t = x_t + (1-\alpha)y_t^+ - (1+\theta)y_t^-. \tag{7}$$

Let x_0 and y_0 be the given initial positions in the liquid assets (the bond and the liquid stock) and the illiquid stock, respectively. We let $\Theta(x_0, y_0)$ denote the set of admissible trading strategies (ξ, D, I) such that (3), (4), and (5) are satisfied.

The fund manager's problem is then[12]

$$\sup_{(\xi,D,I)\in\Theta(x_0,y_0)} E\big[u(A_T)\big], \tag{8}$$

where the utility function is given by

$$u(A) = \frac{A^{1-\gamma} - 1}{1-\gamma}$$

and $\gamma > 0$ is the constant relative risk aversion coefficient. This specification allows us to obtain the corresponding results for the log utility case by letting γ approach 1. Implicitly, we assume that the performance evaluation or incentive fee structure depends on the AUM on paper instead of the liquidation value. This is consistent with common industry practice and avoids trading strategies that lead to liquidation on the performance evaluation or terminal date.

3. Optimal policies without transaction costs

For the purpose of comparison, let us first consider the case without transaction costs (i.e., $\alpha = \theta = 0$). In this case, the fund manager's problem at time t becomes

$$J(A,t) \equiv \sup_{\{\pi_L,\pi\}} E_t\big[u(A_T) \,\big|\, A_t = A\big], \tag{9}$$

[9] Because of a possible misalignment of interests between the fund manager and the investor (e.g., different risk tolerance, different investment horizons, different view of asset characteristics, etc.), the investor may impose constraints on the trading strategy of the fund. See Almazan et al. [4] for more details on why many mutual fund managers are constrained. In this paper, we focus on the case where these constraints are exogenously given and do not consider the optimal contracting issue. This serves as a foundation toward examining the optimal contracting problem that allows endogenous position limits in the presence of transaction costs. Later in this paper, we illustrate the choice of optimal position limits using numerical examples.

[10] Similar arguments to those in Cuoco and Liu [14] imply that the margin requirement for the one-stock case is a special case of this constraint. So our model can also be used to study the effect of margin requirement in the presence of transaction costs.

[11] Choosing the AUM on paper in (5) instead of the AUM after liquidation (as defined in (7)) as the denominator is consistent with common industry practice. Switching the choice does not affect our main results.

[12] It can be shown that as long as $\bar{b} > \underline{b}$, there exist feasible strategies and this problem is well posed. The proof of this claim is omitted to save space.

subject to the self-financing condition

$$dA_s = rA_s\,ds + \pi_{L_s}A_s(\mu_L - r)\,ds + \pi_{L_s}A_s\sigma_L\,dB_{Ls} + \pi_s A_s(\mu_I - r)\,ds \\ \quad + \pi_s A_s\sigma_I\,dB_{Is}, \quad \forall s \geqslant t, \tag{10}$$

and the portfolio constraint

$$\pi_s \in [\underline{b}, \bar{b}], \quad \forall s \geqslant t,$$

where π_L and π represent the fractions of AUM invested in the liquid stock and illiquid stock, respectively.

Let π^M ("Merton line," Merton [31]) be the optimal fraction of AUM invested in the illiquid stock in the unconstrained case in the absence of transaction costs. Then it can be shown that

$$\pi^M = \frac{1}{1-\rho^2}\left(\frac{\mu_I - r}{\gamma\sigma_I^2} - \rho\frac{\mu_L - r}{\gamma\sigma_L\sigma_I}\right). \tag{11}$$

We summarize the main result for this no-transaction-cost case in the following theorem.

Theorem 1. *Suppose that $\alpha = \theta = 0$. Then for any time $t \in [0, T]$, the optimal trading policy is given by*

$$\pi^*(t) = \pi^* \equiv \begin{cases} \bar{b} & \text{if } \pi^M \geqslant \bar{b}, \\ \pi^M & \text{if } \underline{b} < \pi^M < \bar{b}, \\ \underline{b} & \text{if } \pi^M \leqslant \underline{b} \end{cases}$$

and

$$\pi_L^*(t) = \pi_L^* \equiv \begin{cases} \frac{\mu_L - r}{\gamma\sigma_L^2} - \rho\frac{\sigma_I}{\sigma_L}\bar{b} & \text{if } \pi^M \geqslant \bar{b}, \\ \frac{1}{1-\rho^2}(\frac{\mu_L - r}{\gamma\sigma_L^2} - \rho\frac{\mu_I - r}{\gamma\sigma_L\sigma_I}) & \text{if } \underline{b} < \pi^M < \bar{b}, \\ \frac{\mu_L - r}{\gamma\sigma_L^2} - \rho\frac{\sigma_I}{\sigma_L}\underline{b} & \text{if } \pi^M \leqslant \underline{b}, \end{cases}$$

and the value function is

$$J(A, t) = \frac{(e^{\eta(T-t)}A)^{1-\gamma} - 1}{1-\gamma},$$

where

$$\eta = r - \frac{1}{2}\gamma\left(\left(\pi_L^*\sigma_L\right)^2 + 2\pi_L^*\pi^*\rho\sigma_L\sigma_I + \left(\pi^*\sigma_I\right)^2\right) + \pi_L^*(\mu_L - r) + \pi^*(\mu_I - r). \tag{12}$$

Theorem 1 implies that the optimal fractions of AUM invested in each asset are time and horizon independent. In addition, the investor is myopic with respect to the constraints even for a nonlog preference. Specifically, the optimal fraction is equal to a bound if and only if the unconstrained optimal fraction violates the bound. We will show that in the presence of transaction costs, the investor will no longer be myopic even with a log preference. In addition, the optimal trading strategy will be horizon dependent.

4. The transaction cost case without constraints

In the presence of transaction costs, i.e., $\alpha + \theta > 0$, the problem becomes considerably more complicated. In this section, we consider the unconstrained case first. In this case, the investor's problem at time t becomes

$$V(x, y, t) \equiv \sup_{(\xi, D, I) \in \Theta(x,y)} E_t\left[u(A_T) \,\middle|\, x_t = x,\ y_t = y\right] \tag{13}$$

with $\underline{b} = -\frac{1}{\theta}$ and $\bar{b} = \frac{1}{\alpha}$, which is equivalent to the solvency constraint (6). Under some regularity conditions on the value function, we have the following HJB equation:

$$\max\left(V_t + \mathscr{L}V,\ (1-\alpha)V_x - V_y,\ -(1+\theta)V_x + V_y\right) = 0,$$

with the boundary conditions

$$(1-\alpha)V_x - V_y = 0 \quad \text{on } \frac{y}{x+y} = \frac{1}{\alpha}, \qquad -(1+\theta)V_x + V_y = 0 \quad \text{on } \frac{y}{x+y} = -\frac{1}{\theta},$$

and the terminal condition

$$V(x, y, T) = \frac{(x+y)^{1-\gamma} - 1}{1-\gamma},$$

where

$$\begin{aligned}\mathscr{L}V &= \frac{1}{2}\sigma_I^2 y^2 V_{yy} + \mu_I y V_y + rxV_x + \max_{\xi}\left\{\frac{1}{2}\sigma_L^2 \xi^2 V_{xx} + (\mu_L - r)\xi V_x + \rho\sigma_I\sigma_L \xi y V_{xy}\right\} \\ &= \frac{1}{2}\sigma_I^2 y^2 V_{yy} + \mu_I y V_y + rxV_x - \frac{[(\mu_L - r)V_x + \rho\sigma_I\sigma_L y V_{xy}]^2}{2\sigma_L^2 V_{xx}},\end{aligned}$$

and the optimal ξ, denoted by ξ^*, is

$$\xi^* = -\frac{(\mu_L - r)V_x + \rho\sigma_I\sigma_L y V_{xy}}{\sigma_L^2 V_{xx}}.$$

The HJB equation implies that the solvency region for the illiquid stock

$$\mathcal{S} = \left\{(x, y):\ x + (1-\alpha)y^+ - (1+\theta)y^- > 0\right\}$$

at each point in time splits into a buy region (BR), a no-transaction region (NTR), and a sell region (SR), as in Davis and Norman [21]. Formally, we define these regions as follows:

$$\begin{aligned}SR &\equiv \left\{(x, y, t) \in \mathcal{S} \times [0, T):\ (1-\alpha)V_x - V_y = 0\right\}, \\ BR &\equiv \left\{(x, y, t) \in \mathcal{S} \times [0, T):\ (1+\theta)V_x - V_y = 0\right\},\end{aligned}$$

and

$$NTR \equiv \left\{(x, y, t) \in \mathcal{S} \times [0, T):\ (1-\alpha)V_x < V_y < (1+\theta)V_x\right\}.$$

The homogeneity of the function $u + \frac{1}{1-\gamma}$ and the fact that $\Theta(\beta x, \beta y) = \beta\Theta(x, y)$ for all $\beta > 0$ imply that $V + \frac{1}{1-\gamma}$ is concave in (x, y) and homogeneous of degree $1-\gamma$ in (x, y) [cf. Fleming and Soner [22, Lemma VIII.3.2]]. This homogeneity implies that

$$V(x, y, t) \equiv (x+y)^{1-\gamma}\varphi\left(\frac{y}{x+y}, t\right) - \frac{1}{1-\gamma}, \tag{14}$$

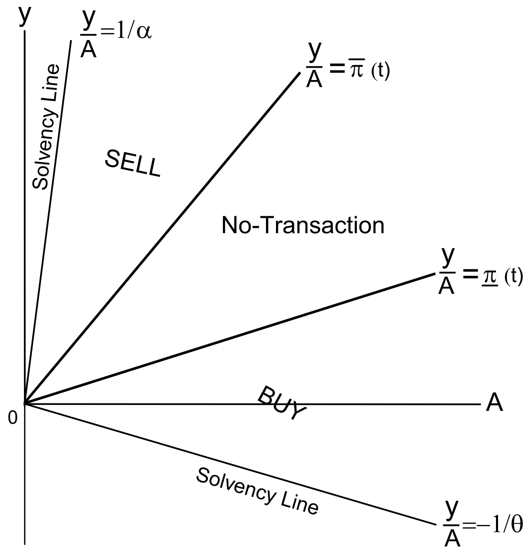

Fig. 1. The solvency region.

for some function $\varphi : (-1/\theta, 1/\alpha) \times [0, T] \to \mathbb{R}$. Let

$$\pi = \frac{y}{x+y} \tag{15}$$

denote the fraction of AUM invested in the illiquid stock. The combination of the concavity of the utility function and the homogeneity property then implies that buy, no-transaction, and sell regions can be described by two functions of time $\underline{\pi}(t)$ and $\bar{\pi}(t)$. The buy region BR corresponds to $\pi \leqslant \underline{\pi}(t)$, the sell region SR to $\pi \geqslant \bar{\pi}(t)$, and the no-transaction region NTR to $\underline{\pi}(t) < \pi < \bar{\pi}(t)$. A time snapshot of these regions is depicted in Fig. 1. Similar to Davis and Norman [21] and Liu and Loewenstein [28], the optimal trading strategy in the illiquid stock is to transact a minimum amount to keep the fraction $\pi(t)$ in the optimal no-transaction region. Therefore the determination of the optimal trading strategy in the illiquid stock reduces to the determination of the optimal no-transaction region. In contrast to the no-transaction-cost case, the optimal fractions of AUM invested in both the illiquid and the liquid stocks change stochastically, since π_t varies stochastically due to no transaction in NTR.

By (14), the HJB equation simplifies into

$$\max\big(\varphi_t + \mathcal{L}_1\varphi, -(1-\alpha\pi)\varphi_\pi - \alpha(1-\gamma)\varphi, (1+\theta\pi)\varphi_\pi - \theta(1-\gamma)\varphi\big) = 0,$$

with the terminal condition

$$\varphi(\pi, T) = \frac{1}{1-\gamma},$$

where

$$\begin{aligned}
\mathcal{L}_1\varphi &= \frac{1}{2}\beta_1\pi^2(1-\pi)^2\varphi_{\pi\pi} + (\beta_2 - \gamma\beta_1\pi)\pi(1-\pi)\varphi_\pi \\
&\quad + (1-\gamma)\left(\beta_3 + \beta_2\pi - \frac{1}{2}\gamma\beta_1\pi^2\right)\varphi - \frac{1}{2}\beta_4\frac{[(\gamma-1)\varphi + \pi\varphi_\pi]^2}{\gamma(\gamma-1)\varphi + 2\gamma\pi\varphi_\pi + \pi^2\varphi_{\pi\pi}}, \\
\beta_1 &= \big(1-\rho^2\big)\sigma_I^2, \qquad \beta_2 = \sigma_I\left(\frac{\mu_I - r}{\sigma_I} - \rho\frac{\mu_L - r}{\sigma_L}\right), \\
\beta_3 &= r - \frac{1}{2}\gamma\rho^2\sigma_I^2 + \rho\sigma_I\frac{\mu_I - r}{\sigma_L}, \qquad \beta_4 = \left(\frac{\mu_L - r}{\sigma_L} - \gamma\rho\sigma_I\right)^2.
\end{aligned} \tag{16}$$

The nonlinearity of this HJB equation and the time-varying nature of the free boundaries (i.e., the buy and sell boundaries) make it difficult to solve directly. Instead, we transform the above problem into a double obstacle problem, which is much easier to analyze. All the analytical results in this paper are obtained through this approach.

Theorem 2 in the next section shows the existence and the uniqueness of the optimal trading strategy in the case with portfolio constraints and also applies to the unconstrained case by choosing constraints that never bind. It also ensures the smoothness of the value function except for a set of measure zero.

Before we proceed further, we make the following assumption to simplify analysis.

Assumption 1. $\alpha > 0$, $\theta > 0$, and $-\frac{1}{\alpha} + 1 < \pi^M < \frac{1}{\theta} + 1$.

Assuming the transaction costs for both purchases and sales are positive reflects the common industry practice. Because α and θ are typically small (e.g., 0.05), the assumption that $-\frac{1}{\alpha} + 1 < \pi^M < \frac{1}{\theta} + 1$ is almost without loss of generality.

Let $\bar{\pi}(t)$ be the optimal sell boundary and let $\underline{\pi}(t)$ be the optimal buy boundary in the (π, t) plane. Then we have the following properties for the no-transaction boundaries in the (π, t) plane.

Proposition 1. *Let π^M be as defined in* (11). *Denote $\bar{\pi}(T^-) = \lim_{t\to T} \bar{\pi}(t)$ and $\underline{\pi}(T^-) = \lim_{t\to T} \underline{\pi}(t)$. Under Assumption* 1, *we have*

(1) *for the sell boundary, $\bar{\pi}(T^-) = \frac{1}{\alpha}$ and*

$$\bar{\pi}(t) \geqslant \frac{\pi^M}{1-\alpha(1-\pi^M)}, \quad \text{for any } t;$$

(2) *for the buy boundary, $\underline{\pi}(T^-) = -\frac{1}{\theta}$ and*

$$\underline{\pi}(t) \leqslant \frac{\pi^M}{1+\theta(1-\pi^M)}, \quad \text{for any } t.$$

This proposition shows that both the buy boundary and the sell boundary tend to the corresponding solvency line as the investment horizon goes to zero. This implies that the entire solvency region becomes the no-transaction region with a very short horizon. This is because when a horizon is short enough, it is almost impossible to recoup the transaction cost incurred from any trading. The lower and upper bounds provided in Proposition 1 are helpful in computing the optimal boundaries. Furthermore, if $\pi^M \in (0, 1)$, then the width of the NTR is bounded below by

$$\frac{(\theta+\alpha)(1-\pi^M)\pi^M}{(1-\alpha(1-\pi^M))(1+\theta(1-\pi^M))}.$$

Let

$$\bar{t}_0 = T - \frac{1}{\beta_2}\log(1-\alpha), \tag{17}$$

$$\underline{t}_0 = T - \frac{1}{\beta_2}\log(1+\theta). \tag{18}$$

Proposition 2. *Under Assumption* 1, *we have*:

(1) *If* $\pi^M < 0$, *then* $\underline{\pi}(t) < 0$ *for all* t, *and* $\bar{\pi}(t)$ *is below 0 for* $t < \bar{t}_0$ *and above* 0 *for* $t \geqslant \bar{t}_0$; *in addition,* $\bar{\pi}(t)$ *is increasing in* t *for* $t > \bar{t}_0$.
(2) *If* $\pi^M > 0$, *then* $\underline{\pi}(t)$ *is above* 0 *for* $t < \underline{t}_0$ *and below 0 for* $t \geqslant \underline{t}_0$; $\bar{\pi}(t)$ *is above* 0 *for all* t; *in addition,* $\underline{\pi}(t)$ *is decreasing in* t *for* $t > \underline{t}_0$.
(3) *If* $\pi^M = 0$, *then* $\underline{\pi}(t) < 0$ *and* $\bar{\pi}(t) > 0$ *for all* t, *and* $\underline{\pi}(t)$ *is decreasing in* t *and* $\bar{\pi}(t)$ *is increasing in* t *for* $t \in [0, T)$.

Proposition 2 shows the presence of transaction costs can make a long position optimal when a short position is optimal in the absence of transaction costs and vice versa. For example, Part (1) of Proposition 2 shows that if the time to horizon is short (i.e., $< T - \bar{t}_0$), then the sell boundary will always be positive even if it is optimal to short the illiquid asset in the absence of transaction costs. This implies that if the fund starts with a large long position in the illiquid asset, then the fund will only sell a part of its position and optimally choose to keep a long position in it. This is because trading the large long position into a short position would incur large transaction costs. Similar results also hold when it is optimal to long in the absence of transaction costs.

We conjecture that the optimal buy boundary is always decreasing in time and the optimal sell boundary is always increasing in time. Unfortunately, we can only show this when $\pi^M = 0$. However, for other cases, we are able to show this property in some scenarios. For example, Part (2) implies that the monotonicity of the buy boundary holds when $t > \underline{t}_0$.

Propositions 1 and 2 indicate that in the absence of position limits, the portfolio chosen by a fund with a short horizon can be far from the portfolio that is optimal for a long horizon investor. Thus, this large deviation can be substantially suboptimal for investors with longer horizons and therefore it may be one of the reasons for investors to impose position limits.

5. The transaction cost case with position limits

Now we examine the case with both transaction costs and position limits. In this case, the investor's problem at time t can be written as

$$V^c(x, y, t) \equiv \sup_{(\xi, D, I) \in \Theta(x,y)} E_t\big[u(A_T) \,\big|\, x_t = x,\ y_t = y\big] \tag{19}$$

with

$$-\frac{1}{\theta} \leqslant \underline{b} \leqslant \frac{y_s}{x_s + y_s} \leqslant \bar{b} \leqslant \frac{1}{\alpha},$$

for all $T \geqslant s \geqslant t$.

Under regularity conditions on the value function, we have the HJB equation

$$\max\big(V_t^c + \mathscr{L} V^c, (1-\alpha)V_x^c - V_y^c, -(1+\theta)V_x^c + V_y^c\big) = 0, \tag{20}$$

with the boundary conditions

$$(1-\alpha)V_x^c - V_y^c = 0 \quad \text{on } \frac{y}{x+y} = \bar{b}, \qquad (1+\theta)V_x^c - V_y^c = 0 \quad \text{on } \frac{y}{x+y} = \underline{b},$$

and the terminal condition

$$V^c(x, y, T) = \frac{(x+y)^{1-\gamma} - 1}{1-\gamma}.$$

The following verification theorem shows the existence and the uniqueness of the optimal trading strategy. It also ensures the smoothness of the value function except for a set of measure zero.

Theorem 2. *Under Assumption* 1*, we have the following conclusions.*

(i) *The HJB equation* (20) *admits a unique viscosity solution, and the value function is the viscosity solution.*
(ii) *The value function is* $C^{2,2,1}$ *in* $\{(x,y,t)\colon\ x+(1-\alpha)y^+-(1+\theta)y^->0,\ \underline{b}<y/(x+y)<\bar{b},\ 0\leqslant t<T\}\setminus(\{y=0\}\cup\{x=0\})$.

Similar to $\bar{\pi}(t)$ and $\underline{\pi}(t)$, let $\bar{\pi}^c(t;\underline{b},\bar{b})$ and $\underline{\pi}^c(t;\underline{b},\bar{b})$ be respectively the optimal sell and buy boundaries in the (π,t) plane in the presence of position limits.

We have the following proposition on the properties of the optimal no-transaction boundaries in the (π,t) plane.

Proposition 3. *Under Assumption* 1*, we have*

(1) *for the sell boundary,* $\bar{\pi}^c(T^-;\underline{b},\bar{b})=\bar{b}$ *and*

$$\bar{\pi}^c(t;\underline{b},\bar{b})\geqslant\max\left(\min\left(\frac{\pi^M}{1-\alpha(1-\pi^M)},\bar{b}\right),\underline{b}\right)\quad\textit{for any } t;\tag{21}$$

(2) *for the buy boundary,* $\underline{\pi}^c(T^-;\underline{b},\bar{b})=\underline{b}$ *and*

$$\underline{\pi}^c(t;\underline{b},\bar{b})\leqslant\min\left(\max\left(\frac{\pi^M}{1+\theta(1-\pi^M)},\underline{b}\right),\bar{b}\right)\quad\textit{for any } t.\tag{22}$$

Corollary 1. *Under Assumption* 1*,*

(1) *if* $\frac{\pi^M}{1-\alpha(1-\pi^M)}\geqslant\bar{b}$*, then* $\bar{\pi}^c(t;\underline{b},\bar{b})=\bar{b}$ *for all* $t\in[0,T]$;
(2) *if* $\frac{\pi^M}{1+\theta(1-\pi^M)}\leqslant\underline{b}$*, then* $\underline{\pi}^c(t;\underline{b},\bar{b})=\underline{b}$ *for all* $t\in[0,T]$.

As stated in Corollary 1, these results imply that if the selling cost adjusted Merton line is higher than the upper bound, then the sell boundary becomes flat throughout the horizon and if the buying cost adjusted Merton line is lower than the lower bound $\underline{b}$, then the buy boundary becomes flat throughout the horizon. Proposition 3 also shows that the buy (sell) boundary converges to the lower bound $\underline{b}$ (upper bound $\bar{b}$) as the remaining horizon approaches 0 irrespective of the level of the Merton line.

Proposition 4. *Under Assumption* 1*, we have*

(1) *both* $\bar{\pi}^c(t;\underline{b},\bar{b})$ *and* $\underline{\pi}^c(t;\underline{b},\bar{b})$ *are increasing in* $\underline{b}$ *and* $\bar{b}$ *for all* $t\in[0,T]$;
(2) *if* $\bar{b}>0$*, then the upper bound does not affect any trading boundary that is below* 0;
(3) *if* $\underline{b}<0$*, then the lower bound does not affect any trading boundary that is above* 0.

Part (1) of Proposition 4 suggests that both the sell boundary and the buy boundary at any point in time shift upward as the lower bound or the upper bound increases. The intuition is straightforward. For example, if a binding lower bound is raised, then obviously one needs to increase the buy boundary to satisfy the more stringent constraint. If the sell boundary remained the same, then the no-transaction region would become narrower and the trading frequency would increase. Therefore, the sell boundary also shifts upward to save transaction costs from too frequent trading.

Parts (2) and (3) of Proposition 4 suggest that the optimal boundaries in either of the two regions $\{\pi < 0\}$ and $\{\pi > 0\}$ are not affected by a constraint that lies in the other region. Intuitively, this is because in NTR, the position in the illiquid asset can never become negative if it is positive at time 0, i.e., the fraction of AUM invested in the illiquid asset cannot cross the $\pi = 0$ line.

Proposition 5. *Under Assumption* 1 *except that either* $\theta = 0$ *or* $\alpha = 0$*, we have*

(1) (21) *and* (22) *remain valid*;
(2) *if* $\theta = 0$ *and* $\alpha > 0$*, then we have*
  (a) *both* $\underline{\pi}^c(t; \underline{b}, \bar{b})$ *and* $\bar{\pi}^c(t; \underline{b}, \bar{b})$ *are increasing in* t;
  (b) $\underline{\pi}^c(T^-; \underline{b}, \bar{b}) = \min(\max(\pi^M, \underline{b}), \bar{b})$ *and* $\bar{\pi}^c(T^-; \underline{b}, \bar{b}) = \bar{b}$;
(3) *if* $\alpha = 0$ *and* $\theta > 0$*, then we have*
  (a) *both* $\underline{\pi}^c(t; \underline{b}, \bar{b})$ *and* $\bar{\pi}^c(t; \underline{b}, \bar{b})$ *are decreasing in* t;
  (b) $\underline{\pi}^c(T^-; \underline{b}, \bar{b}) = \underline{b}$ *and* $\bar{\pi}^c(T^-; \underline{b}, \bar{b}) = \max(\min(\pi^M, \bar{b}), \underline{b})$.

Proposition 5 shows that if the buying cost is zero, then both trading boundaries are monotonically increasing in time. In contrast, if the selling cost is zero, then both trading boundaries are monotonically decreasing in time. Intuitively, if both the buying and selling costs are zero, then it is optimal to stay on the Merton line in the absence of constraints. If the selling cost is positive (and the buying cost is zero), then before maturity there is a chance that the investor needs to sell the stock and thus incurs the selling cost before maturity; therefore, the investor buys less, and thus the buy boundary stays below the Merton line. As the time to horizon shrinks, the probability of future sale decreases (note that there is no liquidation at maturity), thus the investor is willing to buy more and the buy boundary converges to the Merton line. The opposite is true for the sell boundary, because as time to horizon decreases, the benefit of selling (i.e., keeping the optimal risk exposure) decreases to zero and the cost of selling is always positive. Therefore, the investor sells less as the remaining horizon shortens and the sell boundary increases. The intuitive for the case with zero selling cost is similar.

Proposition 5 also suggests that the slopes of the trading boundaries with respect to time can be discontinuous in the transaction costs. For example, Part (1) of Proposition 2 shows that if both selling and buying costs are strictly positive, the sell boundary is increasing in time for $t > \bar{t}_0$. In contrast, Proposition 5 shows that the sell boundary is always decreasing in time if the selling cost is zero. So any positive increase in the selling cost can change the slope of the sell boundary from positive to negative and thus has a discontinuous impact.

Propositions 1–5 provide analytical results that help better understand the properties of the optimal trading strategy. Equally important, these results also serve as effective checks on the validity of numerical results. As shown later, our numerical results all satisfy these analytical properties and thus indicate the reliability of our numerical method.

5.1. An additional no-borrowing constraint

Because some mutual funds are restricted from borrowing, we next extend our analysis to impose an additional no-borrowing constraint. In this case, the investor's problem at time t can be written as

$$V^c(x, y, t) \equiv \sup_{(\xi, D, I)\in\Theta(x,y)} E_t\big[u(A_T) \,\big|\, x_t = x,\ y_t = y\big] \tag{23}$$

with the constraint on the illiquid asset

$$-\frac{1}{\theta} \leqslant \underline{b} \leqslant \frac{y_s}{x_s + y_s} \leqslant \bar{b} \leqslant \frac{1}{\alpha}, \quad \forall T \geqslant s \geqslant t,$$

and the no-borrowing constraint

$$x_s - \xi_s \geqslant 0, \quad \forall T \geqslant s \geqslant t.$$

Let $\bar{\pi}^c(t; \underline{b}, \bar{b})$ be the associated optimal sell boundary and let $\underline{\pi}^c(t; \underline{b}, \bar{b})$ be the optimal buy boundary in the (π, t) plane. Then under some technical conditions (specified in the proof), the following proposition holds and most of our analytical results in the previous section remain valid with the additional constraint.

Proposition 6. *. Let*

$$\pi_0^M = \frac{\mu_I - \mu_L + \gamma\sigma_L(\sigma_L - \rho\sigma_I)}{\gamma(\sigma_I^2 + \sigma_L^2 - 2\rho\sigma_I\sigma_L)}$$

be the Merton line for the market with only the two risky assets such that $-\frac{1}{\alpha} + 1 < \pi_0^M < \frac{1}{\theta} + 1$ *and define*

$$\Pi(\delta) = \big\{\pi\colon \big[(\mu_L - r - \gamma\rho\sigma_I\sigma_L)\delta + \gamma\big(\sigma_L^2 - \rho\sigma_I\sigma_L\big)\big]\pi < -\big(\mu_L - r - \gamma\sigma_L^2\big)\big\}.$$

Under Assumption 1*, we have*

(1) $\bar{\pi}^c(T^-; \underline{b}, \bar{b}) = \bar{b}$*, and*

$$\bar{\pi}^c(t; \underline{b}, \bar{b}) \geqslant \max\left(\min\left(\frac{\pi^M}{1 - \alpha(1 - \pi^M)}, \bar{b}\right), \underline{b}\right) \quad \text{if } \bar{\pi}^c(t; \underline{b}, \bar{b}) \in \Pi(-\alpha),$$

$$\bar{\pi}^c(t; \underline{b}, \bar{b}) \geqslant \max\left(\min\left(\frac{\pi_0^M}{1 - \alpha(1 - \pi_0^M)}, \bar{b}\right), \underline{b}\right) \quad \text{if } \bar{\pi}^c(t; \underline{b}, \bar{b}) \notin \Pi(-\alpha).$$

$\underline{\pi}^c(T^-; \underline{b}, \bar{b}) = \underline{b}$*, and*

$$\underline{\pi}^c(t; \underline{b}, \bar{b}) \leqslant \max\left(\min\left(\frac{\pi^M}{1 + \theta(1 - \pi^M)}, \bar{b}\right), \underline{b}\right) \quad \text{if } \underline{\pi}^c(t; \underline{b}, \bar{b}) \in \Pi(\theta), \tag{24}$$

$$\underline{\pi}^c(t; \underline{b}, \bar{b}) \leqslant \max\left(\min\left(\frac{\pi_0^M}{1 + \theta(1 - \pi_0^M)}, \bar{b}\right), \underline{b}\right) \quad \text{if } \underline{\pi}^c(t; \underline{b}, \bar{b}) \notin \Pi(\theta). \tag{25}$$

(2) *Both* $\underline{\pi}^c(t; \underline{b}, \bar{b})$ *and* $\underline{\pi}^c(t; \underline{b}, \bar{b})$ *are increasing in* $\underline{b}$ *and* $\bar{b}$ *for all* $t \in [0, T]$*; if* $\bar{b} > 0$*, then the upper bound does not affect any trading boundary that is below* 0*; if* $\underline{b} < 0$*, then the lower bound does not affect any trading boundary that is above* 0*.*

(3) *Suppose either $\theta = 0$ or $\alpha = 0$, then* (i) *Part* (1) *remains valid*; (ii) *if $\theta = 0$ and $\alpha > 0$, then both $\underline{\pi}^c(t; \underline{b}, \bar{b})$ and $\bar{\pi}^c(t; \underline{b}, \bar{b})$ are increasing in t, $\bar{\pi}^c(T^-; \underline{b}, \bar{b}) = \bar{b}$,*

$$\underline{\pi}^c\left(T^-; \underline{b}, \bar{b}\right) = \begin{cases} \min(\max(\pi^M, \underline{b}), \bar{b}) & \text{if } \underline{\pi}^c(T^-; \underline{b}, \bar{b}) \in \Pi(0), \\ \min(\max(\pi_0^M, \underline{b}), \bar{b}) & \text{if } \underline{\pi}^c(T^-; \underline{b}, \bar{b}) \notin \Pi(0). \end{cases}$$

(iii) *if $\alpha = 0$ and $\theta > 0$, then both $\underline{\pi}^c(t; \underline{b}, \bar{b})$ and $\bar{\pi}^c(t; \underline{b}, \bar{b})$ are decreasing in t, $\underline{\pi}^c(T^-; \underline{b}, \bar{b}) = \underline{b}$, and*

$$\bar{\pi}^c\left(T^-; \underline{b}, \bar{b}\right) = \begin{cases} \min(\max(\pi^M, \underline{b}), \bar{b}) & \text{if } \bar{\pi}^c(T^-; \underline{b}, \bar{b}) \in \Pi(0), \\ \min(\max(\pi_0^M, \underline{b}), \bar{b}) & \text{if } \bar{\pi}^c(T^-; \underline{b}, \bar{b}) \notin \Pi(0). \end{cases}$$

6. Numerical analysis

In this section, we conduct a numerical analysis of the optimal trading strategy and the cost from following some seemingly intuitive, but suboptimal strategies. We also examine the optimal choice of position limits. For this analysis, we use the following default parameter values: $\gamma = 2$, $T = 5$, $\mu_L = 0.05$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $r = 0.01$, $\rho = 0.3$, $\alpha = 0.01$, $\theta = 0.01$, $\underline{b} = 0.60$, and $\bar{b} = 0.80$, which implies that in the default case the fraction of AUM invested in the illiquid stock is greater than that in the liquid assets, like in a small cap fund. For a large cap fund, we set $\underline{b} = 0.10$ and $\bar{b} = 0.30$ so that the fraction of AUM invested in the liquid (large cap) stock is greater than that in the illiquid (small cap) stock. These parameter values imply that the illiquid stock has higher expected return and volatility than the liquid stock. We also impose the no-borrowing constraint in our numerical examples. We use a finite difference scheme to discretize the HJB equations and then use a projected successive over-relaxation method (and a penalty method in some cases) to solve the resulting nonlinear algebraic system (e.g., Wilmott, Dewynne, and Howison [34] and Dai, Kwok, and Zong [17]).

6.1. Horizon effect

In Fig. 2, we plot π against calendar time t for the constrained case (the solid lines) and the unconstrained case (the dotted lines). The dashed line represents the Merton line in the absence of transaction costs. Consistent with the theoretical results in the previous section, this figure shows that the buy boundary is monotonically decreasing in time and the sell boundary is monotonically (weakly) increasing in time, with or without the position limits. The upper bound of 80% is binding throughout the investment horizon and therefore the sell boundary becomes flat at 80% across all time. In Fig. 2, the no-borrowing constraint is never binding. The lowest riskless asset position is about 0.5% of the AUM, with an average of 4.7% across the investment horizon, consistent with empirical evidence.

The buy boundary reaches the lower bound of 60% at $t = 4.6$. In addition, compared to the unconstrained case, the buy boundary before $t = 3.7$ is moved lower and the portion after $t = 3.7$ is moved higher. Thus, the optimal trading strategy is not myopic in the sense that in anticipation of the constraint becoming binding later, it is optimal to change the early trading strategy. To understand this result, recall that by Proposition 3, as time to horizon decreases to 0, the buy boundary decreases to $-1/\theta$ and the sell boundary increases to $1/\alpha$. A lower bound $\underline{b} < 1$ will then for sure bind if time to horizon is short. For a fund with a long time to horizon, it will therefore change its optimal trading boundaries in anticipation of the fact that when its remaining investment horizon gets short enough, it will be forced to buy the illiquid asset and incur transaction costs. In this sense, the fund's trading strategy is nonmyopic with respect to the portfolio

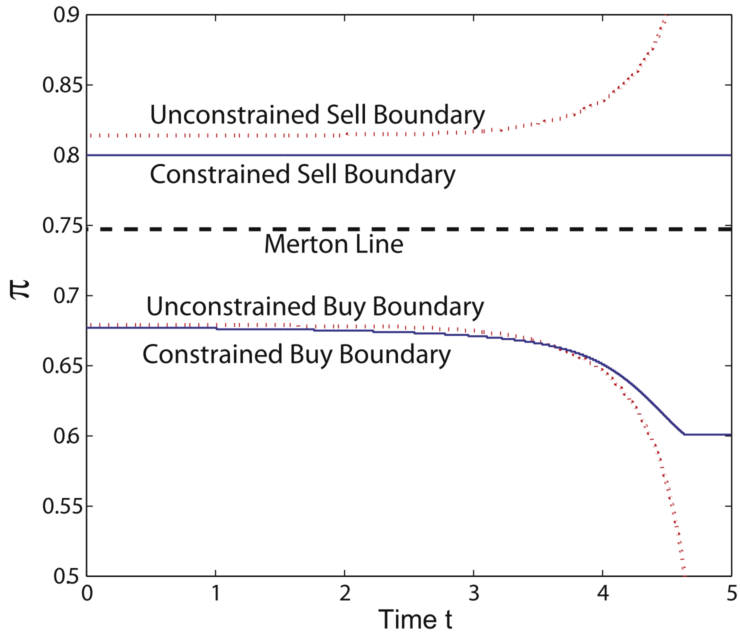

Fig. 2. The optimal trading strategy for the illiquid asset for a small cap fund against time. Parameter default values: $\gamma = 2$, $T = 5$, $\mu_L = 0.05$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $r = 0.01$, $\rho = 0.3$, $\alpha = 0.01$, $\theta = 0.01$, $\underline{b} = 0.60$, and $\bar{b} = 0.80$.

constraints in the presence of transaction costs because what will happen in the future affects the current trading behavior. Since the results in this proposition hold for any risk aversion, it also holds for a log utility (a special case with $\gamma = 1$). Therefore, the optimal trading strategy is nonmyopic even for log preferences by the same intuition. This nonmyopium of the optimal trading strategy with respect to the portfolio constraints is robust and present in all the cases we have numerically solved.

The Merton line is flat through time, implying that in the absence of transaction costs, it is optimal to keep a constant fraction of AUM in the stock. In the presence of transaction costs, however, the optimal fraction becomes a stochastic process because the investor cannot trade continuously to keep the fraction constant.

We present a similar figure (Fig. 3) for the large cap fund case with the expected return for the large cap (liquid) stock changed to 9%. In this case, both the lower bound (10%) and the upper bound (30%) are tight constraints and the upper bound is so restrictive that the sell boundary becomes flat at 30% throughout the horizon. In contrast to the case depicted in Fig. 2, the no-borrowing constraint is always binding and the investor does not invest in the riskless asset at all, because both stocks provide a better risk and return trade-off. This implies that the large cap stock position varies from 70% to 90% of the AUM. The buy boundary also shifts downward significantly through most of the horizon and only shifts upward toward the end of the horizon. In contrast to Fig. 2, the Merton line is outside the optimal no-transaction region for the constrained case. These parameter values for the constraints can be reasonable for investors who are more risk averse than the fund manager.

6.2. *Change in illiquidity*

In Fig. 4, we plot the time 0 optimal boundaries ($\pi(0)$) against the transaction cost rate α for several different cases. In the unconstrained case, as the transaction cost rate increases, the buy boundary decreases and the sell boundary increases and thus the no-transaction region widens

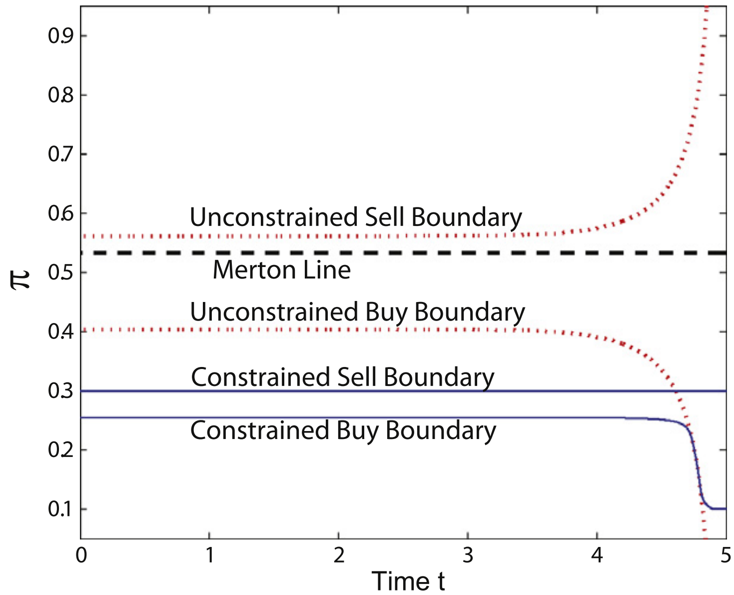

Fig. 3. The optimal trading strategy for the illiquid asset for a large cap fund against time. Parameter default values: $\gamma = 2$, $T = 5$, $\mu_L = 0.09$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $r = 0.01$, $\rho = 0.3$, $\alpha = 0.01$, $\theta = 0.01$, $\underline{b} = 0.10$, and $\bar{b} = 0.30$.

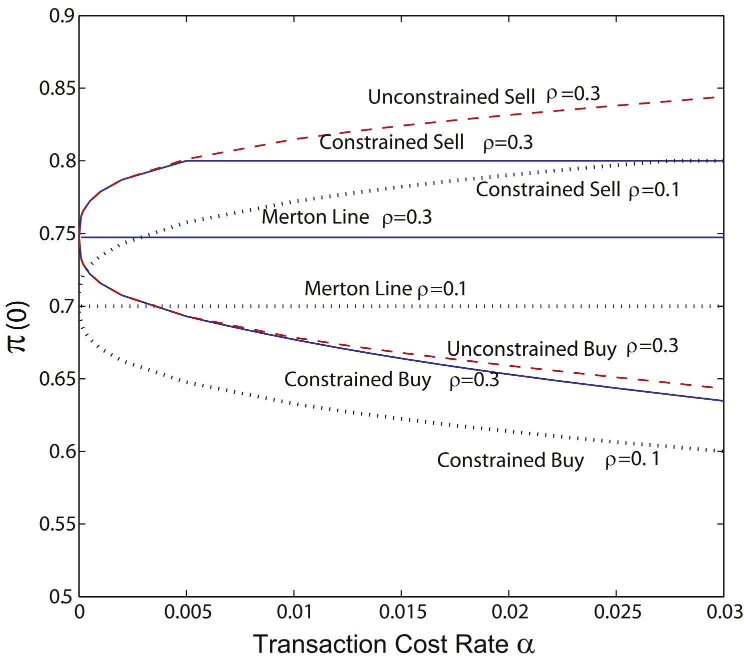

Fig. 4. The initial optimal trading strategy for the illiquid asset for a small cap fund against transaction cost rate. Parameter default values: $\gamma = 2$, $T = 5$, $\mu_L = 0.05$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $r = 0.01$, $\rho = 0.3$, $\theta = \alpha$, $\underline{b} = 0.60$, and $\bar{b} = 0.80$.

to decrease transaction frequency. In contrast, the sell boundary in the presence of constraints first increases and then stays at the upper bound because the upper bound becomes binding. The binding upper bound also drives down the buy boundary and makes it move down more for higher transaction cost rates.

This figure also shows that as the correlation between the liquid and illiquid stock returns decreases, the fraction of AUM invested in the illiquid stock decreases in the absence of transaction costs. This is because the diversification benefit of investing in the large cap stock increases and

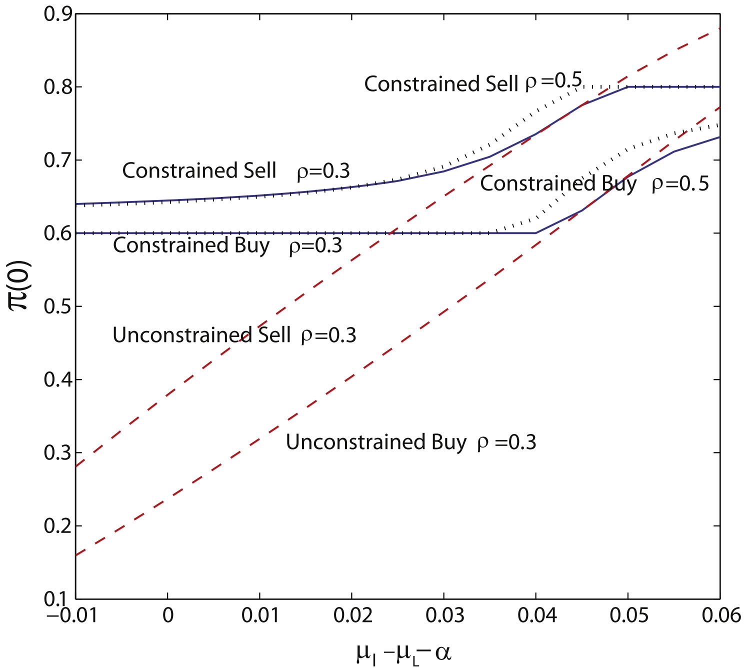

Fig. 5. The initial optimal trading strategy for the illiquid asset for a small cap fund against net excess return over the liquid asset. Parameter default values: $\gamma = 2$, $T = 5$, $\mu_L = 0.05$, $\sigma_L = 0.20$, $\sigma_I = 0.25$, $r = 0.01$, $\rho = 0.3$, $\alpha = 0.01$, $\theta = 0.01$, $\underline{b} = 0.60$, and $\bar{b} = 0.80$.

thus one should invest less in the small cap stock that has a higher risk. In the presence of transaction costs, a decrease of the correlation drives both the sell boundary and the buy boundary downward.

6.3. *Expected return of the illiquid stock*

In Fig. 5, we plot the time 0 optimal boundaries ($\pi(0)$) against the excess return $R_I \equiv \mu_I - \mu_L - \alpha$ (a measure of the excess return over the liquid stock net of illiquidity), varying the expected return of the illiquid stock μ_I. Even when the excess return is negative, it is still optimal to invest in the illiquid asset due to its diversification benefit. The lower bound is binding for low excess returns. This binding constraint makes the buy boundary flat at 60% until it gets close to the buy boundary for the unconstrained case at $R_I = 4.1\%$. It also makes the sell boundary significantly higher than the unconstrained case to balance the cost from over-investment in the illiquid asset and the transaction cost payment.

As the excess return increases, the no-transaction region widens because the cost of over-investment decreases. Between $R_I = 4.1\%$ and $R_I = 4.8\%$, the constraints become less binding and thus the constrained boundaries are close to the unconstrained boundaries. Above $R_I = 4.8\%$, the upper bound becomes binding, which makes the sell boundary flat at 80% for $R_I > 4.8\%$. To reduce transaction costs, the buy boundary is adjusted downward to reduce transaction frequency. An increase in the correlation drives down the optimal boundaries if the excess return is low and drives them up if the excess return is high. Intuitively, if the correlation gets larger, the diversification benefit shrinks and so the fund will shift funds into the asset with a more attractive Sharpe ratio. Therefore, if the excess return is high then the fund will shift into the illiquid asset and vice versa.

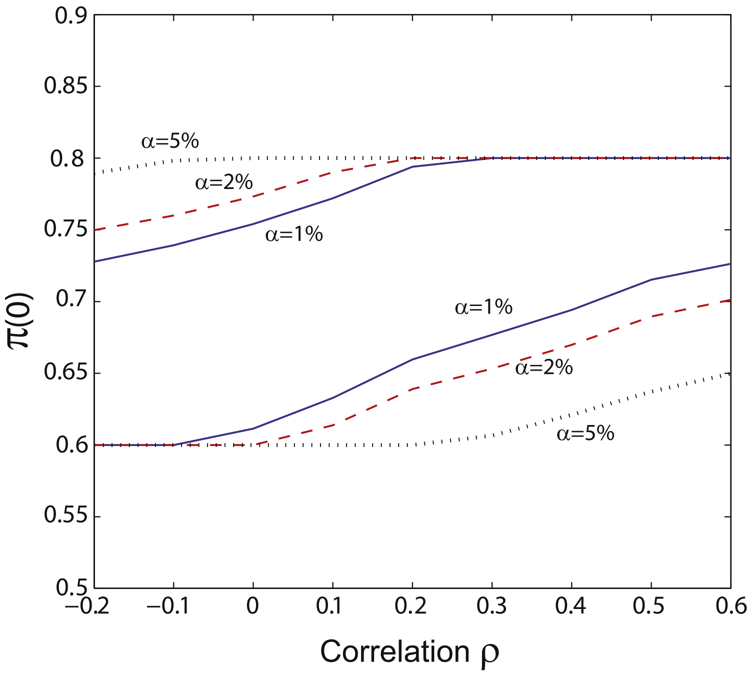

Fig. 6. The initial optimal trading strategy for the illiquid asset for a small cap fund against correlation coefficient. Parameter default values: $\gamma = 2$, $T = 5$, $\mu_L = 0.05$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $r = 0.01$, $\theta = \alpha$, $\underline{b} = 0.60$, and $\bar{b} = 0.80$.

6.4. *Correlation and diversification*

Next we examine more closely the effect of correlation on diversification. In Fig. 6, we plot the time 0 optimal fraction of AUM invested in the illiquid asset ($\pi(0)$) against the correlation coefficient ρ for different levels of transaction cost rates. Consistent with Fig. 4, Fig. 6 verifies that for this set of parameter values such that the Sharpe ratio of the illiquid stock is higher, as the correlation coefficient increases the optimal fraction of AUM invested in the illiquid asset increases, because of the decrease in the diversification effect of the liquid stock investment. In addition, as the transaction cost rate increases, the no-transaction region widens and both the upper bound and the lower bound bind for a larger range of correlation coefficients. For example, the buy boundary is flat at 60% only for $\rho < -0.1$ with $\alpha = \theta = 0.01$. In contrast, if $\alpha = \theta = 0.02$, it remains flat at 60% for all $\rho < 0.01$. The intuition behind this result is that an increase in transaction costs makes the fund lower the buy boundary and increase the sell boundary to reduce transaction cost payment.

6.5. *The cost of a myopic trading strategy*

One intuitive trading strategy in the presence of position limits is to take the optimal trading strategy for the unconstrained case and modify it myopically, i.e., set a trading boundary to the position upper bound or lower bound if and only if the bound is binding. We show that this myopically modified strategy can be costly and thus it is important for an institution to adopt the optimal trading strategy. First, if the unconstrained buy (lower) boundary is greater than the position upper bound, then both the unconstrained sell and the unconstrained buy boundaries violate the position upper bound and thus the myopic strategy would set both the sell and the buy boundaries at the position upper bound, which implies infinite transaction costs because of the implied continuous trading. A similar result obtains if the unconstrained sell (upper) boundary

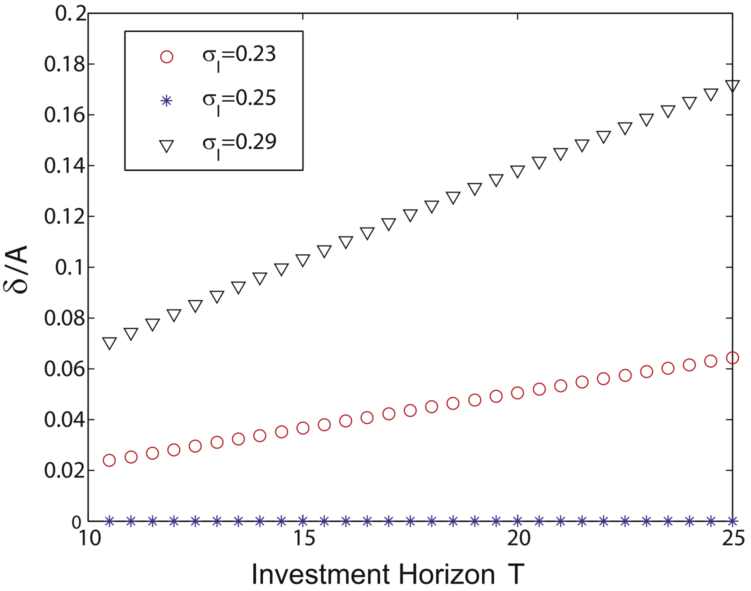

Fig. 7. The fraction of the certainty equivalent AUM loss from the myopic strategy for a small cap fund. Parameter default values: $\gamma = 2$, $\mu_L = 0.05$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $r = 0.01$, $\rho = 0.3$, $\alpha = 0.01$, $\theta = 0.01$, $\underline{b} = 0.60$, and $\bar{b} = 0.80$.

is smaller than the position lower bound. Next, we show that even for many other cases, this myopic strategy can also be costly. We use the certainty equivalent AUM loss δ from the myopic strategy to measure the cost. Fig. 7 plots the ratio of δ to the initial AUM for a small cap fund against investment horizon. Fig. 7 shows that following the myopic strategy can be very costly. For example, if $\sigma_I = 29\%$, then the certainty equivalent AUM loss is as high as 10% of the initial AUM for a 15-year horizon. As the horizon increases to 25 years, the cost increases to 17%. The main reason for this large cost is that with the myopic strategy, the no-trading region is too narrow and thus the investor incurs large transaction costs. Interestingly, Fig. 7 also shows that the cost can be nonmonotonic in the illiquid asset volatility. In particular, the cost when $\sigma_I = 25\%$ is lower than when $\sigma_I = 29\%$ and when $\sigma_I = 23\%$. The reason for this nonmonotonicity is that when $\sigma_I = 29\%$ and when $\sigma_I = 23\%$, the implied myopically modified no-trading regions are narrower than when $\sigma_I = 25\%$. More specifically, when $\sigma_I = 29\%$, only the lower bound is binding and it is close to the unconstrained sell boundary. Similarly, when $\sigma_I = 23\%$, only the upper bound is binding and it is close to the unconstrained buy boundary.

6.6. *Endogenous position limits*

Now we briefly illustrate the optimal choice of the optimal position limits by investors who hire fund managers. There are many possible reasons why investors might constrain their managers, e.g., different preferences, different investment horizons, asymmetric information, moral hazard, etc. In the subsequent analysis, for illustration purposes, we focus on the case where the only difference between investors and managers is risk aversion. Specifically, suppose the investor has the same type of utility function (i.e., CRRA), but with a different risk aversion coefficient (γ_I) from that of the fund manager (γ_M). To compute the optimal constraints, we follow the following steps:

(1) first, for given position limits $\underline{b}$ and $\bar{b}$, compute the optimal strategy of the fund manager for the constrained case and the unconstrained case;

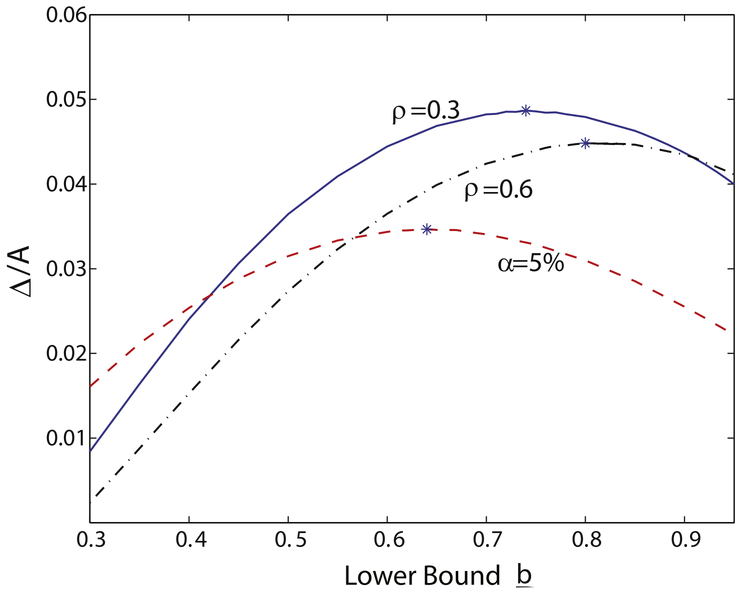

Fig. 8. The optimal choice of the lower bound for an investor. Parameter default values: $\gamma_I = 2$, $\gamma_M = 5$, $T = 5$, $\mu_L = 0.05$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $r = 0.01$, $\rho = 0.3$, $\alpha = 0.01$, $\theta = 0.01$, and $\bar{b} = 0.80$.

(2) then compute the value functions of the investor given the optimal trading strategy of the fund manager for these two cases, denoting the value functions as $V_c(A; \underline{b}, \bar{b})$ and $V_u(A)$, respectively;

(3) then solve $V_c(A - \Delta; \underline{b}, \bar{b}) = V_u(A)$ for Δ to compute the equivalent AUM gain of the investors from imposing the constraints as a measure of the value of constraints. Because of the homogeneity, the ratio Δ/A is independent of A;

(4) Now repeat steps (1)–(3) for different $\underline{b}$ and $\bar{b}$ to find the optimal $\underline{b}$ and $\bar{b}$ that maximize the equivalent AUM gain.

We illustrate the optimal choice through two cases: one case where the investor is less risk averse than the manager (Fig. 8) and the other case where the investor is more risk averse (Fig. 9). Specifically, we set $\gamma_I = 2$ and $\gamma_M = 5$ in Fig. 8 and $\gamma_I = 5$ and $\gamma_M = 2$ in Fig. 9. This implies that in Fig. 8 (Fig. 9) the investor would like the manager to invest more (less) in the illiquid stock than what the manager would choose to. Accordingly, we only consider the imposition of a lower (upper) bound in Fig. 8 (Fig. 9). Loosely speaking, the investor chooses the limits so that the fund portfolio is close to his own optimal portfolio "on average." Fig. 8 plots the ratio Δ/A against $\underline{b}$ and Fig. 9 plots the ratio Δ/A against $\bar{b}$ for different correlation coefficients and transaction cost rates, where the stars in the figures indicate where the ratios are maximized. Fig. 8 shows that the optimal lower bound $\underline{b}$ is equal to 0.73 in the first case (given default parameter values) and Fig. 9 shows that the optimal upper bound $\bar{b}$ is equal to 0.31 in the second case. These figures also show that as transaction cost rate increases, the optimal choice of the lower bound decreases and the optimal upper bound increases. Intuitively, as transaction cost rate increases, the illiquid stock becomes more costly to trade and thus the investor imposes looser constraints.

As correlation increases, the diversification benefit of investing in the liquid asset decreases and therefore it is optimal to increase the investment in the illiquid stock, which has a higher expected return. Accordingly, as these figures suggest, as the correlation increases, both the optimal upper bound and the optimal lower bound increase.

These figures also demonstrate that the benefit of constraining fund managers can be quite significant. Fig. 8 indicates that the investor is willing to pay more than 4.8% of the initial AUM

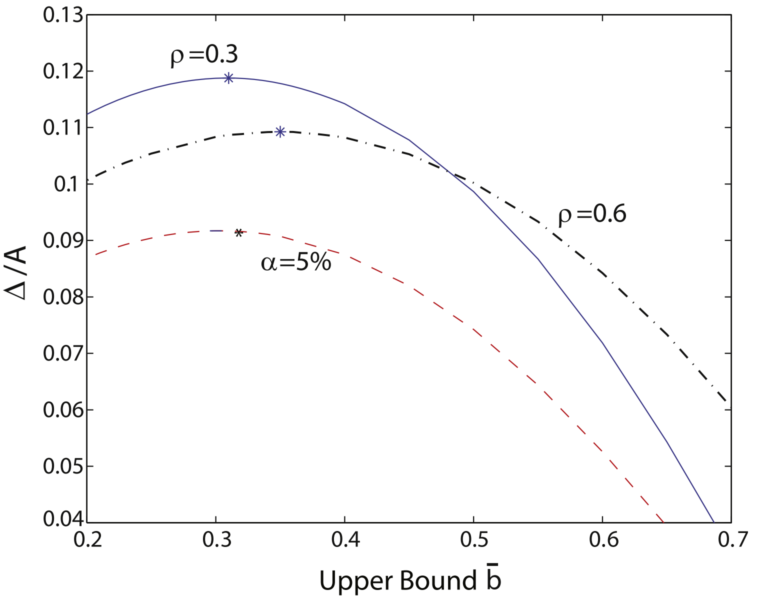

Fig. 9. The optimal choice of the upper bound for an investor. Parameter default values: $\gamma_I = 5$, $\gamma_M = 2$, $T = 5$, $\mu_L = 0.05$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $r = 0.01$, $\rho = 0.3$, $\alpha = 0.01$, $\theta = 0.01$, and $\underline{b} = 0$.

for the right to constrain fund managers. In Fig. 9, the gain from imposing portfolio constraints is as high as 11.7%.

7. Conclusions

Mutual funds are often restricted to allocate certain percentages of fund assets to certain securities that have different degrees of illiquidity. However, the existing literature has largely ignored the coexistence of position limits and differential illiquidity. Therefore, the optimal trading strategy for a typical mutual fund is still largely unknown. This paper is the first to derive and analyze the optimal trading strategy of mutual funds that face both asset illiquidity and position limits. We conduct an extensive analytical and numerical analysis of the optimal trading strategy and provide a fast numerical procedure for solving a large class of similar problems. In addition, we show that adopting other seemingly intuitive strategies can be very costly. We also examine the endogenous choice of position limits, which is a first step toward understanding why it might be optimal for investors to impose position limits for mutual funds and how transaction costs and return correlations affect the optimal position limits.

Appendix A

In this appendix, we present proofs for the propositions and theorems in this paper.

A.1. *Proof of Theorem 1*

Define

$$f(\pi_L, \pi) = r + (\mu_L - r)\pi_L + (\mu_I - r)\pi - \frac{\gamma}{2}\left[\sigma_L^2\pi_L^2 + 2\rho\sigma_L\sigma_I\pi_L\pi + \sigma_I^2\pi^2\right]. \tag{A.1}$$

The following lemma gives the explicit form for the solutions in Theorem 1.

Lemma A.1. *The solution for*

$$\max_{\pi_L \in \mathbb{R},\, \pi \in [\underline{b}, \bar{b}]} f(\pi_L, \pi) \tag{A.2}$$

is

$$\left(\pi_L^*, \pi^*\right) = \begin{cases} \left(\frac{\mu_L - r}{\gamma\sigma_L^2} - \frac{\rho\sigma_I \underline{b}}{\sigma_L}, \underline{b}\right) & \text{if } \pi^M < \underline{b}, \\ (\pi_L^M, \pi^M) & \text{if } \pi^M \in [\underline{b}, \bar{b}], \\ \left(\frac{\mu_L - r}{\gamma\sigma_L^2} - \frac{\rho\sigma_I \bar{b}}{\sigma_L}, \bar{b}\right) & \text{if } \pi^M > \bar{b}, \end{cases}$$

and $\eta = f(\pi_L^*, \pi^*)$*, where*

$$\pi_L^M = \frac{1}{1-\rho^2}\left(\frac{\mu_L - r}{\gamma\sigma_L^2} - \frac{\rho(\mu_I - r)}{\gamma\sigma_L\sigma_I}\right) \tag{A.3}$$

and η *is as defined as in* (12).

Proof of Lemma A.1. This follows from simple constrained bivariate quadratic function maximization. □

Now let us prove Theorem 1. Given any investment strategy (π_{Ls}, π_s), we denote

$$\sigma_s = \sqrt{\pi_{Ls}^2\sigma_L^2 + 2\rho\pi_{Ls}\pi_s\sigma_L\sigma_I + \pi_s^2\sigma_I^2}.$$

Then the budget constraint implies that for any feasible strategy π_{Ls}, π_s,

$$A_T = A_t \exp\left(\int_t^T \left[r + \pi_{Ls}(\mu_L - r) + \pi_s(\mu_I - r) - \frac{1}{2}\sigma_s^2\right] ds + \int_t^T \pi_{Ls}\sigma_L dB_{Ls} + \int_t^T \pi_s\sigma_I\, dB_{Is}\right).$$

Therefore, for $\gamma > 0$, $\gamma \neq 1$, some algebra yields that[13]

$$E_t\left[u(A_T)\right] + \frac{1}{1-\gamma} = \frac{A_t^{1-\gamma}}{1-\gamma} E_t\left[\exp\left((1-\gamma)\int_t^T f(\pi_{Ls}, \pi_s)\, ds\right) Z(T)\right], \tag{A.4}$$

where $f(\pi_L, \pi)$ is as defined in (A.1), and

$$Z(\nu) = \exp\left(-\int_t^\nu \frac{1}{2}(1-\gamma)^2\sigma_s^2\, ds + \int_t^\nu (1-\gamma)\pi_{Ls}\sigma_L\, dB_{Ls} + \int_t^\nu (1-\gamma)\pi_s\sigma_I\, dB_{I_s}\right)$$

is a positive local martingale, and therefore a supermartingale with

$$E_t\left[Z(T)\right] \leqslant Z(t) = 1. \tag{A.5}$$

[13] The proof for the case with $\gamma = 1$ follows similar arguments, but involves more technicality. We omit it to save space.

Lemma A.1 implies that for any feasible strategy (π_{Ls}, π_s), $f(\pi_{Ls}, \pi_s) \leqslant \eta$ and equality holds if and only if $\pi_{Ls} = \pi_L^*, \pi_s = \pi^*$. We then deduce that

$$E_t\big[u(A_T)\big] \leqslant \frac{A_t^{1-\gamma}}{1-\gamma} e^{(1-\gamma)\eta(T-t)} E_t\big[Z(T)\big] - \frac{1}{1-\gamma} \leqslant \frac{(A_t e^{\eta(T-t)})^{1-\gamma}}{1-\gamma} - \frac{1}{1-\gamma},$$

and equalities hold if and only if $\pi_{Ls} = \pi_L^*, \pi_s = \pi^*$, a.s. Therefore $(\pi_{Ls}, \pi_s) \equiv (\pi_L^*, \pi^*)$ is the unique solution. □

A.2. *Proof of Proposition 1 and Proposition 3*

Proposition 1 is a special case of Proposition 3. So, we will only prove Proposition 3. By transformation

$$V^c(x, y, t) \equiv (x+y)^{1-\gamma}\varphi^c(\pi, t) - \frac{1}{1-\gamma}, \quad \pi = \frac{y}{x+y}, \tag{A.6}$$

the HJB equation (20) reduces to

$$\max\big(\varphi_t^c + \mathcal{L}_1\varphi^c, -(1-\alpha\pi)\varphi_\pi^c - \alpha(1-\gamma)\varphi^c, (1+\theta\pi)\varphi_\pi^c - \theta(1-\gamma)\varphi^c\big) = 0,$$

where $\mathcal{L}_1$ is given in (16). The terminal and boundary conditions become

$$\varphi^c(\pi, T) = \frac{1}{1-\gamma},$$
$$(1+\theta\pi)\varphi_\pi^c - \theta(1-\gamma)\varphi^c = 0 \quad \text{on } \pi = \underline{b},$$
$$-(1-\alpha\pi)\varphi_\pi^c - \alpha(1-\gamma)\varphi^c = 0 \quad \text{on } \pi = \bar{b}.$$

Let

$$w = \frac{1}{1-\gamma}\log\big[(1-\gamma)\varphi^c\big]. \tag{A.7}$$

It is easy to see that $w(\pi, t)$ satisfies

$$\max\left\{w_t + \mathcal{L}_2 w, -\frac{\alpha}{1-\alpha\pi} - w_\pi, w_\pi - \frac{\theta}{1+\theta\pi}\right\} = 0 \tag{A.8}$$

in $(\underline{b}, \bar{b}) \times [0, T)$, with the terminal condition $w(\pi, T) = 0$ and the boundary conditions[14]

$$w_\pi(\underline{b}, t) = \frac{\theta}{1+\theta\pi}, \qquad w_\pi(\bar{b}, t) = -\frac{\alpha}{1-\alpha\pi},$$

where

$$\mathcal{L}_2 w = \frac{1}{2}\beta_1\pi^2(1-\pi)^2\big[w_{\pi\pi} + (1-\gamma)w_\pi^2\big] + (\beta_2 - \gamma\beta_1\pi)\pi(1-\pi)w_\pi + \beta_3 + \beta_2\pi - \frac{1}{2}\gamma\beta_1\pi^2 - \frac{1}{2}\beta_4\frac{(\pi w_\pi - 1)^2}{-\gamma + 2\gamma\pi w_\pi + \pi^2(w_{\pi\pi} + (1-\gamma)w_\pi^2)}$$

[14] For Proposition 1, $\underline{b} = -\frac{1}{\theta}$ or $\bar{b} = \frac{1}{\alpha}$, the boundary condition becomes $v(-\frac{1}{\theta}, t) = +\infty$ or $v(\frac{1}{\alpha}, t) = -\infty$. In this case the boundary condition can be removed.

Eq. (A.8) can be rewritten as

$$w_t + \mathcal{L}_2 w = 0, \quad \text{if } -\frac{\alpha}{1-\alpha\pi} < w_\pi < \frac{\theta}{1+\theta\pi}, \tag{A.9}$$

$$w_t + \mathcal{L}_2 w \leqslant 0, \quad \text{if } w_\pi = -\frac{\alpha}{1-\alpha\pi}, \tag{A.10}$$

$$w_t + \mathcal{L}_2 w \leqslant 0, \quad \text{if } w_\pi = \frac{\theta}{1+\theta\pi}. \tag{A.11}$$

Denote

$$v(\pi,t) = w_\pi(\pi,t). \tag{A.12}$$

Note that

$$\begin{aligned}
\frac{\partial}{\partial\pi}(\mathcal{L}_2 w) = {} & \frac{1}{2}\beta_1\pi^2(1-\pi)^2 v_{\pi\pi} + \big[\beta_1+\beta_2-(2+\gamma)\beta_1\pi\big]\pi(1-\pi)v_\pi \\
& + \big[\beta_2 - 2(\beta_2+\gamma\beta_1)\pi + 3\gamma\beta_1\pi^2\big]v \\
& + (1-\gamma)\beta_1\pi(1-\pi)v\big[(1-2\pi)v + \pi(1-\pi)v_\pi\big] + \beta_2 - \gamma\beta_1\pi \\
& + \frac{1}{2}\beta_4\frac{\pi^2(\pi v-1)[(v_{\pi\pi}+2vv_\pi)(\pi v-1) - 2\pi(v_\pi+v^2)]}{[\pi^2(v_\pi+v^2)-\gamma(\pi v-1)^2]^2} \\
\equiv {} & \mathcal{L}v.
\end{aligned} \tag{A.13}$$

The following lemma shows that we can transform the original problem into a double obstacle problem.

Lemma A.2. *$v(\pi,t)$ is the solution to the following parabolic double obstacle problem*:

$$\max\left\{\min\left\{-v_t - \mathcal{L}v, v + \frac{\alpha}{1-\alpha\pi}\right\}, v - \frac{\theta}{1+\theta\pi}\right\} = 0, \tag{A.14}$$

or equivalently,

$$v_t + \mathcal{L}v = 0, \quad \textit{if } -\frac{\alpha}{1-\alpha\pi} < v(\pi,t) < \frac{\theta}{1+\theta\pi}, \tag{A.15}$$

$$v_t + \mathcal{L}v \leqslant 0, \quad \textit{if } v(\pi,t) = -\frac{\alpha}{1-\alpha\pi}, \tag{A.16}$$

$$v_t + \mathcal{L}v \geqslant 0, \quad \textit{if } v(\pi,t) = \frac{\theta}{1+\theta\pi} \tag{A.17}$$

in $(\underline{b},\bar{b})\times[0,T)$, with the terminal condition $v(\pi,T)=0$ and the boundary conditions

$$v(\underline{b},t) = \frac{\theta}{1+\theta\pi}, \qquad v(\bar{b},t) = -\frac{\alpha}{1-\alpha\pi}.$$

Proof of Lemma A.2. In the (π,t) plane, define the associated sell region, buy region, and no-transaction region as follows:

$$SR^c_\pi(\underline{b},\bar{b}) \equiv \left\{(\pi,t)\in(\underline{b},\bar{b})\times[0,T)\colon\ v(\pi,t) = -\frac{\alpha}{1-\alpha\pi}\right\} = \big\{\pi \geqslant \bar{\pi}^c(t;\underline{b},\bar{b})\big\},$$

$$BR^c_\pi(\underline{b},\bar{b}) \equiv \left\{(\pi,t)\in(\underline{b},\bar{b})\times[0,T)\colon\ v(\pi,t) = \frac{\theta}{1+\theta\pi}\right\} = \big\{\pi \leqslant \underline{\pi}^c(t;\underline{b},\bar{b})\big\},$$

and

$$NTR^c_\pi(\underline{b},\bar{b}) \equiv \left\{(\pi,t)\in(\underline{b},\bar{b})\times[0,T]:\ -\frac{\alpha}{1-\alpha\pi}<v<\frac{\theta}{1+\theta\pi}\right\}$$
$$=\left\{\underline{\pi}^c(t;\underline{b},\bar{b})<\pi<\bar{\pi}^c(t;\underline{b},\bar{b})\right\},$$

where the second equalities in the above three expressions can be shown using similar arguments to those in Dai and Yi [18]. For any $t<T$, if $\underline{\pi}^c(t;\underline{b},\bar{b})=\underline{b}$ and $\bar{\pi}^c(t;\underline{b},\bar{b})=\bar{b}$, then differentiating Eq. (A.9) w.r.t. π immediately gives (A.15); if either $\underline{\pi}^c(t;\underline{b},\bar{b})>\underline{b}$ or $\bar{\pi}^c(t;\underline{b},\bar{b})<\bar{b}$, we will use an indirect method. Suppose $\underline{\pi}^c(t;\underline{b},\bar{b})>\underline{b}$,[15] then we can define $w_1(\pi,t)\equiv F(t)+\log(1+\theta\underline{\pi}^c(t;\underline{b},\bar{b}))+\int_{\underline{\pi}^c(t;\underline{b},\bar{b})}^{\pi}v(\xi,t)\,d\xi$, where $F(t)$ is chosen such that

$$w_{1t}+\mathcal{L}_2w_1|_{\pi=\underline{\pi}^c(t;\underline{b},\bar{b})}=0. \tag{A.18}$$

Clearly $w_{1\pi}=v$. Then, by (A.13), we can rewrite (A.15)–(A.17) as

$$\frac{\partial}{\partial\pi}(w_{1t}+\mathcal{L}_2w_1)\geqslant 0,\qquad w_{1\pi}=\frac{\theta}{1+\theta\pi},\quad \text{if } \pi\leqslant\underline{\pi}^c(t;\underline{b},\bar{b}),$$
$$\frac{\partial}{\partial\pi}(w_{1t}+\mathcal{L}_2w_1)=0,\qquad -\frac{\alpha}{1-\alpha\pi}<w_{1\pi}<\frac{\theta}{1+\theta\pi},$$
$$\text{if } \underline{\pi}^c(t;\underline{b},\bar{b})<\pi<\bar{\pi}^c(t;\underline{b},\bar{b}),$$
$$\frac{\partial}{\partial\pi}(w_{1t}+\mathcal{L}_2w_1)\leqslant 0,\qquad w_{1\pi}=-\frac{\alpha}{1-\alpha\pi},\quad \text{if } \pi\geqslant\bar{\pi}^c(t;\underline{b},\bar{b}).$$

This means that $w_{1t}+\mathcal{L}_2w_1$ is increasing in $\pi\leqslant\underline{\pi}^c(t;\underline{b},\bar{b})$, flat in $\underline{\pi}^c(t;\underline{b},\bar{b})<\pi<\bar{\pi}^c(t;\underline{b},\bar{b})$, and decreasing in $\pi\geqslant\bar{\pi}^c(t;\underline{b},\bar{b})$. Combining with (A.18), we then deduce that w_1 satisfies (A.9)–(A.11). Due to the uniqueness of the solution to the problem (A.8), we have $w=w_1$. The desired result follows. This completes the proof of the lemma. □

Now let us use Lemma A.2 to prove Part (1), and the proof of Part (2) is similar. For any $(\pi,t)\in SR^c_\pi$, we have $v=-\frac{\alpha}{1-\alpha\pi}$. By (A.16),

$$0\geqslant\left(\frac{\partial}{\partial t}+\mathcal{L}\right)\left(-\frac{\alpha}{1-\alpha\pi}\right)=\frac{1-\alpha}{(1-\alpha\pi)^3}\left[\beta_2-(\gamma\beta_1-\alpha\gamma\beta_1+\alpha\beta_2)\pi\right]$$
$$=\frac{(1-\alpha)\gamma\beta_1}{(1-\alpha\pi)^3}\left[\pi^M-\left(1-\alpha\left(1-\pi^M\right)\right)\pi\right].$$

Since $\frac{(1-\alpha)\gamma\beta_1}{(1-\alpha\pi)^3}>0$ and $1-\alpha(1-\pi^M)>0$, we obtain $\pi\geqslant\frac{\pi^M}{1-\alpha(1-\pi^M)}$ for any $(\pi,t)\in SR^c_\pi$, which implies that

$$\bar{\pi}^c(t;\underline{b},\bar{b})\geqslant\frac{\pi^M}{1-\alpha(1-\pi^M)}.$$

Clearly $\bar{\pi}^c(t;\underline{b},\bar{b})$ must also be in $[\underline{b},\bar{b}]$; we then obtain (21).

[15] A similar proof goes through if $\bar{\pi}^c(t;\underline{b},\bar{b})<\bar{b}$.

It remains to show that $\bar{\pi}^c(T^-;\underline{b},\bar{b})=\bar{b}$. Suppose not, then there exists a convergent sequence $\{\bar{\pi}^c(t_n;\underline{b},\bar{b})\}_{n=1,2,\ldots}$ such that $\lim_{n\to+\infty} t_n=T$ and $\lim_{n\to+\infty}\bar{\pi}^c(t_n;\underline{b},\bar{b})<\bar{b}$. It follows that

$$\lim_{n\to+\infty} v\big(\bar{\pi}^c(t_n;\underline{b},\bar{b}),t_n\big)=\lim_{n\to+\infty}\left(-\frac{\alpha}{1-\alpha\bar{\pi}^c(t_n;\underline{b},\bar{b})}\right)<0.$$

In contrast, $v(\pi,T)=0$ for all π. So, v is discontinuous at $\pi=\lim_{n\to+\infty}\bar{\pi}^c(t_n;\underline{b},\bar{b})$ and $t=T$. However, according to the regularity theory of solutions to the double obstacle problem (cf. Friedman [23]), $v(\pi,t)$ is continuous for any $\pi\neq\bar{b},\underline{b}$. A contradiction! □

A.3. Proof of Proposition 2

The double obstacle problem transformation remains valid for the unconstrained case, where $\bar{b}=\frac{1}{\alpha}$, $\underline{b}=-\frac{1}{\theta}$. We still denote by $v(\pi,t)$ the solution to the double obstacle problem. Let BR_π, SR_π, and NTR_π be the associated buy, sell, and no-transaction regions and let $\underline{\pi}(t)$ and $\bar{\pi}(t)$ be the associated buy and sell boundaries. Note that the differential operator $\mathcal{L}$ is degenerate at $\pi=0$, where the double obstacle problem reduces to

$$\begin{cases} v_t(0,t)+\beta_2 v(0,t)+\beta_2=0, & \text{if } -\alpha<v(0,t)<\theta,\\ v_t(0,t)+\beta_2 v(0,t)+\beta_2\leqslant 0, & \text{if } v(0,t)=-\alpha,\\ v_t(0,t)+\beta_2 v(0,t)+\beta_2\geqslant 0, & \text{if } v(0,t)=\theta,\\ v(0,T)=0. \end{cases}$$

Solving it, we then obtain

$$v(0,t)=\begin{cases} e^{\beta_2(T-t)}-1, & \text{when } t>\bar{t}_0,\\ -\alpha, & \text{when } t\leqslant\bar{t}_0, \end{cases}\quad \text{if } \beta_2<0, \tag{A.19}$$

$$v(0,t)=\begin{cases} e^{\beta_2(T-t)}-1, & \text{when } t>\underline{t}_0,\\ \theta, & \text{when } t\leqslant\underline{t}_0, \end{cases}\quad \text{if } \beta_2>0, \tag{A.20}$$

$$v(0,t)=0,\quad \text{if } \beta_2=0. \tag{A.21}$$

Now let us prove Part (1). If $\pi^M<0$, then $\beta_2<0$. So, we have (A.19) from which we can see that

$$v(0,t)\leqslant 0<\theta\quad \text{for all } t.$$

Note that $\{\pi=0\}\cap BR_\pi=\{(0,t)\colon\ v(0,t)=\theta\}$. So, $\{\pi=0\}\cap BR_\pi=\emptyset$. Combining with $\underline{\pi}(T^-)=-\frac{1}{\theta}$, we then deduce $\underline{\pi}(t)<0$ for all t. Again by (A.19), we have

$$v(0,t)>-\alpha\quad \text{for } t>\bar{t}_0,\quad \text{and}\quad v(0,t)=-\alpha\quad \text{for } t\leqslant\bar{t}_0.$$

Noticing $\{\pi=0\}\cap SR_\pi=\{(0,t)\colon\ v(0,t)=-\alpha\}$, we get

$$\{\pi=0,\ t>\bar{t}_0\}\notin SR_\pi,$$
$$\{\pi=0,\ t\leqslant\bar{t}_0\}\in SR_\pi.$$

These mean that $\bar{\pi}(t)$ intersects with the line $\{\pi=0\}$ at $\bar{t}_0$. Combining with the fact $\bar{\pi}(T^-)=\frac{1}{\alpha}$, we then infer $\bar{\pi}(t)\leqslant 0$ for $t<\bar{t}_0$, and $\bar{\pi}(t)\geqslant 0$ for $t>\bar{t}_0$.

To show the monotonicity of $\bar{\pi}(t)$ for $t>\bar{t}_0$, let us introduce the comparison principle that plays a critical role in the subsequent proofs.

Comparison principle for double obstacle problem (cf. Friedman [23])

Let v_i, $i = 1, 2$, satisfy the double obstacle problem

$$\max\{\min\{-v_{it} - \mathcal{A}v_i - f_i, v_i - g_i^l\}, v_i - g_i^u\} = 0$$

in $\Omega \times [0, T)$, where $\mathcal{A}$ is an elliptic operator.[16] Assume

$$f_1 \leqslant f_2;\ g_1^l \leqslant g_2^l;\ g_1^u \leqslant g_2^u \quad \text{in } \bar{\Omega} \times [0, T)$$

and

$$v_1 \leqslant v_2 \quad \text{on } t = T \quad \text{and} \quad \partial\Omega \times [0, T).$$

Then

$$v_1 \leqslant v_2 \quad \text{in } \Omega \times [0, T).$$

Now we prove that $\bar{\pi}(t)$, $t > \bar{t}_0$, is increasing in t. It suffices to show if $(\pi, t_1) \in SR_\pi$, then $(\pi, t_2) \in SR_\pi$ for any $t_2 < t_1$, $\pi > 0$. Due to $\bar{\pi}(T^-) = \frac{1}{\alpha}$ and $\underline{\pi}(T^-) = -\frac{1}{\theta}$, we have the equation $v_t + \mathcal{L}v = 0$ as t goes to T. So we can apply the equation at $t = T$ to get

$$v_t|_{t=T} = -\mathcal{L}v|_{t=T} = -\beta_2 + \gamma\beta_1\pi > -\beta_2 > 0 \quad \text{for } \pi > 0,$$

which gives $v(\cdot, T) \geqslant v(\cdot, T - \delta)$, for small $\delta > 0$. By (A.19), $v(0, t) \geqslant v(0, t - \delta)$ for any $t < T$. Since both $v(\pi, t)$ and $v(\pi, t - \delta)$ satisfy the double obstacle problem (A.14), applying the comparison principle gives $v(\cdot, t) \geqslant v(\cdot, t - \delta)$ or $v_t \geqslant 0$ in $\{\pi > 0\}$. Hence, if $(\pi, t_1) \in SR_\pi$, i.e., $v(\pi, t_1) = -\frac{\alpha}{1-\alpha\pi}$, then $v(\pi, t_2) \leqslant v(\pi, t_1) = -\frac{\alpha}{1-\alpha\pi}$. On the other hand, clearly $v(\pi, t_2) \geqslant -\frac{\alpha}{1-\alpha\pi}$. It follows that $v(\pi, t_2) = -\frac{\alpha}{1-\alpha\pi}$, i.e., $(\pi, t_2) \in SR_\pi$, which is desired.

The proof of Parts (2)–(3) is similar. □

A.4. Proof of Theorem 2

The uniqueness of viscosity solution can be obtained by using a similar argument in Akian, Mendaldi, and Sulem [3] (see also Crandall, Ishii, and Lions [11]). Here we highlight that on the boundaries the solution is a viscosity supersolution. In terms of the definition of viscosity solution and Itô's formula for a C^2 function of a stochastic process with jump, we are able to show that the value function is a viscosity solution to the HJB equation (see, for example, Shreve and Soner [32]).

To show the smoothness of the value function, let us examine the regularity of the solution $v(\pi, t)$ to the double obstacle problem. Without loss of generality, we assume $\underline{b} < 0$ and $\bar{b} > 0$. Noticing the differential operator $\mathcal{L}$ is degenerate at $\pi = 0, 1$, by the regularity theory of a double obstacle problem (see Friedman [23]), we know that $v(\pi, t) \in W_p^{2,1}([\underline{b} + \varepsilon, -\varepsilon] \times [0, T] \cup [\varepsilon, \bar{b} - \varepsilon] \times [0, T])$ for any $p > 1$ and small $\varepsilon > 0$, where $W_p^{2,1}$ is the Sobolev space. Thanks to the embedding theorem, we then infer $v(\pi, t) \in C^{1,0}((\underline{b}, \bar{b}) \times [0, T] \setminus (\{\pi = 0\} \cup \{\pi = 1\}))$. Further, we can obtain the smoothness of $\bar{\pi}^c(t; \underline{b}, \bar{b})$ and $\underline{\pi}^c(t; \underline{b}, \bar{b})$, from which we can infer v_t is continuous across $\pi = \bar{\pi}^c(t; \underline{b}, \bar{b})$ and $\pi = \underline{\pi}^c(t; \underline{b}, \bar{b})$. This indicates v_t is continuous except at $\pi = 0, 1$ and $\pi = \underline{b}, \bar{b}$. Hence, $v(\pi, t) \in C^{1,1}((\underline{b}, \bar{b}) \times [0, T] \setminus (\{\pi = 0\} \cup \{\pi = 1\}))$. Owing to (A.12), we conclude $w(\pi, t) \in C^{2,1}((\underline{b}, \bar{b}) \times [0, T] \setminus (\{\pi = 0\} \cup \{\pi = 1\}))$, which implies the desired smoothness of the value function by virtue of (A.6) and (A.7). □

[16] Strictly speaking, the elliptic operator $\mathcal{A}$ is required to satisfy certain conditions. Fortunately, we can show that the operator $\mathcal{L}$ involved in subsequent proofs does satisfy those conditions.

A.5. Proof of Proposition 4

Now we prove Part (1). Let $v(\pi, t; \underline{b}, \bar{b})$ be the solution of the double obstacle problem (A.14), and $BR_\pi^c(\underline{b}, \bar{b})$, $SR_\pi^c(\underline{b}, \bar{b})$, and $NTR_\pi^c(\underline{b}, \bar{b})$ be the associated buy, sell, and no-transaction regions as given in the proof of Lemma A.2. Assume $\bar{b}_1 \geqslant \bar{b}_2$. Because

$$v(\underline{b}, t; \underline{b}, \bar{b}_1) = v(\underline{b}, t; \underline{b}, \bar{b}_2) = \frac{\theta}{1+\theta\underline{b}},$$
$$v(\pi, T; \underline{b}, \bar{b}_1) = v(\pi, T; \underline{b}, \bar{b}_2) = 0,$$

and

$$v(\bar{b}_2, t; \underline{b}, \bar{b}_1) \geqslant -\frac{\alpha}{1-\alpha\bar{b}_2} = v(\bar{b}_2, t; \underline{b}, \bar{b}_2),$$

we apply the comparison principle in $(\underline{b}, \bar{b}_2) \times [0, T)$ to get

$$v(\pi, t; \underline{b}, \bar{b}_1) \geqslant v(\pi, t; \underline{b}, \bar{b}_2) \quad \text{in } (\underline{b}, \bar{b}_2) \times [0, T).$$

So, if $(\pi, t) \in SR_\pi^c(\underline{b}, \bar{b}_1)$, i.e., $v(\pi, t; \underline{b}, \bar{b}_1) = -\frac{\alpha}{1-\alpha\pi}$, then

$$v(\pi, t; \underline{b}, \bar{b}_2) \leqslant v(\pi, t; \underline{b}, \bar{b}_1) = -\frac{\alpha}{1-\alpha\pi}.$$

Because $-\frac{\alpha}{1-\alpha\pi}$ is also the lower bound, we get $v(\pi, t; \underline{b}, \bar{b}_2) = -\frac{\alpha}{1-\alpha\pi}$, i.e., $(\pi, t) \in SR_\pi^c(\underline{b}, \bar{b}_2)$. This indicates $SR_\pi^c(\underline{b}, \bar{b}_1) \subset SR_\pi^c(\underline{b}, \bar{b}_2)$, or equivalently, $\bar{\pi}^c(t; \underline{b}, \bar{b}_1) \geqslant \bar{\pi}^c(t; \underline{b}, \bar{b}_2)$. That is, $\bar{\pi}^c(t; \underline{b}, \bar{b})$ is increasing with $\bar{b}$. We can similarly obtain that $\bar{\pi}^c(t; \underline{b}, \bar{b})$ is increasing with $\underline{b}$.

Next we show that $\underline{\pi}^c(t; \underline{b}, \bar{b})$ is increasing with $\underline{b}$. Assume $\underline{b}_1 \geqslant \underline{b}_2$. Again applying the comparison principle in $(\underline{b}_1, \bar{b}) \times [0, T)$ gives

$$v(\pi, t; \underline{b}_1, \bar{b}) \geqslant v(\pi, t; \underline{b}_2, \bar{b}) \quad \text{in } (\underline{b}_1, \bar{b}) \times [0, T).$$

So, if $(\pi, t) \in BR_\pi^c(\underline{b}_2, \bar{b})$, i.e., $v(\pi, t; \underline{b}_2, \bar{b}) = \frac{\theta}{1+\theta\pi}$, then

$$v(\pi, t; \underline{b}_1, \bar{b}) \geqslant v(\pi, t; \underline{b}_2, \bar{b}) = \frac{\theta}{1+\theta\pi}.$$

Because $\frac{\theta}{1+\theta\pi}$ is also the upper bound, we get $v(\pi, t; \underline{b}_1, \bar{b}) = \frac{\theta}{1+\theta\pi}$, i.e., $(\pi, t) \in BR_\pi^c(\underline{b}_1, \bar{b})$. This indicates $BR_\pi^c(\underline{b}_2, \bar{b}) \subset BR_\pi^c(\underline{b}_1, \bar{b})$, i.e., $\underline{\pi}^c(t; \underline{b}_2, \bar{b}) \leqslant \underline{\pi}^c(t; \underline{b}_1, \bar{b})$, which is the desired result. In a similar way, we can show that $\underline{\pi}^c(t; \underline{b}, \bar{b})$ is increasing with $\bar{b}$.

It remains to prove Parts (2) and (3). We only prove Part (2) because the proof of Part (3) is similar. As in the proof of Proposition 2, we can still derive the boundary condition (A.19)–(A.21) at $\pi = 0$. So, $v(\pi, t; \underline{b}, \bar{b})$ in $\{\pi \leqslant 0\}$ is determined by the double obstacle problem restricted in $\{\pi < 0\}$ with the boundary condition (A.19)–(A.21) at $\pi = 0$. The boundary condition at $b = \bar{b} > 0$ will not affect the solution in $\{\pi \leqslant 0\}$. This yields the desired result. □

A.6. Proof of Proposition 5

If one of θ and α is 0, the HJB equation (20) remains valid and so does the double obstacle problem (A.14). Then we can use the same argument as in the proof of Proposition 3 to show that (21) and (22) remain valid. In the following, we will prove Part (3) only because the proof of Part (2) is similar.

If $\theta > 0$ and $\alpha = 0$, the corresponding double obstacle problem becomes

$$\max\left\{\min\{-v_t - \mathcal{L}v, v\}, v - \frac{\theta}{1+\theta\pi}\right\} = 0,$$

in $(\underline{b}, \bar{b}) \times [0, T)$, with the terminal condition $v(\pi, T) = 0$ and the boundary conditions

$$v(\underline{b}, t) = \frac{\theta}{1+\theta\pi}, \qquad v(\bar{b}, t) = 0.$$

Now the lower obstacle is 0, i.e., $v \geqslant 0$ for any (π, t), which gives $v_t|_{t=T} \leqslant 0$. Applying the comparison principle then leads to

$$v_t \leqslant 0 \quad \text{for any } (\pi, t). \tag{A.22}$$

To prove that the sell boundary $\bar{\pi}^c(t; \underline{b}, \bar{b})$ is decreasing in t, it suffices to show that if $(\pi, t_1) \in SR^c_\pi$, i.e., $v(\pi, t_1) = 0$, then $(\pi, t_2) \in SR^c_\pi$ for any $t_2 > t_1$. By (A.22), we have

$$v(\pi, t_2) \leqslant v(\pi, t_1) = 0.$$

On the other hand $v(\pi, t_2) \geqslant 0$ because the lower obstacle is 0. We then deduce $v(\pi, t_2) = 0$, the desired result. Similarly we can obtain the monotonicity of $\underline{\pi}^c(t; \underline{b}, \bar{b})$.

It remains to prove (3)(b). Using the same argument as in the proof of Proposition 3, we can obtain $\underline{\pi}^c(T^-; \underline{b}, \bar{b}) = \underline{b}$. Now we prove $\bar{\pi}^c(T^-; \underline{b}, \bar{b}) = \max(\min(\pi^M, \bar{b}), \underline{b})$. Suppose not, then we must have $\bar{\pi}^c(T^-; \underline{b}, \bar{b}) > \max(\min(\pi^M, \bar{b}), \underline{b})$ because of $\bar{\pi}^c(T^-; \underline{b}, \bar{b}) \geqslant \max(\min(\pi^M, \bar{b}), \underline{b})$. Clearly $\bar{\pi}^c(T^-; \underline{b}, \bar{b}) \leqslant \bar{b}$, then it follows

$$\pi^M < \bar{b}. \tag{A.23}$$

Noticing $\underline{\pi}^c(T^-; \underline{b}, \bar{b}) = \underline{b}$, we then infer that for any

$$\pi \in \left(\min\left(\max\left(\pi^M, \underline{b}\right), \bar{b}\right), \bar{\pi}^c\left(T^-; \underline{b}, \bar{b}\right)\right),$$

the equation $v_t + \mathcal{L}v = 0$ holds as t goes to T. So we can apply the equation at $t = T$ and $\pi \in (\min(\max(\pi^M, \underline{b}), \bar{b}), \bar{\pi}^c(T^-; \underline{b}, \bar{b}))$ to get

$$v_t|_{t=T} = -\mathcal{L}v|_{t=T} = -\beta_2 + \gamma\beta_1\pi.$$

Due to (A.22), we deduce $-\beta_2 + \gamma\beta_1\pi \leqslant 0$, namely,

$$\pi \leqslant \frac{\beta_2}{\gamma\beta_1} = \pi^M.$$

Combining with (A.23), we get

$$\pi \leqslant \min\left(\pi^M, \bar{b}\right),$$

which contradicts $\pi \in (\min(\max(\pi^M, \underline{b}), \bar{b}), \bar{\pi}^c(T^-; \underline{b}, \bar{b}))$. The proof is complete. □

A.7. *Proof of Proposition 6*

As before, we make a transformation:

$$w(\pi, t) = \frac{1}{1-\gamma}\log\left[(1-\gamma)\varphi(\pi, t)\right].$$

Then $w(\pi, t)$ satisfies

$$\max\left\{w_t + \mathcal{L}^r_2 w, -\frac{\alpha}{1-\alpha\pi} - w_\pi, w_\pi - \frac{\theta}{1+\theta\pi}\right\} = 0,$$

where

$$\begin{aligned}\mathcal{L}_2^r w = {} & \left[\frac{1}{2}\sigma_I^2(1-\pi)^2 + \frac{1}{2}\sigma_L^2 g^2 - \rho\sigma_I\sigma_L(1-\pi)g\right]\pi^2\left[w_{\pi\pi} + (1-\gamma)w_\pi^2\right] \\ & + \left[\left(-\gamma\sigma_I^2\pi + \mu_I - r\right)(1-\pi) \right.\\ & \left. + \left(\gamma\sigma_L^2 g + 2\gamma\rho\sigma_I\sigma_L\pi - \mu_L + r - \gamma\rho\sigma_I\sigma_L\right)g\right]\pi w_\pi \\ & - \frac{1}{2}\gamma\left(\sigma_I^2\pi^2 + \sigma_L^2 g^2\right) + (\mu_I - r - \gamma\rho\sigma_I\sigma_L g)\pi + (\mu_L - r)g + r\end{aligned}$$

with

$$\begin{aligned} g(\pi) &= \min\left(\frac{(\mu_L - r - \gamma\rho\sigma_I\sigma_L)(\pi w_\pi - 1)}{\sigma_L^2[-\gamma + 2\gamma\pi w_\pi + \pi^2(w_{\pi\pi} + (1-\gamma)w_\pi^2)]} + \frac{\rho\sigma_I\sigma_L(1-\pi)}{\sigma_L^2}, 1-\pi\right) \\ &\equiv \min\left(h(\pi; w_\pi), 1-\pi\right).\end{aligned}$$

Denote $v(\pi,t) = w_\pi(\pi,t)$. It can be verified that

$$\frac{\partial}{\partial\pi}\left(\mathcal{L}_2^r w\right) = \mathcal{L}^r v \equiv \begin{cases}\mathcal{L} & \text{if } h(\pi; v) < 1-\pi, \\ \widetilde{\mathcal{L}} & \text{if } h(\pi; v) \geqslant 1-\pi,\end{cases} \tag{A.24}$$

where $\mathcal{L}$ is as given before and

$$\begin{aligned}\widetilde{\mathcal{L}}v = {} & \frac{1}{2}\widetilde{\beta}_1\pi^2(1-\pi)^2 v_{\pi\pi} + \left[\widetilde{\beta}_1 + \widetilde{\beta}_2 - (2+\gamma)\widetilde{\beta}_1\pi\right]\pi(1-\pi)v_\pi \\ & + \left[\widetilde{\beta}_2 - 2(\widetilde{\beta}_2 + \gamma\widetilde{\beta}_1)\pi + 3\gamma\widetilde{\beta}_1\pi^2\right]v \\ & + (1-\gamma)\widetilde{\beta}_1\pi(1-\pi)v\left[(1-2\pi)v + \pi(1-\pi)v_\pi\right] + \widetilde{\beta}_2 - \gamma\widetilde{\beta}_1\pi\end{aligned}$$

with

$$\widetilde{\beta}_1 = \sigma_I^2 + \sigma_L^2 - 2\rho\sigma_I\sigma_L, \quad \widetilde{\beta}_2 = \mu_I - \mu_L + \gamma\sigma_L(\sigma_L - \rho\sigma_I), \quad \widetilde{\beta}_3 = \mu_L - \frac{1}{2}\gamma\sigma_L^2.$$

In the same way, we can consider the following double obstacle problem:

$$\max\left\{\min\left\{-v_t - \mathcal{L}^r v, v + \frac{\alpha}{1-\alpha\pi}\right\}, v - \frac{\theta}{1+\theta\pi}\right\} = 0, \tag{A.25}$$

in $(\underline{b}, \bar{b}) \times [0,T)$, with the terminal condition $v(\pi,T) = 0$ and the boundary conditions

$$v(\underline{b}, t) = \frac{\theta}{1+\theta\pi}, \qquad v(\bar{b}, t) = -\frac{\alpha}{1-\alpha\pi}.$$

It should be pointed out that the differential operator $\mathcal{L}^r$ involved in (A.25) is rather complicated and it is nontrivial to prove the existence of a solution to the associated double obstacle problem. We will assume that the problem has a solution. In addition, we assume the comparison principle holds for the differential operator $\mathcal{L}^r$.

Now let us prove Part (1). From (A.25), we can see that if $v(\pi,t) = \frac{\theta}{1+\theta\pi}$, then

$$-\mathcal{L}^r\left(\frac{\theta}{1+\theta\pi}\right) \leqslant 0.$$

Note that

$$h\left(\pi; \frac{\theta}{1+\theta\pi}\right) = \frac{(\mu_L - r - \gamma\rho\sigma_I\sigma_L)(\theta\pi + 1)}{\gamma\sigma_L^2} + \frac{\rho\sigma_I\sigma_L(1-\pi)}{\sigma_L^2}.$$

We then deduce that

$$\mathcal{L}^r = \begin{cases} \mathcal{L} & \text{if } \pi \in \Pi(\theta), \\ \widetilde{\mathcal{L}} & \text{if } \pi \notin \Pi(\theta). \end{cases}$$

We have seen in the proof of Proposition 3 that $-\mathcal{L}(\frac{\theta}{1+\theta\pi}) \leqslant 0$ leads to

$$\pi \leqslant \frac{\pi^M}{1+\theta(1-\pi^M)}.$$

In the same way, we can verify that $-\widetilde{\mathcal{L}}(\frac{\theta}{1+\theta\pi}) \leqslant 0$ yields

$$\pi \leqslant \frac{\pi_0^M}{1+\theta(1-\pi_0^M)}.$$

(24)–(25) then follow. Similarly we can obtain the result for the sell boundary. Using the same argument as in the proof of Proposition 3, we can prove $\bar{\pi}^c(T^-; \underline{b}, \bar{b}) = \bar{b}$ and $\underline{\pi}^c(T^-; \underline{b}, \bar{b}) = \underline{b}$.

To show Part (2), we note that

$$h(0; v) = \frac{\mu_L - r - \gamma\rho\sigma_I\sigma_L}{\gamma\sigma_L^2} + \frac{\rho\sigma_I\sigma_L}{\sigma_L^2} = \frac{\mu_L - r}{\gamma\sigma_L^2}.$$

Then at $\pi = 0$ the double obstacle problem reduces to

$$\max\{\min\{-v_t - \beta_2 v - \beta_2, v + \alpha\}, v - \theta\}\big|_{\pi=0} = 0 \quad \text{when } \frac{\mu_L - r}{\gamma\sigma_L^2} < 1$$

or

$$\max\{\min\{-v_t - \widetilde{\beta}_2 v - \widetilde{\beta}_2, v + \alpha\}, v - \theta\}\big|_{\pi=0} = 0 \quad \text{when } \frac{\mu_L - r}{\gamma\sigma_L^2} \geqslant 1$$

with $v(0, T) = 0$. It follows that

$$v(0,t) = \begin{cases} \max(\min(e^{\beta_2(T-t)} - 1, \theta), -\alpha) & \text{when } \frac{\mu_L - r}{\gamma\sigma_L^2} < 1, \\ \max(\min(e^{\widetilde{\beta}_2(T-t)} - 1, \theta), -\alpha) & \text{when } \frac{\mu_L - r}{\gamma\sigma_L^2} \geqslant 1. \end{cases}$$

As a consequence, the double obstacle problem can be solved in $\{\pi < 0\}$ and in $\{\pi > 0\}$ independently. The remaining arguments are the same as those in the proof of Proposition 4 and the comparison principle associated with $\mathcal{L}^r$ is used.

It remains to show Part (3). Notice that

$$h(\pi; 0) = \frac{\mu_L - r - \gamma\rho\sigma_I\sigma_L}{\gamma\sigma_L^2} + \frac{\rho\sigma_I\sigma_L(1-\pi)}{\sigma_L^2}.$$

Then $h(\pi; 0) < 1 - \pi$ if and only if $\pi \in \Pi(0)$. By virtue of (A.24) and the same argument as in the proof of Proposition 5, we can obtain the desired results. □

References

[1] V. Acharya, L.H. Pedersen, Asset pricing with liquidity risk, J. Finan. Econ. 77 (2005) 375–410.

[2] Y. Amihud, H. Mendelson, L.H. Pedersen, Liquidity and asset prices, Found. Trends Finan. 1 (2005) 269–364.

[3] M. Akian, J.L. Menaldi, A. Sulem, On an investment-consumption model with transaction costs, SIAM J. Control Optim. 34 (1996) 329–394.

[4] A. Almazan, K.C. Brown, M. Carlson, D.A. Chapman, Why constrain your mutual fund manager? J. Finan. Econ. 73 (2004) 289–321.
[5] P. Balduzzi, A. Lynch. Transaction costs and predictability: Some utility cost calculations, J. Finan. Econ. 52 (1999) 47–78.
[6] S. Basak, A. Pavlova, A. Shapiro, Optimal asset allocation and risk shifting in money management, Rev. Finan. Stud. 50 (2007) 1583–1621.
[7] J. Carpenter, Does option compensation increase managerial risk appetite? J. Finance 55 (2000) 2311–2331.
[8] J.M.R. Chalmers, R.M. Edelen, G.B. Kadlec, An analysis of mutual fund trading costs, Working paper, University of Pennsylvania.
[9] R. Clarke, H. de Silva, S. Thorley, Portfolio constraints and the fundamental law of active management, Financ. Anal. J. 58 (2002) 48–66.
[10] G.M. Constantinides, Capital market equilibrium with transaction costs, J. Polit. Economy 94 (1986) 842–862.
[11] M.G. Crandall, H. Ishii, P.L. Lions, User's guide to viscosity solutions of second order partial differential equations, Bull. Amer. Math. Soc. 27 (1992) 1–67.
[12] D. Cuoco, Optimal consumption and equilibrium prices with portfolio constraints and stochastic income, J. Econ. Theory 72 (1997) 33–73.
[13] D. Cuoco, R. Kaniel, Equilibrium prices in the presence of delegated portfolio management, J. Finan. Econ., doi:10.1016/j.jfineco.2011.02.012, in press.
[14] D. Cuoco, H. Liu, A martingale characterization of consumption choices and hedging costs with margin requirements, Math. Finance 10 (2000) 355–385.
[15] D. Cuoco, H. Liu, Optimal consumption of a divisible durable good, J. Econ. Dynam. Control 24 (2000) 561–613.
[16] J. Cvitanić, I. Karatzas, Convex duality in constrained portfolio optimization, Ann. Appl. Probab. 2 (1992) 767–818.
[17] M. Dai, Y.K. Kwok, J. Zong, Guaranteed minimum withdrawal benefit in variable annuities, Math. Finance 18 (2008) 595–611.
[18] M. Dai, F. Yi, Finite horizon optimal investment with transaction costs: A parabolic double obstacle problems, J. Differential Equations 246 (2009) 1445–1469.
[19] S. Das, R. Sundaram, On the regulation of fee structures in mutual funds, in: M. Avellaneda (Ed.), Quantitative Analysis in Financial Markets, vol. 3, Courant Institute of Mathematical Sciences, New York, 2001.
[20] E.J. Elton, M.J. Gruber, C.R. Blake, Incentive fees and mutual funds, J. Finance 58 (2003) 779–804.
[21] M.H.A. Davis, A.R. Norman, Portfolio selection with transaction costs, Math. Oper. Res. 15 (1990) 676–713.
[22] W.H. Fleming, H.M. Soner, Controlled Markov Processes and Viscosity Solutions, Springer, New York, 1993.
[23] A. Friedman, Variational Principles and Free-Boundary Problems, Wiley, New York, 1982.
[24] J. Karceski, M. Livingston, E.S. O'Neal, Portfolio transactions costs at US equity mutual funds, Working paper, University of Miami, 2009.
[25] N. Gârleanu, L.H. Pedersen, Dynamic trading with predictable returns and transaction costs, Working Paper, New York University, 2009.
[26] B. Jang, H. Koo, H. Liu, M. Loewenstein, Liquidity premia and transaction costs, J. Finance 62 (2007) 2329–2366.
[27] H. Liu, Optimal consumption and investment with transaction costs and multiple risky assets, J. Finance 59 (2004) 289–338.
[28] H. Liu, M. Loewenstein, Optimal portfolio selection with transaction costs and finite horizons, Rev. Finan. Stud. 15 (2002) 805–835.
[29] A. Lynch, S. Tan, Explaining the magnitude of liquidity premia: The roles of return predictability, wealth shocks and state-dependent transaction costs, J. Finance, submitted for publication.
[30] A. Lynch, S. Tan, Multiple risky assets, transaction costs and return predictability: Allocation rules & implications for US investors, J. Finan. Quant. Anal. (2009), forthcoming.
[31] R.C. Merton, Optimum consumption and portfolio rules in a continuous-time model, J. Econ. Theory 3 (1971) 373–413.
[32] S.E. Shreve, H.M. Soner, Optimal investment and consumption with transaction costs, Ann. Appl. Probab. 4 (1994) 609–692.
[33] R. Wermers, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses, J. Finance 55 (2000) 1655–1695.
[34] P. Wilmott, J. Dewynne, S. Howison, Option Pricing: Mathematical Models and Computation, Oxford Financial Press, Oxford, 1993.