Portfolio Liquidity Estimation and Optimal Execution*

Ciamac C. Moallemi
Graduate School of Business
Columbia University
email: ciamac@gsb.columbia.edu

Kai Yuan
Graduate School of Business
Columbia University
email: kyuan17@gsb.columbia.edu

Current Revision: August 2019

Abstract

Estimating liquidity accurately is an important ingredient of portfolio management. Traditionally, liquidity costs are estimated by means of single-asset models. Yet such an approach ignores the fact that, fundamentally, liquidity is a *portfolio* problem: asset prices are correlated. We develop a model to estimate portfolio liquidity costs through a multi-dimensional generalization of the optimal execution model of Almgren and Chriss (1999). Our model allows for the trading of standardized liquid bundles of assets (e.g., ETFs or indices). We show that the benefits of hedging when trading with many assets significantly reduce cost when liquidating a large position. In a "large-universe" asymptotic limit, where the correlations across a large number of assets arise from a relatively few underlying common factors, the liquidity cost of a portfolio is essentially driven by its idiosyncratic risk. Moreover, the additional benefit from trading standardized bundles is roughly equivalent to increasing the liquidity of individual assets. Our method is tractable and can be easily calibrated from market data.

1. Introduction

Estimation of *liquidity costs*, those associated with trading a collection of large positions, is an important issue in modern financial markets. In portfolio management, estimation of liquidity costs is important since these costs can be significant. This is particularly true for investors who are very active (and hence incur significant costs by trading frequently) or are very large (and hence incur significant costs through their size). In such settings, effective portfolio construction decisions cannot be made without considering liquidity costs. Similarly, in risk management, assessment of the risk associated with holding a portfolio depends on both the long-term fluctuations in the value of the underlying assets and the short-term ability to convert the portfolio into cash. This latter effect can be especially important in times of distress, and is fundamentally a question of liquidity costs.

A closely related problem is that of *optimal execution*. In many markets, when an investor seeks to execute a large trade (a so-called "parent order"), it is usually broken into pieces with the help of algorithmic trading systems and executed as a sequence of much smaller trades ("child orders"). Optimal execution problems seek to do this in the most efficient manner by balancing

*The authors wish to thank Paul Glasserman and Albert Menkveld for helpful discussions. Moallemi was supported by NSF grant CMMI-1235023.

two effects. First, there are transaction costs associated with execution, including, for example, commissions, fees, the bid-ask spread, and (most importantly for large investors) the market impact of the trading itself. Second, by spreading out a large trade over time, investors are exposed to risks associated with the movement of market prices over the execution horizon. Traders must evaluate their trading strategies against the transaction costs and market risks. Those who trade too fast incur high transaction costs from market impact while those who trade too slow are exposed to adverse price movements: both trading strategies could potentially result in more than the expected liquidity cost. This trade-off between cost and uncertainty has given rise to a rich literature on optimal execution in general and optimal liquidation of a single risky asset in particular, starting with the work of Almgren and Chriss (2001).

To date, much of the literature on the estimation of liquidity costs and optimal execution has focused on the *single-asset* setting (with several notable exceptions to be discussed shortly). By contrast, we believe that liquidity is fundamentally a *multi-asset* problem that must be addressed at the portfolio level. This is for several reasons:

(i) Investors make trading decisions seldom in isolation on an asset-by-asset basis, but rather jointly to produce a trade list consisting of a portfolio of trades to be made simultaneously in multiple assets. A simple example would be an open-end fund, which, upon an an inflow or outflow, would in effect trade portfolios to maintain proportional holdings. Since the market risk associated with such a trade depends on the joint distribution of correlated assets, the estimation of its liquidity costs will not decompose across assets, nor can optimal trading schedules be determined by considering assets in isolation.

(ii) Even if an investor seeks to trade only a single asset, he may receive significant benefits from simultaneously trading correlated assets for the hedging purposes. For example, an investor unwinding a position in an illiquid asset may seek to hedge the execution risk by establishing positions in correlated but liquid assets, in order to drive down overall liquidity costs.

(iii) Finally, investors may benefit from the multi-asset approach through the trading of what we call *liquid bundles*. These are collections of assets (in effect, portfolios) existing in many markets that can be directly and atomically traded. For example, in equity markets, investors can directly trade exchange-traded funds (ETFs), which are economically (ignoring creation and redemption issues) equivalent to trading a basket of underlying equities. Similarly, in credit markets, trading credit default swap (CDS) indices is equivalent to taking a simultaneous position in a portfolio of underlying credit entities. In futures markets, spread trades, such as calendar spreads, inter-commodity spreads (e.g., crack spreads), and option spreads, are also portfolio trades. Such portfolio instruments can be important both because they provide another mechanism for trading the constituent assets, and because they are often extremely liquid and have little idiosyncratic risk, which makes them excellent candidates as hedging instruments.

In this paper, we develop a multi-asset generalization of the model of Almgren and Chriss (2001), building on the work of Guéant (2015), Kim (2014), and Guéant et al. (2015). Going beyond this earlier work, our model explicitly incorporates the trading of liquid bundles such as ETFs. Our model is easily calibrated and computationally tractable.

The most important contribution of our model, however, is that it enables us to provide a structural analysis of the underlying drivers of liquidity costs. Specifically, we make the assumption of a factor model, where the covariance structure across the universe of tradeable assets decomposes into common, systemic factors (which drive correlations) and individual, idiosyncratic risk. We consider a *large-universe* asymptotic regime, where a large number of assets are available for trading relative to the number of underlying systemic factors. This large-universe setting is consistent with asset pricing theory, particularly the assumptions made in the arbitrage pricing theory (APT), first developed by Ross (1976). It is also consistent with the state of the art in practice, where, for example, commercial risk models for equities (e.g., BARRA) use tens of systemic factors to explain the covariance structure for thousands of assets.

In this asymptotic large-universe setting, under suitable technical assumptions, we develop simple closed-form approximations for liquidity costs. These approximations are useful for computation, but they also highlight two key structural properties of portfolio liquidity costs. First, liquidity costs are primarily driven by idiosyncratic risk. This is because, in a large-universe setting, systemic risk can be hedged very cheaply and asymptotically eliminated. Put differently, the benefit from considering optimal execution at the portfolio level roughly corresponds to reducing risk exposure from total risk to only idiosyncratic risk. Second, introducing a liquid bundle (ETF) is approximately equivalent to commensurately increasing the liquidity of each underlying asset by its implied trading volume in the ETF. In other words, liquid high-volume ETFs can offer significant reductions in liquidity costs.

We explore the practical implications of our model in an empirical example consisting of 29 U.S. equities in the utility sector, along with a sector ETF. There, we demonstrate the above-referenced structure effects and illustrate the magnitude of the benefits of our approach. In particular, the portfolio approach to trading single assets in the utility sector can reduce liquidity costs by a factor of up to five. In addition, use of the sector ETF further reduces costs by 10–20%.

The rest of the paper is organized as follows. In Section 1.1, we review the prior literature. In Section 2, we present our model and characterize the solution of the resulting optimal execution problem. In Section 3, we specialize our results to settings with separable transaction costs that are of particular interest. In Section 4, we introduce the large-universe asymptotic regime and establish our main structural results. In Section 5, we provide empirical examples calibrated to market data. Section 6 concludes. Note that all proofs are provided in the Online Supplement.

1.1. Literature Review

Research on optimal execution has been of particular academic interest in the past two decades. It first started with Bertsimas and Lo (1998), who focused on the minimization of execution costs. The trade-off between transaction cost and market risk was first documented by Grinold and Kahn

(2000), and was then used in the seminal papers of Almgren and Chriss (1999, 2001) to derive the framework of single-asset optimal execution in a mean-variance formulation. Initially in discrete time with linear market impact, the Almgren–Chriss model was extended to continuous time by He and Mamaysky (2005) and Forsyth (2011) using the Hamilton–Jacobi–Bellman approach, and by Almgren (2003) and Guéant (2015) using nonlinear market impact functions. Almgren (2012) further takes into account stochastic volatility and liquidity. Whereas these frameworks are all based on static or deterministic strategies in which the number of shares to be sold at any time is pre-specified, Almgren and Lorenz (2007) improves on them with the more realistic mean-variance formulation of a simple update strategy that accelerates execution when the prices move in favor of the trader. A more detailed discussion of the form of adaptivity is given in Lorenz and Almgren (2011).

Perhaps due to its mathematical difficulties, the portfolio approaches to optimal execution is much less studied. Almgren and Chriss (2001), followed by Engle and Ferstenberg (2007) and Brown et al. (2010), briefly discuss the portfolio approach and provide a solution to a simple case. In recent years, the body of work dedicated to the portfolio approach has grown. Kim (2014) considers the case where market impact is assumed to be minimal and decays sufficiently fast to be negligible in price dynamics. Guéant et al. (2015) present a numerical method to approximate the optimal execution strategy based on convex duality. While the framework used in these two papers is quite similar to that of the present paper, our framework is more general and allows for the trading of liquid bundles. Finally, Tsoukalas et al. (2014) analyze a multi-asset optimal execution problem; however, they confine their attention to the microstructure of cross-asset market impact. One key observation to be drawn from all these papers is that there are large hedging benefits by using the portfolio approach.

Our research is also related to empirical research that conduct cross-sectional regressions to estimate market impact. For example, Chacko et al. (2008) provide empirical evidence that the expected market impact is proportional to the square root of the trading size; see also Bouchaud et al. (2008). However, this approach has two downsides: it is extremely noisy (because it is hard to estimate transaction costs from actual returns) and from our perspective, it is fundamentally a single-asset approach. More recently, a literature has emerged on the topic of cross-asset market impact (see, e.g., Benzaquen et al., 2016; Schneider and Lillo, 2017; Min et al., 2018), which also necessitates a portfolio approach to optimal execution.

2. Model

In this section, we describe our general model setup and characterize the solution of the resulting optimization problem.

2.1. Setup

Portfolio and trading strategies. Consider an agent who wishes to liquidate efficiently a portfolio consisting of positions in up to n assets. The agent's initial holdings are specified by the vector $q \in \mathbb{R}^n$, where component q_i represents the initial position in asset i denominated in shares. In order to liquidate this portfolio, the agent can trade $m \geq n$ possible liquid instruments. The vector $y_i \in \mathbb{R}^n$ specifies the composition of the ith instrument in terms of shares of underlying assets. That is, selling one unit of the ith instrument results in the agent's portfolio components being adjusted according the the vector y_i. Denote by $Y \triangleq [y_1, y_2, ..., y_m] \in \mathbb{R}^{n \times m}$ the *liquidation matrix* that characterizes the available instruments.

In the simplest case, the agent is only allowed to directly trade the underlying assets. Then, the tradeable instruments correspond to the underlying assets ($n = m$) and $Y = I$; i.e., the liquidity matrix is the identity matrix. More generally, our model supports tradeable instruments that are not necessarily individual assets, but can be liquid bundles that are essentially portfolios that can be traded directly. As was discussed in Section 1, examples of such liquid bundles include exchange traded funds[1] (ETFs), credit default swap (CDS) indices, and tradeable futures spreads. As an example, consider the following:

Example 1 (Two-asset ETF). *Suppose an agent starts with a portfolio consisting of two stocks, and can trade those stocks directly. In addition, suppose that there exists an ETF of a portfolio consisting of one share of each stock. In this case, the liquidation matrix is given by*

$$Y = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

We will make the assumption that $\text{rank}(Y) = n$, i.e., that Y is full rank, so that any initial portfolio in $q \in \mathbb{R}^n$ can be liquidated with the instruments available.

Given a liquidation matrix Y, a trading strategy is characterized by the rate at which each of the liquid instruments (the columns of Y) are bought or sold. Specifically, a trading strategy is defined by the control process $u \in L^1([0, \infty); \mathbb{R}^m)$, where $u_i(t)$ represents the rate at (in shares per unit time) at which instrument i is traded at time t. We adopt the convention that positive trading rates correspond to selling, while negative trading rates correspond to buying. Given the control u and the initial position q, the evolution of position over time is given by the position process $x \in C([0, \infty); \mathbb{R}^n)$, where

$$x(0) = q, \qquad \dot{x}(t) = -Yu(t), \quad \forall\ t \geq 0.$$

Equivalently,

$$x(t) = q - \int_0^t Yu(s)\, ds, \quad \forall\ t \geq 0.$$

[1] Strictly speaking, an ETF may not be not exactly equivalent or fungible to its underlying portfolio, but we will assume the existence of efficient creation or redemption mechanisms that make them equivalent for our modeling purposes.

Trading constraints. We consider a *constrained liquidity* setting where the trading rate of each instrument is bounded according to

$$|u_i(t)| \leq \gamma_i, \quad \forall\ 1 \leq i \leq m,\ t \geq 0. \tag{1}$$

Here $\gamma_i > 0$ is a bound on on the absolute trading rate of instrument i. Such restrictions on the trading rate are very common in practice, for several reasons. First, an excessive trading rate will almost certainly lead to unfavorable execution prices due to market impact. We will momentarily introduce transaction costs that depend on the trading rate. However, at very high trading rates, the agent will create a significant supply-demand imbalance in the market, and hence transaction costs will be dominated by effects such as information leakage and are difficult to estimate. Empirical evidence on information leakage of large trades is found by Van Kervel and Menkveld (2015) where the authors show that the high frequency traders "prey" on orders that are large. On the other hand, transaction costs for very low trading rates will be dominated by observable quantities such as the bid-ask spread and easy to estimate. Hence, transaction cost models typically are accurate only for a restricted range of trading rates, and the constraint (1) can enforce this range. Finally, observe that constraints of the form (1) are very common in practice, and can be easily calibrated through market parameters. Typically, one might restrict the trading rate to a certain percentage of the future predicted overall market trading volume for a particular instrument.

Transaction costs. We allow for the possibility of trades to be associated with transaction costs. Such costs may arise from, say, commissions or trading fees, the bid-ask spread, or the distortion of market prices caused by the agent's trading. In all of these cases, transaction costs are related to the trading rate. For example, costs associated per share commissions or the bid-ask spread accumulate as a linear function of the trading rate. Market impact may take a more complicated form, but will still be an increasing function of the trading rate.

Though the sources of transaction costs vary, they are all closely related to the trading rate. In particular, if we look at the rate of transaction cost accumulation, the contributions of commission fees and the costs from bid-ask spreads are linear as a function of the trading rate, whereas market impact costs may take a more complicated form such as that studied in Kyle and Obizhaeva (2016). Here, we will not seek to decompose the transaction costs and will describe the total rate of transaction cost accumulation with some functional $f(\cdot)$ of the trading rate u. We assume that $f\colon \mathbb{R}^n \to \mathbb{R}_+$ is a non-negative convex function that is symmetric around 0, i.e., $f(u) = f(-u)$, for all $u \in \mathbb{R}^n$. Further, we assume that no costs are incurred by not trading, i.e., $f(0) = 0$. By making these assumptions, we are essentially focusing on only the temporary (or instantaneous) market impact, which depends only on how fast you trade. We are not considering permanent or transient market impact, which features the impact of current trade on future execution prices.

Price dynamics. The evolution of the price dynamics is typically determined by a predictable drift component and a random noise component. Since the liquidation process typically happens in a short time horizon, we will neglect the drift and focus only on the unpredictable variations.

Specifically, we assume that the prices of the n assets ($S(t) \in \mathbb{R}^n$) follow a multidimensional Brownian motion given by

$$dS(t) = \Sigma^{\frac{1}{2}} dW(t), \tag{2}$$

where $W(t) \in \mathbb{R}^n$ is an n-dimensional standard Brownian motion, and $\Sigma \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix that characterizes the covariance structure of $W(t)$. We will also assume that there are no tracking errors for the liquid bundles. As a result, the price process of any instrument y_i is given by $y_i^\top S(t)$. We will also make the assumption that the covariance matrix Σ is *constant* over the period of liquidation. This may be a reasonable approximation of since the liquidation process we are considering typically takes a short time horizon ranging from hours to days. It is expected that the covariance structure will not change dramatically over such a short time horizon.

Portfolio value and risk. We now discuss the profit and loss resulting from the liquidation process. For any liquidation process defined by (x, u), let IS_t be the implementation shortfall from liquidating the portfolio up to time t. This is defined to be the difference between the value of the initial portfolio at time 0 and the value of the remaining portfolio at time t (along with any intermediate cashflows resulting from trading between time 0 and time t). That is,

$$\begin{aligned} \mathrm{IS}_t &\triangleq \int_0^t (S(0) - S(s))^\top \, dx(s) + \int_0^t f(u(s))\, ds \\ &= -\int_0^t x(s)^\top \, dS(s) + \int_0^t f(u(s))\, ds. \end{aligned} \tag{3}$$

The first term represents the total effect of price changes during the liquidation process up to time t. The second term is the loss due to transaction costs.

The expected value of IS_t takes the form

$$\mathsf{E}[\mathrm{IS}_t] = \int_0^t f(u(s))\, ds. \tag{4}$$

Notice that, by construction $x \in C([0,\infty); \mathbb{R}^n)$. It follows immediately that

$$\int_0^t x(s)^\top x(s)\, ds < \infty. \tag{5}$$

Then, by Ito's isometry, we have

$$\mathrm{Var}(\mathrm{IS}_t) = \int_0^t x(s)^\top \Sigma x(s)\, ds. \tag{6}$$

Let $\mathrm{IS} \triangleq \lim_{t\to\infty} \mathrm{IS}_t$ denote the implementation shortfall incurred over the entire liquidation process; we have

$$\mathsf{E}[\mathrm{IS}] = \int_0^\infty f(u(t))\, dt, \qquad \mathrm{Var}(\mathrm{IS}) = \int_0^\infty x(t)^\top \Sigma x(t)\, dt. \tag{7}$$

The mean of IS is simply the total transaction costs associated with the liquidation process. The variance of IS provides us with a natural measure of market risk during the liquidation process.

Optimization problem. The optimal liquidation problem can be formulated by minimizing the expected implementation shortfall adjusted for the risk according to a mean-variance objective:

$$
\begin{aligned}
J^*(q) \triangleq \underset{u}{\text{minimize}} \quad & \int_0^\infty f(u(t))\,dt + \mu \int_0^\infty x(t)^\top \Sigma x(t)\,dt \\
\text{subject to} \quad & \dot{x}(t) = -Y u(t), && \forall\ t \geq 0, \\
& |u_i(t)| \leq \gamma_i, && \forall\ 1 \leq i \leq m,\ t \geq 0, \\
& x(0) = q, \\
& u \in L^1([0,\infty);\mathbb{R}^m).
\end{aligned}
\tag{8}
$$

Here, $\mu > 0$ is a parameter capturing the degree of the agent's risk aversion.

The objective value of this dynamic control problem captures an explicit trade-off between transaction costs and market risk. If the agent trades faster, he is more likely to end up with higher transaction costs due to increased market impact; if he trades slower, he will end up facing more market risk over a longer period of time. As we will show in Theorem 1, the optimal liquidation process always has the finite objective value defined in (8). This implies that the asymptotic position is zero as time goes to infinity; otherwise the risk component of the objective value in (8) would be infinite. An alternative would be to explicitly impose a exogenous finite time horizon by which the entire position must be liquidated, and this might be more appropriate in a fire sale setting, for example. Many of the results in this paper would hold in such an alternative, but we will opt for the simplicity of an endogenous time horizon.

Note that explicit in the formulation (8) is the fact that we are restricting attention to only deterministic strategies; in other words, we are requiring that trading rates for each asset at every time to be specified in advance at time $t = 0$. In general, there may be adaptive or stochastic strategies that perform better for our mean-variance objective. For example, Almgren and Lorenz (2007) show that stochastic strategies may outperform the best deterministic strategy; see also Lorenz and Almgren (2011). However, the proper economic motivation for mean-variance objective comes from the problem of maximizing expected utility for exponential, or CARA, utility functions. Schied et al. (2010) found that there is no added utility from adaptive strategies for CARA investors with a finite time horizon. Schöneborn (2011) expands this observation to infinite time horizons. As a result, if we believe the mean-variance objective stems from the optimization of CARA utility functions, the deterministic strategy is optimal. In any case, we will restrict our attention to deterministic strategies. This is consistent with much of the rest of the mean-variance optimal execution literature.

2.2. Optimal Strategy

In this section, we discuss some of the general characteristics of optimal strategies in our formulation.

Theorem 1 (Existence and Convexity). *The dynamic control problem defined in* (8) *is bounded and an optimal solution* u^* *always exists. In addition, the optimal value (the liquidity cost) is convex in initial position.*

The proof of the theorem is given in the Online Supplement and is similar to results of Guéant (2015); Guéant et al. (2015). In our setting, the main technical requirement for existence of an optimal solution is the constrained liquidity assumption (1). This helps us to establish the equi-integrability of the feasible set. Because of this, our result is in some ways simpler than the earlier work. For example, we do not need to impose additional requirements on the transaction cost functional $f(\cdot)$ beyond the convexity. Note that one implication of the existence theorem is that the optimal objective value is finite. This implies that, as $t \to \infty$, $x(t) \to 0$. In other words, the position will be asymptotically liquidated.

A key element in our framework is that we allow for the direct trading of liquid bundles. As such, we may have more instruments than individual assets ($m > n$), and it is possible to have more than one trading strategy $u(\cdot)$ corresponding to any given trajectory of position $x(\cdot)$. Therefore, the uniqueness of the optimal trading strategy may not be guaranteed. However, the optimal trajectory of position $x(\cdot)$ must be unique, i.e., all optimal solutions have the same position at any time. Moreover, as established in the following theorem, under an additional convexity assumption the trading strategy must also be unique:

Theorem 2 (Uniqueness). *All optimal solutions for the the dynamic control problem in* (8) *have a unique optimal position trajectory* $x* \in C([0,\infty);\mathbb{R}^n)$. *Moreover, if the transaction cost functional* $f(\cdot)$ *is strictly convex, the optimal trading strategy* $u^* \in L^1([0,\infty;\mathbb{R}^m)$ *must also be unique.*

In general, it is difficult to come up with closed-form solutions to the dynamic control problem given in (8) (although we will consider some special cases in Section 3). We provide sufficient conditions for optimality by exploiting the convexity of the problem in the following:

Theorem 3 (Sufficiency). *A feasible pair* $(x^*, u^*) \in C([0,\infty);\mathbb{R}^n) \times L^1([0,\infty);\mathbb{R}^m)$ *form an optimal solution of* (8) *if, for all* $t \geq 0$,

$$x^*(t) = q - \int_0^t Y u^*(s)\, ds,$$

$$u^*(t) \in \underset{u:\ -\gamma \leq u \leq \gamma}{\operatorname{argmin}}\ f(u) - 2\int_t^\infty x^*(s)^\top \Sigma Y u\, ds. \tag{9}$$

Theorem 3 provides a sufficient condition for the optimal trading strategy. Intuitively, the optimal trading rate at any given time results from a trade-off between the two components in (9). The first component represents the instantaneous transaction cost and the second component represents the impact on future risks. Note that Theorem 3 gives a sufficient condition, but not a necessary one. If, however, the liquidation process takes only finite time, it can be shown that, (9) is also necessary, using Pontryagin's minimum principle, as we will do later. The necessity is difficult to generalize to an infinite trading horizon, where the corresponding general version of Pontryagin's minimum principle is often pathological (Halkin, 1974).

3. Examples: Separable Transaction Costs

In the optimization problem (8), decision making across multiple assets is coupled. This coupling arises from two sources: (1) asset prices may be correlated impacting risk considerations; and (2) there may cross-asset market impact interactions embedded in the transaction cost functional $f(\cdot)$. In practice, the existence of cross-asset market impact is debatable and, more importantly, it is extremely difficult to measure if it exists. Hence, the most relevant case, which we discuss now, involves the absence of any cross-asset market impact.

Specifically, the class of transaction cost functionals that are of particular interest are what we call *separable transaction costs*. These are transaction cost functionals that take the form of

$$f(u) = \sum_j \nu_j \hat{f}(u_j/\gamma_j), \tag{10}$$

for $u \in \mathbb{R}^n$, where $\hat{f}\colon \mathbb{R} \to \mathbb{R}_+$ is a nonnegative convex function symmetric around 0 with $f(0) = 0$. The scaling constant $\nu_j > 0$ captures the magnitude of the transaction cost of asset j, and γ_j is the maximum trading rate from (1).

The intuition behind this type of functional is that the transaction cost of each asset is driven by similar mechanisms and depends primarily on the relative trading rate (u_j/γ_j). This corresponds to the fact that assets with higher liquidity (higher γ_j) are expected to have a smaller transaction cost given the same trading rate. Additionally, (10) rules out the possibility of cross-asset market impact. Though advocated by some approaches in the literature such as Tsoukalas et al. (2014), cross-asset market impact is extremely difficult to estimate. On the other hand, the transaction costs of the form in (10) can be estimated relatively easily from historical transaction data. For example, a typical model of market impact costs for a single asset may take the form (see, e.g., Almgren et al., 2005),

$$\sigma_j \left| \frac{u_j}{V_j} \right|^\alpha,$$

where σ_j is the asset volatility, V_j is the asset volume, and the exponent α is a model parameter. This is clearly of the form specified by (10).

In what follows, we focus specifically on two examples of separable transaction costs: zero-cost constrained liquidity and linear-cost constrained liquidity.

3.1. Zero-Cost Constrained Liquidity

The simplest case of a separable transaction cost is where there is a constraint on the trading rate, but trading itself doesn't incur any cost. In this case we simply assume that $f(u) = 0$. Thus, we are capturing a setting where trading costs are minimal relative to risk, e.g., when the agent is very patient and tends to trade slowly and passively. A potential such setting would be an agent that trades only passively with mid-point orders, and where there is no cost arising either from the bid-ask spread or from market impact.

Under zero-cost constrained liquidity, (8) is equivalent to

$$\begin{aligned} J^*(q) \triangleq \underset{u}{\text{minimize}} \quad & \int_0^\infty x^\top(t)\Sigma x(t)\,dt \\ \text{subject to} \quad & \dot{x}(t) = -Yu(t), && \forall\ t \geq 0, \\ & |u_i(t)| \leq \gamma_i, && \forall\ 1 \leq i \leq m,\ t \geq 0, \\ & x(0) = q, \\ & u \in L^1([0,\infty);\mathbb{R}^m). \end{aligned} \tag{11}$$

Here, we have assumed without loss of generality that the parameter for risk aversion (μ) takes the value $\mu = 1$. This setup under zero-cost constrained liquidity is very similar to that developed by Kim (2014). The main difference is that we allow for the trading of liquid bundles and hence Y does not need to be an identity matrix.

The zero-cost constrained liquidity model has some interesting features. The first is that the problem scales as a function of initial position.

Theorem 4 (Scaling). *If u^* is optimal for the problem (11) starting from initial position $x(0) = q$, then*

$$\tilde{u}(t) \triangleq u^*(t/\alpha), \quad \forall\ t \geq 0,$$

is optimal for the problem starting with initial position $x(0) = \alpha q$ for all $\alpha > 0$. Furthermore,

$$J^*(\alpha q) = \alpha^3 J^*(q).$$

Interestingly, the optimal object value, which is essentially the variance of the liquidation P&L, scales with initial position to the power of three. This is consistent with the prediction of inventory risk models (e.g., Chapter 16, Grinold and Kahn, 2000) where the total trading cost of a portfolio increases according to the 3/2 power of the amount traded.

Another property of the zero-cost constrained liquidity model is that the optimal liquidation process completes in finite time. Although we do expect the position to be liquidated asymptotically (as discussed in Section 2.1), in theory our framework does not guarantee that liquidation occurs in finite time. It could be that as the position gets smaller, transaction costs dominate the market risk, in which case it may make sense to trade slower and slower to keep transaction costs small. One example is the case where the transaction cost is quadratic and (8) becomes a constrained linear-quadratic control problem, which takes infinite time. However, under zero-cost constrained liquidity, where the finiteness of the liquidation time is guaranteed, the theorem goes as follows.

Theorem 5 (Finite Horizon). *For any initial position q, the optimal position trajectory $x(t)$ is guaranteed to reach zero in finite time.*

Recall that, in general, condition (9) of Theorem 3 provided a sufficient but not necessary condition to characterize an optimal trading policy. In the zero-cost constrained liquidity setting, however, since the optimal liquidation process takes only finite time, Pontryagin's minimum principle can be used to derive the necessity of (9).

Lemma 1 (Optimality). *A feasible control u^* is optimal for* (11) *if and only if, for all $t \geq 0$,*

$$x^*(t) = q - \int_0^t Y u^*(s)\, ds,$$

$$u^*(t) \in \underset{u:-\gamma \leq u \leq \gamma}{\operatorname{argmax}} \left(\int_t^\infty x^*(s)^\top \Sigma Y \, ds \right) u. \tag{12}$$

Condition (12) suggests a so-called "bang-bang" or singular control, where each $u_j(t)$ takes the upper bound γ_j if the jth component of $\int_t^\infty x^*(s)^\top \Sigma Y ds$ is positive, and takes the lower bound $-\gamma_j$ if it is are negative. Properties of singular control problems can be found in Johnson and Gibson (1963), among other places.

In general, it is difficult to come up with a closed-form solution to the problem (11) in high dimensions. However, it is possible to characterize the two-dimensional case. Although it is possible to provide closed-form solutions to all two-dimensional cases, we are particularly interested the case of *high liquidity hedging*, where an initial position in a single asset is liquidated by concurrently trading with a second, hedging asset that is both highly liquid and correlated with the original asset. In practice, this is the most relevant setting, as people tend not to hedge with illiquid assets. The following theorem characterizes this setting.

Theorem 6 (High Liquidity Hedging). *In the two-dimensional case where model parameters are given by*

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_1 & \sigma_2^2 \end{bmatrix}, \qquad Y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

if we further assume that

$$\gamma_2 \geq |\rho| \frac{\sigma_1 \gamma_1}{\sigma_2}, \tag{13}$$

then the optimal liquidity cost of portfolio $q = (q, 0)$ is given by

$$J^*(q) = \frac{1}{3} \frac{q^3}{\gamma_1} \sigma_1^2 \left(1 - \frac{\rho^2}{1 + |\rho| \frac{\sigma_1 \gamma_1}{\sigma_2 \gamma_2}} \right). \tag{14}$$

We refer to condition (13) as the *high-liquidity hedging condition.* It requires that the liquidity of the hedging instrument (γ_2) exceed a certain threshold. We can rewrite (13) as

$$\frac{\gamma_2}{\gamma_1} \geq \frac{|\rho|\sigma_1}{\sigma_2}.$$

The right side is an *optimal hedging ratio.* Given a unit of asset 1, it can be shown that the optimal amount of asset 2 needed to minimize total risk is given by $|\rho|\sigma_1/\sigma_2$. By examining the proof, one finds that the optimal trading strategy is to trade asset 2 as a hedge before unloading the hedged portfolio. And $|\rho|\sigma_1/\sigma_2$ is the optimal quantity of asset 2 needed to hedge every unit of asset 1.

Equation (14) suggests some very intuitive structural properties of liquidity costs. The first term ($\frac{1}{3}\frac{q^3}{\gamma_1}\sigma_1^2$) is the fair liquidity cost of trading asset 1 alone without hedging. The second term can then be interpreted as the benefit from hedging. Given that (13) holds, it is easy to see that the hedging benefit is increasing in $|\rho|$, which captures the correlation between the two assets. This indicates that hedging is more efficient when one use highly correlated assets. Additionally, the hedging benefit is increasing in $\gamma_2\sigma_2$, a fact that can be interpreted as the rate of risk transferred by trading the hedging asset.

3.2. Linear-Cost Constrained Liquidity

Now we consider the case where transaction costs are determined by the following linear function:

$$f(u) = \sum_j \nu_j |u_j|. \tag{15}$$

Notice that this definition is still consistent with (10) if we have

$$\hat{f}(u) = |u|$$

and if we define v_j as the coefficient. Now ν_j can be viewed as the bid-ask spread of asset j. Basically, then, the agent is a liquidity taker and (8) can be written as

$$\begin{aligned} J^*(q) \triangleq \underset{u}{\text{minimize}} \quad & \int_0^\infty \nu_j |u_j(t)|\, dt + \mu \int_0^\infty x^\top(t)\Sigma x(t)\, dt \\ \text{subject to} \quad & \dot{x}(t) = -Y u(t), \qquad \forall\ t \geq 0, \\ & |u_i(t)| \leq \gamma_i, \qquad \forall\ 1 \leq i \leq m,\ t \geq 0, \\ & x(0) = q, \\ & u \in L^1([0,\infty); \mathbb{R}^m). \end{aligned} \tag{16}$$

The scaling property ceases to hold in this case as the transaction costs are typically linear in the position traded, whereas the risk component is at least quadratic. Again, a general closed-form solution is beyond our reach, but we can still explicitly solve the one-dimensional and two-dimensional cases.

Suppose we need to liquidate a certain position in asset 1. In the one-dimensional case where hedging is not possible, the total transaction costs incurred for trading a certain position are fixed and do not depend on the trading rate. As a result, the optimal strategy is to sell the position as fast as possible (at rate γ). Hence we get the following proposition:

Theorem 7 (One Asset). *In the one-dimensional case, the cost of liquidating a position of q with parameters (σ, γ, ν) is given by*

$$J^*(q) = \nu|q| + \mu\frac{|q|^3\sigma^2}{3\gamma}.$$

The first term represents the total transaction costs associated with liquidating a position of q, and the second term represents the market risk of this liquidating process.

The two-dimensional case is more complicated – yet tractable. We consider only the case of "high liquidity hedging," where (13) holds.

Theorem 8 (Two Assets). *Consider the two-dimensional case where model parameters are given by*

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_1 & \sigma_2^2 \end{bmatrix}, \qquad Y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

and assume the portfolio to be liquidated contains q shares of asset 1 and no position in asset 2. If we further assume that

$$\gamma_2 \geq |\rho| \frac{\sigma_1\gamma_1}{\sigma_2},$$

then the asset 2 will only be used to hedge if and only if

$$q^2 \geq \frac{2\gamma_1\nu_2}{\mu\gamma_2\rho\sigma_1\sigma_2}. \tag{17}$$

If (17) *is satisfied, then the optimal liquidity cost of the portfolio is given by*

$$J^*(q) = \frac{\mu}{3}\frac{q^3}{\gamma_1}\sigma_1^2\left(1 - \frac{\rho^2}{1+|\rho|\frac{\sigma_1\gamma_1}{\sigma_2\gamma_2}}\right) + \nu_1 q + 2\nu_2 q \frac{\gamma_2}{\gamma_1}\frac{|\rho|\frac{\sigma_1\gamma_1}{\sigma_2\gamma_2}}{1+|\rho|\frac{\sigma_1\gamma_1}{\sigma_2\gamma_2}} - \frac{4\frac{\nu_2}{\sigma_2}\sqrt{\frac{2\nu_2|\rho|\sigma_1\gamma_1}{\sigma_2}}}{3(1+|\rho|\frac{\sigma_1\gamma_1}{\sigma_2\gamma_2})}. \tag{18}$$

In this case, hedging with other assets comes with transaction costs, which are proportional to the hedging position acquired. Intuitively, if the transaction costs strictly dominate the market risk, the agent will find hedging unattractive. Theorem 8 indicates that hedging is only optimal when the position passes the threshold defined in (17). This provides intuition about the trade-off between hedging benefits and their associated transaction costs. Specifically, hedging is less likely to be beneficial when:

1. The position size is small.

2. The covariance between the two assets is small.

3. The transaction costs for the hedging asset (asset 2 in this case) are large.

4. The agent is not risk averse (i.e., there are smaller μ).

Additionally, the structure of (18) is interesting. We can see that the first term of (18) is exactly the liquidity cost given in (14) in Theorem 6, which is the liquidity cost for the case of a zero transaction cost. The second and third terms are the transaction costs associated with the trading strategy given in Theorem 6. The last term is the penalty that results from the fact that we expect less hedging in the presence of transaction costs. Interestingly, the penalty term is some constant that does not depend on the liquidating position. Thus, if the position is too small, hedging is not worthwhile; otherwise the optimal liquidity cost is the cost associated with the optimal strategy in the zero transaction cost case minus a constant that does not depend on position size.

4. Large Universe

Although deriving a closed-form solution for (8) proves to be difficult, it is not hard to see that the optimal liquidity cost is determined primarily by two factors: the covariance structure of prices and transaction cost functionals. We have discussed several transaction cost models in Section 3; now we consider the covariance structure of prices. Throughout this section, we will assume separable transaction costs.

In the most straightforward case where all asset prices are independent of each other, the liquidation problem of a portfolio consisting of n assets will degenerate to n one-dimensional subproblems where each asset is liquidated on its own. But if the asset prices are correlated, the story is more complicated. First of all, decisions regarding the liquidation of assets across a portfolio are coupled. Second, it might be beneficial to hedge a position's market risk by acquiring some assets that are negatively correlated with the liquidating portfolio, as long as the extra transaction costs are acceptable. However, since the complexity of covariance structure grows with asset numbers, it becomes extremely difficult to provide an intuitive analysis of the liquidity costs. Unless we can somehow decrease the dimensions of the problem, very little can be said. By the same logic, the widely accepted idea that the variations of the prices of a large number of assets can be modeled by a small number of systemic factors becomes appealing.

Various models have been developed in finance to model the structure of the covariance matrix of asset prices. Here, we consider the multi-factor risk model first developed by Ross (1976) and then generalized by Chamberlain and Rothschild (1983). The multi-factor risk is the basis for the arbitrage pricing theory which is well-studied in the finance literature. The main idea is that covariance across asset prices can be decomposed into two components: a systemic one and an idiosyncratic one. The systemic component is then modeled through various systemic factors that characterize different sources of systemic risks. This type of model has been widely used in the industry to predict risk structure in the solution of practical investment problems, e.g., the BARRA model from MSCI.

We define $F(t) \in \mathbb{R}^K$ to be the K-dimensional factor process. Without loss of generality, we assume that the factors are orthonormal[2] and follow a standard K-dimensional Brownian motion. Under a continuous time version of the multiple-factor model, the price dynamics of asset i can be written as

$$dS_i(t) = l_i^\top dF(t) + \varsigma_i dz_i(t), \quad 1 \le i \le n, t \ge 0, \tag{19}$$

where $z_i(t)$ is a standard Brownian motion representing the idiosyncratic shocks for asset i. $l_j \in \mathbb{R}^K$ is the loading vector for asset j. ς_i is the magnitude of the asset's idiosyncratic risk and hence

[2] If the factors are correlated, we can always find a new set of K orthonormal factors with a change of coordinate, as long as the covariance matrix of factors is full rank.

$\varsigma_i z_i(t)$ represents the idiosyncratic disturbances that are zero-mean and independent across assets. In addition, we assume that $z_i(t)$ is independent of the factors.

Usually, the number of assets is much larger than the number of underlying factors. For example, in the BARRA equity multi-factor model, the covariance structure of thousands of U.S. equities is explained by 60 industry factors, 12 style factors, and one country factor. This inspires us to explore the large-universe setting.

Now consider a sequence of problems with an increasing universe of securities, where the nth problem contains the first n assets. The nth problem is characterized by asset price covariance matrix $\Sigma^{(n)}$. Now we can see that the definition in (2) is equivalent to (19) if the following decomposition holds:

$$\Sigma^{(n)} = L^{(n)}(L^{(n)})^\top + \Xi^{(n)}, \tag{20}$$

where $L^{(n)} = (l_1, ..., l_n)$ is the factor loadings of the assets and

$$\Xi^{(n)} \triangleq \mathrm{diag}(\varsigma_1^2, \varsigma_2^2, ..., \varsigma_n^2)$$

captures the idiosyncratic risk contribution.

Now we define $\lambda_{\min}^{(n)}$ to be the smallest eigenvalues of $L^{(n)}(L^{(n)})^\top$. The notion of *large universe* is defined as follows:

Definition 1 (Large Universe). *The sequence of problems is said to satisfy the* large-universe *property if the following conditions hold:*

1. *The magnitude of the idiosyncratic risk for each asset is bounded above,*

$$\sup_j \varsigma_j^2 < \infty.$$

2. *The smallest non-zero eigenvalue of $L^{(n)}(L^{(n)})^\top$ goes to infinity as n goes to infinity:*

$$\liminf_{n\to\infty} \lambda_{\min}^{(n)} = \infty.$$

3. *We will assume that the trading rate of assets satisfy:*

$$\lim_{n\to\infty} \underline{\gamma_n} \sqrt{\lambda_{\min}^{(n)}} = \infty,$$

where $\underline{\gamma_n} = \inf_{j\le n} \gamma_j$.

The first condition indicates that the idiosyncratic risk of each asset is upper bounded by some positive number. This condition basically says that the idiosyncratic risk for each asset is small and hence can be diversified away.

The second condition has two interpretations: firstly, the factors are pervasive, in the sense that each factor affects almost all of the assets; secondly, there have to be enough variations in the factor loadings; otherwise, some factors may become redundant as their loadings can be explicitly calculated from the loadings of other factors. This condition can be linked to the conditions of arbitrage pricing theorem of Chamberlain and Rothschild (1983). The intuition is that we can potentially approximate the return of each factor after diversifying away the idiosyncratic risks. This condition is also related to the literature on estimating factor models, for example Fan et al. (2013). Those works typically make a stronger assumption, which requires the smallest non-zero eigenvalue to be linear on n, in order to asymptotically estimate the factor decomposition.

The third condition requires that there is enough liquidity for each asset.

Proposition 1 (Factor Replicating Portfolio). *If the* large-universe *conditions hold, then for each factor $F_i(t)$, there exists a series of portfolios $\{p^{(i,n)}(t)\}$ defined by weights $\{\beta_j^{(i,n)}\}$ where*

$$p^{(i,n)}(t) \triangleq \sum_{j=1}^{n} \beta_j^{(i,n)} S_j(t),$$

such that

1. *The portfolio $p^{(i,n)}(t)$ has unit exposure on factor $F_i(t)$:*

$$p^{(i,n)}(t) - F_i(t) = \epsilon^{(i,n)}(t),$$

where $\epsilon^{(i,n)}(t)$ is zero mean and independent of all factor-price processes, and has variance upper bounded by

$$Var(\epsilon^{(i,n)}(t)) \leq \frac{\sup_j \varsigma_j^2}{\lambda_{\min}^{(n)}} t.$$

2. *The sum of the squares of the weights converge to* 0*:*

$$\lim_{n\to\infty} \sum_{j=1}^{n} (\beta_j^{(i,n)})^2 = 0.$$

This proposition indicates that in the *large-universe* regime, we can construct a sequence of well-diversified portfolios with returns that eventually converge to the factor returns. The intuition is that as the number of tradeable assets increases, we can potentially take a small position in each asset and the idiosyncratic risks will be canceled out due to diversification. The proposition also provides an upper bound on the idiosyncratic risks for the *factor portfolios*, which is given by the ratio between the maximum idiosyncratic variance and the smallest non-zero eigenvalue of $L^{(n)}(L^{(n)})^\top$. On one hand, if the assets have larger idiosyncratic risks, diversification becomes more difficult. On the other hand, achieving perfect diversification also depends on the assumption that the smallest non-zero eigenvalue of $L^{(n)}(L^{(n)})^\top$ goes to infinity, which is guaranteed by the *large-universe* conditions.

The second part of the proposition implies that $\beta_j^{(i,n)} \to 0$ as $n \to \infty$; hence we can construct those portfolios without trading too much of any asset. Combined with the third condition for *large-universe* regime, this suggests that factor portfolios can be traded very quickly.

4.1. Zero-Cost Constrained Trading

To start with, we will adopt the zero-cost constrained trading model where the transaction cost of each asset is represented by a constraint on its maximum trading rate. For simplicity, we assume that the Y matrix is just the identity matrix; hence only individual assets are traded. We will later expand the results to the case of liquid bundles.

Now, consider a sequence of problems, indexed by n, where in the nth problem there are n tradeable assets and we wish to liquidate a portfolio $q \in \mathbb{R}^n$ with positions in at most the first m assets, i.e.,

$$q_j = 0, \quad \forall j > m. \tag{21}$$

Further, define $J_n^*(q)$ to be the optimal liquidity cost of portfolio q. If we only consider the idiosyncratic risks, this will result in less risk for the portfolio and hence should provide a lower bound for liquidity costs. As there is no correlation between assets, the problem will also be separable and can be solved asset by asset. By applying the results from Section 3.1, we have the following:

Theorem 9 (Lower Bound of Hedging Benefits). *The liquidity cost is lower bounded according to*

$$J_n^*(q) \geq \sum_{j=1}^{m} \frac{\varsigma_j^2}{3\gamma_j} |q_j|^3. \tag{22}$$

The lower bound in Theorem 9 captures the situation where the portfolio has zero exposure to any of the risk factors. In this case, no other assets are needed for hedging and hence the liquidity cost consists only of idiosyncratic risks of assets already in the portfolio. Since this situation is of rare occurrence, the question is whether the lower bound is informative. In the following theorem, we try to prove that the lower bound in Theorem 9 is tight under the *large-universe* regime.

Consider the sequence of problems, indexed by n, discussed in the previous section. As we expand the set of assets that can be used for hedging, the liquidity cost should go down simply because we have more choices for hedging.

Theorem 10 (Large Universe). *If the* large-universe *property is satisfied, then, asymptotically, the liquidity cost of any portfolio consisting of finitely many assets will be driven purely by idiosyncratic risks. More specifically, we have*

$$J_\infty^*(q) = \lim_{n\to\infty} J_n^*(q) = \sum_{j=1}^{m} \frac{\varsigma_j^2}{3\gamma_j} |q_j|^3, \tag{23}$$

where q is a portfolio with positions in at most the first m assets, and $J_n^*(q)$ is the optimal cost of liquidating q while trading at most the first $n \geq m$ assets.

Theorem 10 guarantees the convergence of liquidity cost when the number of tradeable assets goes to infinity. In showing what drives liquidity costs, Theorem 10 is important for two reasons. Firstly, from a risk perspective, only the idiosyncratic risks matter. Secondly, from a computational perspective, we can simply use (23) to approximate actual cost if the large-universe setting is valid, instead of solving some complex dynamic control problem as in (11).

However, this only answers part of the question: (23) is still impractical if the convergence is too slow. The following theorem addresses this problem by explicitly bounding the rate of convergence.

Theorem 11 (Convergence Speed). *Asymptotically, the difference between the liquidity cost and the theoretical limit converges at rate* $1/(\underline{\gamma}_n\sqrt{\lambda_{\min}^{(n)}})$*:*

$$\limsup_{n\to\infty} \underline{\gamma}_n \sqrt{\lambda_{\min}^{(n)}} |J_n^*(q) - J_\infty^*(q)| < \infty. \tag{24}$$

Theorem 11 says that the liquidity cost converges to the theoretical value roughly at the speed of one of $1/\sqrt{\lambda_{\min}^{(n)}}$. For a concrete example, consider a simple case where the factor loadings of assets are drawn in an independent and identically distributed fashion. We then have the following theorem:

Theorem 12 (Random factor loading). *If the asset factor loadings are drawn independently from a* K*-dimensional distribution (with a finite second moment), then, asymptotically, we have*

$$\frac{\lambda_{\min}^{(n)}}{n} \xrightarrow{a.s.} C, \tag{25}$$

where C *is some constant that depends on only the distribution of factor loadings. If we further assume that*

$$\inf_j \gamma_j > 0,$$

then we have

$$\limsup_{n\to\infty} \sqrt{n}|J_n^*(q) - J_\infty^*(q)| < \infty, a.s. \tag{26}$$

Theorem 12 shows that if the asset factor loadings are i.i.d., the liquidity cost converges to the large-universe approximation at a rate of at least $1/\sqrt{n}$.

4.2. Vanishing Bid-Ask Spread

So far, we have explored the asymptotic features of liquidity costs for the model of zero-cost constrained trading. Notice that in this case the only cost incurred by hedging comes from the extra idiosyncratic risks added by trading other assets. The intuition here is that if there are many assets to choose from, we can construct a perfect hedging portfolio by trading only a small amount of each asset. By assuming a certain covariance structure, the *large-universe* conditions ensure the availability of such portfolios.

In reality, however, hedging with other assets is almost always associated with additional costs originating from commission fees, bid-ask spreads, and possibly price impact. In such cases it would be more interesting if the results for the *large-universe* regime could be extended to models with non-zero transaction costs. Fortunately, it can be shown that similar results as in (10) can be extended to a class of models with a *separable transaction cost.* More specifically, we define the *vanishing bid-ask spread* condition as follows.

Definition 2 (Vanishing Bid-Ask Spread). *The* vanishing bid-ask spread *condition holds if the transaction cost functional is twice differentiable with*

$$\hat{f}'(0) = 0, \quad \hat{f}(0) = 0.$$

The idea is that there is no transaction cost for trading very small quantities. One such example is the case of quadratic transaction cost, which is documented in Gârleanu and Pedersen (2013).

Again, we consider the case where Y is the identity matrix. Here, the optimization problem we are considering becomes

$$\begin{aligned} J^*(q) \triangleq \underset{u}{\text{minimize}} \quad & \int_0^\infty \sum_{j=1}^m \nu_j \hat{f}(u_j(t))\, dt + \mu \int_0^\infty x^\top(t) \Sigma x(t)\, dt \\ \text{subject to} \quad & \dot{x}(t) = -Y u(t), \qquad \forall\, t \geq 0, \\ & |u_i(t)| \leq \gamma_i, \qquad \forall\, 1 \leq i \leq m,\ t \geq 0, \\ & x(0) = q, \\ & u \in L^1([0,\infty); \mathbb{R}^m). \end{aligned} \tag{27}$$

For any liquidation model specified by (27), let's consider the one-dimensional case ($n = 1$), where only one asset of size q is traded. Additionally, we assume that the asset has a transaction cost parameter ν and a liquidity parameter γ. If we consider only the idiosyncratic risk of this asset, which is ς^2, the corresponding optimal liquidity cost is given by

$$\begin{aligned} \tilde{J}^*(q,\varsigma,\nu,\gamma) \triangleq \underset{u}{\text{minimize}} \quad & \mu\varsigma^2 \int_0^\infty x^2(t)\, dt + \nu \int_0^\infty \hat{f}(u(t))\, dt \\ \text{subject to} \quad & \dot{x}(t) = -u(t), \qquad \forall\, t \geq 0, \\ & |u(t)| \leq \gamma, \qquad \forall\ \ t \geq 0, \\ & x(0) = q, \\ & u \in L^1([0,\infty); \mathbb{R}^1). \end{aligned} \tag{28}$$

It is easy to check that (28) satisfies the conditions in 1 and 2. Then the optimal liquidity cost is just a function of $q, \varsigma, \gamma, \nu$, and we denote it by $\tilde{J}^*(q,\varsigma,\nu,\gamma)$. The specific form of $\tilde{J}^*(q,\varsigma,\nu,\gamma)$ depends on the corresponding transaction cost function $\hat{f}(\cdot)$, and can be solved through HJB equations.

Theorem 13 (Generalization). *For any liquidation model specified in 27, if the transaction cost* *functional is twice differentiable with*

$$\hat{f}'(0) = 0, \quad \hat{f}(0) = 0,$$

the extended results of Theorem 10 still hold. More specifically, we have

$$\lim_{n\to\infty} J_n^*(q) = \sum_{j=1}^{m} \tilde{J}^*(q_j, \varsigma_j, \nu_j, \gamma_j), \tag{29}$$

where $J_n^*(q)$ represents the optimal liquidity costs for a portfolio q with assets in $\mathcal{A}_n$.

Theorem 13 says that the liquidity cost of any portfolio consisting of only finitely many assets is equal to the cost of liquidating each asset individually but with only idiosyncratic risks.

4.3. Linear Transaction Costs

The previous theorem depends on the assumption that the average transaction cost diminishes when the position is very small. Now we consider the case where transaction costs are determined by the following linear function:

$$f(u) = \sum_j \nu_j |u_j|.$$

In this case, ν_j can be viewed as the bid-ask spread of asset j. Without hedging, the optimal strategy is to sell the position as soon as possible (at trading rate limit γ), as is illustrated in Proposition 7.

Unlike in the setup of the previous model, here hedging with other assets is not cost-free. With linear transaction costs, the agent can no longer make the transaction cost vanish by trading small positions. It is always costly to incur any other positions.

Theorem 8 says that hedging is desirable if the position is large. Now, in the case of a large position, the question is, will a similar version of Theorem 10 hold asymptotically?

Theorem 14 (Linear costs)**.**

$$\lim_{n\to\infty} \lim_{||q||_\infty\to\infty} \frac{\tilde{J}_{LC}^*(q)}{J_{LC,n}^*(q)} = 1 \tag{30}$$

where

$$\tilde{J}_{LC}^*(q) = \mu \sum_{j=1}^{m} \frac{\varsigma_j^2}{3\gamma_j} |q_j|^3.$$

This theorem suggests that when the position is extremely large, the market-risk contribution strictly dominates that of the transaction cost, and the best strategy is to fully hedge.

4.4. Hedging with Liquidity Bundles

So far we have studied the case of trading only individual assets. Let us now expand the result to allow for the trading of liquid bundles. For tractability reasons, we restrict our attention to the model of zero-cost constrained trading and consider trading only one liquidity bundle, such as an ETF. Without loss of generality, suppose this ETF covers the first m assets. Then the liquidation matrix is given by

$$Y = \begin{bmatrix} 1 & 0 & ... & 0 & \alpha_1 \\ 0 & 1 & ... & 0 & \alpha_2 \\ ... & ... & ... & ... & ... \\ 0 & ... & 1 & ... & \alpha_m \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & \gamma_n & 0 \end{bmatrix} \tag{31}$$

In addition, we assume that

$$|\alpha_i \gamma_{\text{ETF}}| < \gamma_i, \forall 1 \leq i \leq m.$$

This assumption suggests that the liquidity of asset i from trading the ETF ($|\alpha_i \gamma_{\text{ETF}}|$) should be less than the liquidity from trading asset i itself (γ_i). The assumption is generally true in practice and enables us to bypass technical difficulties.

In a further attempt to keep things simple, we consider liquidating the position of a single asset.

Theorem 15 (ETF). *If the* large-universe *property is satisfied, then, asymptotically, the cost for liquidating q_j shares of asset $j \leq m$ is given by*

$$\lim_{n\to\infty} J^*_{\text{ETF},n}(q) = \frac{\varsigma_j^2}{3\left(|\alpha_1|\gamma_{\text{ETF}} + \gamma_j\right)} q_j^3, \tag{32}$$

where $q_i = 0, \forall i \neq j$.

We can see that (32) is very similar to the one-dimensional case of (23). First of all, only idiosyncratic risk matters in the large-universe context. Second, the structure of the optimal liquidity cost is the same except for different denominators. In particular, $|\alpha_1|\gamma_{\text{ETF}}$ can be viewed as the liquidity from trading the ETF, while γ_j is the liquidity from trading asset j itself. This theorem provides the intuition that when asset space is large enough, adding an ETF is equivalent to directly increasing the liquidity of the asset.

5. Empirical Results

So far, we have built the framework for estimating liquidity costs for portfolios, and discussed the implications for the drivers of liquidity costs. However, many of our theoretical results rely heavily on assumptions about the structure of price covariance matrix, liquidity level of assets, and so on. For example, one wonders whether the conditions of a large universe are necessarily easy to satisfy in the real world. Also, it would be interesting to illustrate some of our main findings with concrete examples. In order to demonstrate these questions, in the remainder of this paper we will calibrate our model with a small subset of U.S. equities (29 stocks in the utility sector). Specifically, we will fit the factor model using historical returns.

5.1. Overview of the Data Set

As candidates for our calibration, we restrict our attention to the 29 stocks in the Utilities Select Sector Index, which is one of the eleven Select Sector Indices in S&P 500 that track major economic segments and are highly liquid. All the stocks included are from the following industries: electric utilities, water utilities, multi-utilities, independent power producers and energy traders, and gas utilities. We also take into account the Utilities Select Sector SPDR Fund (or XLU), which is an ETF seeking to track the performance of the Utilities Select Sector Index. As a result, the universe of instruments is comprised of 29 individual stocks and one ETF. The market parameters, including prices, daily returns,[3] and daily volume, are obtained from Yahoo Finance for all trading days from January 1, 2012, to April 1, 2016.

Summary statistics of XLU and its asset holdings (as of April 1, 2016) are given in Table 1. As we can see, the weights of assets are somewhat close. This is a result by construction. Quarterly rebalancing ensures that no stock is allowed to have a weight greater than 25%, and that the sum of the stocks with weight greater than 4.8% cannot exceed 50% of the total index weight. All the individual assets are actively traded, with a daily volume ranging from 7.3×10^5 shares to 6.8×10^6 shares. In particular, the ETF (XLU) is highly liquid with an average daily volume of 1.6×10^7 shares, which is much larger than any individual stocks. The average daily volatility of return for XLU is much smaller than its underlying assets, which was to be expected from the diversification it brought. For each asset, the volume traded through the ETF is quite significant and accounts for a sizable portion of the total daily volume ranging from 7.82% for NRG to 31.49% for NEE. We also observe large correlations between individual stocks.

5.2. Model Calibration

The main parameters involved in our model are the liquidation matrix Y, covariance matrix of asset prices Σ, factor loading matrix L, and liquidity parameters γ_j.

Liquidation matrix and liquidity constraints. The liquidation matrix in the example takes the same form as in (31). The key is to determine parameters $\{\alpha_i\}$, where α_i represents the number of shares of asset i contained in one share of the ETF (XLU). More specifically, they are given by the following the formula:

$$\alpha_i = \frac{S_{\text{ETF}} w_i}{S_i},$$

where S_{ETF} is the price of the ETF (XLU), S_i is the price of asset i, and w_i is the dollar weight of asset i in the ETF. XLU is subjected to quarterly rebalancing after the close of business on the second to last calculation day of March, June, September, and December. As a result, the weights for each stock can be modified accordingly. But in our analysis, since the liquidation process takes place only in a short time period, we may safely assume that the structure of the ETF does not

[3] Prices and returns are adjusted for dividends.

Name	Identifier	Weight (%)	Price ($)	Average Daily Volume (Shares, $\times 10^6$)	Total Risk (Daily, %)	Volume Trade Through ETF (%)
Utilities Select Sector SPDR Fund	XLU	-	49.81	16.08	0.85	-
Ameren Corporation	AEE	1.93	50.54	1.68	1.33	19.93
American Electric Power Company Inc.	AEP	5.24	67.01	2.86	1.32	23.90
AES Corporation	AES	1.26	11.57	6.22	2.04	15.26
American Water Works Company Inc.	AWK	2.15	69.50	3.30	1.51	8.19
CMS Energy Corporation	CMS	1.89	42.71	2.50	1.31	15.46
CenterPoint Energy Inc.	CNP	1.55	21.20	4.62	1.59	13.89
Dominion Resources Inc.	D	7.04	75.39	2.56	1.26	31.86
DTE Energy Company	DTE	2.64	91.05	1.01	1.20	25.00
Duke Energy Corporation	DUK	8.25	81.13	3.43	1.32	25.91
Consolidated Edison Inc.	ED	3.60	76.99	1.82	1.36	22.41
Edison International	EIX	3.80	71.94	2.10	1.44	22.01
Eversource Energy	ES	2.96	58.80	1.77	1.30	24.87
Entergy Corporation	ETR	2.25	79.97	1.26	1.39	19.48
Exelon Corporation	EXC	5.38	35.66	6.80	1.63	19.38
FirstEnergy Corp.	FE	2.46	36.03	3.71	1.62	16.09
AGL Resources Inc.	GAS	1.33	65.51	0.73	1.41	24.51
NextEra Energy Inc.	NEE	8.98	118.71	2.10	1.31	31.49
NiSource Inc.	NI	1.22	23.82	2.24	1.46	20.06
NRG Energy Inc.	NRG	0.77	12.77	6.77	3.31	7.82
PG&E Corporation	PCG	4.77	59.83	2.57	1.39	27.08
Public Service Enterprise Group Incorporated	PEG	3.89	47.32	3.07	1.38	23.41
Pinnacle West Capital Corporation	PNW	1.37	75.49	0.83	1.31	19.12
PPL Corporation	PPL	4.16	38.19	3.93	1.31	24.25
SCANA Corporation	SCG	1.60	71.10	0.99	1.34	19.85
Southern Company	SO	7.58	51.70	4.53	1.36	28.30
Sempra Energy	SRE	4.05	105.92	1.27	1.41	26.29
TECO Energy Inc.	TE	1.12	27.56	2.46	1.56	14.47
WEC Energy Group Inc	WEC	3.02	59.97	1.88	1.34	23.40
Xcel Energy Inc.	XEL	3.39	41.94	3.09	1.32	22.86

Table 1: Descriptive statistics for the equity holdings of the assets under discussion. The weights and prices are as of 04/01/2016. The average daily volume is calculated through the period 01/01/2012 – 04/01/2016. The volatility is defined as the standard deviation of percentage daily returns. The volume trade through ETF is calculated as $|\gamma_{XLU}\alpha_j|/\gamma_j$.

change over time; in other words, $\{\alpha_i\}$ is fixed. It is worth noticing that the weights of all the individual stocks do not sum up to 100%. The reason is that the ETFs often put away a small percentage of money in cash. In our analysis, we will neglect those terms as they pose no risk whatsoever.

In our model of zero-cost constrained trading, the liquidity constraint is defined as the maximum rate one can trade without incurring any transaction cost. In general, there is no good way to estimate the threshold without proprietary trading data. For simplicity, we set the threshold at 10% of market trading rate. As we can see from Theorem 10, the exact level of liquidity constraints does not affect the properties of the solutions. Another issue that could complicate the analysis is that the market trading rate can be changing over time. Typically, more trading activities are expected to happen around open and close, and fewer are expected at noon. For tractability, we will just assume that the liquidity constraint is fixed during the liquidation period we are looking at. For example, when AEE is trading at an average daily volume of 1.68×10^6 shares per day, then we set $\gamma_{AEE} = 1.68 \times 10^5$ shares per day.

Covariance structure. We fit a single-factor model (since all the stocks are in the same sector) using historical daily returns from January 1, 2012 to April 1, 2016. In our analysis, it is achieved using the principal component method. An in-depth discussion of this method is given in Chapter

9.4 in Tsay (2005).

5.3. Results

We consider liquidating q shares of one individual stock within the asset universe. Now for each specific stock j, we can consider the following four trading strategies:

1. *No Hedging*: trade stock j only.
2. *Hedging with ETF*: trade stock j and hedge with the ETF (XLU).
3. *Hedging with Basket*: trade stock j and hedge with all other individual stocks.
4. *Hedging with All Assets*: trade all assets including the ETF.

Without loss of generality, for each stock j we consider the liquidation of a position by 10% of its daily volume.

Trading Strategy	Theoretical Liquidity Cost in Closed Form
No Hedging	$\frac{\sigma_j^2}{3\gamma_j}q^3$
Hedging with ETF	$\frac{1}{3}\frac{q^3}{\gamma_j}\sigma_j^2\left(1-\frac{\rho^2}{1+\lvert\rho\rvert\frac{\sigma_j\gamma_j}{\sigma_{\text{ETF}}\gamma_{\text{ETF}}}}\right)$
Hedging with Basket	$\frac{\varsigma_j^2}{3\gamma_j}q^3$
Hedging with All Assets	$\frac{\varsigma_j^2}{3(\lvert\alpha_j\rvert\gamma_{\text{ETF}}+\gamma_j)}q^3$

Table 2: Theoretical results for the four trading strategies.

Table 2 provides the theoretical liquidity costs associated with the four strategies, where the results for the last two strategies are obtained as in the *large-universe* asymptotic limit. If we compare the strategy of *no hedging* with that of *hedging with basket*, we can see that the former is proportional to the total variance σ_j^2, whereas the latter is proportional to the idiosyncratic variance ς_j^2.

Table 3 shows the numerical results of applying the estimated market parameters. For strategy 3 and strategy 4, we provide two sets of numerical results: the one we call *exact* is calculated by solving the discretized version of the optimization problem as in (11); the other one we call *approximated* is calculated using the closed-form equations in the *large-universe* limit as shown in Table 2. From the scaling property in Theorem 4, it is expected that the size of the position we are liquidating should not affect the comparisons between different trading strategies. To better illustrate the results, we normalize the results by setting the *approximated* liquidity cost of strategy 3 (in *large-universe* asymptotic limit) as a benchmark.

Identifier	No Hedging	Hedging with ETF Only	Hedging with Basket		Hedging with All Assets	
			Exact	Approximate	Exact	Approximate
AEE	3.90	1.24	1.06	1.00	0.88	0.83
AEP	4.31	1.52	1.16	1.00	0.92	0.81
AES	1.68	1.15	1.01	1.00	0.87	0.87
AWK	2.04	1.20	1.05	1.00	0.96	0.92
CMS	5.09	1.35	1.11	1.00	0.95	0.87
CNP	2.42	1.22	1.04	1.00	0.90	0.88
D	3.29	1.28	1.10	1.00	0.82	0.76
DTE	5.37	1.31	1.10	1.00	0.87	0.80
DUK	3.61	1.44	1.16	1.00	0.91	0.79
ED	3.21	1.14	1.07	1.00	0.86	0.82
EIX	2.89	1.22	1.07	1.00	0.87	0.82
ES	4.36	1.29	1.09	1.00	0.86	0.80
ETR	2.82	1.16	1.05	1.00	0.87	0.84
EXC	2.10	1.25	1.07	1.00	0.89	0.84
FE	2.12	1.14	1.04	1.00	0.89	0.86
GAS	1.25	1.03	1.00	1.00	0.80	0.80
NEE	3.55	1.44	1.15	1.00	0.87	0.76
NI	2.49	1.12	1.02	1.00	0.85	0.83
NRG	1.35	1.13	1.01	1.00	0.93	0.93
PCG	2.68	1.18	1.06	1.00	0.82	0.79
PEG	3.55	1.34	1.10	1.00	0.88	0.81
PNW	4.33	1.18	1.06	1.00	0.88	0.84
PPL	3.22	1.28	1.08	1.00	0.86	0.80
SCG	4.97	1.18	1.07	1.00	0.89	0.83
SO	3.39	1.26	1.11	1.00	0.86	0.78
SRE	2.73	1.22	1.05	1.00	0.83	0.79
TE	1.67	1.08	1.01	1.00	0.88	0.87
WEC	4.41	1.27	1.10	1.00	0.88	0.81
XEL	5.20	1.40	1.13	1.00	0.91	0.81

Table 3: Numerical results for the utility-sector example.

First of all, the benefit from hedging is quite substantial. By hedging with the ETF alone, we see a significant decrease in all assets, with the ratio of reduction ranging from 16% for NRG to 76% for SCG. This can be explained by the huge liquidity of XLU and the high correlation between XLU and the individual assets.

Secondly, hedging with a basket of individual assets is even better than hedging with the ETF alone for all the assets, though the size of the benefit varies among assets. For AEP, trading with individual assets further reduces the liquidity cost by about 38% from that of hedging with ETF only; by contrast, the number is only 3% for GAS.

Thirdly, for strategy *hedging with basket*, we can see that the *approximate* values are very close to the *exact* ones obtained from solving the dynamic control problem. This shows that the conditions for the *large-universe* regime should be satisfied here and (23) is indeed a good approximation of the actual liquidity cost.

Finally, we see that the benefit of adding ETF to the hedging basket is sizable. In most cases, the reduction of liquidity cost is close to that predicted by (32). This shows that in the *large-universe* regime, trading ETF is equivalent to providing additional liquidity, since the portfolio's market risk exposure has been almost perfectly hedged by the basket of individual stocks.

Finally, we consider the liquidation of a certain position in a representative stock: AEE. To do so, we add other stocks one by one into the stock basket in alphabetic order. Figure 1 shows how the liquidity cost changes as more and more individual stocks are allowed to be used for hedging. The convergence of the liquidity cost to the *large-universe* asymptotic limit is very fast. Figure 2 further shows the evolution of $\underline{\gamma}_n\sqrt{\lambda_{\min}^{(n)}}|J_n^*(q) - \tilde{J}^*(q)|$, as defined in Theorem 11. As expected, the quantity gradually converges to some constant, which shows that the convergence rate of the liquidity cost is roughly converging at the rate of $1/(\underline{\gamma}_n\sqrt{\lambda_{\min}^{(n)}})$.

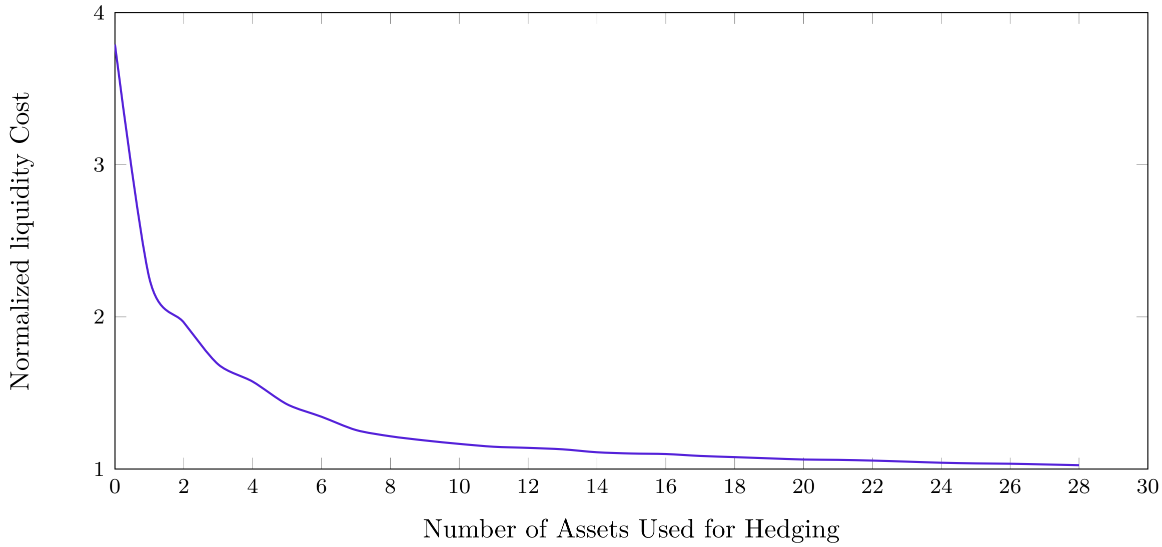

Figure 1: Liquidity cost as the number of assets for hedging increases.

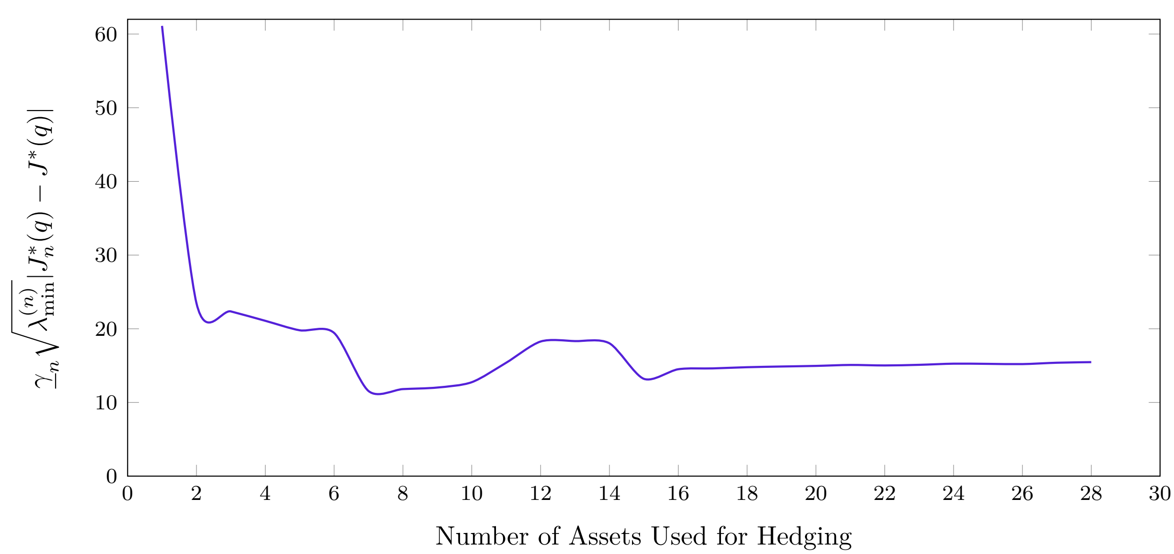

Figure 2: Convergence of the liquidity costs.

6. Concluding Remarks

Accurately estimating liquidity cost is of central importance in portfolio management, and is especially crucial when portfolio managers need to unwind large positions. Additionally, liquidity risk premia can be used to penalize illiquid assets in portfolio construction. We provide a framework to address the multi-asset optimal execution problem, which is far from being a simple extension to the single-asset approach currently adopted in practice.

Our results suggest that managing execution at the portfolio level can substantially reduce liquidity cost by taking advantage of the inter-correlation of asset prices. The complex interaction between asset prices can have a substantial impact on the aggregate portfolio execution cost and risk. We find that traders can improve execution efficiency by hedging the market risk by trading correlated assets simultaneously. This advanced strategy is also true even for the execution of single assets.

An even more compelling takeaway is that in the *large-universe* setting where the covariance structure of asset prices can be explained by only a handful of factors, the liquidity cost is almost purely driven by idiosyncratic risks. This implies that portfolio managers need to pay more attention to an asset's idiosyncratic risk as it not only impacts the risk of security return but also plays a key role in the liquidation process. Additionally, we are able to provide a good closed-form approximation of the liquidity costs in non-asymptotic situations. This can potentially save one the trouble of solving a large-scale dynamic control problem.

Finally, our results signify the importance of trading liquid bundles such as ETFs in optimal liquidation. While previous works are mainly focusing on the hedging benefits of trading liquid bundles, we are the first to recognize its contribution in terms of liquidity provision. In fact, the contribution of liquidity provision is often larger than that of hedging risks. In the *large-universe* context, we manage to show that trading liquid bundles is almost equivalent to providing an additional source of liquidity to the underlying asset.

References

R. Almgren. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10(1):1–18, 2003.

R. Almgren. Optimal trading with stochastic liquidity and volatility. *SIAM Journal on Financial Mathematics*, 3(1):163–181, 2012.

R. Almgren and N. Chriss. Value under liquidation. *Journal of Risk*, 12(12):61–63, 1999.

R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2001.

R. Almgren and J. Lorenz. Adaptive arrival price. *Trading*, 2007(1):59–66, 2007.

R. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *Risk*, 18(5752):10, 2005.

M. Benzaquen, I. Mastromatteo, Z. Eisler, and J.-P. Bouchaud. Dissecting cross-impact on stock markets: An empirical analysis. Working paper, 2016.

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

J. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. Working paper, 2008.

D. B. Brown, B. I. Carlin, and M. S. Lobo. Optimal portfolio liquidation with distress risk. *Management Science*, 56(11):1997–2014, 2010.

G. C. Chacko, J. W. Jurek, and E. Stafford. The price of immediacy. *The Journal of Finance*, 63 (3):1253–1290, 2008.

G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304, 1983.

C. Dellacherie and P. Meyer. *Probabilities and Potential, C: Potential Theory for Discrete and Continuous Semigroups*. Elsevier, 2011.

R. Engle and R. Ferstenberg. Execution risk. *Journal of Portfolio Management*, 33(2):34–45, 2007.

J. Fan, Y. Liao, and M. Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4): 603–680, 2013.

P. A. Forsyth. A Hamilton–Jacobi–Bellman approach to optimal trade execution. *Applied Numerical mathematics*, 61(2):241–265, 2011.

N. Gârleanu and L. H. Pedersen. Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, 68(6):2309–2340, 2013.

R. C. Grinold and R. Kahn. *Active Portfolio Management: a Quantitative Approach for Producing Superior Returns and Controlling Risk*. McGraw-Hill: New York, 2000.

O. Guéant. Optimal execution and block trade pricing: A general framework. *Applied Mathematical Finance*, 22(4):336–365, 2015.

O. Guéant, J. Lasry, and J. Pu. A convex duality method for optimal liquidation with participation constraints. *Market Microstructure and Liquidity*, 1(01):1550002, 2015.

H. Halkin. Necessary conditions for optimal control problems with infinite horizons. *Econometrica*, 42(2):267–272, 1974.

H. He and H. Mamaysky. Dynamic trading policies with price impact. *Journal of Economic Dynamics and Control*, 29(5):891–930, 2005.

C. D. Johnson and J. Gibson. Singular solutions in problems of optimal control. *IEEE Transactions on Automatic Control*, 8(1):4–15, 1963.

H. Kim. *Optimal execution under liquidity constraints.* PhD thesis, Courant Institute of Mathematical Sciences New York, 2014.

Albert S Kyle and Anna A Obizhaeva. Market microstructure invariance: Empirical hypotheses. *Econometrica*, 84(4):1345–1404, 2016.

J. Lorenz and R. Almgren. Mean–variance optimal adaptive execution. *Applied Mathematical Finance*, 18(5):395–422, 2011.

S. Min, C. Maglaras, and C. C. Moallemi. Cross-sectional variation of intraday liquidity, cross-impact, and their effect on portfolio execution. Working paper, 11 2018.

S. A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3): 341–360, 1976.

A. Schied, T. Schöneborn, and M. Tehranchi. Optimal basket liquidation for CARA investors is deterministic. *Applied Mathematical Finance*, 17(6):471–489, 2010.

M. Schneider and F. Lillo. Cross-impact and no dynamic arbitrage. Working paper, 2017.

T. Schöneborn. Adaptive basket liquidation. *preprint*, 2011.

R. S. Tsay. *Analysis of Financial Time Series.* John Wiley & Sons, 2005.

G. Tsoukalas, J. Wang, and K. Giesecke. Dynamic portfolio execution. Working paper, 2014.

Vincent Van Kervel and Albert J Menkveld. High-frequency trading around large institutional orders. Working paper, 2015.

Online Supplement

A. Proofs for Section 2

Theorem 1 (Existence and Convexity). *The dynamic control problem defined in* (8) *is bounded and an optimal solution* u^* *always exists. In addition, the optimal value (the liquidity cost) is convex in initial position.*

Proof. Given q and u, x is then given by

$$x(t) = q - \int_0^t Yu(s)\,ds. \tag{33}$$

By substituting (33) in (8), we obtain the reduced optimal control problem

$$\begin{aligned} J_{red}(u) \triangleq \underset{u}{\text{minimize}} \quad & \int_0^\infty f(u(t))\,dt + \mu \int_0^\infty (q - \int_0^t Yu(s)ds)^\top \Sigma (q - \int_0^t Yu(s)ds)\,dt \\ \text{subject to} \quad & |u_j(t)| \le \gamma_j, \qquad \forall j. \end{aligned} \tag{34}$$

It is easy to see that the functional J_{red} is convex in u. Now define

$$J^* \triangleq \inf_u J_{red}(u).$$

Notice that J^* is well defined since $J_{red}(u)$ is lower bounded by 0. Then there exists a sequence of feasible controls $S = \{u^{(i)} | i = 1, 2, ...\}$ such that

$$J_{red}(u^{(i)}) \to J^*. \tag{35}$$

Without loss of generality, we assume that $J_{red}(u^{(i)}) < \infty$. Combined with the fact that $u^{(i)}$ is bounded, this suggests that $x^{(i)}$ must be bounded.

Next we will prove that $S \subset L^{\mathbf{1}}([0,\infty); \mathbb{R}^m)$ is equi-integrable. Since $u_j(t)$ are bounded, $||u(t)||_1$ must be bounded, and we can simply take constant C such that

$$||u(t)||_1 \le C, \quad \forall u \in L^{\mathbf{1}}([0,\infty); \mathbb{R}^m), t \ge 0.$$

Now consider any measurable set $A \subset [0,\infty)$ such that

$$\int_A ||u(s)||_1\,ds \le C\mu(A), \forall\ u \in S.$$

The equi-integrability then follows trivially. Next, by the Dunford–Pettis theorem (Chapter II, Theorem T25 in Dellacherie and Meyer (2011)), $S \subset L^{\mathbf{1}}([0,\infty)$ is relatively compact for the weak topology. Then there exists a weakly convergent subsequence $\{\hat{u}^{(i)}\}$ of $\{u^{(i)}\}$ that converges to some $u^* \in S$ such that

$$\hat{u}^{(i)} \xrightarrow{w} u^*, \quad \hat{u}^{(i)} \in S. \tag{36}$$

Due to its convexity, the reduced functional J_{red} is lower semicontinuous with respect to the weak topology and hence

$$\liminf_{i\to\infty} J_{red}(\hat{u}^{(i)}) \leq J_{red}(u^*),$$

which allows us to conclude that u^* is a minimizer.

■

Theorem 2 (Uniqueness). *All optimal solutions for the the dynamic control problem in* (8) *have a unique optimal position trajectory* $x* \in C([0,\infty);\mathbb{R}^n)$. *Moreover, if the transaction cost functional* $f(\cdot)$ *is strictly convex, the optimal trading strategy* $u^* \in L^1([0,\infty;\mathbb{R}^m)$ *must also be unique.*

Proof. Denote $J(u,x)$ to be the liquidity cost associated with feasible trading rate u and position process x.

Suppose the optimal solution is not path-unique; then there exist $(u_1,x_1),(u_2,x_2)$ that are both optimal and such that

$$J(u_1,x_1) = J(u_2,x_2) = J^*, \quad x_1 \neq x_2.$$

Now consider

$$u_3 = \frac{u_1+u_2}{2}, \quad x_3 = \frac{x_1+x_2}{2}.$$

It is easy to see that (u_3,x_3) is also feasible for (8).

Then we have

$$\begin{aligned} J(u_3,x_3) &= \int_0^\infty f(u_3(t))dt + \mu\int_0^\infty x_3(t)^\top \Sigma x_3(t)dt \\ &= \int_0^\infty f(\frac{u_1(t)+u_2(t)}{2})dt + \mu\int_0^\infty (\frac{x_1(t)+x_2(t)}{2})^\top \Sigma (\frac{x_1(t)+x_2(t)}{2})dt \\ &< \int_0^\infty (f(u_1(t))+f(u_2(t)))\,dt/2 + \mu\int_0^\infty \Big(x_1(t)^\top\Sigma x_1(t) + x_2(t)^\top \Sigma x_2(t)\Big)\,dt/2 \\ &= J(u_1,x_1)/2 + J(u_2,x_2)/2 = J^*. \end{aligned} \tag{37}$$

The strict inequality is provided by the strict convexity of $x^\top\Sigma x$. Notice that this contradicts the optimality of $(u_1,x_1),(u_2,x_2)$.

If $f(\cdot)$ is strictly convex, we will also require $u_1 = u_2$ such that

$$\int_0^\infty f(\frac{u_1(t)+u_2(t)}{2})dt = \int_0^\infty (f(u_1(t))+f(u_2(t)))\,dt/2.$$

This provides the uniqueness of the optimal solution.

■

Theorem 3 (Sufficiency). *A feasible pair* $(x^*, u^*) \in C([0,\infty);\mathbb{R}^n) \times L^1([0,\infty);\mathbb{R}^m)$ *form an optimal solution of* (8) *if, for all* $t \geq 0$,

$$x^*(t) = q - \int_0^t Y u^*(s)\, ds,$$

$$u^*(t) \in \underset{u:\ -\gamma \leq u \leq \gamma}{\operatorname{argmin}}\ f(u) - 2\int_t^\infty x^*(s)^\top \Sigma Y u\, ds. \tag{9}$$

Proof. Define $J_{red}(u)$ to be the optimal liquidity cost associated with trading strategy u. Now suppose u^* satisfies (9) but is not optimal. Then there must exist some feasible trading strategy $\tilde{u} \neq u^*$ such that

$$J_{red}(\tilde{u}) < J_{red}(u^*).$$

By algebra, we also have

$$\begin{aligned} J_{red}(\tilde{u}) = J_{red}(u^*) &+ \int_0^\infty \int_0^t (\tilde{u}(s) - u^*(s))^\top Y^\top \Sigma Y (\tilde{u}(s) - u^*(s))\, dsdt \\ &- 2\int_0^\infty \int_0^t (x^*(t))^\top \Sigma Y (\tilde{u}(s) - u^*(s))\, dsdt + \int_0^\infty (f(\tilde{u}(t)) - f(u^*(t)))\, dt \\ &\geq J_{red}(u^*) + \int_0^\infty \left(f(\tilde{u}(t)) - f(u^*(t)) - 2\int_s^\infty (x^*(s))^\top \Sigma Y (\tilde{u}(s) - u^*(s))\, ds \right) dt. \end{aligned} \tag{38}$$

Since u^* satisfies (9), it is easy to see that the second term is always positive. Hence we reach a contradiction. ■

B. Proofs for Section 3

Theorem 4 (Scaling). *If* u^* *is optimal for the problem* (11) *starting from initial position* $x(0) = q$, *then*

$$\tilde{u}(t) \triangleq u^*(t/\alpha), \quad \forall\, t \geq 0,$$

is optimal for the problem starting with initial position $x(0) = \alpha q$ *for all* $\alpha > 0$. *Furthermore,*

$$J^*(\alpha q) = \alpha^3 J^*(q).$$

Proof. Under trading strategy $\tilde{u}$,

$$\tilde{x}(t) = \alpha q - \int_0^t Y\tilde{u}(s)ds = \alpha q - \alpha\int_0^{t/\alpha} Yu(s)ds = \alpha x(t/\alpha).$$

Suppose that $\tilde{x}$ is not the optimal path for the problem starting with αq; then there exists a better path y such that

$$\int_0^\infty (y(t))^\top \Sigma y(t)dt < \int_0^\infty (\tilde{x}(t))^\top \Sigma \tilde{x}(t)dt.$$

Then we have

$$\int_0^\infty (y(\alpha t))^\top \Sigma y(\alpha t)dt < \int_0^\infty (x(t))^\top \Sigma x(t)dt.$$

It is easy to see that $y_\alpha(t) = y(\alpha t)$ is feasible for the problem starting with q, a contradiction:

$$J^*(\alpha q) = \int_0^\infty (\tilde{x}(t))^\top \Sigma \tilde{x}(t) dt = \alpha^3 \int_0^\infty (x(t))^\top \Sigma x(t) dt = \alpha^3 J^*(q).$$

■

Theorem 5 (Finite Horizon). *For any initial position q, the optimal position trajectory $x(t)$ is guaranteed to reach zero in finite time.*

Proof. First of all, define the norm $||\cdot||_\Sigma$ as

$$||q||_\Sigma \triangleq \sqrt{q^\top \Sigma q}, \quad \forall q \in \mathbb{R}^n.$$

By Theorem 4, we know that it suffices to prove the theorem for $\forall q \in \mathbb{R}^n$ with

$$||q||_\Sigma = 1.$$

Now, for any q, let $x^*(t)$ be the position associated with its optimal execution strategy at time t. Clearly, we have $x^*(0) = q$. Now define $T(q)$ to be the first time that the norm of the position is less than 1/2:

$$T(q) \triangleq \inf\{t : ||x^*(t)||_\Sigma \leq \frac{1}{2}||q||_\Sigma\}.$$

Given that Y is full rank, we know that the set $\mathcal{A} = \{Yu | u \in \mathbb{R}^m, |u_i| \leq \gamma_i\}$ is an n-dimensional polytope in R^n. Notice that $0 \in \mathcal{A}$. Therefore, there must exist $\epsilon > 0$ such that $\{q \in \mathbb{R}^n | ||q||_\Sigma \leq \epsilon\} \subset \mathcal{A}$.

Now consider any q such that $||q||_\Sigma = 1$, and let $x^*(t)$ be its position at time t in the optimal execution strategy. Define

$$\tau_i = \inf\{t \geq 0 : ||x^*(t)||_\Sigma \leq \frac{1}{2^i}\}.$$

Lemma 2 shows that $\tau_1 - \tau_0 \leq T^*$. By applying Theorem 4, we know that

$$T(q) \leq T^*/2, \quad \forall ||q|| \leq 1/2.$$

It follows that

$$\tau_{i+1} - \tau_i \leq \frac{T^*}{2^i}.$$

Then

$$\begin{aligned}
\lim_{i\to\infty} \tau_i &= \lim_{i\to\infty} \left[\tau_0 + \sum_{j=1}^{i} (\tau_j - \tau_{j-1})\right] \\
&\leq 0 + \lim_{i\to\infty} \sum_{j=1}^{i} \frac{1}{2^{j-1}} T^* \\
&= 2T^*.
\end{aligned}$$

Then, $\{\tau_i\}$ is increasing and bounded from above and so there exists a limit τ^* such that

$$\tau^* = \lim_{i\to\infty} \tau_i \leq 2T^*.$$

By continuity of $x^*(t)$, we have $x^*(\tau^*) = 0$. Then, it must be that

$$x^*(t) = 0, \quad t \geq \tau^*.$$

otherwise the liquidity cost could be reduced. ■

Lemma 2.

$$T^* \triangleq \sup\{T(q) : ||q||_\Sigma = 1\} < \infty.$$

Proof. $\forall q$ such that $||q||_\Sigma = 1$, consider a trading strategy where we set

$$u(t) = \begin{cases} u^*, & t \leq \frac{1}{\epsilon} \\ 0, & t > \frac{1}{\epsilon} \end{cases},$$

where $Yu^* = \epsilon q$. Notice that this strategy is clearly feasible.

The liquidity cost of this strategy $J(q)$ is easily given by

$$J(q) = \int_0^{\frac{1}{\epsilon}} ||q - t\epsilon q||_\Sigma dt < \frac{1}{\epsilon}.$$

Now, in the optimal trading strategy, we have

$$\frac{1}{2}T(q) \leq \int_0^{T(q)} ||x^*(t)||_\Sigma dt < \int_0^\infty ||x^*(t)||_\Sigma dt \leq J(q) < \frac{1}{\epsilon},$$

which leads to

$$T(q) < \frac{2}{\epsilon}.$$

Since the upper bound does not depend on q, we have

$$T^* < \frac{2}{\epsilon} < \infty.$$

■

Lemma 1 (Optimality). *A feasible control u^* is optimal for* (11) *if and only if, for all $t \geq 0$,*

$$x^*(t) = q - \int_0^t Yu^*(s)\, ds,$$

$$u^*(t) \in \operatorname*{argmax}_{u:-\gamma\leq u\leq\gamma} \left(\int_t^\infty x^*(s)^\top \Sigma Y\, ds\right) u. \tag{12}$$

Proof. The sufficiency is given by Theorem 3. By Theorem 5, given the initial position q, there exists some T such that the optimal execution ends before time T. As a result, (11) is equivalent to the following:

$$\begin{aligned}
J^*(q) \triangleq \underset{u}{\text{minimize}} \quad & \int_0^T x^\top(t)\Sigma x(t)\, dt \\
\text{subject to} \quad & \dot{x}(t) = -Yu(t), && \forall\ t \geq 0, \\
& |u_i(t)| \leq \gamma_i, && \forall\ 1 \leq i \leq m,\ t \geq 0, \\
& x(0) = q, \\
& u \in L^1([0,\infty);\mathbb{R}^m).
\end{aligned} \tag{39}$$

We can obtain the necessity through Pontrjagin's minimum principle. The convexity assumption and regularity assumption are satisfied trivially in this case due to the linear control. The Hamiltonian function of (39) can be written as

$$H(x,u,p) = x^\top \Sigma x - p^\top Y u. \tag{40}$$

Suppose that $x^*(t), u^*(t)$ is optimal; then there must exist an optimal adjoint state $p^*(t), t \in [0,\infty)$ that satisfies the following:

$$\dot{p}^*(t) = -\bigtriangledown_x H(x^*(t), u^*(t), p^*(t)) = -2\Sigma x^*(t),$$

$$p^*(T) = 0.$$

We can then solve for $p^*(t)$ as

$$p^*(t) = 2\Sigma Y \int_t^T x^*(s)ds. \tag{41}$$

Moreover, we have

$$u^*(t) \in \underset{u:\ -\gamma \leq u \leq \gamma}{\text{argmin}} \ H(x^*(t), u(t), p^*(t)), \tag{42}$$

which can also be written as

$$u^*(t) \in \underset{u:\ -\gamma \leq u \leq \gamma}{\text{argmin}} \left(-2 \int_t^T (x^*(s))^\top \Sigma Y u ds \right), \forall t \in [0,T].$$

Since we know that $x^*(t) = 0$ for $\forall t \geq T$, the condition above can be written as

$$u^*(t) \in \underset{u:\ -\gamma \leq u \leq \gamma}{\text{argmax}} \left(\int_t^\infty (x^*(s))^\top \Sigma Y u ds \right), \forall t \in [0,\infty).$$

■

Theorem 6 (High Liquidity Hedging). *In the two-dimensional case where model parameters are given* *by*

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_1 & \sigma_2^2 \end{bmatrix}, \qquad Y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

if we further assume that

$$\gamma_2 \geq |\rho| \frac{\sigma_1 \gamma_1}{\sigma_2}, \tag{13}$$

then the optimal liquidity cost of portfolio $q = (q, 0)$ *is given by*

$$J^*(q) = \frac{1}{3}\frac{q^3}{\gamma_1}\sigma_1^2 \left(1 - \frac{\rho^2}{1 + |\rho| \frac{\sigma_1\gamma_1}{\sigma_2\gamma_2}}\right). \tag{14}$$

Proof. Without loss of generality, we assume that $\rho > 0$. In order to simplify the notations and to provide better intuitions, we define the following:

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = Y^\top \Sigma Y,$$

where a, b, c can be easily determined by the model parameters:

$$a = \sigma_1^2, \quad b = \rho\sigma_1\sigma_2, \quad c = \sigma_2^2.$$

Now consider the following trading strategy:

1. For $0 \leq t \leq \frac{b}{b+c}\frac{q}{\gamma_1}$, trade at rate $u^*(t) = (\gamma_1, -\gamma_2)^\top$.
2. For $\frac{b}{b+c}\frac{q}{\gamma_1} \leq t \leq \frac{q}{\gamma_1}$, trade at rate $u^*(t) = (\gamma_1, \frac{b}{c}\gamma_1)^\top$.

The assumption that $\gamma_2 \geq |\rho|\frac{\sigma_1\gamma_1}{\sigma_2}$ guarantees the feasibility of this strategy.

The liquidation path can now be calculated as

$$x^*(t) = \begin{cases} (q - \gamma_1 t, \gamma_2 t)^\top, & 0 \leq t \leq \frac{b}{b+c}\frac{q}{\gamma_1} \\ \left(q - \gamma_1 t, -\frac{b}{c}(q - \gamma_1 t)^\top\right), & \frac{b}{b+c}\frac{q}{\gamma_1} \leq t \leq \frac{q}{\gamma_1} \\ (0, 0)^\top, & t > \frac{q}{\gamma_1} \end{cases}. \tag{43}$$

Then it's easy to justify that (43) satisfies the optimality condition in Lemma 1.

Hence the optimal solution is given by

$$J^*(q) = \int_0^\infty (x^*(t))^\top \Sigma x^*(t) dt = \frac{1}{3}\frac{q^3}{\gamma_1}\sigma_1^2 \left(1 - \frac{\rho^2}{1 + \rho\frac{\sigma_1\gamma_1}{\sigma_2\gamma_2}}\right).$$

■

Theorem 7 (One Asset). *In the one-dimensional case, the cost of liquidating a position of q with* *parameters* (σ, γ, ν) *is given by*

$$J^*(q) = \nu|q| + \mu\frac{|q|^3\sigma^2}{3\gamma}.$$

Proof. In the one-dimensional case, it is easy to see that the transaction cost is only a function of the total position traded and does not depend on the trading rate. As a result, the optimal trading strategy is simply to unload the position as fast as possible.

With this in mind, the agent should trade the asset at a constant rate γ if $q > 0$ and $-\gamma$ if $q \leq 0$. Then the liquidity cost is given by:

$$J^*(q) = \nu|q| + \mu\frac{|q|^3\sigma^2}{3\gamma}.$$

■

Theorem 8 (Two Assets). *Consider the two-dimensional case where model parameters are given by*

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_1 & \sigma_2^2 \end{bmatrix}, \qquad Y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

and assume the portfolio to be liquidated contains q shares of asset 1 and no position in asset 2. If we further assume that

$$\gamma_2 \geq |\rho|\frac{\sigma_1\gamma_1}{\sigma_2},$$

then the asset 2 will only be used to hedge if and only if

$$q^2 \geq \frac{2\gamma_1\nu_2}{\mu\gamma_2\rho\sigma_1\sigma_2}. \tag{17}$$

If (17) *is satisfied, then the optimal liquidity cost of the portfolio is given by*

$$J^*(q) = \frac{\mu}{3}\frac{q^3}{\gamma_1}\sigma_1^2\left(1 - \frac{\rho^2}{1+|\rho|\frac{\sigma_1\gamma_1}{\sigma_2\gamma_2}}\right) + \nu_1 q + 2\nu_2 q\frac{\gamma_2}{\gamma_1}\frac{|\rho|\frac{\sigma_1\gamma_1}{\sigma_2\gamma_2}}{1+|\rho|\frac{\sigma_1\gamma_1}{\sigma_2\gamma_2}} - \frac{4\frac{\nu_2}{\sigma_2}\sqrt{\frac{2\nu_2|\rho|\sigma_1\gamma_1}{\sigma_2}}}{3(1+|\rho|\frac{\sigma_1\gamma_1}{\sigma_2\gamma_2})}. \tag{18}$$

Proof. Without loss of generality, we assume that $\rho > 0, q > 0$, in which case the two assets are positively correlated. Notice that Theorem 6 can be viewed as a special case where the transaction cost parameters ν_1, ν_2 are zero. Accordingly, we first establish a short position in asset 2 in order to hedge the market risk. Since there is no transaction cost, asset 2 is traded at full rate γ_2 until the ratio of position in the two assets reaches $\rho\frac{\sigma_1}{\sigma_2}$, and this ratio is maintained till the end of the liquidation process. Intuitively, by shorting asset 2 we are hedging the market risk but introducing another source of idiosyncratic risk, and the ratio $\rho\frac{\sigma_1}{\sigma_2}$ is the balance point of such a trade-off. Now we have transaction costs for trading asset 2, and it is expected that the agent will trade asset 2 less.

Now consider the following trading strategy:

- If $q^2 < \frac{2\gamma_1\nu_2}{\mu\gamma_2\rho\sigma_1\sigma_2}$:

1. Sell asset 1 as fast as possible until the entire position is unloaded.

- If $q^2 \geq \frac{2\gamma_1\nu_2}{\mu\gamma_2\rho\sigma_1\sigma_2}$:

  1. For $0 \leq t \leq T_1$, trade at rate $u^*(t) = (\gamma_1, -\gamma_2)^\top$.
  2. For $T_1 < t \leq T_2$, trade at rate $u^*(t) = (\gamma_1, 0)^\top$.
  3. For $T_2 < t \leq \frac{q}{\gamma_1}$, trade at rate $u^*(t) = (\gamma_1, \frac{\rho\sigma_1}{\sigma_2}\gamma_1)^\top$,

  where

$$T_1 = \frac{q\rho\sigma_1\gamma_2\sigma_2 - \sqrt{2\nu_2\rho\sigma_1\gamma_1\sigma_2\gamma_2}}{\rho\sigma_1\gamma_1\sigma_2\gamma_2 + \sigma_2^2\gamma_2^2},$$

$$T_2 = \frac{q\gamma_1\rho\sigma_1 + \sqrt{\frac{2\nu_2\gamma_1\gamma_2\sigma_2}{\rho\sigma_1}}}{\gamma_1^2\rho\sigma_1 + \gamma_1\gamma_2\sigma_2}.$$

  The cost induced by this strategy is given by (18). This strategy can be shown to be optimal by checking (9) in Theorem 3. We omit the details here as the algebra is cumbersome.

■

C. Proofs for Section 4

Proposition 1 (Factor Replicating Portfolio). *If the* large-universe *conditions hold, then for each factor $F_i(t)$, there exists a series of portfolios $\{p^{(i,n)}(t)\}$ defined by weights $\{\beta_j^{(i,n)}\}$ where*

$$p^{(i,n)}(t) \triangleq \sum_{j=1}^{n} \beta_j^{(i,n)} S_j(t),$$

such that

1. *The portfolio $p^{(i,n)}(t)$ has unit exposure on factor $F_i(t)$:*

$$p^{(i,n)}(t) - F_i(t) = \epsilon^{(i,n)}(t),$$

   where $\epsilon^{(i,n)}(t)$ is zero mean and independent of all factor-price processes, and has variance upper bounded by

$$Var(\epsilon^{(i,n)}(t)) \leq \frac{\sup_j \varsigma_j^2}{\lambda_{\min}^{(n)}} t.$$

2. *The sum of the squares of the weights converge to 0:*

$$\lim_{n\to\infty} \sum_{j=1}^{n} (\beta_j^{(i,n)})^2 = 0.$$

Proof. For each factor i, we want to find the portfolio that has unit exposure on factor i and has minimum idiosyncratic risks. This can be done by solving the following optimization problem:

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \frac{1}{2} w^\top \Xi^{(n)} w \\ \text{subject to} \quad & (L^{(n)})^\top w = e_i, \\ & w \in \mathbb{R}^n. \end{aligned} \tag{44}$$

where e_i is the ith column of the $K \times K$ identity matrix.

For simplicity, we will assume that the idiosyncratic risk for each individual asset is strictly positive ($\varsigma_j^2 > 0$), and will simply ignore the assets with $\varsigma_j^2 = 0$.

Now denote $z = (\Xi^{(n)})^{1/2} w$, and consider the singular value decomposition of $L^{(n)}$:

$$L^{(n)} = (U^{(n)})^\top \Lambda^{(n)} V^{(n)}. \tag{45}$$

Then (44) is equivalent to

$$\begin{aligned} \min_z \frac{1}{2} \quad & z^\top z \\ s.t. \quad & (V^{(n)})^\top \Lambda^{(n)} U^{(n)} (\Xi^{(n)})^{-1/2} z = e_i, \\ & z \in \mathbb{R}^n. \end{aligned} \tag{46}$$

The Lagrangian of (46) is

$$L = \frac{1}{2} z^\top z - \mu^\top \left((L^{(n)})^\top (\Xi^{(n)})^{-1/2} w - e_i \right).$$

The optimal solution z^* is given by solving

$$\begin{aligned} & z^* - (\Xi^{(n)})^{-1/2} L^{(n)} \mu = 0, \\ & L^{(n)} (\Xi^{(n)})^{-1} L^{(n)} \mu = e_i. \end{aligned} \tag{47}$$

Then we have

$$\begin{aligned} (w^*)^\top \Xi^{(n)} w^* &= (z^*)^\top z^* = \mu^\top L^{(n)} (\Xi^{(n)})^{-1} L^{(n)} \mu \\ &= e_i^\top \left(L^{(n)} (\Xi^{(n)})^{-1} L^{(n)} \right)^{-1} e_i \\ &\leq \lambda_{max} \left(\left(L^{(n)} (\Xi^{(n)})^{-1} L^{(n)} \right)^{-1} \right) \\ &\leq \frac{\sup_{j \leq n} \varsigma_j^2}{\lambda_{\min}^{(n)}}. \end{aligned} \tag{48}$$

We can now set $\beta^{i,n}$ as w^* solved from above. As n goes to infinity, we have

$$\lim_{n\to\infty}\sum_{j=1}^{n}\beta_j^{(i,n)} \le \lim_{n\to\infty}\frac{(w^*)^\top \Xi^{(n)} w^*}{\sup_{j\le n}\varsigma_j^2} \le \lim_{n\to\infty}\frac{1}{\lambda_{\min}^{(n)}} = 0.$$

■

Lemma 3. *Consider two liquidation problems that differ only in their covariance matrices (Σ_1 and Σ_2, respectively). Suppose that their optimal liquidity costs are given by $J_1^*(q)$ and $J_2^*(q)$. If we have*

$$\Sigma_1 \preceq \Sigma_2,$$

where $\preceq$ is the positive definite ordering, then

$$J_1^*(q) \le J_2^*(q), \quad \forall q \in \mathbb{R}^n.$$

Proof. Suppose that $u^{(2)} \in L_1([0,\infty);\mathbb{R}^m)$ is the optimal solution to problem 2, and $x^{(2)} \in C([0,\infty);\mathbb{R}^n)$ is the corresponding position process. Since problems 1 and 2 differ only in their covariance matrices, $(u^{(2)}, x^{(2)})$ is also feasible for problem 1. If we denote $J_q^*(q)$ as the optimal liquidity cost for problem 1 and $J_2^*(q)$ as that for problem 2, then we have

$$\begin{aligned} J_2^*(q) &= \int_0^\infty f(u^{(2)}(t))dt + \mu\int_0^\infty x^{(2)}(t)^\top \Sigma_2 x^{(2)}(t)dt \\ &\ge \int_0^\infty f(u^{(2)}(t))dt + \mu\int_0^\infty x^{(2)}(t)^\top \Sigma_1 x^{(2)}(t)dt \\ &\ge J_1^*(q), \end{aligned} \tag{49}$$

where the first inequality comes from the fact that $\Sigma_1 \preceq \Sigma_2$. ■

Theorem 9 (Lower Bound of Hedging Benefits)**.** *The liquidity cost is lower bounded according to*

$$J_n^*(q) \ge \sum_{j=1}^{m}\frac{\varsigma_j^2}{3\gamma_j}|q_j|^3. \tag{22}$$

Proof. Consider the following problem where we replace the covariance matrix $\Sigma^{(n)}$ with $\Xi^{(n)}$:

$$\begin{aligned} \tilde{J}_n^*(q) = \underset{u}{\text{minimize}} \quad & \int_0^\infty x^\top(t)\Xi^{(n)}x(t)\,dt = \sum_{i=1}^{n}\int_0^\infty \varsigma_i^2 x_i^2(t), dt \\ \text{subject to} \quad & \dot{x}(t) = -Yu(t), \qquad \forall\ t\ge 0, \\ & |u_i(t)| \le \gamma_i, \qquad \forall\ 1\le i\le m,\ t\ge 0, \\ & x(0) = q, \\ & u \in L^1([0,\infty);\mathbb{R}^m). \end{aligned} \tag{50}$$

Since there are no correlations, it is easy to see that the optimal execution strategy in this case is to liquidate each asset separately at full rate. Hence the optimal solution to the above problem is given by

$$\tilde{J}_n^*(q) = \sum_{j=1}^{m} \frac{\varsigma_j^2}{3\gamma_j} |q_j|^3.$$

Notice that $\Xi^{(n)} \preceq \Sigma^{(n)}$; then, by applying Lemma 3 stated above, we always have

$$J_n^*(q) \geq \tilde{J}_n^*(q) = \sum_{j=1}^{m} \frac{\varsigma_j^2}{3\gamma_j} |q_j|^3.$$

■

Theorem 10 (Large Universe). *If the* large-universe *property is satisfied, then, asymptotically, the liquidity cost of any portfolio consisting of finitely many assets will be driven purely by idiosyncratic risks. More specifically, we have*

$$J_\infty^*(q) = \lim_{n\to\infty} J_n^*(q) = \sum_{j=1}^{m} \frac{\varsigma_j^2}{3\gamma_j} |q_j|^3, \tag{23}$$

where q is a portfolio with positions in at most the first m assets, and $J_n^*(q)$ is the optimal cost of liquidating q while trading at most the first $n \geq m$ assets.

Proof. The key to the proof is finding a trading strategy that converges to the lower bound asymptotically. We assume that the chosen factor portfolios don't contain assets in the liquidation portfolio. By Proposition 1, for each factor $F_i(t)$ there exists a sequence of portfolios $\{p^{(i,n)}(t)\}$ characterized by $\{\beta^{(i,n)}\}$ such that $p^{(i,n)}(t) \to F_i(t)$. More specifically, we have $\sum_{j=1}^{n}(\beta_j^{(i,n)})^2 \to 0$.

Now, for each asset j in the portfolio to be liquidated, we construct a sequence of portfolios $\{z^{(j,n)}\}$ characterized by $\{\hat{\beta}^{(j,n)}\}$ such that

$$\hat{\beta}^{(j,n)} = \sum_{i=1}^{K} l_{ji} \beta^{(i,n)}, \tag{51}$$

where l_{ji} is the factor exposure of asset j on factor i:

$$\begin{aligned}
\sum_{k=1}^{n} (\hat{\beta}_k^{(j,n)})^2 &= \sum_{k=1}^{n} \left(\sum_{i=1}^{K} l_{ji} \beta_k^{(i,n)} \right)^2 \\
&\leq \bar{l}^2 K \sum_{i=1}^{K} \sum_{k=1}^{n} (\beta_k^{(i,n)})^2 \\
&\leq \bar{l}^2 K^2 \frac{1}{\lambda_{\min}^{(n)}} \to 0,
\end{aligned}$$

where $\bar{l} = \max_{i\leq m, j\leq K} l_{ij}$.

The exposure of $z^{(j,n)}$ on factor $F_i(t)$ is given by:

$$l_{ji} \sum_{k=1}^{n} \beta_k^{(i,n)} l_{ki} = l_{ji}.$$

Essentially, we have created a sequence of portfolios that has the same factor exposure as that of asset j, but whose idiosyncratic risk converges to 0.

Further, define

$$N(n) = \sqrt{\frac{\lambda_{\min}^{(n)}}{\bar{l}^2 K^2}}. \tag{52}$$

Notice that for each n and $1 \leq j \leq m$, we have

$$|\hat{\beta}_k^{(j,n)}| \leq \sqrt{\sum_{i=1}^{n} (\hat{\beta}_i^{(j,n)})^2} \leq \frac{1}{N(n)}, \quad \forall k \leq n.$$

Intuitively $\frac{1}{N(n)}$ can be viewed as the upper bound of the weight of each asset in every portfolio we constructed. Given that $\sum_{k=1}^{n} (\hat{\beta}_k^{(j,n)})^2 \to 0$, we have $N(n) \to \infty$.

We consider the following "dumb" strategy for the problem indexed by n:

1. For $0 \leq t \leq \frac{|q_j|}{\gamma_j}$, buy asset j at a rate of $-\frac{q_j}{|q_j|}\gamma_j$.
2. For $0 \leq t \leq \frac{\sum_{j=1}^{m} |q_j|}{N(n)\underline{\gamma}_n}$, buy portfolio $z^{(j,n)}$ at a rate of $\frac{-q_j}{\sum_{i=1}^{m} |q_i|} N(n)\underline{\gamma}_n$. Do it for all $1 \leq j \leq m$.
3. For $\frac{\sum_{j=1}^{m} |q_j|}{N(n)\underline{\gamma}_n} \leq t \leq \frac{|q_j|}{\gamma_j}$, buy $z^{(j,n)}$ at a rate of $\frac{q_j}{|q_j|}\gamma_j$. Do it for all $1 \leq j \leq m$.

Let's first try to understand this "dumb" strategy. First of all, we notice that $z^{(j,n)}$ approximates the factor risk exposure of asset j, and so step 2 is the hedging process. Basically, we acquire a certain amount of portfolio $z^{(j,n)}$ in order to hedge the factor risks contributed by asset j. Step 3 is the liquidation process: we sell each asset together with its hedging portfolio as soon as possible.

We still need to check whether this strategy violates the liquidity constraints. Consider a particular asset k whose weight in each portfolio is at most $1/N(n)$. In step 2, its trading rate is upper bounded by

$$\frac{1}{N(n)} \sum_{j=1}^{m} \frac{-q_j}{\sum_{i=1}^{m} |q_i|} N(n)\underline{\gamma}_n \leq \underline{\gamma}_n.$$

Hence the liquidity constraint is satisfied in step 1.

In step 3, the trading rate for asset $j > m$ is upper bounded by

$$\sum_{i=1}^{m} \gamma_i / N(n) < \underline{\gamma}_n,$$

given n is large enough. As a result, given n is large enough, all the liquidity constraints are satisfied.

Following this trading strategy, the risk of the position at time t ($V_n(t)$) is

$$V_n(t) = \begin{cases} \left[\sum_{j=1}^m (q_j - \gamma_j t)(l_j - \frac{t}{t_n} l_{z^{(j,n)}})\right]^\top \left[\sum_{j=1}^m (q_j - \gamma_j t)(l_j - \frac{t}{t_n} L_{z^{(j,n)}})\right] \\ \quad + \sum_{j=1}^m (q_j - \gamma_j t)^2 \varsigma_j^2 + \sum_{j=1}^m (q_j - \gamma_j t_n)^2 \frac{t^2}{t_n^2} \varsigma^2_{z^{(j,n)}}, \qquad 0 \le t \le t_n, \\ \left[\sum_{j=1}^m \frac{(t-T_j)^+}{T_j}(q_j - \gamma_j t)(l_j - l_{z^{(j,n)}})\right]^\top \left[\sum_{j=1}^m \frac{(t-T_j)^+}{T_j}(q_j - \gamma_j t)(l_j - l_{z^{(j,n)}})\right] \\ \quad + \sum_{j=1}^m \left(\frac{(t-T_j)^+}{T_j}(q_j - \gamma_j t)\right)^2 \left(\varsigma_j^2 + \varsigma^2_{z^{(j,n)}}\right), \qquad \textit{otherwise.} \end{cases}$$

where

$$t_n = \frac{\sum_{j=1}^m |q_j|}{N(n)\underline{\gamma}_n}, \quad T_j = \frac{|q_j|}{\gamma_j}, \tag{53}$$

and $l_{z^{(j,n)}} \in \mathbb{R}^K$ is the factor exposure of portfolio $z^{(j,n)}$, and $\varsigma^2_{z^{(j,n)}}$ is its idiosyncratic risk exposure.

The liquidity cost $J_n(q)$ in this case can be written as

$$J_n(q) = \int_0^{t_n} V_n(t)dt + \int_{t_n}^\infty V_n(t)dt. \tag{54}$$

Notice that by construction, we have the following:

$$l_{z^{(j,n)}} = l_j, \quad \lim_{n\to\infty} \varsigma^2_{z^{(j,n)}} = \lim_{n\to\infty} \sum_{i=1}^n \varsigma_i^2 (\hat{\beta}_k^{(j,n)})^2 \le \sup_n \varsigma_n^2 \lim_{n\to\infty} \sum_{i=1}^n (\hat{\beta}_k^{(j,n)})^2 = 0. \tag{55}$$

By using (55), it is easy to show that

$$\begin{aligned} \int_0^{t_n} V_n(t)dt &= \int_0^{t_n} \left(\sum_{j=1}^m (q_j - \gamma_j t)\right)^2 l_j^\top l_j (1 - \frac{t}{t_n})^2 dt \\ &\quad + \int_0^{t_n} \left(\sum_{j=1}^m (q_j - \gamma_j t)^2 \varsigma_j^2 + \sum_{j=1}^m (q_j - \gamma_j t_n)^2 \frac{t^2}{t_n^2} \varsigma^2_{z^{(j,n)}}\right) dt, \\ &\le \frac{1}{3} || \sum_{j=1}^m q_j l_j ||_2^2 t_n + \int_0^{t_n} \left(\sum_{j=1}^m (q_j - \gamma_j t)^2 \varsigma_j^2 + \sum_{j=1}^m (q_j - \gamma_j t_n)^2 \frac{t^2}{t_n^2} \varsigma^2_{z^{(j,n)}}\right) dt, \end{aligned} \tag{56}$$

and that

$$\int_{t_n}^\infty V_n(t)dt = \int_{t_n}^\infty \sum_{j=1}^m \left(\frac{(t-T_j)^+}{T_j}(q_j - \gamma_j t)\right)^2 \left(\varsigma_j^2 + \varsigma^2_{z^{(j,n)}}\right) dt. \tag{57}$$

If we define

$$\tilde{J}^*(q) = \sum_{j=1}^m \frac{\varsigma_j^2}{3\gamma_j} |q_j|^3,$$

then we have

$$\begin{aligned}J_n(q) &= \int_0^{t_n} V_n(t)dt + \int_{t_n}^{\infty} V_n(t)dt. \\ &\leq \frac{1}{3}||\sum_{j=1}^{m} q_j l_j||_2^2 t_n + \int_0^{\infty} \sum_{j=1}^{m} ((q_j - \gamma_j t))^2 \left(\varsigma_j^2 + \varsigma_{z(j,n)}^2\right) dt \\ &= \frac{1}{3}||\sum_{j=1}^{m} q_j l_j||_2^2 t_n + \sum_{j=1}^{m} \frac{\varsigma_{z(j,n)}^2}{3\gamma_j}|q_j|^3 + \sum_{j=1}^{m} \frac{\varsigma_j^2}{3\gamma_j}|q_j|^3.\end{aligned} \tag{58}$$

From the third condition in definition 1, we have

$$\lim_{n\to\infty} t_n = \lim_{n\to\infty} \frac{\sum_{j=1}^{m} |q_j|}{N(n)\underline{\gamma}_n} = \lim_{n\to\infty} \frac{\sum_{j=1}^{m} |q_j|\sqrt{l^2 K^2}}{\sqrt{\lambda_{\min}^{(n)}}\underline{\gamma}_n} = 0$$

By using (55) and the fact that $\lim_{n\to\infty} t_n = 0$, we have

$$\lim_{n\to\infty} J_n(q) \leq \tilde{J}^*(q).$$

Since the optimal cost should be less than or equal to any feasible trading strategy, we have

$$J_n^*(q) \leq J_n(q).$$

Thus, we have proved that

$$\lim_{n\to\infty} J_n^*(q) \leq \lim_{n\to\infty} J_n(q) \leq \tilde{J}^*(q).$$

Combined with Theorem 9, this yields

$$\tilde{J}^*(q) \leq \lim_{n\to\infty} J_n^*(q) \leq \tilde{J}^*(q),$$

and the proof of the theorem is complete. ■

Theorem 11 (Convergence Speed)**.** *Asymptotically, the difference between the liquidity cost and the theoretical limit converges at rate* $1/(\underline{\gamma}_n\sqrt{\lambda_{\min}^{(n)}})$*:*

$$\limsup_{n\to\infty} \underline{\gamma}_n \sqrt{\lambda_{\min}^{(n)}} |J_n^*(q) - J_\infty^*(q)| < \infty. \tag{24}$$

Proof. According to (48), (52), (55), and (58), we have

$$
\begin{aligned}
J_n(q) - \tilde{J}^*(q) &\le \frac{1}{3}||\sum_{j=1}^m q_j l_j||_2^2 t_n + \sum_{j=1}^m \frac{\varsigma_{z(j,n)}^2}{3\gamma_j}|q_j|^3 \\
&= \frac{1}{3}||\sum_{j=1}^m q_j l_j||_2^2 \frac{\sum_{j=1}^m |q_j|}{N(n)\underline{\gamma}_n} + \sum_{j=1}^m \frac{\varsigma_{z(j,n)}^2}{3\gamma_j}|q_j|^3 \\
&\le A\frac{1}{\underline{\gamma}_n\sqrt{\lambda_{\min}^{(n)}}} + B\frac{1}{\lambda_{\min}^{(n)}},
\end{aligned}
\tag{59}
$$

where A, B are constants that are not related to n:

$$
A = \frac{1}{3}||\sum_{j=1}^m q_j l_j||_2^2 \sum_{j=1}^m |q_j|\sqrt{\bar{l}^2 K^2}, \quad B = \sum_{j=1}^m \frac{|q_j|^3}{3\gamma_j}.
$$

Given that $\lambda_{\min}^{(n)} \to \infty$, we have

$$
\limsup_{n\to\infty} \sqrt{\lambda_{\min}^{(n)}}|J_n^*(q) - \tilde{J}^*(q)| \le \limsup_{n\to\infty} \sqrt{\lambda_{\min}^{(n)}}|J_n(q) - \tilde{J}^*(q)| \le A.
$$

■

Theorem 12 (Random factor loading). *If the asset factor loadings are drawn independently from a K-dimensional distribution (with a finite second moment), then, asymptotically, we have*

$$
\frac{\lambda_{\min}^{(n)}}{n} \stackrel{a.s.}{\to} C, \tag{25}
$$

where C is some constant that depends on only the distribution of factor loadings. If we further assume that

$$
\inf_j \gamma_j > 0,
$$

then we have

$$
\limsup_{n\to\infty} \sqrt{n}|J_n^*(q) - J_\infty^*(q)| < \infty, a.s. \tag{26}
$$

Proof. Now suppose that the factor loadings are i.i.d., and define $\hat{G} \in \mathbb{R}^{K\times K}$ as

$$
\hat{G}_{ij} = \begin{cases} E[l_{ki}l_{kj}], & i \neq j, \\ E[l_{ki}^2], & i = j. \end{cases}
$$

Given the matrix $G^{(n)} = (L^{(n)})^\top L^{(n)}$, we have

$$
G_{ij}^{(n)} = \begin{cases} \sum_{k=1}^n l_{ki}l_{kj}, & i \neq j, \\ \sum_{k=1}^n l_{ki}^2. \end{cases}
$$

Then, by adopting the strong law of large numbers, we have

$$\frac{G^{(n)}}{n} \stackrel{a.s.}{\rightarrow} \hat{G}.$$

Suppose that $\hat{\lambda}_{\min}$ is the smallest eigenvalue of $\hat{G}$. It is easy to see that $\det(\hat{G}) > 0$ if there is no perfect linearity in the factor loadings. As a result, we have

$$\frac{\lambda_{\min}^{(n)}}{n} \stackrel{a.s.}{\rightarrow} \hat{\lambda}_{\min}. \tag{60}$$

The theorem is proved by plugging (58) into Theorem 11.

■

Theorem 13 (Generalization). *For any liquidation model specified in 27, if the transaction cost functional is twice differentiable with*

$$\hat{f}'(0) = 0, \quad \hat{f}(0) = 0,$$

the extended results of Theorem 10 still hold. More specifically, we have

$$\lim_{n\to\infty} J_n^*(q) = \sum_{j=1}^{m} \tilde{J}^*(q_j, \varsigma_j, \nu_j, \gamma_j), \tag{29}$$

where $J_n^*(q)$ represents the optimal liquidity costs for a portfolio q with assets in $\mathcal{A}_n$.

Proof. We first prove the following lemma.

Lemma 4. *Suppose that f is a convex function:*

$$f : [-1, 1] \to \mathbb{R}_+ \cup \{+\infty\}.$$

If $f(\cdot)$ is twice differentiable and

$$f(0) = 0, \quad f'(0) = 0, \quad f''(0) > 0, \tag{61}$$

then there exists $\delta > 0$ such that for any $|\beta| < \delta$,

$$f(\beta x) \leq C\beta^2 f(x), \quad \forall |x| \leq 1, \tag{62}$$

where C is some constant.

Proof. The case for $x = 0$ is trivial. Turning to the case where $x \neq 0$, by convexity we know that $f(x) > 0$:

$$\lim_{\beta\to 0} \frac{f(\beta x)}{(\beta x)^2} = f''(0). \tag{63}$$

Hence there exists δ such that for $\forall \beta < \delta$ we have

$$\frac{1}{2}f''(0) \leq \frac{f(\beta x)}{(\alpha x)^2} \leq \frac{3}{2}f''(0).$$

Denote

$$C_1 = \max_x(\frac{x^2}{f(x)}).$$

Then, $\forall \alpha < \delta$, we have

$$\frac{f(\alpha x)}{\alpha^2 f(x)} = \frac{f(\alpha x)}{\alpha^2 x^2}\frac{x^2}{f(x)} \leq \frac{3}{2}f''(0)C_1.$$

This proves the lemma. ■

Following the proof in Theorem 10, we construct a series of portfolios $\{z^{(j,n)}\}$ for each asset $j \leq m$. Moreover, we denote $\tilde{u}_j^*(t)$ to be the optimal trading strategy of liquidating asset j alone without hedging with other assets and only considering its idiosyncratic risk.

To simplify the notations, we denote $\bar{\beta}_n^{(j,n)} = \sum_{j=1}^m \frac{-q_j}{\sum_{s=1}^m |q_s|}\hat{\beta}_k^{(j,n)}$, where $\hat{\beta}_k^{(j,n)}$ is defined in (51). It is easy to see that we also have $\sum_{k=1}^n (\bar{\beta}_k^{(j,n)})^2 \to 0$.

Define

$$N(n) = \sqrt{\frac{1}{\sum_{k=1}^n (\bar{\beta}_k^{(j,n)})^2}};$$

hence we have $N(n) \to \infty$.

Now consider the following trading strategy:

1. For $0 \leq t \leq t_n$, buy portfolio $z^{(j,n)}$ at a rate of $\frac{-q_j}{\sum_{j=1}^m |q_j|}\sqrt{N(n)}$, where t_n is given by

$$t_n = \frac{\sum_{j=1}^m |q_j|}{\sqrt{N(n)}}.$$

2. For $t > t_n$, trade asset j at a rate of $\tilde{u}_j^*(t - t_n)$ and trade $z^{(j,n)}$ at a rate of $-\tilde{u}_j^*(t - t_n)$. Do it for all $1 \leq j \leq m$.

In this case, the total cost is made up of two parts, namely, transaction costs and market risks. Let us look at them separately.

First, following this trading strategy $(V_n(t))$, the market risk contribution of the position at time t is given by

$$V_n(t) = \begin{cases} \left[\sum_{j=1}^m q_j(l_j - \frac{t}{t_n} l_{z(j,n)})\right]^\top \left[\sum_{j=1}^m q_j(l_j - \frac{t}{t_n} l_{z(j,n)})\right] \\ \quad + \sum_{j=1}^m (q_j - \gamma_j t)^2 \varsigma_j^2 + \sum_{j=1}^m (q_j - \gamma_j t_n)^2 \frac{t^2}{t_n^2} \varsigma_{z(j,n)}^2, \qquad 0 \le t \le t_n, \\ \left[\sum_{j=1}^m \left(q_j - \int_0^{t-t_n} \tilde{u}_j^*(s)ds\right)(l_j - l_{z(j,n)})\right]^\top \left[\sum_{j=1}^m \left(q_j - \int_0^{t-t_n} \tilde{u}_j^*(s)ds\right)(l_j - l_{z(j,n)})\right] \\ \quad + \sum_{j=1}^m \left(q_j - \int_0^{t-t_n} \tilde{u}_j^*(s)ds\right)^2 \left(\varsigma_j^2 + \varsigma_{z(j,n)}^2\right), \qquad otherwise. \end{cases}$$

where $\varsigma_{z(j,n)}^2, l_{z(j,n)}$ are defined in the same way. Further, we have

$$l_{z(j,n)} = l_j, \quad \lim_{n\to\infty} \varsigma_{z(j,n)}^2 = 0, \quad \lim_{n\to\infty} t_n = 0. \tag{64}$$

Similarly to the proofs in Theorem 10, it can be shown that

$$\lim_{n\to\infty} \int_0^\infty V_n(t)dt = \sum_{j=1}^m \int_0^\infty \left(q_j - \int_0^t \tilde{u}_j^*(s)ds\right)^2 \varsigma_j^2 dt.$$

Now let's consider the transaction costs. The transaction costs at time t ($T_n(t)$) are given by

$$T_n(t) = \begin{cases} \sum_{k=1}^n \nu_k \hat{f}\left(\sum_{j=1}^m \frac{-q_j\sqrt{N(n)}}{\sum_{j=1}^m |q_j|} \hat{\beta}_k^{(j,n)}/\gamma_k\right) = \sum_{k=1}^n \nu_k \hat{f}\left(\bar{\beta}_k^{(j,n)}\sqrt{N(n)}/\gamma_k\right), \qquad 0 \le t \le t_n, \\ \sum_{k=1}^n \nu_k \hat{f}\left(-\sum_{j=1}^m \hat{\beta}_k^{(j,n)} \tilde{u}_j^*(t-t_n)/\gamma_k\right) \\ \quad + \sum_{j=1}^m \nu_j \hat{f}\left(\tilde{u}_j^*(t-t_n)/\gamma_j + \sum_{k=1}^m \hat{\beta}_j^{(k,n)} \tilde{u}_k^*(t-t_n)/\gamma_j\right), \qquad otherwise. \end{cases}$$

Notice that $|\bar{\beta}_k^{(j,n)}| \le \sqrt{\sum_{k=1}^n (\hat{\beta}_{nk}^{(j)})^2} = 1/N(n)$; hence we have $\bar{\beta}_k^{(j,n)}\sqrt{N(n)} \to 0$. Then we can take the Taylor expansion of $\hat{f}(\cdot)$ around 0:

$$\hat{f}\left(\bar{\beta}_k^{(j,n)}\sqrt{N(n)}/\gamma_k\right) = \frac{1}{2}N(n)(\bar{\beta}_k^{(j,n)})^2/\gamma_k^2 + o(N(n)(\bar{\beta}_k^{(j,n)})^2).$$

We have

$$
\begin{aligned}
\lim_{n\to\infty}\int_0^{t_n} T(t)dt &= \lim_{n\to\infty}\int_0^{t_n}\sum_{k=1}^{n}\nu_k\hat{\hat{f}}\left(\bar{\beta}_k^{(j,n)}\sqrt{N(n)}/\gamma_k\right)dt \\
&= \lim_{n\to\infty}\int_0^{t_n}\sum_{k=1}^{n}\nu_k\left(\frac{1}{2}N(n)(\bar{\beta}_k^{(j,n)})^2/\gamma_k^2 + o(N(n)(\bar{\beta}_k^{(j,n)})^2)\right)dt \\
&\le \lim_{n\to\infty}\frac{\bar{\nu}}{2\underline{\gamma}_n^2}t_n\left(N(n)\sum_{k=1}^{n}(\bar{\beta}_k^{(j,n)})^2\right) \\
&= \lim_{n\to\infty}\frac{\bar{\nu}}{2\underline{\gamma}_n^2}t_n \\
&= 0.
\end{aligned} \tag{65}
$$

Then,

$$
\begin{aligned}
\lim_{n\to\infty}\int_{t_n}^{\infty} T(t)dt = \lim_{n\to\infty}&\int_{t_n}^{\infty}\sum_{k=1}^{n}\nu_k\hat{f}\left(-\sum_{j=1}^{m}\hat{\beta}_{nk}^{(j)}\tilde{u}_j^*(t-t_n)/\gamma_k\right)dt \\
&+ \lim_{n\to\infty}\int_{t_n}^{\infty}\sum_{j=1}^{m}\nu_j\hat{f}\left(\tilde{u}_j^*(t-t_n)/\gamma_j + \sum_{k=1}^{m}\hat{\beta}_{nj}^{(k)}\tilde{u}_k^*(t-t_n)/\gamma_j\right)dt.
\end{aligned} \tag{66}
$$

Notice that the second term is

$$
\lim_{n\to\infty}\int_{t_n}^{\infty}\sum_{j=1}^{m}\nu_j\hat{f}\left(\tilde{u}_j^*(t-t_n)/\gamma_j + \sum_{k=1}^{m}\hat{\beta}_{nj}^{(k)}\tilde{u}_k^*(t-t_n)/\gamma_j\right)dt = \int_0^{\infty}\sum_{j=1}^{m}\nu_j\hat{f}(\tilde{u}_j^*(t)/\gamma_j)dt. \tag{67}
$$

Now it remains to show that the first term converges to 0:

$$
\begin{aligned}
&\lim_{n\to\infty}\int_{t_n}^{\infty}\sum_{k=1}^{n}\nu_k\hat{f}\left(-\sum_{j=1}^{m}\hat{\beta}_{nk}^{(j)}\tilde{u}_j^*(t-t_n)/\gamma_k\right)dt \\
&\le \bar{\nu}\sum_{j=1}^{m}\lim_{n\to\infty}\int_{t_n}^{\infty}\sum_{k=1}^{n}\hat{f}\left(\hat{\beta}_{nk}^{(j)}\tilde{u}_j^*(t-t_n)/\underline{\gamma}_n\right)dt \\
&\le \bar{\nu}\sum_{j=1}^{m}\lim_{n\to\infty}\sum_{k=1}^{n}(\hat{\beta}_{nk}^{(j)})^2\int_{t_n}^{\infty}\hat{f}\left(\tilde{u}_j^*(t-t_n)/\underline{\gamma}_n\right)dt \\
&= 0.
\end{aligned} \tag{68}
$$

The first inequality is a direct application of 4, and the last equality is due to the fact that $\sum_{k=1}^{n}(\hat{\beta}_{nk}^{(j)})^2 \to 0$.

Notice that the total liquidity cost of this strategy ($J_n(q)$) is given by

$$
J_n(q) = \int_0^{\infty} T_n(t)dt + \mu\int_0^{\infty} V_n(t). \tag{69}
$$

Given that this cost should never be smaller than the optimal cost, together with (65), (66), (67), (68), and (69), we have

$$\begin{aligned}\lim_{n\to\infty} J_n^*(q) \le \lim_{n\to\infty} J_n(q) &= \int_0^\infty \sum_{j=1}^m \nu_j \hat{f}(\tilde{u}_j^*(t)/\gamma_j)dt + \mu \sum_{j=1}^m \int_0^\infty \left(q_j - \int_0^t \tilde{\gamma}_j u_j^*(s)ds\right)^2 \varsigma_j^2 \\ &= \sum_{j=1}^m \tilde{J}^*(q_j, \varsigma_j, \nu_j, \gamma_j).\end{aligned}$$

■

Theorem 14 (Linear costs).

$$\lim_{n\to\infty} \lim_{||q||_\infty\to\infty} \frac{\tilde{J}_{LC}^*(q)}{J_{LC,n}^*(q)} = 1 \tag{30}$$

where

$$\tilde{J}_{LC}^*(q) = \mu \sum_{j=1}^m \frac{\varsigma_j^2}{3\gamma_j} |q_j|^3.$$

Proof. From Theorem 9, we know that the term in the denominator is the lower bound of the liquidity cost of the portfolio when we neglect the transaction costs. Thus, it also has to be the lower bound of $J_n^*(q)$:

$$\frac{\mu \sum_{j=1}^m \frac{\varsigma_j^2}{3\gamma_j} |q_j|^3}{J_n^*(q)} \le 1. \tag{70}$$

Now consider the trading strategy given in the proof of Theorem 10; the only difference here is that we also have to calculate the transaction cost from this strategy. The transaction cost comes from two sources: the transaction cost of selling the position in the portfolio that is given by $\sum_{j=1}^m \nu_j |q_j|$ and the transaction cost of establishing and liquidating the hedging positions. More specifically, the trading cost of trading portfolio $z^{(j,n)}$ is given by

$$\nu_{z^{(j,n)}} = \sum_{k=1}^n |\hat{\beta}_k^{(j,n)}| \nu_k. \tag{71}$$

By Cauchy−Schwarz inequality, we have

$$\left(\sum_{k=1}^n |\hat{\beta}_k^{(j,n)}|\right)^2 \le n \sum_{k=1}^n |\hat{\beta}_k^{(j,n)}|^2 \to 0, \quad n \to \infty, \tag{72}$$

Combining (71) and (72), we know that when n is large enough,

$$\nu_{z^{(j,n)}} < \bar{\nu}\sqrt{n}. \tag{73}$$

The total transaction costs for the nth problem are given by

$$TC_n \le 2\sum_{j=1}^{m} \nu_{z(j,n)}|q_j| + \sum_{j=1}^{m} \nu_j|q_j| < \sum_{j=1}^{m}(2\bar{\nu}\sqrt{n}+\nu_j)|q_j| < m(2\bar{\nu}\sqrt{n}+\bar{\nu})||q||_\infty.$$

Thus we have

$$\lim_{||q||_\infty\to\infty} \frac{TC_n}{||q||_\infty} = 0.$$

$$\begin{aligned}\lim_{n\to\infty}\lim_{||q||_\infty\to\infty} \frac{\mu\sum_{j=1}^{m}\frac{\varsigma_j^2}{3\gamma_j}|q_j|^3}{J^*_{LC,n}(q)} &\ge \lim_{n\to\infty}\lim_{||q||_\infty\to\infty} \frac{\mu\sum_{j=1}^{m}\frac{\varsigma_j^2}{3\gamma_j}|q_j|^3}{J_n(q)+TC_n} \\ &\ge \lim_{n\to\infty} \frac{\mu\sum_{j=1}^{m}\frac{\varsigma_j^2}{3\gamma_j}|q_j|^3}{J_n(q)} = 1.\end{aligned} \tag{74}$$

Combining (70) and (74), we complete the proof of the theorem.

■

Theorem 15 (ETF). *If the* large-universe *property is satisfied, then, asymptotically, the cost for liquidating q_j shares of asset $j \le m$ is given by*

$$\lim_{n\to\infty} J^*_{\mathrm{ETF},n}(q) = \frac{\varsigma_j^2}{3\left(|\alpha_1|\gamma_{\mathrm{ETF}}+\gamma_j\right)} q_j^3, \tag{32}$$

where $q_i = 0, \forall i \ne j$.

Proof. We start with the following observations:

1. For asset j, the fastest trading rate attainable is $|\alpha_j|\gamma_{\mathrm{ETF}}+\gamma_j$. It is obtained by selling asset j and the ETF at full rate, but at the same time buying back other assets in the ETF, so that the net liquidity contribution from trading the ETF is just $|\alpha_j|\gamma_{\mathrm{ETF}}$. The feasibility of this strategy is guaranteed by the assumption that $|\alpha_i\gamma_{\mathrm{ETF}}| < \gamma_i$.

2. At any time point during the liquidation process, the total risk is made up of three components: the idiosyncratic risk of asset j, which is given by $\varsigma_j^2 x_j^2(t)$; idiosyncratic risks of the hedging positions; and the entire portfolio's market risk. Since the latter two terms are nonnegative, the total risk term is always greater than or equal to $\varsigma_j^2 x_j^2(t)$.

From these two observations, we have

$$J^*_{\mathrm{ETF},n}(q) \ge \int_0^\infty \varsigma_j^2 x_j^2(t)dt \ge \frac{\varsigma_j^2}{3\left(|\alpha_j|\gamma_{\mathrm{ETF}}+\gamma_j\right)} q_j^3.$$

As a result, the right-hand side of (32) is actually the lower bound of the liquidity cost.

As discussed above, it is possible to trade asset j at rate $|\alpha_j|\gamma_{\text{ETF}} + \gamma_j$ by essentially trading a portfolio p containing the ETF and all the assets in it (this portfolio has only a net position in asset j). By viewing this portfolio as a single asset and using the results in Theorem 10, we have

$$\lim_{n\to\infty} J^*_{\text{ETF},n}(q) = \frac{\varsigma_j^2}{3\gamma_p} q_j^3,$$

where $\gamma_p = |\alpha_1|\gamma_{\text{ETF}} + \gamma_j$, and this completes the proof of the theorem. ■