Fast or Slow: Optimal Trading Strategies with Speed-Dependent Transaction Cost*

Hong Liu

Washington University in St. Louis and CHIEF

Jing Xu

Renmin University of China

This Version: February 24, 2021

*For helpful comments, we thank Zhi Da, Phil Dybvig, Tony He, and seminar participants at the 2018 Asian Quantitative Finance Conference, Renmin University of China, and Washington University in St. Louis. Authors can be reached at liuh@wustl.edu and jing.xu@ruc.edu.cn.

Fast or Slow: Optimal Trading Strategies with Speed-Dependent Transaction Cost

Abstract

The effective transaction cost rates (TCRs) facing large institutional investors often depend on their trading speeds. We propose a continuous time work-horse model to study optimal trading strategies with speed-dependent TCRs. Unlike the existing literature, our model allows the TCRs to be a general function of trading speeds. We apply our framework to the optimal portfolio rebalancing problem and the optimal liquidation/acquisition problem of financial institutions. Our model implies the commonly observed order-shredding strategy and U-shaped trading speeds against time. We show that adopting some naive trading strategies can be economically costly.

Journal of Economic Literature Classification Numbers: C02, G11

Keywords: Portfolio Choice, Transaction Cost, Liquidation, Acquisition

1 Introduction

Transaction costs are prevalent in financial markets and economically important for investors. For example, Wermers (2000) concludes that transaction costs drag down net mutual fund returns by as much as 0.8%, about the same impact as fund expenses. The prevalence of turnover constraints for mutual funds also suggests the importance of transaction costs (e.g., Clarke et al. (2002)). Prior studies on portfolio selection with transaction costs assume that the transaction cost rate (TCR) at which an investor pays the trading cost per share is independent of the trading speed.[1] While this assumption is reasonable for retail investors who trade a small number of shares, it does not represent well the transaction cost structure faced by large institutional investors, because they consume a significant portion of market liquidity and thus the effective TCR they face can increase significantly with trading speeds (e.g., Keim and Madhavan (1996), Dufour and Engle (2000), Almgren et al. (2005)).[2]

To fill this gap, in this paper, we propose a portfolio choice model for a financial institution ("fund" hereafter) that aims to maximize its expected utility from final wealth in the presence of speed-dependent transaction costs. The fund can invest in a risk-free bond and a risky stock. Unlike most of the existing literature, the transaction cost rate for the stock in our model is a general step function of the trading speed, which is consistent with the cost structure reflected in typical limit order books and what is found in the empirical literature (e.g., Niemeyer and Sandas (1993) and Weber and Rosenow (2005)). In addition, because it is largely nonparametric, it is suitable for calibrations to approximate a general function of the trading speed.[3] As a result of the speed-dependent

[1] See, for example, Constantinides (1986), Davis and Norman (1990), Shreve and Soner (1994), Vayanos (1998), Liu and Loewenstein (2002), Liu (2004), Lo, Mamaysky, and Wang (2004), Lynch and Tan (2010, 2011), Dai, Jin, and Liu (2011), Chen and Dai (2013), Garleanu and Pedersen (2013), Dai et al. (2016), Chen et al. (2020).

[2] For example, if a large investor trades against a limit order book or negotiates a block trade with a counterparty, she will have to trade at worse prices as her trading size per unit of time increases to compensate the liquidity providers for taking on increased risks.

[3] For example, studies on price impact of large orders usually assume the magnitude of price impact follows a power law. The step transaction cost rate function used in our model can be used to approximate

transaction cost rates, the fund optimally chooses its trading speed dynamically to trade off risk exposure, risk premium, and transaction costs.

We first apply the model to an optimal portfolio rebalancing problem of a fund that maximizes its expected utility from the before-transaction-cost asset under management (AUM) at a fixed terminal time. This is motivated by the fact that the compensation of fund managers often increases linearly with the before-transaction-cost AUM. The optimal trading strategy in this application displays the following decreasing trading speed pattern: when the risk exposure is far from the optimal risk exposure in the absence of transaction costs (i.e., the Merton line), the fund trades at an infinite speed (i.e., performs a lump sum trade); when the risk exposure gets close enough to the Merton line, the fund reduces trading speed in steps until the risk exposure is sufficiently close to the Merton line so that the fund stops trading. This trading pattern is consistent with the order-shredding behavior for large orders commonly observed in practice.[4] In contrast, in the existing literature on portfolio choice with linear transaction costs, it is optimal to either trade the full amount at the infinite speed or not to trade at all, and thus the existing literature cannot explain the order-shredding strategy.

The intuition behind the above results can be explained as follows. When choosing its trading speed, the fund optimally trades off the risk exposure effect (i.e., bearing the target level of the market price risk), the risk premium effect (i.e., earning higher returns by holding stocks rather than holding cash), and the transaction cost effect (i.e., reducing transaction costs). The risk exposure effect makes the fund trade fast when the exposure is far from the target position (i.e., the Merton line), the risk premium effect increases the purchasing speed but reduces the selling speed, while the transaction cost effect reduces the trading speed. Thus, the optimal trading speed depends on the time

well the power law function. In addition, if a large investor is trading against a limit order book, then the transaction cost rate is exactly a step function of trading speed (see, e.g., Niemeyer and Sandas (1993) and Weber and Rosenow (2005)).

[4] See, for example, Chan and Lakonishok (1995), Keim and Madhavan (1995, 1996), and Kervel and Menkveld (2019).

to the investment horizon, the distance between the current position and the target, the market liquidity, the expected return, and the volatility. When the fund over-invests in the stock, it would like to sell to reduce risk exposure as soon as possible, but because transaction cost rates increase with the trading speed, the fund's optimal selling speed is determined by the balance between the benefit and the cost. In particular, it depends on the degree of overinvestment and the number of shares to sell in order to reach an optimal exposure. The more the fund overinvests, the stronger the exposure effect, and hence the faster the fund should sell. Similar intuition applies to the case where the fund underinvests in the stock.

We then apply the model to an optimal liquidation/acquisition problem of a financial institution. This application is motivated by the observation that institutional investors often need to trade a large number of shares by a certain date in markets where liquidity is limited and trading is costly[5] and that the related literature typically assumes special functional forms for the transaction cost rates that may not represent well the transaction cost structure in practice.[6] Thus, it is important to understand how financial institutions should optimally execute large trades in these markets where the transaction cost structure is not well captured by the existing functional forms considered in the literature.

In particular, we consider a fund that needs to liquidate or acquire a large number of shares of a risky asset by a fixed date.[7] The optimal liquidation/acquisition strategy implied by our model can be described as follows. When the fund's position is far away from the target, it is optimal for the fund to first trade a lump sum of the shares (i.e.,

[5] Even for relatively liquid markets, significant cost incurred by large block trading has also been found. See, for example, Biais, Hillion, and Spatt (1995), Almgren et al. (2005), and Coval and Stafford (2007). For an estimate of the effective trading cost incurred by institutional investors, see, for example, Chan and Lakonishok (1997).

[6] See, for example, Almgren and Chriss (2000), Ting, Warachka, and Zhao (2007), and Gatheral, Schied and Slynko (2012).

[7] Specifically, as an example, we assume the fund's target is to liquidate (or acquire) 1 million shares of a small cap stock in a week.

at an infinite speed), then at the highest finite speed, then at a lower speed, and so on, and finally liquidate/acquire all the remaining shares (if any) on the terminal date to reach the target. Thus, the optimal execution strategy also displays the order-shredding pattern as in the optimal rebalancing problem.[8] Loosely speaking, when the number of shares to be liquidated/acquired is large, the optimal trading speed implied by our model is a U-shape: trading fast first, then slow, and then fast again. These results also apply to the optimal trading of informed traders.

The economic forces behind the above results for the optimal liquidation/acquisition problem are similar to those for the optimal portfolio rebalancing problem. We take the optimal liquidation problem as an example. When the number of shares remaining is too large relative to the target, the risk exposure effect dominates the transaction cost effect and the risk premium effect, and thus the fund sells fast. As the distance from the target decreases, the risk exposure effect decreases, and thus the fund decreases its trading speed to reduce the transaction cost payment and earns the higher expected return along the way. Different from the optimal rebalancing problem, there is a target level that the fund must achieve on the terminal date. As a result, when the time-to-horizon is short, the fund starts to increase its trading speed to reach the target in time to save transaction costs from a large lump sum trade at the terminal date.[9] In contrast, in the optimal rebalancing problem, the fund does not have such a target level to achieve and thus its trading speed does not increase as the time-to-horizon shrinks.[10]

We also calculate the costs incurred from adopting some suboptimal trading strategies. For the optimal portfolio rebalancing problem, we show that assuming a constant

[8] If, on the other hand, the fund has an initial position close to the target or the fund is risk-neutral (i.e., it maximizes expected net revenue from liquidation or minimizes expected net cost of acquisition), it can be optimal to directly start trading at a finite speed to reduce the transaction costs.

[9] If the risk premium effect dominates, it can even be optimal to trade against the intended target when the time-to-horizon is long and the market condition is favorable. For example, purchasing shares before the terminal date can be optimal in a liquidation program if the stock price is sufficiently low.

[10] If the fund manager's compensation was based on the after-liquidation AUM at the investment horizon, then it would also increase trading speed towards the end of the horizon.

transaction cost rate, as does the existing literature, can be costly for a fund that aims to maximize its expected utility from the final wealth. For example, in our numerical analysis, for a fund with an investment horizon of 5 years and zero initial stock allocation, such an incorrect assumption can result in a certainty equivalent wealth loss (CEWL) of 104 basis points in terms of the fund's initial wealth. For the optimal liquidation/acquisition problem, as noted by Bertsimas and Lo (1998), some naive trading strategies, such as trading at a constant speed over the execution horizon or liquidating/acquiring all position instantly to avoid any market price risk, are not uncommon in practice. We find, for example, for a relatively risk-averse fund with an absolute risk aversion coefficient of 5×10^{-6}, when the fund aims at liquidating 1 million shares within one week, liquidation at a constant speed can result in a loss of about 5,000 shares, while the immediate complete liquidation can cost the fund about 4,000 shares.[11] These results clearly suggest the economic importance of adopting the optimal trading strategy.

Our optimal liquidation model nests a special case in which the fund's objective is to maximize its expected revenue from liquidation, because it is equivalent to assuming the fund is risk neutral. Because a risk-neutral fund does not care about price risk, to reduce transaction costs and earn stock risk premium, the fund tends to liquidate shares faster only when the final horizon is approaching. This is in contrast to the risk-averse fund which tends to liquidate faster at the beginning of the liquidation horizon. We find that the risk-averse fund can achieve a substantially higher Sharpe ratio. This suggests that the optimal liquidation strategy derived from our model can be useful for managing revenue risk.

Beyond the analysis presented in this paper, we believe that the approach proposed here can find potential applications in many interesting cases when transaction costs are present. For example, how should a large investor optimally change positions in response

[11] Note that these costs are accrued within a short period of one week, and they can be even larger if the fund's trading horizon becomes longer.

to changing market conditions? How should a fund liquidate large blocks of managed assets to meet its investors' redemption flows, or build up new large positions with its investors' inflows? How should a CEO liquidate her granted stocks when they vest?

The remainder of the paper proceeds as follows. We briefly review the studies related to ours in the next section. We present our general theoretical framework in Section 3. In Section 4, we apply our framework to an optimal portfolio-rebalancing problem and an optimal liquidation/acquisition problem faced by a fund, and provide a comprehensive analysis of the solutions. We conclude in Section 5. All proofs and technical details are in the Appendix.

2 Related Literature

Our paper contributes to the large body of literature on optimal portfolio choice with transaction costs (e.g., Constantinides (1986); Davis and Norman (1990); Shreve and Soner (1994); Vayanos (1998); Liu and Loewenstein (2002); Liu (2004); Lo, Mamaysky, and Wang (2004); Lynch and Tan (2010, 2011); Dai, Jin, and Liu (2011); Chen and Dai (2013); Garleanu and Pedersen (2013); Dai et al. (2016); Chen et al. (2020)). These studies usually assume the investor's trading incurs transaction costs at a constant rate. As a result, the optimal strategies only involve trading at an infinite speed or zero speed (i.e., not trade at all).[12] One implication of these models is that a large order is executed in an all or nothing manner, which is inconsistent with the wide-spread order shredding practice. Our paper generalizes existing models by allowing the transaction cost rates to be dependent on trading speeds. As a result, our model predicts that a large order may be divided into smaller orders and executed gradually over time. In addition, the model is able to capture a more flexible transaction cost structure observed in the trading data.

[12] Lynch and Tan (2011) allow the transaction cost to be state-dependent, but not trading speed–dependent.

Price impact can be an important source of trading costs. Although we do not attempt to directly model the dynamics of price impact, our model can represent the case of temporary price impact in an illiquid market. This is because (1) the magnitude of price impact usually increases in the size of trade, and (2) if price reverts quickly after a trade, then the realized cost due to price impact is similar to transaction costs. Therefore, our paper is also related to the rich literature on optimal trade execution with price impact, including Cuoco and Cvitanić (1998), Almgren and Chriss (2000), Ting, Warachka, and Zhao (2007), Gatheral, Schied, and Slynko (2012), Obizhaeva and Wang (2013), Lokka (2014), Gueant and Lehalle (2015), Kratz and Schoneborn (2015, 2018), Curato, Gatheral, and Lillo (2017), and Tsoukalas, Wang, and Giesecke (2019). These studies typically assume some parametric forms of the price impact function. By contrast, the transaction cost rate function in our model is largely nonparametric and flexible. It allows for a concave impact function for one range of trading speeds but a convex one for another range. This makes our model more easily calibrated to data, where the realized cost due to price impact typically displays this concave–convex shape (see, e.g., Niemeyer and Sandas (1994), Keim and Madhavan (1997), and Weber and Rosenow (2005)). Unlike some models with permanent or transient (but decaying at a finite rate) price impact (e.g., Huberman and Stanzl (2004)), no price manipulation exists in our model, because in our model trades only affect the transaction price for the large trader but do not affect the market price.

The optimal liquidation/acquisition speed derived from our model is a U-shaped function of time. Obizhaeva and Wang (2013) examine optimal trading execution in a limit order book with time-varying supply and demand. The optimal trade size implied by their model is also a U-shaped function of time. However, the mechanism of their model is fundamentally different from ours. In their model, the purpose of making a large trade at the initial time is to push the order distribution away from its steady state so that liquidity providers can be induced to trade. By comparison, in our model, the purpose of conducting a large initial trade is to achieve a certain desirable risk exposure to the risky asset as soon as possible. In addition, in Obizhaeva and Wang's (2013) model, after the initial large trade, an investor makes another large trade only at the terminal time to reach the target. In contrast, in our model, it can be optimal to increase trading speed before the terminal date, as this allows the fund to reduce transaction costs that would potentially be incurred by a final lump sum trade.

3 The General Framework

In this section, we describe our general framework.

3.1 Assets and Transaction Cost Structure

Throughout this paper we assume a complete probability space $(\Omega, \mathcal{F}, P)$ on which a filtration $\{\mathcal{F}_t\}$ is defined. We consider a financial institution (fund) that can invest in two assets. The first asset ("the bond" hereinafter) is a money market account growing at a continuously compounded, constant rate of r. The second asset ("the stock" hereinafter) is risky. The stock price S_t follows

$$dS_t = \mu S_t dt + \sigma S_t dw_t, \tag{1}$$

where constants μ is the expected return and $\sigma > 0$ is the volatility, and w_t is a standard Brownian motion process adapted to the filtration $\{\mathcal{F}_t\}$.

We assume that the fund does not have private information about the stock and the stock market liquidity is limited. Unlike the existing literature, we allow the transaction cost rate to be increasing in the fund's trading speed, measured by the number of shares traded per unit of time. However, because the fund does not have private information about the fundamentals of the stock, its trades do not have any permanent price impact.

In the presence of transaction cost to be modeled below, the fund's trading strategy can be characterized by two nondecreasing, right continuous adapted processes D_t and I_t, with $D_{0=} = I_{0-} = 0$, representing the cumulative *number of shares* sold and purchased up to time t, respectively. We assume that to purchase dI_t shares of the stock, the fund has to pay a total dollar amount of

$$\left(1 + \sum_{i=0}^{n_1} \left(\theta_i 1_{\{dI_t > \eta_i dt\}}\right)\right) S_t dI_t, \tag{2}$$

where $n_1 \geq 0$ is a nonnegative integer; $0 = \eta_0 < \eta_1 < ... < \eta_{n_1} < \infty$ represents the sequence of the "threshold" purchasing speeds, beyond which purchasing shares will incur transaction costs at greater rates; the vector $(\theta_1, ..., \theta_{n_1})$, with $\theta_j \geq 0$ and $\theta \equiv \sum_{j=0}^{n_1} \theta_j < \infty$, measures the magnitude of transaction cost incurred by purchasing; and $\sum_{j=0}^{i} \theta_j$ represents the transaction cost rate when the fund's purchasing speed satisfies $\eta_i dt < dI_t \leq \eta_{i+1} dt$. Similarly, we assume that to sell dD_t shares of the stock, the fund receives a net proceed of

$$\left(1 - \sum_{i=0}^{n_2} \left(\alpha_i 1_{\{dD_t > \xi_i dt\}}\right)\right) S_t dD_t, \tag{3}$$

where $n_2 \geq 0$ is a nonnegative integer; $0 = \xi_0 < \xi_1 < ... < \xi_{n_2} < \infty$ represents the sequence of the "threshold" selling speeds, beyond which selling will incur transaction cost at greater rates; the vector $(\alpha_1, ..., \alpha_{n_2})$, with $\alpha_j \geq 0$ and $\alpha \equiv \sum_{j=0}^{n_2} \alpha_j < 1$, measures the magnitude of transaction cost incurred by selling; and $\sum_{j=0}^{i} \alpha_j$ represents the transaction cost rate when the fund's selling speed satisfies $\xi_i dt < dD_t \leq \xi_{i+1} dt$.[13]

Discussion of the Model. The piecewise linear transaction cost structure with positive cost coefficients stated above implies positive monotonicity of the transaction cost

[13] $(1 - \alpha_0)S_t$ and $(1 + \theta_0)S_t$ represent the bid and ask, respectively. We show later that, although the fund can trade at all nonnegative speeds, the optimal strategy will only involve trading at these "threshold" speeds.

rate in the trading speed, which is consistent with all the empirical findings on trading cost patterns.[14] In addition, the transaction cost function is largely nonparametric, and thus can be calibrated to approximate any monotonic transaction cost structure. With this generality, we do not need to assume a constant or concave or convex cost structure, as the existing literature does. In particular, our model can be applied to cases where transaction cost is convex in a range but concave in a different range, as shown by the empirical literature.[15]

Our model generalizes the existing portfolio choice models with proportional transaction costs, such as Davis and Norman (1990) and Liu and Loewenstein (2002), because effectively, these papers only allow either an infinite speed or a zero speed. If we set $n_1 = n_2 = 0$ and $\alpha_0 + \theta_0 > 0$, then our model reduces to the standard proportional transaction cost model. If $n_1 = n_2 = 0$ and $\alpha_0 = \theta_0 = 0$, then our model further reduces to the standard Merton model with no friction (e.g., Merton (1971)). Our model also generalizes portfolio choice models with a finite trading speed, such as Longstaff (2001), to include multiple trading speeds and corresponding transaction costs.

3.2 The Trading Problem and Solution Approach

Let x_t be the dollar amount invested in the bond, and y_t be the number of stock shares held by the fund, both at time t. When $\theta + \alpha > 0$, the above model gives rise to the following equations that govern the evolution of x_t and y_t:

$$dx_t = rx_{t-}dt - \left(1 + \sum_{i=0}^{n_1} \theta_i 1_{\{dI_t > \eta_i dt\}}\right) S_t dI_t + \left(1 - \sum_{i=0}^{n_2} \alpha_i 1_{\{dD_t > \xi_i dt\}}\right) S_t dD_t, \quad (4)$$

$$dy_t = dI_t - dD_t. \quad (5)$$

[14] It can be easily shown that if transaction cost decreases in the trading speed, the fund will always trade at the speed that corresponds to the lowest transaction cost, and therefore one can ignore the speeds at which the transaction cost is decreasing without loss of generality.

[15] In the baseline calibration that we adopt from Keim and Madhavan (1997), the transaction cost rates are concave for low speeds and convex for high speeds. See the analyses in Section 4.

Given the initial positions and stock price (x_{0-}, y_{0-}, S_0), the fund's problem is to choose the optimal trading strategies $(I^*, D^*) \equiv \{(I_t^*, D_t^*) : 0 \leq t \leq T\}$, among all admissible trading strategies,[16] to maximize the expected utility derived from the wealth level at time T, that is,

$$E\left[u(\phi(x_T, y_T, S_T))\right], \tag{6}$$

subject to budget constraints (4), (5), and stock price evolution (1), where $u(\cdot)$ is an increasing and concave utility function, and $\phi(x_T, y_T, S_T)$ is the fund's terminal wealth level.[17]

We denote the fund's value function as $V(x, y, S, t)$. Then, the associated Hamilton-Jacobi-Bellman (HJB) equation can be characterized as follows. When it is optimal to buy the stock at an infinite speed (i.e., buy a lump sum of shares at once), using an argument similar to Shreve and Soner (1994), we have

$$V_y - (1+\theta)SV_x = 0. \tag{7}$$

Similarly, when it is optimal to sell the stock at an infinite speed (i.e., sell a lump sum of shares at once), we have

$$(1-\alpha)SV_x - V_y = 0. \tag{8}$$

When it is optimal to trade at a finite speed (including the zero speed), we let $dI_t = i_t dt$ and $dD_t = d_t dt$, with $0 \leq i_t \leq \eta_{n_1}$ and $0 \leq d_t \leq \xi_{n_2}$. In this case, we have

$$\mathcal{M}V + \sup_{0\leq i\leq \eta_{n_1}} [V_y - (1+\sum_{j=0}^{n_1-1} \theta_j 1_{\{i>\eta_j\}})SV_x]i + \sup_{0\leq d\leq \xi_{n_2}} [(1-\sum_{j=0}^{n_2-1} \alpha_j 1_{\{d>\xi_j\}})SV_x - V_y]d = 0, \tag{9}$$

[16] Here, admissible trading strategies refer to those that preclude arbitrage opportunity.

[17] We do not specify the functional form of $\phi(x, y, S)$ at this moment. It will be specified in concrete applications later. Moreover, our model can be extended to the case with intertemporal consumption. The inclusion of intertemporal consumption may decrease the transaction costs paid at the terminal date, but would not change our qualitative results. Therefore, we focus our analyses on the simpler case without consumption to reveal the fundamental trade-off between speedy trading and trading cost.

where the supremum over i and d indicates that the fund can optimally choose any buying speed in $[0, \eta_{n_1}]$ and any selling speed in $[0, \xi_{n_2}]$ (not necessarily the threshold speeds), and

$$\mathcal{M}V = rxV_x + \mu SV_S + \frac{1}{2}\sigma^2 S^2 V_{SS} + V_t. \tag{10}$$

Equations (7), (8), and (9) form a linear complementarity problem (LCP). It is convenient for later analysis to rewrite this LCP in the following compact form

$$\begin{aligned}
&\max\Bigg\{\mathcal{M}V + \sup_{0\le i\le \eta_{n_1}} [V_y - (1 + \sum_{j=0}^{n_1-1} \theta_j 1_{\{i>\eta_j\}})SV_x]i \\
&+ \sup_{0\le d\le \xi_{n_2}} [(1 - \sum_{j=0}^{n_2-1} \alpha_j 1_{\{d>\xi_j\}})SV_x - V_y]d, \\
&V_y - (1+\theta)SV_x,\ \ (1-\alpha)SV_x - V_y\Bigg\} = 0.
\end{aligned} \tag{11}$$

Lastly, the value function also satisfies the following terminal condition

$$V(x, y, S, T) = u(\phi(x, y, S)). \tag{12}$$

Identify the Optimal Trading Speed. Now, we discuss how to identify the optimal trading speed, given the solution to the HJB equations (11)–(12). Intuitively, conditional on trading, the fund should choose the trading speed that maximizes the marginal utility gains. Therefore, the procedure of finding the optimal selling or buying strategy (D^*, I^*) involves a linear search that can be formally described as follows: First, if $V_y - (1+\theta)SV_x = 0$, then it is optimal to buy at infinite speed. On the other hand, if there exists an integer k_1 such that $1 \le k_1 \le n_1$ and

$$(1 + \sum_{j=0}^{k_1-1} \theta_j)SV_x \le V_y < (1 + \sum_{j=0}^{k_1} \theta_j)SV_x, \tag{13}$$

then $i_t^* = \eta_{k^*}$, where

$$k^* = \arg \max_{1 \le k \le k_1} \left\{ \left[V_y - (1 + \sum_{j=0}^{k-1} \theta_j) SV_x \right] \eta_k \right\}. \tag{14}$$

This is because inequality (13) implies that the fund should not buy stock shares at any speed higher than η_{k_1}, and thus it should choose a speed between 0 and η_{k_1} that generates maximal utility gains, as indicated by equation (14). Finally, the fund should not buy at all, if $V_y - (1 + \theta_0) SV_x < 0$.

Similarly, it is optimal to sell at infinite speed if $-V_y + (1 - \alpha) SV_x = 0$. Otherwise, if there exists an integer k_2 such that $1 \le k_2 \le n_2$ and

$$(1 - \sum_{j=0}^{k_2} \alpha_j) SV_x < V_y \le (1 - \sum_{j=0}^{k_2 - 1} \alpha_j) SV_x, \tag{15}$$

then $d_t^* = \xi_{k^*}$, where

$$k^* = \arg \max_{1 \le k \le k_2} \left\{ \left[-V_y + (1 - \sum_{j=0}^{k-1} \alpha_j) SV_x \right] \xi_k \right\}. \tag{16}$$

The fund should not sell, if $-V_y + (1 - \alpha_0) SV_x < 0$.

Next, we provide a verification theorem that formally characterizes the optimal strategy and verifies its optimality.

Proposition 1. *(Verification theorem.) Let $v(x, y, S, t)$ be a sufficiently smooth solution to equation (11) with the terminal condition (12). Define the following regions:*

$$URS = \{(x, y, S, t) : (1 - \alpha) Sv_x - v_y = 0\}, \tag{17}$$

$$RS_k = \{(x, y, S, t) : \mathcal{M}v + [(1 - \sum_{j=0}^{k-1} \alpha_j) Sv_x - v_y] \xi_k = 0\}, \tag{18}$$

$$URB = \{(x, y, S, t) : v_y - (1 + \theta) Sv_x = 0\}, \tag{19}$$

$$RS_k = \{(x, y, S, t) : \mathcal{M}v + [v_y - (1 + \sum_{j=0}^{k-1} \theta_j) Sv_x] \eta_k = 0\}, \tag{20}$$

where the operator $\mathcal{M}$ is defined in (10).

Given the initial bond position x_0, the initial number of stock shares y_0, and the initial stock price S_0, let $\hat{d} = \inf\{d \geq 0 : (x_0 + (1-\alpha)d, y_0 - d, S_0, 0) \notin int(URS)\}$ and $\hat{i} = \inf\{i \geq 0 : (x_0 - (1+\theta)i, y_0 + i, S_0, 0) \notin int(UBS)\}$, where $int(A)$ refers to the interior of set A. Then, the optimal trading strategy is

$$D_t^* = D_t^\infty + D_t^f, \quad I_t^* = I_t^\infty + I_t^f, \tag{21}$$

where

$$D_t^\infty = \hat{d} + \int_0^t 1_{\{(x_u, y_u, S_u, u) \in \partial URS\}} dD_t^\infty, \tag{22}$$

$$D_t^f = \int_0^t \sum_{j=1}^{n_2} 1_{\{(x_u, y_u, S_u, u) \in RS_j\}} \xi_j dt, \tag{23}$$

$$I_t^\infty = \hat{i} + \int_0^t 1_{\{(x_u, y_u, S_u, u) \in \partial URB\}} dI_t^\infty, \tag{24}$$

$$I_t^f = \int_0^t \sum_{j=1}^{n_1} 1_{\{(x_u, y_u, S_u, u) \in RB_j\}} \eta_j dt. \tag{25}$$

Moreover, $v(x, y, S, t)$ coincides with the value function $V(x, y, S, t)$.

In Proposition 1, $\hat{d}$ ($\hat{i}$, resp.) is the initial lump sum sale (purchase, resp.). It is worth noting that the optimal trading policy implied by our model involves two distinct components: the singular control component (I_t^∞, D_t^∞), which characterizes trading at infinite speed, and the regular control component (I_t^f, D_t^f), which characterizes trading at a finite speed. This is in contrast to the portfolio choice model with constant proportional transaction cost rates, in which the optimal trading policy only consists of the singular control component. Moreover, Proposition 1 suggests that multiple free boundaries can arise endogenously to separate trading regions with different trading speeds.

So far, we have proposed a general framework for studying optimal portfolio choice problems with speed-dependent transaction cost rates. Our setting is mostly relevant for financial institutions, which tend to trade in large quantities, making the variations in the transaction cost rates important. In the next section, we examine two applications of our framework.

4 Applications

We consider optimal trading of a small cap stock with 25 million shares outstanding. We adopt the estimates of trading cost for small cap stocks, at NYSE and AMEX, in Keim and Madhavan (1997). To reasonably reflect the decline in stock trading costs in the recent years, we reduce Keim and Madhavan (1997)'s estimate by half.[18] Specifically, there are four levels of transaction cost rates for purchase, with $n_1 = 3$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps (basis points), and $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ (annual volume, in millions). For sale, there are three levels of transaction cost rates, with $n_2 = 2$, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, and $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$.[19] This implies that, for example, in order to purchase stock shares at a speed of 24,000 (resp. 128,000) shares per day, the fund needs to pay a transaction cost of 18 (resp. 24) bps, and so on. Note that the transaction cost rate is concave for low trading speeds and convex for high trading speeds, which cannot be captured by the power law commonly assumed in prior studies.

Consistent with the existing empirical evidence for a typical small cap stock, we assume that the stock has an expected return of $\mu = 0.1$ and a return volatility of

[18] We will show that a higher level of trading cost will yield stronger results.

[19] Keim and Madhavan (1997) estimate trading costs for trade with different sizes, which are calculated as dividing the number of shares traded by the number of outstanding shares. They also report that the average time period it takes to finish a trade is 1.7 days, which allows us to convert the size measure into the speed measure. Also note that according to the estimates in Keim and Madhavan (1997), selling shares at a speed below 11.86 million shares per annum incurs a higher transaction cost rate than selling shares at a speed of 62.63 million shares per annum does. Hence, it is never optimal to sell at the lower speed.

Table 1: Default Parameter Values

This table summarizes our baseline parameter values for asset returns, admissible trading speeds, and transaction costs.

Parameter	Symbol	Baseline value
Risk-free rate	r	0.02
Expected return of the stock	μ	0.1
Volatility of stock returns	σ	0.2
Speed for buying (million shares per annum)	$(\eta_0, \eta_1, \eta_2, \eta_3, \eta_4)$	(0, 5.93, 32.24, 209.75, ∞)
Speed for selling (million shares per annum)	$(\xi_0, \xi_1, \xi_2, \xi_3)$	(0, 62.63, 287.21, ∞)
Marginal transaction cost for buying (bps)	$(\theta_0, \theta_1, \theta_2, \theta_3)$	(18, 6, 10, 41)
Marginal transaction cost for selling (bps)	$(\alpha_0, \alpha_1, \alpha_2)$	(33, 2, 49)

$\sigma = 0.2$. We set the risk-free rate at $r = 0.02$.[20] We summarize these parameter values in Table 1.[21]

Due to the presence of transaction cost, our model cannot be solved in closed-form. As a result, we solve both of the following applications numerically.

4.1 Application 1: Optimal Portfolio Rebalancing

In this subsection, we consider a case in which a fund aims to maximize its expected utility from before-transaction-cost wealth (i.e., asset under management) at time T. This is motivated by the common practice of paying a fund manager in an amount that is linear in the mark-to-market value of the AUM before transaction cost. Let the wealth function be $\phi(x, y, S) = x + yS$. We assume the fund has a constant relative risk aversion (CRRA) utility function with a relative risk aversion (RRA) coefficient $\gamma > 0$, that is,

$$u(W) = \begin{cases} \frac{W^{1-\gamma}}{1-\gamma}, & \text{if } \gamma \neq 1, \\ \ln W, & \text{if } \gamma = 1. \end{cases} \tag{26}$$

We also impose the solvency condition that the $\phi(x_t, y_t, S_t) \geq 0$ for all $t \geq 0$.

[20] Since $\mu > r$, it is never optimal to sell short the stock.

[21] We demonstrate our main results with these parameter values. However, our results hold for a wide range of parameter values.

In this application, we assume the fund maximizes the expected continuously compounded return, which is equivalent to assuming an RRA coefficient of $\gamma = 1$. Moreover, we assume the fund has an investment horizon of 5 years. The detailed description of the solution method is presented in Appendix B.

4.1.1 Optimal Rebalancing Strategy

The Case without Transaction Cost. We begin our analysis with the case in which the market is perfectly liquid. In other words, there is no transaction cost for any trading speed (i.e., $\alpha = \theta = 0$). This analysis reveals the optimal exposure to the risky asset that allows the fund to extract the excess return of the asset while controlling for price risk. In this case, it is well known that the optimal strategy is to continuously rebalance the portfolio to maintain a constant stock–wealth ratio of $\pi^* = \frac{\mu - r}{\gamma\sigma^2}$ (see, e.g., Merton (1971)), which we term the "Merton line" for the fund. In this case, the fund only trades off expected return and market price risk in deciding how many shares to hold before the terminal time.

The Case with Speed-Dependent Transaction Cost Rates. When transaction cost is present, it is optimal to rebalance the portfolio only when the portfolio composition deviates sufficiently away from the Merton line such that the benefit from rebalancing exceeds the incurred transaction cost. Thus, the optimal trading speed in our model is a function of the fund's stock allocation (i.e., $\pi \equiv yS/(x + yS)$) and the stock price–gross wealth ratio (i.e., $b \equiv S/(x + yS)$). The variable π measures the fraction of wealth allocated to the stock, and the variable b measures each share's contribution to the fund's wealth, which is relevant because the transaction cost depends on the number of shares traded per unit of time.

In the presence of transaction costs, when choosing its trading speed, the fund optimally trades off the risk exposure effect (i.e., achieving optimal exposure to the market price risk), the risk premium effect (i.e., earning higher returns by holding stocks rather than holding cash), and the transaction cost effect (i.e., reducing transaction costs). The risk exposure effect makes the fund trade fast when the exposure is different from the target position (i.e., the Merton line); the risk premium effect increases the purchasing speed but reduces the selling speed, while the transaction cost effect reduces the trading speed. These effects jointly determine the optimal speed at which the fund should move its portfolio towards the target position.

When the fund's exposure to the stock is too high, it would like to sell stock shares to reduce risk exposure as soon as possible, but transaction cost rates increase with the trading speed, and therefore the fund's optimal selling speed is determined by the trade-offs between the costs and the benefits. In particular, the optimal trading speed depends on the degree of excess exposure and the number of shares required to sell in order to reach an optimal exposure. The greater the excess exposure, the stronger the exposure effect, and hence the faster the fund should sell. Similarly, when the fund's risk exposure is too low, it would like to increase risk exposure as soon as possible for a greater risk premium, but needs to trade off the greater transaction cost from trading too fast. Following this intuition, we postulate and later verify that, at each point in time, the domain $\mathcal{S} = \Big\{(b, \pi) : b > 0, \pi \geq 0\Big\}$ splits into *at most* $n_1 + n_2 + 3$ regions: an unrestrained Buy region in which the fund purchases shares at an infinite speed, n_1 restrained Buy regions in which the fund purchases shares at a finite speed, a No-Transaction region in which the fund does not trade, n_2 restrained Sell regions in which the fund sells shares at a finite speed, and an unrestrained Sell region in which the fund sells shares at an infinite speed.[22] These regions are separated by at most $n_1 + n_2 + 2$ free boundaries endogenously determined by the solution to Equation (A-5).

[22] Note that some of these regions can be empty sets. In this case, trading at the corresponding speeds is not optimal.

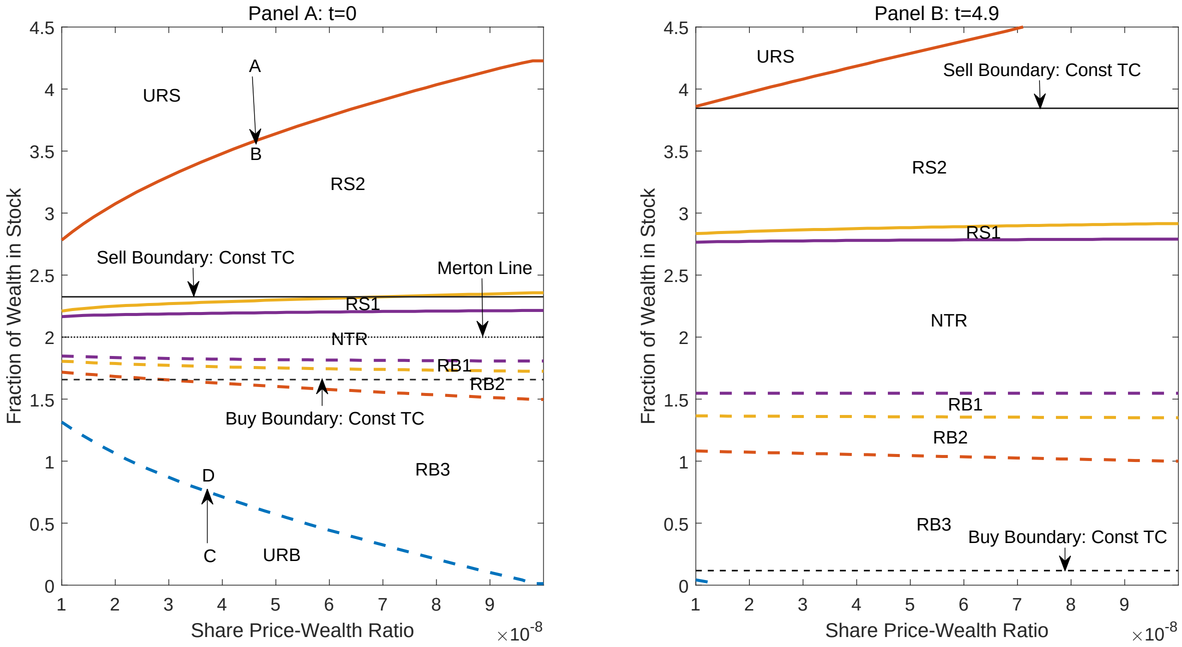

Figure 1: Optimal Trading Boundaries

This figure shows the optimal trading boundaries of the fund, at time $t = 0$ (left panel) and $t = 4.9$ (right panel). Baseline parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\gamma = 1$, $T = 5$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, and $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions.

In Figure 1, we plot the free boundaries that separate regions with different optimal trading speeds. The left panel shows the case at the initial time $t = 0$. When the stock allocation is sufficiently higher than the Merton line (e.g., lies in the region labeled "URS"), it is optimal for the fund to sell stock shares at an infinite speed, which instantly brings the fund's stock allocation to the upper boundary of the region labeled "RS2," signified by the arrow from A to B. After such a lump sum trade, it is optimal for the fund to sell stock shares at the highest finite speed until its stock allocation reaches the upper boundary of the first restrained Sell region labeled "RS1" (i.e., the solid yellow line), in which the fund should further reduce its selling speed to the second highest finite speed to reduce the transaction cost. Finally, the fund should stop selling when its stock allocation reaches the upper boundary of the No-Transaction region labeled "NTR" (i.e., the solid purple line). The NTR arises because maintaining the stock allocation within a tolerable range enables the fund to extract the positive excess return offered by the stock and avoid transaction costs while keeping the exposure to stock price risk in a certain range.

Similarly, the fund needs to purchase stock shares when its stock allocation is sufficiently lower than the Merton line. This is driven by the fund's incentive to extract risk premium from the stock. In the region labeled "URB", it is optimal to purchase shares at an infinite speed, as labeled by the arrow from C to D. Afterwards, it is optimal to purchase shares at the highest finite speed, then the second highest finite speed, and so on, until the stock allocation reaches the lower boundary of the No-Transaction region (i.e., the dashed purple line).

It can be observed from Figure 1 that the free boundaries are functions of the price–fund wealth ratio. The smaller the ratio is, the wider the unconstrained trading regions are. This is intuitive: a smaller price–fund wealth ratio implies a larger amount of shares to be traded to rebalance the fund's portfolio; hence, the fund has a stronger need to trade quickly.

In Figure 1, we also show the optimal trading boundaries in a setting similar to Liu and Loewenstein (2002), in which the transaction cost rates are constant (i.e., independent of the fund's trading speed). In this case, there is only one unrestrained sell region inside which it is optimal to sell stock shares at an infinite speed, one unrestrained buy region inside which it is optimal to buy stock shares at an infinite speed, and a no-transaction region inside which it is optimal not to trade. Because the boundaries between these regions are independent of the share price–wealth ratio, we show these boundaries with two black lines in Figure 1. We find that, compared with the model with constant transaction cost rates, our model implies smaller unrestrained buy and sell regions; however, the no transaction region is smaller. This is intuitive: the presence of speed-dependent transaction costs makes it optimal for the fund to break large lump sum trades into smaller ones so that fewer transaction costs will be incurred; meanwhile, lower transaction costs incurred by smaller trades allow the fund to tilt its portfolio composition closer to the Merton line.

The optimal trading boundaries exhibit little time-variation, except when the terminal time is approached. In the right panel of Figure 1, we show the optimal trading regions at time $t = 4.9$ years, which is close to the 5-year horizon. As time approaches the final horizon, the No-Transaction region (NTR) expands, and the trading regions shrink. The reason is that prior to the terminal time, the risk premium and risk exposure effect tend to vanish, and the fund should make less transactions to save transaction costs.

4.1.2 Utility Loss from Assuming Constant Transaction Cost Rates

If the fund mistakenly assumes the transaction cost rates are constant, then it will only trade at an infinite speed when it rebalances its portfolio, as described in Liu and Loewenstein (2002). Because such a rebalancing strategy is suboptimal in our model, adopting it will be costly. In Figure 2, we report results for the certainty equivalent wealth loss (CEWL) from wrongly assuming constant transaction cost rates of $\alpha = \sum_{i=1}^{n_2} \alpha_i = 84$ bps and $\theta = \sum_{i=1}^{n_1} \theta_i = 75$ bps. The CEWL Δ is calculated through the following equation

$$V(x_0 - \Delta(x_0 + y_0 S_0), y_0, S_0, 0) = V_0(x_0, y_0, S_0, 0), \tag{27}$$

where $V_0(x, y, S, t)$ is the fund's value function if it only trades at the infinite speed. Δ can be interpreted as the fraction of portfolio value that the fund is willing to give up to have access to the optimal trading strategy.[23]

In Figure 2, we plot the CEWL against the fund's investment horizon, ranging from 1 to 5 years, for two levels of the transaction cost rates and two types of initial stock allocation. Figure 2 suggests that ignoring the variations in transaction cost rates can be costly to the fund, even when these rates are small. For instance, Panel B suggests

[23] When calculating CEWL using equation (27), we assume a share price–fund wealth ratio of $b_0 = \frac{S_0}{x_0 + y_0 S_0} = 10^{-6}$. Using other initial values of b_0 yields similar results.

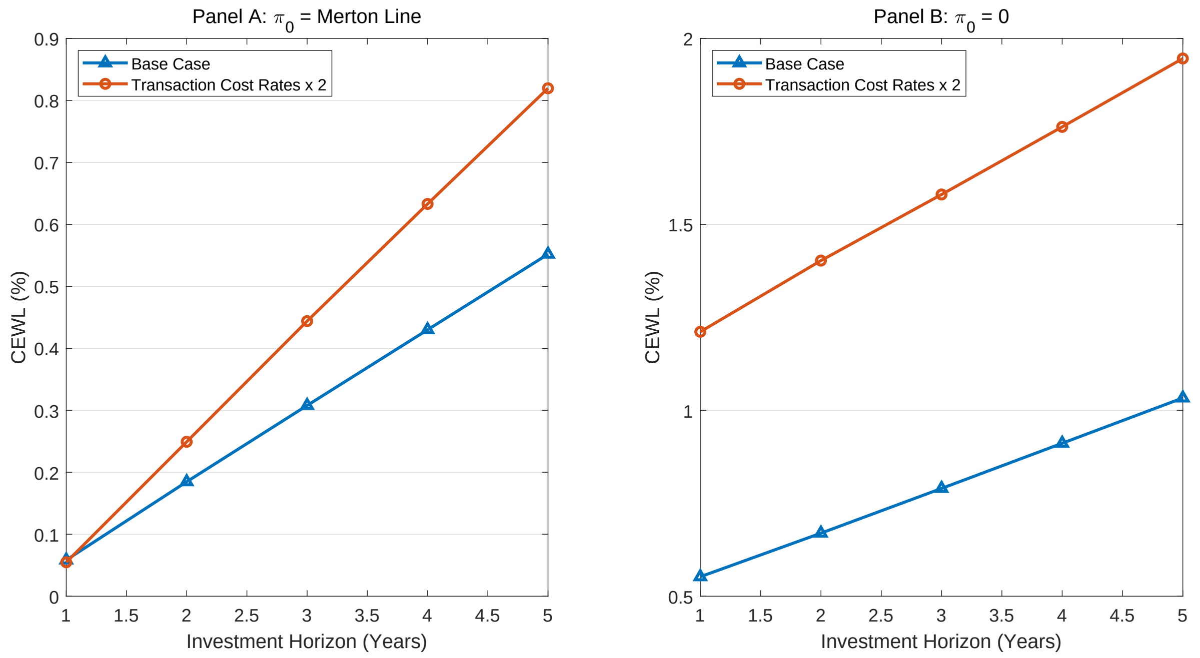

Figure 2: Certainty Equivalent Wealth Loss

This figure shows the fund's certainty equivalent wealth loss if it mistakenly assumes constant transaction cost rates. Baseline parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\gamma = 1$, $T = 5$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, and $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions.

that in our base case, mistakenly assuming constant transaction cost rates will result in a loss of about 104 bps in terms of the fund's initial wealth if the initial stock allocation is zero. This finding explains why breaking a larger order into multiple smaller orders ("shredding") is important, as we commonly observe in practice.

In general, the CEWL increases in the investment horizon. This is intuitive because a longer horizon implies more chances of rebalancing; thus, the fund will incur more losses from adopting a suboptimal rebalancing strategy. Figure 2 also suggests that the utility losses can be much greater with higher transaction cost rates. Therefore, adopting the optimal rebalancing strategy can be particularly important for funds that invest in more illiquid assets.

4.2 Application 2: Optimal Liquidation and Acquisition

In this subsection, we consider a case in which a fund needs to liquidate or acquire a certain number of stock shares by some fixed finite time $T > 0$. In order to treat the optimal liquidation and optimal acquisition problems in a unified framework, we use the following functional form for the terminal wealth level

$$\phi(x, y, S) = x + yS - \alpha(y_T^* - y)^- S - \theta(y_T^* - y)^+ S. \tag{28}$$

In equation (28), y_T^* is the fund's target asset position at time T, with $y_T^* = 0$ and $y_0 = N$ for the case where the fund needs to liquidate N shares by time T, and $y_T^* = N$ and $y_0 = 0$ for the case where the fund needs to acquire N shares by time T. If y_T is less (greater) than the target y_T^*, then the fund must buy (sell) the difference at an infinite speed to reach y_T^*, and pay the corresponding transaction cost.[24]

For preferences that have infinite marginal utility at zero wealth (e.g., CRRA preference), given a feasible but finite amount of initial capital, the only feasible acquisition strategy is to purchase all the shares immediately if the stock price is unbounded above (e.g., the geometric Brownian motion stock price process (1). Therefore, to make the problem nontrivial, we assume that the fund has a CARA preference, namely,

$$u(W) = -e^{-\beta W}, \tag{29}$$

where $\beta > 0$ is the fund's constant absolute risk aversion coefficient. Using the CARA preference still reflects the trade-off among price risk, expected return, and transaction cost. It also captures the idea that the fund may care about not only the expected revenue

[24] Note that our model allows the fund to trade against the intended target when market conditions are favorable. On the other hand, in our model we can also forbid trading against the intended target by imposing the no-purchase constraint $I_t = 0$ for the liquidation case (resp. no-sale constraint $D_t = 0$ for the acquisition case) for any t. See later analysis.

from the liquidation or the expected cost from the purchase, but also the risk of having low revenue or high cost in some states of the world.[25]

We focus on the optimal liquidation case because this case is more commonly studied in the literature, which makes it more meaningful for comparisons. The optimal acquisition case is examined later in Section 4.2.4 to show the qualitative similarity.

We assume the fund aims at liquidating 1 million shares of the stock in five trading days, that is, $y_0 = 1$ million shares and $T = 5/252$ years. We set the initial stock price at $S_0 = 10$ dollars. We assume the fund has an absolute risk aversion coefficient of $\beta = 5 \times 10^{-7}$.[26]

Note that the fund can choose to approach its intended target at a constant intermediate speed of 0.2 million shares per day. Thus, liquidating at a speed faster than this intermediate speed reveals the fund's incentive to control price risk, while liquidating at a speed slower than this intermediate speed indicates the fund's incentive to reduce transaction cost.

We discretize time into hours, that is, $\Delta t = 1$ hour. With CARA utility, we choose y_t and S_t as the effective state variables. The details of the solution method are presented in Appendix C.

4.2.1 Optimal Trading Policies

The No-Transaction Cost Case. Using a dynamic programming approach similar to that of Merton (1971), it can be shown that the optimal trading strategy of the fund is to invest a constant dollar amount (with risk-free discounting) in the stock until time T, when the entire stock position is liquidated. Using the notation of our model, the

[25] CARA preferences with an unbounded stock price, as in our model, provide qualitatively the same optimal acquisition strategy as other preferences with infinite marginal utility at zero wealth (e.g., CRRA preferences) and a bounded stock price. In Appendix D, we present the results obtained from an optimal liquidation model with a CRRA utility function to show the robustness of our results.

[26] This seemingly small absolute risk aversion can translate into a large relative risk aversion when the value of the stock that needs to be liquidated or purchased is large. For example, at a value level of 10 million dollars, the corresponding relative risk aversion coefficient becomes 5.

optimal number of shares invested in the stock at time t, that is, y_t^*, before liquidation time T is equal to

$$y_t^* = \frac{\mu - r}{\beta\sigma^2 S_t} e^{-r(T-t)}. \tag{30}$$

The optimal trading strategy can be described as follows: the fund first sells a lump sum of the stock shares to reach the initial optimal position $y_0^* = 399,800$ shares, and then trades continuously (at an infinite speed) to maintain the above optimal number of shares invested in the stock at each time $t < T$. At time T, the fund liquidates the remaining shares.[27] In what follows, we refer to the position specified by (30) as the "Merton line" for CARA preference.

The Case with Transaction Cost. In the presence of transaction cost, similar to the first application, we postulate that, at each point in time, the domain $\mathcal{S} = \Big\{(y, S) : y \geq 0, S > 0\Big\}$ splits into at most n_1+n_2+3 regions: an unrestrained Buy region, n_1 restrained Buy regions, a No-Transaction region, n_2 restrained Sell regions, and an unrestrained Sell region.

In Figure 3, we plot these optimal trading regions at three points in time: $t = 0$, $t = T/2$, and $t = T - \Delta t$. The Merton line in the absence of any transaction cost is shown in the middle panel by the dotted black line. The meaning of these regions is similar to that in our previous application. Figure 3 suggests that the trading boundaries decrease over time, implying that, for a given number of remaining shares, the speed of liquidation increases as time approaches the liquidation horizon. Interestingly, we find that the NT region converges to zero when the liquidation date approaches (see, e.g., Panel C of Figure 3). This pattern does not show up in classic portfolio choice models with transaction costs, such as Liu and Loewenstein (2002). This convergence suggests

[27] Given the short liquidation horizon of 5 trading days, this optimal liquidation strategy can be approximated by a 60% lump sum sale at the beginning and another 40% lump sum sale in the end. We will shortly show that the optimal liquidation strategy is qualitatively different in the presence of speed-dependent transaction cost.

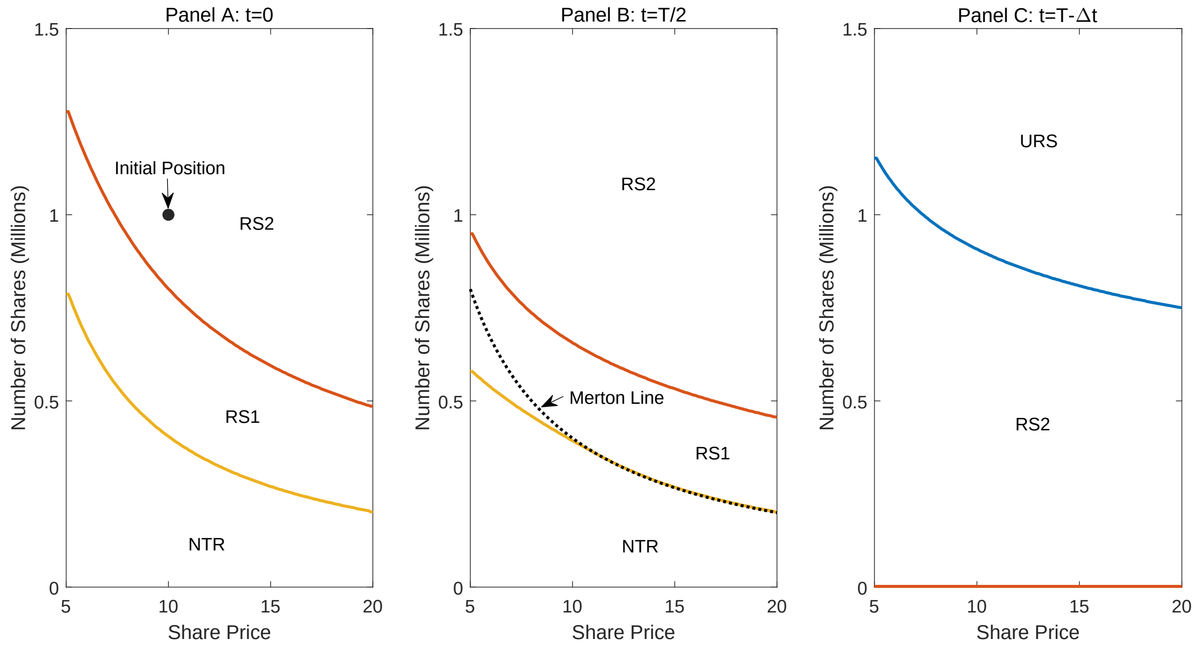

Figure 3: Optimal Trading Boundaries

This figure shows the optimal trading boundaries of the fund at three points in time: $t = 0$, $t = T/2$, and $t = T - \Delta t$. Baseline parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\beta = 5 \times 10^{-7}$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions, and $T = 5/252$ years.

that, as the liquidation date approaches, the fund should liquidate as many shares as possible at finite speed. This can reduce the amount of lump sum liquidation at terminal date T, which would involve significantly greater transaction costs.

4.2.2 Expected Time to Reach a Target

In our model, the optimal trading speed depends on the fund's remaining position to liquidate, the fund's risk preference, the stock's risk-return profile, and the structure of the transaction costs. In this section, we measure the fund's average trading speed by calculating the expected time it takes to liquidate a certain number of shares.[28]

[28] We use this measure because the fund in our model can make lump sum trades at an infinite speed, and the expected time measure can incorporate this situation. Specifically, if the fund trades at an infinite speed so that its position jumps from one level to another, then the expected time to reach these two levels is the same.

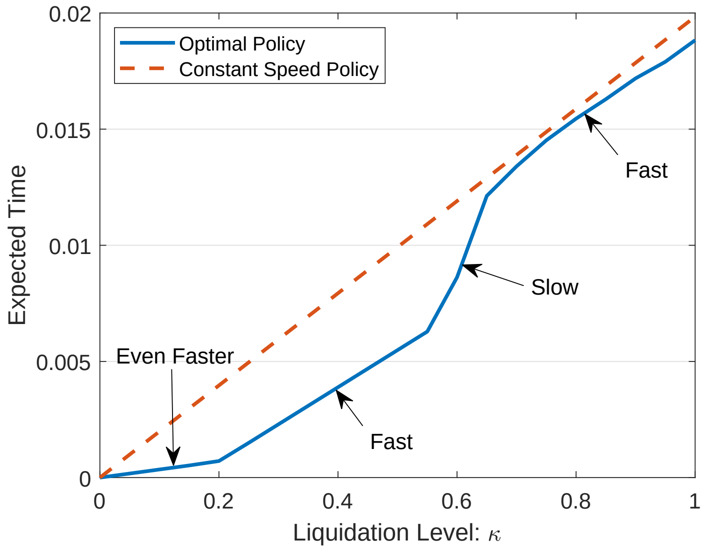

Figure 4: Expected Time for Liquidation

This figure shows the expected time to liquidate a certain proportion of shares. Baseline parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\beta = 5 \times 10^{-7}$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions, and $T = 5/252$ years.

Formally, for $0 \leq \kappa \leq 1$, we define the following stopping time

$$\tau_\kappa^l = \min\{t \geq 0 : y_t \leq (1 - \kappa) y_0\}, \tag{31}$$

which is the first time the fund has already liquidated a fraction κ of its initial stock position. We can then measure the fund's average trading speed by calculating the expected value $E[\tau_\kappa^l]$.[29]

In Figure 4, we show the expected time to liquidate a certain fraction of the initial position. Compared with the naive strategy of liquidating at a constant speed (depicted by the red dashed line), the optimal liquidation strategy (depicted by the blue line) sells shares at time-varying speeds to optimally trade off the market risk exposure and the transaction costs. Note that the steeper the slope of the curve, the more slowly the fund liquidates its position. Hence, Figure 4 indicates that the optimal liquidation strategy is to first liquidate shares at a higher speed, then reduce the liquidation speed; as time

[29] The approach to compute $E[\tau_\kappa^l]$ is presented in Appendix C.

approaches the final horizon, the fund increases its liquidation speed again. In other words, the fund's optimal liquidation speed is a U-shaped function of time.

Obizhaeva and Wang (2013) examine optimal trading execution in a limit order book with time-varying supply and demand. The optimal trade size implied by their model also displays a similar U-shaped pattern. However, the mechanism of their model is fundamentally different from ours. In their model, the purpose of making a large trade at the initial time is to push the order distribution away from its steady state. This could attract liquidity providers to trade against later. By comparison, in our model, the purpose of trading quickly during early hours is to reduce the market price risk exposure of the position. In Obizhaeva and Wang's (2013) model, the purpose of making a large trade at the terminal time is just to (mechanically) reach the target. In our model, a lump-sum trade at the terminal date is not always optimal, because it can be optimal to increase trading speed to reach the target before the terminal date, as this allows the fund to reduce transaction costs potentially incurred by a final lump sum trade.[30]

In Figure 5, we show how the fund's average trading speed changes with various model parameters, including the stock's expected return, the stock's return volatility, the fund's risk aversion coefficient, and the magnitude of transaction cost rates. When the expected return of the stock decreases or its volatility increases or the risk aversion level of the fund increases, the fund is willing to invest less in the stock. As a result, the fund should liquidate the stock at a higher speed. When the transaction cost rates faced by the fund increase, it is optimal to liquidate at a lower speed at the beginning to save transaction costs; however, it is not optimal to further reduce the trading speed (e.g., by stopping trading) half way, because the fund needs to reach its target before the final time to avoid a much costlier lump-sum trade. This explains why the optimal trading speed plotted in Panel D exhibits less time-variation when the transaction cost rates increase.

[30] The necessity of a lump-sum trade at the terminal date depends on model parameters. Generally, the higher the Merton line or the slower the trading speeds, the more likely that a final lump-sum trade is necessary.

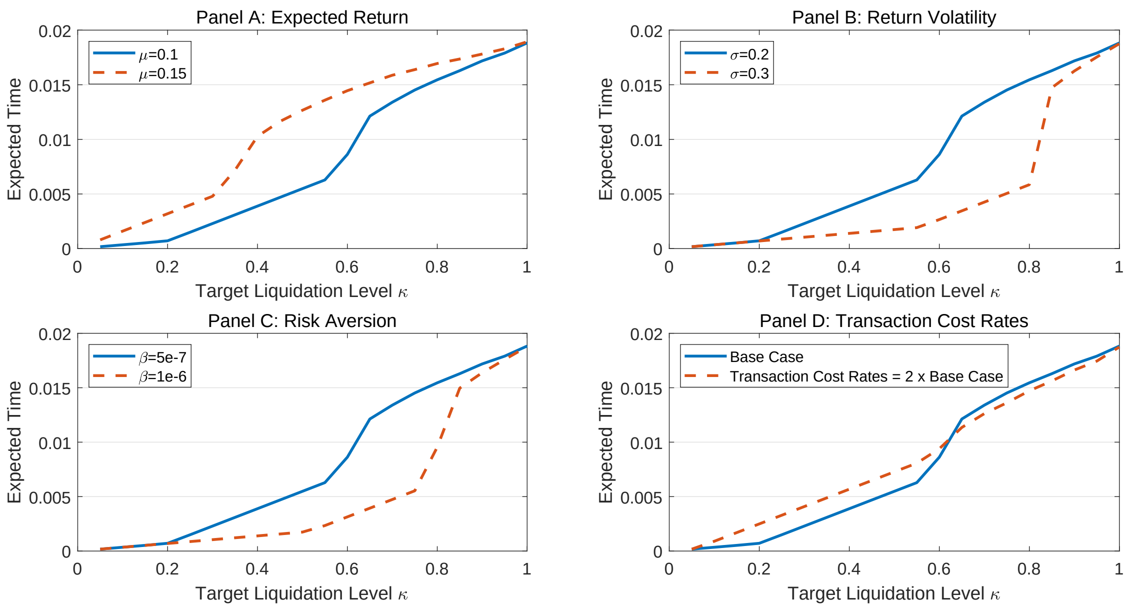

Figure 5: Expected Time for Liquidation: Comparative Statics

This figure shows the comparative statics results on the expected time to liquidate a certain proportion of shares. Baseline parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\beta = 5 \times 10^{-7}$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions, and $T = 5/252$ years.

4.2.3 Utility Loss from Adopting Naive Strategies

As noted by Bertsimas and Lo (1998), some naive liquidation strategies, such as liquidating at a constant speed over the execution horizon or liquidating all position instantly to avoid any market price risk, are not uncommon in practice. In this section we examine the welfare implication of adopting these naive liquidation strategies.

The constant speed strategy can be characterized by the following conditions

$$dI_t = 0, \quad dD_t = \frac{y_0}{T} dt, \tag{32}$$

and the immediate complete liquidation strategy can be characterized by

$$dI_t = 0, \quad dD_0 = y_0 = N, dD_t = 0, \forall t > 0. \tag{33}$$

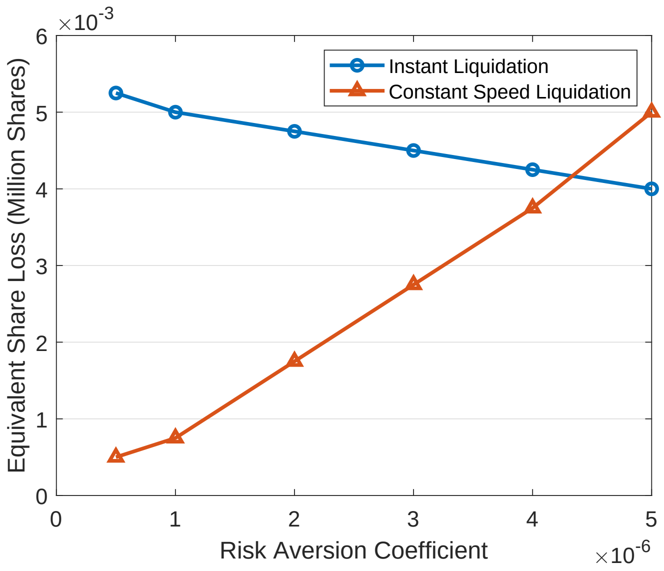

Figure 6: Certainty Equivalent Share Loss from Adopting Naive Strategies
This figure shows the certainty equivalent shares loss from adopting the naive liquidation strategies. Baseline parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\beta = 5 \times 10^{-7}$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions, and $T = 5/252$ years.

Therefore, both strategies are feasible but suboptimal in our model. We measure the utility loss from adopting a particular naive liquidation strategy through the certainty equivalent share loss (CESL) Δ_1, which solves the following equation

$$V^N(x_0, y_0, S_0, 0) = V(x_0, y_0 - \Delta_1, S_0, 0), \tag{34}$$

where $V^N(x, y, S, t)$ is the indirect utility function associated with a naive liquidation strategy.

In Figure 6, we show the CESLs from adopting the naive strategies and how they change with respect to the fund's risk aversion coefficient. Intuitively, the constant speed liquidation strategy exposes the fund to greater market price risk, while the instant complete liquidation strategy incurs larger transaction costs. When the fund's risk aversion level is low, transaction costs tends to dominate market price risk. Thus, the constant speed strategy is likely to be less costly to adopt than the instant liquidation strategy. The reverse is true when the fund is highly risk averse. Figure 6 confirms such an intuition.

Figure 6 also suggests that the losses from adopting the naive strategies can be economically large. For example, with a risk aversion coefficient of 5×10^{-6}, the constant liquidation speed strategy can result in a loss that accounts for 5,000 shares, while the immediate complete liquidation can result in a loss that accounts for 4,000 shares. These certainty equivalent cost calculations clearly show the economic importance of adopting the optimal liquidation strategy.[31]

4.2.4 The Acquisition Case

In this subsection, we briefly analyze an optimal acquisition problem that can be important in practice (e.g., acquiring toeholds for merger and acquisition purposes). In this case, we assume the fund's objective is to acquire $y_T^* = 1$ million shares in one week.

In Figure 7, we show the optimal acquisition strategy of the fund. Panel A (B, C, resp.) shows the optimal trading regions at $t = 0$ ($T/2$, $T - \Delta t$, resp.). At $t = 0$, the solution domain is split into four regions: a No-Transaction region (NTR), the first restrained Buy region (RB1), the second restrained Buy region (RB2), and the third restrained Buy region (RB3). Similar to the liquidation case, the Merton line represents the optimal stock exposure absent any transaction costs. Panel A shows that the fund does not purchase a lump sum at time 0.[32] Instead, the fund buys at a speed of 0.83 million shares per day toward its target and starts to earn the risk premium. As the fund acquires more shares, it decreases the purchasing speed to reduce transaction costs. After the fund's position enters the No-Transaction region, the fund does not acquire more shares, even if its current position has not reached the target position yet. This is

[31] The results are even stronger when the liquidation horizon is longer or when the transaction cost effect is larger.

[32] With a higher Merton line, however, the optimal acquisition strategy may require an initial lump sum purchase.

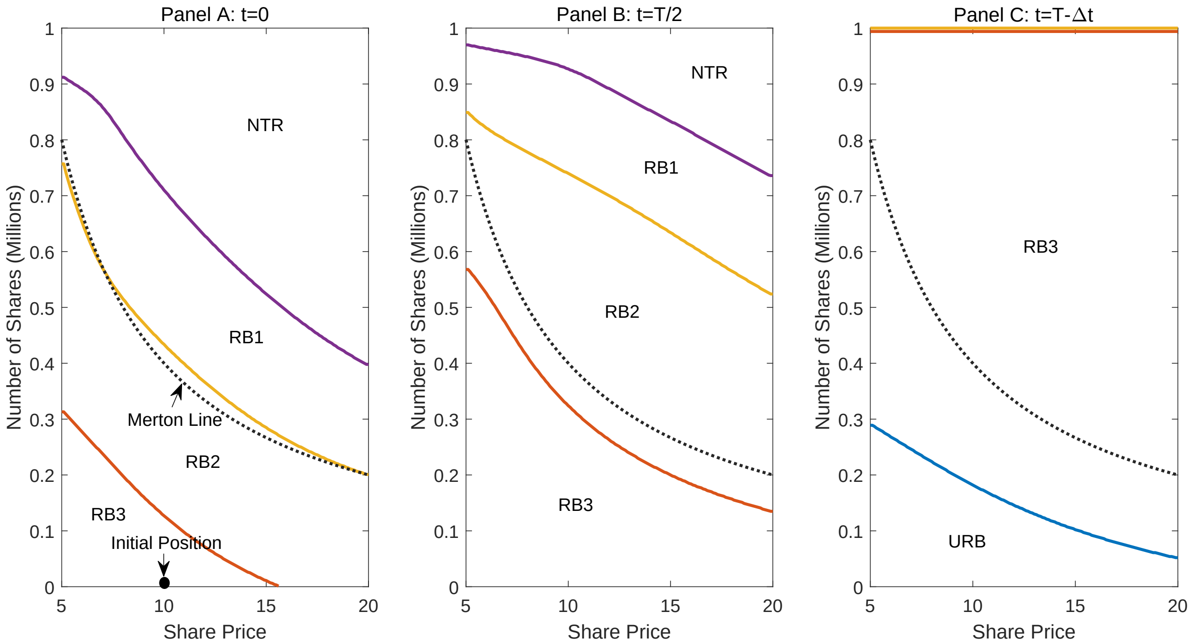

Figure 7: Optimal Trading Boundaries: Acquisition Case

This figure shows the optimal trading boundaries of the fund when the fund's objective is to acquire $y_T^* = 1$ million shares by time T. Baseline parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\beta = 5 \times 10^{-7}$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions, and $T = 5/252$ years.

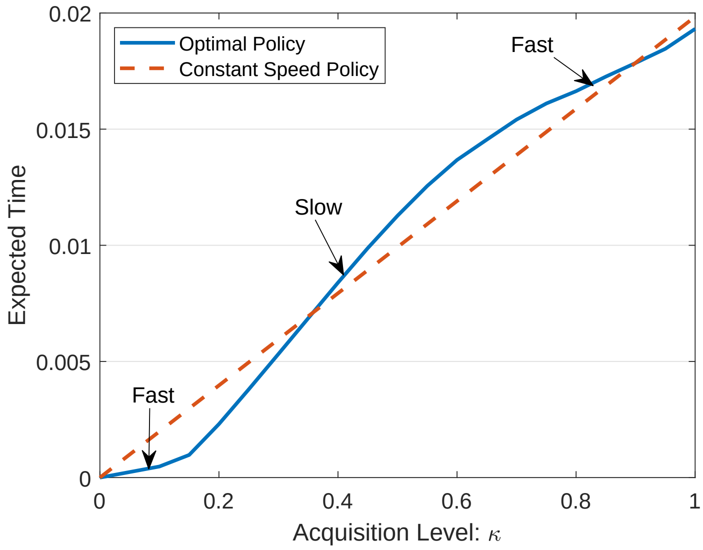

Figure 8: Expected Time for Acquisition

This figure shows the expected time required to acquire a certain proportion of the target share. Baseline parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\beta = 5 \times 10^{-7}$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions, and $T = 5/252$ years.

because the fund needs to control market price risk by avoiding too much exposure to the asset during early periods.

Right before the terminal date $t = T - \Delta t$, an unrestrained Buy region (UBR) shows up for sufficiently low stock exposures. This is because it is possible that the fund needs to conduct a lump sum acquisition at the final date T, and doing part of such an acquisition at $T - \Delta t$ enables the fund to extract the expected return of the stock longer at the same transaction costs. In addition, at $t = T - \Delta t$, the third restrained Buy region becomes much wider than at $t = 0$. This is because the fund needs to reach its target position at the shorter remaining horizon. Thus, to save the transaction costs, it is better to purchase as much as possible with a finite speed to reduce the amount that has to be purchased at an infinite speed at T.

In Figure 8, we show the expected time required to acquire a certain fraction of the target share. It demonstrates that the optimal acquisition strategy is to first acquire shares at a higher speed, then reduce the acquisition speed; as time approaches the final horizon, the fund increases its acquisition speed again. This implies that, like the liquidation case, the optimal acquisition speed is also a U-shaped function of time.

4.2.5 Revenue Maximization and Informed Trading

In this section, we examine two other cases related to this application. Specifically, we examine a case in which the fund's objective is to maximize expected revenue from liquidation without controlling revenue risk and a case in which the fund has private information regarding the final true value of the stock, as in Kyle (1985).

The Revenue Maximization Case. One alternative objective of the fund might be to maximize its expected revenue from the liquidation without taking into account the revenue risk (e.g., Ting et al. (2007) and Bertsimas and Lo (1998) for the acquisition case). In this case, the fund's objective can be represented by

$$\max_{D_t} E\left[\int_0^T e^{-rt}\left(1-\sum_{i=0}^{n_2}\alpha_i 1_{\{dD_t>\xi_i dt\}}\right)S_t dD_t + e^{-rT}(1-\alpha)y_T S_T\right], \tag{35}$$

subject to $dy_t = -dD_t$.[33] Note that this is the present value of total revenue at the initial time 0. It can be shown that the optimal liquidation strategy can be obtained from our main model for the CARA fund by setting the risk-aversion coefficient to $\beta = 0$ with the additional constraints.[34] Therefore, in what follows we refer to this fund as a risk-neutral fund.

The optimal liquidation strategy of the risk-neutral fund has some distinct features. Different from the risk-averse fund case, the optimal liquidation speed of a risk-neutral fund only depends on time-to-maturity and the number of remaining shares, and is inde-

[33] Restricting the fund from purchasing is a sufficient condition to ensure that the risk-neutral fund's optimization problem (35) is well-defined. This assumption also facilitates a fair comparison between the risk-neutral case and our baseline case where there is no purchasing.

[34] This can be shown by deriving the HJB equation in this case and comparing it with the HJB equation in our baseline model.

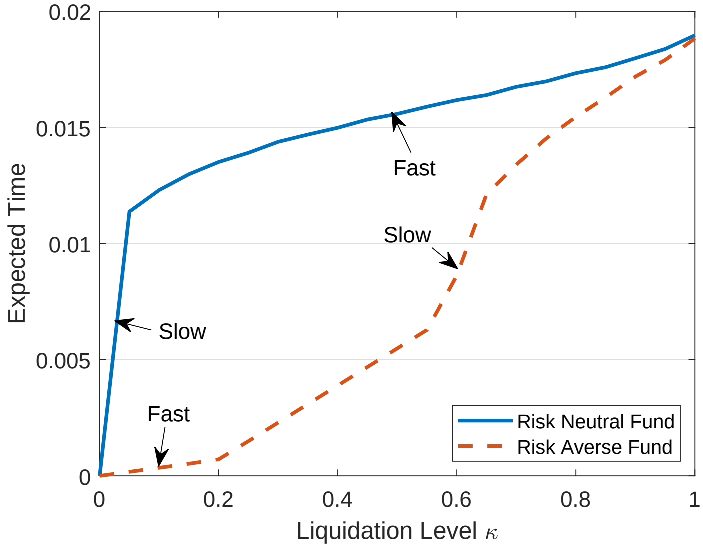

Figure 9: Optimal Trading Speed of Risk-Neutral and Risk-Averse Fund

This figure shows the average trading speed of a fund that aims at maximizing its expected revenue over the liquidation horizon. Parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\beta = 0$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions, and $T = 5/252$ years.

Table 2: Statistics of Total Revenue

This table shows the average value and standard deviation of the total revenue after liquidation. Baseline parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\beta = 5 \times 10^{-7}$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions, and $T = 5/252$ years.

	Risk-Averse Fund	Risk-Neutral Fund
Average Revenue (Millions)	9.975	9.981
Standard Deviation (Millions)	0.105	0.234
Sharpe Ratio	0.526	0.263

pendent of the stock price.[35] This is because the risk-neutral fund is not concerned about market price risk.

In Figure 9, we show the average liquidation speed when the fund is risk-neutral (vs. risk-averse). Because the risk-neutral fund does not care about price risk, to reduce transaction costs and earn stock risk premium, the fund tends to liquidate shares faster only when the final horizon is approaching. This is different from the risk-averse fund which tends to liquidate faster at the beginning of the liquidation horizon.

[35] In other words, the optimal trading boundaries are similar to those in Figure 3, except that they are flat. We do not show them to save space. They are available from authors upon request.

In the risk-averse fund case, the fund needs to control the price risk to avoid excessive stock exposure. In the risk-neutral fund case, the fund tends to maintain a high level of stock exposure to extract the risk premium. As a result, the total revenue should have a riskier distribution in the risk-neutral fund case. In Table 2, we report the average value and the standard deviation of the total revenue after liquidation. The results are obtained by performing 10,000 Monte-Carlo simulations of the model. We find that the average revenue is only slightly decreased in the risk-averse fund case, while the standard deviation in this case is much smaller. As a result, the risk-averse fund can achieve a higher Sharpe ratio.[36] These results suggest that the optimal liquidation strategy derived from our model can substantially reduce revenue risk while maintaining a similar expected revenue.

The Informed Trading Case. In this part, we examine a case of informed trading that is similar to Kyle (1985) but in the presence of speed-dependent transaction cost. In particular, we assume that the fund has private information regarding the true value of the asset at time T, which is denoted by V_T. For simplicity, we assume that V_T is a constant. As such, the fund's terminal gross wealth function becomes $\phi(x, y, S) = x + yV_T$. We assume the fund's objective is still to maximize (6).[37]

In Figure 10, we show the optimal trading boundaries of the fund at initial time $t = 0$. If the stock's spot price is sufficiently high, then the fund's optimal strategy is to liquidate shares in a way that is similar to our baseline model. This allows the fund to obtain more proceeds before the stock's true value is revealed. Specifically, the optimal liquidation

[36] The Sharpe ratio is calculated as follows. We first calculate the total revenue obtained by implementing the instant liquidation strategy, and use this as the risk-free benchmark. In the baseline case, it is equal to $10 \times (1 - 0.0084) \times e^{0.02 \times 5/252} = 9.920$. We then calculate the average revenue in excess of the risk-free benchmark. Finally, we divide this excess revenue by the standard deviation.

[37] In order to prevent arbitrage opportunities, we restrict the fund from purchasing the stock shares at an infinite speed. Otherwise, when the spot price is well below the true value V_T, the fund can borrow to purchase an unlimited number of shares and sell them at time T to earn unlimited profit. One could endogenize such an effect by assuming that a large block purchase order of the fund will reveal its private information and cause a large increase in the market price of the stock, which eliminates the fund's incentive to conduct such a purchase. We impose this restriction on trading exogenously to simplify our analysis without altering the main mechanism of the model.

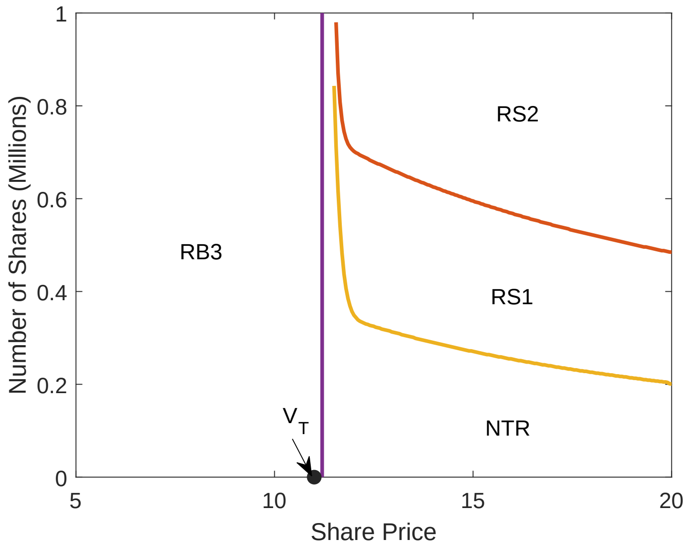

Figure 10: Optimal Trading Boundaries in a Kyle-Type Model

This figure shows the optimal trading boundaries at initial time $t = 0$ in a Kyle-type model. Baseline parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\beta = 5 \times 10^{-7}$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions, and $T = 5/252$ years. The stock's true terminal value is $V_T = 11$ dollars.

speed decreases in the number of shares held by the fund. For example, within the region signified by "RS1" in Figure 10, it is optimal to sell shares at the speed of 0.25 million shares per day; within the region signified by "NTR," it is optimal to stop selling. If the stock's spot price is low, then the fund's optimal strategy is to acquire shares. This allows the fund to accumulate more wealth on paper when the stock's true value is revealed. In the case shown in Figure 10, it is optimal to purchase shares at the speed of 0.83 million shares per day when the stock is undervalued.

Interestingly, we find that it can be optimal to purchase shares even when the spot price is slightly above its true value $V_T = 11$ (e.g., when $S_0 = 11.1$). The intuition is that the private information on V_T provides the fund with an insurance that limits its potential loss on the stock at time T. As a result, the fund has a stronger incentive to purchase the stock to extract its risk premium, and hopes to sell it before the true terminal value is revealed.

To summarize, this case suggests that the main intuition developed in our baseline model also applies to the optimal trading of informed traders.

5 Conclusion

The effective transaction cost rates faced by large investors, such as financial institutions, can be significant and typically depend on their trading speeds. The existing literature on optimal portfolio choice with transaction costs does not incorporate such empirically documented speed-dependency. In this study, we propose a framework to study how speed-dependent transaction costs affect optimal trading strategies. The transaction cost rate, as a function of the trading speed, can have flexible shapes. Thus, our framework generalizes the existing literature on optimal rebalancing with transaction costs and complements the existing literature on optimal liquidation/acquisition for a large institution in a unified setting. In addition, the flexible transaction cost rate structure facilitates calibration of our model to empirical data.

We then apply our framework to the optimal rebalancing problem and an optimal liquidation and acquisition problem for an institution in the presence of speed-dependent transaction costs. We characterize and numerically solve the optimal trading strategies. Unlike the existing literature on the optimal rebalancing strategy in the presence of constant transaction costs, we find that the optimal rebalancing startegy in our model is to break up a large order into multiple small orders and trade at a decreasing speed across time toward the Merton line, which is largely consistent with empirical evidence. For the optimal liquidation or acquisition strategy, our model shows that it is optimal to transact at a decreasing speed when the time-to-horizon is long, and to transact at an increasing speed when the time-to-horizon is short. In other words, the optimal liquidation or acquisition speed is a U-shaped function of time. Moreover, our insights apply to informed trading when transaction cost is dependent on trading speed.

References

Almgren, R., and N. Chriss, 2000, Optimal execution of portfolio transactions, *Journal of Risk* 3, 5–39.

Almgren, R., C. Thum, E. Hauptmann, and H. Li, 2005, Direct estimation of equity market impact, *Journal of Risk* 18, 57–62.

Bertsimas, D., and A. W. Lo, 1998, Optimal control of execution costs, *Journal of Financial Markets* 1, 1–50.

Biais, B., P., Hillion, and C. Spatt, 1995, An empirical analysis of the limit order book and the order flow in the Paris Bourse, *Journal of Finance* 50, 1655–1689.

Chan, L., and J. Lakonishok, 1995, The behavior of stock prices around institutional trades, *Journal of Finance* 50, 1147–1174.

Chan, L., and J. Lakonishok, 1997, Institutional equity trading costs: NYSE versus Nasdaq, *Journal of Finance* 52, 713–735.

Chen, X., and M. Dai, 2013, Characterization of optimal strategy for multi-asset investment and consumption with transaction costs, *SIAM Journal on Financial Mathematics* 4, 857–883.

Chen, Y., M. Dai, L. Goncalves-Pinto, J. Xu, and C. Yan, 2020, Incomplete information and the liquidity premium puzzle, *Management Science*, Forthcoming.

Clarke, R., H. de Silva, and S. Thorley, 2002, Portfolio constraints and the fundamental law of active management, *Financial Analysts Journal* 58, 48–66.

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

Coval, J., and E. Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 479–512.

Cuoco, D., and J. Cvitanić, 1998, Optimal consumption choices for a "large" investor, *Journal of Economic Dynamics and Control* 22, 401–436.

Curato, G., J. Gatheral, and F. Lillo, 2017, Optimal execution with nonlinear transient market impact, *Quantitative Finance* 17, 41–54.

Dai, M., H. Jin, and H. Liu, 2011, Illiquidity, position limits, and optimal investment for mutual funds, *Journal of Economic Theory* 146, 1598–1630.

Dai, M., P. Li, H. Liu, and Y. Wang, 2016, Portfolio choice with market closure and implications for liquidity premia, *Management Science* 62, 368–386.

Dai, M., and Y. Zhong, 2010, Penalty methods for continuous-time portfolio selection with proportional transaction costs, *Journal of Computational Finance* 13, 1–31.

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

Dufour, A., and R. Engle, Time and the price impact of a trade, *Journal of Finance* 55, 2467–2498.

Garleanu, N., and L. H. Pedersen, 2013, Dynamic trading with predictable returns and transaction costs, *Journal of Finance* 68, 2309–2340.

Gatheral, J., A. Schied, and A. Slynko, 2012, Transient linear price impact and Fredholm integral equations, *Mathematical Finance* 22, 445–474.

Gueant, O., and C. Lehalle, 2015, General intensity shapes in optimal liquidation, *Mathematical Finance* 25, 457–495.

Huberman, G., and W. Stanzl, 2004, Price manipulation and quasi-arbitrage, *Econometrica* 72, 1247–1275.

Keim, D., and A. Madhavan, 1995, The anatomy of the trading process, *Journal of Financial Economics* 37, 371–398.

Keim, D., and A. Madhavan, 1996, The upstairs market for large-block transactions: Analysis and measurement of price effects, *Review of Financial Studies* 9, 1–36.

Keim, D., and A. Madhavan, 1997, Transaction costs and investment style: An inter-exchange analysis of institutional equity trades, *Journal of Financial Economics* 46, 265–292.

Kervel, V., and A. Menkveld, 2019, High-frequency trading around large institutional orders, *Journal of Finance* 74, 1091–1137.

Kratz, P., and T. Schoneborn, 2015, Portfolio liquidation in dark pools in continuous time, *Mathematical Finance* 25, 496–544.

Kratz, P., and T. Schoneborn, 2018, Optimal liquidation and adverse selection in dark pools, *Mathematical Finance* 28, 177–210.

Kyle, A. S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1335.

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

Liu, H., and M. Loewenstein, 2002, Optimal portfolio selection with transaction costs and finite horizons, *Review of Financial Studies* 15, 805–835.

Lo, A. W., H. Mamaysky, and J. Wang, 2004, Asset prices and trading volume under fixed transaction costs, *Journal of Political Economy* 112, 1054–1090.

Lokka, A., 2014, Optimal liquidation in a limit order book for a risk-averse investor, *Mathematical Finance* 24, 696–727.

Longstaff, F. A., 2001, Optimal portfolio choice and the valuation of illiquid securities, *Review of Financial Studies* 14, 407–431.

Lynch, A. W., and S. Tan, 2010, Multiple risky assets, transaction costs, and return predictability: Allocation rules and implications for U.S. investors, *Journal of Financial and Quantitative Analysis* 45, 1015–1053.

Lynch, A. W., and S. Tan, 2011, Explaining the magnitude of liquidity premia: The roles of return predictability, wealth shocks, and state-dependent transaction costs, *Journal of Finance* 66, 1329–1368.

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

Niemeyer, J., and P. Sandas, 1993, An empirical analysis of the trading structure at the Stockholm Stock Exchange, *Journal of Multinational Financial Management* 3, 63–101.

Obizhaeva, A., and J. Wang, 2013, Optimal trading strategy and supply/demand dynamics, *Journal of Financial Markets* 16, 1–32.

Perold, A., 1988, The implementation shortfall: paper versus reality, *Journal of Portfolio Management* 14, 4–9.

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

Ting, C., M. Warachka, and Y. Zhao, 2007, Optimal liquidation strategies and their implications, *Journal of Economic Dynamics and Control* 31, 1431–1450.

Tsoukalas, G., J. Wang, and K. Giesecke, 2019, Dynamic portfolio execution, *Management Science* 65, 2015–2040.

Vayanos, D., 1998, Transaction costs and asset prices: A dynamic equilibrium model, *Review of Financial Studies* 11, 1–58.

Weber, P., and B. Rosenow, 2005, Order book approach to price impact, *Quantitative Finance* 5, 357–364.

Appendix

The content of this appendix is as follows. In Appendix A, we sketch the proof of Proposition 1. In Appendix B, we present the detailed solution method to the optimal portfolio rebalancing problem. In Appendix C, we present the details of the optimal liquidation/acquisition problem. In Appendix D, we present the solution to an optimal liquidation problem with a CRRA utility function.

A Proof of Proposition 1

Proof. Because the proof is similar to those in the literature on portfolio choice with transaction costs (see, e.g., Shreve and Soner (1994), and Liu (2004)), we only provide the main steps and skip some technical details.

(i) For any $t \in [0, T]$, without loss of generality, let us assume that $(x_t, y_t, S_t, t) \notin int(URS) \bigcup int(UBS)$. Let $(I, D) = \{(I_u, D_u) : t \leq u \leq T\}$ be an arbitrary admissible strategy, and (x_u, y_u, S_u) be the state process associated with this admissible strategy for $u \in [t, T]$. Under some regularity conditions, we can apply the generalized Itô's lemma to the process $v(x_u, y_u, S_u, u)$ and use equation (11) to conclude

$$v(x_t, y_t, S_t, t) \geq E[v(x_T, y_T, S_T, T)|\mathcal{F}_t] = E[u(\phi(x_T, y_T, S_T))|\mathcal{F}_t]. \tag{A-1}$$

Due to the arbitrariness of (x_t, y_t, S_t, t) and (I, D), we must have $v(x, y, S, t) \geq V(x, y, S, t)$.

(ii) For the strategy (I^*, D^*) as specified in the theorem, denote by (x_u^*, y_u^*, S_u) the state process associated with this strategy for $u \in [t, T]$. In this case, we have

$$v(x_t, y_t, S_t, t) = E[v(x_T, y_T, S_T, T)|\mathcal{F}_t] = E[u(\phi(x_T, y_T, S_T))|\mathcal{F}_t]. \tag{A-2}$$

Due to the definition of the value function, we must also have $V(x, y, S, t) \geq v(x, y, S, t)$.

As a result of (i) and (ii), we must have $v(x, y, S, t) = V(x, y, S, t)$, and the optimality of the stated strategy is established. $\square$

B Details of the Optimal Rebalancing Problem

In this case, the value function has the following homogeneity property

$$V(ax, y, aS, t) = \begin{cases} a^{1-\gamma} V(x, y, S, t) & \text{if } \gamma \neq 1 \\ \ln a + V(x, y, S, t) & \text{if } \gamma = 1 \end{cases} \tag{A-3}$$

for any $a > 0$. This motivates us to use the following transformation to reduce the dimensionality of the problem

$$V(x, y, S, t) = \begin{cases} \frac{(x+yS)^{1-\gamma}}{1-\gamma} e^{(1-\gamma)\varphi(b,\pi,t)} & \text{if } \gamma \neq 1 \\ \ln(x + yS) + \varphi(b, \pi, t) & \text{if } \gamma = 1, \end{cases} \tag{A-4}$$

where $b = \frac{S}{x+yS}$ denotes the share price–fund wealth ratio, and $\pi = \frac{yS}{x+yS}$ denotes the fund's stock allocation. Then, it can be verfied that the function $\varphi(b, \pi, t)$ satisfies

$$\max \left\{ \mathcal{M}_1\varphi + \varphi_t, \ \ \mathcal{S}\varphi, \ \ \mathcal{B}\varphi \right\} = 0, \tag{A-5}$$

$$\varphi(b, \pi, T) = 0, \tag{A-6}$$

where

$$\begin{aligned} \mathcal{M}_1\varphi \; = \; & A_0 + A_b\varphi_b + A_\pi\varphi_\pi + A_{bb}(\varphi_{bb} + (1-\gamma)\varphi_b^2) \\ & + A_{\pi\pi}(\varphi_{\pi\pi} + (1-\gamma)\varphi_\pi^2) + A_{b\pi}(\varphi_{b\pi} + (1-\gamma)\varphi_b\varphi_\pi) \\ & + \sup_{0 \le i \le \eta_{n_1}} [B_0 + B_b\varphi_b + B_\pi\varphi_\pi] i + \sup_{0 \le d \le \xi_{n_2}} [C_0 + C_b\varphi_b + C_\pi\varphi_\pi] d, \end{aligned} \tag{A-7}$$

$$\mathcal{S}\varphi = -\alpha b + \alpha b^2\varphi_b - b(1 - \alpha\pi)\varphi_\pi, \tag{A-8}$$

$$\mathcal{B}\varphi = -\theta b + \theta b^2\varphi_b + b(1 + \theta\pi)\varphi_\pi, \tag{A-9}$$

with the following coefficients

$$A_0 = r + (\mu - r)\pi - \frac{1}{2}\gamma\sigma^2\pi^2, \tag{A-10}$$

$$A_b = b(1-\pi)(\mu - r - \gamma\sigma^2\pi), \tag{A-11}$$

$$A_\pi = \pi(1-\pi)(\mu - r - \gamma\sigma^2\pi), \tag{A-12}$$

$$A_{bb} = \frac{1}{2}\sigma^2 b^2(1-\pi)^2, \tag{A-13}$$

$$A_{\pi\pi} = \frac{1}{2}\sigma^2\pi^2(1-\pi)^2, \tag{A-14}$$

$$A_{b\pi} = \sigma^2 b\pi(1-\pi)^2, \tag{A-15}$$

$$B_0 = b(-\sum_{j=1}^{n_1}\theta_j 1_{i>\eta_j}), \tag{A-16}$$

$$B_b = b^2(\sum_{j=1}^{n_1}\theta_j 1_{i>\eta_j}), \tag{A-17}$$

$$B_\pi = b(1 + \sum_{j=1}^{n_1}\theta_j 1_{i>\eta_j}\pi), \tag{A-18}$$

$$C_0 = b(-\sum_{j=1}^{n_2}\alpha_j 1_{d>\xi_j}), \tag{A-19}$$

$$C_b = b^2(\sum_{j=1}^{n_2}\alpha_j 1_{d>\xi_j}), \tag{A-20}$$

$$C_\pi = -b(1 - \sum_{j=1}^{n_2}\alpha_j 1_{d>\xi_j}\pi). \tag{A-21}$$

We numerically solve equation (A-5) using the penalty method (see, for example, Dai and Zhong (2010)) combined with a finite differences scheme.

C Details of the Optimal Liquidation/Acquisition Problem

In this case, we have the following proposition that helps us simplify the HJB equation.

Proposition 2. *(Separability of the value function.) The value function $V(x, y, S, t)$ has the following property:*

$$V(x, y, S, t) = e^{-\beta x e^{r(T-t)}} V(0, y, S, t). \tag{A-22}$$

Proof. For any state $(0, y_t, S_t, t)$, let $\Theta = \{(I_s, D_s) : t \le s \le T\}$ be an admissible trading strategy that generates a terminal state of $(x_T^\Theta, y_T^\Theta, S_T)$. Then, for the state (x_t, y_t, S_t, t), it is easy to see that the same strategy Θ will generate terminal state $(x_T^\Theta + x_t e^{r(T-t)}, y_T^\Theta, S_T)$. Therefore, we have

$$\begin{aligned} V(x_t, y_t, S_t, t) &\ge E_t\left[-e^{-\beta(\phi(x_T^\Theta + x e^{r(T-t)}, y_T^\Theta, S_T))}\right] \\ &= e^{-\beta x_t e^{r(T-t)}} E_t\left[-e^{-\beta(\phi(x_T^\Theta, y_T^\Theta, S_T))}\right]. \end{aligned} \tag{A-23}$$

Due to the arbitrariness of Θ, we have $V(x_t, y_t, S_t, t) \ge e^{-\beta x e^{r(T-t)}} V(0, y_t, S_t, t)$.

Similarly, for any state (x_t, y_t, S_t, t), let $\tilde{\Theta} = \{(\tilde{I}_s, \tilde{D}_s) : t \le s \le T\}$ be an admissible trading strategy that generates terminal state $(x_T^{\tilde{\Theta}}, y_T^{\tilde{\Theta}}, S_T)$. Then, for the state $(0, y_t, S_t, t)$, consider the following strategy: borrow x_t dollars at time t, follow the strategy $\tilde{\Theta}$ afterwards, and repay the borrowed position at time T. It is easy to see that this strategy will generate a terminal state of $(x_T^{\tilde{\Theta}} - x_t e^{r(T-t)}, y_T^{\tilde{\Theta}}, S_T)$. Therefore, we have

$$\begin{aligned} V(0, y_t, S_t, t) &\ge E_t\left[-e^{-\beta(\phi(x_T^{\tilde{\Theta}} - x_t e^{r(T-t)}, y_T^{\tilde{\Theta}}, S_T))}\right] \\ &= e^{\beta x_t e^{r(T-t)}} E_t\left[-e^{-\beta(\phi(x_T^{\tilde{\Theta}}, y_T^{\tilde{\Theta}}, S_T))}\right]. \end{aligned} \tag{A-24}$$

Due to the arbitrariness of $\tilde{\Theta}$, we have $V(0, y_t, S_t, t) \ge e^{\beta x_t e^{r(T-t)}} V(x_t, y_t, S_t, t)$. The result then follows. □

Proposition 2 suggests that if we define a new function

$$f(y, S, t) = -\frac{1}{\beta} \ln(-V(0, y, S, t)) \tag{A-25}$$

in the domain $\{y \in R, S > 0, t \in [0, T]\}$, then the solution function $V(x, y, S, t)$ takes the following functional form

$$V(x, y, S, t) = -e^{-\beta(xe^{r(T-t)} + f(y,S,t))}. \tag{A-26}$$

Direct substitution shows that the function $f(y, S, t)$ satisfies the following equation

$$\begin{aligned} \max \Bigg\{ \mathcal{L}f \quad &+ \quad \sup_{0 \le i \le \eta_{n_1}} [f_y - (1 + \sum_{j=0}^{n_1 - 1} \theta_j 1_{\{i > \eta_j\}}) S e^{r(T-t)}] i \\ &+ \quad \sup_{0 \le d \le \xi_{n_2}} [(1 - \sum_{j=0}^{n_2 - 1} \alpha_j 1_{\{d > \xi_j\}}) S e^{r(T-t)} - f_y] d, \\ &\qquad f_y - S e^{r(T-t)}(1 + \theta), \;\; S e^{r(T-t)}(1 - \alpha) - f_y \Bigg\} = 0, \end{aligned} \tag{A-27}$$

with terminal condition

$$f(y, S, T) = yS - \alpha(y_T^* - y)^- S - \theta(y_T^* - y)^+ S, \tag{A-28}$$

where the differential operator in (A-27) is given by

$$\mathcal{L}f = \mu S f_S + \frac{1}{2}\sigma^2 S^2 \left(f_{SS} - \beta f_S^2\right) + f_t. \tag{A-29}$$

Calculating the Expected Time to Reach a Target. In order to compute the expected time $E[\tau_\kappa^l]$, where τ_κ^l is defined by (31), we consider the following auxiliary function defined on $\{(y, S, t) : (1 - \kappa)y_0 \le y \le y_0, S > 0, 0 \le t \le T\}$

$$g(y, S, t) = E[\tau_\kappa^l | y_t = y, S_t = S]. \tag{A-30}$$

Then the Feynman-Kac theorem implies that $g(y, S, t)$ satisfies the following equation in the restrained Sell/Buy regions

$$g_t + (i^* - d^*)g_y + \mu S g_S + \frac{1}{2}\sigma^2 S^2 g_{SS} = 0, \tag{A-31}$$

where $i^* = i^*(y, S, t)$ and $d^* = d^*(y, S, t)$ are the optimal buying and selling speed, respectively. In the unrestrained selling region or the unrestrained buying region, $g(y, S, t)$ satisfies the following Neumann condition

$$g_y(y, S, t) = 0. \tag{A-32}$$

This is because in these regions the stock shares are sold/purchased at infinite speed. The boundary condition on $y = (1 - \kappa)y_0$ is clearly given by

$$g((1 - \kappa)y_0, S, t) = t, \tag{A-33}$$

and the boundary condition on $t = T$ is given by

$$g(y, S, T) = T \tag{A-34}$$

because all remaining shares are liquidated at T.

After we solve for the function $g(y, S, t)$, we can then calculate $E[\tau_\kappa^l]$ by the equation

$$E[\tau_\kappa^l] = g(y_0, S_0, 0). \tag{A-35}$$

For the acquisition case, we define

$$\tau_\kappa^a = \min\{t \geq 0 : y_t \geq \kappa y_T^*\}. \tag{A-36}$$

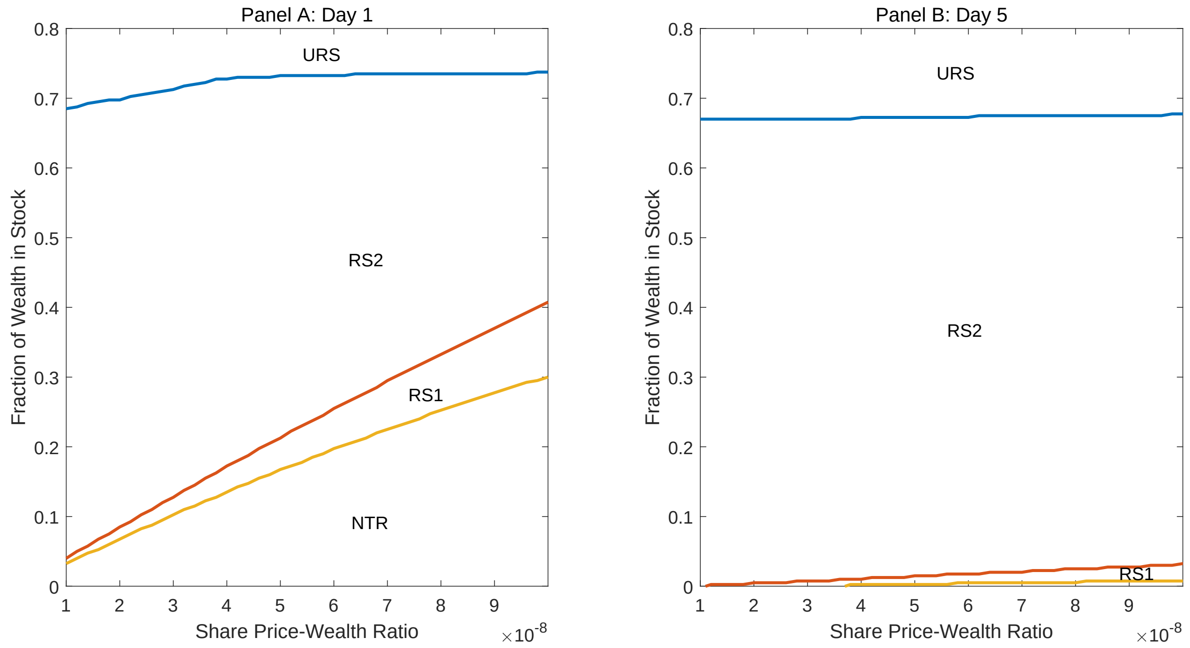

Figure 11: Optimal Liquidation Boundaries with CRRA Preference

This figure shows the optimal trading boundaries of the fund, at the beginning of the first day (left panel) and at the beginning of the last day (right panel). Baseline parameter values: $r = 0.02$, $\mu = 0.1$, $\sigma = 0.2$, $\gamma = 3$, $n_1 = 3$, $n_2 = 2$, $(\theta_0, \theta_1, \theta_2, \theta_3) = (18, 6, 10, 41)$ bps, $(\alpha_0, \alpha_1, \alpha_2) = (33, 2, 49)$ bps, $(\eta_1, \eta_2, \eta_3, \eta_4) = (0, 5.93, 32.24, 209.75)$ millions, $(\xi_1, \xi_2, \xi_3) = (0, 62.63, 287.21)$ millions, and $T = 5/252$ years.

In order to compute $E[\tau_\kappa^a]$, we can similarly consider the following auxiliary function defined on $\{(y, S, t) : 0 \le y \le \kappa y_T^*, S > 0, 0 \le t \le T\}$

$$h(y, S, t) = E[\tau_\kappa | y_t = y, S_t = S]. \tag{A-37}$$

It can be shown that $h(y, S, t)$ satisfies the same Feynmann-Kac equation, with the following boundary condition on κy_T^*: $h(\kappa y_T^*, S, t) = t$.

D Optimal Liquidation with CRRA Preference

In this Appendix, we present the solution to an optimal liquidation problem with a CRRA utility function. We assume the fund has a relative risk aversion coefficient of $\gamma = 3$, and its objective is to liquidate a large number of shares within 5 trading days.

In Figure 11, we plot the optimal liquidation boundaries of the fund at two time points: the beginning of the first day and the beginning of the last day. It suggests that the main insights we derive from the model with CARA preference still carry over to the model with CRRA preference. For instance, it is optimal for the fund to liquidate shares at a state-dependent speed to trade off risk exposure, risk premium and trading cost, and the optimal liquidation speed is a decreasing function of the fund's stock allocation.