Illiquidity, Portfolio Constraints, and Diversification *

Min Dai, Hanqing Jin, and Hong Liu

This revision: March 5, 2008

*Dai and Jin are from Department of Mathematics of National University of Singapore (NUS) and Liu is from the Olin Business School of Washington University in St. Louis. We thank Phil Dybvig, Bill Marshall, and the seminar participants at Washington University for helpful comments. We thank Yifei Zhong for excellent research assistance. Authors can be reached at matdm@nus.edu.sg, matjinh@nus.edu.sg, and liuh@wustl.edu. Dai and Jin are partially supported by the NUS RMI research grant and the NUS academic research grant. All errors are ours.

Illiquidity, Portfolio Constraints, and Diversification

Extended Abstract

Mutual funds are often restricted to allocate certain percentages of fund assets to certain securities that have different degrees of illiquidity. The coexistence of these restrictions and asset illiquidity and the interactions among them are important for the optimal trading strategy of a mutual fund. However, the existing literature ignores this coexistence and the interactions.

In this paper, we consider a fund that can trade a liquid stock and an illiquid stock that is subject to proportional transaction costs. The percentage of capital allocated to the illiquid stock is restricted to remain between a lower bound and an upper bound. We use a novel approach to characterize the value function and to provide extensive analytical comparative statics on the optimal trading strategy. The optimal trading strategy for the illiquid stock is determined by the optimal buy boundary and the optimal sell boundary which are easy to compute numerically using a penalty method. We also show the existence and uniqueness of the optimal trading strategy.

We also conduct numerical analysis on trading strategies, liquidity premium, and diversification. Constantinides (1986) concludes that transaction costs only have a second-order effect on liquidity premia. We find that the presence of portfolio constraints can significantly magnify the effect of transaction costs on liquidity premium and can make it more than a first-order effect. In addition, somewhat surprisingly, the liquidity premium can increase when constraints are less stringent. We show that even for log preferences, the optimal trading strategy is nonmyopic with respect to portfolio constraints, in the sense that a constraint can affect current trading strategy even when it is not binding now. Correlation coefficient between the two stocks affects the efficiency of diversification and thus can significantly alter the optimal trading strategy in both stocks. We also examine the endogenous choice of the portfolio bounds. Our analysis shows that the optimal upper (lower) bound is increasing (decreasing) in transaction costs.

Journal of Economic Literature Classification Numbers: D11, D91, G11, C61.

Keywords: Illiquidity, Portfolio Constraints, Portfolio Selection, Transaction Costs

Mutual funds are often restricted to allocate certain percentages of fund assets to certain securities that have different degrees of illiquidity. As stated in Almazan, Brown, Carlson, and Chapman (2004), over 90% funds are restricted from buying-on-margin and about 70% are prevented from short selling (see also, Clarke, de Silva, and Thorley (2002)). These constraints are often specified in a funds 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. U.S. stock funds commonly state in their prospectus an obligation to hold less than 20% of non U.S. stocks. In Switzerland, regulations require that at least two thirds of a fund's assets must 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 illiquidity in trading securities in some asset classes. Chalmers, Edelen, and Kadlec (1999) 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 inter-quartile range of 0.59%. Delib and Varma (2002) find that transaction costs concerns can affect a fund's permissible investment policy. There is a large literature on 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 portfolio constraints imposed upon mutual funds. As is well known, the presence of transaction costs and portfolio constraints can have drastic impact on the optimal trading strategy and thus the fund performance.[2] The coexistence of the portfolio constraints and asset illiquidity and the interactions among them are

[1]See, for example, Carpenter (2000), Basak, Pavlova and Shapiro (2006)

[2]See, for example, Davis and Norman (1990), Cuoco (1997), Liu and Loewenstein (2002), and Liu (2004).

important for the optimal trading strategy of a mutual fund. However, the existing literature ignores this coexistence and the interactions.

In this paper, we consider a fund that can trade a liquid stock and an illiquid stock that is subject to proportional transaction costs.[3] The percentage of fund assets allocated to the illiquid stock is restricted to remain between a lower bound and an upper bound.[4] Since the implied Hamilton-Jacobi-Bellman equation is highly nonlinear and difficult to analyze, we convert the original problem into a double obstacle problem which 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 smooth except on a measure zero set. The optimal trading strategy for the illiquid stock is determined by the optimal buy boundary and the optimal sell boundary between which no transaction occurs. Both the buy boundary and the sell boundary are monotonically decreasing in the portfolio bounds. Intuitively, when a positive lower bound is raised, it is going to bind for sure for the buy boundary when the time is close enough to the horizon.[5] To partly make up for the extra transaction costs from the time T liquidation of this over-investment, one also holds more at the buy boundary throughout the investment horizon. On the other hand, when a binding upper bound is reduced, obviously one needs to decrease the sell boundary. If the buy boundary remained the same, then the no-transaction region would become narrower and the trading frequency would

[3]Vast majority of mutual funds are restricted from holding any significant amount of cash. We therefore simply assume that they can only invest in stocks. The less relevant case that allows the fund to hold a riskless asset can be similarly analyzed without much technical difficulty.

[4]Obviously, this implies that the percentage allocated to the liquid stock is also restricted.

[5]This is because without a lower bound, the optimal buy boundary is always 0 with short enough time to horizon, as shown in Section III..

increase. Therefore the buy boundary also shifts downward to save transaction costs from frequent trading.

We find that the optimal buy (sell) boundary is monotonically decreasing (increasing) in calendar time. As time passes and thus the remaining horizon shrinks, it becomes less likely that the stock return over the remaining horizon can make up for the transaction costs paid at transaction. Thus the fund increases the sell boundary and lowers the buy boundary to decrease trading frequency.

We also conduct an extensive numerical analysis on trading strategies, liquidity premium, and diversification. The above analytical results are useful for improving the precision and robustness of the numerical procedure. Constantinides (1986) concludes that transaction costs only have a second-order effect on liquidity premia. We find that the presence of portfolio constraints can significantly magnify the effect of transaction costs on liquidity premia and can make it more than a first-order effect. Intuitively, the presence of constraints can force the fund to trade too frequently and to substantially distort its trading strategy. In addition, surprisingly, the liquidity premium can increase when constraints are *less* stringent. Lowering the binding lower bound (and increasing the upper bound) decreases the optimal investment in the illiquid stock and thus the value function becomes less sensitive to a change in the expected return of the illiquid stock. Also, the value function in the presence of transaction costs increases less as the constraints become less and less binding. Therefore it requires a greater reduction in the expected return (i.e., liquidity premium) in the no transaction cost case to make the fund willing to face transaction costs.

Our numerical analysis shows that even for log preferences, the optimal trading strategy is nonmyopic with respect to portfolio constraints, in the sense that constraints can affect current trading strategy even when they are not binding now.

Intuitively, even though the constraints are not binding now, they will for sure bind when time to horizon becomes short enough. In anticipation of this future binding of the constraints, the fund changes its current trading boundaries. 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.

To partly address the endogeneity of the portfolio constraints, we also examine the endogenous choice of the portfolio bounds by investors who have different risk preferences from the fund managers.[6] Our analysis show that the optimal upper (lower) bound is increasing (decreasing) in transaction costs. This is because loosing the constraints reduces transaction frequency and hence transaction costs.

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 (1986), Davis and Norman (1990), Liu and Loewenstein (2002), and Liu (2004)) finds that the presence of transaction costs can dramatically change the optimal trading strategy and even a small transaction cost can significantly decrease the trading frequency. The second strand of literature (e.g., Cvitanić and Karatzas (1992), Cuoco (1997), Cuoco and Liu (2000)) shows that portfolio constraints can also have 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. We find that the presence of portfolio constraints can significantly magnify the impact of transaction costs on the liquidity premium. In addition, the presence of transaction costs in general makes

[6]There are obviously other reasons for imposing constraints, e.g., different investment horizons, asymmetric information, etc.

the optimal trading strategy no longer myopic with respect to portfolio constraints.

The rest of the paper is organized as follows. In Section I., we describe the model. We solve the first benchmark case without transaction costs in Section II.. We solve the second benchmark case with transaction costs but without portfolio constraints in Section III.. Section IV. provides a verification theorem and an analysis of the main problem. In Section V. we conduct an extensive numerical analysis. Section VI. concludes. All the proofs are contained in the Appendix.

I. The Model

We consider a fund manager who has a finite horizon $T \in (0, \infty)$ and maximizes his constant relative risk averse (CRRA) utility from terminal wealth.[7] The fund can invest in two assets. One is a liquid risky asset ("the liquid stock," e.g., a large cap stock, S&P index) whose price process S_{Lt} evolves as [8]

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

where μ_L and $\sigma_L > 0$ are both constants and B_{Lt} is a one-dimensional Brownian motion. The other 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 $S_{It}^A = (1 + \theta)S_{It}$ and sell the stock at the bid price $S_{It}^B = (1 - \alpha)S_{It}$, where $\theta \geq 0$ and $0 \leq \alpha < 1$ represent the proportional transaction cost rates and S_{It} follows

[7]This form of utility function is consistent with a linear fee structure predominantly adopted by mutual fund companies and is also commonly used in the literature (e.g., Carpenter (2000), Basak, Pavlova, and Shapiro (2006)). As in Huang and Liu (2007), generalization to the class of hyperbolic absolute risk averse utility is straightforward.

[8]Extension to a case with a risk-free asset and/or multiple liquid assets is trivial and does not change any of the qualitative results. Since most funds are prohibited from making any significant amount of risk-free investment, we assume that the fund cannot hold any risk-free asset. In fact, we have numerically solved the case where investment in the risk free asset is allowed but the fraction of wealth invested in the risk free asset is restricted to be small in addition to the constraints on the illiquid asset holdings, all the qualitative results hold.

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 Brownian motion that has a correlation of ρ with B_{Lt} with $|\rho| < 1$.[9]

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

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

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

where 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 $D(0) = I(0) = 0$.

Let $W_t = x_t + y_t$ be the fund's wealth (on paper) at time t. The fund is subject to the following exogenously given constraints on its trading strategy:[10]

$$\underline{b} \leq \frac{y_t}{W_t} \leq \bar{b}, \quad \forall t \geq 0, \tag{5}$$

where $-\frac{1}{\theta} \leq \underline{b} < \bar{b} \leq \frac{1}{\alpha}$ are constants.[11] These constraints restrict the fraction of wealth (on paper) that must be invested in the illiquid asset and imply that the fund

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

[10]Because of 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 portfolio constraints on the trading strategy of the fund. See Almazan et. al. (2004) for more details on why many mutual fund managers are constrained. In this paper we focus on the case where this constraint is exogenously given and do not consider the optimal contracting issue. This serves as a foundation toward examining the optimal contracting problem in the presence of transaction costs and endogenous portfolio constraints. Later in this paper, we illustrate the choice of optimal constraints using numerical examples.

[11]Similar arguments to those in Cuoco and Liu (2000) 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.

is always solvent after liquidation, i.e., the liquidation wealth[12]

$$\hat{W}_t \geq 0, \quad \forall t \geq 0, \tag{6}$$

where

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

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

The fund manager's problem is then[13]

$$\sup_{(D,I)\in\Theta(x_0,y_0)} E\left[u(W_T)\right], \tag{8}$$

where the utility function is given by

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

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 wealth on paper instead of the liquidation wealth. This is consistent with common industry practice and avoids trading strategies that lead to liquidation on the terminal date.

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

[13] 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.

II. Optimal Policies without Transaction Costs

For 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

$$J(W,t) \equiv \sup_{\{\pi_L,\pi_I\}} E_t\left[u(W_T)|W_t = W\right], \tag{9}$$

subject to the self financing condition

$$dW_s = (1-\pi_{I_s})\, W_s\mu_L ds + (1-\pi_{I_s})\, W_s\sigma_L dB_{Ls} + \pi_{I_s}W_s\mu_I ds + \pi_{I_s}W_s\sigma_I dB_{Is}, \quad \forall s \geq t, \tag{10}$$

and the portfolio constraint (5), where π_{I_s} represents the fraction of wealth invested in the illiquid stock.

Let π_I^M ("Merton line") be the optimal fraction of wealth invested in the illiquid stock in the unconstrained case in the absence of transaction costs. Then it can be shown that

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

We summarize the main result for this case of no transaction costs in the following theorem.

Theorem 1 *Suppose that $\alpha = \theta = 0$. Then the optimal trading policy is given by*

$$\pi_I^* = \begin{cases} \bar{b} & \text{if } \pi_I^M \geq \bar{b} \\ \pi_I^M & \text{if } \underline{b} < \pi_I^M < \bar{b} \\ \underline{b} & \text{if } \pi_I^M \leq \underline{b} \end{cases}, \quad \pi_L^* = 1 - \pi_I^*$$

and the value function is

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

where

$$\eta = \mu_I\pi_I^* + \mu_L\left(1-\pi_I^*\right) - \frac{1}{2}\gamma\left[\sigma_I^2\pi_I^{*2} + \sigma_L^2\left(1-\pi_I^*\right)^2 + 2\rho\sigma_I\sigma_L\pi_I^*(1-\pi_I^*)\right].$$

PROOF: see Appendix.

Theorem 1 implies that the optimal fractions of wealth 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.

III. The Transaction Cost Case without Constraints

In the case where $\alpha + \theta > 0$, the problem is considerably more complicated. In this section, we consider the unconstrained case first. In this case, the investor's problem at time t becomes

$$V(x, y, t) \equiv \sup_{(D,I) \in \Theta(x,y)} E_t \left[u(W_T) | x_t = x, y_t = y \right] \tag{12}$$

with $\underline{b} = -\frac{1}{\theta}$ and $\bar{b} = \frac{1}{\alpha}$. Under regularity conditions on the value function, we have the following HJB equation:

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

with the boundary conditions

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

and the terminal condition

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

where

$$\mathscr{L}V = \frac{1}{2}\sigma_I^2 y^2 V_{yy} + \frac{1}{2}\sigma_L^2 x^2 V_{xx} + \rho\sigma_I\sigma_L xy V_{xy} + \mu_I y V_y + \mu_L x V_x$$

As we show later, the HJB equation implies that the solvency region for the illiquid stock

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

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 (1990).

The homogeneity of the utility function u 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 (1993), Lemma VIII.3.2]. This homogeneity implies that

$$V(x, y, t) \equiv (x+y)^{1-\gamma}\varphi\left(\frac{y}{x+y}, t\right) - \frac{1}{1-\gamma}, \tag{13}$$

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

$$\pi = \frac{y}{x+y} \tag{14}$$

denote the fraction of wealth invested the illiquid stock. The homogeneity property then implies that Buy, No-transaction, and Sell regions can be described by two functions of time $\underline{\pi}_I(t)$ and $\overline{\pi}_I(t)$. The Buy region BR corresponds to $\pi \leq \underline{\pi}_I(t)$, the Sell region SR to $\pi \geq \overline{\pi}_I(t)$, and the No-Transaction region NTR to $\underline{\pi}_I(t) < \pi < \overline{\pi}_I(t)$. A time snapshot of these regions is depicted in Figure 1. As we show later, the optimal trading strategy in the illiquid stock is to transact a minimum amount to keep the ratio π_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 the liquidated wealth invested in both the illiquid and the liquid stocks change stochastically, since π_t varies stochastically due to no transaction in NTR.

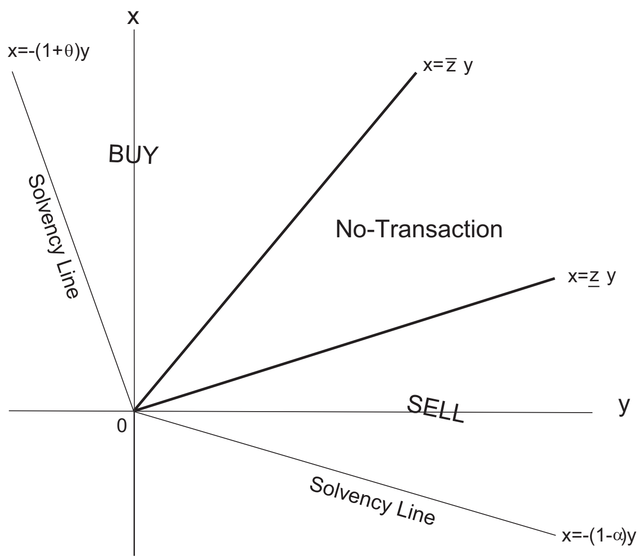

Figure 1: The Solvency Region

Using (13), the HJB equation simplifies into:

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

with the terminal condition

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

where

$$\mathcal{L}_1\varphi = \frac{1}{2}\beta_1\pi^2\left(1-\pi\right)^2\varphi_{\pi\pi} + \left(\beta_2 - \gamma\beta_1\pi\right)\pi\left(1-\pi\right)\varphi_\pi + \left(1-\gamma\right)\left(\beta_3 + \beta_2\pi - \frac{1}{2}\gamma\beta_1\pi^2\right)\varphi,$$

$$\begin{aligned} \beta_1 &= \sigma_I^2 + \sigma_L^2 - 2\rho\sigma_I\sigma_L, \\ \beta_2 &= \mu_I - \mu_L + \gamma\sigma_L\left(\sigma_L - \rho\sigma_I\right), \\ \beta_3 &= \mu_L - \frac{1}{2}\gamma\sigma_L^2. \end{aligned} \tag{15}$$

The nonlinearity of this HJB equation and the time-varying nature of the free boundaries make it difficult to solve directly. Instead, as in Dai and Yi (2006), 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 $-\frac{1}{\alpha} + 1 < \pi_I^M < \frac{1}{\theta} + 1$.*

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

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

Proposition 1 *Let π_I^M is as defined in (11). Under Assumption 1, we have $\forall t \in [0, T]$,*

1. *for the sell boundary, there exists $\overline{t} < T$ such that*

$$\frac{1}{\alpha} = \overline{\pi}_I(s) \geq \overline{\pi}_I(t) \geq \frac{\pi_I^M}{1 - \alpha\left(1 - \pi_I^M\right)}, \text{ for any } t \text{ and all } s > \overline{t};$$

2. *for the buy boundary, there exists $\underline{t} < T$ such that*

$$-\frac{1}{\theta} = \underline{\pi}_I(s) \leq \underline{\pi}_I(t) \leq \frac{\pi_I^M}{1 + \theta\left(1 - \pi_I^M\right)}, \text{ for any } t \text{ and all } s > \underline{t}.$$

PROOF: see Appendix.

This proposition shows that both the buy boundary and the sell boundary become the solvency line when the investment horizon is short enough. In addition, if $\pi_I^M \in (0,1)$, then the width of the NTR is bounded below by

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

Let

$$\beta_4 = \mu_I - \mu_L - \gamma\sigma_I\,(\sigma_I - \rho\sigma_L)\,,$$

$$\bar{t}_0 = T - \frac{1}{\beta_2}\log(1-\mu)\,,\ \bar{t}_1 = T - \frac{1}{\beta_4}\log(1-\mu)\,, \tag{16}$$

$$\underline{t}_0 = T - \frac{1}{\beta_2}\log(1+\theta)\,,\ \underline{t}_1 = T - \frac{1}{\beta_4}\log(1+\theta)\,. \tag{17}$$

Proposition 2 *Under Assumption 1, we have:*

1. If $\pi_I^M < 0$, then $\underline{\pi}_I(t) < 0$ for all t and $\overline{\pi}_I(t)$ is below 0 for $t < \bar{t}_0$, between 0 and 1 for $t \in [\bar{t}_0, \bar{t}_1]$, and above 1 for $t > \bar{t}_1$; In addition, $\overline{\pi}_I(t)$ is increasing in t for $t > \bar{t}_0$.

2. If $0 < \pi_I^M < 1$, then $\underline{\pi}_I(t)$ is between 0 and 1 for $t < \underline{t}_0$ and below 0 for $t \geq \underline{t}_0$; $\overline{\pi}_I(t)$ is between 0 and 1 for $t < \bar{t}_1$ and above 1 for $t \geq \bar{t}_1$; In addition, $\underline{\pi}_I(t)$ is decreasing in t for $t > \underline{t}_0$ and $\overline{\pi}_I(t)$ is increasing in t for $t > \bar{t}_1$.

3. If $\pi_I^M > 1$, then $\overline{\pi}_I(t) > 1$ for all t, and $\underline{\pi}_I(t)$ is above 1 for $t < \underline{t}_1$, between 0 and 1 for $t \in [\underline{t}_1, \underline{t}_0]$, and below 0 for $t > \underline{t}_0$; In addition, $\underline{\pi}_I(t)$ is decreasing in t for $t > \underline{t}_1$.

4. If $\pi_I^M = 0$, then $\underline{\pi}_I(t) < 0$ and $\overline{\pi}_I(t) > 0$ for all t, and $\underline{\pi}_I(t) = \overline{\pi}_I(t) = 0$ if $T = \infty$. Similarly, if $\pi_I^M = 1$, then $\underline{\pi}_I(t) < 1$ and $\overline{\pi}_I(t) > 1$ for all t, and $\underline{\pi}_I(t) = \overline{\pi}_I(t) = 1$ if $T = \infty$. In addition, if $\pi_I^M = 0$ or $\pi_I^M = 1$, then $\underline{\pi}_I(t)$ is decreasing in t and $\overline{\pi}_I(t)$ is increasing in t for $t \in [0, T]$.

Proof: see Appendix.

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 1 shows that if the time to horizon is short (i.e., $< T - \bar{t}_0$), then the sell boundary will be always 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 position in the illiquid asset, then the fund will only sell a part 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 $\pi_I^M = 0$ or $\pi_I^M = 1$. However, for other cases, we are able to show this property when the horizon is short enough. For example, Part 2 implies that this monotonicity holds when $t > \max(\bar{t}_1, \underline{t}_0)$.

Propositions 1 and 2 imply that in the absence of position limits, the portfolio chosen by the fund can be far from the optimal portfolio that is optimal without transaction costs. This large deviation is suboptimal for investors with longer horizons and therefore it may be one of the reasons for investors to impose position limits.

IV. The Transaction Cost Case With Portfolio Constraints

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

$$V^c(x, y, t) \equiv \sup_{(D,I)\in\Theta(x,y)} E\left[u(W_T)|x_t = x, y_t = y\right] \tag{18}$$

with $\underline{b} \leq y_s/(x_s + y_s) \leq \bar{b}$ for all $T \geq s \geq t$.

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

$$\max(V_t^c + \mathscr{L}V^c, (1-\alpha)V_x^c - V_y^c, -(1+\theta)V_x^c + V_y^c) = 0, \tag{19}$$

with the boundary conditions[14]

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

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 *(i) The HJB equation (19) admits a unique viscosity solution, and the value function is the viscosity solution.*

[14]It should be pointed out that the boundary conditions should be slightly modified when $\underline{b} = 0$, or $\bar{b} = 1$. For example, if $\pi_I^M > 0$, we then infer from Proposition 2 that $\{\pi = 0, \underline{t}_0 < t < T\}$ belongs to NTR, and the corresponding value function there

$$V^c(x, 0, t) \quad = \quad E\left[u\left(x_T\right)|x_t = x\right] = \frac{x^{1-\gamma}e^{(1-\gamma)(\mu_L - \frac{1}{2}\gamma\sigma_L^2)(T-t)} - 1}{1-\gamma}, \text{ for } \underline{t}_0 \leq t < T,$$

which is the boundary condition at $\underline{b}$ for $t \in [\underline{t}_0, T)$ if $\underline{b} = 0$.

(ii) The value function is $C^{2,2,1}$ in $\{(x,y,t) : x + (1-\alpha)y^+ - (1+\theta)y^- > 0,\ \underline{b} < y/(x+y) < \bar{b},\ 0 \le t < T\} \setminus (\{y = 0\} \cup \{x = 0\})$.

Proof: see Appendix.

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

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

Proposition 3 *Under Assumption 1, we have*

1. *for the sell boundary, there exists $\bar{t}_b < T$ such that*

$$\bar{b} = \overline{\pi}_I^c(s;\underline{b},\bar{b}) \ge \overline{\pi}_I^c(t;\underline{b},\bar{b}) \ge \max\left(\min\left(\frac{\pi_I^M}{1-\alpha\left(1-\pi_I^M\right)}, \bar{b}\right), \underline{b}\right),\ \text{for any } t \text{ and all } s > \bar{t}_b; \tag{20}$$

2. *for the buy boundary, there exists $\underline{t}_b < T$ such that*

$$\min\left(\max\left(\frac{\pi_I^M}{1+\theta\left(1-\pi_I^M\right)}, \underline{b}\right), \bar{b}\right) \ge \underline{\pi}_I^c(t;\underline{b},\bar{b}) \ge \underline{\pi}_I^c(s;\underline{b},\bar{b}) = \bar{b},\ \text{for any } t \text{ and all } s > \underline{t}_b. \tag{21}$$

Proof: The proof is similar to that of Proposition 1.

Corollary 1 *Under Assumption 1,*

1. *if $\frac{\pi_I^M}{1-\alpha\left(1-\pi_I^M\right)} \ge \bar{b}$, then $\overline{\pi}_I^c(t;\underline{b},\bar{b}) = \bar{b}$ for all $t \in [0,T]$;*
2. *if $\frac{\pi_I^M}{1+\theta\left(1-\pi_I^M\right)} \le \underline{b}$, then $\underline{\pi}_I^c(t;\underline{b},\bar{b}) = \underline{b}$ for all $t \in [0,T]$.*

As stated in Corollary 1, these results imply that if the adjusted Merton line is higher than the upper bound, then the sell boundary becomes flat throughout the horizon and that if the 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 remain flat at the lower bound $\underline{b}$ (upper bound $\overline{b}$) when the remaining horizon is short enough irrespective of the level of the Merton line. This is the same as in the unconstrained case.

Proposition 4 *Under Assumption 1, we have:*

1. *Both $\overline{\pi}_I^c(t;\underline{b},\overline{b})$ and $\underline{\pi}_I^c(t;\underline{b},\overline{b})$ are increasing in $\underline{b}$ and $\overline{b}$ for all $t \in [0,T]$;*

2. *If $\overline{b} > 0$, then the upper bound does not affect the sell/buy boundary that is below 0; If $\overline{b} > 1$, then the upper bound does not affect the sell/buy boundary that is below 1;*

3. *If $\underline{b} < 0$, then the lower bound does not affect the sell/buy boundary that is above 0; If $\underline{b} < 1$, then the lower bound does not affect the sell/buy boundary that is above 1;*

PROOF: see Appendix.

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. Intuitively, when a positive lower bound is raised, it is going to bind for sure for the buy boundary when the time is close enough to the horizon. To partly make up for the extra transaction costs from the time T liquidation of this over-investment, one also holds more at the buy boundary throughout the investment horizon. When a binding upper bound is reduced, obviously one needs to decrease the sell boundary. If the buy boundary remained the same, then the no-transaction region would become narrower and the trading frequency would increase. Therefore the buy boundary also shifts downward to save transaction costs from frequent trading.

Parts 2 and 3 of Proposition 4 suggest that the optimal boundaries in the each of the three regions: $\{\pi \leq 0\}$, $\{0 < \pi < 1\}$, and $\{\pi \geq 1\}$ are not affected by an constraint that lies in a different region. Intuitively, this is because in NTR, the position in the illiquid asset can never become levered if it is unlevered at time 0 and can never become negative if it is positive at time 0, i.e., the fraction of wealth invested in the illiquid asset cannot cross the $\pi = 1$ line or the $\pi = 0$ line.

Proposition 5 *Under Assumption 1, suppose $\underline{b} < 0$ and $\bar{b} > 0$. Then we have:*

1. *If $\pi_I^M < 0$, then $\underline{\pi}_I(t) < 0$ for all t and $\overline{\pi}_I(t)$ is below 0 for $t < \bar{t}_0$, between 0 and $\min(\bar{b}, 1)$ for $t \geq \bar{t}_0$; In addition, $\overline{\pi}_I(t)$ is increasing in t for $t > \bar{t}_0$.*

2. *If $0 < \pi_I^M < 1$, then $\underline{\pi}_I(t)$ is between 0 and $\min(\bar{b}, 1)$ for $t < \underline{t}_0$ and below 0 for $t \geq \underline{t}_0$; $\overline{\pi}_I(t)$ is between 0 and $\min(\bar{b}, 1)$ for all t; In addition, $\underline{\pi}_I(t)$ is decreasing in t for $t > \underline{t}_0$.*

3. *If $\pi_I^M \geq 1$, then $\overline{\pi}_I(t) \geq \min(\bar{b}, 1)$ for all t, and $\underline{\pi}_I(t)$ is between 0 and $\min(\bar{b}, 1)$ for $t \leq \underline{t}_0$, and below 0 for $t > \underline{t}_0$; In addition, $\underline{\pi}_I(t)$ is decreasing in t for $t > \underline{t}_0$.*

4. *If $\pi_I^M = 0$, then $\underline{\pi}_I(t) < 0$ and $\overline{\pi}_I(t) > 0$ for all t, and $\underline{\pi}_I(t) = \overline{\pi}_I(t) = 0$ if $T = \infty$. In addition, $\underline{\pi}_I(t)$ is decreasing in t and $\overline{\pi}_I(t)$ is increasing in t for $t \in [0, T]$.*

Proof: see Appendix.

Proposition 6 *Under Assumption 1, suppose $0 \leq \underline{b} < 1$ and $\bar{b} > 1$. Then we have:*

1. *If $\pi_I^M \leq 0$, then $\underline{\pi}_I(t) = \underline{b}$ for all t, and $\overline{\pi}_I(t)$ is between $\underline{b}$ and 1 for $t \geq \bar{t}_1$, and above 1 for $t > \bar{t}_1$; In addition, $\overline{\pi}_I(t)$ is increasing in t for $t > \bar{t}_1$.*

2. *If* $0 < \pi_I^M < 1$*, then* $\underline{\pi}_I(t)$ *is between* $\underline{b}$ *and 1 for all* t *and* $\overline{\pi}_I(t)$ *is between* $\underline{b}$ *and 1 for* $t < \bar{t}_1$ *and above 1 for* $t \geq \bar{t}_1$*; In addition,* $\overline{\pi}_I(t)$ *is increasing in* t *for* $t > \bar{t}_1$.

3. *If* $\pi_I^M > 1$*, then* $\overline{\pi}_I(t) > 1$ *for all* t*, and* $\underline{\pi}_I(t)$ *is above 1 for* $t < \underline{t}_1$*, between* $\underline{b}$ *and 1 for* $t > \underline{t}_1$*; In addition,* $\underline{\pi}_I(t)$ *is decreasing in* t *for* $t > \underline{t}_1$.

4. *If* $\pi_I^M = 1$*, then* $\underline{\pi}_I(t) < 1$ *and* $\overline{\pi}_I(t) > 1$ *for all* t*, and* $\underline{\pi}_I(t) = \overline{\pi}_I(t) = 1$ *if* $T = \infty$*. In addition,* $\underline{\pi}_I(t)$ *is decreasing in* t *and* $\overline{\pi}_I(t)$ *is increasing in* t *for* $t \in [0, T]$.

PROOF: see Appendix.

Propositions 5 and 6 shows similar results on the properties of the optimal trading boundaries to those in Proposition 2. In particular, monotonicity patterns remain the same in the presence of constraints for the two cases in these two propositions. The only difference is that the portfolio constraints may be binding in some regions. Interestingly, even when the portfolio constraints are binding, the times that optimal boundaries cross 0 and 1 remain unchanged. For example, Part 1 of Proposition 6 suggests the time that the sell boundary reaches 1 (i.e., $\bar{t}_1$) is the same as that in the unconstrained case.

V. Numerical Analysis

In this section, we conduct numerical analysis of the optimal trading strategy, the diversification, and the liquidity premium. For this analysis we use the following default parameter values: $\gamma = 2$, $T = 5$, $\mu_L = 0.06$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $\rho = 0.2$, $\alpha = 0.01$, $\theta = 0.01$, $\underline{b} = 0.60$, and $\bar{b} = 0.80$, which implies that the fraction of wealth invested in the illiquid (small cap) stock is greater than that in the liquid (large cap) stock. For a large cap fund, we set $\underline{b} = 0.10$ and $\bar{b} = 0.30$ so that the fraction of wealth invested in the liquid (large cap) stock is greater than that in the illiquid (small cap) stock. We use a higher expected return and a higher volatility for the small cap stock than those for a large cap stock. We use a penalty method to solve the HJB equations (see Dai, Kwok, and Zong (2007) for example).

In Figure 2, we plot π_I 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 are monotonically increasing in time. The lower bound of 60% is binding throughout the investment horizon and therefore the buy boundary becomes flat at 60% across all time. The sell boundary reaches the upper bound of 80% at $t = 4.74$. In addition, compared to the unconstrained case, the sell boundary before $t = 3.98$ is moved higher and the portion after $t = 3.98$ is moved lower. 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. Intuitively, since the fund will be forced to sell some of the illiquid stock later on and the buy boundary is forced to move higher than the unconstrained case, the fund needs to increase the sell boundary early on to reduce transaction frequency and transaction cost payment. The Merton line is flat through time, implying that in the absence of transaction costs, it is optimal to keep a constant fraction of wealth in the stock. In the presence of transaction costs, however, the optimal fraction becomes a stochastic process.

We present a similar figure (Figure 3) for the large cap fund case. In this case, both the lower bound (10%) and the upper bound (30%) are tight constraints and the upper

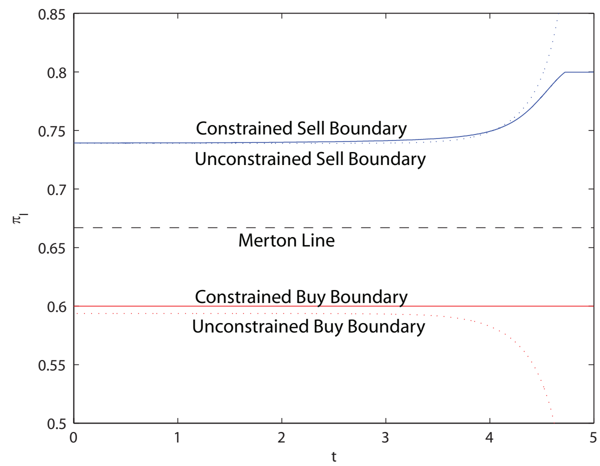

Figure 2. The optimal trading strategy for the illiquid asset for a small cap fund against time. Parameter default values: $\gamma = 2$, $T = 5$, $\mu_L = 0.06$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $\rho = 0.2$, $\alpha = 0.01$, $\theta = 0.01$, $\underline{b} = 0.60$, and $\bar{b} = 0.80$.

bound becomes so restrictive that the sell boundary becomes flat at 30% throughout the horizon. The buy boundary also shifts downward significantly through most of the horizon and only shifts upward towards the end of the horizon. In contrast to Figure 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.

In Figure 4, we plot the time 0 optimal boundaries ($\pi_I(0)$) 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 to decrease transaction frequency. In contrast, the buy boundary in the presence of constraints first decreases and then stays at the lower bound because the lower bound becomes binding. The binding lower bound also drives up the sell boundary and makes it move up more for higher transaction cost rates.

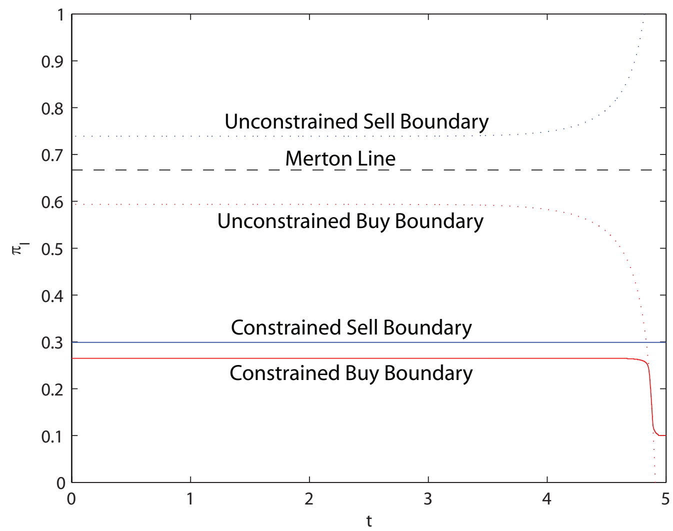

Figure 3. The optimal trading strategy for the illiquid asset for a large cap fund against time. Parameter default values: $\gamma = 2$, $T = 5$, $\mu_L = 0.06$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $\rho = 0.2$, $\alpha = 0.01$, $\theta = 0.01$, $\underline{b} = 0.10$, and $\bar{b} = 0.30$.

This figure also shows that as the correlation between the liquid and illiquid stock returns increases, the fraction of wealth invested in the illiquid stock increases in the absence of transaction costs. This is because the diversification benefit of investing in the large cap stock decreases and thus one should invest more in the small cap stock that has a higher Sharpe ratio. In the presence of transaction costs, an increase of the correlation drives both the sell boundary and the buy boundary upward. In addition, the upper bound becomes binding for the sell boundary for transaction cost above 0.2%.

In Figure 5, we plot the time 0 optimal boundaries ($\pi_I(0)$) against the difference $R_I \equiv \mu_I - \alpha - \mu_L$ (a measure of the excess return net of illiquidity over the liquid stock), varying the expected return of the illiquid stock μ_I. In the absence of constraints, 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

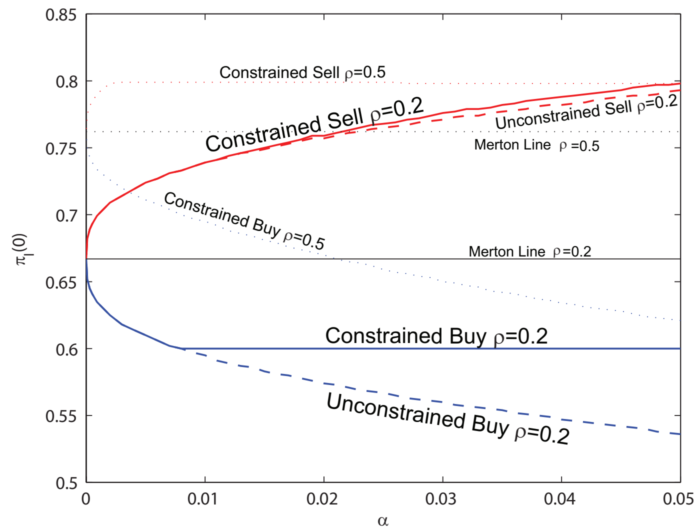

Figure 4. The initial optimal trading strategy for the illiquid asset for a small cap fund against transaction cost rate. Parameter default values: $\gamma = 2$, $T = 5$, $\mu_L = 0.06$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $\rho = 0.2$, $\theta = \alpha$, $\underline{b} = 0.60$, and $\bar{b} = 0.80$.

returns. This binding constraint makes the buy boundary flat at 60% until it gets close the buy boundary for the unconstrained case at $R_I = 4.1\%$. It also changes the sell boundary to be only slightly above the buy boundary to balance the cost from over-investment in the illiquid asset and the transaction cost payment. When the unconstrained buy boundary is at 60%, the unconstrained sell boundary is well below the upper bound 80%. Therefore the portfolio constraints are not binding at time 0. However, Figure 5 shows that the constrained buy boundary is strictly above 60%. To understand this result, recall that by Proposition 3, as time to horizon decreases to 0, the buy boundary decreases to $-1/\alpha$ and the sell boundary increases to $1/\theta$. An upper bound $\bar{b} < 1$ then will 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 sell the illiquid asset and incur transaction costs.

In this sense, the fund's trading strategy is non-myopic with respect to the portfolio constraints in the presence of transaction costs since 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. This nomyopium of the optimal trading strategy with respect to the portfolio constraints is robust and present in all the cases we have numerically solved.

As the excess return increases, the no transaction widens because the cost of over-investment decreases. Between $R_I = 4.1\%$ and $R_I = 4.8\%$, the constraints become not 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 widen the no transaction region. 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 more attractive expected returns. Therefore, if the excess return is low then the fund will shift into the liquid asset and vice versa.

Next we examine more closely the effect of correlation on diversification. In Figure 6, we plot the time 0 optimal fraction of wealth invested in the illiquid asset ($\pi_I(0)$) against the correlation coefficient ρ for different levels of transaction cost rates. Consistent with Figure 4, Figure 6 verifies that for this set of parameter values, as the correlation coefficient increases the optimal fraction of wealth 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

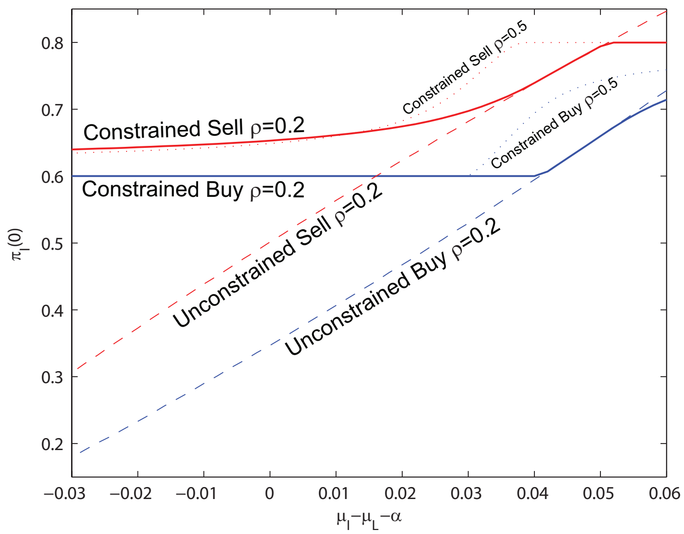

Figure 5. The initial optimal trading strategy for the illiquid asset for a small cap fund against net excess return over the liquid asset. Parameter default values: $\gamma = 2$, $T = 5$, $\mu_L = 0.06$, $\sigma_L = 0.20$, $\sigma_I = 0.25$, $\rho = 0.2$, $\alpha = 0.01$, $\theta = 0.01$, $\underline{b} = 0.60$, and $\bar{b} = 0.80$.

region widens. This is because that the trading in the illiquid asset becomes more costly. Since the fraction of wealth invested in the liquid asset is always equal to $1 - \pi_I$, the fluctuation in the liquid asset investment goes up for higher transaction costs. Also, as the transaction cost increases, 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 $\rho < 0.24$ with $\alpha = \theta = 0.01$. In contrast, if $\alpha = \theta = 0.02$, it remains flat at 60% for all $\rho < 0.31$. The intuition behind this result is again that an increase in transaction costs makes the fund lower the buy boundary and increase the sell boundary.

Next, we analyze the impact of portfolio constraints on the magnitude of the liquidity premium in our model. To simplify analysis, we set $\underline{b} = \pi_I^M - \frac{1}{2}\beta$ and $\bar{b} = \pi_I^M + \frac{1}{2}\beta$, where π_I^M is the Merton line and β measures the portfolio constraint

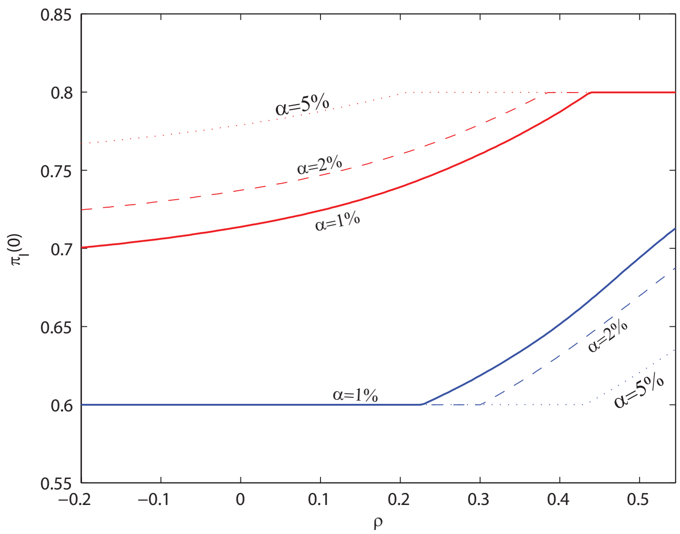

Figure 6. The optimal trading strategy for the liquid asset for a small cap fund against correlation coefficient. Parameter default values: $\gamma = 2$, $T = 5$, $\mu_L = 0.06$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $\theta = \alpha$, $\underline{b} = 0.60$, and $\bar{b} = 0.80$.

stringency. As β decreases, the constraint becomes more stringent. We compute the liquidity premium using a similar approach to that of Constantinides (1986). Specifically, let $v(W; \mu_I, \beta)$ be the value function at time 0 for the case with constraints and transaction costs. Let $J(W; \mu_I, \beta)$ be the value function at time 0 for the case with constraints but without transaction costs. Then we solve $J(W; \mu_I - \delta, \beta) = v(W; \mu_I, \beta)$ for the liquidity premium δ.

In Figure 7, we plot the liquidity premium to transaction cost ratio δ/α against the constraint stringency β. This figure shows that when the constraint is very stringent, the ratio can be much greater than what Constantinides (1986) finds (where it is typically around 0.1). Thus in contrast to Constantinides (1986), transaction costs can have a first-order effect in the presence of stringent portfolio constraints. This is because imposing stringent constraints can force more frequent transactions and also can significantly distort the investment strategy. This figure also shows that for a given constraint bandwidth β, the liquidity premium is much higher if the volatility of the illiquid stock is high. For example, the ratio becomes as high as 0.95 when $\sigma_I = 0.4$ and $\beta = 0.2$. Surprisingly, the dashed line shows that the liquidity premium can increase when the constraints become less stringent (i.e, as β increases). The main reason for this counterintuitive result is the presence of binding constraints. When $\sigma_I = 0.4$, the Merton line for this case is 0.2917. The lower bound at $\beta = 0.02$, for example, is $\underline{b} = 0.6427$ and thus binding in the no transaction cost case. This implies that the value function shifts upward in both the case with transaction costs and the case without transaction costs as β increases. More importantly, as β increases from 0.02 and thus $\underline{b}$ decreases, the investment in the illiquid asset decreases. It follows that it takes a larger reduction in the expected return μ_I to decrease the value function J by the same amount. It is this effect that makes the liquidity premium go up in this region when the constraints become less binding. Figure 7 also shows that as the constraint becomes more relaxed, the liquidity premium starts to decline again. To understand this eventual decreasing pattern, it is helpful to consider the extreme case where the constraints do not bind for the no transaction cost case. Therefore, as the constraints become more relaxed, the value function with transaction costs increases and thus the liquidity premium decreases until the constraints become non-binding even for the case with transaction costs, then the liquidity premium stays constant.

Now we briefly examine the issue of the optimal choice of the constraints 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

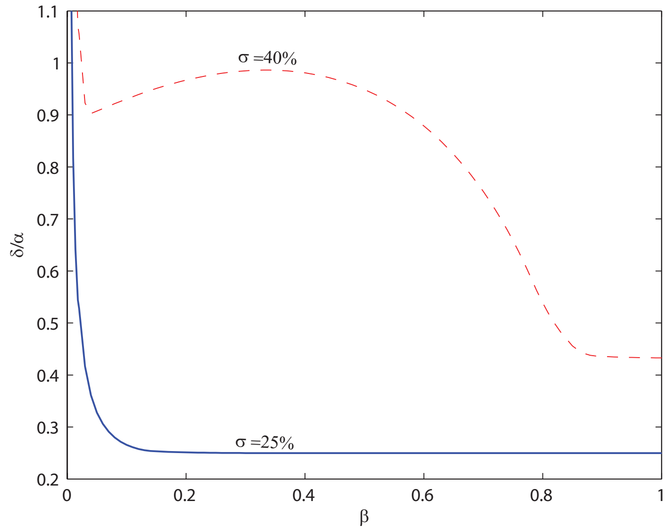

Figure 7. The liquidity premium to transaction cost ratio against weights bandwidth β. Parameter default values: $\gamma = 2$, $T = 5$, $\mu_L = 0.06$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $\rho = 0.2$, $\theta = 0.01$, $\underline{b} = 0.6667 - \frac{1}{2}\beta$, and $\bar{b} = 0.6667 + \frac{1}{2}\beta$.

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 compute the optimal strategy of the fund manager for the constrained case and the unconstrained case;

(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 $V_c(W; \underline{b}, \bar{b})$ and $V_u(W)$;

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

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

We illustrate the optimal choice through two cases: One case where the investor is less risk averse than the manager (Figure 8) and the other case where the investor is more risk averse (Figure 9). Specifically, we set $\gamma_I = 2$ and $\gamma_M = 5$ in Figure 8 and $\gamma_I = 5$ and $\gamma_M = 2$ in Figure 9. This implies that in Figure 8 (Figure 9) the investor would like the manager to invest more (less) in the illiquid stock than what the manager would choose to. So we only consider the imposition of a lower (upper) bound in Figure 8 (Figure 9). Figure 8 plots the ratio Δ/W against $\underline{b}$ and Figure 9 plots the ratio Δ/W against $\bar{b}$ for different correlation coefficients and transaction cost rates, where the stars in the figures indicate where the ratios are maximized. Figure 8 shows that the optimal lower bound $\underline{b}$ is equal to 0.624 in the first case (given default parameter values) and Figure 9 shows that the optimal upper bound $\bar{b}$ is equal to 0.520 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.

Interestingly, these figures also suggest that as the correlation increases while the optimal upper bound decreases, the optimal lower bound increases. This is driving by the fact that how the correlation coefficient affects the trading strategy depends on the magnitude of the risk aversion. If the risk aversion is high, then as the correlation increases, the investor decreases the investment in the illiquid asset to reduce the risk. As correlation increases, portfolio risk increases because diversification is less effective. To counter this adverse effect, an investor can either increase its illiquid asset to achieve a higher expected return or reduce it to decrease risk. If the risk aversion is low, then it is optimal for the investor to increase the investment in the illiquid asset. If the risk aversion is high, then it is optimal for the investor to decrease

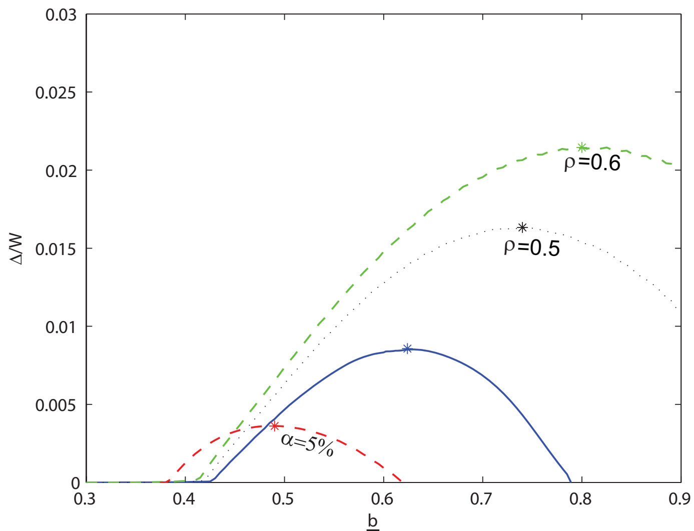

Figure 8. The optimal choice of the lower bound for an investor. Parameter default values: $\gamma_I = 2$, $\gamma_M = 5$, $T = 5$, $\mu_L = 0.06$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $\rho = 0.2$, $\alpha = 0.01$, $\theta = 0.01$, and $\bar{b} = 0.80$.

it. In Figure 8, the investor's risk aversion is low, so as the correlation increases, the optimal level of illiquid asset investment becomes higher, and thus he imposes a higher lower bound. In Figure 9, the investor's risk aversion is high, so as the correlation increases, the optimal level of illiquid asset investment becomes lower, and thus he imposes a lower upper bound.

These figures also demonstrate that the benefit of constraining fund managers can be quite significant. Figure 8 suggest that the investor is willing to pay more than 0.86% of the initial wealth for the right to constrain fund managers. In Figure 9, the gain from imposing portfolio constraints is as high as 2.71%. The right to impose constraints becomes even greater the correlation is high (e.g., 5.4% when $\rho = 0.6$ in Figure 9).

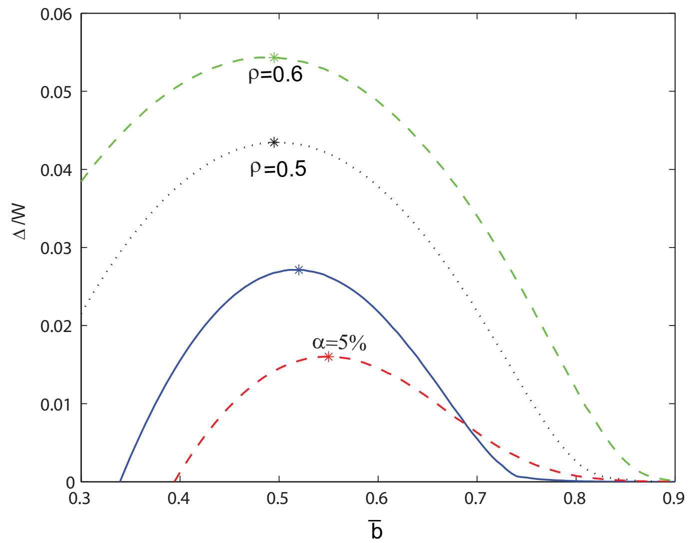

Figure 9. The optimal choice of the upper bound for an investor. Parameter default values: $\gamma_I = 5$, $\gamma_M = 2$, $T = 5$, $\mu_L = 0.06$, $\sigma_L = 0.20$, $\mu_I = 0.11$, $\sigma_I = 0.25$, $\rho = 0.2$, $\alpha = 0.01$, $\theta = 0.01$, and $\underline{b} = 0$.

VI. Conclusions

We use a novel approach to examine the optimal investment problem of a mutual fund who faces transaction costs and portfolio constraints. We show that both the buy boundary and the sell boundary for the illiquid stock are monotonically decreasing in the portfolio bounds.

We find that the presence of portfolio constraints can significantly magnify the effect of transaction costs on liquidity premium and can make it a first-order effect for a reasonable set of parameter values. Surprisingly, the liquidity premium can increase when binding constraints become less stringent. We also show that even for log preferences, the optimal trading strategy is nonmyopic with respect to the constraints, in the sense that currently nonbinding constraints can affect the current optimal trading strategy. In addition, the optimal buy boundary is monotonically decreasing in calendar time. We also examine the endogenous choice of the portfolio bounds. Our analysis shows that the optimal upper is increasing in transaction costs and the optimal lower bound is decreasing in transaction costs.

A APPENDIX

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

A.1 Proof of Theorem 1

Given π_{I_s}, we denote

$$\sigma(s) = \sqrt{\frac{1}{2}\left[(1-\pi_{I_s})^2\sigma_L^2 + 2\rho(1-\pi_{I_s})\pi_{I_s}\sigma_L\sigma_I + \pi_{I_s}^2\sigma_I^2\right]}.$$

Then

$$W_T = W_t e^{\int_t^T [(1-\pi_{I_s})\mu_L + \pi_{I_s}\mu_I - \sigma^2(s)ds + \int_t^T (1-\pi_{I_s})\sigma_L dB_{Ls} + \int_t^T \pi_{I_s}\sigma_I dB_{I_s}}.$$

Therefore

$$\begin{aligned}
& EW_T^{1-\gamma} \\
= \;& W_t^{1-\gamma}E[e^{(1-\gamma)\int_t^T[(1-\pi_{I_s})\mu_L+\pi_{I_s}\mu_I-\sigma^2(s)]ds+\int_t^T(1-\gamma)(1-\pi_{I_s})\sigma_L dB_{Ls}+\int_t^T(1-\gamma)\pi_{I_s}\sigma_I dB_{I_s}}] \\
= \;& W_t^{1-\gamma}E[e^{(1-\gamma)\int_t^T[(1-\pi_{I_s})\mu_L+\pi_{I_s}\mu_I-\gamma\sigma^2(s)]ds}Z(T)] \\
\equiv \;& W_t^{1-\gamma}E[e^{(1-\gamma)\int_t^T f(\pi_{I_s})ds}Z(T)],
\end{aligned}$$

where $Z(\nu) = e^{-\int_t^\nu (1-\gamma)^2\sigma^2(s)ds + \int_t^\nu (1-\gamma)(1-\pi_{I_s})\sigma_L dB_{Ls} + \int_t^\nu (1-\gamma)\pi_{I_s}\sigma_I dB_{I_s}}$ is a nonnegative local martingale, therefore supermartingale with $E[Z(T)] \le Z(t) = 1$, and

$$f(\xi) = (1-\xi)\mu_L + \xi\mu_I - \frac{\gamma}{2}[(1-\xi)^2\sigma_L^2 + 2\rho(1-\xi)\xi\sigma_L\sigma_I + \xi^2\sigma_I^2].$$

Denote $\eta = \max_{x\in\boldsymbol{R},\xi\in[\underline{b},\bar{b}]} f(\xi)$, and denote the maximizer as ξ^*. It is easy to see

$$\xi^* = \pi_I^*, \qquad \eta = f(\pi_I^*).$$

We then deduce

$$\begin{aligned}
EW_T^{1-\gamma} &\le W_t^{1-\gamma}e^{(1-\gamma)\eta(T-t)}E[Z(T)] \\
&\le (W_t e^{\eta(T-t)})^{(1-\gamma)},
\end{aligned}$$

and the equality holds if and only if $\pi_{I_s} \equiv \pi_I^*$. □

A.2 Proof of Proposition 1

We only prove part 1, and the case of part 2 is similar. Let

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

It is not hard to see that $w(\pi, t)$ satisfies

$$\begin{cases} \max\left\{w_t + \mathcal{L}_2 w, -\frac{\alpha}{1-\alpha\pi} - w_\pi, w_\pi - \frac{\theta}{1+\theta\pi}\right\} = 0 \\ w(\pi, T) = 0, \text{ in } \left(-\frac{1}{\theta}, \frac{1}{\alpha}\right) \times [0, T), \end{cases}$$

where

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

Denote

$$v(\pi, t) = w_\pi(\pi, t).$$

Clearly

$$\begin{aligned} & \frac{\partial}{\partial \pi}\left(\mathcal{L}_2 w\right) \\ = \; & \frac{1}{2}\beta_1\pi^2(1-\pi)^2 v_{\pi\pi} + \left[\beta_1 + \beta_2 - (2+\gamma)\beta_1\pi\right]\pi(1-\pi) v_\pi \\ & + \left[\beta_2(1-2\pi) - \gamma\beta_1\pi(2-3\pi)\right] v \\ & + (1-\gamma)\beta_1\pi(1-\pi) v\left[(1-2\pi)v + \pi(1-\pi)v_\pi\right] + \beta_2 - \gamma\beta_1\pi \\ \triangleq \; & \mathcal{L}v. \end{aligned}$$

Using the same technique as in Dai and Yi (2006), we are able to show that $v(\pi, t)$ satisfies the following parabolic double obstacle problems:

$$\begin{cases} v_t + \mathcal{L}v = 0 & \text{if } -\frac{\alpha}{1-\alpha\pi} < v < \frac{\theta}{1+\theta\pi}, \\ v_t + \mathcal{L}v \leq 0 & \text{if } v = -\frac{\alpha}{1-\alpha\pi}, \\ v_t + \mathcal{L}v \geq 0 & \text{if } v = \frac{\theta}{1+\theta\pi}, \\ v(\pi, T) = 0, \end{cases} \tag{A-1}$$

in $\left(-\frac{1}{\theta}, \frac{1}{\alpha}\right) \times [0, T)$.

It is not hard to verify

$$
\begin{aligned}
0 &\geq \left(\frac{\partial}{\partial t}+\mathcal{L}\right)\left(-\frac{\alpha}{1-\alpha\pi}\right)=\frac{1-\alpha}{(1-\alpha\pi)^3}\left[\beta_2-(\gamma\beta_1-\alpha\gamma\beta_1+\alpha\beta_2)\,\pi\right] \\
&= \frac{(1-\alpha)\,\gamma\beta_1}{(1-\alpha\pi)^3}\left[\pi_I^M-\left(1-\alpha+\alpha\pi_I^M\right)\pi\right].
\end{aligned}
$$

Since $1-\alpha+\alpha\pi_I^M>0$, it follows

$$
\overline{\pi}_I(t)\geq\frac{\pi_I^M}{1-\alpha\left(1-\pi_I^M\right)}.
$$

It remains to show that there exists $\bar{t}<T$ such that $\overline{\pi}_I(s)=\frac{1}{\alpha}$ for $s\in(\bar{t},T)$. Suppose not, we then infer $v(\pi,t)$ is smooth across $\pi=\overline{\pi}_I(t)$, $t<T$. It follows

$$
\begin{aligned}
\overline{\pi}_I'(t)v_\pi+v_t\big|_{\pi=\overline{\pi}_I(t)} &= \frac{d}{dt}v\left(\overline{\pi}_I(t),t\right)=\frac{d}{dt}\left(-\frac{\alpha}{1-\alpha\overline{\pi}_I(t)}\right) \\
&= -\frac{\alpha^2\overline{\pi}_I'(t)}{(1-\alpha\overline{\pi}_I(t))^2}=\overline{\pi}_I'(t)v_\pi\big|_{\pi=\overline{\pi}_I(t)},\ \text{for } t<T,
\end{aligned}
$$

which yields

$$
v_t\big|_{\pi=\overline{\pi}_I(t)}=0,\ \text{for } t<T.
$$

On the other hand, $v(\pi,t)$ clearly has a singularity at $\left(\frac{1}{\alpha},T\right)$, i.e.

$$
\lim_{\pi\to\frac{1}{\alpha},\,t\to T} v_t\left(\pi,t\right)=+\infty.
$$

A contradiction. □

A.3 Proof of Proposition 2

Note that the differential operator $\mathcal{L}$ is degenerate at $\pi=0,1$, which yields two ordinary differential inequalities there:

$$
\begin{cases}
v_t(0,t)+\beta_2v(0,t)+\beta_2=0 \text{ if } -\alpha<v(0,t)<\theta \\
v_t(0,t)+\beta_2v(0,t)+\beta_2\leq 0 \text{ if } v(0,t)=-\alpha \\
v_t(0,t)+\beta_2v(0,t)+\beta_2\geq 0 \text{ if } v(0,t)=\theta \\
v(0,T)=0.
\end{cases}
$$

and

$$\begin{cases} v_t(1,t) - \beta_4 v(1,t) + \beta_4 = 0 \text{ if } -\frac{\alpha}{1-\alpha} < v(1,t) < \frac{\theta}{1+\theta} \\ v_t(1,t) - \beta_4 v(1,t) + \beta_4 \leq 0 \text{ if } v(1,t) = -\frac{\alpha}{1-\alpha} \\ v_t(1,t) - \beta_4 v(1,t) + \beta_4 \geq 0 \text{ if } v(1,t) = \frac{\theta}{1+\theta} \\ v(1,T) = 0. \end{cases}$$

Solving them, we then obtain

$$v(0,t) = \begin{cases} e^{\beta_2(T-t)} - 1, \text{ when } t > \bar{t}_0 \\ -\alpha, \text{ when } t \leq \bar{t}_0 \end{cases} \text{ if } \beta_2 < 0 \tag{A-2}$$

$$v(0,t) = \begin{cases} e^{\beta_2(T-t)} - 1, \text{ when } t > \underline{t}_0 \\ \theta, \text{ when } t \leq \underline{t}_0 \end{cases} \text{ if } \beta_2 > 0 \tag{A-3}$$

$$v(0,t) = 0 \text{ if } \beta_2 = 0 \tag{A-4}$$

and

$$v(1,t) = \begin{cases} 1 - e^{-\beta_4(T-t)}, \text{ when } t > \bar{t}_1 \\ -\frac{\alpha}{1-\alpha}, \text{ when } t \leq \bar{t}_1 \end{cases} \text{ if } \beta_4 < 0 \tag{A-5}$$

$$v(1,t) = \begin{cases} 1 - e^{-\beta_4(T-t)}, \text{ when } t > \underline{t}_1 \\ \frac{\theta}{1+\theta}, \text{ when } t \leq \underline{t}_1 \end{cases} \text{ if } \beta_4 > 0 \tag{A-6}$$

$$v(1,t) = 0 \text{ if } \beta_4 = 0. \tag{A-7}$$

Now let us prove part 1. If $\pi_I^M < 0$, then $\beta_2 < 0$ and $\beta_4 < 0$. So, we have (A-2) and (A-5), which implies that $\underline{\pi}_I(t) < 0$ for all t, and $\overline{\pi}_I(t)$ intersects with the lines $\pi = 0$ and $\pi = 1$ at $\bar{t}_0$ and $\bar{t}_1$ respectively. We then infer $\overline{\pi}_I(t) \leq 0$ for $t < \bar{t}_0$, $0 \leq \overline{\pi}_I(t) \leq 1$ for $t \in \left[\bar{t}_0, \bar{t}_1\right]$, and $\overline{\pi}_I(t) \geq 1$ for $t > \bar{t}_1$.

To show the monotonicity of $\overline{\pi}_I(t)$ 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 (1982))

Let v_i, $i = 1, 2$, satisfy

$$\begin{aligned} \frac{\partial v_i}{\partial t} + \mathcal{L}v_i + f_i &= 0, \text{ if } g_i^l < v_i < g_i^u, \\ \frac{\partial v_i}{\partial t} + \mathcal{L}v_i + f_i &\leq 0, \text{ if } v_i = g_i^l, \\ \frac{\partial v_i}{\partial t} + \mathcal{L}v_i + f_i &\geq 0, \text{ if } v_i = g_i^u, \end{aligned}$$

in $\Omega \times [0,T)$. Here $\mathcal{L}$ is an elliptic operator. Assume

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

and

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

Then

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

Note that

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

Applying the comparison principle gives $v_t \ge 0$ in $\{\pi > 0\}$, which yields the desired result.

The proof of part 2-4 is similar for finite T. In part 4, if $T = \infty$, we then follow Dai and Yi (2006) to take into account the corresponding stationary problem, from which we can infer $\underline{\pi}_I(t) = \overline{\pi}_I(t) = 0$ when $\pi_I^M = 0$ and $\underline{\pi}_I(t) = \overline{\pi}_I(t) = 1$ when $\pi_I^M = 1$. $\square$

A.4 Proof of Theorem 2

The uniqueness of viscosity solution can be obtained by using a similar argument in Akian, Mendaldi and Sulem (1996) (see also Crandal, Ishii and Lions (1992)). Here we highlight that on the boundaries the solution is a viscosity supersolution. In terms of the definition of viscosity solution and the 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 (1994)).

Part ii) can be obtained using the same technique as in Dai and Yi (2006). $\square$

A.5 Proof of Proposition 4

1. For the constrained case, we can similarly obtain the following double obstacle problem:

$$\begin{cases} v_t + \mathcal{L}v = 0 & \text{if } -\frac{\alpha}{1-\alpha\pi} < v < \frac{\theta}{1+\theta\pi}, \\ v_t + \mathcal{L}v \le 0 & \text{if } v = -\frac{\alpha}{1-\alpha\pi}, \\ v_t + \mathcal{L}v \ge 0 & \text{if } v = \frac{\theta}{1+\theta\pi}, \\ v(z,T) = 0, \end{cases} \tag{A-8}$$

subject to boundary conditions

$$\begin{cases} v\left(\underline{b}, t\right) = \frac{\theta}{1+\theta\pi}, \\ v\left(\overline{b}, t\right) = -\frac{\alpha}{1-\alpha\pi}, \end{cases} \tag{A-9}$$

in $\left(\underline{b}, \overline{b}\right) \times [0,T)$. Denote the solution of the above problem by $v(\pi, t; \underline{b}, \overline{b})$. Assume $\overline{b}_1 \ge \overline{b}_2$. Since

$$v(\overline{b}_2, t; \underline{b}, \overline{b}_1) \le \frac{\theta}{1+\theta\underline{b}} = v(\overline{b}_2, t; \underline{b}, \overline{b}_2),$$

we get by the comparison principle

$$v(\pi, t; \underline{b}, \overline{b}_1) \le v(\pi, t; \underline{b}, \overline{b}_2) \text{ in } \left(\max\left(\underline{b}, 0\right), \overline{b}_2\right) \times [0,T).$$

So, if $v(\pi, t; \underline{b}, \overline{b}_2) < \frac{\theta}{1+\theta\pi}$, then

$$v(\pi, t; \underline{b}, \overline{b}_1) < \frac{\theta}{1+\theta\pi},$$

which implies $\overline{\pi}_I^c(t; \underline{b}, \overline{b}_1) \ge \overline{\pi}_I^c(t; \underline{b}, \overline{b}_2)$, namely, $\overline{\pi}_I^c(t; \underline{b}, \overline{b})$ is increasing with $\overline{b}$. In a similar way, we can show that $\overline{\pi}_I^c(t; \underline{b}, \overline{b})$ is increasing with $\underline{b}$ and $\underline{\pi}_I^c(t; \underline{b}, \overline{b})$ is increasing with $\overline{b}$ and $\underline{b}$.

2 and 3. We will only prove one case $\overline{b} > 0$, and the other cases are similar. As in the proof of Proposition 2, we can still derive the boundary conditions (A-2)-(A-4) at $\pi = 0$. So, we can deal with the problem in $\{\pi \le 0\}$ and $\{0 \le \pi \le \overline{b}\}$ independently. This yields the desired result. □

A.6 Proof of Propositions 5 and 6

Still, we use the fact that the problem can be dealt with in $\{\pi \leq 0\}$, $\{0 \geq \pi \leq 1\}$ and $\{\pi \geq 1\}$ independently. The remaining argument is similar to the proof of Proposition 2. $\square$

References

Akian, Marianne, Jose Luis Menaldi, and Agnes Sulem, 1996, "On an investment-consumption model with transaction costs," *SIAM Journal on Control and Optimization* 34, 329–394.

Almazan, Andres, Keith C. Brown, Murray Carlson, and David A. Chapman, 2004, "Why constrain your mutual fund manager?" *Journal of Financial Economics* 73, 289-321.

Basak, S., A. Pavlova, and A. Shapiro, 2006, "Optimal asset allocation and risk shifting in money management," forthcoming, *Review of Financial Studies*.

Carpenter, J., 2000, Does option compensation increase managerial risk appetite, *Journal of Finance* 55, 2311–2331.

Chalmers, John M. R. , Roger M. Edelen, and Gregory B. Kadlec, "An analysis of mutual fund trading costs," working paper, University of Pennsylvania.

Chevalier, Judith, and Glen Ellison, 1997, "Risk taking by mutual funds as a response to incentives," *Journal of Political Economy* 105, 1167-1200.

Clarke, Roger, Harindra de Silva, and Steven Thorley, 2002. "Portfolio constraints and the fundamental law of active management," *Financial Analysts Journal* 58, 48-66.

Constantinides, George M., 1986, "Capital market equilibrium with transaction costs," *The Journal of Political Economy* 94, 842–862.

Cuoco, Domenico, 1997, "Optimal consumption and equilibrium prices with portfolio constraints and stochastic income," *Journal of Economic Theory* 72, 33–73.

Cuoco, Domenico and Hong Liu, 2000, "A martingale characterization of consumption choices and hedging costs with margin requirements, " *Mathematical Finance* 10, 355–385.

Cvitanić, Jasak and Ioannis Karatzas, 1992, "Convex duality in constrained portfolio optimization," *Annals of applied probability* 2, 767–818.

Dai, Min, and Fahuai Yi, 2006, "Finite horizon optimal investment with transaction costs: a parabolic double obstacle problems," working paper, National University of Singapore.

Dai, Min, Lishang Jiang and Fahuai Yi, 2006, "A parabolic variational inequality with gradient constraint arising from optimal investment and consumption with transaction costs," working paper, National University of Singapore.

Dai, Min, Yue Kuen Kwok and Jianping Zong, 2007, Guaranteed minimum withdrawal benefit in variable annuities, forthcoming, *Mathematical Finance*.

Davis, Mark H.A., and Andrew R. Norman, 1990, "Portfolio selection with transaction costs," *Mathematics of Operations Research* 15, 676–713.

Delib, Daniel N., and Raj Varma, 2002, "Contracting in the investment management industry: evidence from mutual funds," *Journal of Financial Economics* 63, 79–98.

Fleming, Wendell H. and H. Mete Soner, 1993, *Controlled Markov Processes and Viscosity Solutions*, New York, Springer.

Forsyth, Peter A. and K. R. Vetzal, 2002, "Quadratic convergence of a penalty method for valuing American options," *SIAM Journal of Scientific Computations* 23, 2095–2122.

Friedman, Avner, 1982, *Variational principles and free-boundary problems*, Wiley, New York.

Huang, Lixin and Hong Liu, 2007, "Rational inattention and portfolio selection," forthcoming, *Journal of Finance.*

Liu, Hong, 2004, "Optimal consumption and investment with transaction costs and multiple risky assets," *Journal of Finance* 59, 289–338.

Liu, Hong, and Mark Loewenstein, 2002, "Optimal portfolio selection with transaction costs and finite horizons," *The Review of Financial Studies* 15, 805–835.

Merton, Robert C., 1971, "Optimum consumption and portfolio rules in a continuous-time model," *Journal of Economic Theory* 3, 373–413.

Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions, 1992, "User's guide to viscosity solutions of second order partial differential equations," *Bulletin of The American Mathematical Society* 27, 1–67.

Royden, Halsey L., 1988, *Real Analysis*, Macmillan, New York.

Shreve, Steven E., and H. Mete Soner, 1994, "Optimal investment and consumption with transaction costs," *The Annals of Applied Probability* 4, 609–692.