Optimal Trading Strategy and Supply/Demand Dynamics

Anna Obizhaeva and Jiang Wang*

First Draft: November 15, 2004. This Draft: April 8, 2006

Abstract

The supply/demand of a security in the market is an intertemporal, not a static, object and its dynamics is crucial in determining market participants' trading behavior. In this paper, we show that the dynamics of the supply/demand, rather than its static properties, is of critical importance to the optimal trading strategy of a given order. Using a limit-order-book market, we develop a simple framework to model the dynamics of supply/demand and its impact on execution cost. We show that the optimal execution strategy involves both discrete and continuous trades, not only continuous trades as previous work suggested. The cost savings from the optimal strategy over the simple continuous strategy can be substantial. We also show that the predictions about the optimal trading behavior can have interesting implications on the observed behavior of intraday volume, volatility and prices.

*Obizhaeva is from MIT Sloan School of Management, tel: (617)253-3919, email: obizhaeva@mit.edu, and Wang is from MIT Sloan School of Management, CCFR and NBER, tel: (617)253-2632, email: wangj@mit.edu. The authors thank Robert Fernstenberg, Thierry Foucault, William Goetzmann, Christine Parlour, Ioanid Rosu, participants of the 2006 AFA Meetings, the 2005 EFA Meetings and seminars at MIT, Morgan Stanley, and NYSE for helpful comments.

Contents

1 Introduction

It has being well documented that the supply/demand of a security in the market is not perfectly elastic.[1] The limited elasticity of supply/demand or liquidity can significantly affect how market participants trade, which in turn will influence security prices through the changes in their supply/demand.[2] Thus, to study how market participants trade is important to our understanding of how securities markets function, how liquidity is provided and consumed, and how it affects the behavior of security prices.[3] In the paper, we approach this problem by focusing on the optimal strategy to execute a given order, leaving aside its underlying motive. This is also referred to as the optimal execution problem. We show that it is the dynamic properties of supply/demand such as its time evolution after trades, rather than its static properties such as the instantaneous price impact function, that are central to the cost of trading and the optimal strategy.

We consider a limit-order-book market, in which the supply/demand of a security is represented by the limit orders posted to the "book," i.e., a trading system and trade occurs when buy and sell orders match. We propose a simple framework to describe the limit-order-book and how it evolves over time. By incorporating several salient features of the book documented empirically, we attempt to capture the dynamics of supply/demand a trader faces. We show that the optimal trading strategy crucially depends on how the limit-order book responds to a sequence of trades and it involves complex trading patterns including both discrete and continuous trades.

In particular, the optimal strategy consists of an initial discrete trade, followed by a sequence of continuous trades. The initial discrete trade is aimed at pushing the limit order book away from its steady state in order to attract new orders onto the book. The size of the initial trade is chosen to draw sufficient new orders at desirable prices. The subsequent continuous trades will then pick off the new orders and keep the inflow coming. A discrete trade finishes off any remaining order at the end of trading horizon when future demand/supply is no longer of concern. The combination of discrete and continuous trades

[1] See, for example, Holthausen, Leftwitch and Mayers (1987, 1990), Shleifer (1986), Scholes (1972). For the more recent work, see also Greenwood (2004), Kaul, Mehrotra and Morck (2000), Wugler and Zhuravskaya (2002). There is also extensive theoretical work in justifying an imperfect demand/supply in securities market based on market frictions and asymmetric information. See, for example, Grossman and Miller (1998), Kyle (1985) and Vayanos (1999, 2001).

[2] Many empirical studies have shown that this is a problem confronted by institutional investors who need to execute large orders and often break up trades in order to manage the trading cost. See, for example, Chan and Lakonishok (1993, 1995, 1997), Keim and Madhavan (1995, 1997).

[3] For example, Kyle (1985) and Wang (1993) examine the behavior of traders with superior information and how it affects liquidity and asset prices and Vayanos (1999, 2001) considers the trading behavior of large traders with risk-sharing needs and its impact on market behavior.

for the optimal execution strategy is in sharp contrast to simple strategies of splitting a order into small trades as suggested in the literature. Moreover, we find that the optimal strategy and the cost saving depends primarily on the dynamic properties of supply/demand and is not very sensitive to the instantaneous price-impact function, which has been the main focus in previous work. Especially, the speed at which the limit order book rebuilds itself after being hit by a trade, which is also referred to as the resilience of the book, plays a critical role in determining the optimal execution strategy and the cost it saves.

Our predictions about optimal trading strategies lead to interesting implications about the behavior of trading volume, liquidity and security prices. For example, it suggests that the trading behavior of large institutional traders may contribute to the observed U-shaped patterns in intraday volume, volatility and bid-ask spread. It also suggests that these patterns can be closely related to institutional ownership and the resilience of the supply/demand of each security.

The problem of optimal execution takes the order to be executed as given. Ideally, we should consider both the optimal size of an order and its execution, taking into account the underlying motives to trade (e.g., return and risk, preferences and constraints) and the costs to execute trades.[4] The diversity in trading motives makes it difficult to tackle such a problem as a general level. Given that in practice the execution of trades is often separated from the decisions on the trades, in this paper we focus on the execution problem as an important and integral part of the general problem of optimal trading behavior.

Several authors have studied the problem of optimal execution. For example, Bertsimas and Lo (1998) propose a linear price impact function and solve for the optimal execution strategy to minimize the expected cost of executing a given order. Almgren and Chriss (1999, 2000) include risk considerations in a similar setting using a mean-variance objective function.[5] The framework adopted in these papers share two main features. First, it uses a discrete-time setting so that the times to trade are fixed at given intervals. Second, it relies on price impact functions to describe how a sequence of trades affects prices at which trades are executed. A discrete-time setting is clearly undesirable for the execution problem because the timing of trades is an important choice variable and should be determined optimally. A natural way to address this issue would be to take a continuous-time limit of the discrete-time formulation. But such a limit leads to degenerate solutions with the simple price impact functions considered previously. In particular, Lo and Bertsimas (1998) consider the permanent price impact by assuming a static, linear impact function. As a result, the

[4]For example, many authors have considered the problem of optimal portfolio choices in the presence of transactions costs, e.g., Constantinides (1986), Davis and Norman (1990), and Leland (2000).

[5]See also, Almgren (2003), Dubil (2002), Huberman ans Stanzl (2005), Subramanian and Jarrow (2001), among others.

price impact of a sequence of trades depends only on their total size and is independent of their distribution over time. In this case, the execution cost becomes strategy independent in the continuous-time limit. Almgren and Chris (1999, 2000) and Huberman and Stanzl (2005) also allow temporary price impact, which depends on the pace of trades. Introducing temporary price impact adds a dynamic element to the price impact function by penalizing speedy trades. But it restricts the execution strategy to continuous trades in the continuous-time limit, which is in general sub-optimal.

The simple price impact functions used in previous work do not fully capture the intertemporal nature of supply/demand in the market. In particular, it limits the extent to which the allocation of trades over time, given their sizes, influences current and future supply/demand and the resulting execution cost. Yet, it is clear that how to allocate trades over time is at the heart of the problem. Thus, modelling the intertemporal properties of supply/demand is essential in analyzing the optimal execution strategy. Taking these considerations into account, our framework attempts to capture these intertemporal aspects of the supply/demand by directly modelling the liquidity dynamics in a limit-order-book market. We show that when the timing of trades is chosen optimally, the optimal execution strategy differs significantly from those suggested in earlier work and yields substantial cost reduction. It involves a mixture of discrete and continuous trades. Moreover, the characteristics of the optimal execution strategy are mostly determined by the dynamic properties of the supply/demand rather than its static properties as described by the price impact function.

In modelling the supply/demand dynamics, we choose the limit-order-book market mainly for concreteness. Our description of the limit-order-book dynamics relies on an extensive empirical literature.[6] We choose the shape of the limit-order-book to yield a linear price-impact function, which is widely adopted in previous work. More importantly, we explicitly model the resilience of the book, which several empirical studies document as an important property of the book (see, e.g., Biais, Hillion and Spatt (1995) and Harris (1990)).

Our analysis is partial equilibrium in nature. We take the dynamics of the limit-order-book as given and do not attempt to provide an equilibrium justification for the specific limit-order-book dynamics used in the paper. Nonetheless, it is worth pointing out that in addition to the empirical motivation mentioned above, the supply/demand dynamics we consider is also consistent with several equilibrium models (e.g., Kyle (1985) and Vayanos (1999, 20001)). In particular, Vayanos (2001) analyzes the optimal trading behavior of a large trader who trades with a set of competitive market makers for risk sharing. He shows

[6]See, for example, Ahn, Bae and Chan (2000) for a study on the Hong Kong Stock Exchange, Biais, Hillion and Spatt (1995) on the Paris Bourse, Chung, Van Ness and Van Ness (1999) on the NYSE, Hasbrouck and Saar (2002) on the Island ECN, Hollfield, Miller and Sandas (2003) on the Stockholm Stock Exchange and Griffiths, Smith, Turnbull and White (2000) on the Toronto Stock Exchange.

that the price impact of the large trader is linear in his trades and the supply/demand by the market makers exhibits certain form of resilience. Although his analysis relies on specific assumptions on traders' trading motives and preferences, it does provide additional theoretical basis for the qualitative properties of supply/demand dynamics we consider.

Several authors have also considered equilibrium models for the limit-order-book market, including Foucault, Kadan and Kandel (2004), Goettler, Parlour and Rajan (2005) and Rosu (2005). For tractability, the set of order-placement strategies allowed in studies are severely limited to obtain an equilibrium. For example, Foucault, Kadan and Kandel (2004) and Rosu (2005) only allow orders of a fixed size. Goettler, Parlour and Rajan (2005) focus on one-shot strategies. These simplifications are helpful when we are interested in certain properties of the book, but quite restrictive when analyzing the optimal trading strategy. A more general and realistic equilibrium model must allow general strategies. From this perspective, our analysis, namely to solve the optimal execution strategy under general supply/demand dynamics, is an unavoidable step in this direction.

The rest of the paper is organized as follows. Section 2 states the optimal execution problem. Section 3 introduces the limit-order-book market and a model for the limit order book dynamics. In Section 4, we show that the conventional setting in previous work can be viewed as a special case of our limit-order-book framework. We also explain why the stringent assumptions in the conventional setting lead to its undesirable properties. In Section 5, we solve the discrete-time version of the problem within our framework. We also consider its continuous-time limit and show that it is economically sensible and properly behaved. Section 6 provides the solution of the optimal execution problem in the continuous-time setting. In Section 7, we analyze the properties of the optimal execution strategy and their dependence on the dynamics of the limit order book. We also compare it with the strategy predicted by the conventional setting. In addition, we examine the empirical implications of the optimal execution strategy. Section 8 discusses possible extensions of the model. Section 9 concludes. All proofs are given in the appendix.

2 Statement of the Problem

The problem we are interested in is how a trader optimally executes a given order. To fix ideas, let us assume that the trader has to buy X_0 units of a security over a fixed time period $[0,\ T]$. Suppose that the trader ought to complete the order in $N+1$ trades at times $t_0, t_1, \ldots, t_N$, where $t_0 = 0$ and $t_N = T$. Let x_{t_n} denote the trade size for the trade at t_n. We then have

$$\sum_{n=0}^{N} x_{t_n} = X_0. \tag{1}$$

A strategy to execute the order is given by the number of trades, $N+1$, the set of times to trade, $\{0 \le t_0, t_1, \ldots, t_{N-1}, t_N \le T\}$ and trade sizes $\{x_{t_0}, x_{t_1}, \ldots, x_{t_N} : x_{t_n} \ge 0 \; \forall \; n \text{ and } (1)\}$. Let Θ_D denote the set of these strategies:

$$\Theta_D = \left\{ \{x_{t_0}, x_{t_1}, \ldots, x_{t_N}\} : \; 0 \le t_0, t_1, \ldots, t_N \le T; \; x_{t_n} \ge 0 \; \forall \; n; \; \sum_{n=0}^{N} x_{t_n} = X_0 \right\}. \tag{2}$$

Here, we have assumed that the strategy set consists of execution strategies with finite number of trades at discrete times. This is done merely for easy comparison with previous work. Later we will expand the strategy set to allow uncountable number of trades over time.

Let $\bar{P}_n$ denote the average execution price for trade x_{t_n}. We assume that the trader chooses his execution strategy to minimize the expected total cost of his purchase:

$$\min_{x \in \Theta_D} \; \mathrm{E}_0 \left[\sum_{n=0}^{N} \bar{P}_n x_n \right]. \tag{3}$$

For simplicity, we have assumed that the trading horizon T is fixed and the trader is risk-neutral who cares only about the expected value not the uncertainty of the total cost. We will incorporate risk considerations later (in Section 8), which also allows us to endogenize the trading horizon.

The solution to the trader's optimal execution strategy crucially depends on how his trades impact the prices. It is important to recognize that the price impact of a trade has two key dimensions. First, it changes the security's *current* supply/demand. For example, after a purchase of x units of the security at the current price of $\bar{P}$, the remaining supply of the security at $\bar{P}$ in general decreases. Second, a change in current supply/demand can lead to evolutions in future supply/demand, which will affect the costs for future trades. In other words, the price impact is determined by the full dynamics of supply/demand in response to a trade. Thus, in order to fully specify the optimal execution problem, we need to model the supply/demand dynamics.

3 Limit Order Book and Supply/Demand Dynamics

The actual supply/demand of a security in the market place and its dynamics depend on the actual trading process. From market to market, the trading process varies significantly, ranging from a specialist market or a dealer market to a centralized electronic market with a limit order book. In this paper, we consider the limit-order-book market, which is arguably the closest, at least in form, to the text-book definition of a centralized market.

3.1 Limit Order Book (LOB)

A limit order is a order to trade a certain amount of a security at a given price. In a market operated through a limit-order-book, thereafter LOB for short, traders post their supply/demand in the form of limit orders to a electronic trading system.[7] A trade occurs when an order, say a buy order, enters the system at the price of an opposite order on the book, in this case a sell order, at the same price. The collection of all limit orders posted can be viewed as the total demand and supply in the market.

Let $q_A(P)$ be the density of limit orders to sell at price P and $q_B(P)$ the density of limit orders to buy at price P. The amount of sell orders in a small price interval $[P,\ P{+}dP)$ is $q_A(P)(P{+}dP)$. Typically, we have

$$q_A(P) = \begin{cases} +, & P \geq A \\ 0, & P < A \end{cases} \qquad \text{and} \qquad q_B(P) = \begin{cases} 0, & P > B \\ +, & P \leq B \end{cases}$$

where $A \geq B$ are the best ask and bid prices, respectively. We define

$$V = (A{+}B)/2, \quad s = A{-}B \tag{4}$$

where V is the mid-quote price and s is the bid-ask spread. Then, $A = V + s/2$ and $B = V{-}s/2$. Because we are considering the execution of a large buy order, we will focus on the upper half of the LOB and simply drop the subscript A.

In order to model the execution cost for a large order, we need to specify the initial LOB and how it evolves after been hit by a series of buy trades. Let the LOB (the upper half of it) at time t be $q(P; F_t; Z_t; t)$, where F_t denotes the fundamental value of the security and Z_t represents the set of state variables that may affect the LOB such as past trades. We will consider a simple model for the LOB, to capture its dynamic nature and to illustrate their importance in analyzing the optimal execution problem, and return to its extensions to better fit the empirical LOB dynamics later. In particular, we assume that the fundamental value the security F_t follows a Brownian motion, reflecting the fact that in absence of any trades, the mid-quote price may change due to news about the fundamental value of the

[7] The number of exchanges adopting an electronic trading system with posted orders has been increasing. Examples include NYSE's OpenBook program, Nasdaq's SuperMontage, Toronto Stock Exchange, Vancouver Stock Exchange, Euronext (Paris, Amsterdam, Brussels), London Stock Exchange,Copenhagen Stock Exchange, Deutsche Borse, and Electronic Communication Networks such as Island. For the fixed income market, there are, for example, eSpeed, Euro MTS, BondLink and BondNet. Examples for the derivatives market include Eurex, Globex, and Matif.

security. Thus, $V_t = F_t$ in absence of any trades and the LOB maintains the same shape except that the mid-point, V_t, is changing with F_t. In addition, we assume that the only set of relevant state variables is the history of past trades, which we denote by $x_{[0,\,t]}$, i.e., $Z_t = x_{[0,\,t]}$.

At time 0, we assume that the mid-quote is $V_0 = F_0$ and LOB has a simple block shape

$$q_0(P) \equiv q(P; F_0; 0; 0) = q\, 1_{\{P \geq A_0\}}$$

where and $A_0 = F_0 + s/2$ is the initial ask price and $1_{\{z \geq a\}}$ is an indicator function:

$$1_{\{z\geq a\}} = \begin{cases} 1, & z \geq a \\ 0, & z < a \end{cases}$$

In other words, q_0 is a step function of P with a jump from zero to q at the ask price $A_0 = V_0 + s/2 = F_0 + s/2$. The first panel in Figure 1 shows the shape of the book at time 0.

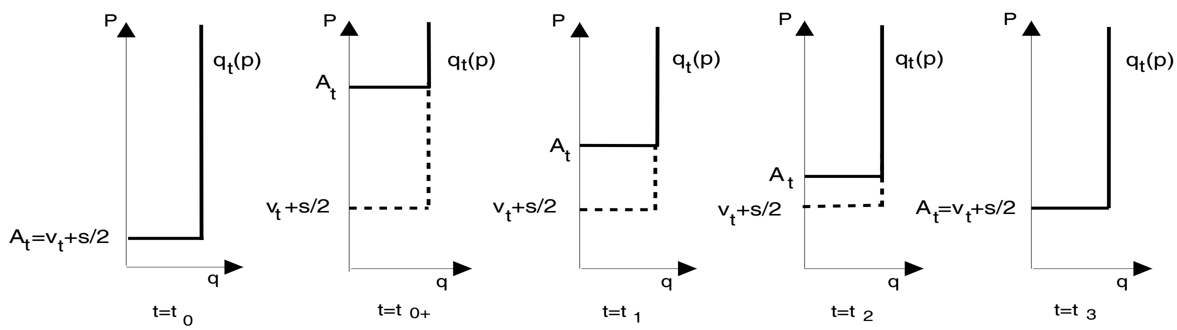

Figure 1: The limit order book and its dynamics. This figure illustrates how the sell side of limit order book evolves over time in response to a sale trade. Before the trade at time $t_0 = 0$, the limit order book is full at the ask price $A_0 = V_0 + s/2$, which is shown in the first panel from the left. The trade of size x_0 at $t = 0$ "eats off" the orders on the book with lowest prices and pushes the ask price up to $A_{0+} = (F_0 + s/2) + x_0/q$, as shown in the second panel. During the following periods, new orders will arrive at the ask price A_t, which fill up the book and lower the ask price until it converges to its new steady state $A_t = F_t + \lambda x_0 + s/2$, as shown in the last panel on the right. For clarity, we assume that there are no fundamental shocks.

Now we consider a trade of size x_0 at $t = 0$. The trade will "eat off" all the sell orders with prices from $F_0 + s/2$ up to A_{0_+}, where A_{0_+} is given by

$$\int_{F_0+s/2}^{A_{0_+}} q dP = x_0$$

or $A_{0_+} = F_0 + s/2 + x_0/q$. The average execution price is $\bar{P} = F_0 + s/2 + x_0/(2q)$, which is linear in the size of the trade. Thus, the shape of the LOB we propose is consistent with the linear price impact function assumed in previous work. This is also the main reason we adopted it here.

Right after the trade, the limit order book becomes:

$$q_{0_+}(P) \equiv q(P; F_0; Z_{0_+}; 0_+) = q\,1_{\{P \geq A_{0_+}\}}.$$

$A_{0_+} = F_0 + s/2 + x_0/q$ is the new ask price. Orders at prices below $A_{0_+} = (F_0 + s/2) + x_0/q$ have all been executed. The book is left with limit sell orders at prices above (including) A_{0_+}. The second panel of Figure 1 plots the limit order book right after the trade.

3.2 Limit Order Book Dynamics

What we have to specify next is how the LOB evolves over time after being hit by a trade. Effectively, this amounts to describing how the new sell orders arrive to fill in the gap in the LOB eaten away by the trade. First, we need to specify the impact of the trade on the mid-quote price, which will determine the prices of the new orders. In general, the mid-quote price will be shifted up by the trade. We assume that the shift in the mid-quote price will be linear in the size of the total trade. That is,

$$V_{0^+} = F_0 + \lambda x_0$$

where $0 \leq \lambda \leq 1/q$ and λx_0 gives the permanent price impact the trade x_0 has. If there are no more trades after the initial trade x_0 at $t = 0$ and there are no shocks to the fundamental, the limit order book will eventually converge to its new steady state

$$q_t(P) = q\,1_{\{P \geq A_t\}}$$

where t is sufficiently large, $A_t = V_t + s/2$ and $V_t = F_0 + \lambda x_0$. Next we need to specify how the limit order book converges to its steady-state. Note that right after the trade, the ask price is $A_{0+} = F_0 + s/2 + x_0/q$, while in the steady-state it is $A_\infty = F_0 + s/2 + \lambda x_0$. The difference between the two is $A_{0+} - A_\infty = x_0(1/q - \lambda)$. We assume that the limit order book converges to its steady state exponentially:

$$q_t(P) = q\,1_{\{P \geq A_t\}} \tag{5}$$

where

$$A_t = V_t + s/2 + x_0 \kappa e^{-\rho t}, \quad \kappa = 1/q - \lambda \tag{6}$$

and $\rho \geq 0$ gives the convergence speed and $V_t = V_{0_+}$ in absence of new trades and changes in F_t, which measures the "resilience" of the LOB.[8]

[8] A number of empirical studies documented the existence of the resiliency of LOB. See, for example, Biais, Hillion and Spatt (1995), Hamao and Hasbrouck (1995), Coppejans, Domowitz and Madhavan (2001), and Ranaldo (2004). Moreover, the idea of liquidity being exhausted by a trade and then replenished as traders

Equations (5) and (6) imply that after a trade x_0, the new sell orders will start coming in at the new ask price A_t at the rate of $\rho q(A_t - V_t - s/2)$. For convenience, we define

$$D_t = A_t - V_t - s/2 \tag{7}$$

which stands for the deviation of current ask price A_t from its steady state level $V_t + s/2$.

We can easily extend the LOB dynamics described above for a single trade to allow multiple trades and shocks to the fundamental value. Let $n(t)$ denote the number of trades during interval $[0,\ t)$, $t_1, \ldots, t_{n(t)}$ the times for these trades, and x_{t_i} their sizes, respectively. Let X_t be the remaining order to be executed at time t, before trading at t. We have

$$X_t = X_0 - \sum_{t_n < t} x_{t_n}. \tag{8}$$

with $X_{T_+} = 0$. Let

$$V_t = F_t + \lambda(X_0 - X_t) = F_t + \lambda \sum_{i=0}^{n(t)} x_{t_i} \tag{9}$$

where $X_0 - X_t$ is the total amount of purchase during $[0,\ t)$. The ask price at any time t is

$$A_t = V_t + s/2 + \sum_{i=0}^{n(t)} x_{t_i} \kappa e^{-\rho(t-t_i)} \tag{10}$$

and the limit order book at any time t is given by (5). Panels 2 to 5 in Figure 1 illustrates the time evolution of the LOB after a trade. We can easily extend the above description to include sell orders which may occur in the mean time and can shift the mid-quote V_t. If not predictable, they are not important to our analysis. Thus, we omit them here.

Before we go ahead with the LOB dynamics and examines its implications on execution strategy, several comments are in order. We note that the simple LOB dynamics described above is assumed to be given, without further economic justification. Presumably, it is driven by the optimizing behavior of those who submit the orders and thus provide liquidity to the market.[9] In addition, the LOB dynamics may be further affected by the strategic interactions among market participants (see, for example, Vayanos (1999, 2001)). To describe the actual

take advantage of profit opportunities is behind most of the dynamic equilibrium frameworks of LOB. See, for example, Foucault, Kadan and Kandel (2004), Goettler, Parlour and Rajan (2005), Parlour (1998), Rosu (2005).

[9] Several recent work show that traders do use the rich information revealed by the books when deciding on their order submission strategies. See, for example, Cao, Hansch, and Wang (2003), Harris and Panchapagesan (2005), Bloomfield, O'Hara, and Saar (2003), Ranaldo (2004), among others. Many authors have developed models for optimal order placement in markets with limit orders. See Foucault (1999), Foucault, Kadan, and Kandel (2001), Glosten (1994), Goettler, Parlou and Rajan (2005), Harris(1998), Parlour (1998), Parlour and Seppi(2003), Rock (1996), Rosu (2005), Sandas (2001), and Seppi (1997). However, as mentioned earlier, most of these models impose strong restrictions on the strategies allowed.

LOB dynamics will require an equilibrium framework. However, for any equilibrium analysis, we first need to study the optimal trading strategy under general LOB dynamics. Thus, our analysis can be viewed as a necessary step along this direction. Clearly, our setting is general enough for this purpose. It should be emphasized that the goal of this paper is to demonstrate the importance of supply/demand dynamics in determining the optimal trading strategy. The specific model we use mainly helps us to make the point in a simple and revealing way. Its partial equilibrium nature as well as its quantitative features are not crucial to our main conclusions.

3.3 Execution Cost

Given the above description of the LOB dynamics, we can now describe the total cost of an execution strategy for a given order X_0. Let x_{t_n} denote the trade at time t_n and A_{t_n} the ask price at t_n prior to the trade. The evolution of ask price A_t as given in (10) is not continuous. For clarity, A_t always denotes the left limit of A_t, $A_t = \lim_{s \to t^-} A_s$, i.e., the ask price before the trade at t. The same convention is followed for V_t. The cost for x_{t_n} is then

$$c(x_{t_n}) = \int_0^{x_{t_n}} P_{t_n}(x)dx \tag{11}$$

where $P_t(x)$ is defined by equation

$$x = \int_{A_t}^{P_t(x)} q_t(P)dP. \tag{12}$$

For block-shaped LOB given in (5), we have

$$P_t(x) = A_t + x/q$$

and

$$c(x_{t_n}) = \left[A_{t_n} + x_{t_n}/(2q)\right] x_{t_n}. \tag{13}$$

The total cost is $\sum_{n=0}^{N} c(x_{t_n})$. Thus, the the optimal execution problem (3) now reduces to

$$\min_{x \in \Theta_D} \mathrm{E}_0 \left[\sum_{n=0}^{N} \left[A_{t_n} + x_{t_n}/(2q)\right] x_{t_n}\right] \tag{14}$$

under our dynamics of the limit order book given in (9) and (10).

4 Conventional Models As A Special Case

Previous work on optimal execution strategy uses a discrete-time setting with fixed time intervals and relies on a specific price-impact function to describe supply/demand (e.g., Bertsimas and Lo (1998) and Almgren and Chriss (1999, 2000)). Such a setting, however, avoids the question of optimal trading times. In this section, we briefly describe the setting used in previous work and its limitations. We then show that the conventional setting can be viewed as a special case of our framework with specific restrictions on the LOB dynamics. We further point out why these restrictions are unrealistic when the timing of trades is determined optimally.

4.1 Conventional Setup

We first consider the setup proposed by Bertsimas and Lo (1998). We adopt a simple version of their framework which captures the basic features of the models used in earlier work.

In a discrete-time setting, the trader trades at fixed time intervals, $n\tau$, where $\tau = T/N$ and $n = 0, 1, \ldots, N$ are given. Each trade will have an impact on the price, which will affect the total cost of the trade and future trades. Most models assume a linear price-impact function of the following form:

$$\bar{P}_n = \bar{P}_{n-1} + \lambda x_n + u_n = (F_n + s/2) + \lambda \sum_{i=0}^{n} x_i \tag{15}$$

where the subscript n denotes the n-th trade at $t_n = n\tau$, $\bar{P}_n$ is the average price at which trade x_n is executed with $\bar{P}_{0_-} = F_0 + s/2$, λ is the price impact coefficient and u_n is i.i.d. random variable, with a mean of zero and a variance of $\sigma^2\tau$.[10] In the second equation, we have set $F_n = F_0 + \sum_{i=0}^{n} u_i$. Clearly, λ captures the permanent price impact a trade has. The trader who has to execute an order of size X_0 solves the following problem:

$$\min_{\{x_0, x_1, \ldots, x_N\}} \mathrm{E}_0 \left[\sum_{n=0}^{N} \bar{P}_n x_n \right] = (F_0 + s/2) X_0 + \lambda \sum_{n=0}^{N} X_n (X_{n+1} - X_n). \tag{16}$$

where $\bar{P}_n$ is defined in (15) and X_n is a number of shares left to be acquired at time t_n (before trade x_{t_n}) with $X_{N+1} = 0$.

As Bertsimas and Lo (1998) show, given that the objective function is quadratic in x_n, it is optimal for the trader to split his order into small trades of equal sizes and execute them at regular intervals over the fixed period of time:

$$x_n = \frac{X_0}{N+1} \tag{17}$$

where $n = 0, 1, \ldots, N$.[11]

[10] Huberman and Stanzl (2004) have argued that in the absence of quasi-arbitrage, permanent price-impact functions must be linear.

[11] If the trader is risk averse, he will trade more aggressively at the beginning, trying to avoid the uncertainty in execution cost in later periods.

4.2 The Continuous-Time Limit

Although the discrete-time setting with a linear price impact function gives a simple and intuitive solution, it leaves a key question unanswered, namely, what determines the time-interval between trades. An intuitive way to address this question is to take the continuous-time limit of the discrete-time solution, i.e., to let N goes to infinity. However, as Huberman and Stanzl (2005) point out, the solution to the discrete-time model (16) does not have a well-defined continuous-time limit. In fact, as $N \to \infty$, the cost of the trades as given in (16) approaches the following limit:

$$(F_0+s/2)X_0 + (\lambda/2)X_0^2$$

which is strategy-independent. Thus, for a risk-neutral trader, the execution cost with continuous trading is a fixed number and any continuous strategy is as good as another. Therefore, the discrete-time model as described above does not have a well-behaved continuous-time limit.[12] For example, without increasing the cost the trader can choose to trade intensely at the very beginning and complete the whole order in an arbitrarily small period. If the trader becomes slightly risk-aversion, he will choose to finish all the trades right at the beginning, irrespective of their price impact.[13] Such a situation is clearly undesirable and economically unreasonable.

This problem has led several authors to propose different modifications to the conventional setting. He and Mamaysky (2001), for example, directly formulate the problem in continuous-time and impose fixed transaction costs to rule out any continuous trading strategies. Similar to the more general price impact function considered by Almgren and Chriss (1999, 2000), Huberman and Stanzl (2005) proposes a temporary price impact of a particular form to penalize high-intensity continuous trading. Both of these modifications limit us to a subset of feasible strategies, which is in general sub-optimal. Given its closeness to our paper, we now briefly discuss the modification with temporary price impact.

[12] In taking the continuous-time limit, we have held λ constant. This is, of course, unrealistic. For different τ, λ can well be different. But the problem remains as long as λ has a finite limit when $\tau \to 0$.

[13] As $N \to \infty$, the objective function to be minimized for a risk-averse trader with a mean-variance preference approaches the following limit

$$C\big(x_{[0,\,T]}\big) = \mathrm{E}\left[\int_0^T P_t dX_t\right] + \tfrac{1}{2}a\mathrm{Var}\left[\int_0^T P_t dX_t\right] = (F_0+s/2)X_0 + (\lambda/2)X_0^2 + \tfrac{1}{2}a\sigma^2\int_0^T X_t^2 dt$$

where $a > 0$ is the risk-aversion coefficient and σ is the price volatility of the security. The trader cares not only about the expected execution cost but also its variance, which is given by the last term. Only the variance of the execution cost depends on the strategy. It is easy to see that the optimal strategy is to choose an L-shaped profile for the trades, i.e., to trade with infinite speed at the beginning, which leads to a value of zero for the variance term in the cost function.

4.3 Temporary Price Impact

Almgren and Chriss (1999, 2000) include a temporary component in the price impact function, which can in general depend on the trading interval τ. The temporary price impact gives additional flexibility in dealing with the continuous-time limit of the problem. In particular, they specify the following dynamics for the execution prices of trades:

$$\hat{P}_n = \bar{P}_n + G(x_n/\tau) \tag{18}$$

where $\bar{P}_n$ is the same as given in (15), $\tau = T/N$ is the time between trades, and $G(\cdot)$ describes a temporary price impact, which reflects temporary price deviations from "equilibrium" caused by trading. With $G(0) = 0$ and $G'(\cdot) > 0$, the temporary price impact penalizes high trading volume per unit of time, x_n/τ. Using a linear form for $G(\cdot)$, $G(z) = \theta z$, it is easy to show that as N goes to infinity the expected execution cost approaches to

$$(F_0+s/2)X_0 + (\lambda/2)X_0^2 + \theta \int_0^T \left(\frac{dX_t}{dt}\right)^2 dt$$

(see, e.g., Grinold and Kahn (2000) and Huberman and Stanzl (2005)). Clearly, with the temporary price impact, the optimal execution strategy has a continuous-time limit. In fact, it is very similar to its discrete-time counterpart: It is deterministic and the trade intensity, defined by the limit of x_n/τ, is constant over time.[14]

The temporary price impact reflects an important aspect of the market, the difference between short-term and long-term supply/demand. If a trader speeds up his buy trades, as he can do in the continuous-time limit, he will deplete the short-term supply and increase the immediate cost for additional trades. As more time is allowed between trades, supply will gradually recover. However, as a heuristic modification, the temporary price impact does not provide an accurate and complete description of the supply/demand dynamics, which leads to several drawbacks. First, the temporary price impact function in the form considered in Almgren and Chriss (2000) and Huberman and Stanzl (2005) rules out the possibility of discrete trades. This is not only artificial but also undesirable. As we show later, in general the optimal execution strategy does involve both discrete and continuous trades. Moreover, introducing the temporary price impact does not capture the full dynamics of supply/demand.[15] Also, simply specifying a particular form for the temporary price impact function says little about the underlying economic factors that determine it.

[14] If the trader is risk-averse with a mean-variance preference, the optimal execution strategy has a decreasing trading intensity over time. See Almgren and Chriss (2000) and Huberman and Stanzl (2005).

[15] For example, two sets of trades close to each other in time versus far apart will generate different supply/demand dynamics, while in Huberman and Stanzl (2005) they lead to the same dynamics.

4.4 A Special Case of Our Framework

In the conventional setting, the supply/demand of a security is described by a price impact function at fixed times. This is inadequate when we need to determine the optimal timing of the execution strategy. We show in Section 3, using a simple limit order book framework, that the supply/demand is an intertemporal object which exhibits rich dynamics. The simple price impact function, even with the modification proposed by Almgren and Chriss (1999, 2000) and Humberman and Stanzl (2005), misses important intertemporal aspects of the supply/demand that are crucial to the determination of optimal execution strategy.

We can see the limitations of the conventional model by considering it as a special case of our general framework. Indeed, we can specify the parameters in the LOB framework so that it will be equivalent to the conventional setting. First, we set the trading times at fixed intervals: $t_n = n\tau$, $n = 0, 1, \ldots, N$. Next, we make the following assumptions on the LOB dynamics as described in (5) and (9):

$$q = 1/(2\lambda), \quad \lambda = \lambda, \quad \rho = \infty \tag{19}$$

where the second equation simply states that the price impact coefficient in the LOB framework is set to be equal to its counterpart in the conventional setting. These restrictions imply the following dynamics for the LOB. As it follows from (10), after the trade x_n at t_n ($t_n = n\tau$) the ask price A_{t_n} jumps from $V_{t_n}+s/2$ to $V_{t_n}+s/2+2\lambda x_n$. Over the next period, it comes all the way down to the new steady state level of $V_{t_n}+s/2+\lambda x_n$ (assuming no fundamental shocks from t_n to t_{n+1}). Thus, the dynamics of ask price A_{t_n} is equivalent to dynamics of $\bar{P}_{t_n}$ in (15).

For the parameters specified in (19), the cost for trade x_{t_n}, $c(x_{t_n}) = [A_{t_n}+x_{t_n}/(2q)]\, x_{t_n}$, becomes

$$c(x_{t_n}) = [F_{t_n}+s/2 + \lambda(X_0-X_{t_n}) + \lambda x_{t_n}]\, x_{t_n}$$

which is the same as the trading cost in the conventional model (16). Thus, the conventional model is a special case of LOB framework for parameters in (19).

The main restrictive assumption we have to make to obtain the conventional setup is that $\rho = \infty$ and the limit order book always converges to its steady state before the next trading time. This is not crucial if the time between trades is held fixed. But if the time between trades is allowed to shrink, this assumption becomes unrealistic. It takes time for the new limit orders to come in to fill up the book again. The shape of the limit order book after a trade depends on the flow of new orders as well as the time elapsed. As the time between trades shrinks to zero, the assumption of infinite recovery speed becomes less reasonable and it gives rise to the problems in the continuous-time limit of the conventional model.

5 Discrete-Time Solution

We now return to our general framework and solve the model for the optimal execution strategy when trading times are fixed, as in the conventional model. We then show that in contrast to the conventional setting, our framework is robust for studying convergence behavior as time between trades goes to zero. Taking the continuous-time limit we examine the resulting optimal execution strategy which turns out to include both discrete and continuous trading.

Suppose that trade times are fixed at $t_n = n\tau$, where $\tau = T/N$ and $n = 0, 1, \ldots, N$. We consider the corresponding strategies $x_{[0,\,T]} = \{x_0, x_1, \ldots, x_n\}$ within the strategy set Θ_D defined in Section 2. The optimal execution problem, defined in (3), now reduces to

$$J_0 \;=\; \min_{\{x_0,\ldots,x_N\}} \quad \mathrm{E}_0\left[\sum_{n=0}^{N} \left[A_{t_n} + x_n/(2q)\right] x_n\right] \tag{20}$$

$$\text{s.t.} \qquad A_{t_n} = F_{t_n} + \lambda(X_0 - X_{t_n}) + s/2 + \sum_{i=0}^{n-1} x_i \kappa e^{-\rho\tau(n-i)}$$

where F_t follows a random walk. This problem can be solved using dynamic programming. We have the following result:

Proposition 1 *The solution to the optimal execution problem (20) is*

$$x_n = -\tfrac{1}{2}\delta_{n+1}\left[D_{t_n}\big(1 - \beta_{n+1}e^{-\rho\tau} + 2\kappa\gamma_{n+1}e^{-2\rho\tau}\big) - X_{t_n}\big(\lambda + 2\alpha_{n+1} - \beta_{n+1}\kappa e^{-\rho\tau}\big)\right] \tag{21}$$

with $x_N = X_N$, where $D_t = A_t - V_t - s/2$. The expected cost for future trades under the optimal strategy is

$$J_{t_n} = (F_{t_n} + s/2)X_{t_n} + \lambda X_0 X_{t_n} + \alpha_n X_{t_n}^2 + \beta_n D_{t_n} X_{t_n} + \gamma_n D_{t_n}^2 \tag{22}$$

where the coefficients α_{n+1}, β_{n+1}, γ_{n+1} and δ_{n+1} are determined recursively as follows

$$\alpha_n = \alpha_{n+1} - \tfrac{1}{4}\delta_{n+1}(\lambda + 2\alpha_{n+1} - \beta_{n+1}\kappa e^{-\rho\tau})^2 \tag{23a}$$

$$\beta_n = \beta_{n+1}e^{-\rho\tau} + \tfrac{1}{2}\delta_{n+1}(1 - \beta_{n+1}e^{-\rho\tau} + 2\kappa\gamma_{n+1}e^{-2\rho\tau})(\lambda + 2\alpha_{n+1} - \beta_{n+1}\kappa e^{-\rho\tau}) \tag{23b}$$

$$\gamma_n = \gamma_{n+1}e^{-2\rho\tau} - \tfrac{1}{4}\delta_{n+1}(1 - \beta_{n+1}e^{-\rho\tau} + 2\gamma_{n+1}\kappa e^{-2\rho\tau})^2 \tag{23c}$$

with $\delta_{n+1} = \left[1/(2q) + \alpha_{n+1} - \beta_{n+1}\kappa e^{-\rho\tau} + \gamma_{n+1}\kappa^2 e^{-2\rho\tau}\right]^{-1}$ and terminal condition

$$\alpha_N = 1/(2q) - \lambda, \quad \beta_N = 1, \quad \gamma_N = 0. \tag{24}$$

Proposition 1 gives the optimal execution strategy when we fix the trade times at a certain interval τ. But it is only optimal among strategies with the same fixed trading interval. In principle, we want to choose the trading interval to minimize the execution cost.

One way to allow different trading intervals is to take the limit $\tau \to 0$, i.e., $N \to \infty$, in the problem (20). Figure 2 plots the optimal execution strategy $\{x_n,\ n = 0, 1, \ldots, N\}$ for $N = 10, 25, 100$, respectively. Clearly, it is very different from the strategy given in (17) and obtained previously when the dynamics of demand/supply is ignored. Moreover, as N becomes large, the strategy splits into two parts, large trades at both ends of the horizon (the beginning and the end) and small trades in between.

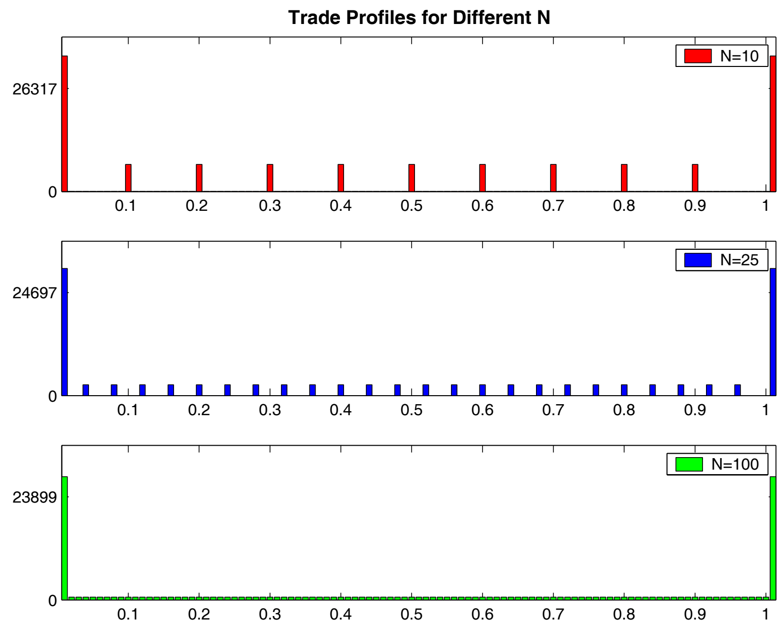

Figure 2: Optimal execution strategy with fixed discrete trading intervals. This figure plots the optimal trades for N fixed intervals, where N is 10, 25 and 100 for respectively the top, middle and bottom panels. The initial order to trade is set at $X_0 = 100,000$ units, the time horizon is set at $T = 1$ day, the market depth is set at $q = 5,000$ units, the price-impact coefficient is set at $\lambda = 1/(2q) = 10^{-4}$ and the resiliency coefficient is set at $\rho = 2.231$.

The next proposition describes the continuous-time limit of the optimal execution strategy and the expected cost:

Proposition 2 *In the limit of $N \to \infty$, the optimal execution strategy becomes*

$$\lim_{N\to\infty} x_0 = x_{t=0} = \frac{X_0}{\rho T + 2} \tag{25a}$$

$$\lim_{N\to\infty} x_n/(T/N) = \dot{X}_t = \frac{\rho X_0}{\rho T + 2}, \quad t \in (0,\ T) \tag{25b}$$

$$\lim_{N\to\infty} x_N = x_{t=T} = \frac{X_0}{\rho T + 2} \tag{25c}$$

and the expected cost is

$$J_t = (F_0 + s/2)X_t + \lambda X_0 X_t + \alpha_t X_t^2 + \beta_t X_t D_t + \gamma_t D_t^2$$

where coefficients α_t, β_t, γ_t are given by

$$\alpha_t = \frac{\kappa}{\rho(T-t)+2} - \frac{\lambda}{2}, \quad \beta_t = \frac{2}{\rho(T-t)+2}, \quad \gamma_t = -\frac{\rho(T-t)}{2\kappa[\rho(T-t)+2]}. \tag{26}$$

The optimal execution strategy given in Proposition 2 is different from those obtained in the conventional setting. In fact, it involves both discrete and continuous trades. This clearly indicates that the timing of trades is a critical part of the optimal strategy. It also shows that ruling out discrete or continuous trades ex ante is in general suboptimal. More importantly, it demonstrates that both the static and dynamic properties of supply/demand, which are captured by the LOB dynamics in our framework, are important in analyzing the optimal execution strategy. We return in Section 7 to examine in more detail the properties of the optimal execution strategy and their dependence on the LOB dynamics.

6 Continuous-Time Solution

The nature of the continuous-time limit of the discrete-time solution suggests that limiting ourselves to discrete strategies can be suboptimal. We should in general formulate the problem in continuous-time setting and allow both continuous and discrete trading strategies. In this section, we present the continuous-time version of the LOB framework and derive the optimal strategy.

The uncertainty in model is fully captured by fundamental value F_t. Let $F_t = F_0 + \sigma Z_t$ where Z_t is a standard Brownian motion defined on $[0, T]$. $\mathcal{F}_t$ denotes the filtration generated by Z_t. A general execution strategy can consist of two components, a set of discrete trades at certain times and a flow of continuous trades. A set of discrete trades is also called an "impulse" trading policy.

Definition 1 *Let $N_+ = \{1, 2, \ldots\}$. An impulse trading policy $(\tau_k, x_k) : k \in N_+$ is a sequence of trading times τ_k and trade amounts x_k such that: (1) $0 \le \tau_k \le \tau_{k+1}$ for $k \in N_+$, (2) τ_k is a stopping time with respect to $\mathcal{F}_t$, and (3) x_k is measurable with respect to $\mathcal{F}_{t_k}$.*

The continuous trades can be defined by a continuous trading policy described by the intensity of trades $\mu_{[0,\,t]}$, where μ_t is measurable with respect to $\mathcal{F}_t$ and $\mu_t dt$ gives the trades during time interval $[t,\, t+dt)$. Let us denote $\hat{T}$ the set of impulse trading times. Then, the set of admissible execution strategies for a buy order is

$$\Theta_C = \left\{ \mu_{[0,\,T]}, x_{\{t\in\hat{T}\}} :\ \mu_t,\ x_t \ge 0,\ \int_0^T \mu_t dt + \sum_{t\in\hat{T}} x_t = X_0 \right\} \tag{27}$$

where μ_t is the rate of continuous buy trades at time t and x_t is the discrete buy trade for $t \in \hat{T}$. The dynamics of X_t, the number of shares to acquire at time t, is then given by the following equation:

$$X_t = X_0 - \int_0^t \mu_s ds - \sum_{s \in \hat{T},\ s<t} x_s.$$

Now let us specify the dynamics of ask price A_t. Similar to the discrete-time setting, we have $A_0 = F_0 + s/2$ and

$$A_t = A_0 + \int_0^t \left[\, dV_s - \rho D_s ds - \kappa dX_s \,\right] \tag{28}$$

where $V_t = F_t + \lambda(X_0 - X_t)$ as in (9) and $D_t = A_t - V_t - s/2$ as in (7). The dynamics of A_t captures the evolution of the limit order book, in particular the changes in V_t, the inflow of new orders and the continuous execution of trades.

Next, we compute the execution cost, which consists of two parts: the costs from continuous trades and discrete trades, respectively. The execution cost from t to T is

$$C_t = \int_t^T A_s \mu_s ds + \sum_{s \in \hat{T},\ t \le s \le T} \left[A_s + x_s/(2q)\right] x_s. \tag{29}$$

Given the dynamics of the state variables in (9), (28), and cost function in (29), the optimal execution problem now becomes

$$J_t \equiv J(X_t, A_t, V_t, t) = \min_{\{\mu_{[0,\,T]},\, \{x_{t \in \hat{T}}\}\} \in \Theta_C} \mathrm{E}_t\left[C_t\right] \tag{30}$$

where J_t is the value function at t, the expected cost for future trades under the optimal execution strategy. At time T, the trader is forced to buy all of the remaining order X_T, which leads to the following boundary condition:

$$J_T = \left[A_T + 1/(2q) X_T\right] X_T.$$

The next proposition gives the solution to the problem:

Proposition 3 *The value function for the optimization problem (30) is*

$$J_t = (F_t + s/2) X_t + \lambda X_0 X_t + \alpha_t X_t^2 + \beta_t D_t + \gamma_t D_t^2$$

where $D_t = A_t - V_t - s/2$. The optimal execution strategy is

$$x_0 = x_T = \frac{X_0}{\rho T + 2}, \quad \mu_t = \frac{\rho X_0}{\rho T + 2} \quad \forall\ t \in (0,\ T) \tag{31}$$

where the coefficients α_t, β_t, and γ_t are the same as given in Proposition 2.

Obviously, the solution we obtained with the continuous-time setting is identical to the continuous-time limit of the solution in the discrete-time setting. The optimal strategy consists of both continuous and discrete trades.

7 Optimal Execution Strategy and Cost

In contrast with previous work, the optimal execution strategy includes discrete and continuous trading. We now analyze the properties of the optimal execution strategy in more detail. Interestingly, while it does not depend on parameters λ and q, which determine static supply/demand, it crucially depends on parameter ρ, which describes the LOB dynamics, and the horizon for execution T. Further in this section we quantify the cost reduction which the optimal execution strategy brings and discuss its empirical implications.

7.1 Properties of Optimal Execution Strategy

The first thing to notice is that the execution strategy does not depend on λ and q. Coefficient λ captures the permanent price impact of a trade. Given the linear form, the permanent price impact gives an execution cost of $(F_0+s/2)X_0 + (\lambda/2)X_0^2$, which is independent of the execution strategies. This is a rather striking result given that most of the previous work focus on λ as the key parameter determining the execution strategy and cost. As we show earlier, λ affects the execution strategy when the times to trade are exogenously set at fixed intervals. When the times to trade are determined optimally, the impact of λ on execution strategy disappears. Given the linear form of the price impact function, λ fully describes the instantaneous supply/demand, or the static supply/demand. Our analysis clearly shows that the static aspects of the supply/demand does not fully capture the factors that determining the optimal execution strategy.

Coefficient q captures the depth of the market. In the simple model for the limit order book we have assumed, market depth is constant at all price levels above the ask price. In this case, the actual value of the market depth does not affect the optimal execution strategy. For more general (and possibly more realistic) shapes of the limit order book, the optimal execution strategy may well depend on the characteristics of the book.

The optimal execution strategy depends on two parameters, the resilience of the limit order book ρ and the horizon for execution T. We consider these dependencies separately.

Panel (a) of Figure 3 plots the optimal execution strategy, or more precisely the time path of the remaining order to be executed. Clearly, the nature of the optimal strategy is different from those proposed in the literature, which involve a smooth flow of small trades. When the timing of trades is determined optimally, the optimal execution strategy consists of both large discrete trades and continuous trades. In particular, under the LOB dynamics

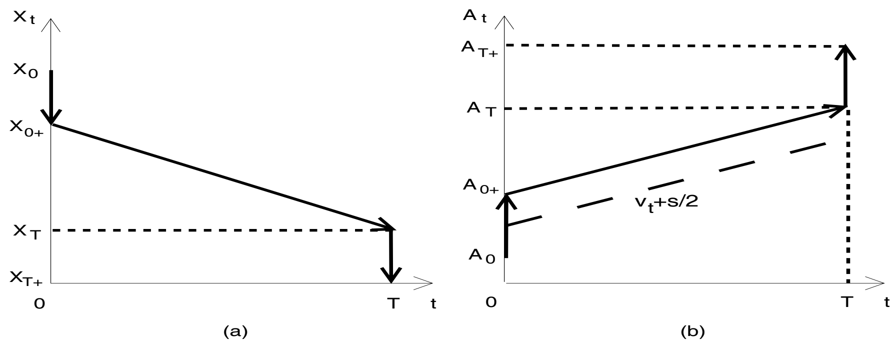

Figure 3: Profiles of the optimal execution strategy and ask price. Panel (a) plots the profile of optimal execution policy as described by X_t. Panel (b) plots the profile of realized ask price A_t. After the initial discrete trade, continuous trades are executed as a constant fraction of newly incoming sell orders to keep the deviation of the ask price A_t from its steady state $V_t+s/2$, shown with grey line in panel (b), at a constant. A discrete trade occurs at the last moment T to complete the order.

we consider here, the optimal execution involves a discrete trade at the beginning, followed by a flow of small trades and then a discrete terminal trade. Such a strategy seems intuitive given the dynamics of the limit order book. The large initial trade pushes the limit order book away from its stationary state so that new orders are lured in. The flow of small trades will "eat up" these new orders and thus keep them coming. At the end, a discrete trade finishes the remaining part of the order. The final discrete trade is determined by two factors. First, the order has to be completed within the given horizon. Second, the evolution of supply/demand afterwards no longer matters. In practice, both of these two factors can take different forms. For example, the trading horizon T can be endogenously determined rather than exogenously given. We consider this extension in Section 8.

The size of the initial trade determines the prices and the intensity of the new orders. If too large, the initial trade will raise the average prices of the new orders. If too small, an initial trade will not lure in enough orders before the terminal time. The trade off between these two factors largely determines the size of the initial trade.

The continuous trades after the initial trade are intended to maintain the flow of new orders at desirable prices. To see how this works, let us consider the path of the ask price A_t under the optimal execution strategy. It is plotted in panel (b) of Figure 3. The initial discrete trade pushes up the ask price from $A_0 = V_0+s/2$ to $A_{0+} = V_0+s/2+X_0/(\rho T+2)/q$. Afterwards, the optimal execution strategy keeps $D_t = A_t-V_t-s/2$, the deviation of the current ask price A_t from its steady state $V_t+s/2$, at a constant level of $\kappa X_0/(\rho T+2)$. Consequently, the rate of new sell order flow, which is given by ρD_t, is also maintained at a constant level. The ask price A_t goes up together with $V_t+s/2$, the steady-state "value" of the security, which is shown with the grey line in Figure 3(b). As a result, from (28) with $dA_t = dV_t$ for $0 < t < T$, we have $\rho D_t = \kappa\mu_t$ or $\mu_t = (1/\kappa)\rho D_t$. In other words, under the optimal execution strategy a constant fraction of $1/\kappa$ of the new sell orders is executed to maintain a constant order flow.

Our discussion above shows that the dynamics of the limit order book, which is captured by the resilience parameter ρ, is the key factor in determining optimal execution strategy. In order to better understand this link, let us consider two extreme cases, when $\rho = 0$ and ∞. When $\rho = 0$, we have no recovery of the limit order book after a trade. In this case, the cost of execution will be strategy independent and it does not matter when and at what speed the trader eats up the limit order book. This result is also true in a discrete setting with any N and in its continuous-time limit. When $\rho = \infty$, the limit order book rebuilds itself immediately after a trade. As we discussed in Section 4, this corresponds to the conventional setting. Again, the execution cost becomes strategy independent. It should be pointed out that even though in the limit of $\rho \to 0$ or ∞, the optimal execution strategy given in Proposition 3 converges to a pure discrete strategy or a pure continuous strategy, other strategies are equally good given the degeneracy in these two cases.

When $0 < \rho < \infty$, the resiliency of the limit order book is finite, the optimal strategy is a mixture of discrete and continuous trades. The fraction of the total order executed through continuous trades is $\int_0^T \mu_t dt/X_0 = \rho T/(\rho T+2)$, which increases with ρ. In other words, it is more efficient to use small trades when the limit order book is more resilient. This is intuitive because discrete trades do less in taking full advantage of new order flows than continuous trades.

Another important parameter in determining the optimal execution strategy is the time-horizon to complete the order T. From Proposition 3, we see that as T increases, the size of the two discrete trades decreases. This result is intuitive. The more time we have to execute the order, the more we can continuous trades to benefit from the inflow of new orders and to lower the total cost.

7.2 Minimum Execution Cost

So far, we have focused on the optimal execution strategy. We now examine how important the optimal execution is, as measured by the execution cost it saves. For this purpose, we use the strategy obtained in the conventional setting and its cost as the benchmark. The total expected execution cost of a buy order of size X_0 is equal to its fundamental value $(F_0+s/2)X_0$, which is independent of the execution strategy, plus the extra cost from the price impact of trading, which does depend on the execution strategy. Thus, we will only consider the execution cost, net of the fundamental value, or the net execution cost.

As shown in Section 4, the strategy from the conventional setting is a constant flow of trades with intensity $\mu_\infty = X_0/T$, $t \in [0, T]$. Under this simple strategy, we have $V_t = F_t + \lambda(t/T)X_0$, $D_t = [\kappa X_0/(\rho T)](1 - e^{-\rho t})$ and $A_t = V_t + D_t + s/2$. The expected net execution cost for the strategy with constant rate of execution μ_∞ is given by

$$\tilde{J}_0^{\mathrm{CM}} = \mathrm{E}_0\left[\int_0^T (A_t - F_t - s/2)(X_0/T)dt\right] = (\lambda/2)X_0^2 + \kappa\frac{\rho T - (1 - e^{-\rho T})}{(\rho T)^2}X_0^2$$

where the superscript stands for the "Conventional Model". From Proposition 3, the expected net cost under the optimal execution strategy is

$$\tilde{J}_0 = J_0 - (F_0 + s/2)X_0 = (\lambda/2)X_0^2 + \frac{\kappa}{\rho T + 2}X_0^2$$

(note that at $t = 0$, $D_0 = 0$). Thus, the improvement in expected execution cost by the optimal strategy is $J_0^{\mathrm{CM}} - J_0$, which is given by

$$\tilde{J}_0^{\mathrm{CM}} - \tilde{J}_0 = \kappa\frac{2\rho T - (\rho T + 2)(1 - e^{-\rho T})}{(\rho T + 2)(\rho T)^2}X_0^2$$

and is always non-negative. The relative gain can be defined as $\Delta = (\tilde{J}^{\mathrm{CM}} - \tilde{J}_0)/\tilde{J}^{\mathrm{CM}}$.

In order to calibrate the magnitude of the cost reduction by the optimal execution strategy, we consider some numerical examples. Let the size of the order to be executed be $X_0 = 100,000$ shares and the initial security price be $A_0 = F_0 + s/2 = \$100$. We choose the width of the limit order book, which gives the depth of the market, to be $q = 5,000$. This implies that if the order is executed at once, the ask price will move up by 20%. Without losing generality, we consider the execution horizon to be one day, $T = 1$.[16] The other parameters, especially ρ, may well depend on the security under consideration. In absence of an empirical calibration, we with consider a range of values for them.

Table 1 reports the numerical values of the optimal execution strategy for different values of ρ. As discussed above, for small values of ρ, most of the order is executed through two discrete trades, while for large values of ρ, most of the order is executed through a flow of continuous trades as in the conventional models. For intermediate ranges of ρ, a mixture of discrete and continuous trades is used.

Table 2 reports the relative improvement in the expected net execution cost by the optimal execution strategy over the simple strategy of the conventional setting. Let us first consider the extreme case in which the resilience of the LOB is very small, e.g., $\rho = 0.001$

[16] Chan and Lackonishok (1995) documented that for institutional trades T is usually between 1 to 4 days. Keim and Madhavan (1995) found that the duration of trading is surprisingly short, with almost 57% of buy and sell orders completed in the first day. Keim and Madhavan (1997) reported that average execution time is 1.8 days for a buy order and 1.65 days for a sell order.

ρ	Half-life ($\log 2/\rho$)	Trade x_0	Trade over $(0,T)$	Trade x_N
0.001	693.15 day	49,975	50	49,975
0.01	69.31 day	49,751	498	49,751
0.5	1.39 day	40,000	20,000	40,000
1	270.33 min	33,333	33,334	33,333
2	135.16 min	25,000	50,000	25,000
4	67.58 min	16,667	66,666	16,667
5	54.07 min	14,286	71,428	14,286
10	27.03 min	8,333	83,334	8,333
20	13.52 min	4,545	90,910	4,545
50	5.40 min	1,921	96,153	1,921
300	0.90 min	331	99,338	331
1000	0.20 min	100	99,800	100
10000	0.03 min	10	99,980	10

Table 1: Profiles of the optimal execution strategy for different levels of LOB resiliency ρ. The table reports values of optimal discrete trades x_0 and x_T at the beginning and the end of the trading horizon and the intensity of continuous trades in between for an order of $X_0 = 100,000$ for different values of the LOB resilience parameter ρ or the half-life of an LOB disturbance $\tau_{1/2}$, which is defined as $\exp\{-\rho\,\tau_{1/2}\} = 1/2$. The initial ask price is \$100, the market depth is set at $q = 5,000$ units, the (permanent) price-impact coefficient is set at $\lambda = 1/(2q) = 10^{-4}$, and the trading horizon is set at $T = 1$ day, which is 6.5 hours (390 minutes).

		λ				
ρ	Half-life	$\frac{1}{2q}$	$\frac{1}{10q}$	$\frac{1}{50q}$	$\frac{1}{100q}$	0
0.001	693.15 day	0.00	0.01	0.02	0.02	0.02
0.01	69.31 day	0.08	0.15	0.16	0.16	0.17
0.5	1.39 day	2.82	5.42	5.99	6.06	6.13
1	270.33 min	3.98	8.16	9.14	9.26	9.39
2	135.16 min	4.32	9.97	11.51	11.71	11.92
4	67.58 min	3.19	9.00	11.05	11.35	11.65
5	54.07 min	2.64	8.07	10.21	10.53	10.86
10	27.03 min	1.13	4.58	6.65	7.01	7.41
20	13.52 min	0.37	1.98	3.54	3.89	4.31
50	5.40 min	0.07	0.49	1.24	1.50	1.88
300	0.90 min	0.00	0.02	0.08	0.13	0.33
1000	0.20 min	0.00	0.00	0.01	0.02	0.10
10000	0.03 min	0.00	0.00	0.00	0.00	0.09

Table 2: Cost savings by the optimal execution strategy from the simple trading strategy. Relative improvement in expected net execution cost $\Delta = (\tilde{J}^{\mathrm{CM}} - \tilde{J}_0)/\tilde{J}^{\mathrm{CM}}$ is reported for different values of LOB resiliency coefficient ρ and the permanent price-impact coefficient. The order size is set at 100,000, the market depth is set at $q = 5,000$ and the horizon for execution is set at $T = 1$ day (equivalent of 390 minutes).

and the half-life for the LOB to rebuild itself after being hit by a trade is 693.15 days. In this case, even though the optimal execution strategy looks very different from the simple execution strategy, as shown in Figure 4, the improvement in execution cost is minuscule. This is not surprising as we know the execution cost becomes strategy independent when $\rho = 0$. For a modest value of ρ, e.g. $\rho = 2$ with a half life of 135 minutes (2 hours and 15 minutes), the improvement in execution cost ranges from 4.32% for $\lambda = 1/(2q)$ to 11.92% for $\lambda = 0$. When ρ becomes large and the LOB becomes very resilient, e.g., $\rho = 300$ and the half-life of LOB deviation is 0.90 minute, the improvement in execution cost becomes small again, with a maximum of 0.33% when $\lambda = 0$. This is again expected as we know that the simple strategy is close to the optimal strategy when $\rho \to \infty$ (as in this limit, the cost becomes strategy independent).

In order to see the difference between the optimal strategy and the simple strategy obtained in conventional settings, we compare them in Figure 4. The solid line shows the optimal execution strategy of the LOB framework and the dashed line shows the execution strategy of the conventional setting. Obviously, the difference between the two strategies are more significant for smaller values of ρ.

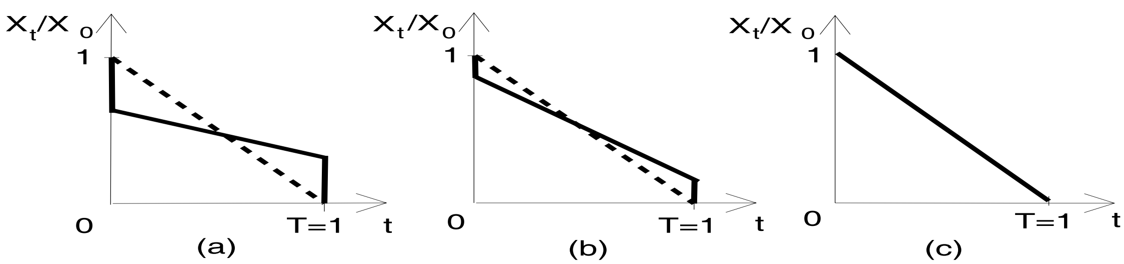

Figure 4: Optimal execution strategy versus simple execution strategy from the conventional models. The figure plots the time paths of remaining order to be executed for the optimal strategy (solid line) and the simple strategy obtained from the conventional models (dashed line), respectively. The order size is set at $X_0 = 100,000$, the initial ask price is set at \$100, the market depth is set at $q = 5,000$ units, the (permanent) price-impact coefficient is set at $\lambda = 1/(2q) = 10^{-4}$, and the trading horizon is set at $T = 1$ day, which is assumed to be 6.5 hours (390 minutes). Panels (a), (b) and (c) plot the strategies for $\rho = 0.001, 2$ and $1,000$, respectively.

Table 2 also reveals an interesting result. The relative savings in execution cost by the optimal execution strategy is the highest when $\lambda = 0$, i.e., when the permanent price impact is zero.[17]

[17] Of course, the magnitude of net execution cost becomes very small as λ goes to zero.

7.3 Empirical Implications

Optimality of discrete trades at the beginning and the end of the trading period leads to interesting empirical implications. It is well documented that there is a U-shaped pattern in the intraday trading volume, price volatility and average bid-ask spread.[18] Several authors have proposed theoretical models that can help to explain the intraday price and volume patterns.[19] Most of these models generate the intraday patterns from the time variation in information asymmetry and/or trading opportunities associated with market closures.

Our model suggests an alternative source for such patterns. Namely, they can be generated by the optimal execution of block trades. It is well known that large-block transactions have become a substantial fraction of the total trading volume for common stocks. According to Keim and Madhaven(1996), block trades represented almost 54% of New York Stock Exchange share volume in 1993 while in 1965 the corresponding figure was merely 3%. Thus, the execution strategies of institutional traders can influence the intraday variation in volume and prices. It is often the case that institutional investors have daily horizons to complete their orders, for example to accommodate the inflows and outflows in mutual funds. For reasonable values of the LOB recovery speed ρ, our optimal execution strategy implies large trades at the beginning and at the end of trading period. If execution horizon of institutional traders coincides with a trading day, their trading can cause the increase in trading volume and bid-ask spread at the beginning and the end of a trading day.

Our model predicts higher variation in the optimal trading profile for stocks with lower ρ. This implies that stocks with low resilience in its LOB (low ρ) and high institutional holdings should exhibit more intraday volume variation. We leave the empirical tests of these predictions for future research.

8 Extensions

So far, we have used a parsimonious LOB model to analyze the impact of supply/dynamics on optimal execution strategy. Obviously, the simple characteristics of the model does not reflect the richness in the LOB dynamics observed in the market. However, the framework we developed is quite flexible to allow for extensions in various directions. In this section, we

[18]Intraday patterns in volume and prices in the U.S. markets have been documented by Jain and Joh(1988), Gerety and Mulherin (1992), Chan, Christie, Schultz (1995), among others. They are also present in other markets. See McInish and Wood(1991) for the Toronto Stock Exchange, Hamao and Hasbrouck (1995) for the Tokyo Stock Exchange, Niemayer and Sandas (1993) for the Stockholm Stock Exchange, and Kleidon and Werner (1996) for the London Stock Exchange.

[19]See, for example, Admati and Pfleiderer (1988), Back and Baruch (2004), Brock and Kleidon (1992), Foster and Viswanathan(1990, 1995), and Hong and Wang (2000).

briefly discuss some of these extensions. First, we consider the case where the resilience of the LOB is time-varying. Next, we discuss the possibility of allowing more general shapes of the static limit order book. Finally, we include risk considerations in optimization problem. This also allows us to endogenize the trading horizon T, which is taken as given above.

8.1 Time Varying LOB Resiliency

Our model can easily incorporate time-variation in LOB resiliency. It has been documented that trading volume, order flows and transaction costs all exhibit a U-shaped intraday pattern, high at the opening of the trading day, then falling to lower constant levels during the day and finally rising again towards the close of trading day. This suggests that the liquidity in the market may well vary over a trading day. Monch (2004) has attempted to incorporate such a time-variation in implementing the conventional models.

We can easily allow time-variation in LOB and its dynamics in our model. In particular, we can allow the resilience coefficient to be time dependent, $\rho = \rho_t$ for $t \in [0, T]$. The results in Proposition 1, 2, 3 still hold if we replace ρ by ρ_t, ρT by $\int_0^T \rho_t dt$ and $\rho(T-t)$ by $\int_t^T \rho_t dt$.

8.2 Different Shapes for LOB

We have considered a simple shape for the LOB, which is a step function. As we showed in Section 3, this form of the LOB is consistent with the static linear price impact function widely used in the literature. Huberman and Stanzl (2005) have provided theoretical arguments in support of the linear price impact functions. However, empirical literature has suggested that the shape of the LOB can be more complex (see, e.g., Hopman (2003)). Addressing this issue, we can allow more general shapes of the LOB in our framework. For example, we may extend our analysis to LOB with a density of placed limit orders defined by power function. This will also make the LOB dynamics more complex. As a trade eats away the tip of the LOB, we have to specify how the LOB converges to its steady state. With a complicated shape for the LOB, this convergence process can take many forms which involves assumptions about the flow or new orders at a range of prices. For certain specifications of this convergence process, our model is still tractable. For brevity, we do not present these cases here. But beyond certain point, closed form solutions become hard to find. Although the actual strategy can be quite complex and depends on the specifics of the LOB shape and its dynamics, we expect its qualitative features to be the same as that under the simple LOB dynamics we considered.

8.3 Risk Aversion

Let us consider the optimal execution problem for a risk-averse trader. For tractability, we assume that he has a mean-variance objective function with a risk-aversion coefficient of a. The optimization problem (30) now becomes

$$J_t \equiv J(X_t, A_t, V_t, t) = \min_{\{\mu_{[0,\,T]},\,\{x_{t\in\hat{T}}\}\}\in\Theta_C} \mathrm{E}_t\left[C_t\right] + \tfrac{1}{2}\, a\, \mathrm{Var}_t\left[C_t\right] \tag{32}$$

with (9), (28), (29) and the same terminal condition $J_T = \left[A_T + 1/(2q)X_T\right] X_T$. Since the only source of uncertainty is F_t and only the trades executed in interval $[t,\ t+dt)$ will be subject to this uncertainty, we can rewrite (32) in a more convenient form:

$$J_t = \min_{\{\mu_{[0,\,T]},\,\{x_{t\in\hat{T}}\}\}\in\Theta_C} \mathrm{E}_t\left[C_t\right] + \tfrac{1}{2}\, a \int_t^T \sigma^2 X_s^2 ds. \tag{33}$$

At time T, the trader is forced to buy all of the remaining order X_T. This leads to the following boundary condition:

$$J_T = \left[A_T + 1/(2q)X_T\right] X_T.$$

The next proposition gives the solution to the problem for a risk averse trader:

Proposition 4 *The solution to the optimization problem (33) is*

$$\begin{aligned}
x_0 &= X_0 \frac{\kappa f'(0) + a\sigma^2}{\kappa\rho f(0) + a\sigma^2} \\
\mu_t &= \kappa x_0 \frac{\rho g(t) - g'(t)}{1 + \kappa g(t)} e^{-\int_0^t \frac{\kappa g'(s)+\rho}{1+\kappa g(s)} ds}, \quad \forall\ t \in (0,\ T) \\
x_T &= X_0 - x_0 - \int_0^T \mu_s ds
\end{aligned}$$

and the value function is

$$J_t = (F_t + s/2)X_t + \lambda X_0 X_t + \alpha_t X_t^2 + \beta_t D_t + \gamma_t D_t^2$$

where $D_t = A_t - V_t - s/2$ and the coefficients are given by

$$\alpha_t = \frac{\kappa f(t) - \lambda}{2}, \quad \beta_t = f(t), \quad \gamma_t = \frac{f(t) - 1}{2\kappa}$$

and

$$\begin{aligned}
f(t) &= (v - a\sigma^2)/(\kappa\rho) + \left[-\frac{\kappa\rho}{2v} + e^{\frac{2\rho v}{2\kappa\rho + a\sigma^2}(T-t)}\left(\frac{\kappa\rho}{2v} - \frac{\kappa\rho}{v - a\sigma^2 - \kappa\rho}\right)\right]^{-1} \\
g(t) &= -\frac{f'(t) - \rho f(t)}{k f'(t) + a\sigma^2}
\end{aligned}$$

with $v = \sqrt{a^2\sigma^4 + 2a\sigma^2\kappa\rho}$.

It can be shown that as risk aversion coefficient goes to 0 the coefficients α_t, β_t, and γ_t converge to the ones given in Proposition 2, which presents the results for the risk neutral trader.

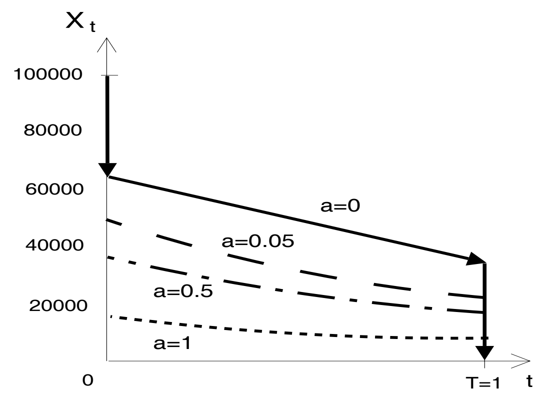

Figure 5: Profiles of the optimal execution strategies for different coefficients of risk aversion. This figure shows the profiles of optimal execution policies X_t for the traders with different coefficients of risk aversion $a = 0$ (solid line), $a = 0.05$ (dashed line), $a = 0.5$ (dashed-dotted line) and $a = 1$ (dotted line), respectively. Variable X_t indicates how much shares still has to be executed before trading at time t. The order size is set at $X_0 = 100,000$, the market depth is set at $q = 5,000$ units, the permanent price-impact coefficient is set at $\lambda = 0$, and the trading horizon is set at $T = 1$, the resiliency coefficient is set at $\rho = 1$.

The nature of the optimal strategy remains qualitatively the same under risk aversion: discrete trades at the two ends of the trading horizon with continuous trades in the middle. The effect of trader's risk aversion on his optimal trading profile is shown in Figure 5. The more risk averse is the trader, the larger the initial trade more trades he shifts to the beginning.

So far, we have assumed the execution horizon, $[0,\ T]$, to be exogenously given, and ignored any time preference for execution a trade. Risk aversion, however, introduces a natural preference for such a preference: Trading sooner reduces uncertainty in execution prices. Such a preference is clearly reflected in the optimal policy as shown in Figure 5. Such a time preference provides a mechanism to endogenize the execution horizon. For example, T is sufficiently large, when the trader is risk-averse enough, he may optimally finish the whole order soon before T.

9 Conclusion

In this paper, we analyze the optimal trading strategy to execute a large order. We show that the static price impact function widely used in previous work fails to capture the intertemporal nature of a security's supply/demand in the market. We construct a simple dynamic model for a limit order book market to capture the intertemporal nature of supply/demand and solve for the optimal execution strategy. We show that when trading times are chosen optimally, the dynamics of the supply/demand is the key factor in determining the optimal execution strategy. Contrary to previous work, the optimal execution strategy involves discrete trades as well as continuous trades, instead of merely continuous trades. This trading behavior is consistent with the empirical intraday volume and price patterns. Our results on the optimal execution strategy also suggest testable implications for these intraday patterns and provide new insight into the demand of liquidity in the market.

The specific model we used for the LOB dynamics is very simple since our goal is mainly to illustrate its importance. The actual LOB dynamics can be much more complex. However, the framework we developed is fairly general to accommodate rich forms of LOB dynamics. Moreover, with the current increase in the number of open electronic limit order books, our LOB model can be easily calibrated and used to address real world problems.

Appendix

Proof of Proposition 1

From (7), we have

$$D_{t_n} = A_{t_n} - V_{t_n} - s/2 = \sum_{i=0}^{n-1} x_{t_i} \kappa e^{-\rho\tau(n-i)} \tag{A.1}$$

From (A.1), the dynamics of D_t between trades will be

$$D_{t_{n+1}} = \left(D_{t_n} + x_{t_n}\kappa\right) e^{-\rho\tau} \tag{A.2}$$

with $D_0 = 0$. We can then re-express the optimal execution problem (20) in terms of variables X_t and D_t:

$$\min_{x \in \Theta_D} \mathrm{E}_0 \sum_{n=0}^{N} \left[(F_{t_n} + s/2) + \lambda(X_0 - X_{t_n}) + D_{t_n} + x_{t_n}/(2q)\right] x_{t_n}. \tag{A.3}$$

under dynamics of D_t given by (A.2).

First, by induction we prove that value function for (A.3) is quadratic in X_t and D_t and has a form implied by (22):

$$J(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = (F_{t_n} + s/2)X_{t_n} + \lambda X_0 X_{t_n} + \alpha_n X_{t_n}^2 + \beta_n X_{t_n} D_{t_n} + \gamma_n D_{t_n}^2. \tag{A.4}$$

At time $t = t_N = T$, the trader has to finish the order and the cost is

$$J(X_T, D_T, F_T, T) = (F_T + s/2)X_T + [\lambda(X_0 - X_T) + D_T + X_T/(2q)]X_T.$$

Hence, $\alpha_N = 1/(2q) - \lambda$, $\beta_N = 1$, $\gamma_N = 0$. Recursively, the Bellman equation yields

$$\begin{aligned} J_{t_{n-1}} &= \min_{x_{n-1}} \Big\{ \left[(F_{t_{n-1}} + s/2) + \lambda(X_0 - X_{t_{n-1}}) + D_{t_{n-1}} + x_{n-1}/(2q)\right] x_{n-1} \\ &\qquad + \mathrm{E}_{t_{n-1}} J\left[X_{t_{n-1}} - x_{n-1}, (D_{t_{n-1}} + \kappa x_{n-1})e^{-\rho\tau}, F_{t_n}, t_n\right] \Big\}. \end{aligned}$$

Since F_{t_n} follows Brownian motion and value function is linear in F_{t_n}, it immediately follows that the optimal x_{n-1} is a linear function of $X_{t_{n-1}}$ and $D_{t_{n-1}}$ and the value function is a quadratic in $X_{t_{n-1}}$ and $D_{t_{n-1}}$ satisfying (A.4), which leads to the recursive equation (23) for the coefficients. Q.E.D.

Proof of Proposition 2

First, we prove the convergence of the value function. As $\tau = T/N \to 0$, the first order approximation of the system (23) in τ leads to the following restrictions on the coefficients:

$$\begin{aligned} &\lambda + 2\alpha_t - \beta_t\kappa = 0 \\ &1 - \beta_t + 2\kappa\gamma_t = 0 \end{aligned} \tag{A.5}$$

and

$$\begin{aligned}
\dot{\alpha}_t &= \tfrac{1}{4}\kappa\rho\beta_t^2 \\
\dot{\beta}_t &= \rho\beta_t - \tfrac{1}{2}\rho\beta_t(\beta_t - 4\kappa\gamma_t) \\
\dot{\gamma}_t &= 2\rho\gamma_t + \tfrac{1}{4\kappa}\rho(\beta_t - 4\kappa\gamma_t)^2.
\end{aligned} \tag{A.6}$$

It is easy to verify that α_t, β_t and γ_t given in (26) are the solution of (A.6), satisfying (A.5) and the terminal condition (24). Thus, as $\tau \to 0$ the coefficients of the value function (23) converge to (26).

Next, we prove the convergence result for the optimal execution policy $\{x_t\}$. Substituting α_t, β_t, γ_t into (21), we can show that as $\tau \to 0$, the execution policy converges to

$$x_t = \left\{ X_t \frac{1}{\rho(T-t)+2} - D_t \frac{1+\rho(T-t)}{\kappa[\rho(T-t)+2]} \right\} \left[1 - \tfrac{1}{2}\rho^2(T-t)\tau\right] + \tfrac{1}{2}(\rho/\kappa)D_t\tau + o(\tau) \tag{A.7}$$

where $o(\tau)$ denotes terms to the higher order of τ. At $t = 0$, $D_0 = 0$ and we have $\lim_{\tau\to 0} x_0 = \frac{X_0}{\rho T+2}$. Moreover, after the initial discrete trade x_0 all trades will be the continuous (except possibly at T) and equal to

$$x_t = \tfrac{1}{\kappa}\rho D_t\tau + o(\tau), \qquad t = n\tau, \qquad n = 1, \ldots, N-1. \tag{A.8}$$

We prove this by induction. First, using (A.7), where $X_\tau = X_0 - x_0$ and $D_\tau = kx_0(1-\rho\tau)$, it is easy to check that (A.8) holds for x_τ. Second, let us assume that (A.8) holds for some x_t, where $t = n\tau$, then we can show that $x_{t+\tau}$ will satisfy it as well. In fact, the dynamics of X_t and D_t is defined by

$$X_{t+\tau} = X_t - x_t, \qquad D_{t+\tau} = (D_t + kx_t)(1-\rho\tau), \qquad t = n\tau, \qquad n = 0, \ldots, N-1. \tag{A.9}$$

Substituting these into (A.7) and using the induction assumption, we get that

$$x_{t+\tau} = (\rho/\kappa)D_{t+\tau}\tau + o(\tau).$$

Thus, after the discrete trade x_0 at time $t = 0$ all consequent trades will be the continuous. Moreover, (A.8) implies the following form of X_t and D_t dynamics:

$$X_{t+\tau} = X_t - \tfrac{1}{\kappa}\rho D_t\tau + o(\tau), \qquad D_{t+\tau} = D_t + o(\tau). \tag{A.10}$$

Taking into account the initial condition right after the trade at time 0, we find that

$$D_t = D_\tau = \frac{kX_0}{\rho T + 2} + o(\tau).$$

Thus, from (A.8) as $\tau \to 0$ for any $t \in (0, T)$ trade x_t converges to $\frac{\rho X_0}{\rho T+2}\tau$. Since all shares X_0 should be acquired by time T, it is obvious that $\lim_{\tau\to 0} x_T = \frac{X_0}{\rho T+2}$. Q.E.D.

Proof of Propositions 3 and 4

We give the proof of Proposition 4 along with Proposition 3 as a special case. Let us first formulate problem (33) in terms of variables X_t and $D_t = A_t - V_t - s/2$ whose dynamics similar to (A.2) is

$$dD_t = -\rho D_t dt - \kappa dX_t \tag{A.11}$$

with $D_0 = 0$. If we write the cost of continuous and discrete trading as following:

$$dC_t^c = (F_t + s/2)\mu_t dt + \lambda(X_0 - X_t)\mu_t dt + D_t \mu_t dt \tag{A.12}$$

$$\Delta C_t^d = 1_{\{t\in\hat{T}\}} \left[(F_t + s/2)x_t + \lambda(X_0 - X_t)x_t + D_t x_t + x_t^2/(2q)\right]. \tag{A.13}$$

then (33) is equivalent to

$$\min_{\{\mu_{[0,\,T]},\,\{x_{t\in\hat{T}}\}\}\in\Theta_C} \mathrm{E}_t \left[\int_0^t dC_t^c + \sum_{t\in\hat{T}} \Delta C_t^d\right] + (a/2)\int_t^T \sigma^2 X_s^2 ds \tag{A.14}$$

with (A.11), (A.12) and (A.13).

This is the optimal control problem with a single control variable X_t. We can now apply standard methods to find its solution. In particular, the solution will be characterized by three regions where it will be optimal to trade discretely, continuously and do not trade at all. We can specify the necessary conditions for each region which any value function should satisfy. In fact, under some regularity conditions on the value function we can use Ito's lemma together with dynamic programming principle to derive Bellman equation associated with (A.14). For this problem, Bellman equation is a variational inequality involving first-order partial differential equation with gradient constraints. Moreover, the value function should also satisfy boundary conditions. Below we will heuristically derive the variational inequalities and show the candidate function which satisfies them. To prove that this function is a solution we have to check the sufficient conditions for optimality using verification principle.[20]

We proceed with the proof of Proposition 4 in three steps. First, we heuristically define the variational inequalities (VI) and the boundary conditions for the optimization problem (A.14). Second, we show that the solution to the VI exists and implies a candidate value function and a candidate optimal strategy. Third, we verify that candidate value function and optimal strategy are indeed solution to optimization problem. Finally, we will discuss the properties of optimal strategies.

[20]For detailed treatment of similar problems see Hindy, Huang and Zhu(1997), Shreve and Soner (1994), Eastham and Hastings(1988).

A. Variational Inequalities

Let $J(X_t, D_t, F_t, t)$ be a value function for our problem. Then, under some regularity conditions it has to satisfy the necessary conditions for optimality or Bellman equation associated with (A.14). For this problem, Bellman equation is a variational inequality involving first-order partial differential equation with gradient constraints, i.e.,

$$\min\left\{J_t - \rho D_t J_D + \tfrac{1}{2}\sigma^2 J_{FF} + a\sigma^2 X_t^2,\ (F_t{+}s/2) + \lambda(X_0{-}X_t) + D_t - J_X + \kappa J_D\right\} = 0.$$

Thus, the space can be divided into three regions. In the discrete trade (DT) region, the value function J has to satisfy

$$J_t - \rho D_t J_D + \tfrac{1}{2}\sigma^2 J_{FF} + a\sigma^2 X_t^2 > 0, \quad (F_t{+}s/2) + \lambda(X_0{-}X_t) + D_t - J_X + \kappa J_D = 0. \quad \text{(A.15)}$$

In the no trade (NT) region, the value function J satisfies:

$$J_t - \rho D_t J_D + \tfrac{1}{2}\sigma^2 J_{FF} + a\sigma^2 X_t^2 = 0, \quad (F_t{+}s/2) + \lambda(X_0{-}X_t) + D_t - J_X + \kappa J_D > 0. \quad \text{(A.16)}$$

In the continuous trade (CT) region, the value function J has to satisfy:

$$J_t - \rho D_t J_D + \tfrac{1}{2}\sigma^2 J_{FF} + a\sigma^2 X_t^2 = 0, \quad (F_t{+}s/2) + \lambda(X_0{-}X_t) + D_t - J_X + \kappa J_D = 0. \quad \text{(A.17)}$$

In addition, we have the boundary condition at terminal point T:

$$J(X_T, D_T, F_T, T) = (F_T{+}s/2)X_T + \lambda(X_0{-}X_T)X_T + D_T X_T + X_T^2/(2q). \quad \text{(A.18)}$$

Inequalities (A.15)-(A.18) are the so called variational inequalities (VI's), which are the necessary conditions for any solutions to the problem (A.14).

B. Candidate Value Function

Basing on our analysis of discrete-time case we can heuristically derive the candidate value function which will satisfy variational inequalities (A.15)-(A.18). Thus, we will be searching for the solution in a class of quadratic in X_t and D_t functions. Note that it is always optimal to trade at time 0. Moreover, the nature of the problem implies that there should be no NT region. In fact, if we assume that there exists a strategy with no trading at period (t_1, t_2), then it will be always suboptimal with respect to the similar strategy except that the trade at t_1 is reduced by sufficiently small amount ϵ and ϵ trades are continuously executed over period (t_1, t_2). Thus, the candidate value function has to satisfy (A.17) in CT region and (A.15) in any other region.

Since there is no NT region, $(F_t{+}s/2) + \lambda(X_0{-}X_t) + D_t - J_X + \kappa J_D = 0$ holds for any point

(X_t, D_t, F_t, t). This implies a particular form for the quadratic candidate value function:

$$\begin{aligned} J(X_t, D_t, F_t, t) &= (F_t+s/2)X_t + \lambda X_0 X_t \\ &+ [\kappa f(t) - \lambda)]X_t^2/2 + f(t)X_t D_t + [f(t) - 1]D_t^2/(2\kappa) \end{aligned} \tag{A.19}$$

where $f(t)$ is a function which depends only on t. Substituting (A.19) into $J_t - \rho D_t J_D + \frac{1}{2}\sigma^2 J_{FF} + a\sigma^2 X_t^2 \geq 0$ we have:

$$(\kappa f' + a\sigma^2)X_t^2/2 + (f' - \rho f)X_t D_t + (f' + 2\rho - 2f)D_t^2/(2\kappa) \geq 0 \tag{A.20}$$

which holds with an equality for any point of the CT region.

Minimizing with respect to X_t, we show that the CT region is specified by:

$$X_t = -\frac{f' - \rho f}{\kappa f' + a\sigma^2} D_t. \tag{A.21}$$

For $(X_t,\ D_t)$ in the CT region (A.20) holds with the equality. Thus, function $f(t)$ can be found from the Riccati equation:

$$f'(t)(2\rho\kappa + a\sigma^2) - \kappa\rho^2 f^2(t) - 2a\sigma^2\rho f(t) + 2a\sigma^2\rho = 0. \tag{A.22}$$

This guarantees that $J_t - \rho D_t J_D + \frac{1}{2}\sigma^2 J_{FF} + a\sigma^2 X_t^2$ is equal to zero for any points in CT region and greater then zero for any other points. Taking in account terminal condition $f(T) = 1$, we can solve for $f(t)$. As a result, if the trader is risk neutral and $a = 0$, then

$$f(t) = \frac{2}{\rho(T - t) + 2}.$$

Substituting the expression for $f(t)$ into (A.19) we get the candidate value function of Proposition 3. If the trader is risk averse and $a \neq 0$, then

$$f(t) = \tfrac{1}{\kappa\rho}(v - a\sigma^2) - \left[\frac{\kappa\rho}{2v} + \left(\frac{\kappa\rho}{v - a\sigma^2 - \kappa} - \frac{\kappa\rho}{2v}\right) e^{\frac{2\rho v}{2\rho\kappa + a\sigma^2}(T - t)}\right]^{-1}$$

where v is the constant defined in Proposition 4. From (A.19) this results in the candidate value function specified in Proposition 4.

C. Verification Principle

Now we verify that the candidate value function $J(X_0, D_0, F_0, 0)$ obtained above is greater or equal to the value achieved by any other trading policy. Let $X_{[0,\,T]}$ be an arbitrary feasible policy from Θ_C and $V(X_t, D_t, F_t, t)$ be the corresponding value function. We have

$$X(t) = X(0) - \int_0^t \mu_t dt - \sum_{s \in \hat{T},\ s < t} x_s$$

where $\mu_t \geq 0$ and $x_t \geq 0$ for $t \in \hat{T}$. For any τ and X_0, we consider a hybrid policy which follows policy X_t on the interval $[0, \tau]$ and the candidate optimal policy on the interval $[\tau, T]$.

The value function for this policy is

$$V_\tau(X_0,D_0,F_0,0) = \mathrm{E}_0\Bigg[\int_0^\tau [(F_t{+}s/2){+}\lambda(X_0{-}X_t)+D_t]\,\mu_t dt$$
$$+\sum_{t_i<\tau,\ t_i\in\hat{T}} \big[(F_{t_i}{+}s/2)x_{t_i}{+}\lambda(X_0{-}X_{t_i})x_{t_i}+D_{t_i}x_{t_i}+x_{t_i}^2/(2q)\big]+J(X_\tau,D_\tau,F_\tau,\tau)\Bigg]. \tag{A.23}$$

For any function, e.g., $J(X_t,D_t,F_t,t)$ and any (X_t,D_t,F_t,t), we have

$$\begin{aligned} J(X_t,D_t,F_t,t) =& J(X_0,D_0,F_0,0)+\int_0^t J_s ds+\int_0^t J_X dX+\int_0^t J_D dD \\ &+\int_0^t J_F dF+\int_0^t \tfrac{1}{2}J_{FF}(dF)^2+a\sigma^2\int_0^t X_s^2 ds+\sum_{t_i<t,\ t_i\in\hat{T}}\Delta J. \end{aligned} \tag{A.24}$$

Use $dD_t=-\rho D_t dt-\kappa dX_t$ and substitute (A.24) for $J(X_\tau,D_\tau,F_\tau,\tau)$ into (A.23), we have

$$\begin{aligned} V_\tau(X_0,D_0,F_0,0) =& J(X_0,D_0,F_0,0) \\ &+E_0\int_0^\tau \Big[F_t+\frac{s}{2}+\lambda(X_0{-}X_t)+D_t-J_X+\kappa J_D\Big]\mu_t dt \\ &+E_0\int_0^\tau \big(J_t-\rho D_t J_D+\tfrac{1}{2}\sigma^2 J_{FF}+a\sigma^2 X_t^2\big)\,dt \\ &+E_0\sum_{t_i<t,\ t_i\in\hat{T}}\Big[\Delta J+\Big(F_t+\frac{s}{2}+\lambda(X_0{-}X_t)+D_t+x_{t_i}/(2q)\Big)x_{t_i}\Big] \\ =& J(X_0,D_0,F_0,0)+I_1+I_2+I_3 \end{aligned} \tag{A.25}$$

Now we are ready to show that for any arbitrary strategy X_t and for any moment τ it is true that

$$V_\tau(X_0,D_0,F_0,0)\geq J(X_0,D_0,F_0,0). \tag{A.26}$$

It is clear that VI (A.15)-(A.17) implies non-negativity of I_1 and I_2 in (A.25). Moreover, it implies that $I_3\geq 0$. It is easy to be shown if you rewrite $\Delta J(X_{t_i},D_{t_i},F_{t_i},t_i)$ as $J(X_{t_i}-x_{t_i},D_{t_i}+\kappa x_{t_i},F_{t_i}+\sigma Z_{t_i},t_i)-J(X_{t_i},D_{t_i},F_{t_i},t_i)$. This complete the proof of $(A.26)$.

Use it for $\tau=0$ to see that $J(X_0,D_0,F_0,0)\leq V(X_0,D_0,F_0,0)$. Moreover there is an equality if our candidate optimal strategy is used. This complete the proof of Proposition 3.

D. Properties of the Optimal Execution Policy

We now analyze the properties of optimal execution strategies. First, let us consider the risk neutral trader with $a=0$. Substituting the established expression for $f(t)$ into (A.21), we find that the CT region is given by

$$X_t=\frac{\rho(T{-}t){+}1}{\kappa}D_t.$$

This implies that after the initial trade $x_0 = \frac{X_0}{\rho T+2}$ which pushes the system from its initial state X_0 and $D_0 = 0$ into CT region, the trader trades continuously at the rate $\mu_t = \frac{\rho X_0}{\rho T+2}$ staying in CT region and executes the rest $x_T = \frac{X_0}{\rho T+2}$ at the end of trading horizon. In fact, this is the same solution as we had for continuous time limit of solution of problem (20).

If the trader is risk averse then the CT region is given by

$$X_t = g(t)D_t, \qquad \text{where} \qquad g(t) = -\frac{f'(t) - \rho f(t)}{f'(t)\kappa + a\sigma^2}.$$

This implies that after discrete trade $x_0 = X_0 \frac{\kappa f'(0)+a\sigma^2}{\rho\kappa f(0)+a\sigma^2}$ at the beginning which pushes the system from its initial state into CT region, the trader will trade continuously at the rate

$$\mu_t = \kappa x_0 \frac{\rho g(t) - g'(t)}{1+\kappa g(t)} e^{-\int_0^t \frac{\kappa g'(s)+\rho}{1+\kappa g(s)} ds}.$$

This can be shown taking in account the dynamics of D_t given in (A.2) and specification of CT region. At the end the trader finishes the order. Q.E.D.

References

Admati, Anat R., and Paul C. Pfleiderer, 1988, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies* 1, 3-40.

Ahn, Hee-Joon, Kee-Hong Bae and Kalok Chan, 2001, Limit orders, depth, and volatility, *Journal of Finance*, 56, 767–788.

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

Almgren, Robert, and Neil Chriss, 1999, Value under liquidation, *Risk*, 12 (12).

Almgren, Robert, and Neil Chriss, 2000, Optimal execution of portfolio transactions, *University of Toronto Working Paper.*

Back, Kerry, and Shmuel Baruch, 2004, Information in securities markets: Kyle meets Glosten and Milgrom, Econometrica, 72, 433–465.

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

Biais, Bruno, Pierre Hillion, and Chester S. Spatt, 1995, An empirical analysis of the limit order book and order flow in Paris Bourse, *Journal of Finance* 50, 1655–1689.

Biais, Bruno, Christophe Bisiere, and Chester S. Spatt, 2003, Imperfect competition in financial markets, *Carnegie Mellon University Working Paper.*

Bloomfield, Robert, Maureen O'Hara, and Gideon Saar, 2004, The "Make or Take" decision in an electronic market: Evidence on the evolution of liquidity, *Journal of Financial Economics*, forthcoming.

Brock, William A., and Allan W. Kleidon, 1992, Periodic market closure and trading volume: A model of intraday bids and asks, *Journal of Economic Dynamics and Control*, 16, 451–489.

Cao, Charles, Oliver Hansch and Xiaoxin Wang, 2003,The informational content of an open limit order book, *EFA 2004 Maastricht Meetings Paper No. 4311.*

Chan, Kakeung, Wiliam Christie, and Paul Schultz, 1995, Market structure and the intraday pattern of bid-ask spreads for Nasdaq securities, *Journal of Business*, 68(1), 35—60.

Chan, Louis K., and Josef Lakonishok, 1993, Institutional trades and intraday stock price behavior, *Journal of Financial Economics*, 33, 173-199.

Chan, Louis K., and Josef Lakonishok, 1995, The behavior of stock prices around institutional trades, *Journal of Finance*, Vol.50, No.4, 1147–1174.

Chan, Louis K., and Josef Lakonishok, 1997, Institutional equity trading costs: NYSE versus Nasdaq, *Journal of Finance*, Vol.52, No.2, 713–735.

Chung, Kee H., Bonny Van Ness, and Robert Van Ness, 1999, Limit orders and the bid-ask spread, *Journal of Financial Economics*, Vol. 53, No 2, 255–287.

Constantinides, George, 1986, Capital market equilibrium with transactions costs, *Journal of Political Economy*, 94, 842–862.

Coppejans, Mark T., Ian H. Domowitz, and Anath Madhavan, 2001, Liquidity in an automated auction, *AFA 2002 Atlanta Meetings*.

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

Dubil, Robert, 2002, Optimal liquidation of large security holdings in thin markets, *Research in International Entrepreneurial Finance and Business Ventures at the Turn of the Third Millennium, Editor: Yochanan Shachmurove, Academy of Entrepreneurial Finance.*

Easley, David, and Maureen O'Hara, 1992, Time and process of security price adjustment, *Journal of Finance*, 47, 577–605.

Eastham, Jerome, and Kevin J. Hastings, 1988, Optimal impulse control of portfolios, *Mathematics of Operations Research*, 13(4), 588–605.

Foster, Douglas, and S. Viswanathan, 1990, A theory of the interday variations in volume, varience, and trading costs in securities markets, *Review of Financial Studies*, 3, 592–624.

Foster, Douglas, and S. Viswanathan, 1995, Can speculative trading explain the volume-volatility relation?, *Journal of Business and Economic Statistics*, 13, 379–396.

Foucault, Thierry, 1999, Order flow composition and trading costs in a dynamic limit order market, *Journal of Financial Markets*, 2, 99–134.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2001, Limit order book as a market for liquidity, *Review of Financial Studies*, forthcoming.

Gerety, Mason, and Harold Mulherin, 1992, Trading halts and market activity: an analysis of volume at the open and the close, *Journal of Finance*, 47, 1765–1784.

Glosten, Lawrence R., 1994, Is the electronic open limit order book inevitable?, *Journal of Finance*, Vol.49, No.4, 1127–1161.

Goettler, Ron, Christine Parlour, Uday Rajan, 2005, Equilibrium in a dynamic limit order market, *Journal of Finance*, Vol.60, No. 5, 2149–2192.

Greenwood, Robin, 2004, Short– and long–term demand curves for stock: theory and evidence on the dynamics of arbitrage, *Journal of Financial Economics*, forthcoming.

Griffiths, Mark, and Brian Smith, Alasdair Turnbull and Robert W. White, 2000, The Cost and determinants of order aggressiveness, *Journal of Financial Economics* 56, 65–88.

Grinold, Richard, and Ronald Kahn, 2000, *Active Portfolio Management*, McGraw-Hill, New York.

Hamao, Yasushi, and Joel Hasbrouck, 1995, Securities trading in the absence of dealers: Trades and quotes on the Tokyo Stock Exchange, *The Review of Financial Studies*, Vol. 8, No. 3, 849–878.

Handa, Puneet, and Robert Schwartz, 1996, Limit order trading, *Journal of Finance*, 51, 1835–1861.

Harris, Lawrence, 1986, A transactions data study of weekly and intradaily patterns in stock prices, *Journal of Financial Economics*, 16, 99–117.

Harris, Lawrence, 1990, Liquidity, trading rules and electronic trading systems, *NYU Salomon Center Series in Finance and Economics, Monograph* 1990–4.

Harris, Lawrence, 1998, Optimal dynamic order submission strategies in some stylized trading problems, *Financial Markets, Institutions and Instruments*, 7, 1–76.

Harris, Lawrence, and Venkatesh Panchapagesan, 2005, The information-content of the limit order book: Evidence from NYSE specialist trading decisions, *Journal of Financial Markets*, 8, forthcoming.

Hasbrouck, Joel, 1996, Order characteristics and stock price evolution: An application to program trading, *Journal of Financial Economics*, 41(1), 129–149.

Hasbrouck, Joel, and Duane J. Seppi, 2001, Common factors in prices, order flows, and liquidity, *Journal of Financial Economics*, 59, 383–411.

Hasbrouck, Joel and Gideon Saar, 2002, Limit orders and volatility in a hybrid market: the Island ECN, *NYU Working paper.*

He, Hua, and Harry Mamaysky, 2001, Dynamic trading policies with price impact, *Yale ICF Working Paper* No. 00–64.

Hindy, Ayman, Chi-fu Huang, and Steven Zhu, 1997, Optimal consumption and portfolio rules with durability and habit formation, *Journal of Economic Dynamics and Control* 21.

Hollifield, Burton, Robert A. Miller, and Patrik Sandas, 2004, Empirical analysis of limit order markets, *Review of Economic Studies*, 71, 1027-1063.

Holthausen, Robert W., Richard W. Leftwich, and David Mayers, 1987, The effect of large block transactions on security prices: a cross-sectional analysis, *Journal of Financial Economics*, 19, 237–267.

Holthausen, Robert W., Richard W. Leftwitch, and David Mayers, 1990, Large block transactions, the speed of response of temporary and permanent stock-price effect, *Review of Financial Studies*, 9, 71–95.

Hong, Harrison and Jiang Wang, Trading and returns under periodic market closures, 2000, *Journal of Finance*, 55, 297–354.

Huberman, Gur, and Werner Stanzl, 2004, Arbitrage-free price update and price-impact functions, *Econometrica*, 72, No 4, 1247–1275.

Huberman, Gur, and Werner Stanzl, 2005, Optimal liquidity trading, *Review of Finance* 9, 165–200.

Irvine, Paul J., George J. Benston, and Eugene Kandel, 2000, Liquidity beyond the inside spread: measuring and using information in the limit order book, *Working Paper*.

Jain, Prem, and Gun-Ho Joh, 1992, The dependence between hourly prices and trading volume, *Journal of Financial and Quantitative Analysis* 23, 269–284.

Kaul, Aditya, Vikas Mehrotra, and Randall Morck, 2000, Demand curves for stocks do slope down: new evidence from an index weights adjustment, *Journal of Finance* 55 (2), 893-912.

Keim, Donald B., and Ananth Madhavan, 1995, Anatomy of the trading process: Empirical evidence on the behavior of institutional traders, *Journal of Financial Economics* 37, 371- 398.

Keim, Donald B. and Ananth Madhavan, 1996, The upstairs market for large-block transactions, *Review of Financial Studies* 9, 1–36.

Keim, Donald B., and Ananth Madhavan, 1997, Transactions costs and investment style: An inter-exchange analysis of institutional equity trades, *Journal of Financial Economics* 46, 265-292.

Kleidon, Allan, and Ingrid Werner, 1996, UK and US trading of British cross-listed stocks: An intraday analysis of market integration, *Review of Financial Studies*, 9, 619–664.

Koski, Jennifer, and Roni Michaely, 2000, Prices, liquidity, and the information content of trades, *Review of Financial Studies*, 13, 659–696.

Leland, Hayne, 2000, Optimal portfolio implementation with transactions costs and capital gains taxes, *University of California at Berkeley Working Paper*.

McInish, Thomas, and Robert Wood, 1991, Hourly returns, volume, trade size and number of trades, *Journal of Financial Research*, 14(4), 303–315.

Monch, Burkart, 2004, Optimal liquidation strategies, *Goethe University Working Paper.*

Parlour, Christine, 1998, Price dynamics in limit order markets, *Review of Financial Studies*, 11, 789–816.

Parlour, Christine, and Duane Seppi, 2003, Liquidity-based competition for order flow, *Review of Financial Studies*, 16, 301–303.

Perold, Andre, 1988, The implementation shortfall: Paper versus reality, *Journal of Portfolio Management*, 14, 4-9.

O'Hara, Maureen, 1995, Market microstructure theory, *Blackwell Publishers, Cambridge, Mass.*

O'Hara, Maureen, 2003, Liquidity and price discovery, *Journal of Finance*, 58, 1335–1354.

Ranaldo, Angelo, 2004, Order aggressiveness in limit order book markets, *Journal of Financial Markets*, 7, 53–74.

Rock, Kevin, 1996, The specialists order book and price anomalies, *Review of Financial Studies.*

Rosu, Ioanid, 2005, A dynamic model of the limit order book, *University of Chicago Working Paper.*

Sandas, Patrick, 2001, Adverse selection and competitive market making: evidence from a pure limit order book, *Review of Financial Studies*, 14, 705–734.

Scholes, Myron, 1972, The market for securities: substitution versus price pressure and the effects of information on share price, *Journal of Business*, 45, 179–211.

Seppi, Duane, 1997, Liquidity provision with limit orders and a strategic specialist, *Review of Financial Studies*, 10, 103–150.

Shleifer, Andrei, 1986, Do demand curves for stocks slope down?, *Journal of Finance*, 41, 579–590.

Shreve, Steven, and Halil Soner, 1994, Optimal investment and consumption with transaction costs, *Annals of Applied Probability*, No.4, 609–692.

Subramanian, Ajai, and Robert A. Jarrow, 2001, The liquidity discount, *Mathematical Finance*, 11(4), 447–474.

Vayanos, Dimitri, 1999, Strategic Trading and Welfare in a Dynamic Market, *Review of Economic Studies*, 66, 219–254.

Vayanos, Dimitri, 2001, Strategic Trading in a Dynamic Noisy Market, *Journal of Finance*, 56, 131–171.

Wurgler, Jeffrey, and Ekatherina Zhuravskaya, 2002, Does arbitrage flatten demand curves for stock?, *Journal of Business*, vol. 7, 583–608.