Optimal Execution of Security Trading

Hizuru Konishi

Submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

Graduate School of Systems Management

University of Tsukuba

April   2002

Abstract

As the investment business becomes increasingly competitive, it has turned out that the trade execution significantly affects investment performance. As a result, many practitioners have started recognizing trade execution as an independent task. In contrast, few theoretical research papers have been published on the subject. Therefore, this study analyzes the optimal trade execution strategies that minimize trading costs for uninformed traders (traders without specific information) whose trading needs are given exogenously.

In practice, there are two main approaches for saving trading costs:

1. To balance market impact and volatility risk by referring a fixed price in portfolio or block trades.

2. To mitigate impact of trades by referring volume weighted average price (VWAP henceforth) in VWAP trades.

These approaches are chosen according to trade size, traders' objectives, circumstances, and so forth. In Approach 1, we analyze a long-term execution scheduling problem and short-term order placement problem since it is difficult to analyze Approach 1 as a whole. Also, in Approach 2, we analyze both static and dynamic optimization since static optimization works well for numerous small orders while dynamic optimization is more suitable for few large orders. Therefore, the four cases below are analyzed.

1. Fixed–Price Trade

   - Execution Schedule

   - Order Placement

2. VWAP Trade

   - Static Optimization

   - Dynamic Optimization

In practice, all securities trading is carried out through some kind of trade execution, regardless of recognition. At execution, one of the four approaches above is chosen according to the assets' characteristics and the traders' objectives and circumstances. In order to make an appropriate choice, traders have to deeply understand the nature of the alternatives.

Although a few analyses have been made about the optimal trade execution based on investigations of the market microstructure, this study derives the optimal execution strategy in each approach, analyzes its characteristics, and provides the guidance to the selection of these approaches. Therefore, an extensive range of applications can be expected for any security trading.

Acknowledgments

I greatly appreciate the support of my chief thesis advisor Professor Naoki Makimoto who has been an invaluable resource over the past three years. My understanding of both the topics relating to this thesis and the directions taken by the research have been shaped to a large extent by his guidance. The reported work represents an outcome of our collaboration.

I would like to extend special thanks to my associate thesis advisors Professor Hisatoshi Suzuki and Professor Hiroe Tsubaki, who have also offered me a great deal of encouragement and advice. They have introduced me to several interesting related areas of research. I also wish to thank anonymous referees at the University of Tsukuba for numerous suggestions that have sharpened the focus of the article.

Interactions with colleagues at Tokyo-Mitsubishi Securities and Tokyo-Mitsubishi International have also been fruitful. Also, I have benefited greatly from participation in the Trading Technology Research Group.

Finally, I would like to dedicate this work to my wife, Satomi, and two daughters, Sayumi and Megumi, who supported and encouraged me throughout this program at the University of Tsukuba.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Most of the existing research papers on optimal trading such as Black and Litterman (1991) analyze how informed traders (traders with specific information) construct trading strategies based on their price prediction because the investment performance was considered to be explained mostly by the degree to which the prediction was correct. However, as the investment business becomes increasingly competitive, it has turned out that the trade execution also significantly affects investment performance. As a result, many practitioners have started recognizing trade execution as an independent task. For example, the responsibilities of fund managers and traders are clearly separated in most institutional investors. Principal trades in which security brokers take responsibility of trade execution have become popular for complicated trading needs while execution risk remains investors' responsibility in conventional agency trades. In contrast, few theoretical research papers can be found about trade execution. Therefore, this study analyzes the optimal trade execution strategies that minimize trading costs for uninformed traders (traders without specific information) whose trading needs are given exogenously.

The behavior of uninformed traders depends more on short-term characteristics of price movement than on any long-term forecast. Therefore, trade execution strategy is largely affected by the market microstructure, the process and outcomes of exchanging assets among financial intermediations and institutions of exchange.

According to the "implementation shortfall method" by Perold (1988), the standard framework of market microstructure analysis, trading costs are defined as difference between the market price at the decision-making and evaluating price, which consists of the executed price of filled order and the evaluating price of unfilled order. Further, trading costs are decomposed into four components, fixed commisions, timing cost, market impact, and opportunity cost as we see in Figure 1.1. Among them, timing cost is the difference between market price at the decision-making and order placement, market impact is the difference between market price at the order placement and executed price, which is caused by the relationship between trading volume and price change, and opportunity cost is the difference between the executed price and the evaluating price of unfilled order. Strictly speaking, the nature of timing cost

and opportunity cost is the price movement risk and therefore not necessarily a loss. However, the price movement risk is often recognized as the cost after being multiplied with a conversion factor, where the factor is a risk premium, because most traders are risk averse. The trade execution strategy can control any market impact and price movement risk.
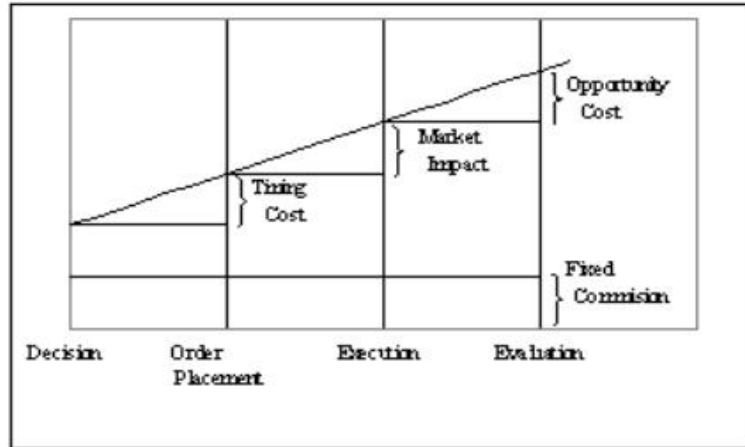
Figure 1.1: **Decomposition of Implementation Shortfall Method.**   Trading costs are decomposed into four components, fixed commisions, timing cost, market impact, and opportunity cost. (x-axis is time and y-axis is stock price.)

In practice, there are several trading methods in which trading cost occurs in different fashions as shown in Table 1.1. We can also classify these trading methods into fixed-price trades and VWAP trades. Fixed-price trades are regarded as trades in which execution price is determined explicitly at the trade execution. Therefore, fixed-price trades include portfolio and block trades, market orders, and limit orders. Since fixed-price trades tend to suffer larger market impact, execution of large trades should be divided into small orders. However, fixed-price trades are frequently used by traders who can monitor and respond to the market quickly because traders can optimise fixed-price trades strategy explicitly.

Alternatively, VWAP trades are the second best approach because VWAP trades provide small trading costs and low price movement risks while not necessarily being optimal. Although VWAP trades cannot be submitted and cancelled quickly because counterparts must agree with the terms and conditions of VWAP before the trades, VWAP trades are convenient methods for certain traders such as foreign investors and individual investors. Therefore, there are mainly two approaches for saving trading costs:

1. To balance the market impact and volatility risk by referring to a fixed price in portfolio or block trades.

2. To mitigate the impact of trades by referring to a volume weighted average price (VWAP henceforth) in VWAP trades.

These approaches are chosen according to trade size, traders' objectives, circumstances, and so forth. For example, while Approach 1 has a direct effect on cost reduction, implementation has to be closely monitored during execution. Therefore, foreign investors rather prefer Approach 2 in order to avoid market impact by completion of large orders under time difference . Figure 1.2 shows a sample flow chart for a choice of trading approaches and methods.

	Cost				Stability / transparency	
	Commission	Timing cost	Market impact	Opportunity cost		
(Agency trade) Market order	medium	small	large	small	bad	
Limit order	medium	small	small	large	good	
VWAP trade	medium	medium	small–large	small	good–moderate	
(Principal trade) Portfolio trade	large	large	N.A.	N.A.	moderate	
Block trade	large	small	N.A.	N.A.	moderate	
VWAP trade	medium	medium	small–large	N.A.	good–moderate	

Handling / optimality	Competitiveness	Others	
			(Agency trade)
bad	bad		Market order
good	bad		Limit order
good	moderate	measure performance	VWAP order
			(Principal trade)
good	good		Portfolio trade
good	good		Block trade
moderate	good		VWAP trade

Table 1.1: **Comparison of trading method.** This table shows the advantages and disadvantages of the trading methods. Trading cost is decomposed according to the implementation shortfall method. Stability means that it is hard to affect the execution price by instantaneous supply and demand. Transparency means that execution price is hard to manipulate. Handling means that it is easy to add discretion. Optimality means that optimal execution can be expected. Competitiveness means that commission is competitive.

This study proposes several optimal trade execution strategies based on the both approaches. In Approach 1, we analyze a long-term execution scheduling problem and short-term order placement problem since it is difficult to analyze Approach 1 as a whole. Also, in Approach 2, we analyze both static and dynamic optimization since static optimization works well for numerous small orders while dynamic optimization is more suitable for few large orders. Therefore, the four cases below are analyzed in the subsequent chapters.

1. Fixed-Price Trade

$$\text{Trading Needs} \begin{cases} \text{Time difference} \begin{cases} \text{Large Trade} \begin{cases} \text{Completion Required} & \rightarrow \textbf{VWAP Trade} \\ \text{Completion Not Required} & \rightarrow \textbf{Limit Order} \end{cases} \\ \\ \text{Small Trade} \begin{cases} \text{Completion Required} & \rightarrow \textbf{VWAP Trade} \text{ or } \textbf{Market Order} \\ \text{Completion Not Required} & \rightarrow \textbf{Limit Order} \text{ or } \textbf{Market Order} \end{cases} \end{cases} \\ \\ \text{No Time difference} \begin{cases} \text{Confident in Execution} \begin{cases} \text{Large Trade} \begin{cases} \text{Small Tracking Error} & \rightarrow \textbf{Portfolio Trade} \\ \text{Large Tracking Error} & \rightarrow \textbf{Block Trade} \end{cases} \\ \text{Medium Trade} \quad \rightarrow \textbf{VWAP Trade} \\ \text{Small Trade} \quad \rightarrow \textbf{Market Order} \end{cases} \\ \\ \text{Unconfident in Execution} \begin{cases} \text{Large Trade} & \rightarrow \textbf{Any Trade} \\ \text{Small Trade} & \rightarrow \textbf{Any Agency Trade} \end{cases} \end{cases} \end{cases}$$

Figure 1.2: **Flow chart for trading method selection.**   This diagram shows a sample flow chart for a selection of the trading methods, based on comparison of the trading methods.

- Execution Schedule

- Order Placement

2. VWAP Trade

- Static Optimization

- Dynamic Optimization

In practice, all securities trading are carried out through some type of trade execution, regardless of recognition. At execution, one of the four approaches above is chosen according to the assets' characteristics and traders' objectives and circumstances. In order to make an appropriate choice, traders have to deeply understand the nature of the alternatives.

Although a few analyses have been made about the optimal trade execution based on findings of market microstructure, this study derives the optimal execution strategy in both fixed-price trades and VWAP trades, analyzes their characteristics, and provides guidance to the selection of these approaches. Therefore, an extensive range of applications can be expected for any security trading activity.

1.1   Fixed-Price Trade

1.1.1   Execution Schedule

Konishi and Makimoto (2001) and Chapter 3 of this study, Optimal Slice of a Block Trade, analyze the static optimal execution schedule of fixed-price trades by variational methods. Traders generally suffer large market impact in immediate execution while they avoid price movement risk. Therefore, block trades are executed over a considerable duration to balancing market impact and price movement risk. This study sets objective function as the sum of linear market impact and price movement risk multiplied

by a conversion factor. This value has two economic insights: total cost for traders where the conversion factor is the risk premium, the extent of risk averseness, and value at risk (VaR) for risk managers where the conversion factor is set according to the confidence interval. This study then, derives analytically the minimal trading cost strategy of liquidation of a portfolio by variational methods. VaR type risk measurement is widely used because the economic implication is intuitive and because risks of different assets are comparable. However, since VaR cannot be additive in time, we employ static optimization but not dynamically optimizing.

The optimal strategies derived by existing studies such as Almgren and Chriss (1999) have the unrealistic property that the time scale of optimal trading is independent of the initial portfolio size although intuition suggests that larger portfolios should trade more slowly and that total execution costs should grow superlinearly with portfolio size. In contrast, our model has several advantages over existing studies. First, we succeeded in making optimal execution duration an endogenous increasing function of whole trading size. Second, the transaction cost of our strategy grows superlinearly with whole trading size, consistent with intuition. Third, our results are immediately applicable to existing risk management frameworks and provides an explicit solution for minimum liquidity adjusted VaR (L-VaR). Although this study uses an example of liquidation of a block of common stock, results can also be extended over wider problems on liquidation of any large block of assets.

1.1.2 Order Placement

Chapter 4 of this study, Selection of Market and Limit Order, analyzes the dynamic optimal order placement strategy of fixed-price trades by dynamic programming. Market and limit orders are frequently submitted and canceled and sometimes switched with each other as market price moves and probability of limit order execution changes. However, since the execution price and volume of limit orders are non-linear, the economic effect and the optimal order placing strategy is not trivial, and few preceding studies can be found. Therefore, this study analyses the optimal selection of market and limit orders in a series of single price batch auctions when the expectation of the limit order book is given.

This study derives the analytical solution of a single limit order model for risk neutral traders by dynamic programming, and then extends the result for risk averse traders. We find that the limit order size is independent of the whole trade size while the limit order price and the market order size are linear functions and that market orders replace limit orders as time passes. Further, we calculate the economic value of monitoring the limit order book.

1.2   VWAP Trade

1.2.1   Static Optimization

Konishi (2002) and Chapter 5 of this study, Optimal Slice of a VWAP Trade, analyze the static optimal execution strategy of VWAP trades by an iteration of a single variable optimization. VWAP trades are overtaking fixed-price trades in numerous trades including foreign investments because of low cost, high transparency, and capability of performance measurement. Because VWAP trades are labor intensive, automatic execution of a static optimal strategy is often employed for trades with small risk.

Although a few analyses have been made on the optimal execution of VWAP trades, this study derives analytically the static optimal execution strategy that minimizes the expected squared execution error when market price, price volatility, and market volume is stochastic. This method is powerful because the optimal execution strategy is determined by an iteration of a single variable optimization, rather than by a multivariable optimization. Analytical solutions are derived in some cases. We show that optimal execution times tend to lag behind expected market trading volume distribution since price volatility has a positive correlation with market trading volume. In a basket trade, execution error can be reduced by spreading out execution times according to the correlation of price movement. We confirm our theoretical results with actual trading data and simulations.

1.2.2   Dynamic Optimization

Chapter 6 of this study, Dynamic Optimal Slice of a VWAP Trade, further analyzes the dynamic optimal execution strategy of VWAP trades by dynamic programming. It is feasible to reduce the execution error with a dynamic optimal strategy if a trader picks up a few stocks, monitors VWAP(of the market and the trader), and forecasts the future price movement and trading volume. This study models the stochastic structure of market volume, and derives an approximation on optimal execution strategy with dynamic programming.

This analysis confirms execution error reduction by actual trading data. We find that even if market trading volume surges due to news, or other reasons, the trader should hold his execution rather than follow market execution. If either buy or sell order is not allowed, execution becomes slower than in the case in which both are allowed.

1.3   Closing Remarks

Although a few analyses have been made about the optimal trade execution based on findings of market microstructure, this study derives the optimal execution strategy in both fixed-price trades and VWAP

trades, analyzes their characteristics, and provides guidance to the selection of these approaches. Therefore, an extensive range of applications can be expected for any security trading activity.

This study is organized as follows. First, Chapter 2 surveys relevant existing studies. Chapter 3 to Chapter 6 analyze the optimal execution strategies of the four cases described above. Finally, Chapter 7 concludes this study.

Chapter 2

Research Survey

2.1 Introduction

Market microstructure is the study of the process and outcomes of exchanging assets among financial intermediations and institutions of exchange. As is suggested by its name, the study of market microstructure has been developed along with the expansion of practical business needs under explicit trading rules and objectives. Although the study of market microstructure spreads over various issues from macroscopic to microscopic as well as from theoretical to empirical, we would like to illustrates two large issues, market design in Section 2.2 and optimal strategy in Section 2.3, which are most relevant to this study.

Looking at the history of investment business and research, we can see several stages. First, the modern portfolio theory has been established in 1960's and developed in 1970's. The greatest academic achievements include studies on the long-term equilibrium of financial markets, such as Portfolio Selection by Markoviz (1959), Capital Asset Pricing Model by Sharpe (1970), and Arbitrage Pricing Theory by Ross (1976). These concepts provided great progress on the investment management, and as a consequence, portfolio investment was introduced and gradually took over investment on individual securities. However, this is also the start of the long lasting argument on the gap between academic research and practical business partly because the practical market is sometimes far away from the equilibrium.

Since the deregulation of the Securities markets such as May Day in 1975 in the U.S.A. and Big Bang in 1986 in the U.K., trading activity has burgeoned. This has led to a widening of the gap between the ideal equilibrium market and the practical market to the extent that it can no longer be considered negligible. Consequently, the need for efficient market design was inevitable, and this gave birth to the study of market microstructure; an analysis of the dynamics of price formation.

2.2 Market Design

In this study area, empirical studies preceded theoretical studies in order to verify practitioners' feelings that the real market could not be explained by existing finance theories. Then, theoretical studies of

general equilibrium followed to provide descriptive explanations to these empirical findings.

Regardless of empirical and theoretical, the main parameters in market microstructure are price and quantity, just as the existing studies on price formation in micro economics. Especially, market microstructure analyzes dynamical mechanisms, these parameters often appears as the change of price or volatility, and trading volume flow, which is henceforth referred as trading volume for simplicity.

Due to the restriction of data and analytical tools, earlier studies observe these two parameters separately in the form of volatility and trading volume, which we call studies on market pattern and explain in Subsection 2.2.1. Later studies tied individual trades and price changes, and observe market impact, which are reflected in Subsection 2.2.2. Further, most of the existing studies are on quote driven markets because market data on order driven markets is too complex. However, we pick up some among a few papers on order driven markets in Subsection 2.2.3 since the understanding of quote driven market is necessary for Chapter 4 of this study.

2.2.1 Market Pattern

It is known that some patterns exist in actual price and liquidity, which cannot be explained by existing finance theories. For example, Jain and Joh (1988) analyze trading data of NYSE stocks from 1979 to 1983 and report the "U-shaped effect" that trading volume is concentrated at the opening and closing of trading sessions, that trading volume on Monday is lower than on other days of the week while weekday pattern is smaller than intraday pattern, and that trading volume is closely related to price movement four hours before, which suggests effect of price movement lasts for a while.

Further, McInish and Wood (1992) analyze the intraday pattern of quote spread on the NYSE. They find that quote spread tightens when trading is active. Also, quote spread is larger for larger orders and smaller for more competitive markets to market makers. While they also confirm the "U-shaped effect" in the intraday pattern at NYSE, quote spread in a quote driven market such as NASDAQ and London Stock Exchange (LSE) remains large throughout the day, and tightens rapidly towards the closing. Therefore, they conclude that difference in trading mechanism causes intraday pattern of quote spread.

Then, Chan et al. (1995) study quote spread of individual market makers and inside spread (difference between best quotes) for NASDAQ stocks. They find that individual quote spread remains stable throughout the day and reveal that decline of inside spread at closing is due to heterogeneity of quote spread among market makers to unwind their positions. Further, Barclay et al. (1999) find that after NASDAQ reform in which public traders became able to submit limit orders, NASDAQ shows the "U-shaped effect" as NYSE.

Numerous studies have been done on the relationship between price volatility and market trading volume since Clark's (1973) seminal paper, including Epps and Epps (1976), Tauchen and Pitts (1983). Karpoff (1987) summarizes these results, and Andersen (1996) proposes a modified model. These studies,

in general, support the existence of a positive correlation between price volatility and market trading volume and the autocorrelation of themselves because surprising news boosts both price volatility and market trading volume.

Regarding Japanese markets, Kawahara and Murase (1993) analyze intraday data of TSE stocks in 1993. They attribute factors of price movement within a day into quote spread, execution period, and trading volume.

On the theory side, trade models are devised in order to provide customary market designs and patterns with theoretical grounds. Market participants in the real world do not always have all the public information and behave rationally, but act with different information, objectives, and circumstances. Therefore, real markets are too complex to analyze. As a result, existing trade models employ significantly simplified assumptions on the roles and objectives of market participants, and try to obtain qualitative implications.

One of the pioneering works is done by Kyle (1985), who assumed that traders submit market orders, and then market makers settle the surplus of traders' orders and determine the price. According to this model, the excess demand of the informed trader is larger when volatility of settlement price is lower and when uncertainty of noise trader is higher. Also, quote spread is smaller when the volatility of settlement price is lower and when uncertainty of the noise trader is higher. Further, this study specifies the abstract idea of market liquidity as 1) existence of quote, 2) small quote spread, 3) capability of successive trades, and 4) capability of instantaneous execution of large trade. Liquid market thus represents an efficient market in the sense that orders of any size are tradable instantaneously around market prices.

Admati and Pfeiderer (1988) analyze a multiple period model in order to explain the "U-shaped effect." They show that orders by uninformed traders tend to cluster in the period with more informed traders such as opening and closing, and price volatility also increases.

Subrahmanyam (1991) assumes riskaverse informed traders and finds that informed traders reduce orders to avoid risks from noise traders, and therefore, price does not reflect information effectively.

2.2.2 Market Impact

Market impact measures the extent to which short-term liquidity makes execution prices diverge from long-term fundamental prices. Holthausen et al. (1987) analyse block trades, and find that prices move in the opposite direction after block trades. They conclude that block orders not only introduce new information to adjust price levels but also bring temporary price turbulence due to low liquidity and speculations. As a result, they distinguish temporary impact and permanent impact.

Also, Hausman et al. (1992) focuses on discreteness of execution price and estimate market impact precisely with an ordered probit model. Further, BARRA (1997) develops a practical market impact model whose impact is an increasing function of trade size and the marginal increase is a decreasing

function. BARRA's model attributes market impact to individual factors (price elasticity of quote, volatility, number of ticks, and trading size) and market factors.

Since analysis on market impact was born in the United States, where quote driven markets are prevailing, there are few studies on order driven markets. It is difficult to identify block orders from trading data in an order driven market since trading data is merely a sequence of times, prices and volumes of ticks which are split into single limit orders on the limit order book. In addition trading data does not show lumps or the direction of orders.

Uno and Yamada (1993) analyze intraday data of TSE stocks. They regard down-ticks as sell orders and up-ticks as buy orders and aggregate net buy orders into five-minute intervals. They explained market impact is determined by long-term characteristics of the asset such as average trading volume and volatility and short-term characteristics of trading such as tick trend, trading volume, and the depth of the limit order book.

2.2.3 Limit Order

Most of the existing studies are on quote driven markets because market data in order driven markets is too complex. However, we pick up some among few papers on order driven markets in this subsection since the understanding of quote driven market is necessary for Chapter 4 of this study.

Market and limit orders are two essential instruments of order placement. Progress in information technology has improved order processing capability and made order placement strategy a critical issue in investment. For example, order driven markets such as the Tokyo Stock Exchange (TSE) have excluded floor traders and adopted automatic matching systems, which makes order processing faster and less error prone. Also, several markets such as New York Stock Exchange (NYSE) and National Association of Securities Dealers Automated Quotation System (NASDAQ) started allowing limit orders, to broadening traders' order placement options. Further, emergence of alternative trading systems such as crossing networks and electronic communication networks have enabled traders to tactically submit and cancel orders across markets. In practice, traders recognize the advantage of limit orders: cost efficiency, and disadvantages: uncertain execution and free option, and choose either market or limit order or sometimes both as it fits their objectives and circumstances.

Several empirical studies confirm traders' behavior in the real market. For example, Biais et al. (1995) analyze trading data of Paris Stock Exchange, an order driven market, and find that traders are affected by limit order books and trades just before the order submission. Especially, more market orders are submitted when the bid-ask spread is small, and more limit orders are submitted inside the bid-ask spread when the bid-ask spread is large, which suggests that traders sufficiently consider the execution probability of limit orders.

Harris and Hasbrouck (1996) analyze data of NYSE and calculate that execution probability of limit

orders within the bid-ask spread is 60%. They show that trading cost of limit orders is lower than other strategies when bid-ask spread is wide, which is consistent with the result of Biais et al. (1995).

Regarding the Japanese market, Kawahara (1994) analyzes TSE stock data in 1994 and finds that the trading cost of limit orders outside the best bid and ask prices is 0.15% lower than that of market orders. Also, that the trading cost is larger for sell orders, large orders, and limit orders with shorter exposure to the market.

In contrast, theoretical studies are limited for the selection of market and limit orders. For example, although Bertisimas and Lo (1998) and Chapter 3 of this study derive the optimal order slicing strategy in a large portfolio liquidation, only market orders but no limit orders are allowed in their models. Also, Parlour (1999) analyses how market conditions affect the selection of market and limit orders when a trader places just one unit of order. This analysis shows the interaction between trader's order placing strategy and market conditions, but fails to study how traders split orders for large trades. Further, Chakravarty and Holden (1995) provide a theoretical model which explains how traders select market and limit orders in a quote driven market. However, existence of a market maker is crucial in their model, and the results are not directly applicable to an order driven market. Accordingly, Chapter 4 of this study analyzes the optimal selection of market and limit orders, sizes, prices, and times in a series of single price batch auctions.

2.2.4 Closing Remarks

Empirical studies successfully reveal practical market anomalies. Especially, it is shown that market and limit orders should be chosen depending on market circumstances and traders' objectives although these orders should be indifferent theoretically in an efficient market. However, analysis on order driven markets has just been started, and numerous issues are left for further analysis since the trading data is complex and disclosure has not been sufficient and also since most exchanges in the United States where research on market microstructure was born are quote driven market.

On the theory side, trade models successfully reveal the roles of market participants and show the conditions in which market stability is retained. However, the models are significantly simplified in order to avoid complexity, and therefore, reflection of the practical market is left for succeeding researches of the optimal strategy, which are explained in Section 2.3.

2.3 Optimal Strategy

In the 1990's, world equity market parcipitants prosperedwhile market prices trended higher. These conditions favoured the birth of numerous hedge funds. Investment business has become more competitive, and trade execution has attracted more attention. With the deep understanding of market mechanisms brought by previous studies on market designs, the study of market microstructure has been revealed to

be another large issue of the optimal strategy. Whereas theoretical models in market design tend to use qualitative analysis and general equilibrium, theories in optimal strategy lend themselves more towards quantitative analysis and the partial equilibrium approach.

The first major task is the conceptual definition of trading cost by Perold (1988), which is explained in Subsection 2.3.1. With this trading cost in mind, earlier theoretical studies try to determine the optimal portfolio with specific forecast, which is explained in Subsection 2.3.2. After trade execution is recognized as an independent function, we saw several studies on the optimal execution as explained in Subsection 2.3.3.

2.3.1 Trading Cost

According to the "implementation shortfall method" by Perold (1988), the standard framework of market microstructure analysis, trading costs are defined as difference between the market price at the time of decision-making and evaluating price. The evaluating price consists of the executed price of the filled order and the evaluating price of the unfilled order. Further, trading costs are decomposed into four components, fixed commisions, timing cost, market impact, and opportunity cost as we see in Figure 1.1. Among them, timing cost is the difference between market price at the time of decision-making and time of order placement. Market impact is the difference between market price at the time of order placement and the executed price, which is caused by the relationship between trading volume and price change. Lastly opportunity cost is the difference between the executed price and the evaluating price of unfilled order. Strictly speaking, the nature of timing cost and opportunity cost is the price movement risk but is not necessarily a loss. However, the price movement risk is often recognized as cost after being mulitiplied with a conversion factor, a risk premium, because most traders are risk averse. Trade execution strategy can control market impact and price movement risk.

2.3.2 Optimal Portfolio Selection

Konno and Wijayanayake (1998) develop a mean absolute difference model in order to build optimal rebalance portfolio under a non-linear trading cost function. They take into account trading costs for rebalance (difference between existing and optimized portfolio) in addition to expected return and risk as usual. Generally, average trading cost is large for small orders, and decreases as trade size increases. However, over large trades tighten order supplies and increase trading cost again. As a result, average trading cost is a "reverse S-shaped" function of trade size.

In order to solve this complicated optimization numerically, this study measures price movement risk in absolute difference and approximates trading cost as a piecewise linear function, which simplifies the problem into piecewise linear programming.

2.3.3   Optimal Execution

In various aspects of financial decision-making from risk management to trading, it is commonly assumed that any trade can be executed around a prevailing price within a short period of time. However in practice, we sometimes face situations in which market cannot fully absorb all the trading needs of a large portfolio. In such cases, asset liquidity risk, defined as the potential deviation between market price and executed price, becomes significant. Liquidity risk has attracted much attention since the LTCM 1998 disaster, as described in Jorion (2000). Conventional VaR measures should be extended to account for liquidity risk as well as optimal trading strategies. This leads to the concept of L-VaR, a liquidity-adjusted risk measure.

Since asset liquidity risk is a highly complicated problem, there used to be a large gap between practice and theory. On the practice side, it is known that in liquidation of a block trade portfolio, traders generally divide the whole trade into small orders in order to reduce the cost. In doing so, it is common to take a longer time for larger trades. Also, BARRA (1997)'s analysis on quote driven market and Mannen and Uno (2000)'s analysis on order driven market show that the quoted premium of a block trade in the real market is an increasing function of whole trading size, and that the marginal increment is a decreasing function (i.e., the first derivative is positive and the second derivative is negative).

On the theory side, Collins and Fabozzi (1991) proposed a conceptual framework, which decomposes trading cost into execution cost and opportunity cost. According to their research, quick execution suffers huge execution cost, while slow execution creates large opportunity cost from price movement risk. Therefore a trader chooses the best strategy that balances these two factors in determining execution schedule.

Subsequent studies investigated more detailed theory. Bertsimas and Lo (1998) derived the optimal execution strategy of a large block of common stock over a fixed time horizon by dynamic programming. However, they analyzed only risk neutral traders because of computational complexity of dynamic programming. So, they just minimized the expectation value of execution cost, but did not consider price movement risk, and therefore, no reasonable guidance on the duration of execution was provided. On the other hand, Grinold and Kahn (1999) and Almgren and Chriss (1999) took the sum of execution cost and opportunity cost of the whole trade, as the objective function to minimize, and derived a static optimal execution strategy. Although they studied risk averse traders, optimal execution duration and average cost are still independent of whole trading size, which contradicts the result of BARRA (1997) and Mannen and Uno (2000). Further, Grinold and Kahn (1999) contradicts their own statement that risk increases as square root of time, since their risk measured in variance, in fact, increases linearly. This is crucial because the time dependency of cost and risk determines the optimal execution strategy.

To the contrary, in order to bridge the gap between theory and practice, this chapter modifies objective

function in a practical manner, which reflects the reality of the order driven market, and obtains a reasonable relationship between transaction cost and whole trading size. Specifically, we extend the results on optimal execution strategy derived by Grinold and Kahn (1999) and Almgren and Chriss (1999). As mentioned above, the explicit solutions obtained in the earlier work have the unrealistic properties in that the time scale of optimal execution strategy is independent of whole trading size, and the total cost of trading is linear in whole trading size.

There are two possible ways to capture these effects in a quantitative model. One way is to modify the cost functions to be nonlinear. Another way is to change the weighting of cost uncertainty from a mean-variance model based on a smooth utility function, to a penalty based on standard deviation, which is more in the spirit of VaR models. This approach was briefly discussed in Almgren and Chriss (1999) where it was observed that solutions could be obtained graphically on the efficient frontier, but explicit solutions were not obtained due to the nonlinearity of the problem, nor were the properties of these solutions explored.

For the standard deviation model, our study in Chapter 3 derives explicit solutions of optimal execution strategy under asset liquidity risk. It extends the work of Almgren and Chriss (1999), who derive a measure of L-VaR. We obtain explicit solutions both for the single-asset case and for a multi-asset portfolio model in the form of pure exponentials. The explicit form obtained for the coefficients shows that the characteristic time increases as whole trading size to the two-thirds power, and total cost increases as whole trading size to the four-thirds power, both consistent with intuition.

Optimal execution is very important in both fixed-price and VWAP trades. It is true that VWAP has become an industry standard, and that numerous studies have been done on market pattern as we have seen in Subsection 2.2.1, and that several studies have been made on optimal strategies for general trade execution. However, little academic work can be found on the optimal execution of VWAP trades. Therefore, Chapter 5 and 6 analyze the static and dynamic optimal execution of VWAP trades, respectively.

2.3.4 Closing Remarks

Previous studies successfully balance market friction and traders' objectives, which is not addressed by existing finance theory. However, for analytical feasibility, most studies are too simplified and sometimes inconsistent with market practice. Further, there are several trading methods in practice. Policies which suggests how to select appropriate trading method according to asset characteristics and traders' objectives and circumstances are left for further analysis. In practice, all the security trading is carried out through some trade execution, regardless of recognition. At execution, one of the four approaches above is chosen according to assets' characteristics and traders' objectives and circumstances. In order to make an appropriate choice, traders have to deeply understand the nature of the alternatives.

Although few analyses have been made about the optimal trade execution based on findings of market microstructure, this study derives the optimal execution strategy in each approach, analyzes its characteristics, and provides the guidance to the selection of these approaches. Therefore, an extensive range of application can be expected for any security trading.

Chapter 3

Optimal Slice of a Block Trade

Abstract   In liquidation of a block trade, it is common to divide execution in order to minimize total transaction cost, balancing market impact costs and volatility effects. This chapter formulates the execution scheduling as a static optimization problem assuming linear market impact and derives explicit solutions. We show that optimal execution duration and average cost are increasing functions of the initial portfolio size, which is consistent with trading practice and empirical studies. Our model provides explicit solutions for minimum liquidity adjusted value at risk (L-VaR), which is particularly useful for risk management.

3.1   Introduction

In various aspects of financial decision-making from risk management to trading, it is commonly assumed that any trade can be executed around a prevailing price within a short period of time. However in practice, we sometimes face situations in which market cannot fully absorb all the trading needs of a large portfolio. In such cases, asset liquidity risk, defined as the potential deviation between market price and executed price, becomes significant. Liquidity risk has attracted much attention since the LTCM 1998 disaster, as described in Jorion (2000). Conventional VaR measures should be extended to account for liquidity risk as well as optimal trading strategies. This leads to the concept of L-VaR, a liquidity-adjusted risk measure.

   This chapter derives explicit solutions of optimal execution strategy under asset liquidity risk. It extends the work of Almgren and Chriss (1999), who derive a measure of L-VaR. Almgren and Chriss obtain optimal execution strategies by considering the trade-off between immediate liquidation, which entails no price risk but has high market impact, and slow liquidation, which minimizes market impact but exposes the portfolio to price risk. Their approach, however, has the unrealistic property that the time scale of optimal trading is independent of the initial portfolio size. Intuition suggests that larger

portfolios should trade more slowly and that total execution costs should grow superlinearly with portfolio size.

In contrast, our model has several advantages. First, we succeeded in making optimal execution duration an endogenous increasing function of whole trading size. Second, our transaction cost grows superlinearly with whole trading size, consistent with intuition. Third, our results are immediately applicable to existing risk management framework and provides explicit solution for minimum L-VaR. Also, although this chapter uses an example of liquidation of a block of common stock, results can be extended over wider problems on liquidation of any large block of assets.

Since asset liquidity risk is a highly complicated problem, there used to be a large gap between practice and theory. On the practice side, it is known that in liquidation of a block trade portfolio, traders generally divide the whole trade into small orders in order to reduce his cost. In doing so, it is common to take a longer time for larger trades. Also, BARRA (1997)'s analysis on quote driven market and Mannen and Uno (2000)'s analysis on order driven market show that the quoted premium of a block trade in the real market is an increasing function of whole trading size, and the marginal increment is a decreasing function (i.e., the first derivative is positive and the second derivative is negative).

On the theory side, Collins and Fabozzi (1991) proposed a conceptual framework, which decomposes trading cost into execution cost and opportunity cost. According to their research, quick execution suffers huge execution cost, while slow execution creates large opportunity cost from price movement risk. Therefore a trader chooses the best strategy that balances these two factors in determining execution schedule.

Subsequent studies investigated more detailed theory. Bertsimas and Lo (1998) derived the optimal execution strategy of a large block of common stock over a fixed time horizon by dynamic programming. However, they analyzed only risk neutral traders because of computational complexity of dynamic programming. So, they just minimized the expectation value of execution cost, but did not consider price movement risk, and therefore, no reasonable guidance on the duration of execution was provided. On the other hand, Grinold and Kahn (1999) and Almgren and Chriss (1999) took the sum of execution cost and opportunity cost of the whole trade, as the objective function to minimize, and derived a static optimal execution strategy. Although they studied risk averse traders, optimal execution duration and average cost are still independent of whole trading size, which contradicts the result of BARRA (1997) and Mannen and Uno (2000). Further, Grinold and Kahn (1999) contradicts their own statement that risk increases as square root of time, since their risk measured in variance, in fact, increases linearly. This is crucial because the time dependency of cost and risk determines the optimal execution strategy.

To the contrary, in order to bridge the gap between theory and practice, this chapter modifies objective function in a practical manner, which reflects the reality of the order driven market, and obtains a reasonable relationship between transaction cost and whole trading size. Specifically, we extend the

results on optimal execution strategy derived by Grinold and Kahn (1999) and Almgren and Chriss (1999). As mentioned above, the explicit solutions obtained in the earlier work have the unrealistic properties that the time scale of optimal execution strategy is independent of whole trading size, and the total cost of trading is linear in whole trading size.

There are two possible ways to capture these effects in a quantitative model. One way is to modify the cost functions to be nonlinear. Another way is to change the weighting of cost uncertainty from a mean-variance model based on a smooth utility function, to a penalty based on standard deviation, which is more in the spirit of VaR models. This approach was briefly discussed in Almgren and Chriss (1999) where it was observed that solutions could be obtained graphically on the efficient frontier, but explicit solutions were not obtained due to the nonlinearity of the problem, nor were the properties of these solutions explored.

For the standard deviation model, we obtain explicit solutions both for the single-asset case and for a multi-asset portfolio model in the form of pure exponentials. The explicit form obtained for the coefficients shows that the characteristic time increases as whole trading size to the two-thirds power, and total cost increases as whole trading size to the four-thirds power, both consistent with intuition.

The biggest problem with this VaR-type weighting of risk is that the solutions do not satisfy the "principle of optimality": solutions computed statically at the start of trading are no longer optimal if reevaluated in the course of execution. Although formally this makes the solutions invalid, and calls for the use of a dynamic programming method, these solutions are useful in practice because of the following advantages:

- Statically optimized execution can be used as a benchmark, i.e., actual trading strategy might be changed as new information arrives and the performance of a strategy change can be evaluated by comparing the cost to that of static optimal execution strategy.

- Statically optimized execution is sometimes a reasonable choice in program trading, i.e., it is technically difficult to record all the history of price and profit and change the strategy dynamically especially in a portfolio trade. Therefore, static optimal execution strategy is a good alternative in practical trading.

This chapter is organized as follows. In the next section, we formulate the cost minimization problem. Section 3.3 is devoted to derive the optimal execution strategy for general case. The case when we have limit to the volume executable instantaneously is studied in Section 3.4. Section 3.5 discusses the principle of optimality. Finally in Section 3.6, we conclude the analysis.

3.2   Formulation of Optimization

Let $(\Omega, \mathcal{F}, Q)$ be a probability space with a filtration $\{\mathcal{F}_t\}$ which satisfies usual conditions. Suppose we would like to liquidate a portfolio of N assets and whole trading size of asset i is V_i $(i = 1, \cdots, N)$. Let $P_i(t)$ $(i = 1, \cdots, N)$ denote $\{\mathcal{F}_t\}$-adapted stock's fair value process. Similar to the standard option pricing theory, the stock price process is assumed to follow standard Brownian motion with volatility σ_i and correlation ρ_{ij}. Drift term is considered to be zero because the time frame is short. So, the stock's fair value processes are represented as

$$dP_i(t) = \sigma_i dB_i(t), \quad i = 1, \cdots, N,$$

$$dB_i(t)dB_j(t) = \rho_{ij}dt, \quad i, j = 1, \cdots, N.$$

Let $v_i(t)$ $(i = 1, \cdots, N)$ denote the selling volume of asset i at time t. Because of the order quantity constraints,

$$\int_0^\infty v_i(t)dt = V_i, \qquad i = 1, \cdots, N. \tag{3.1}$$

In the rest of the chapter, we consider total transaction cost as the sum of execution cost and opportunity cost, and study the static minimization of transaction cost under this framework. More precisely, we look for $v_i(t)$ which minimizes transaction cost among all trading strategy that is pre-determined at $t = 0$.

Regarding execution cost, the main factor that can be controlled by order slicing is market impact, the difference between observed price at the market and the stock's fair value. Holthausen et al. (1987) and Uno and Yamada (1993) say that market price tends to overshoot with a block trade, and therefore, market impact should be divided into temporary and permanent components. Further, several features are known regarding the market impact in an order driven market, such as the Japanese market. First, according to Copeland and Galai (1983), a limit order can be considered as a free option contract, and so, only a part of demand and supply are placed as limit orders at one time. Therefore, a trader has to wait until orders on the opposite side appear. Actually, Biais et al. (1995) observed that there are few limit orders at prices far from the market price. Second, while a quote driven market allows a trader to negotiate prices with a market maker according to trading volume, an order driven market does not have a function to negotiate price and trading volume, which forces traders to repeat small execution. Also, if anonymity is guaranteed in an order driven market, repeating execution is thought to leak less information than a bulk execution.

In summary, the advantages of order slicing are 1) avoiding temporary market impact, 2) waiting for all the potential orders to appear, and 3) concealing one's trade. Therefore, this chapter separates market impact into temporary and permanent component, sets a limit to the volume executable instantaneously, and assumes each component to be a linear function of trading volume within moderate volume and time.

Choice between market order and limit order is certainly another important factor in optimal execution. However, this chapter regards it as a tactical decision after the order slice is fixed, and excludes it from the scope of the research since it is common to place orders only around a prevailing price as mentioned above. Specifically, we introduce coefficients below.

$a_i > 0 \ (i = 1, \cdots, N)$: temporary impact coefficient of asset i,

$b_i > 0 \ (i = 1, \cdots, N)$: permanent impact coefficient of asset i.

Since temporary impact depends only on the momentary trade and permanent impact depends on the accumulated volume traded up to that moment, total market impact at time t can be represented as

$$\sum_{i=1}^{N} v_i(t) \left\{ a_i v_i(t) + b_i \int_0^t v_i(s) ds \right\}.$$

Therefore, market impact over the execution period becomes

$$
\begin{aligned}
(\text{market impact}) \quad &= \quad \int_0^\infty \sum_{i=1}^{N} v_i(t) \left\{ a_i v_i(t) + b_i \int_0^t v_i(s) ds \right\} dt \\
&= \quad \int_0^\infty \boldsymbol{v}(t)^\top \boldsymbol{A} \boldsymbol{v}(t) dt + \frac{1}{2} \boldsymbol{V}^\top \boldsymbol{B} \boldsymbol{V}
\end{aligned}
\tag{3.2}
$$

where $\boldsymbol{v}(t) = (v_1(t), \cdots, v_N(t))^\top$, $\boldsymbol{V} = (V_1, \cdots, V_N)^\top$, $\boldsymbol{A} = \mathrm{Diag}(a_1, \cdots, a_N)$, $\boldsymbol{B} = \mathrm{Diag}(b_1, \cdots, b_N)$. Here, $\mathrm{Diag}(z_1, \cdots, z_N)$ denotes an $N \times N$ diagonal matrix whose diagonal element is $z_i \ (i = 1, \cdots, N)$ and $\top$ denotes transpose.

Meanwhile, regarding opportunity cost, price movement risk is often measured by a multiple of standard deviation as VaR, which has the same dimension as the price itself. This risk measurement has the following advantages:

- A trader can evaluate the appropriateness of the trade by estimating risk as VaR, i.e., he compares the maximum loss of the block trade and his risk tolerance, and judges before the trade whether he can assume the risk.

- A trader can evaluate the profitability of the trade, i.e., if a trader requires a profit greater than Sharp ratio (return over standard deviation) λ, the premium p should be

$$p > (\text{market impact}) + \lambda(\text{opportunity cost})$$

since

$$\frac{p - (\text{market impact})}{(\text{opportunity cost})} > \lambda.$$

Specifically, we define opportunity cost as standard deviation of fair sales value. Total fair value of the whole execution is

$$\sum_{i=1}^{N} \int_0^\infty v_i(t) P_i(t) dt \quad = \quad \sum_{i=1}^{N} \int_0^\infty v_i(t) \left\{ P_i(0) + \int_0^t dP_i(s) \right\} dt$$

$$= \sum_{i=1}^{N} \left\{ V_i P_i(0) + \int_0^\infty x_i(t) dP_i(t) \right\}$$

$$= \sum_{i=1}^{N} \left\{ V_i P_i(0) + \sigma_i \int_0^\infty x_i(t) dB_i(t) \right\}$$

where $x_i(t) = \int_t^\infty v_i(s) ds$. So, standard deviation is given by

$$\sqrt{\mathrm{E}\left(\left(\sum_{i=1}^{N} \sigma_i \int_0^\infty x_i(t) dB_i(t)\right)^2\right)} = \sqrt{\int_0^\infty \sum_{i,j=1}^{N} \sigma_i \sigma_j \rho_{ij} x_i(t) x_j(t) dt}$$

$$= \sqrt{\int_0^\infty \boldsymbol{x}(t)^\top \boldsymbol{S} \boldsymbol{x}(t) dt} \qquad (3.3)$$

where $\boldsymbol{S} = (\sigma_i \sigma_j \rho_{ij}; i,j = 1, \cdots, N)$ is a variance-covariance matrix of price processes and $\boldsymbol{x}(t) = (x_1(t), \cdots, x_N(t))^\top$. Therefore, opportunity cost can be defined as $\lambda \sqrt{\int_0^\infty \boldsymbol{x}(t)^\top \boldsymbol{S} \boldsymbol{x}(t) dt}$ where λ is the risk premium.

Total transaction cost (TC) of a trader is the sum of execution cost and opportunity cost, and from (3.2) and (3.3),

$$TC = \int_0^\infty \boldsymbol{x}'(t)^\top \boldsymbol{A} \boldsymbol{x}'(t) dt + \frac{1}{2} \boldsymbol{V}^\top \boldsymbol{B} \boldsymbol{V} + \lambda \sqrt{\int_0^\infty \boldsymbol{x}(t)^\top \boldsymbol{S} \boldsymbol{x}(t) dt}$$

which can be divided into fixed cost (FC) and variable cost (VC),

$$FC = \frac{1}{2} \boldsymbol{V}^\top \boldsymbol{B} \boldsymbol{V},$$

$$VC(\boldsymbol{x}) = \int_0^\infty \boldsymbol{x}'(t)^\top \boldsymbol{A} \boldsymbol{x}'(t) dt + \lambda \sqrt{\int_0^\infty \boldsymbol{x}(t)^\top \boldsymbol{S} \boldsymbol{x}(t) dt}.$$

Here, $\boldsymbol{x} = (\boldsymbol{x}(t), 0 \le t < \infty)$ and $\boldsymbol{x}'(t) = -\boldsymbol{v}(t)$ denotes element-wise derivative. In what follows, we assume $\boldsymbol{x} = (\boldsymbol{x}(t), 0 \le t < \infty) \in \mathcal{C}_\infty^2$ meaning that each component of $\boldsymbol{x}(t) = (x_1(t), \cdots, x_N(t))^\top$ has continuous second derivative $x_i''(t)$ and vanishes at infinity, i.e., $\lim_{t \to \infty} x_i(t) = 0$ and hence $\lim_{t \to \infty} x_i'(t) = 0$. Therefore, regarding $VC(\boldsymbol{x})$ as a functional of $\boldsymbol{x}$, our problem can be formulated as follows:

$$(\mathrm{P}) \quad \left| \begin{array}{ll} \text{minimize} & VC(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} = (\boldsymbol{x}(t), 0 \le t < \infty) \in \mathcal{C}_\infty^2, \quad \boldsymbol{x}(0) = \boldsymbol{V}. \end{array} \right.$$

This definition is more practical than that of Almgren and Chriss (1999) in that the opportunity cost is measured by estimating standard deviation similar to VaR. As we will see later, dimension of the market impact becomes larger than that of opportunity cost because of this definition, and the execution schedule becomes a function of whole trading size.

3.3   Derivation of the Optimal Execution Strategy

In order to derive the optimal slicing, we first prepare some preliminary results. First, choose arbitrary strategy $\boldsymbol{x} = (\boldsymbol{x}(t), 0 \le t < \infty) \in \mathcal{C}_\infty^2$ and let $\boldsymbol{y} = (\boldsymbol{y}(t), 0 \le t < \infty) \in \mathcal{C}_{0,\infty}^2$ be any perturbation of $\boldsymbol{x}$

where $\boldsymbol{y} = (\boldsymbol{y}(t), 0 \leq t < \infty) \in \mathcal{C}_{0,\infty}^2$ means that each element $y_i(t)$ of $\boldsymbol{y}(t) = (y_1(t), \cdots, y_N(t))^\top$ has continuous second derivative and vanishes at $t = 0$ and infinity, i.e., $y_i(0) = \lim_{t \to \infty} y_i(t) = 0$. Then, a direct calculation shows

$$\frac{d}{d\epsilon} VC(\boldsymbol{x} + \epsilon\boldsymbol{y}) = 2C + 2D\epsilon + \frac{\lambda(F + G\epsilon)}{\sqrt{E + F\epsilon + G\epsilon^2}},$$

$$\frac{d^2}{d\epsilon^2} VC(\boldsymbol{x} + \epsilon\boldsymbol{y}) = 2D + \frac{\lambda(EG - F^2)}{(E + F\epsilon + G\epsilon^2)^{3/2}}$$

where

$$C = \int_0^\infty \boldsymbol{x}'(t)^\top \boldsymbol{A} \boldsymbol{y}'(t) dt, \qquad D = \int_0^\infty \boldsymbol{y}'(t)^\top \boldsymbol{A} \boldsymbol{y}'(t) dt,$$

$$E = \int_0^\infty \boldsymbol{x}(t)^\top \boldsymbol{S} \boldsymbol{x}(t) dt, \qquad F = \int_0^\infty \boldsymbol{x}(t)^\top \boldsymbol{S} \boldsymbol{y}(t) dt, \qquad G = \int_0^\infty \boldsymbol{y}(t)^\top \boldsymbol{S} \boldsymbol{y}(t) dt.$$

Lemma 3.3.1 *The equality*

$$\frac{d}{d\epsilon} VC(\boldsymbol{x} + \epsilon\boldsymbol{y}) \bigg|_{\epsilon=0} = 0 \tag{3.4}$$

holds for any $\boldsymbol{y} \in \mathcal{C}_{0,\infty}^2$ *if and only if* $\boldsymbol{x}$ *satisfies*

$$\boldsymbol{x}''(t) = \frac{\lambda}{2\sqrt{E}} \boldsymbol{A}^{-1} \boldsymbol{S} \boldsymbol{x}(t), \qquad \forall t \in (0, \infty). \tag{3.5}$$

Proof. Since $\lim_{t \to \infty} \boldsymbol{x}'(t) = \boldsymbol{0}$ and $\boldsymbol{y}(0) = \boldsymbol{0}$,

$$C = [\boldsymbol{x}'(t) \boldsymbol{A} \boldsymbol{y}(t)]_{t=0}^\infty - \int_0^\infty \boldsymbol{x}''^\top(t) \boldsymbol{A} \boldsymbol{y}(t) dt = - \int_0^\infty \boldsymbol{x}''^\top(t) \boldsymbol{A} \boldsymbol{y}(t) dt.$$

Then, we get

$$\frac{d}{d\epsilon} VC(\boldsymbol{x} + \epsilon\boldsymbol{y}) \bigg|_{\epsilon=0} = 2C + \frac{\lambda F}{\sqrt{E}} = 2 \int_0^\infty \left\{ -\boldsymbol{x}(t)''^\top \boldsymbol{A} + \frac{\lambda}{2\sqrt{E}} \boldsymbol{x}(t)^\top \boldsymbol{S} \right\} \boldsymbol{y}(t) dt.$$

Therefore, (3.4) holds for arbitrary $\boldsymbol{y} \in \mathcal{C}_{0,\infty}^2$ if and only if (3.5) holds since $\boldsymbol{A}$ is non-singular and both $\boldsymbol{A}$ and $\boldsymbol{S}$ are symmetry. $\square$

The next lemma is concerned with the second order condition.

Lemma 3.3.2 *For any* $\boldsymbol{y} \in \mathcal{C}_{0,\infty}^2$,

$$\frac{d^2}{d\epsilon^2} VC(\boldsymbol{x} + \epsilon\boldsymbol{y}) \geq 0, \qquad \forall \epsilon \in (-\infty, \infty).$$

Proof. See Section 3.7. $\square$

From Lemmas 3.3.1 and 3.3.2, and the optimality criterion of the variational method, the optimization problem is reduced to finding $\boldsymbol{x}$ that satisfies (3.5) and the boundary condition. When there is only single asset, this problem is relatively easy. Indeed, standard theory of differential equations shows that

$$x(t) = V e^{-\alpha t}, \qquad t \geq 0 \tag{3.6}$$

where α can be identified as

$$\alpha = \left(\frac{\lambda^2\sigma^2}{2a^2V^2}\right)^{1/3}$$

by substituting (3.6) into (3.5) and solving it in terms of α. So, we have the following result.

Proposition 3.3.1   *For single asset case, the optimal strategy becomes*

$$x(t) = Ve^{-\alpha t}, \quad t \geq 0; \qquad \alpha = \left(\frac{\lambda^2\sigma^2}{2a^2V^2}\right)^{1/3}.$$

Also, average variable cost is

$$\frac{VC}{V} = \left(\frac{27}{16}\lambda^2\sigma^2 aV\right)^{1/3}.$$

In contrast to single asset case, however, the problem is more involved for multiple asset case. Let η_i $(i = 1, \cdots, N)$ denote eigenvalues of $\boldsymbol{A}^{-1}\boldsymbol{S}$. The facts that $\boldsymbol{A}^{-1}\boldsymbol{S}$ and $\boldsymbol{A}^{-1/2}\boldsymbol{S}\boldsymbol{A}^{-1/2}$ with $\boldsymbol{A}^{-1/2} =$ $\text{Diag}(a_1^{-1/2}, \cdots, a_N^{-1/2})$ have common eigenvalues and that $\boldsymbol{S}$ as well as $\boldsymbol{A}^{-1/2}\boldsymbol{S}\boldsymbol{A}^{-1/2}$ are non-negative definite ensure $\eta_i \geq 0$ $(i = 1, \cdots, N)$. In what follows, we make the following assumption:

Assumption 3.3.1   *All eigenvalues η_i $(i = 1, \cdots, N)$ of $\boldsymbol{A}^{-1}\boldsymbol{S}$ are distinct and positive.*

This assumption is not critical in practice since if it happens that the assumption does not hold, all we have to do is to perturbate some parameter slightly so as to make the assumption satisfied.

Under this assumption, $\boldsymbol{A}^{-1}\boldsymbol{S}$ is represented as

$$\boldsymbol{A}^{-1}\boldsymbol{S} = \sum_{i=1}^{N} \eta_i \boldsymbol{r}_i \boldsymbol{\ell}_i^{\top}$$

where $\boldsymbol{r}_i$ $(\boldsymbol{\ell}_i^{\top}$, respectively) is a right (left) eigenvector of $\boldsymbol{A}^{-1}\boldsymbol{S}$ associated with η_i normalized so as to satisfy

$$\boldsymbol{\ell}_i^{\top}\boldsymbol{r}_j = \left\{ \begin{array}{ll} 1, & i = j, \\ 0, & i \neq j. \end{array} \right.$$

So, if we define

$$\boldsymbol{R} = \sum_{i=1}^{N} \sqrt{\eta_i}\, \boldsymbol{r}_i \boldsymbol{\ell}_i^{\top}, \tag{3.7}$$

then, $\boldsymbol{R}^2 = \boldsymbol{A}^{-1}\boldsymbol{S}$. To proceed further, consider a system of linear equations

$$\boldsymbol{R}^{\top}\boldsymbol{T} + \boldsymbol{T}\boldsymbol{R} = \boldsymbol{S}. \tag{3.8}$$

Lemma 3.3.3   *(3.8) has a unique solution $\boldsymbol{T} = (T_{ij}; i, j = 1, \cdots, N)$.*

Proof.   Since all eigenvalues $\sqrt{\eta_i}$ $(i = 1, \cdots, N)$ of $\boldsymbol{R}$ are distinct under Assumption 3.3.1, there exists a non-singular matrix $\boldsymbol{P}$ for which

$$\boldsymbol{K} \overset{\text{def}}{=} \boldsymbol{P}^{-1}\boldsymbol{R}\boldsymbol{P} = \text{Diag}(\sqrt{\eta_1}, \cdots, \sqrt{\eta_N}). \tag{3.9}$$

From (3.8) and (3.9), we obtain

$$\boldsymbol{P}^{\top}\boldsymbol{S}\boldsymbol{P} = \boldsymbol{K}(\boldsymbol{P}^{\top}\boldsymbol{T}\boldsymbol{P}) + (\boldsymbol{P}^{\top}\boldsymbol{T}\boldsymbol{P})\boldsymbol{K}. \tag{3.10}$$

If we denote $\boldsymbol{P}^{\top}\boldsymbol{S}\boldsymbol{P} = (\theta_{ij}; i,j = 1, \cdots, N)$ and $\boldsymbol{P}^{\top}\boldsymbol{T}\boldsymbol{P} = (\tau_{ij}; i,j = 1, \cdots, N)$, (3.10) is reduced to

$$\theta_{ij} = \sqrt{\eta_i}\tau_{ij} + \sqrt{\eta_j}\tau_{ij}, \qquad i,j = 1, \cdots, N. \tag{3.11}$$

Since all η_i's are positive, τ_{ij} is easily obtained from (3.11) and thus $\boldsymbol{T}$ exists uniquely. $\square$

Now we are in a position to obtain the optimal execution strategy. Let

$$\boldsymbol{z}(t) = \exp(-\alpha \boldsymbol{R}t)\boldsymbol{V} = \sum_{i=1}^{N}(\boldsymbol{\ell}_i^{\top}\boldsymbol{V})e^{-\alpha\sqrt{\eta_i}t}\boldsymbol{r}_i, \qquad t \geq 0$$

where

$$\alpha = \left(\frac{\lambda^2}{4\boldsymbol{V}^{\top}\boldsymbol{T}\boldsymbol{V}}\right)^{1/3}.$$

Then, we have the following proposition.

Proposition 3.3.2 $\boldsymbol{z} = (\boldsymbol{z}(t), 0 \leq t < \infty)$ *is the optimal solution for the problem (P).*

Proof. That $\boldsymbol{z} \in \mathcal{C}_{\infty}^{2}$ and $\boldsymbol{z}(0) = \boldsymbol{V}$ is obvious. Thus, it is enough to confirm that $\boldsymbol{z}$ satisfies (3.5). Since

$$\boldsymbol{z}''(t) = \alpha^2 \boldsymbol{R}^2 \boldsymbol{z}(t) = \alpha^2 \boldsymbol{A}^{-1}\boldsymbol{S}\boldsymbol{z}(t),$$

this is equivalent to check $\alpha^2 = \lambda/2\sqrt{E}$. Now note that

$$\frac{d}{dt}(\boldsymbol{z}(t)^{\top}\boldsymbol{T}\boldsymbol{z}(t)) = -\alpha\boldsymbol{z}(t)^{\top}(\boldsymbol{R}^{\top}\boldsymbol{T} + \boldsymbol{T}\boldsymbol{R})\boldsymbol{z}(t) = -\alpha\boldsymbol{z}(t)^{\top}\boldsymbol{S}\boldsymbol{z}(t).$$

Then,

$$E = \alpha^{-1}\left[\boldsymbol{z}(t)^{\top}\boldsymbol{T}\boldsymbol{z}(t)\right]_{t=\infty}^{0} = \alpha^{-1}\boldsymbol{V}^{\top}\boldsymbol{T}\boldsymbol{V}$$

from which we obtain

$$\alpha^2 = \frac{\lambda}{2\sqrt{E}},$$

completing the proof. $\square$

Similar to Almgren and Chriss (1999), $\boldsymbol{z}$ is not always a monotone decreasing non-negative function. For example, when the portfolio consists of two highly correlated assets, one with high liquidity and the other with low liquidity, it sometimes makes sense to go short in the high liquid security at first so as to avoid price movement risk, and then liquidate the portfolio slowly afterwards. This strategy is often taken for portfolio with cash equity and its future.

Further, we observe the diversification effect, using actual intraday data provided by Nikkei Quick Information Co. Ltd. Empirical analysis in the rest of this chapter uses minute-by-minute data from the Tokyo Stock Exchange in April 2000. σ is estimated as the standard deviation of log return, and a as the regression coefficient between excess buy orders (amounts of trades at ask prices minus those at bid prices) and log return. Figure 3.1 shows how variable cost changes according to the correlation coefficient. Specifically, we used the average market impact coefficients and volatility among two very different sectors: electricity, a typical value sector for stock 1 and service, a typical growth sector for stock 2 in order to see the diversification effect ($V_1 = 100, V_2 = 150, a_1 = 0.0002, a_2 = 0.0003, \sigma_1 = 0.001, \sigma_2 = 0.0015, \lambda = 1$). Due to the diversification, when we construct the optimal execution strategy, considering portfolio effect, transaction cost becomes smaller than that when we optimize each stock separately. Although this is an extreme sample with two very different sectors, we also see similar diversification effect even if the parameters of each stocks are the same. Therefore, we can expect the diversification effect does not depend on parameters.
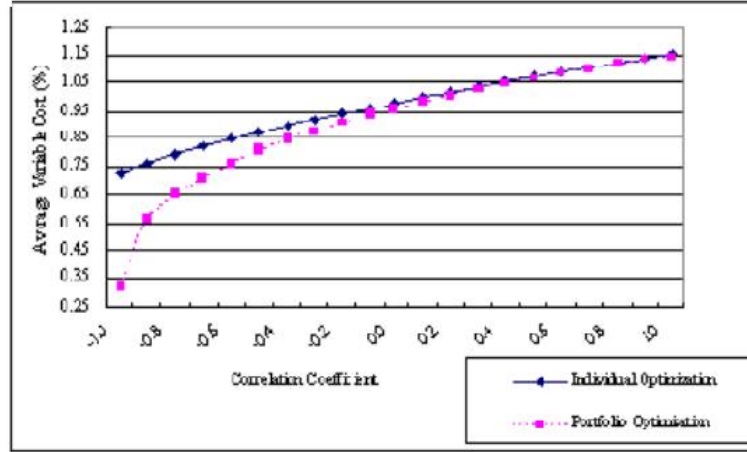
Figure 3.1: **Correlation coefficient and variable cost in two asset case.**   This graph shows how variable cost (y-axis) changes according to the correlation coefficient (x-axis), using market impact coefficients and volatility of typical two stocks. Due to diversification effect, simultaneous optimization reduces cost than iterative optimization ($V_1 = 100$, $V_2 = 150$, $a_1 = 0.0002$, $a_2 = 0.0003$, $\sigma_1 = 0.001$, $\sigma_2 = 0.0015$, $\lambda = 1$).

3.4   Limit to Instantaneous Executable Volume

1) Solution with upper limit

So far, we have assumed market impact is linear. However, in practice, this does not always hold without any restrictions, but only within moderate trading volume and time. Therefore, in this section, we consider the effect of limit to the volume executable instantaneously. For simplicity, we study only

single asset case. Without limit, recall that the optimal execution strategy as well as average variable cost have been obtained in Proposition 3.3.1.

Now, assume that there is an upper limit V_0 to the volume executable instantaneously. The solution in the above violates the limit when

$$v(0) = \left(\frac{\lambda^2 \sigma^2 V}{2a^2} \right)^{1/3} > V_0, \qquad \text{or equivalently,} \qquad V > \frac{2a^2 V_0^3}{\lambda^2 \sigma^2}. \tag{3.12}$$

If this is the case, it seems difficult to get an explicit form of the optimal solution. However, since the price movement risk is an increasing function of time while market impact coefficient remains constant, $v(t)$ is considered to be a decreasing function of time. Therefore, the optimal execution strategy is expected to take a constant value V_0 on $[0, t_0]$ and then decrease monotonically to approach zero asymptotically. Thus, it is natural to approximate the optimal execution strategy as

$$v_0(t) = \begin{cases} V_0, & 0 \le t < t_0, \\ V_0 e^{-\beta(t-t_0)}, & t_0 \le t. \end{cases} \tag{3.13}$$

From the constraint (3.1),

$$V = \int_0^\infty v_0(t)dt = V_0 \left(t_0 + \frac{1}{\beta} \right)$$

so that the relation

$$\beta = \frac{V_0}{V - V_0 t_0} \tag{3.14}$$

holds. The following proposition identifies optimal t_0 (and hence β) for this approximation.

Proposition 3.4.1 *Within function class that satisfies (3.13) and (3.14), the optimal t_0 is given as a unique solution of*

$$(t_0 - \gamma)^4 + \frac{2a^2 V_0^2}{3\lambda^2 \sigma^2} \left\{ (t_0 - \gamma)^3 - 2\gamma^3 \right\} = 0, \quad t_0 \in (0, \gamma); \qquad \gamma = \frac{V}{V_0}. \tag{3.15}$$

Proof.   See Section 3.7.   □

In order to see the precision of the approximation, we compare the numerical solution and this approximation with real market data. Figure 3.2 shows the optimal execution strategy of stock 1 in the Section 3.3. The limit to instantaneous execution volume is assumed to be 10 units since the average trading volume was 32 units in that period, and we regard market share over 30% is too much ($\sigma = 0.001$, $a = 0.0002$, $\lambda = 1$, $V = 200$, $V_0 = 10$). Figure 3.2 supports this approximation.

For multiple asset case, we can expect an exponential function as we did in the one asset case. However, the formula of opportunity cost depends on the integration range, and so we need to calculate case by case according to the time when the execution volume of each asset becomes variable.
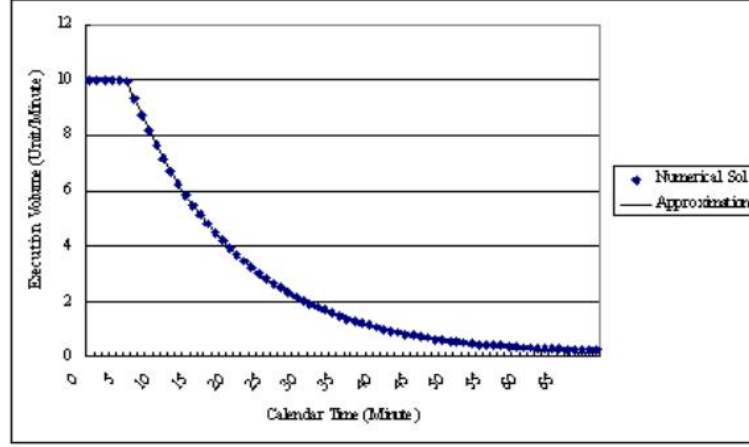
2) Numerical examples

Figure 3.2: **Numerical solution and approximation for the case with upper limit.**   Instantaneous execution volume of numerical solution (dashed curve) and approximation (solid curve) of the optimal execution strategy of the stock in the Section 3.3 is on y-axis, and time is on x-axis. The limit to instantaneous execution volume is assumed to be 10. Approximation is good enough to be almost indistinguishable with numerical solution. ($\sigma = 0.001$, $a = 0.0002$, $\lambda = 1$, $V = 200$, $v_0 = 10$).

It is difficult to identify implications on sensitivity of parameters and constraints from the optimal strategies since they require numerical calculation. So, we study the characteristics of the optimal execution strategy of a single asset case. Specifically, using parameters of stock 1 in Section 3.3, we observe how average variable cost changes according to whole trading size, and compare each solution. In this comparison, we also show the behavior of optimal execution strategy in which orders are executed at a constant rate (constant execution henceforth) presented in the following proposition.

Proposition 3.4.2   *Within constant executions, the optimal execution volume v_c is*

$$v_c = \begin{cases} \left(\frac{\lambda^2 \sigma^2 V}{12a^2}\right)^{1/3}, & V \leq \frac{12a^2 V_0^3}{\lambda^2 \sigma^2}, \\ V_0, & V > \frac{12a^2 V_0^3}{\lambda^2 \sigma^2} \end{cases}$$

where, as before, V_0 is the upper limit to the volume executable instantaneously. In this case, the average variable cost is

$$\frac{VC}{V} = \begin{cases} \left(\frac{9}{4}a\lambda^2\sigma^2 V\right)^{1/3}, & V \leq \frac{12a^2 V_0^3}{\lambda^2 \sigma^2}, \\ \left(\frac{\lambda^2 \sigma^2 V}{3V_0}\right)^{1/2} + aV_0, & V > \frac{12a^2 V_0^3}{\lambda^2 \sigma^2}. \end{cases}$$

Proof.   See Section 3.7. $\square$

Comparing with Proposition 3.3.1, the average variable cost under constant execution looks quite similar to that of the general optimal execution strategy (general execution henceforth) for small trading size, but the size is $(4/3)^{1/3}$ times larger. Figure 3.3 and Figure 3.4 show the relationship with whole trading

size and average variable cost, using parameters of stock 1 in Section 3.3 ($\sigma = 0.001$, $a = 0.0002$, $\lambda = 1$, $V_0 = 10$).
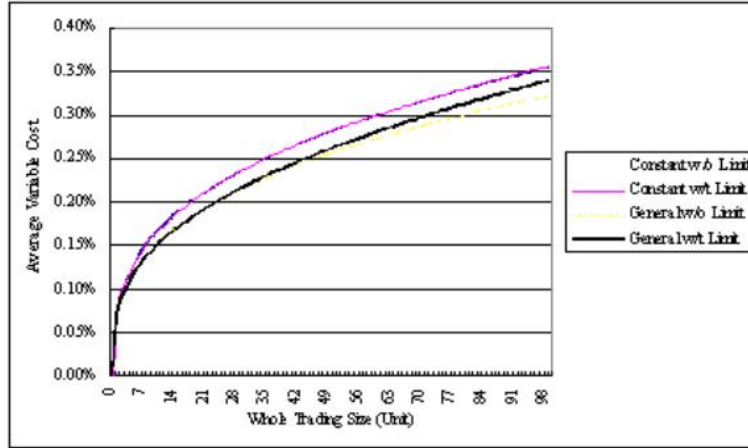

Figure 3.3: **Whole trading size and average variable cost (small trades).** Average variable cost (y-axis) is an increasing function of whole trading size (x-axis), while the marginal increment is a decreasing function, which are consistent with intuition. Constant execution with limit coincides with constant execution without limit for small trades. The same parameters of the stock in Section 3.3 are used ($\sigma = 0.001$, $a = 0.0002$, $\lambda = 1$, $V_0 = 10$).

In Figure 3.3, as long as optimal execution volume is below the limit, average variable costs with and without limit coincide and increase as whole trading size to the power of 1/3.

As whole trading size grows in Figure 3.4, average variable cost of general execution with limit departs from that of general execution without limit and approaches that of constant execution with limit, which increases as whole trading size to the power of 1/2.

Further, it is of great interest to compare our optimal execution strategy with those in previous studies such as Almgren and Chriss (1999). To do so, using parameters on Section 4.4 in Almgren and Chriss (1999), we calculate optimal strategies of both ours and theirs. They assumed whole trading size V as 10^6 shares, daily volatility σ as $0.95(\$/\text{share})/\text{day}^{1/2}$, temporary market impact coefficient a as $2.5 \times 10^{-6}(\$/\text{share})(\text{share}/\text{day})$, and risk premium λ as 1.645 for VaR-model and $10^{-6}/\$$ for Variance-model. Also, the number of time periods was set to five. We neglect other parameters for simplicity. Optimal execution strategies are shown in Figure 3.5.

We can see that, in optimal execution strategy of Almgren and Chriss (1999), trades which would have best fitted periods after 6 are somewhat awkwardly pushed into period $3 - 5$ since the number of time periods is arbitrarily set to five. Further, average variable cost is \$2.18 for our strategy and \$2.60 for that of Almgren and Chriss (1999), and therefore, we can confirm that our model indeed provides lower average cost.

In practice, we suffer statistical errors in estimating parameters. However, empirical studies such as

Figure 3.4: **Whole trading size and average variable cost (large trades).**   Same graph as Figure 3.3, but for large trades. As whole trading size grows, general execution with limit departs from general execution without limit and approaches constant execution with limit.

Figure 3.5: **Comparison of optimal execution strategy.**   Daily execution volume is on y-axis, and calendar time is on x-axis. We can see that, in strategy of Almgren and Chriss (1999), trades which would have best fitted periods after 6 are somewhat awkwardly pushed into period 3–5 since the number of time periods is arbitrarily set to five ($V = 10^6$ shares, $\sigma = 0.95$(dollar/share)/day$^{1/2}$, $a = 2.5 \times 10^{-6}$(dollar/share)/day$^{1/2}$, $\lambda = 1.645$).

Uno and Yamada (1993) reports that the market impact coefficients and the price volatilities are rather stable. Also, it is more realistic to regard parameters stochastic. However, according to the analysis by Hisata and Yamai (2000) on constant execution, this change does not affect the result significantly.

3.5 Principle of Optimality

In general, optimal policy satisfies the "principle of optimality," which means that the remaining decision must be the same as the decision taken at the beginning. However, if we recalculate the optimal execution strategy in Proposition 3.3.2 at halfway, the result differs from the strategy calculated at the beginning. This is because VaR is not additive regarding time, or standard deviation of price movement is not additive. This problem might be resolved by introducing a new risk measure although it is often difficult to provide such measure with an exact economical meaning. For example, standard deviation type measure which ignores historical profit and loss can be defined as

$$VC = a \int_0^\infty x'(t)^2 dt + \lambda \sigma \int_0^\infty x(t) dt.$$

Here, by Euler's Equation of the calculus of variations,

$$-2ax''(t) + \lambda\sigma = 0.$$

So, with the initial condition $x(0) = V$,

$$v(t) = -x'(t) = \begin{cases} -\frac{\lambda\sigma}{2a}t + \sqrt{\frac{\lambda\sigma V}{a}}, & 0 \le t < \sqrt{\frac{4aV}{\lambda\sigma}}, \\ 0, & \sqrt{\frac{4aV}{\lambda\sigma}} \le t. \end{cases}$$

We can see that this function satisfies the principle of optimality. The average variable cost is,

$$\frac{VC}{V} = \frac{4}{3}\sqrt{a\lambda\sigma V}.$$

Further, if there is a limit V_0 to instantaneous execution volume,

$$v(t) = \begin{cases} V_0, & 0 \le t < \frac{V}{V_0} - \frac{aV_0}{\lambda\sigma}, \\ -\frac{\lambda\sigma}{2a}t + \frac{\lambda\sigma V}{2aV_0} + \frac{V_0}{2}, & \frac{V}{V_0} - \frac{aV_0}{\lambda\sigma} \le t < \frac{V}{V_0} + \frac{aV_0}{\lambda\sigma}, \\ 0, & \frac{V}{V_0} + \frac{aV_0}{\lambda\sigma} \le t \end{cases}$$

and

$$\frac{VC}{V} = \frac{\lambda\sigma V}{2V_0} + aV_0 - \frac{a^2 V_0^3}{6\lambda\sigma V}.$$

This value increases almost linearly as V increases, which still contradicts the empirical results of BARRA (1997) and Mannen and Uno (2000).

3.6   Closing Remarks

This chapter has studied execution scheduling of liquidation of a block trade and formulated it as a static optimization problem. Based on empirical studies on market microstructure, we have distinguished between temporary market impact and permanent market impact, and assumed that each component to be a linear function of trading volume within moderate volume and time. Also, we have defined the measure of opportunity cost as standard deviation of price movement, and minimized transaction cost, sum of execution cost and opportunity cost. Under this framework, we have obtained explicit solutions both for the single-asset case and a multiple-asset portfolio model in the form of pure exponentials. Further, we have derived an approximation formula when there is a limit to the trading volume executable instantaneously.

As a result, we have shown that the optimal execution strategy is independent of permanent impact, that optimal execution duration is an increasing function of whole trading size, and that the optimal average cost is an increasing function of whole trading size, price volatility, and temporary impact, and a decreasing function of the limit to instantaneous execution volume, which are all consistent with trading practice and empirical studies. Specifically, optimal average cost increases as whole trading size to the power of $1/3$ for small orders and $1/2$ for large orders. Further, it is shown that the transaction cost can be lowered by making order execution late when price movement risk of a portfolio is well diversified.

Our results are immediately applicable to existing risk management framework and provides explicit solution for minimum L-VaR. Also, although this chapter uses an example of liquidation of a block of common stock, results can be extended over wider problems on liquidation of any large block of assets.

While this chapter has focused on static optimization, decision change in accordance with new information might be handled by dynamic programming. Also, this model can reflect more reality by making volatility and market impact time dependent. Further, it is an interesting issue how to relate order slicing strategy and choice between market order and limit order, and how to control information effect of orders. These issues are left for further analysis.

3.7   Appendix

3.7.1   Proof of Lemma 3.3.2

Since $\boldsymbol{S}$ is non-negative definite, $E \geq 0$ is clear. Thus, it is enough to prove $EG \geq F^2$. Non-negative definiteness of $\boldsymbol{S}$ also implies each eigenvalue ξ_i $(i = 1, \cdots, N)$ of $\boldsymbol{S}$ is non-negative (c.f., Bellman (1997), Ch.4.8, Theorem 3). This and symmetricity of $\boldsymbol{S}$ ensure the existence of an orthogonal matrix $\boldsymbol{H}$ for which $\boldsymbol{S} = \boldsymbol{H}\mathrm{Diag}(\xi_1, \cdots, \xi_N)\boldsymbol{H}^\top$ holds. Therefore, if we define $\boldsymbol{Q} = \boldsymbol{H}\sqrt{\boldsymbol{J}}\boldsymbol{H}^\top$ with

$\sqrt{\boldsymbol{J}} = \mathrm{Diag}(\sqrt{\xi_1}, \cdots, \sqrt{\xi_N})$, then $\boldsymbol{Q}$ is symmetric and $\boldsymbol{Q}^2 = \boldsymbol{S}$. Let $\boldsymbol{u}(t) = \boldsymbol{Q}\boldsymbol{x}(t)$ and $\boldsymbol{w}(t) = \boldsymbol{Q}\boldsymbol{y}(t)$. Then, we get

$$\boldsymbol{x}(t)^\top \boldsymbol{S}\boldsymbol{x}(t) = \boldsymbol{u}(t)^\top \boldsymbol{u}(t), \qquad \boldsymbol{y}(t)^\top \boldsymbol{S}\boldsymbol{y}(t) = \boldsymbol{w}(t)^\top \boldsymbol{w}(t), \qquad \boldsymbol{x}(t)^\top \boldsymbol{S}\boldsymbol{y}(t) = \boldsymbol{u}(t)^\top \boldsymbol{w}(t).$$

By Cauchy-Schwartz's inequality (cf., Kolmogorov and Fomin (1970)),

$$
\begin{aligned}
\left(\int_0^\infty \boldsymbol{x}(t)^\top \boldsymbol{S}\boldsymbol{x}(t)dt \right) \left(\int_0^\infty \boldsymbol{y}(t)^\top \boldsymbol{S}\boldsymbol{y}(t)dt \right) &= \left(\int_0^\infty \boldsymbol{u}(t)^\top \boldsymbol{u}(t)dt \right) \left(\int_0^\infty \boldsymbol{w}(t)^\top \boldsymbol{w}(t)dt \right) \\
&\geq \left(\int_0^\infty \boldsymbol{u}(t)^\top \boldsymbol{w}(t)dt \right)^2 \\
&= \left(\int_0^\infty \boldsymbol{x}(t)^\top \boldsymbol{S}\boldsymbol{y}(t)dt \right)^2,
\end{aligned}
$$

completing the proof.

3.7.2 Proof of Proposition 3.4.1

For notational simplicity, we calculate VC/V_0^2 instead of VC. Let $\gamma = V/V_0$. Then, from (3.13) and (3.14),

$$\int_0^\infty \frac{v_0(t)^2}{V_0^2} dt = \left[\int_0^{t_0} dt + \int_{t_0}^\infty e^{-2\beta(t-t_0)} dt \right] = \frac{t_0 + \gamma}{2}.$$

Further, if we let $x_0(t) = \int_t^\infty v_0(s)ds$,

$$\frac{x_0(t)}{V_0} = \begin{cases} \gamma - t, & 0 \leq t < t_0, \\ \beta^{-1} e^{-\beta(t-t_0)}, & t_0 \leq t. \end{cases}$$

So,

$$\int_0^\infty \frac{x_0(t)^2}{V_0^2} dt = \int_0^{t_0} (t-\gamma)^2 dt + \int_{t_0}^\infty \beta^{-2} e^{-2\beta(t-t_0)} dt = -\frac{1}{6}(t_0 - \gamma)^3 + \frac{1}{3}\gamma^3.$$

Therefore,

$$\frac{VC}{V_0^2} = \frac{a(t_0 + \gamma)}{2} + \frac{\lambda\sigma}{\sqrt{6}\, V_0} \sqrt{-(t_0 - \gamma)^3 + 2\gamma^3} =: f(t_0). \tag{3.16}$$

Note that the range of $f(t_0)$ is $t_0 \in (0, \gamma)$ from (3.13). Since the first term of $f(t_0)$ linearly increases as t_0 increases while the second term is a convex decreasing function, $f(t_0)$ is convex. Moreover,

$$f'(t_0) = \frac{1}{4} \left\{ 2a - \frac{\lambda\sigma}{\sqrt{6}\, V_0} \frac{(t_0 - \gamma)^2}{\sqrt{-(t_0 - \gamma)^3 + 2\gamma^3}} \right\}.$$

Then,

$$f'(0) = \frac{1}{4} \left(2a - \frac{\sqrt{2\gamma}\lambda\sigma}{V_0} \right) < 0, \qquad \text{(from (3.12))}$$

$$f'(\gamma) = \frac{a}{2} > 0.$$

Summarizing these arguments, we conclude that t_0 that minimizes VC/V_0^2 is given as a unique solution of $f'(t_0) = 0$, $t_0 \in (0, \gamma)$. Rearranging terms, the desired result is obtained.

3.7.3 Proof of Proposition 3.4.2

Consider a constant execution in which orders are executed at a constant volume v_c until time t_0. According to the constraint in (3.1), v_c and t_0 satisfy

$$\int_0^{t_0} v_c dt = v_c t_0 = V, \qquad \text{or} \qquad v_c = \frac{V}{t_0}. \tag{3.17}$$

Noting that

$$x(t) = \begin{cases} (t_0 - t)v_c, & 0 \le t \le t_0, \\ 0, & t_0 < t, \end{cases}$$

the total cost can be calculated as

$$VC = a\frac{V^2}{t_0} + \lambda \sigma V \sqrt{\frac{t_0}{3}}.$$

Therefore, the optimal t_0 that minimized the variable cost is

$$t_0 = \left(\frac{12a^2 V^2}{\lambda^2 \sigma^2}\right)^{1/3},$$

and, from (3.17), the optimal execution volume is

$$v_c = \left(\frac{\lambda^2 \sigma^2 V}{12a^2}\right)^{1/3}.$$

Also, average variable cost becomes

$$\frac{VC}{V} = \left(\frac{9}{4}\lambda^2 \sigma^2 aV\right)^{1/3}.$$

When there is a limit V_0 to instantaneous execution volume, t_0 needs to satisfy $t_0 \ge V/V_0$. In summary, the optimal execution period is given as

$$t_0 = \max\left(\left(\frac{12a^2 V^2}{\lambda^2 \sigma^2}\right)^{1/3}, \frac{V}{V_0}\right)$$

and the desired results can be obtained by a straightforward calculation.

Chapter 4

Selection of Market and Limit Order

Abstract   This chapter analyses the optimal selection of market and limit orders in a series of single price batch auctions when the expectation of the limit order book is given. We derive the analytical solution of a single limit order model for risk neutral traders, and then extend the result for risk averse traders. We find that the limit order size is independent of the whole trade size while the limit order price and the market order size are linear functions and that market orders replace limit orders as time passes. Further, we calculate economic value of monitoring limit order book.

4.1   Introduction

Market and limit orders are two essential instruments of order placement. Progress in information technology has improved order processing capability and made order placement strategy a critical issue in investment. For example, order driven markets such as the Tokyo Stock Exchange (TSE) have excluded floor traders and adopted an automatic matching system, which makes order processing faster and more certain. Also, several markets such as the New York Stock Exchange (NYSE) and the National Association of Securities Dealers Automated Quotation System (NASDAQ) start receiving limit orders to broaden traders' freehand for order placement. Further, emergence of alternative trading systems such as crossing networks and electronic communication networks enables traders to tactically submit and cancel orders across markets. In practice, traders recognize the advantage of limit orders: cost efficiency, and disadvantages: uncertain execution and free option, and choose either market or limit order or sometimes both as fits their objectives and circumstances.

Several empirical studies confirm traders' behavior in the real market. For example, Biais et al. (1995) analyzed trading data from the Paris Stock Exchange and find that traders actively respond to public orders (orders submitted by other traders). Specifically, market orders are preferred when bid-ask spread is small because the execution probability of limit order is low, while limit orders are preferred when

bid-ask spread is large. Also, Harris and Hasbrouck (1996) analyzed data from NYSE and found that 60% of limit orders which lay inside the best bid and ask prices were executed. They show that the trading cost can be lowered by limit orders especially when bid-ask spread is large, which is consistent with the result of Biais et al. (1995). Regarding Japanese market, Kawahara (1994) analyzes TSE stock data in 1994 and finds that trading cost with limit orders is 0.15% lower than that with only market orders. Also, trading cost turns out to be larger for sell orders, large orders, and limit orders that are exposed to the market at shorter period of time.

In contrast, theoretical studies are limited for the selection of market and limit orders. For example, although Bertisimas and Lo (1998) and Konishi and Makimoto (2001) derive the optimal order slicing strategy in a large portfolio liquidation, only market orders but no limit orders are allowed in their models. Also, Parlour (1999) analyses how market conditions affect the selection of market and limit orders when a trader places just one unit of order. This analysis shows the interaction between trader's order placing strategy and market conditions, but fails to study how traders split orders for large trades. Further, Chakravarty and Holden (1995) provide a theoretical model which explains how traders select market and limit orders in a quote driven market. However, existence of a market maker is crucial in their model, and the results are not directly applicable to an order driven market.

Accordingly, this chapter analyzes the optimal selection of market and limit orders, sizes, prices, and times in a series of single price batch auctions. Specifically, we derive the analytical solution of a single limit order model for risk neutral traders, and then extend the result for risk averse traders. Regarding the execution at each period, we find that the optimal limit order size is independent of the whole trade size so as to balance the non-linearity of execution volume and price. In contrast, the limit order price and the market order size are linear functions in order to balance trading costs across trading periods. Regarding the execution throughout the trading session, market orders replace limit orders as time passes, and necessity of completing execution grows. Further, if trading volume and price volatility are larger at the opening and closing of the trading session as in the practical markets, more limit orders are tried at the opening and closing. These results are consistent with intuitive trading behaviors. Further, although most traders know the public order book as expectations as in our model, member security brokers are allowed to monitor the public order book without time lags. So, we evaluate the value of monitoring public order books and provide criteria for selecting stocks to monitor.

The rest of this chapter is organized as follows. First, by dynamic programming, Section 4.2 derives the analytical solution of a single limit order model for risk neutral traders. Then, Section 4.3 shows several variations of the model in Section 4.2 for risk averse traders and when public market orders are known. Further, Section 4.4 analyzes a numerical solution of a multiple limit order model. Finally, section 4.5 concludes the analysis.

4.2 Standard Single Limit Order Model

4.2.1 Formulization of the Strategy

Imagine a case in which a trader buys a certain number of shares R_1 in a series of T single price batch auctions. The trader tries to minimize the expected value of the purchase price with market and limit orders. We assume that the trader has just the expectation of the public limit orders and knows the distribution of the public market orders but not concrete order sizes since it is costly or prohibited for most traders to monitor public order books. Lehman and Modest (1994) find that stocks with low liquidity tend to be traded at *Itayose*, single price batch auctions at the opening and the closing of the trading sessions on the TSE. Although our model is applicable to any stock it has a better fit with low liquidity stocks. We make the following assumptions about market environments, which are given exogenously.

Assumption 4.2.1

1. *Price change is driven by two uncertainties, change of fundamental value, Q_i, and size of public market orders, M_i. Q_i is given as*

$$Q_i = Q_1 + \sum_{j=2}^{i} \delta_j \epsilon_j$$

   where Q_1 and δ_j ($j = 2, \cdots, T$) are constants and ϵ_j ($j = 2, \cdots, T$) are serially uncorrelated random variables that follow N(0,1). M_i are serially uncorrelated random variables that follow $N(0, \sigma_i^2)$ and are independent of ϵ_i ($i = 2, \cdots, T$) also.

2. *Market impact at the ith period P_i is a linear function of execution volume with a constant market impact coefficient a_i, and r ($0 \leq r \leq 1$) of market impact is passed to subsequent periods as the permanent impact.*

Regarding the price change, a positive (negative) number of M_i represents a buy (sell) market order, respectively. The assumption that M_i is serially uncorrelated and independent of Q_i is natural since market orders by the uninformed liquidity traders tend to be numerous and small, and therefore, excess market orders follow independent normal distribution by the Central Limit Theorem.

Regarding market impact, we can justify the linear function assumption both empirically and theoretically. From the empirical point of view, Jain and Joh (1988) and Holthausen et al. (1987) observe that the market impacts are rather stable linear functions and can be divided into temporary and permanent components. From the theoretical point of view, linear market impact is realized when informed traders submit limit orders while uninformed liquidity traders submit market orders. Specifically, assume that risk averse informed traders try to take advantage of temporary market impact as contrarians and submit limit orders. Let λ denote traders' risk premium, P execution price, and V execution size, and

the traders' objective function can be written as

$$U = (1 - r)PV - \lambda V^2$$

where r is the temporary market impact ratio as defined in Assumption 4.2.1.

So, the optimal position size $\overline{V}$ satisfies

$$\left.\frac{\partial U}{\partial V}\right|_{V=\overline{V}} = (1 - r)P - 2\lambda\overline{V} = 0.$$

And therefore,

$$\overline{V} = \frac{(1 - r)P}{2\lambda}.$$

This is attained by submitting limit orders with equal size for any prices, since

$$\frac{\partial \overline{V}}{\partial P} = \frac{(1 - r)}{2\lambda}.$$

Further, Bertisimas and Lo's (1998) analysis on optimal execution without limit orders assumes linear market impact and randomness with normal distribution. Therefore, we can evaluate economical value of limit orders by comparing with our results and those of Bertisimas and Lo (1998).

In a single price batch auction, market orders, buy limit orders at higher prices than the execution price, and sell limit orders at lower prices than the execution price are executed. Market impact P_1 is determined so as to balance total buy orders and sell orders. Therefore, linear market impact is realized by public limit orders. Specifically, $Q_i + r\sum_{j=1}^{i-1} P_j$ is passed to the subsequent periods since the ratio of permanent impact is r. Therefore, we define base price at the ith period as

$$B_i = \begin{cases} Q_1, & i = 1, \\ Q_i + r\sum_{j=1}^{i-1} P_j, & i = 2, \cdots, T. \end{cases} \tag{4.1}$$

Since market impact coefficient in the ith period is a_i, $1/a_i$ of public sell limit orders exist at any prices above B_i, and $1/a_i$ of public buy limit orders exist at any prices below B_i.

Regarding the trader's strategy, we assume that the trader knows constant parameters and Q_i but not M_i when he decides the ith period strategy. Also, we assume that the trader can submit only one market and limit order for simplicity. As a result, our model has three controllable variables: the trader's market order size u_i, sum of the trader's market and limit order sizes v_i, and deviation of limit order price from the base price p_i, which we call a base limit order price. Note that the limit order size is given by $w_i \overset{\text{def}}{=} v_i - u_i$ and that actual limit order price is $B_i + p_i$. We allow u_i and p_i to take negative values while the size of limit order w_i must not be negative because negative w_i is just a cancellation of a buy limit order but not equivalent with a sell limit order. Instead, without loss of generality, we can regard limit order as "buy" because a sell limit order can be synthesized by a combination of a sell market order

and a buy limit order. Note that all the remaining orders at the last period are submitted as market orders in order to complete the execution. These assumptions are summarized as follows.

Assumption 4.2.2 *Execution strategy is specified as $\boldsymbol{\psi} = \{\psi_i; i = 1, \cdots, T\}$ where $\psi_i = (u_i, v_i, p_i)$, and satisfies the following conditions.*

1. $u_i \leq v_i$ $(i = 1, \cdots, T)$.

2. ψ_i is $\mathcal{G}_i$-measurable where $\mathcal{G}_i = \sigma\{\epsilon_2, \cdots, \epsilon_i, M_1, \cdots, M_{i-1}\}$, the smallest σ-algebra with respect to which $\epsilon_2, \cdots, \epsilon_i$ and $M_1, \cdots, M_{i-1}$ are measurable.

3. Only market order is allowed at the Tth period.

Once ψ_i is chosen and M_i is realized, the execution price $B_i + P_i$ and the execution volume D_i are determined uniquely as a function $P_i = \pi_i(u_i, v_i, p_i, M_i)$ and $D_i = \kappa_i(u_i, v_i, p_i, M_i)$. Specifically, if $p_i > 0$, since volume of public sell limit orders below $B_i + p_i$ is p_i/a_i, trader's limit order starts being executed when the sum of public and the trader's market order (the total market order henceforth) is smaller than p_i/a_i, and is completely executed when the total market order is smaller than $-w_i + p_i/a_i$. Next, if $p_i \leq 0$, since volume of public buy limit orders above $B_i + p_i$ is $-p_i/a_i$, trader's limit order starts being executed when the total market order is smaller than p_i/a_i, and is completely executed when the total market order is smaller than $-w_i + p_i/a_i$. Consequently,

$$\kappa_i(u, v, p, m) = \begin{cases} u, & -u + p/a_i < m, \\ -m + p/a_i, & -v + p/a_i < m \leq -u + p/a_i, \\ v, & m \leq -v + p/a_i. \end{cases} \tag{4.2}$$

Similar arguments show that

$$\pi_i(u, v, p, m) = \begin{cases} a_i(u + m), & -u + p/a_i < m, \\ p, & -v + p/a_i < m \leq -u + p/a_i, \\ a_i(v + m), & m \leq -v + p/a_i. \end{cases} \tag{4.3}$$

If there are no limit orders, we define $u_i = v_i$, $p_i = -\infty$.

There are two non-linearities in (4.2) and (4.3). The first is the non-linearity of execution volume since limit orders are not executed when the execution price is higher than the limit order price. The advantage of non-linearity of execution volume is more significant for larger limit orders. The second non-linearity is that of execution price. Execution price becomes higher because of the trader's limit order, which makes the execution price like a "vertical spread", combination of buy and sell of put options. The disadvantage of non-linearity of execution price is more significant for larger limit orders. Therefore, it is expected that the optimal limit order size balances these two non-linearities.

In summary of Assumptions 4.2.1 and 4.2.2, the execution at the ith period proceeds as follows:

1. Base price B_i is given by (4.1).

2. $1/a_i$ of public sell limit orders are submitted at any prices above B_i, and public buy limit orders below B_i.

3. The trader selects u_i, v_i, and $B_i + p_i$.

4. Volume of the public market order $m_i = M_i$ is realized according to $N(0, \sigma_i^2)$, and the execution price and the execution volume is set to $B_i + P_i$ and D_i by (4.2) and (4.3), respectively.

Let $R_i \stackrel{\text{def}}{=} \sum_{j=i}^{T} D_j = R_1 - \sum_{j=1}^{i-1} D_j$ denote the size of trades left at the ith period. Then, the total purchase cost is given as

$$
\begin{aligned}
(\text{Total Cost}) &= \sum_{i=1}^{T} \left(Q_i + r \sum_{j=1}^{i-1} P_j + P_i \right) D_i \\
&= \sum_{i=1}^{T} \left(Q_1 + \sum_{j=2}^{i} \delta_j \epsilon_j + r \sum_{j=1}^{i-1} P_j + P_i \right) D_i \\
&= Q_1 \sum_{i=1}^{T} D_i + \sum_{i=2}^{T} \sum_{j=2}^{i} \delta_j \epsilon_j D_i + r \sum_{i=1}^{T} \sum_{j=1}^{i-1} P_j D_i + \sum_{i=1}^{T} P_i D_i \\
&= Q_1 R_1 + \sum_{j=2}^{T} \delta_j \epsilon_j \sum_{i=j}^{T} D_i + r \sum_{j=1}^{T-1} P_j \sum_{i=j+1}^{T} D_i + \sum_{i=1}^{T} P_i D_i \\
&= Q_1 R_1 + \sum_{j=2}^{T} \delta_j \epsilon_j R_j + \sum_{i=1}^{T-1} r P_i (R_i - D_i) + \sum_{i=1}^{T} P_i D_i \\
&= Q_1 R_1 + \sum_{j=2}^{T} \delta_j \epsilon_j R_j + \sum_{i=1}^{T} P_i \{(1-r)D_i + rR_i\}.
\end{aligned}
$$

If we define $\mathcal{F}_i \stackrel{\text{def}}{=} \sigma\{\epsilon_2, \cdots, \epsilon_i, M_1, \cdots, M_i\}$, R_j is $\mathcal{F}_{j-1}$-measurable by Assumption 4.2.2. Since ϵ_j and R_j are independent, the expected total purchase cost is

$$
\text{E}\left((\text{Total Cost})\right) = Q_1 R_1 + \sum_{i=1}^{T} \text{E}\left(P_i\{(1-r)D_i + rR_i\}\right).
$$

Therefore, our problem can be written as

$$
\text{minimize} \quad \sum_{i=1}^{T} \text{E}\left(P_i\{(1-r)D_i + rR_i\}\right). \tag{4.4}
$$

Define the cost at the ith period as

$$
\text{E}\left(P_i\{(1-r)D_i + rR_i\}\right) = \text{E}\left(\pi_i(\psi_i, M_i)\{(1-r)\kappa_i(\psi_i, M_i) + rR_i\}\right)
$$

where $\psi_i = (u_i, v_i, p_i)$, and $\pi_i(\psi_i, m)$ and $\kappa_i(\psi_i, m)$ stand for $\pi_i(u_i, v_i, p_i, m_i)$ and $\kappa_i(u_i, v_i, p_i, m_i)$ respectively for the simplicity of representaion.

Note that the accumulated market impact $r\sum_{j=1}^{i-1} P_j$ is recognized as the cost at the jth period but not at the ith period. Since ϵ_i $(i = 2, \cdots, T)$, M_i $(i = 1, \cdots, T)$ are serially uncorrelated and independent with each other, execution strategy at the ith period affects the cost after the $i + 1^{th}$ period only through R_{i+1}. Therefore, in order to search for an optimal execution strategy, all we have to consider is Markov policies in which ψ_i is determined solely by R_i.

Specifically, (4.4) is the Markov decision process whose state is represented by R_i $(i = 1, \cdots, T)$. The order selection rule is described by the policy $\psi_i = (u_i, v_i, p_i)$. Once R_i and ψ_i are given, the state at the $i + 1^{th}$ period is $R_i - \kappa_i(\psi_i, m)$ with transition density $f_i(m) = \dfrac{1}{\sqrt{2\pi}\sigma_i}\exp(-m^2/2\sigma_i^2)$. Let the superscript $*$ denote the optimal solution henceforth. Then, the optimal cost after the ith period with R_i is

$$C_i^*(R_i) = \min_{\boldsymbol{\psi}} \sum_{j=i}^{T} \mathrm{E}\left(P_j\{(1 - r)D_j + rR_j\} \mid R_i\right), \tag{4.5}$$

where $\boldsymbol{\psi}$ denotes the policy space. Bellman Equation for $i = 1, \cdots, T - 1$ is given as

$$
\begin{aligned}
C_i^*(R_i) &= \min_{\psi_i} \mathrm{E}\left(\pi_i(\psi_i, M_i)\{(1 - r)\kappa_i(\psi_i, M_i) + rR_i\} + C_{i+1}^*(R_i - \kappa_i(\psi_i, M_i))\right) \\
&= \min_{\psi_i} \int_{-\infty}^{\infty} \left[\pi_i(\psi_i, m)\{(1 - r)\kappa_i(\psi_i, m) + rR_i\} + C_{i+1}^*(R_i - \kappa_i(\psi_i, m))\right] f_i(m)dm. \tag{4.6}
\end{aligned}
$$

Since only market order is allowed at the Tth period, $C_T^*(R_T)$ can be calculated explicitly as

$$C_T^*(R_T) = a_T R_T^2. \tag{4.7}$$

Solving (4.6) backward with the initial condition (4.7), we can derive the optimal execution strategy at each period.

4.2.2 Derivation of the Optimal Strategy

We first consider two period case, and then extend it to multiple period case in Section 4.2.3 by Markov decision process. In order to complete the execution, all the remaining orders are submitted as market orders in the second period. In this case, since $\mathrm{E}\left(P_2 \mid R_2\right) = a_2 R_2$,

$$C_1(R_1) = \mathrm{E}\left(P_1\{(1 - r)D_1 + rR_1\} + a_2(R_1 - D_1)^2\right).$$

Further calculation shows the following proposition.

Proposition 4.2.1

$$
\begin{aligned}
C_1(R_1) =\ & \{(1 - r)a_1 + a_2\}v_{01}^2 + (ra_1 - 2a_2)R_1 v_{01} + a_2 R_1^2 \\
& - \left[\{(1 - r)a_1 + a_2\}(u_1 + \frac{p_1}{a_1}) + (ra_1 - 2a_2)R_1\right]\left\{(u_1 - \frac{p_1}{a_1})\Phi(-\frac{1}{\sigma_1}(u_1 - \frac{p_1}{a_1})) - \sigma_1\phi(-\frac{1}{\sigma_1}(u_1 - \frac{p_1}{a_1}))\right\} \\
& + \left[\{(1 - r)a_1 + a_2\}(v_1 + \frac{p_1}{a_1}) + (ra_1 - 2a_2)R_1\right]\left\{(v_1 - \frac{p_1}{a_1})\Phi(-\frac{1}{\sigma_1}(v_1 - \frac{p_1}{a_1})) - \sigma_1\phi(-\frac{1}{\sigma_1}(v_1 - \frac{p_1}{a_1}))\right\} \\
& + a_2\sigma_1^2\left\{\Phi(-\frac{1}{\sigma_1}(u_1 - \frac{p_1}{a_1})) - \Phi(-\frac{1}{\sigma_1}(v_1 - \frac{p_1}{a_1}))\right\}
\end{aligned}
$$

where $\phi(x) = \dfrac{1}{\sqrt{2\pi}} exp(-\dfrac{x^2}{2})$ *and* $\Phi(x) = \int_{-\infty}^{x} \phi(y)dy$.

Proof.   See Section 4.6.   □

Further, the following proposition holds for the corresponding optimal solution.

Proposition 4.2.2   *Optimal u_1^*, w_1^*, and p_1^* are uniquely determined to satisfy*

$$
\begin{cases}
p_1^* = -\dfrac{(ra_1 - 2a_2)a_1 R_1}{2\{(1-r)a_1 + a_2\}}, \\[3mm]
\dfrac{w_1^*}{2\sigma_1} = \dfrac{(1-r)a_1}{2\{(1-r)a_1 + a_2\}} \dfrac{\phi(\frac{w_1^*}{2\sigma_1})}{\Phi(-\frac{w_1^*}{2\sigma_1})}, \\[3mm]
u_1^* = \dfrac{p_1^*}{a_1} - \dfrac{w_1^*}{2}.
\end{cases}
\tag{4.8}
$$

Proof.   See Section 4.6.   □

Remark 4.2.1   *The second equation in (4.8) has a unique solution w_1^*. See Section 4.6 for detail.*

The first equation in (4.8) states that the base limit order price is chosen at the peak of the distribution of market order. This is because the value of the non-linearity of execution volume is maximum at the peak of the distribution. Besides, $\dfrac{w_1^*}{2\sigma_1}$ is a constant which depends only on a_1, a_2, and r. This is because a trader should not place large limit orders due to the non-linearity of execution price. Further, u_1^*, v_1^*, and p_1^* are linear functions of R_1.

Minimal $C_1^*(R_1)$ is given by

$$
C_1^*(R_1) = \frac{\{4a_2 - r^2 a_1\}a_1}{4\{(1-r)a_1 + a_2\}}R_1^2 - \frac{\{(1-r)a_1 + 2a_2\}(1-r)a_1\sigma_1^2\phi^2(\frac{w_1^*}{2\sigma_1})}{2\{(1-r)a_1 + a_2\}\Phi(-\frac{w_1^*}{2\sigma_1})} + a_2\sigma_1^2\{2\Phi(\frac{w_1^*}{2\sigma_1}) - 1\}.
$$

Minimal cost is a sum of the square of R_1 and a constant because the second and third term do not depend on R_1. Besides, the first term is equal to the minimal cost without limit orders. Therefore, the difference, sum of the second and third term, is the economical value of limit orders, which is negative and independent of R_1.

It seems that even if R_1 is zero, positive expected return can be made by placing small buy and sell limit order at price zero. This is partly because we ignore bid-ask spread, which makes market impact in the second period negligible. This approximation error might be significant if the trade is too small.

If sell orders are not allowed, all we have to do is to replace negative value of u_1^* and v_1^* with zero because of the positive second order condition of optimality. Figures 4.1 and 4.2 show the relationship between trade size and u_1^*, w_1^*, and p_1^* for both with and without non-negative constraints when $a_1 = a_2 = 1$, $\sigma_1 = 10$, and $r = 0$. We can see that limit orders are preferred in small trades even when sell orders are not allowed.
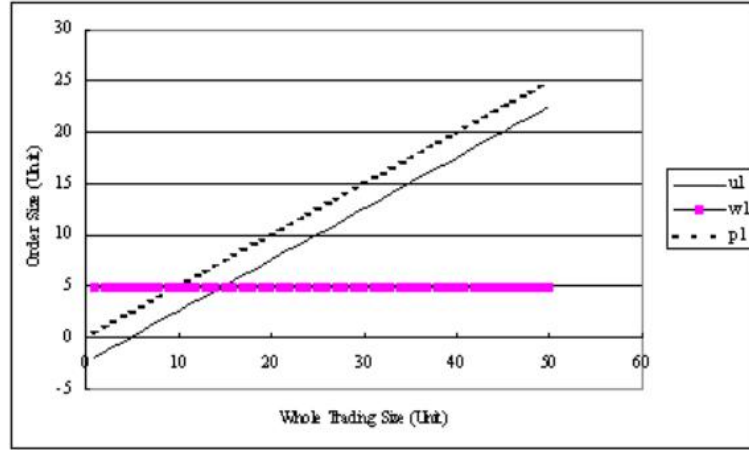
Figure 4.1: R_1 (*x*-axis) and u_1^*, w_1^*, and p_1^* (*y*-axis) **without constraints.** u_1^*, w_1^*, and p_1^* are linear functions of R_1.
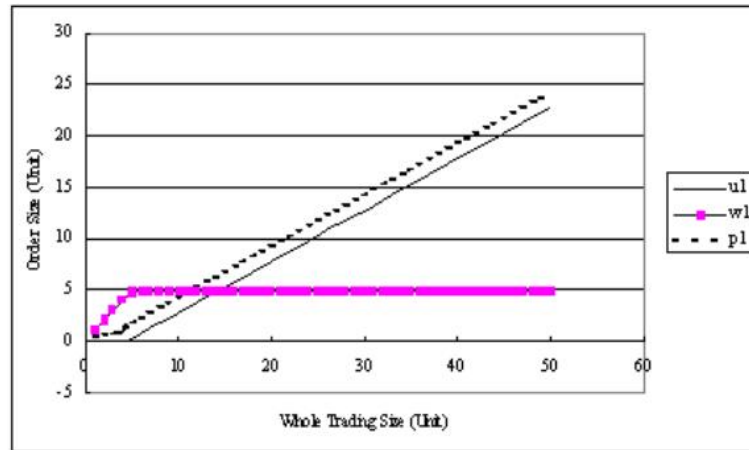

Figure 4.2: R_1 (*x*-axis) and u_1^*, w_1^*, and p_1^* (*y*-axis) **with non-negative constraint.** u_1^*, w_1^*, and p_1^* are linear functions of R_1 for large R_1.

4.2.3   Multiple Period Case

The results in two period case can be extended to T period case by Markov decision process since $C_1^*(R_1)$ is sum of square of R_1 and a constant. Specifically, immediate application of two period case provides $T-1^{th}$ period strategy as

$$
\begin{cases}
p_{T-1}^* = -\dfrac{(ra_{T-1} - 2a_T)a_{T-1}R_{T-1}}{2\{(1-r)a_{T-1} + a_T\}}, \\[2mm]
\dfrac{w_{T-1}^*}{2\sigma_{T-1}} = \dfrac{(1-r)a_{T-1}}{2\{(1-r)a_{T-1} + a_T\}}\dfrac{\phi(\frac{w_{T-1}^*}{2\sigma_{T-1}})}{\Phi(-\frac{w_{T-1}^*}{2\sigma_{T-1}})}, \\[2mm]
u_{T-1}^* = \dfrac{p_{T-1}^*}{a_{T-1}} - \dfrac{w_{T-1}^*}{2},
\end{cases}
$$

and the optimal cost is $C_{T-1}^*(R_{T-1}) = \dfrac{\{4a_T - r^2 a_{T-1}\}a_{T-1}}{4\{(1-r)a_{T-1} + a_T\}}R_{T-1}^2 + constant$. Since the optimal cost in the two period case is $a_2 R_2^2$, i^{th} period strategy ($i = T-2, \cdots, 1$) is calculated similarly as

$$
\begin{cases}
p_i^* = -\dfrac{(ra_i - 2A_i)a_i R_i}{2\{(1-r)a_i + A_i\}}, \\[2mm]
\dfrac{w_i^*}{2\sigma_i} = \dfrac{(1-r)a_i}{2\{(1-r)a_i + A_i\}}\dfrac{\phi(\frac{w_i^*}{2\sigma_i})}{\Phi(-\frac{w_i^*}{2\sigma_i})}, \\[2mm]
u_i^* = \dfrac{p_i^*}{a_i} - \dfrac{w_i^*}{2}
\end{cases}
$$

where A_i ($i = 1, \cdots, T-1$) is defined backward recursively as $A_i = \dfrac{\{4A_{i+1} + r^2 a_i\}a_i}{4\{(1-r)a_i + A_{i+1}\}}$ ($i = T-2, \cdots, 1$) with $A_{T-1} = \dfrac{\{4a_T + r^2 a_{T-1}\}a_{T-1}}{4\{(1-r)a_{T-1} + a_T\}}$.

Figure 4.3 shows a numerical example which illustrates changes of optimal market and limit order size over time. In this numerical example, we assume that market order, all the realized M_i turn out to be zero ($T = 11, R_1 = 100, r = 0, a_i = 1, \sigma_i = 10$ for $i = 1, \cdots, 11$).

We can see that more limit orders are tried at the early stage of execution, and they are gradually replaced with market orders if limit orders are not executed. Also, a case when a_i and σ_i depend on period is worth observing in order to analyze a whole day execution, since Jain and Joh (1988) find that a_i and σ_i are high at the opening and closing of a day. Figure 4.4 shows a numerical example in which $a_1 = a_{11} = 1.5$ and $\sigma_1 = \sigma_{11} = 15$, and all the other parameters are the same as in Figure 4.3. Comparing to Figure 4.3, we can see that market orders are avoided when a_i and σ_i are high due to high expected cost, and limit orders are preferred due to high option values.

Regarding statistical errors in estimating parameters, empirical studies such as Uno and Yamada (1993) and Jain and Joh (1988) report that the patterns of market impact coefficients, the price volatility, and the market trading volume are rather stable.

Figure 4.3: **Time dependency of limit order and market order** ($T = 11$, $R_1 = 100$, $a_i = 1$, $\sigma_i = 10$ for $i = 1, \cdots, 11$). We assumed that market order, M_i, was revealed to be zero all the time. We can see that more limit orders are tried at the early stage of execution, and they are gradually replaced with market orders.

Figure 4.4: **Time dependency of limit order and market order** ($T = 11$, $R_1 = 100$, $a_1 = a_{11} = 1.5$, $\sigma_1 = \sigma_{11} = 15$, $a_i = 1$, $\sigma_i = 10$ for $i = 2, \cdots, 10$). We assumed that market order, M_i, was revealed to be zero all the time. Comparing to Figure 4.3, we can see that limit orders are avoided when a_i and σ_i are high due to high expected cost, and market orders are preferred due to high option values.

4.3   Variations of Single Limit Order Model

In this section, we analyze several variations of the standard model in Section 4.2. Although two period models are analyzed for clarity of comparison with the standard model, multiple period extension is possible in each case.

4.3.1   Risk Averse Traders

Since limit orders have uncertainty in execution, risk averse traders may prefer market orders. When we measure risk by variance of total purchase amount, cost function can be defined as

$$
\begin{aligned}
C_1(R_1) &= \mathrm{E}\left(\sum_{j=2}^{2}\delta_j\epsilon_j R_j + \sum_{i=1}^{2}P_i\{(1-r)D_i + rR_i\}\right) \\
&\quad + \lambda\mathrm{V}\left(\sum_{j=2}^{2}\delta_j\epsilon_j R_j + \sum_{i=1}^{2}P_i\{(1-r)D_i + rR_i\}\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{2}P_i\{(1-r)D_i + rR_i\}\right) + \lambda\mathrm{V}\left(\sum_{j=2}^{2}\delta_j\epsilon_j R_j + \sum_{i=1}^{2}P_i\{(1-r)D_i + rR_i\}\right) \quad (4.9)
\end{aligned}
$$

where the definitions of P_i and D_i are the same as those in risk neutral case, λ is a risk premium, and V is the variance at the first period.

   Although it is difficult to solve (4.9), changes of fundamental cost tends to be much larger than the uncertainly of market impact in practice. Therefore, we ignore uncertainty of market impact, which makes the solution a linear function as in the risk neutral case. That is, cost function can be approximated by

$$
\begin{aligned}
C_1(R_1) &\approx \mathrm{E}\left(\sum_{i=1}^{2}P_i\{(1-r)D_i + rR_i\}\right) + \lambda\mathrm{V}\left(\delta_2\epsilon_2 R_2\right) \\
&= \mathrm{E}\left(P_1\{(1-r)D_1 + rR_1\} + a_2(R_1 - D_1)^2\right) + \lambda\delta_2^2\mathrm{E}\left((R_1 - D_1)^2\right).
\end{aligned}
$$

Therefore, all we have to do is to replace a_2 of the solution in the risk neutral case with $(a_2 + \lambda\delta_2^2)$.

   Our optimal execution duration for risk averse traders is not very sensitive to trade size as in Almgren and Chriss (2001), which contradicts practical intuition. This is because variance type risk measure is employed in order to implement dynamic programming. As a result, it is recommended in practice to use both Konishi and Makimoto (2001) for execution scheduling problem and this analysis for order placing problem.

4.3.2   Visible Public Market Orders

Up to this section, we assume that the trader knows just the expectation of the public limit orders and the distribution of the public market orders, but not concrete order sizes. In this section however, we analyze the case when the trader monitors the public order book, and all the current period information is

available. Regarding the future periods, we continue to assume that the trader knows only the distribution and not the size of public market orders as in the previous cases.

As we have mentioned, not all traders can monitor all the public limit order books at TSE, only member security brokers are allowed to watch without time lag. Also, even member security brokers have to choose which stocks to monitor among thousands of listed stocks. This section estimates the value of monitoring public order books, comparing cost with that in Section 4.2.

Let $b_1(p)$ and $s_1(p)$ denote volume of public buy and sell limit orders at price p in the first period, respectively, and M_i denote volume of excess public market order at the ith period. Without loss of generality, we can assume that

$$\begin{cases} b_1(p) = 0, & \text{if } p > 0, \\ s_1(p) = 0, & \text{if } p < 0 \end{cases}$$

because buy limit order is synthesized by buy market order and sell limit order, and sell limit order is synthesized by sell market order and buy limit order.

Market orders, buy limit orders at higher price than execution price, and sell limit orders at lower price than execution price are executed, and the trader do not distinguish market and limit orders since not only market orders but also limit orders are surely executed. Therefore, let V_1 denote the sum of the trader's market order and buy limit orders at higher prices than execution price. Also, in order to complete the execution, we assume that all the remaining orders are placed as market order in the second period. In this case, market impact P_1 is determined so as to balance total buy orders and sell orders. Among them, since the trader's execution volume is V_1, expected purchase cost is $\mathrm{E}\left((P_1 + Q_1)V_1\right) = P_1 V_1$. Also, execution volume of public buy and sell orders is $\int_{P_1}^{\infty} b_1(p)dp$ and $\int_{-\infty}^{P_1} s_1(p)dp$, respectively. Therefore,

$$V_1 + M_1 + \int_{P_1}^{\infty} b_1(p)dp - \int_{-\infty}^{P_1} s_1(p)dp = 0.$$

Regarding the second period, we assume that limit order exists at any prices for the constant amount $1/a_i$, and that the ratio of permanent market impact is r as in the previous cases. Consequently, there are sell limit orders of $1/a_i$ above price $Q_2 + rP_1$ and buy limit orders of $1/a_i$ below price $Q_2 + rP_1$ in the second period. In this case, our problem can be formulated as follows:

$$\text{(P)} \quad \left| \begin{array}{ll} \text{minimize} & C_1(R_1) = P_1\{(1-r)V_1 + rR_1\} + a_2(R_1 - V_1)^2 \\ \text{s.t.} & V_1 + M_1 + \int_{P_1}^{\infty} b_1(p)dp - \int_{-\infty}^{P_1} s_1(p)dp = 0. \end{array} \right. \tag{4.10}$$

We can easily derive the optimal V_1^* at least numerically once Q_1, R_1, a_2, $b_1(p)$, $s_1(p)$, and M_1 are given.

In the rest of this section, we analyze a linear market impact case in order to compare the results with those in Section 4.2:

$$b_1(p) = \begin{cases} 0, & \text{if } p > 0, \\ \frac{1}{a_1}, & \text{if } p < 0, \end{cases}$$

$$s_1(p) = \begin{cases} \frac{1}{a_1}, & \text{if } p > 0, \\ 0, & \text{if } p < 0. \end{cases}$$

Since the constraint in (4.10) is

$$V_1 + M_1 = \frac{P_1}{a_1},$$

our objective function is

$$
\begin{aligned}
C_1(R_1) &= P_1\{(1-r)V_1 + rR_1\} + a_2(R_1 - V_1)^2 \\
&= \{(1-r)a_1 + a_2\}V_1^2 + [\{(1-r)M_1 + rR_1\}a_1 - 2Ra_2]V_1 + (rM_1a_1 + 2Ra_2). \quad (4.11)
\end{aligned}
$$

V_1 which minimizes (4.11) satisfies

$$\frac{\partial C_1}{\partial V_1} = 2\{(1-r)a_1 + a_2\}V_1 + \{(1-r)M_1 + rR_1\}a_1 - 2Ra_2 = 0.$$

Therefore, the optimal order size V_1^* is

$$V_1^* = \frac{-\{(1-r)M_1 + rR_1\}a_1 + 2R_1a_2}{2\{(1-r)a_1 + a_2\}},$$

and the optimal expected cost $C_1^*(R_1)$ is

$$C_1^*(R_1) = \frac{-(1-r)^2a_1^2M_1^2 + 2\{r(1-r)a_1 + 2a_2\}R_1a_1M_1 + (4a_2 - r^2a_1)a_1R_1^2}{4\{(1-r)a_1 + a_2\}}.$$

If we collect large samples in order to compare the optimal expected cost with the result in Section 4.2, M_i follows normal distribution $N(0, \sigma_i)$. Therefore, the average of $C_1^*(R_1)$ becomes

$$\overline{C_1^*(R_1)} = \frac{(4a_2 - r^2a_1)a_1R_1^2 - (1-r)^2\sigma_1^2a_1^2}{4\{(1-r)a_1 + a_2\}}.$$

Difference between $\overline{C_1^*(R_1)}$ and the minimal cost in Section 4.2 is

$$\frac{-(1-r)^2\sigma_1^2a_1^2}{4\{(1-r)a_1 + a_2\}} + \frac{\{(1-r)a_1 + 2a_2\}(1-r)a_1\sigma_1^2\phi^2(\frac{w_1}{2\sigma_1})}{2\{(1-r)a_1 + a_2\}\Phi(\frac{w_1}{2\sigma_1})} - a_2\sigma_1^2\{2\Phi(\frac{w_1}{2\sigma_1}) - 1\},$$

which corresponds to economical value of monitoring the public order book, which is significant when a_1 and σ_1 are large. Since the trader is assumed to be risk neutral, economical value of monitoring seems independent of R_1. However in practice, traders closely monitor stocks of large trades because risk averse traders try to avoid uncertainty of market impact, and because execution of larger trades takes a longer time, which makes the economical value of monitoring large.

As we have mentioned, at TSE, only member security brokers can monitor the public order book without time lag, but not general traders. Therefore, the economical value of monitoring is a component of the membership value. Also, even brokers have to pay some "cost" of infrastructure and labors to monitor the public order book, and therefore, choose stocks to monitor among thousands of listed stocks. Our formula evaluates the value of monitoring in order to justify costs of monitoring and provides criteria for selecting stocks to monitor.

4.4 Multiple Limit Order Model

Result of the standard model in Section 4.2 can be extended to a case with N limit orders. However, since calculation is extremely complicated, it is difficult to derive explicit solution. Therefore, in this section, we just explain some characteristics of the optimal execution strategy derived numerically. We study two period case for simplicity.

Let p_{ij} $(i = 1, \cdots, N, \ j = 1, \cdots, T)$ and w_{ij} $(i = 1, \cdots, N, \ j = 1, \cdots, T)$ denote i^{th} base limit order prices at the jth period and size of corresponding limit order sizes where $p_{ij} > p_{kj}$ for $i < k$. Also, let u_j denote market order size, and define $v_{0j} \overset{\text{def}}{=} u_j$ and $v_{ij} \overset{\text{def}}{=} v_{0j} + \sum_{k=1}^{i} w_{kj}$. Then, similar calculation leads to the expected cost at the first period C_1 as

$$
\begin{aligned}
C_1 \ = \ & \{(1-r)a_1 + a_2\}v_{01}^2 + (ra_1 - 2a_2)R_1 v_{01} + a_2 R_1^2 + Q_1 R_1 \\
& - \sum_{i=1}^{N}[\{(1-r)a_1 + a_2\}(v_{i-1,1} + \frac{p_{i1}}{a_1}) + (ra_1 - 2a_2)R_1] \\
& \quad \{(v_{i-1,1} - \frac{p_{i1}}{a_1})\Phi(-\frac{1}{\sigma_1}(v_{i-1,1} - \frac{p_{i1}}{a_1})) - \sigma_1 \phi(-\frac{1}{\sigma_1}(v_{i-1,1} - \frac{p_{i1}}{a_1}))\} \\
& + \sum_{i=1}^{N}[\{(1-r)a_1 + a_2\}(v_{i1} + \frac{p_{i1}}{a_1}) + (ra_1 - 2a_2)R_1] \\
& \quad \{(v_{i1} - \frac{p_{i1}}{a_1})\Phi(-\frac{1}{\sigma_1}(v_{i1} - \frac{p_{i1}}{a_1})) - \sigma_1 \phi(-\frac{1}{\sigma_1}(v_{i1} - \frac{p_{i1}}{a_1}))\} \\
& + a_2 \sigma_1^2 \sum_{i=1}^{N}\{\Phi(-\frac{1}{\sigma_1}(v_{i-1,1} - \frac{p_{i1}}{a_1})) - \Phi(-\frac{1}{\sigma_1}(v_{i1} - \frac{p_{i1}}{a_1}))\}.
\end{aligned} \tag{4.12}
$$

Also, the first order conditions of optimality is given as

$$
\begin{aligned}
\frac{\partial C_1}{\partial v_{01}} \ = \ & 2\{(1-r)a_1 + a_2\}v_{01} + (ra_1 - 2a_2)R_1 - [2\{(1-r)a_1 + a_2\}v_{01} + (ra_1 - 2a_2)R_1] \\
& \times \Phi(-\frac{1}{\sigma_1}(v_{01} - \frac{p_{11}}{a_1})) + (1-r)a_1 \sigma_1 \phi(-\frac{1}{\sigma_1}(v_{01} - \frac{p_{11}}{a_1})),
\end{aligned} \tag{4.13}
$$

$$
\begin{aligned}
\frac{\partial C_1}{\partial v_{k1}} \ = \ & -[2\{(1-r)a_1 + a_2\}v_{k1} + (ra_1 - 2a_2)R_1]\Phi(-\frac{1}{\sigma_1}(v_{k1} - \frac{p_{k+1,1}}{a_1})) \\
& + (1-r)a_1 \sigma_1 \phi(-\frac{1}{\sigma_1}(v_{k1} - \frac{p_{k+1,1}}{a_1})) + [2\{(1-r)a_1 + a_2\}v_{k1} + (ra_1 - 2a_2)R_1] \\
& \times \Phi(-\frac{1}{\sigma_1}(v_{k1} - \frac{p_{k1}}{a_1})) - (1-r)a_1 \sigma_1 \phi(-\frac{1}{\sigma_1}(v_{k1} - \frac{p_{k1}}{a_1})),
\end{aligned} \tag{4.14}
$$

$$
\begin{aligned}
\frac{\partial C_1}{\partial v_{N1}} \ = \ & [2\{(1-r)a_1 + a_2\}v_{N1} + (ra_1 - 2a_2)R_1]\Phi(-\frac{1}{\sigma_1}(v_{N1} - \frac{p_{N1}}{a_1})) \\
& - (1-r)a_1 \sigma_1 \phi(-\frac{1}{\sigma_1}(v_{N1} - \frac{p_{N1}}{a_1})),
\end{aligned} \tag{4.15}
$$

$$
\begin{aligned}
a_1 \frac{\partial C_1}{\partial p_{k1}} \ = \ & +[2\{(1-r)a_1 + a_2\}\frac{p_{k1}}{a_1} + (ra_1 - 2a_2)R_1]\Phi(-\frac{1}{\sigma_1}(v_{k-1,1} - \frac{p_{k1}}{a_1})) \\
& + \{(1-r)a_1 + 2a_2\}\sigma_1 \phi(-\frac{1}{\sigma_1}(v_{k-1,1} - \frac{p_{k1}}{a_1})) - [2\{(1-r)a_1 + a_2\}\frac{p_{k1}}{a_1} + (ra_1 - 2a_2)R_1] \\
& \times \Phi(-\frac{1}{\sigma_1}(v_{k1} - \frac{p_{k1}}{a_1})) - \{(1-r)a_1 + 2a_2\}\sigma_1 \phi(-\frac{1}{\sigma_1}(v_{k1} - \frac{p_{k1}}{a_1})).
\end{aligned} \tag{4.16}
$$

Figure 4.5 compares the optimal costs derived numerically from (4.13) to (4.16) for various numbers of limit orders when $a_1 = a_2 = 1$, $\sigma_1 = 10$, $r = 0$, and $R_1 = 50$. Although the optimal cost is a decreasing function of the number of limit orders, cost reduction decreases as the number of limit orders increases. This result supports the effectiveness of the single limit order model.

Figure 4.6 shows the corresponding optimal base limit order prices and their sizes. We can see that the limit order size is symmetrical. Although sizes increase as limit order prices depart from the center of symmetry, this is because limit orders are dense around the center. The density of limit orders is unimodal, similar to the single limit order model.

Figure 4.5: **Number of limit orders (x-axis) and the expected cost (y-axis).**   This graph compares the optimal costs for various numbers of limit orders. Although the optimal cost is a decreasing function of the number of limit orders, cost reduction decreases as the number of limit orders increases. This result supports the effect of the single limit order model.

Numerical solution suggests the following conjecture about the optimal solution although the proof of the optimality has not yet been completed.

Conjecture 4.4.1

$$v_{i1}^* + v_{N-i,1}^* = \frac{p_{i+1,1}^*}{a_1} + \frac{p_{N-i,1}^*}{a_1} = -\frac{(ra_1 - 2a_2)R_1}{2(1-r)a_1 + a_2}, \quad \text{for even } N, \tag{4.17}$$

and

$$v_{i+1,1}^* + v_{N-i,1}^* = \frac{p_{i1}^*}{a_1} + \frac{p_{N-i,1}^*}{a_1} = -\frac{(ra_1 - 2a_2)R_1}{2(1-r)a_1 + a_2}, \quad \text{for odd } N. \tag{4.18}$$

Conjecture 4.4.1 can be explained as follows. First of all, because (4.12) does not change even if we switch $v_{i1} - \frac{p_{i+1,1}}{a_1}$ and $-\left(v_{N-i,1} - \frac{p_{N-i,1}}{a_1}\right)$,

$$v_{i1}^* - \frac{p_{i+1,1}^*}{a_1} = -\left(v_{N-i,1}^* - \frac{p_{N-i,1}^*}{a_1}\right).$$

Figure 4.6: **Optimal base limit order prices (x-axis) and their sizes (y-axis).** We can see that the limit order size is indeed symmetrical. Although sizes increase as limit orders depart from the center of symmetry, this is because limit orders are dense arround the center. The density of limit orders is unimodal, similar to single limit order model.
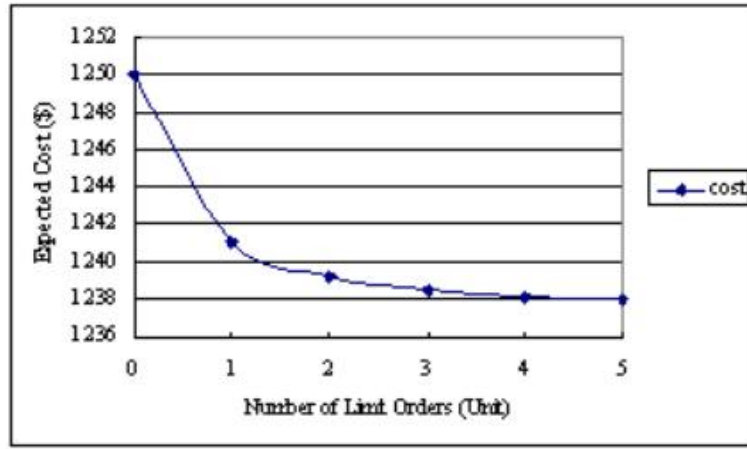
If N is even, because $v^*_{a\frac{N}{2}} - \frac{p^*_{\frac{N}{2}+1}}{a_1} = -(v^*_{a\frac{N}{2}} - \frac{p^*_{\frac{N}{2}}}{a_1})$ and $\frac{\partial C_1}{\partial v_{a\frac{N}{2}}} = 0$,

$$v^*_{a\frac{N}{2}} = -\frac{(ra_1 - 2a_2)R_1}{2\{(1-r)a_1 + a_2\}}.$$

Also, because $v^*_{a\frac{N}{2}} - \frac{p^*_{\frac{N}{2}+1}}{a_1} = -(v^*_{a\frac{N}{2}} - \frac{p^*_{\frac{N}{2}}}{a_1})$,

$$p^*_{\frac{N}{2}} + p^*_{\frac{N}{2}+1} = -\frac{(ra_1 - 2a_2)a_1 R_1}{2\{(1-r)a_1 + a_2\}}.$$

Further, because $v^*_{a\frac{N}{2}-1} - \frac{p^*_{\frac{N}{2}}}{a_1} = -(v^*_{a\frac{N}{2}+1} - \frac{p^*_{\frac{N}{2}+1}}{a_1})$,

$$v^*_{a\frac{N}{2}-1} + v^*_{a\frac{N}{2}+1} = -\frac{(ra_1 - 2a_2)R_1}{2\{(1-r)a_1 + a_2\}}.$$

Similarly, because $v^*_{a\frac{N}{2}+i} - \frac{p^*_{\frac{N}{2}+i+1}}{a_1} = -(v^*_{a\frac{N}{2}-i} - \frac{p^*_{\frac{N}{2}-i}}{a_1})$,

$$p^*_{\frac{N}{2}-i} + p^*_{\frac{N}{2}+i+1} = -\frac{(ra_1 - 2a_2)a_1 R_1}{2\{(1-r)a_1 + a_2\}}.$$

Also, since $v^*_{a\frac{N}{2}-i-1} - \frac{p^*_{\frac{N}{2}-i}}{a_1} = -(v^*_{a\frac{N}{2}+i+1} - \frac{p^*_{\frac{N}{2}+i+1}}{a_1})$,

$$v^*_{a\frac{N}{2}-i-1} + v^*_{a\frac{N}{2}+i+1} = -\frac{(ra_1 - 2a_2)R_1}{2\{(1-r)a_1 + a_2\}}.$$

Similar argument holds for the case if N is odd.

By (4.17) and (4.18), we can reason by analogy that p_{i1} and v_{i1} are linear functions of R_1, and therefore, w_{i1} are constants just as in the single limit order model.

4.5   Closing Remarks

This chapter analyses the optimal selection of market and limit orders in a series of single price batch auctions when the expectation of the limit order book is given. We derive the analytical solution of a single limit order model for risk neutral traders, and then extend the result for risk averse traders. Regarding the execution at each period, we find that the optimal limit order size is independent of the whole trade size so as to balance the non-linearity of execution volume and price. In contrast, the limit order price and the market order size are linear functions in order to balance trading costs across trading periods. Regarding the execution throughout the trading session, market orders replace limit orders as time passes, and necessity of completing execution grows. Further, if trading volume and price volatility are larger at the opening and closing of the trading session as in the practical markets, more limit orders are tried at the opening and closing.

Although most traders know the public order book as expectations as in our model, member security brokers are allowed to monitor the public order book without time lags. So, we evaluate the value of monitoring public order books and provide criteria for selecting stocks to monitor.

Although our results can be extended to multiple limit order models mathematical analysis is left for further research. Besides, while this chapter analyzes batch auctions for simplicity, continuous auction is also of interest. Also, although this chapter assumes randomness comes only from the number of market orders, the number of limit orders are also stochastic in practice. Further, we can extend our research to risk averse traders who uses VWAP as a reference price. Besides, while this chapter performs partial equilibrium type analysis in which public orders are given exogenously, general equilibrium analysis which studies how each trader's strategy interferes is left for further research.

4.6   Appendix

4.6.1   Proof of Proposition 4.2.1

In two period case,

$$
\begin{aligned}
C_1(R_1) &= \mathrm{E}\left(\pi_1(\psi_1)\{(1-r)\kappa_1(\psi_1) + rR_1\} + a_2(R_1 - \kappa_1(\psi_1, M_1)^2)\right) \\
&= \int_{-(u_1 - \frac{p_1}{a_1})}^{\infty} \{a_1(m_1 + u_1)\{(1-r)u_1 + rR_1\} + a_2(R_1 - u_1)^2\}f_1(m_1)dm_1 \\
&\quad + \int_{-(v_1 - \frac{p_1}{a_1})}^{-(u_1 - \frac{p_1}{a_1})} \{p_1\{(1-r)(-m_1 + \frac{p_1}{a_1}) + rR_1\} + a_2(R_1 + m_1 - \frac{p_1}{a_1})^2\}f_1(m_1)dm_1 \\
&\quad + \int_{-\infty}^{-(v_1 - \frac{p_1}{a_1})} \{a_1(m_1 + v_1)\{(1-r)v_1 + rR_1\} + a_2(R_1 - v_1)^2\}f_1(m_1)dm_1 \\
&= \int_{-(u_1 - \frac{p_1}{a_1})}^{\infty} [a_1\{rR_1 + (1-r)u_1\}m_1 + (a_1 + a_2 - ra_1)u_1^2 - (2a_2 - ra_1)R_1u_1]f_1(m_1)dm_1
\end{aligned}
$$

$$+\int_{-(v_1-\frac{p_1}{a_1})}^{-(u_1-\frac{p_1}{a_1})} a_2 m_1^2 + \{2a_2 R_1 + ((r-1)a_1 - 2a_2)\frac{p_1}{a_1}\}m_1$$

$$+\{(1-r)a_1 + a_2\}(\frac{p_1}{a_1})^2 + (ra_1 - 2a_2)R_1\frac{p_1}{a_1}]f_1(m_1)dm_1$$

$$+\int_{-\infty}^{-(v_1-\frac{p_1}{a_1})} [a_1\{rR_1 + (1-r)v_1\}m_1 + (a_1 + a_2 - ra_1)v_1^2 - (2a_2 - ra_1)R_1v_1]f_1(m_1)dm_1$$

$$+a_2 R_1^2.$$

Using the following equations,

$$\int_{-\infty}^{-a} xf_1(x)dx = -\sigma_1^2 f_1(a),$$

$$\int_{a}^{b} x^2 f_1(x)dx = \int_{a}^{b} \sigma_1^2 f_1(x)dx - [\sigma_1^2 x f_1(x)]_a^b,$$

we obtain the following result,

$$\begin{aligned}
C_1(R_1) &= \{(1-r)a_1 + a_2\}v_{01}^2 + (ra_1 - 2a_2)R_1 v_{01} + a_2 R_1^2 \\
&\quad -[\{(1-r)a_1 + a_2\}(u_1 + \frac{p_1}{a_1}) + (ra_1 - 2a_2)R_1]\{(u_1 - \frac{p_1}{a_1})F_1(-(u_1 - \frac{p_1}{a_1})) - \sigma_1^2 f_1(-(u_1 - \frac{p_1}{a_1}))\} \\
&\quad +[\{(1-r)a_1 + a_2\}(v_1 + \frac{p_1}{a_1}) + (ra_1 - 2a_2)R_1]\{(v_1 - \frac{p_1}{a_1})F_1(-(v_1 - \frac{p_1}{a_1})) - \sigma_1^2 f_1(-(v_1 - \frac{p_1}{a_1}))\} \\
&\quad +a_2\sigma_1^2\{F_1(-(u_1 - \frac{p_1}{a_1})) - F_1(-(v_1 - \frac{p_1}{a_1}))\}
\end{aligned}$$

where $F_1(x) = \int_{-\infty}^{x} f_1(y)dy$. The desired result now follows by noting

$$F_1(w) = \Phi(\frac{w}{\sigma_1}), \qquad \sigma_1^2 f_1(w) = \sigma_1\phi(\frac{w}{\sigma_1})$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}exp(-\frac{x^2}{2})$ and $\Phi(x) = \int_{-\infty}^{x} \phi(y)dy$.

4.6.2 Proof of Proposition 4.2.2

Define $\overline{u}_1 = \dfrac{u_1 - \frac{p_1}{a_1}}{\sigma_1}$ and $\overline{v}_1 = \dfrac{v_1 - \frac{p_1}{a_1}}{\sigma_1}$. Since

$$F_1(-(u_1 - \frac{p_1}{a_1})) = \Phi(-\overline{u}_1), \tag{4.19}$$

$$F_1(-(v_1 - \frac{p_1}{a_1})) = \Phi(-\overline{v}_1), \tag{4.20}$$

$$f_1(-(u_1 - \frac{p_1}{a_1})) = \frac{1}{\sigma_1}\phi(-\overline{u}_1), \tag{4.21}$$

$$f_1(-(v_1 - \frac{p_1}{a_1})) = \frac{1}{\sigma_1}\phi(-\overline{v}_1), \tag{4.22}$$

$C_1(R_1)$ can be rewritten as,

$$C_1(R_1) = [\{(1-r)a_1 + a_2\}(\frac{p_1}{a_1})^2 + (ra_1 - 2a_2)R_1\frac{p_1}{a_1} + a_2\sigma_1^2]$$

$$+[\{(1-r)a_1 + a_2\}(\sigma_1^2 \overline{u}_1^2 + 2\sigma_1 \frac{p_1}{a_1}\overline{u}_1) + (ra_1 - 2a_2)R_1\sigma_1\overline{u}_1 - a_2\sigma_1^2]\Phi(\overline{u}_1)$$

$$+[\{(1-r)a_1 + a_2\}(\sigma_1^2 \overline{v}_1^2 + 2\sigma_1 \frac{p_1}{a_1}\overline{v}_1) + (ra_1 - 2a_2)R_1\sigma_1\overline{v}_1 - a_2\sigma_1^2]\Phi(-\overline{v}_1)$$

$$+[\{(1-r)a_1 + a_2\}(\sigma_1^2 \overline{u}_1 + 2\sigma_1 \frac{p_1}{a_1}) + (ra_1 - 2a_2)R_1\sigma_1]\phi(\overline{u}_1)$$

$$-[\{(1-r)a_1 + a_2\}(\sigma_1^2 \overline{v}_1 + 2\sigma_1 \frac{p_1}{a_1}) + (ra_1 - 2a_2)R_1\sigma_1]\phi(-\overline{v}_1)$$

$$+a_2 R_1^2. \tag{4.23}$$

We then prove that the optimality and the uniqueness of the solution. First, define

$$a = (1-r)a_1 + a_2, \qquad A = a_2, \qquad \sigma = \sigma_1,$$

$$\beta = (ra_1 - 2a_2)R_1, \qquad \gamma = p_1/a_1, \qquad x = \overline{u}_1, \qquad y = \overline{v}_1,$$

and

$$G(x) \stackrel{\text{def}}{=} x\Phi(x) + \phi(x).$$

Also, we represent function parameters by subscripts as $\Phi_x = \Phi(x)$ and $G_x = G(x)$. According to (4.23),

$$
\begin{aligned}
C_1(R_1) &= a\gamma^2 + \beta\gamma + A\sigma^2 + [a(\sigma^2 x^2 + 2\sigma x\gamma) + \beta\sigma x - A\sigma^2]\Phi_x + [a(\sigma^2 y^2 + 2\sigma y\gamma) + \beta\sigma y - A\sigma^2]\Phi_{-y} \\
&\quad + [a(\sigma^2 x + 2\sigma\gamma) + \beta\sigma]\phi_x - [a(\sigma^2 y + 2\sigma\gamma) + \beta\sigma]\phi_y + AR_1^2 \\
&= A\sigma^2 + AR_1^2 + a\gamma^2 + 2a\sigma\left(x\Phi_x + \phi_x + y\Phi_{-y} - \phi_y + \frac{\beta}{2a\sigma}\right)\gamma \\
&\quad + \sigma\left[\left(a\sigma x^2 + \beta x - A\sigma\right)\Phi_x + (a\sigma x + \beta)\phi_x + \left(a\sigma y^2 + \beta y - A\sigma\right)\Phi_{-y} - (a\sigma y + \beta)\phi_y\right] \\
&= A\sigma^2 + AR_1^2 + a\gamma^2 + 2a\sigma\left(G_x - G_{-y} + \frac{\beta}{2a\sigma}\right)\gamma \\
&\quad + \sigma\left[(a\sigma x + \beta)G_x - (a\sigma y + \beta)G_{-y} - A\sigma\{\Phi_x + \Phi_{-y}\}\right] \\
&= A\sigma^2 + AR_1^2 + a\left[\gamma + \sigma\left(G_x - G_{-y} + \frac{\beta}{2a\sigma}\right)\right]^2 \\
&\quad + a\sigma^2\left[xG_x - yG_{-y} + \frac{\beta}{a\sigma}(G_x - G_{-y}) - \frac{A}{a}(\Phi_x + \Phi_{-y})\right] - a\sigma^2\left(G_x - G_{-y} + \frac{\beta}{2a\sigma}\right)^2 \\
&= A\sigma^2 + AR_1^2 - \frac{\beta^2}{4a} + a\left[\gamma + \sigma\left(G_x - G_{-y} + \frac{\beta}{2a\sigma}\right)\right]^2 \\
&\quad + a\sigma^2\left[xG_x - yG_{-y} - \frac{A}{a}(\Phi_x + \Phi_{-y}) - (G_x - G_{-y})^2\right]. \tag{4.24}
\end{aligned}
$$

Define the quotient of the last term of (4.24) by $a\sigma^2$ as

$$H(x,y) \stackrel{\text{def}}{=} xG_x - yG_{-y} - \frac{A}{a}(\Phi_x + \Phi_{-y}) - (G_x - G_{-y})^2.$$

We minimize C_1 on $\mathcal{S} = \{(x,y) \ : \ -\infty < x < \infty, y \geq x\}$.

Lemma 4.6.1   *The minimum of H exists on $\mathcal{S}$.*

Proof. Since $H(x, y) = H(-y, -x)$, $H(x, y)$ is symmetrical about $y = -x$. Therefore, we have only to analyze on $\mathcal{S}' = \{(x, y) : -\infty < x < \infty, y \geq |x|\}$. According to Abramowitz and Stegun (1972) p.932,

$$\frac{\phi_x}{x}\left(1 - \frac{1}{x^2} + \frac{3}{x^4} - \frac{15}{x^6}\right) < \Phi_{-x} < \frac{\phi_x}{x}\left(1 - \frac{1}{x^2} + \frac{3}{x^4} + \frac{15}{x^6}\right), \qquad x > 0,$$

and therefore,

$$x + \phi_x\left(\frac{1}{x^2} - \frac{3}{x^4} - \frac{15}{x^6}\right) < G_x < x + \phi_x\left(\frac{1}{x^2} - \frac{3}{x^4} + \frac{15}{x^6}\right), \qquad x > 0,$$

$$\phi_x\left(\frac{1}{x^2} - \frac{3}{x^4} - \frac{15}{x^6}\right) < G_{-x} < \phi_x\left(\frac{1}{x^2} - \frac{3}{x^4} + \frac{15}{x^6}\right), \qquad x > 0.$$

Consequently, limits along the point $\mathcal{S}'$ where (sy, y) $(-1 \leq s \leq 1)$ is given by

$$\lim_{y \to +\infty} H(sy, y) = \begin{cases} -A/a, & s > 0, \\ -A/2a + \phi_0^2, & s = 0, \\ 0, & s < 0. \end{cases}$$

Therefore, $H(x, y)$ is a function bounded from the bottom. Further, since $H(0, 0) = -A/a \leq \lim_{y \to +\infty} H(sy, y)$, $H(x, y)$ takes the minimum at a finite (x, y). □

Since γ can take any value, the minimization of C by (x, y, γ) is equivalent to the minimization of H by (x, y) with $\gamma = -\sigma(G_x - G_{-y} + \beta/2a\sigma)$. From the first order condition of the optimality,

$$\frac{\partial H}{\partial x} = G_x + x\Phi_x - \frac{A}{a}\phi_x - 2(G_x - G_{-y})\Phi_x = 0, \qquad (4.25)$$

$$\frac{\partial H}{\partial y} = -G_{-y} + y\Phi_{-y} + \frac{A}{a}\phi_y - 2(G_x - G_{-y})\Phi_{-y} = 0 \qquad (4.26)$$

are necessary conditions. By canceling A/a from (4.25) and (4.26),

$$\phi_y\left[G_x + x\Phi_x - 2(G_x - G_{-y})\Phi_x\right] + \phi_x\left[-G_{-y} + y\Phi_{-y} - 2(G_x - G_{-y})\Phi_{-y}\right]$$

$$= 2\phi_y\Phi_x[x - (G_x - G_{-y})] + 2\phi_x\Phi_{-y}[y - (G_x - G_{-y})]$$

$$= 2[\phi_y\Phi_x(G_{-y} - G_{-x}) + \phi_x\Phi_{-y}(G_y - G_x)]$$

$$= 0.$$

Define

$$I(x, y) \stackrel{\text{def}}{=} \phi_x\Phi_{-y}(G_y - G_x) + \Phi_x\phi_y(G_{-y} - G_{-x}),$$

and we have the following lemma.

Lemma 4.6.2 *For any x, $I(x, x) = I(x, -x) = 0$. Further, for any x and $y > |x|$, $I(x, y) > 0$.*

Lemma 4.6.2 shows that the first order condition of the optimality is satisfied only by $y = x$ and $y = -x$. If we substitute $y = x$ (4.25), we have

$$\left.\frac{\partial H}{\partial x}\right|_{y=x} = G_x + x\Phi_x - \frac{A}{a}\phi_x - 2(G_x - G_{-x})\Phi_x$$

$$
\begin{aligned}
&= G_x + x\Phi_x - \frac{A}{a}\phi_x - 2x\Phi_x \\
&= \phi_x - \frac{A}{a}\phi_x \\
&= 0,
\end{aligned}
$$

which contradicts $0 < A/a < 1$. Next, if we substitute $y = -x$, we have

$$
\left.\frac{\partial H}{\partial x}\right|_{y=-x} = 2x\Phi(x) + \left(1 - \frac{A}{a}\right)\phi(x) = 0. \tag{4.27}
$$

Regarding the solution of (4.27), we have the following lemma.

Lemma 4.6.3   *Define $M(x) \stackrel{\text{def}}{=} (ax + b)\Phi(x) + c\phi(x)$ where $a > 0$ and $a > c$. Then, $M(x) = 0$ has a unique solution with positive derivative.*

Proof.   Direct calculation shows that

$$
M'(x) = a\Phi(x) + b\phi(x) + (a - c)x\phi(x),
$$

and

$$
M''(x) = \phi(x)\{-(a - c)x^2 - bx + (2a - c)\}.
$$

Therefore, $M''(x) = 0$ has two real solutions across which $M''(x)$ turns from negative to positive once, and then negative again. Together with $\lim_{x\to-\infty} M'(x) = 0$ and $\lim_{x\to\infty} M'(x) = a > 0$, $M'(x)$ has a unique solution x^* and,

$$
\begin{cases}
M'(x) < 0, & x < x^*, \\
M'(x) > 0, & x > x^*.
\end{cases} \tag{4.28}
$$

Further, together with $\lim_{x\to-\infty} M(x) = 0$ and $\lim_{x\to\infty} M(x) = \infty$, $M(x) = 0$ has a unique solution where $M'(x) > 0$. $\qquad\square$

By Lemma 4.6.3, x that satisfies (4.27) is uniquely determined. Consequently, $x = x^*$ is determined as the unique solution of (4.27), and then, $y^* = -x^*$ and $\gamma^* = -\beta/2a$.

Proof of Lemma 4.6.2   It is easy to show $I(x, x) = I(x, -x) = 0$. We consider the case of $y > |x|$. We first calculate partial derivatives of I by y.

$$
\begin{aligned}
I_1(x, y) &\stackrel{\text{def}}{=} \frac{\partial}{\partial y} I(x, y) \\
&= \Phi_x(-y\phi_y G_{-y} - \phi_y\Phi_{-y}) + \phi_x(-\phi_y G_y + \Phi_{-y}\Phi_y) + \Phi_x G_{-x}y\phi_y + \phi_x G_x\phi_y, \\
I_2(x, y) &\stackrel{\text{def}}{=} \frac{\partial}{\partial y} I_1(x, y) \\
&= \phi_y[\Phi_x\{(y^2 - 1)G_{-y} + 2y\Phi_{-y} + \phi_y\} + \phi_x(yG_y + 1 - 3\Phi_y) - (y^2 - 1)\Phi_x G_{-x} - \phi_x G_x y], \\
J_2(x, y) &\stackrel{\text{def}}{=} \Phi_x\{(y^2 - 1)G_{-y} + 2y\Phi_{-y} + \phi_y\} + \phi_x(yG_y + 1 - 3\Phi_y) - (y^2 - 1)\Phi_x G_{-x} - \phi_x G_x y,
\end{aligned}
$$

$$J_3(x,y) \stackrel{\text{def}}{=} \frac{\partial}{\partial y} J_2(x,y)$$

$$= \Phi_x\{2yG_{-y} - (y^2-3)\Phi_{-y} - 3y\phi_y\} + \phi_x(G_y + y\Phi_y - 3\phi_y) - 2\Phi_x G_{-x}y - \phi_x G_x,$$

$$J_4(x,y) \stackrel{\text{def}}{=} \frac{\partial}{\partial y} J_3(x,y)$$

$$= 2[\Phi_x\{G_{-y} - 2y\Phi_{-y} + (2y^2-3)\phi_y\} + \phi_x(\Phi_y + 2y\phi_y) - \Phi_x G_{-x}],$$

$$J_5(x,y) \stackrel{\text{def}}{=} \frac{\partial}{\partial y} J_4(x,y)$$

$$= 2[\Phi_x\{-3\Phi_{-y} - (2y^3-9y)\phi_y\} - \phi_x(2y^2-3)\phi_y],$$

$$J_6(x,y) \stackrel{\text{def}}{=} \frac{\partial}{\partial y} J_5(x,y)$$

$$= 2\phi_y\{(2y^4 - 15y^2 + 12)\Phi_x + (2y^3 - 7y)\phi_x\}.$$

Since

$$I(x,y) \approx \phi_x\Phi_{-y}G_y - \Phi_x G_{-x}\phi_y \approx \Phi_{-x}G_x\phi_y$$

for $y \to +\infty$, we obtain[1]

$$I(x,+\infty) = 0 + 0.$$

Together with $I(x,x) = I(x,-x) = 0$, if we prove the existence and the uniqueness of $y > |x|$ where $I_1(x,y) = 0$ for any x, we can prove that $I(x,y) > 0$ for $y > |x|$.

In order to see the sign change of J_6, we first analyze

$$K(y) \stackrel{\text{def}}{=} \frac{2y^4 - 15y^2 + 12}{2y^3 - 7y} = y - \frac{8y^2 - 12}{2y^3 - 7y}.$$

Then,

$$\frac{dK}{dy} = 1 + \frac{16y^4 - 16y^2 + 84}{(2y^3 - 7y)^2}$$

$$= 1 + \frac{4(2y^2 - 1)^2 + 80}{(2y^3 - 7y)^2}$$

$$> 0, \qquad (\forall x \neq 0, \pm\sqrt{7/2}).$$

Therefore, $K(y)$ is monotone increasing for $y \neq 0, \pm\sqrt{7/2}$, and shifts according to the following table.

y	$-\infty$	$\cdots$	$-\dfrac{\sqrt{15+\sqrt{129}}}{2}$	$\cdots$	$-\sqrt{\dfrac{7}{2}}$	$\cdots$	$-\dfrac{\sqrt{15-\sqrt{129}}}{2}$	$\cdots$	0	
$2y^4 - 15y^2 + 12$	$+\infty$	$+$	0	$-$	$-$	$-$	0	$+$	$+$	
$2y^3 - 7y$	$-\infty$	$-$	$-$	$-$	0	$+$	$+$	$+$	0	
$K(y)$	$-\infty$	$\nearrow$	0	$\nearrow$	$+\infty$	$-\infty$	$\nearrow$	0	$\nearrow$	$+\infty$ $-\infty$

y	0	$\cdots$	$\dfrac{\sqrt{15-\sqrt{129}}}{2}$	$\cdots$	$\sqrt{\dfrac{7}{2}}$	$\cdots$	$\dfrac{\sqrt{15+\sqrt{129}}}{2}$	$\cdots$	$+\infty$
$2y^4 - 15y^2 + 12$	$+$	$+$	0	$-$	$-$	$-$	0	$+$	$+\infty$
$2y^3 - 7y$	0	$-$	$-$	$-$	0	$+$	$+$	$+$	$+\infty$
$K(y)$	$+\infty$ $-\infty$	$\nearrow$	0	$\nearrow$	$+\infty$ $-\infty$	$\nearrow$	0	$\nearrow$	$+\infty$

---

[1] For a function f, we define $f(\pm\infty) = \lim_{t\to\pm\infty} f(t)$. Also, $a+0$ ($a-0$) represents conversion to a from the above (from the below)

Since $J_6(x, y) = 2\phi_y \Phi_x (2y^3 - 7y)[K(y) + \phi_x/\Phi_x]$, $J_6(x, y)$ turns from "positive, negative, positive, negative, and positive" for $y = -\infty \to +\infty$. Henceforth, we represent this shifts as[2]

$$\text{sgn}(J_6) = (+ - + - +). \tag{4.29}$$

Next, we analyze the sign change of J_5.

$$J_5(x, -\infty) \quad = \quad -6\Phi_x \quad < 0, \tag{4.30}$$

$$J_5(x, +\infty) \quad = \quad \lim_{y \to +\infty} -4\Phi_x y^3 \phi_y = 0 - 0. \tag{4.31}$$

By (4.29), (4.30), and (4.31), three cases are possible for the sign change of J_5, $\text{sgn}(J_5) = (-)$, $(- + -)$, $(- + - + -)$. We next show that $\text{sgn}(J_5) = (- + - + -)$ is not viable. When x is given, we define the smallest y as y_* among four y that satisfy $J_6(x, y) = 0$. Then, $y_* < -\sqrt{15 + \sqrt{129}}/2$.[3] Since $J_6(x, y_*) = 0$,

$$J_5(x, y_*) \quad = \quad -2\Phi_x(2y_*^2 - 3)\phi_{y_*}\left\{ \frac{3\Phi_{-y_*} + (2y_*^3 - 9y_*)\phi_{y_*}}{(2y_*^2 - 3)\phi_{y_*}} + \frac{\phi_x}{\Phi_x} \right\}$$

$$= \quad -2\Phi_x(2y_*^2 - 3)\phi_{y_*}\left\{ \frac{3\Phi_{-y_*} + (2y_*^3 - 9y_*)\phi_{y_*}}{(2y_*^2 - 3)\phi_{y_*}} - \frac{2y_*^4 - 15y_*^2 + 12}{2y_*^3 - 7y_*} \right\}.$$

Lemma 4.6.4   *For any x, $J_5(x, y^*) < 0$.*

Proof.   Since $2y_*^2 - 3 > 0$, it is sufficient to show $L(y^*) > 0$ where

$$L(y) \stackrel{\text{def}}{=} \frac{3\Phi_{-y} + (2y^3 - 9y)\phi_y}{(2y^2 - 3)\phi_y} - \frac{2y^4 - 15y^2 + 12}{2y^3 - 7y}.$$

Since $y_* < -\sqrt{15 + \sqrt{129}}/2 < -2$, it is sufficient to show $L(y) > 0$ for $y < -2$.

$$L(y) = \frac{1}{2y^2 - 3}\left(\frac{3\Phi_{-y}}{\phi_y} + \frac{4y^4 - 6y^2 + 36}{2y^3 - 7y} \right) = \frac{1}{2y^2 - 3}\left(\frac{3\Phi_{-y}}{\phi_y} + 2y + \frac{4}{y} + \frac{8}{2y^3 - 7y} \right).$$

Since both $4/y$ and $8/(2y^3 - 7y)$ are decreasing for $y < -2$,

$$\frac{4}{y} + \frac{8}{2y^3 - 7y} > \frac{4}{-2} + \frac{8}{2 \cdot (-2)^3 - 7 \cdot (-2)} = -4.$$

Consequently,

$$L(y) > \frac{1}{2y^2 - 3}\left(\frac{3\Phi_2}{\sqrt{2\pi}}e^{y^2/2} + 2y - 4 \right) > \frac{1}{2y^2 - 3}\left(1.137e^{y^2/2} + 2y - 4 \right), \qquad y < -2.$$

When we define

$$M(y) \stackrel{\text{def}}{=} 1.136e^{y^2/2} + 2y - 4,$$

---

[2]Although we are interested only the area of $y > |x|$, we analyze the change of the sign for $y = -\infty \to \infty$

[3]Although y_* depends on x, we do not write x explicitly for the simplicity of the representation. Besides, $y_* < -\sqrt{15 + \sqrt{129}}/2$ holds for any x.

we obtain

$$M(-2) > 0, \qquad \frac{dM}{dy} = 1.136ye^{y^2/2} + 2 < 0, \qquad \forall y < -2.$$

Therefore, we have shown $M(y) > 0$ for $y < -2$, i.e., $L(y) > 0$. □

Since Lemma 4.6.4 shows that the first relative minimum of J_5 is negative, $\mathrm{sgn}(J_5) = (-+-+-)$ never happens, and therefore, only $\mathrm{sgn}(J_5) = (-)$, $(-+-)$ are possible cases.

Next, regarding the sign change of J_4,

$$J_4(x, -\infty) = \lim_{y\to-\infty} -6\Phi_x y = +\infty,$$
$$J_4(x, +\infty) = 2\Phi_{-x}G_x > 0.$$

Together with $\mathrm{sgn}(J_5) = (-)$, $(-+-)$, there are two possible cases, $\mathrm{sgn}(J_4) = (+)$, $(+-+)$.

Next, regarding the sign change of J_3,

$$J_3(x, -\infty) = \lim_{y\to-\infty} -3\Phi_x y^2 = -\infty,$$
$$J_3(x, +\infty) = \lim_{y\to+\infty} 2\Phi_{-x}G_x y = +\infty.$$

Together with $\mathrm{sgn}(J_4) = (+)$, $(+-+)$, there are two possible cases, $\mathrm{sgn}(J_3) = (-+)$, $(-+-+)$.

Next, regarding the sing change of J_2,

$$J_2(x, -\infty) = \lim_{y\to-\infty} -\Phi_x y^3 = +\infty,$$
$$J_2(x, +\infty) = \lim_{y\to+\infty} \Phi_{-x}G_x y^2 = +\infty.$$

Together with $\mathrm{sgn}(J_3) = (-+)$, $(-+-+)$, there are two possible cases,

$$\mathrm{sgn}(J_2) = (+), \quad (+-+), \quad (+-+-+). \tag{4.32}$$

Finally, regarding the sign change of I_1,

$$I_1(x, -\infty) = \lim_{y\to-\infty} \Phi_x y^2 \phi_y = 0 + 0, \tag{4.33}$$
$$I_1(x, x) = 0, \tag{4.34}$$
$$I_1(x, +\infty) = \lim_{y\to+\infty} -\Phi_{-x}G_x y\phi_y = 0 - 0. \tag{4.35}$$

(I) When $x \geq 0$ The following lemma holds for

$$J_2(x, x) = 2x\Phi_x\Phi_{-x} + \phi_x - 2\phi_x\Phi_x.$$

Lemma 4.6.5

$$J_2(x, x) \begin{cases} > 0, & x > 0, \\ = 0, & x = 0, \\ < 0, & x < 0. \end{cases} \tag{4.36}$$

Proof.   When we define

$$A(x) \stackrel{\text{def}}{=} J_2(x,x) = 2x\Phi_x\Phi_{-x} + \phi_x - 2\phi_x\Phi_x,$$

$$A_1(x) \quad \stackrel{\text{def}}{=} \quad \frac{d}{dx}A(x) = 2\Phi_x\Phi_{-x} + 2x\phi_x\Phi_{-x} - x\phi_x - 2\phi_x^2,$$

$$A_2(x) \quad \stackrel{\text{def}}{=} \quad \frac{d}{dx}A_1(x) = \phi_x(4\Phi_{-x} - 2\Phi_x - 2x^2\Phi_{-x} + 2x\phi_x - 1 + x^2),$$

$$B(x) \quad \stackrel{\text{def}}{=} \quad 4\Phi_{-x} - 2\Phi_x - 2x^2\Phi_{-x} + 2x\phi_x - 1 + x^2,$$

$$B_1(x) \quad \stackrel{\text{def}}{=} \quad \frac{d}{dx}B(x) = -2\left(2\phi_x + 2x\Phi_{-x} - x\right) = -4x\left(\frac{\phi_x}{x} + \Phi_{-x} - \frac{1}{2}\right),$$

$$C(x) \quad \stackrel{\text{def}}{=} \quad \frac{\phi_x}{x} + \Phi_{-x} - \frac{1}{2},$$

$$C_1(x) \quad \stackrel{\text{def}}{=} \quad \frac{d}{dx}C(x) = -\phi_x\left(2 + \frac{1}{x^2}\right).$$

Since $C_1(x) < 0$ for any $x \neq 0$, and

$$
\begin{aligned}
C(-\infty) &= 1/2, \\
C(0-0) &= -\infty, \\
C(0+0) &= +\infty, \\
C(+\infty) &= -1/2,
\end{aligned}
$$

and toghether with $B_1(x) = -4xC(x)$,

$$\text{sgn}(B_1) = (+ - +), \qquad x = -\infty \to +\infty. \tag{4.37}$$

With (4.37) and

$$
\begin{aligned}
B(-\infty) &= -\infty, \\
B(0) &= 0, \\
B(+\infty) &= +\infty, \\
B_1(0) &= -4\phi_0 < 0,
\end{aligned}
$$

we obtain

$$\text{sgn}(B) = \text{sgn}(A_2) = (- + -+), \qquad x = -\infty \to +\infty. \tag{4.38}$$

With (4.38) and

$$
\begin{aligned}
A_1(-\infty) &= 0 - 0, \\
A_1(0) &= \frac{1}{2} - \frac{1}{\pi} > 0, \\
A_1(+\infty) &= 0 - 0,
\end{aligned}
$$

we obtain

$$\text{sgn}(A_1) = (-+-). \tag{4.39}$$

With (4.39) and

$$
\begin{aligned}
A(-\infty) &= 0 - 0, \\
A(0) &= 0, \\
A(+\infty) &= 0 + 0,
\end{aligned}
$$

we finally obtain

$$
A(x) \begin{cases}
> 0, & x > 0, \\
= 0, & x = 0, \\
< 0, & x < 0.
\end{cases}
$$

□

By Lemma 4.6.5, when $x > 0$, $I_1(x, y)$ crosses the x-axis from the negative to positive at $y = x$. Also, when $I_1(x, y)$ moves from positive to positive at $y = x$. In both cases, $I_1(x, y) = 0$ has the unique solution on $y > |x| = x$ by equations from (4.32) to (4.35).[4]

(II) When $x < 0$ The following lemma holds for

$$I_1(x, -x) = \phi_x \{ x^2 \Phi_x + x\phi_x + \Phi_x(\Phi_{-x} - \Phi_x) \}.$$

Lemma 4.6.6   *For any $x \neq 0$, $I_1(x, -x) > 0$.*

Proof.

$$
\begin{aligned}
A(x) &\overset{\text{def}}{=} \frac{I_1(x, -x)}{\phi_x} = xG_x + \Phi_x - 2\Phi_x^2, \\
A_1(x) &\overset{\text{def}}{=} \frac{d}{dx} A(x) = 2(G_x - 2\phi_x \Phi_x), \\
A_2(x) &\overset{\text{def}}{=} \frac{d}{dx} A_1(x) = 2(\Phi_x + 2x\phi_x \Phi_x - 2\phi_x^2), \\
A_3(x) &\overset{\text{def}}{=} \frac{d}{dx} A_2(x) = 2\phi_x(1 + 2\Phi_x - 2x^2\Phi_x + 6x\phi_x), \\
B(x) &\overset{\text{def}}{=} 1 + 2\Phi_x - 2x^2\Phi_x + 6x\phi_x, \\
B_1(x) &\overset{\text{def}}{=} \frac{d}{dx} B(x) = 4(2\phi_x - x\Phi_x - 2x^2\phi_x), \\
B_2(x) &\overset{\text{def}}{=} \frac{d}{dx} B_1(x) = 4[(2x^3 - 7x)\phi_x - \Phi_x], \\
B_3(x) &\overset{\text{def}}{=} \frac{d}{dx} B_2(x) = -4\phi_x(2x^4 - 13x^2 + 8)
\end{aligned}
$$

1) Regarding B_3

---

[4]By this argument, $\text{sgn}(J_2) = (+), (+ - +)$ never happens actually.

Since sgn($2x^4 - 13x^2 + 8$) = $(+ - + - +)$, we obtain sgn(B_3) = $(- + - + -)$. Also, $B_3(x) = 0$ when

$$x = \pm \frac{\sqrt{13 \pm \sqrt{105}}}{2}.$$

2) Regarding B_2

Two relative maximums are

$$B_2\left(-\frac{\sqrt{13 - \sqrt{105}}}{2}\right) \approx 1.1159,$$
$$B_2\left(\frac{\sqrt{13 + \sqrt{105}}}{2}\right) \approx -0.7488.$$

Also, since $B_2(-\infty) = 0 - 0$, and $B_2(+\infty) = -4 + 0$, we obtain sgn(B_2) = $(- + -)$.

3) Regarding B_1

Since $B_1(-\infty) = 0 - 0$, $B_1(0) = 8\phi_0$, and $B_1(+\infty) = -\infty$, we obtain sgn(B_1) = $(- + -)$.

4) Regarding B

Since $B(-\infty) = 1 - 0$, $B(0) = 2$, and $B(+\infty) = -\infty$, we obtain either sgn(B) = sgn(A_3) = $(+-)$, or $(+ - +-)$.

5) Regarding A_2

Since $A_2(-\infty) = 0 + 0$, and $A_2(+\infty) = 2 + 0$, we obtain either sgn(A_2) = $(+)$, or $(+ - +)$.

6) Regarding A_1

Since $A_1(-\infty) = 0 + 0$, $A_1(0) = 0$, $A_1(+\infty) = +\infty$, and $A_2(0) > 0$, we obtain sgn(A_1) = $(+ - +)$. [5]

7) Regarding A

Since $A(-\infty) = 0 + 0$, $A(0) = 0$, $A(+\infty) = +\infty$, and $A_1(0) = 0$, we obtain sgn(A) = $(+)$.   □

By Lemma 4.6.6 and (4.34), there exists an area where $I_1(x, y)$ increases on $y \in (x, -x)$. By equations from (4.32) to (4.35), $I_1(x, y) = 0$ has a unique solution over $y > |x| = -x$.[6]

[5]Therefore, sgn(A_2) = $(+)$ never happens.
[6]By this argument, sgn(J_2) = $(+)$, $(+ - +)$ never happens actually.

Chapter 5

Optimal Slice of a VWAP Trade

Abstract   This chapter derives a static optimal execution strategy of a VWAP trade, in which the optimal execution strategy can be calculated by an iteration of a single variable optimization, rather than by a multivariable optimization. Analytical solutions are derived in some cases. We show that optimal execution times lag behind expected market trading volume distribution since price volatility tends to have a positive correlation with market trading volume. In a basket trade, execution error can be reduced by spreading out execution times according to the correlation of price movement. We confirm our theoretical results with actual trading data and simulations.

5.1   Introduction

VWAP stands for volume-weighted average price during a certain trading period, and a VWAP trade refers a trade that uses VWAP as a benchmark. This chapter analyzes optimal execution of a VWAP trade since little academic work can be found on this issue although VWAP has become an industry standard, and several studies have been made on optimal strategies for general trade execution.

VWAP trades are overtaking fixed-price trades in stock markets these days for several reasons. First, foreign investors and individual investors reduce trading costs by VWAP trades. For example, foreign investors are often forced to place large orders before the market opens due to the time lag. In a fixed-price trade, brokers, who are generally risk averse, charge a large risk premium because of the size of the risk. However, investors can save on the risk premium by using VWAP as a benchmark because market directional risk remains the investors' responsibility.

Second, a block trade and determination of the option strike price require high price transparency, which can be achieved with a VWAP benchmark. In a block and option trade, brokers try to cover their exposure to price movement risk as closely to a benchmark as possible in order to avoid possible losses. If

investors use a price at a fixed time in the future as a benchmark, the closing price for instance, brokers are forced to trade intensively towards closing, which results in a huge market impact, and burdens investors with considerable costs. However, a VWAP benchmark can reduce the market impact of hedging trades and prevent manipulation of market prices.

Third, VWAP orders can be used to evaluate traders' performance since the effect of market directional movement, which is considered to be noise in performance measurements, is excluded in comparisons with VWAP. As a result, a number of institutional investors such as pension funds and mutual funds place VWAP orders.

While investors can place fixed-price orders at any time they like, VWAP orders are gathered by brokers before a trading period starts. Brokers then match buy and sell orders, and only the surplus is executed in the market. Note that aggregated VWAP orders tend to involve various stocks and consist of a few large orders and numerous small orders due to the characteristics mentioned above. Further, small orders for which buyers or sellers take the trouble to use a VWAP benchmark generally have low liquidity.

In executing market trades, a VWAP trader wishes to get the average execution price as close to the market VWAP as possible to avoid price movement risk. To do so, a VWAP trader slices a whole trade into small executions and tries to spread the executions in a well-balanced manner over the trading period, which makes VWAP trades rather labor intensive. Therefore, it often makes sense to build a standard strategy beforehand, and execute trades automatically according to it. A standard strategy is especially good for small orders and orders with small weights in a basket trade whose economic impact is small. Consequently, this chapter focuses on the optimal strategy for automatic execution of small VWAP orders with low liquidity.

As the investment business becomes increasingly competitive and understanding of the market microstructure improves, trading costs attract greater attention from both practitioners and academic researchers. As a consequence, several studies have examined optimal trade execution strategy. For example, Harris and Hasbrouck (1996) and Parlour (1998) analyze the choice of market and limit orders, and Bertsimas and Lo (1998) and Chapter 3 of this study derive an optimal slicing strategy for block trades. Although this chapter has similar objectives to those of preceding studies, both market price and trading volume are stochastic in a VWAP trade, while only the price process is stochastic in a block trade.

Numerous studies have been done on the relationship between price volatility and market trading volume since Clark's (1973) seminal paper, including Epps and Epps (1976), Tauchen and Pitts (1983). Karpoff (1987) summarizes these results, and Andersen (1996) proposes a modified model. These studies, in general, support the existence of a positive correlation between price volatility and market trading volume and the autocorrelation of themselves because surprising news boosts both price volatility and market trading volume. Besides, Jain and Joh (1988), Admati and Pfleiderer (1988), Foster and Viswanathan

(1990), and Chan et al. (1995) study intraday data and find a "U-shaped effect," i.e., within a single day, price volatility and market trading volume tend to be higher at the opening and closing than in the middle of a trading session. Naturally, our VWAP model reflects these findings.

Dynamic optimization, in which the actual trading strategy might be changed as new information arrives, is certainly a potential VWAP execution strategy. However, static optimization without strategy changes works well enough, especially for small VWAP orders with low liquidity. This is because observed variables such as the price volatility and market trading volume of stocks with low liquidity may contain statistical errors too large to be used in forecasting, and therefore, dynamic optimization is considered to be inaccurate. For example, when market trading volume suddenly surges, it is difficult to predict whether high market trading volume is temporary or permanent even judging qualitatively. In spite of inaccuracy and trouble related to dynamic optimization, the impact of small orders on cost efficiency is insignificant. Consequently, static optimization is considered a reasonable alternative.

Further, a static optimal execution strategy can be used as a benchmark for a dynamic execution strategy. A strategy change can be evaluated by comparing the realized cost to that of a static optimal execution strategy.

For these reasons, this chapter focuses on the static optimal execution strategy of a VWAP trade. Specifically, optimal trade decisions are made before the trading period, based on traders' expectations regarding price volatility, market trading volume, and the correlation between these values at different time of the day. No revision is allowed in the middle of the day after the true market trading volume is observed.

The rest of this chapter is organized as follows. Section 5.2 formulizes the optimal strategy. Section 5.3 derives solutions for single-stock cases, analyzes characteristics of the optimal strategy, and tests them against historical data. Section 5.4 analyzes a multiple-stock case. Finally, Section 5.5 concludes the analysis.

5.2 Formulization of the Optimal Slicing

Imagine a case in which a trader buys a certain number of shares and tries to get the average purchase price as close as possible to the market VWAP during the trading period from time 0 to time T. Let $(\Omega, \mathcal{F}, Q)$ be a probability space with a filtration $\{\mathcal{F}_t\}$, which satisfies usual conditions. Let $V(t)$ denote the accumulated market trading volume at time t excluding the trader's, and $v(t)$ the trader's accumulated trading volume. Here, $\{V(t)\}$ is an $\{\mathcal{F}_t\}$-adapted process in L^2 space, and $v(t)$ is a controllable variable. At time 0, the trader knows $v(0) = V(0) = 0$ and the whole trade size $v(T)$ which is given exogenously, but not $V(t)$ for $t > 0$. We do not allow sell orders, and therefore, $v(t)$ is a non-negative increasing function as well as $V(t)$. Further, minimum trading unit is given exogenously and normalized to one for

simplicity, which makes $V(t)$ and $v(t)$ step functions, and we assume left continuous.

Let $\{P(s)\}$ denote a non-negative $\{\mathcal{F}_t\}$-adapted stock price process in L^2 space. A trader's own VWAP is expressed as

$$vwap = \frac{\int_0^T P(s)dv(s)}{v(T)}.$$

Also, the market VWAP is

$$VWAP = \frac{\int_0^T P(s)d\{V(s) + v(s)\}}{V(T) + v(T)} = \frac{\int_0^T P(s)dV(s) + v(T) \times vwap}{V(T) + v(T)}.$$

As we have mentioned, a VWAP trader tries to get his VWAP as close to the market VWAP as possible. Generally speaking, the market impact of a VWAP trade is insignificant because the whole trade is divided into small executions. Therefore, this chapter ignores the market impact effect and minimizes the expected squared error of the VWAP execution due to price movement. Mathematically, our objective is represented as

$$\min_{v(t)} \mathrm{E}\left((VWAP - vwap)^2\right).$$

The following proposition holds regarding the expected squared error of the VWAP execution.

Proposition 5.2.1 *The expected squared error of the VWAP execution is calculated as*

$$W\mathrm{E}\left(\left\{\int_0^T (X(t) - x(t))dP(t)\right\}^2\right) + \mathrm{Cov}\left(\left(\frac{V(T)}{V(T) + v(T)}\right)^2, \left\{\int_0^T (X(t) - x(t))dP(t)\right\}^2\right)$$

where

$$X(t) = \frac{V(T) - V(t)}{V(T)}, \quad x(t) = \frac{v(T) - v(t)}{v(T)}, \quad W = \mathrm{E}\left(\left(\frac{V(T)}{V(T) + v(T)}\right)^2\right).$$

Proof.   See Section 5.6. $\square$

In this proposition, $X(t)$ and $x(t)$ stand for the ratio of remaining trading volume of the market and of the trader, respectively, and both of them are left continuous decreasing step functions. We call $x(t)$ an execution strategy function henceforth.

Further, when $V(T)$ is sufficiently large compared to $v(T)$, $\frac{V(T)}{V(T)+v(T)}$ has a value close to one, which makes its variation small and the covariance term negligible. Thus, we make the following assumption.

Assumption 5.2.1 *The expected squared error can be approximated as*

$$W\mathrm{E}\left(\left\{\int_0^T (X(t) - x(t))dP(t)\right\}^2\right). \tag{5.1}$$

Since this formula is still too complicated to generate implications, further assumptions are made which are considered reasonable based on standard theories and empirical analyses. First, the stock price

process is assumed to follow a Brownian motion, similar to standard option pricing theory. The drift term is considered to be zero because the time frame is short. Second, according to the "Mixture of Distribution Hypothesis" by Clark (1973) and empirical analysis by Jain and Joh (1988), there exists a relationship between price volatility and market trading volume which can be written as follows.

Assumption 5.2.2 *The stock price process is represented as*

$$dP(t) = \sigma(t, V(t))dB(t),$$

in which $\sigma(t, V(t))$ is a positive $\{\mathcal{F}_t\}$-adapted process in L^2 space, and $B(t)$ is a standard Brownian motion on $(\Omega, \mathcal{F}, Q)$.

We can assume that the probability of a negative $P(t)$ is negligible because the time frame is short.

Under these assumptions, (5.1) becomes

$$W \int_0^T \mathrm{E}\left(\sigma(t, V)^2 (X(t) - x(t))^2\right) dt. \tag{5.2}$$

This equation gives great insight into the difficulty of certain VWAP trades: a trade with a small expected squared error is a trade large enough to make W small and $w(t)$ smooth, with a stable market trading volume distribution over time, and with low price volatility.

Since our objective is to minimize the value of (5.2), our problem can be further modified as,

$$
\begin{aligned}
\min \int_0^T &\mathrm{E}\left(\sigma(t, V)^2 (X(t) - x(t))^2\right) dt \\
= \int_0^T &\mathrm{E}\left(\sigma(t, V)^2 X(t)^2\right) dt + \min \int_0^T \{-2\mathrm{E}\left(\sigma(t, V)^2 X(t)\right) x(t) + \mathrm{E}\left(\sigma(t, V)^2\right) x(t)^2]\} dt \\
= \int_0^T &\left\{ \mathrm{E}\left(\sigma(t, V)^2 X(t)^2\right) - \frac{\mathrm{E}\left(\sigma(t, V)^2 X(t)\right)^2}{\mathrm{E}\left(\sigma(t, V)^2\right)} \right\} dt \\
&+ \min \int_0^T \left\{ \frac{\mathrm{E}\left(\sigma(t, V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t, V)^2\right)} - x(t) \right\}^2 \mathrm{E}\left(\sigma(t, V)^2\right) dt
\end{aligned}
\tag{5.3}
$$

in which $x(t)$ is a controllable variable and others are given parameters.

In practice, expectations of $\sigma(t, V)^2$ and $X(t)$ vary gradually across time except at scheduled events such as batch auctions and news announcements. Therefore, $\dfrac{\mathrm{E}\left(\sigma(t, V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t, V)^2\right)}$ in (5.3) can be assumed to be continuous almost everywhere with finite jumps. Also, we can assume that $\dfrac{\mathrm{E}\left(\sigma(t, V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t, V)^2\right)}$ is left continuous without loss of generality, in the same way as $X(t)$ and $x(t)$.

Let t_k $(k = 1, \cdots, v(T))$ denote $v(T)$ of execution times, and define $t_0 = 0$ and $t_{v(T)+1} = T$. Then, $x(t)$ is a step function with $v(T) + 1$ values as follows.

$$x(t) = 1 - \frac{k}{v(T)} \quad \text{if} \quad t_k < t \le t_{k+1} \quad (k = 0, \cdots, v(T)). \tag{5.4}$$

Therefore, determination of the optimal execution strategy function $x^*(t)$ is equivalent to determination of optimal execution times $\{t_k^*\}$. Further, a careful examination of (5.3) reveals that this problem is considered to be an issue of how to approximate a function, continuous almost everywhere, with a step function as in Figure 5.1.

Figure 5.1: **Image of** $\dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$ **and** $x(t)$ (Curve: $\dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$, Step: $x(t)$).   Our problem is considered to be an issue of how to approximate $\dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$, a function differentiable almost everywhere with finite jumps, with $x(t)$, a left continuous step function with $v(T)+1$ values.

It is possible that t_k^* is equal to t_{k+1}^* where $\dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$ jumps at batch auctions. In this case, multiple units are executed simultaneously. This does not detract from our argument since we ignore trading costs, including the market impact. Note that if $v(T)$ is large enough, $x^*(t)$ can be approximated by

$$x^*(t) \approx \frac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}.$$

Mathematically, the following proposition holds.

Proposition 5.2.2

$$\min \int_0^T \left\{-2\mathrm{E}\left(\sigma(t,V)^2 X(t)\right) x(t) + \mathrm{E}\left(\sigma(t,V)^2\right) x(t)^2]\right\} dt$$

$$= \quad \min \sum_{k=1}^{v(T)} \int_0^{t_k} \left[\left\{ -\frac{2\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{v(T)\mathrm{E}\left(\sigma(t,V)^2\right)} + \frac{2(v(T)-k)+1}{v(T)^2} \right\} \mathrm{E}\left(\sigma(t,V)^2\right) \right] dt. \qquad (5.5)$$

Proof.   See Section 5.6.   □

Note that we can optimize each t_k separately since it appears only as the upper bound of the integral on the right hand side of (5.5). The specific solution is derived in the following section.

5.3 Derivation of the Optimal Strategy: Single-Stock Case

In this section, we derive solutions both when price volatility is uncorrelated and correlated with market trading volume, and compare these results.

5.3.1 Uncorrelated Case: σ is uncorrelated with V

In this case, the optimal execution strategy can be solved explicitly as the following proposition.

Proposition 5.3.1 *Under Assumptions 5.2.1 and 5.2.2, and if price volatility is uncorrelated with market trading volume, optimal execution times $\{t_k^*\}$ are determined to satisfy*

$$\mathrm{E}\left(X(t_k^*)\right) \geq \frac{2(v(T)-k)+1}{2v(T)} \geq \lim_{t\downarrow t_k^*} \mathrm{E}\left(X(t)\right).$$

Further, the optimal execution strategy function $x^*(t)$ is solved as

$$x^*(t) = \frac{\lfloor (2v(T)\mathrm{E}\left(X(t)\right)+1)/2 \rfloor}{v(T)},$$

where $\lfloor x \rfloor$ represents the integer part of x.

Proof.   See Section 5.6.   □

Therefore, $\{t_k^*\}$ is determined only by $\mathrm{E}\left(X(t)\right)$, and is independent of expectations regarding the magnitude and term structure of price volatility. Note that $x^*(t)$ is solved directly without solving $\{t_k^*\}$.

In practice, it used to be common to divide trading periods into equal time intervals and to allocate executions proportional to $\mathrm{E}\left(X(t)\right)$. However, this strategy (Old Strategy henceforth) comes with attendant problems, namely that as intervals are shortened in an attempt to track market trading volume more closely, more orders are placed at the opening and closing when instantaneous market trading volume is larger than in the middle of the trading session. In contrast, our optimal strategy (New Strategy henceforth) is derived by dividing $\mathrm{E}\left(X(t)\right)$ into the number of shares and executing at the middle of $\mathrm{E}\left(X(t)\right)$ in each interval. In other words, the New Strategy takes price movement risk into account and minimizes time differences in order execution. Note that the interval with high market trading volume is not necessarily the optimal execution time. Consequently, the New Strategy provides not only an optimal slice with the least expected squared error but also an unbiased slice for any orders even if the time intervals are infinitesimally small.

Next, we confirm our theoretical results empirically, using actual intraday data from the Tokyo Stock Exchange provided by Nikkei Quick Information Co. Ltd. All empirical analyses in this chapter are based on trading data from October 2000 to March 2001 on the 10 stocks with the largest market capitalization. Data from December 29, 2000 and January 4, 2001 are excluded since there was no afternoon session

on these days. We chose these large stocks in order to retain statistical accuracy since their ticks are frequent and stable. We believe that results can be applied to other stocks since the relationship between price volatility and market trading volume is ubiquitous for all stocks as Chan et al. (1995) point out.

Figure 5.2 shows the standard deviation of VWAP execution errors using the Old Strategy and the New Strategy when the whole day is divided into nine intervals of 30 minutes each. Specifically, we first estimate $E(X(t))$ of each stock as the average of past trading data. Then, in the Old Strategy, fixed number of executions are allocated proportional to the increment of $E(X(t))$ for the nine intervals with the smallest rounding errors. In the New Strategy, executions are allocated by the prescribed method. For comparison, we assume that trades are executed at VWAP of each interval in both Old and New Strategy. This allocation is fixed during the whole sample period, and VWAP execution errors are calculated everyday. Finally, standard deviation of VWAP execution errors is calculated over sample stocks and dates.

According to Figure 5.2, the superiority of the New Strategy, especially for small orders, is evident. The superiority is more conspicuous when time intervals are shortened, although we do not show the result here.

Figure 5.2: **Comparison of execution strategies.**   This graph shows the standard deviation of VWAP execution errors using the *old strategy* and the *new strategy* (y-axis) for different order sizes (x-axis) when the whole day is divided into nine intervals of 30 minutes each. The superiority of the new strategy, especially for small orders, is evident.

5.3.2   Correlated Case: σ is a function of V

In this case, the integrand in (5.5) is not necessarily increasing. However, what we have to do is essentially the same; to find the upper bounds of the integrals one by one so as to minimize the area between the straight line and the curved line in Figure 5.3: negative if the curved line is below the straight line,

positive if the curved line is above the straight line. As we can see in Figure 5.3, the upper bounds fall on one of the countable points where the curved line rises to intersect the straight line (A or G for t_k^*, and B or H for t_{k+1}^* in Figure 5.3). Therefore, we have only to evaluate the integral at these points and choose one that minimizes the integral.

Figure 5.3: **Optimization image** (curve: $\dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$, $L(k)$: $-\dfrac{v(T)-k+1/2}{v(T)}$, $L(k+1)$: $-\dfrac{v(T)-(k+1)+1/2}{v(T)}$). Although the integrand is not necessarily increasing, we still want to find the upper bounds of integrals one by one so as to minimize the area between the straight line and the curved line. It turns out that su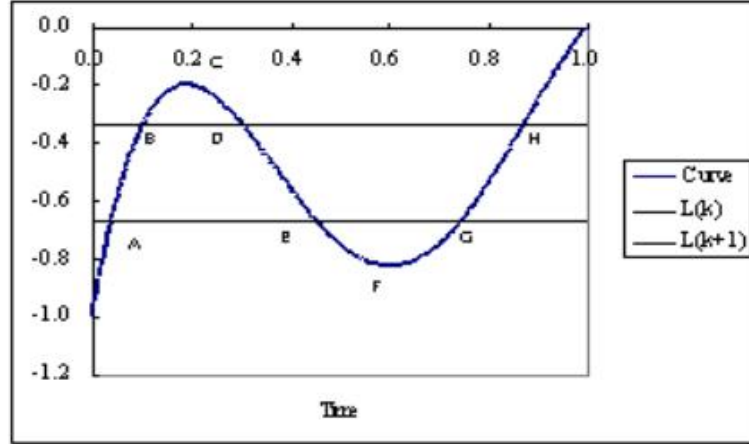ch a point falls on one of the countable points where the curved line rises to intersect the straight line (A or G for t_k^*, and B or H for t_{k+1}^*). Therefore, we have only to evaluate the integral at these points and choose one that minimizes the integral.

Note that t_k^* and t_{k+1}^* chosen separately in the above procedure always satisfy $t_k^* \le t_{k+1}^*$. This is obvious when t_k^* falls on A. If t_k^* falls on G, Area(ACE) < Area(EFG) holds. Together with Area(BCD) < Area(ACE) and Area(EFG) < Area(DFH), Area(BCD) < Area(DFH). Consequently, t_{k+1}^* falls on H and is larger than t_k^*.

As a matter of fact, $\dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$ is usually stable and monotone decreasing in practice, as we confirm with actual data later. Therefore, we make the following assumption in the rest of Section 5.3.

Assumption 5.3.1 $\dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$ *is monotone decreasing.*

Under this assumption, the term in the bracket in (5.5)

$$\left\{ -\frac{2\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{v(T)\mathrm{E}\left(\sigma(t,V)^2\right)} + \frac{2(v(T)-k)+1}{v(T)^2} \right\}$$

is increasing and turns from negative to positive just once. Further, even if the integrand itself

$$\left\{ -\frac{2\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{v(T)\mathrm{E}\left(\sigma(t,V)^2\right)} + \frac{2(v(T)-k)+1}{v(T)^2} \right\} \mathrm{E}\left(\sigma(t,V)^2\right)$$

is not increasing, the integrand turns from negative to positive only once since $E[\sigma(t, V)^2]$ is positive. Consequently, we can obtain the optimal solution explicitly at such a unique point just as in the Uncorrelated Case, and the following proposition holds.

Proposition 5.3.2   *Under Assumptions 5.2.1 – 5.3.1, and if price volatility is correlated with market trading volume, optimal execution times $\{t_k^*\}$ are determined to satisfy*

$$\mathrm{E}\left(\sigma(t_k^*, V)^2 X(t_k^*)\right) \geq \frac{2(v(T) - k) + 1}{2v(T)} \mathrm{E}\left(\sigma(t_k^*, V)^2\right) \geq \lim_{t \downarrow t_k^*} \mathrm{E}\left(\sigma(t, V)^2 X(t)\right).$$

Further, the optimal execution strategy function $x^*(t)$ is solved as

$$x(t) = \frac{\lfloor (\frac{2v(T)\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}(\sigma(t,V)^2)} + 1)/2 \rfloor}{v(T)}$$

where $\lfloor x \rfloor$ represents the integer part of x.

According to Clark (1973), Jain and Joh (1988), and others, $\sigma(t, V)$ is considered to be an increasing function of market trading volume, or equivalently, an increment of $V(t)$ because surprising news boosts both $\sigma(t, V)$ and the increment of $V(t)$. Also, since the increment of $V(t)$ tends to have a positive autocorrelation, $X(t)$ is an increasing function of the increment of $V(t)$. Therefore, in general, the correlation between $\sigma(t, V)^2$ and $X(t)$ is positive, or equivalently, $\mathrm{E}\left(\sigma(t, V)^2 X(t)\right) > \mathrm{E}\left(\sigma(t, V)^2\right) \mathrm{E}\left(X(t)\right)$. Consequently, $x^*(t)$ in the Correlated Case tends to become larger than in the Uncorrelated Case, which delays trade execution. This is because once price volatility and market trading volume surge, traders have to make considerable trades to track market trading volume while execution times do not matter when these parameters remain small throughout the day.

There might be two approaches to check the monotonicity of $\frac{\mathrm{E}\left(\sigma(t, V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t, V)^2\right)}$: theoretical and empirical. In the theoretical approach, we can find sufficient conditions with existing models such as Clark (1973), Admati and Pfleiderer (1988), Andersen (1996), and others, and check that actual data satisfies such conditions. While these conditions may take various forms, depending on parameter specifications in the models, their implications are similar: stable correlation, reasonable volatility, etc. On the other hand, we can show the same results in the empirical approach without assuming any specific models. Since the empirical approach is more robust, we check with past data that actual $\frac{\mathrm{E}\left(\sigma(t, V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t, V)^2\right)}$ is indeed decreasing.

Table 5.1 shows the optimal allocation ratio of the VWAP execution for individual stocks in the Correlated Case, using the previous 10-stock trading data. Prices are normalized so that those at the end of March are one. Also, $v(t)$ is assumed to be large enough that $x^*(t) \approx \frac{\mathrm{E}\left(\sigma(t, V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t, V)^2\right)}$. $x^*(t)$ here is calculated from the average of $\sigma(t, V)^2 X(t)$, and $\sigma(t, V)^2$ over the sample period. Specifically, let $P_{d,t}$ and $V_{d,t}$ denote price and accumulated market trading volume in d^{th} day t^{th} interval ($d = 1, \cdots, D$,

$t = 1, \cdots, T$). Price change can be represented as

$$\Delta P_{d,t} = \begin{cases} P_{d,t} - P_{d,j-1}, & \text{for} \quad t = 2, \cdots, T, \\ P_{d,1} - P_{d-1,T}, & \text{for} \quad t = 1 \end{cases}$$

and realized ratio of remaining trading volume as $X_{d,t} = \dfrac{V_{d,T} - V_{d,t-1}}{V_{d,T}}$. We estimate $\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)$ as

$\dfrac{1}{D}\displaystyle\sum_{d=1}^{D}\Delta P_{d,t}^2 X_{d,t}$, and estimate $\mathrm{E}\left(\sigma(t,V)^2\right)$ as $\dfrac{1}{D}\displaystyle\sum_{d=1}^{D}\Delta P_{d,t}^2$. Let x_t^* denote the optimal ratio of remaining

execution in the tth interval, then the optimal allocation ratio of the VWAP execution is given as $-(x_{t+1}^* - x_t^*)$.

Period	1	2	3	4	5	6	7	8	9
Takeda	0.22	0.09	0.07	0.10	0.14	0.04	0.09	0.07	0.17
Matsushita	0.22	0.09	0.07	0.10	0.14	0.04	0.09	0.07	0.17
Sony	0.21	0.09	0.10	0.09	0.11	0.07	0.06	0.10	0.16
Toyota	0.23	0.04	0.12	0.05	0.13	0.09	0.06	0.01	0.27
Honda	0.21	0.10	0.09	0.09	0.12	0.03	0.09	0.09	0.18
Mizuho	0.19	0.08	0.08	0.09	0.11	0.10	0.08	0.08	0.18
Tokyo-Mitsubishi	0.21	0.06	0.09	0.06	0.13	0.02	0.14	0.08	0.20
Nomura	0.22	0.09	0.08	0.08	0.13	0.07	0.09	0.09	0.15
NTT	0.21	0.08	0.09	0.07	0.12	0.06	0.10	0.07	0.21
NTT Docomo	0.19	0.11	0.11	0.08	0.10	0.09	0.05	0.09	0.18

Table 5.1: **Optimal allocation ratio of execution.** This table shows the optimal allocation ratio of the VWAP execution for individual stocks in the correlated case, using the previous 10 stock trading data. $v(t)$ is assumed to be large enough that $x^*(t) \approx \dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$. $X^*(t)$ here is calculated from the average of $\sigma(t,V)^2 X(t)$ and $\sigma(t,V)^2$ over the sampling period. As we expected, all the optimal allocation ratios of execution in the correlated case are positive, which means $\dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$ is indeed monotone decreasing.

As expected, all of the optimal execution allocation ratios of execution in Table 5.1 are positive, which means $\dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$ is indeed monotone decreasing.

Also, Figure 5.4 shows the average of $x^*(t)$, which coincides with $\mathrm{E}\left(X(t)\right)$ in the Uncorrelated Case and $\dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$ in the Correlated Case. $\mathrm{E}\left(X(t)\right)$ here is also estimated as an average over the sample period.

We can see that $x^*(t)$ in the Correlated Case is, in fact, larger than in the Uncorrelated Case, which delays trade execution. Besides, according to the same data, standard deviations of VWAP execution error are 0.127% in the Uncorrelated Case and 0.119% in the Correlated Case, which shows that a VWAP trade can be executed more accurately by taking into account the correlation between price volatility and market trading volume.

Main sources of these errors are considered to be statistical errors in estimating parameters and stochastic changes of parameters. Regarding the statistical errors, empirical studies such as Uno and

Yamada (1993) and Jain and Joh (1988) report that the parameters of the market impact coefficents, the price volatility, and the market trading volume are rather stable. Regarding stochastic changes, execution errors can be reduced by dynamic optimization in Chapter 6. In practice, we further suffer price changes within the interval and discritization errors due to the minimum trading unit, both of which are considered to be proportional to the price volatility.
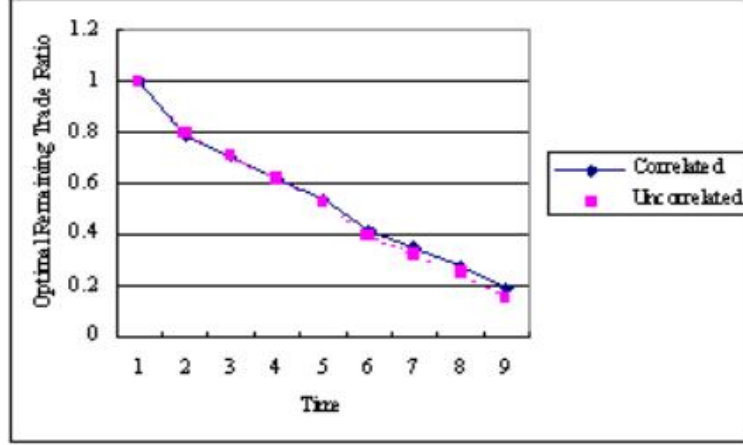

Figure 5.4: **Comparison of average optimal execution strategy functions.** This graph shows the average of $x^*(t)$, which coincides with $\mathrm{E}\left(X(t)\right)$ in the uncorrelated case and $\dfrac{\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{\mathrm{E}\left(\sigma(t,V)^2\right)}$ in the correlated case. We can see that $x^*(t)$ in the correlated case is, in fact, larger than in the uncorrelated case, which delays trade execution.

5.4 Derivation of the Optimal Strategy: Multiple-Stock Case

5.4.1 Formulization of the Optimal Slicing

Imagine a case with N stocks. Price, price volatility, the accumulated trading volume of the market and the trader, etc. of ith stock are represented by subscripts: $P_i(t)$, $\sigma_i(t)$, $V_i(t)$, $v_i(t)$, respectively, and satisfy similar conditions as in single-stock case. Also, let t_{ik} $(i = 1, \cdots, N, k = 1, \cdots, v_i(T))$ denote $v_i(T)$ of execution times of ith stock. Further, define the weights of ith stock in the whole trade as

$$w_i = \frac{v_i(T)}{\sum_{j=1}^{N} v_j(T)}.$$

The VWAP of the whole trade can be defined as the weighted average of the individual VWAP. In this case, the following equation holds as Proposition 5.2.1 in the single-stock case,

$$\mathrm{E}\left(\{\sum_{i=1}^{N} w_i(VWAP_i - vwap_i)\}^2\right)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} \mathrm{E} \left(\int_0^T (X_i(t) - x_i(t)) dP_i(t) \int_0^T (X_j(t) - x_j(t)) dP_j(t) \right)$$

$$+ \mathrm{Cov} \left(w_i \left(\frac{V_i(T)}{V_i(T) + v_i(T)} \right) w_j \left(\frac{V_j(T)}{V_j(T) + v_j(T)} \right), \int_0^T (X_i(t) - x_i(t) dP_i(t) \int_0^T (X_j(t) - x_j(t)) dP_j(t) \right)$$

where

$$X_i(t) = \frac{V_i(T) - V_i(t)}{V_i(T)}, \quad x_i(t) = \frac{v_i(T) - v_i(t)}{v_i(T)}, \quad W_{ij} = \mathrm{E} \left(w_i \left(\frac{V_i(T)}{V_i(T) + v_i(T)} \right) w_j \left(\frac{V_j(T)}{V_j(T) + v_j(T)} \right) \right).$$

As in the single-stock case, when $V_i(T)$ is large enough compared to $v_i(T)$, $\dfrac{V_i(T)}{V_i(T) + v_i(T)}$ has a value close to one, which makes its variation small. Therefore, the covariance term becomes negligible, and the expected squared error of the whole VWAP execution can be approximated as

$$\sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} \mathrm{E} \left(\int_0^T (X_i(t) - x_i(t)) dP_i(t) \int_0^T (X_j(t) - x_j(t)) dP_j(t) \right).$$

Let stock prices follow a Brownian motion with a correlation $\rho_{ij}(t, V_i(t), V_j(t))$, i.e.,

$$dP_i(t) = \sigma_i(t, V_i(t)) dB_i(t),$$
$$dB_i(t) dB_j(t) = \rho_{ij}(t, V_i(t), V_j(t)) dt,$$

in which $\{\rho_{ij}(t, V_i(t), V_j(t))\}$ is an $\{\mathcal{F}_t\}$-adapted process in L^2 space, and $\{B_i(t)\}$ is a standard Brownian motion on $(\Omega, \mathcal{F}, Q)$. In this case, equation above becomes

$$\sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} \int_0^T \mathrm{E} \left(\rho_{ij}(t, V_i(t), V_j(t)) \sigma_i(t, V_i(t)) \sigma_j(t, V_j(t)) (X_i(t) - x_i(t))(X_j(t) - x_j(t)) \right) dt \qquad (5.6)$$

and the following proposition holds for our optimization problem.

Proposition 5.4.1

$$\min \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} \int_0^T \mathrm{E} \left(\rho_{ij} \sigma_i \sigma_j (X_i - x_i)(X_j - x_j) \right) dt$$

$$= \min \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} \sum_{k=1}^{v_i(T)+v_j(T)} \int_0^{t_k^{ij}} \left\{ -\frac{\mathrm{E} \left(\rho_{ij} \sigma_i \sigma_j X_j \right)}{v_i \mathrm{E} \left(\rho_{ij} \sigma_i \sigma_j \right)} + \frac{x_i(t_{k-1}^{ij}) x_j(t_{k-1}^{ij}) - x_i(t_k^{ij}) x_j(t_k^{ij})}{2} \right\} \mathrm{E} \left(\rho_{ij} \sigma_i \sigma_j \right) dt$$

$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} \int_0^T \mathrm{E} \left(\rho_{ij} \sigma_i \sigma_j X_i X_j \right) dt$$

in which $\{t_k^{ij}\}$ is a sequence which renumbers $\{\{t_{ik}\}, \{t_{jk'}\}\}$, and arguments in $\rho_{ij}(t, V_i(t), V_j(t))$, $\sigma_i(t, V_i(t))$, and $X_i(t)$ are omitted due to the simplicity of the representation.

Proof.   See Section 5.6.   □

It is again a problem to determine the upper bounds of the integrals one by one, as in the single-stock case. Therefore, once orders among $\{t_k^{ij}\}$ are assumed, we can find the solution by evaluating the integrals at countable points where the integrand turns from negative to positive, and choosing one that minimizes the integral. However, if the order of execution times contradicts this assumption, we have to adjust them so as to minimize the corresponding integrals simultaneously. In this case, other countable points where the sum of these integrands turns from negative to positive are also candidates.

To determine the optimal order among $\{t_k^{ij}\}$, we need to calculate optimal solutions in several cases. The order with the minimum expected squared error of execution is then the global optimal strategy. In practice, we can divide the whole trade into small clusters that are hardly correlated with each other, and then find the order among highly correlated assets within each cluster. Another effective way is to start with a stock with a large weight and then choose other parameters recursively.

Because this equation is too general to produce deep implications, we analyze a specific numerical example in the following subsection.

5.4.2   Numerical Example

If $t_{11} \leq t_{12}$, (5.6) becomes

$$\int_0^{t_{11}} W_1^2 \sigma_1^2 (1 - 2\mathrm{E}\,(X_1(s))) + 2W_{12}\rho_{12}\sigma_1\sigma_2(1 - \mathrm{E}\,(X_2(s)))ds$$
$$+ \int_0^{t_{21}} W_2^2 \sigma_2^2 (1 - 2\mathrm{E}\,(X_2(s))) + 2W_{12}\rho_{12}\sigma_1\sigma_2(-\mathrm{E}\,(X_1(s)))ds \tag{5.7}$$

where $W_1 = \sqrt{W_{11}}$, and $W_2 = \sqrt{W_{22}}$. Since σ_1, σ_2, and ρ_{12} are constants, the integrands in (5.7) become monotone increasing functions, and the optimal strategy can be solved explicitly as in Section 5.3. Specifically, t_{11}^* and t_{12}^* are chosen so that

$$\begin{cases} W_1^2 \sigma_1^2 (1 - 2\mathrm{E}\,(X_1(t_{11}^*))) + 2W_{12}\rho_{12}\sigma_1\sigma_2\{1 - \mathrm{E}\,(X_2(t_{11}^*))) = 0, \\ W_2^2 \sigma_2^2 (1 - 2\mathrm{E}\,(X_2(t_{21}^*))) + 2W_{12}\rho_{12}\sigma_1\sigma_2\{-\mathrm{E}\,(X_1(t_{21}^*))) = 0. \end{cases} \tag{5.8}$$

For example, if market trading volume is constant,

$$\mathrm{E}\,(X_1(t)) = \mathrm{E}\,(X_2(t)) = 1 - \frac{1}{T}.$$

Further, if $W_{12} = W_1 W_2$ holds, (5.8) then becomes,

$$\begin{cases} W_1^2 \sigma_1^2 (1 - 2\mathrm{E}\,(X_1(t_{11}^*))) + 2\rho_{12}W_1\sigma_1 W_2\sigma_2\{1 - \mathrm{E}\,(X_2(t_{11}^*))\} \\ = 2W_1\sigma_1(W_1\sigma_1 + \rho_{12}W_2\sigma_2)\dfrac{t_{11}^*}{T} - W_1^2\sigma_1^2 = 0, \\ W_2^2 \sigma_2^2 (1 - 2\mathrm{E}\,(X_2(t_{21}^*))) + 2\rho_{12}W_1\sigma_1 W_2\sigma_2\{-\mathrm{E}\,(X_1(t_{21}^*))\} \\ = 2W_2\sigma_2(W_2\sigma_2 + \rho_{12}W_1\sigma_1)\dfrac{t_{21}^*}{T} - W_2^2\sigma_2^2 - 2\rho_{12}W_1\sigma_1 W_2\sigma_2 = 0. \end{cases}$$

So,

$$\begin{cases} t_{11}^* = \dfrac{W_1\sigma_1 T}{2(W_1\sigma_1 + \rho_{12}W_2\sigma_2)}, \\ t_{21}^* = \dfrac{(W_2\sigma_2 + 2\rho_{12}W_1\sigma_1)T}{2(W_2\sigma_2 + \rho_{12}W_1\sigma_1)}. \end{cases} \tag{5.9}$$

Besides, in this case, (5.7) is

$$\rho_{12}W_1^4\sigma_1^4 + (1 + 4\rho_{12}^2)W_1^3\sigma_1^3W_2\sigma_2 + 4\rho_{12}(1 + \rho_{12}^2)W_1^2\sigma_1^2W_2^2\sigma_2^2 + (1 + 4\rho_{12}^2)W_1\sigma_1W_2^3\sigma_2^3 + \rho_{12}W_2^4\sigma_2^4.$$

Therefore, the solution for $t_{21}^* \leq t_{11}^*$ has the same execution error due to symmetry, and so, (5.9) is the optimal strategy.

Figure 5.5 shows t_{11}^* and t_{21}^* for different price correlations when $W_1\sigma_1 = 0.2$, $W_2\sigma_2 = 0.1$, and $T = 1$. If $\rho_{12} = 0$, optimal execution times coincide with the solution optimized independently ($t_{11}^* = t_{21}^* = 0.5$). If $\rho_{12} > 0$, optimal execution times are spread over the trading period according to correlation. In this case, the adjustment will be larger for the stock with the smaller $W\sigma$.

Figure 5.5: **Optimal execution times for two stocks.** This graph shows t_{11}^* and t_{21}^* for different price correlations when $W_1\sigma_1 = 0.2$, $W_2\sigma_2 = 0.1$, and $T = 1$. If $\rho_{12} = 0$, optimal execution times coincide with the solution optimized independently ($t_{11}^* = t_{21}^* = 0.5$). If $\rho_{12} > 0$, optimal execution times are spread over the trading period according to correlation. In this case, the adjustment will be larger for the stock with the smaller $W\sigma$.
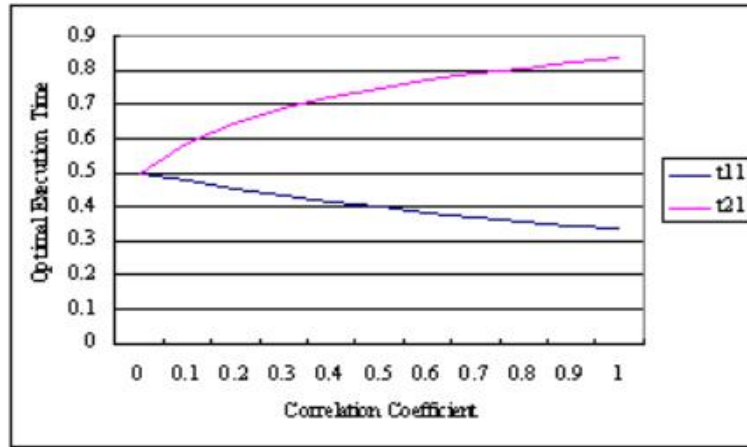
Figure 5.6 shows the standard deviation of VWAP execution errors. Evidently the VWAP execution error becomes smaller when execution times are spread out, considering the portfolio effect. It is interesting to note that although the total risk of a portfolio with a positive correlation is larger than that without correlation, this is not the case for VWAP execution risk.

5.5 Closing Remarks

This chapter derives the static optimal execution strategy of a VWAP trade that minimizes the expected squared execution error. This method is powerful because the optimal execution strategy is determined by an iteration of a single variable optimization, rather than by a multivariable optimization. Analytical
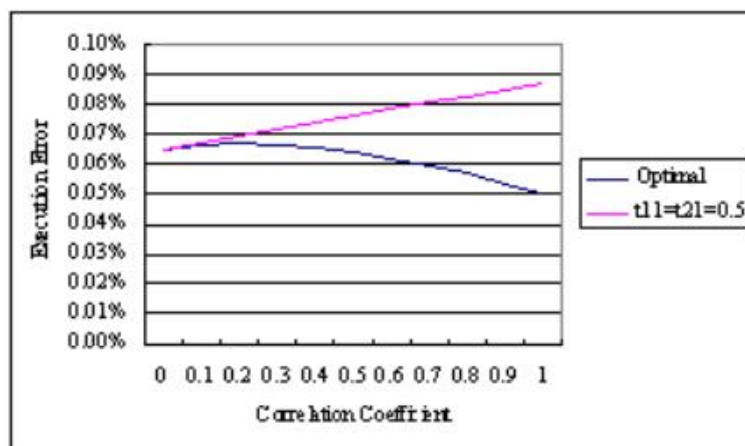
Figure 5.6: **Execution error for two stocks.**   This graph shows the standard deviation of VWAP execution errors (y-axis) for different price correlations (x-axis). Evidently the VWAP execution error becomes smaller when execution times are spread, considering the portfolio effect. It is interesting to note that although the total risk of a portfolio with a positive correlation is larger than that without correlation, this is not the case for VWAP execution risk.

solutions are derived in some cases. The following results are obtained through our analysis. For a single-stock trade, if price volatility is independent of market trading volume, the optimal execution strategy is determined only by the expected market trading volume distribution and is independent of expectations regarding the magnitude and time dependency of price volatility. If price volatility is positively correlated with market trading volume, optimal execution times turn out to lag behind the expected market trading volume distribution. This is because once price volatility and market trading volume surge, traders have to make considerable trades to track market trading volume while execution times do not matter when these parameters remain small throughout the day. In a basket trade, execution error can be reduced by spreading out execution times according to the correlation of price movement. Further, we examine these theoretical results with actual trading data and simulations.

This chapter focuses on static optimization since observed variables such as price volatility and the market trading volume of small orders with low liquidity may contain statistical errors too large to be used in forecasting, and there are some concerns that dynamic optimization is inaccurate. However, we might be able to predict the behavior of stocks with frequent trades to some extent, and dynamic optimization may reduce execution error further in such cases. Dynamic optimization is analyzed in Chapter 6.

For simplicity, this chapter minimizes execution error caused by price movement. In the real world, however, the direct cost of market impact and the indirect cost of information leakage are certainly significant factors in a VWAP trade. This issue is left for further analysis.

5.6 Appendix

5.6.1 Proof of Proposition 5.2.1

The expected square error of the whole VWAP transaction is

$$
\begin{aligned}
\mathrm{E}\left((VWAP - vwap)^2\right) &= \mathrm{E}\left(\left(\frac{\int_0^T P(s)dV(s) + v(T) \times vwap}{V(T) + v(T)} - vwap\right)^2\right) \\
&= \mathrm{E}\left(\left(\frac{\int_0^T P(s)dV(s) - V(T) \times vwap}{V(T) + v(T)}\right)^2\right) \\
&= \mathrm{E}\left(\left\{\left(\frac{V(T)}{V(T) + v(T)}\right)\left(\frac{\int_0^T P(s)dV(s)}{V(T)} - \frac{\int_0^T P(s)dv(s)}{v(s)}\right)\right\}^2\right).
\end{aligned}
$$

Since $P(s) = \int_0^s dP(t) + P(0)$,

$$
\begin{aligned}
\int_0^T P(s)dV(s) &= \int_0^T \int_0^s dP(t)dV(s) + \int_0^T P(0)dV(s) \\
&= \int_0^T \int_t^T dV(s)dP(t) + P(0)V(T).
\end{aligned}
$$

So,

$$
\begin{aligned}
\frac{\int_0^T P(s)dV(s)}{V(T)} &= \frac{\int_0^T \int_t^T dV(s)dP(t) + P(0)V(T)}{V(T)} \\
&= \frac{\int_0^T \{V(T) - V(t)\}dP(t)}{V(T)} + P(0) \\
&= \int_0^T X(t)dP(t) + P(0).
\end{aligned}
$$

Therefore, the equation above becomes

$$
\begin{aligned}
&\mathrm{E}\left(\left\{\left(\frac{V(T)}{V(T) + v(T)}\right)\int_0^T (X(t) - x(t))dP(t)\right\}^2\right) \\
&= \mathrm{E}\left(\left(\frac{V(T)}{V(T) + v(T)}\right)^2\right)\mathrm{E}\left(\left\{\int_0^T (X(t) - x(t))dP(t)\right\}^2\right) \\
&\qquad + \mathrm{Cov}\left(\left(\frac{V(T)}{V(T) + v(T)}\right)^2, \left\{\int_0^T (X(t) - x(t))dP(s)\right\}^2\right).
\end{aligned}
$$

5.6.2 Proof of Proposition 5.2.2

$$
\begin{aligned}
&\min \int_0^T \{-2\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)x(t) + \mathrm{E}\left(\sigma(t,V)^2\right)x(t)\}dt \\
&= \min \sum_{k=1}^{v(T)} \int_{t_{k-1}}^{t_k} \{-2\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)x(t) + \mathrm{E}\left(\sigma(t,V)^2\right)x(t)^2\}dt.
\end{aligned}
$$

The first term is calculated as

$$\sum_{k=1}^{v(T)} \int_{t_{k-1}}^{t_k} \left\{ -2\mathrm{E}\left(\sigma(t,V)^2 X(s)\right) x(t) \right\} dt$$

$$= \sum_{k=1}^{v(T)} \int_{t_{k-1}}^{t_k} \left\{ -2\mathrm{E}\left(\sigma(t,V)^2 X(s)\right) \left(1 - \frac{k-1}{v(T)}\right) \right\} dt$$

$$= \sum_{k=1}^{v(T)} (v(T) - k + 1) \int_{t_{k-1}}^{t_k} \left\{ -\frac{2\mathrm{E}\left(\sigma(t,V)^2 X(s)\right)}{v(T)} \right\} dt$$

$$= \sum_{k=1}^{v(T)-1} (v(T) - k) \int_{t_{k-1}}^{t_k} \left\{ -\frac{2\mathrm{E}\left(\sigma(t,V)^2 X(s)\right)}{v(T)} \right\} dt + \int_{0}^{t_v} \left\{ -\frac{2\mathrm{E}\left(\sigma(t,V)^2 X(s)\right)}{v(T)} \right\} dt$$

$$= \sum_{k=1}^{v(T)-i} (v(T) - k + 1 - i) \int_{t_{k-1}}^{t_k} \left\{ -\frac{2\mathrm{E}\left(\sigma(t,V)^2 X(s)\right)}{v(T)} \right\} dt + \sum_{k=v(T)+1-i}^{v(T)} \int_{0}^{t_k} \left\{ -\frac{2\mathrm{E}\left(\sigma(t,V)^2 X(s)\right)}{v(T)} \right\} dt$$

$$= \sum_{k=1}^{v(T)} \int_{0}^{t_k} \left\{ -\frac{2\mathrm{E}\left(\sigma(t,V)^2 X(s)\right)}{v(T)} \right\} dt.$$

Also, the second term becomes

$$\sum_{k=1}^{v(T)} \int_{t_{k-1}}^{t_k} \left\{ \mathrm{E}\left(\sigma(t,V)^2\right) x(t)^2 \right\} dt$$

$$= \sum_{k=1}^{v(T)} \int_{0}^{t_k} \mathrm{E}\left(\sigma(t,V)^2\right) x(t_{k-1})^2 dt - \sum_{k=1}^{v(T)} \int_{0}^{t_{k-1}} \mathrm{E}\left(\sigma(t,V)^2\right) x(t_{k-1})^2 dt$$

$$= \sum_{k=1}^{v(T)} \int_{0}^{t_k} \mathrm{E}\left(\sigma(t,V)^2\right) x(t_{k-1})^2 dt - \sum_{k=1}^{v(T)-1} \int_{0}^{t_k} \mathrm{E}\left(\sigma(t,V)^2\right) x(t_k)^2 dt$$

$$= \sum_{k=1}^{v(T)} \int_{0}^{t_k} \left\{ \mathrm{E}\left(\sigma(t,V)^2\right) x(t_{k-1})^2 \right\} dt$$

$$\quad - \left[\sum_{k=1}^{v(T)} \int_{0}^{t_k} \mathrm{E}\left(\sigma(t,V)^2\right) x(t_k)^2 dt + \int_{0}^{t_0} \mathrm{E}\left(\sigma(t,V)^2\right) x(t_0)^2 dt - \int_{0}^{t_v} \mathrm{E}\left(\sigma(t,V)^2\right) x(t_v)^2 dt \right]$$

$$= \sum_{k=1}^{v(T)} \int_{0}^{t_k} \mathrm{E}\left(\sigma(t,V)^2\right) \left\{ x(t_{k-1})^2 - x(t_k)^2 \right\} dt$$

$$= \sum_{k=1}^{v(T)} \int_{0}^{t_k} \mathrm{E}\left(\sigma(t,V)^2\right) \left\{ \left(\frac{v(T) - k + 1}{v(T)}\right)^2 - \left(\frac{v(T) - k}{v(T)}\right)^2 \right\} dt$$

$$= \sum_{k=1}^{v(T)} \int_{0}^{t_k} \mathrm{E}\left(\sigma(t,V)^2\right) \frac{2(v(T) - k) + 1}{v(T)^2} dt.$$

Therefore, (5.10) becomes

$$\min \sum_{k=1}^{v(T)} \int_{0}^{t_k} \left\{ -\frac{2\mathrm{E}\left(\sigma(t,V)^2 X(t)\right)}{v(T)\mathrm{E}\left(\sigma(t,V)^2\right)} + \frac{2(v(T) - k) + 1}{v(T)^2} \right\} \mathrm{E}\left(\sigma(t,V)^2\right) dt.$$

5.6.3 Proof of Proposition 5.3.1

Since $X(t)$ and $x(t)$ are uncorrelated, (5.5) becomes

$$\min \sum_{k=1}^{v} \int_0^{t_k} \left\{ -\frac{2\mathrm{E}\left(X(t)\right)}{v(T)} + \frac{2(v(T)-k)+1}{v(T)^2} \right\} \mathrm{E}\left(\sigma(t)^2\right) dt.$$

If we represent the integrand as $f_k(t)$, we can prove that this function takes a negative value in $[0, t_{k-1}]$ and turns from negative to positive just once in $[t_{k-1}, T]$, and therefore, we can choose $\{t_k^*\}$ at such point. This is because the function is monotone increasing with respect to t, and the sign of the integrand can be checked by

$$f_1(0) = -\frac{1}{v(T)^2} < 0,$$
$$f_1(T) = \frac{2v(T)-1}{v(T)^2} > 0$$

when $k = 1$. Also, assuming that the relationship holds for $k-1$, it can be proved for k also since

$$f_k(t_{k-1}^*) = f_{k-1}(t_{k-1}^*) - \frac{1}{v(T)^2} < f_{k-1}(t_{k-1}^*) \leq 0,$$
$$f_k(T) = \frac{2(v(T)-k)+1}{v(T)} > 0.$$

So, the relationship holds for all k. Therefore, $\{t_k^*\}$ is determined to satisfy

$$-\frac{2\mathrm{E}\left(X(t_k^*)\right)}{v(T)} + \left\{\frac{v(T)-k+1}{v(T)}\right\}^2 - \left\{\frac{v(T)-k}{v(T)}\right\}^2 \leq 0,$$

and

$$\lim_{t \downarrow t_k^*} -\frac{2\mathrm{E}\left(X(t_k)\right)}{v(T)} + \left\{\frac{v(T)-k+1}{v(T)}\right\}^2 - \left\{\frac{v(T)-k}{v(T)}\right\}^2 \geq 0.$$

Solving this equation,

$$\mathrm{E}\left(X(t_k^*)\right) \geq \frac{2(v(T)-k)+1}{2v(T)} \geq \lim_{t \downarrow t_k^*} \mathrm{E}\left(X(t)\right).$$

We then first assume that $E[X(t)]$ is continuous at t_k^*. In this case,

$$\mathrm{E}\left(X(t_k^*)\right) = \frac{2(v(T)-k)+1}{2v(T)}.$$

Also, from (5.4),

$$x(t_k^*) = 1 - \frac{k}{v(T)} = \frac{2v(T)\mathrm{E}\left(X(t_k^*)\right)+1}{2v(T)}.$$

According to Figure 5.1, $x^*(t)$ takes a constant value over the t range of

$$\mathrm{E}\left(X(t_k^*)\right) - \frac{1}{v(T)} < \mathrm{E}\left(X(t)\right) \leq \mathrm{E}\left(X(t_k^*)\right)$$

which is equivalent to

$$\mathrm{E}\left(X(t)\right) \leq \mathrm{E}\left(X(t_k^*)\right) < \mathrm{E}\left(X(t)\right) + \frac{1}{v(T)}.$$

Therefore,

$$\frac{2v(T)\mathrm{E}\left(X(t)\right)-1}{2v(T)} \leq x^*(t) < \frac{2v(T)\mathrm{E}\left(X(t)\right)+1}{2v(T)}.$$

Or equivalently,

$$x^*(t) = \frac{\lfloor (2v(T)\mathrm{E}\left(X(t)\right)+1)/2 \rfloor}{v(T)}.$$

Next, if $\mathrm{E}\left(X(t)\right)$ is not continuous at t_k^*, the same argument can be applied. Specifically,

$$x^*(t_k^*) = \frac{\lfloor (2v(T)\mathrm{E}\left(X(t_k^*)\right)+1)/2 \rfloor}{v(T)}$$

from left continuity, and also,

$$\lim_{t\downarrow t_k^*} x^*(t) = \lim_{t\downarrow t_k^*} \frac{\lfloor (2v(T)\mathrm{E}\left(X(t)\right)+1)/2 \rfloor}{v(T)}.$$

Therefore, we can conclude that for all t,

$$x^*(t) = \frac{\lfloor (2v(T)\mathrm{E}\left(X(t)\right)+1)/2 \rfloor}{v(T)}.$$

5.6.4   Proof of Proposition 5.4.1

$$\min \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij} \int_0^T \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j(X_i - x_i)(X_j - x_j)\right) dt$$

$$= \min \left[\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij} \int_0^T \{-\mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j X_i\right)x_j - \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j X_j\right)x_i + \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j\right)x_i x_j\}dt \right.$$

$$+ \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij} \int_0^T \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j X_i X_j\right) dt \Bigg]$$

$$= \min \left[\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij} \sum_{k=1}^{v_i(T)} \int_0^{t_{ik}} \left\{ -\frac{2\mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j X_j\right)}{v_i \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j\right)} \right\} \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j\right) dt + \int_0^T \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j\right) x_i x_j dt \right.$$

$$+ \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij} \int_0^T \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j X_i X_j\right) dt. \Bigg]$$

Because each series of $\{t_{ik}\}$ is double counted in the series of $\{t_k^{ij}\}$,

$$\min \left[\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij} \left\{ \sum_{k=1}^{v_i(T)} \int_0^{t_{ik}} \left(-\frac{2\mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j X_j\right)}{v_i E[\rho_{ij}\sigma_i\sigma_j]} \right) \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j\right) dt \right. \right.$$

$$+ \sum_{k=1}^{v_i(T)+v_j(T)} \int_0^{t_k^{ij}} \frac{x_i(t_{k-1}^{ij})x_j(t_{k-1}^{ij}) - x_i(t_k^{ij})x_j(t_k^{ij})}{2} \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j\right) dt \Bigg\} \Bigg]$$

$$+ \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij} \int_0^T \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j X_i X_j\right) dt$$

$$
\begin{aligned}
= \quad & \min \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} \sum_{k=1}^{v_i(T)+v_j(T)} \int_0^{t_k^{ij}} \left\{ -\frac{\mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j X_j\right)}{v_i \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j\right)} + \frac{x_i(t_{k-1}^{ij})x_j(t_{k-1}^{ij}) - x_i(t_k^{ij})x_j(t_k^{ij})}{2} \right\} \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j\right) dt \\
& + \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} \int_0^T \mathrm{E}\left(\rho_{ij}\sigma_i\sigma_j X_i X_j\right) dt.
\end{aligned}
$$

Chapter 6

Dynamic Optimal Slice of a VWAP Trade

Abstract   This chapter analyzes an optimal execution strategy of a VWAP trade by dynamic control and derives an approximating solution. Non- negative constraint plays an important role in a dynamic strategy because the market trading volume ratio may decrease after big news arrives. If sell order is not allowed, the optimal execution delays in order to avoid excessive execution. Also, if the market trading volume surges, the trader should hold his execution rather than follow the market trading volume. We confirm execution error reduction by actual trading data.

6.1   Introduction

VWAP stands for volume-weighted average price during a certain trading period, and a VWAP trade refers a trade that uses VWAP as a benchmark. This chapter analyzes the dynamic optimal execution strategy of a VWAP trade. It extends the work of Chapter 5 of this study, which analyzes VWAP trade statically.

As VWAP has become an industry standard, related issues acquire popularity. From an empirical side, numerous studies have been done on the relationship between price volatility and market trading volume since Clark's (1973) seminal paper, including Epps and Epps (1976), Tauchen and Pitts (1983). Karpoff (1987) summarizes these results, and Andersen (1996) proposes a modified model. These studies, in general, support the existence of a positive correlation between price volatility and market trading volume and the autocorrelation of themselves because surprising news boosts both price volatility and market trading volume. Besides, Jain and Joh (1988), Admati and Pfleiderer (1988), Foster and Viswanathan (1990), and Chan et al. (1995) study intraday data and find a "U-shaped effect," i.e., within a single day, price volatility and market trading volume tend to be higher at the opening and closing than in the middle of a trading session.

Based on these findings, Chapter 5 of this study analyzes optimal execution of a VWAP trade. In executing market trades, a VWAP trader wishes to get the average execution price as close to the market VWAP as possible to avoid price movement risk. To do so, a VWAP trader slices a whole trade into small executions and tries to spread the executions in a well-balanced manner over the trading period, which makes VWAP trades rather labor intensive. Therefore, it often makes sense to build a standard strategy, and execute trades automatically according to it. Although Chapter 5 of this study obtains several important implications caused by the correlation between trading volume and price volatility, the strategy is optimized statically due to statistical, computational, and analytical difficulty.

However, we might be able to predict the behavior of stocks with frequent trades to some extent. Further, because economic impact tends to be large for such stocks, dynamic optimization may reduce considerable execution error further in such cases. Also, it is computationally feasible to apply dynamic control for few selected large orders. Besides, if sell order is not allowed, it is expected that the dynamic optimal execution delays in order to avoid excessive execution. Therefore, this chapter employs dynamic programming and allows traders to change their strategies in the middle of the day after the true market trading volume is observed.

This chapter is organized as follows. Section 6.2 formulizes the optimal strategy, and Section 6.3 derives the solution. Finally, Section 6.4 concludes the analysis.

6.2 Formulization of the Optimal Slicing

Imagine a discrete time frame $t = 0, 1, \cdots, T$ and a case in which a trader buys a certain number of shares and tries to get the average purchase price as close as possible to the market VWAP during the trading period from time 1 to time T. Let P_t denote the market price at t $(t = 0, \cdots, T)$, and define

$$\Delta P_t \stackrel{\text{def}}{=} P_t - P_{t-1}, \qquad t = 1, 2, \cdots, T.$$

Let Δv_t denote the trader's execution volume at time t, and define

$$v_t \stackrel{\text{def}}{=} \sum_{s=1}^{t} \Delta v_s, \qquad t = 1, 2, \cdots, T; \qquad v_0 = 0.$$

Also, let ΔV_t denote the market trading volume at time t $(t = 1, \cdots, T)$ excluding the trader's, and define

$$V_t \stackrel{\text{def}}{=} \sum_{s=1}^{t} \Delta V_s, \qquad t = 1, 2, \cdots, T; \qquad V_0 = 0, \quad a.e.$$

Define the filtration generated by $\{P_t; \ t = 0, 1, \cdots, T\}$ and $\{V_t; \ t = 0, 1, \cdots, T\}$ as

$$\mathcal{F}_t \stackrel{\text{def}}{=} \sigma\{(P_s, V_s); \ s = 0, 1, \cdots, t\}, \qquad t = 0, 1, \cdots, T.$$

Under this framework, VWAP trades are executed with the following constraints.

(Constraint 1) At time 0, the trader knows the exogenous whole trade size v_T, which has to be completed by time $t = T$.

(Constraint 2) The trader can submit only buy orders but not sell orders, which we call non-negative constraint henceforth.

$$0 = v_0 \le v_1 \le v_2 \le \cdots \le v_{T-1} \le v_T.$$

(Constraint 3) When the trader determines the controllable variable, the execution volume Δv_t at time t, only information up to time $t-1$ is available, i.e., $\{v_t\}$ should be $\{\mathcal{F}_t\}$-predictable.

A trader's VWAP is expressed as

$$vwap = \frac{\sum_{t=1}^{T} P_t \Delta v_t}{v_T}.$$

Also, the market VWAP excluding the traders' is

$$VWAP = \frac{\sum_{t=1}^{T} P_t \Delta V_t}{V_T}.$$

Therefore, the VWAP of the whole market is

$$TVWAP = \frac{\sum_{t=1}^{T} P_t (\Delta V_t + \Delta v_t)}{V_T + v_T}.$$

Assume that the trader's execution volume is sufficiently small compared to market tradng volume. Thus, we make the following approximation.

Approximation 6.2.1 $V_T/(V_T + v_T) \approx 1, \quad a.e.$

Under this approximation,

$$
\begin{aligned}
TVWAP - vwap &= \frac{1}{V_T + v_T} \sum_{t=1}^{T} P_t \Delta V_t - \frac{V_T}{(V_T + v_T) v_T} \sum_{t=1}^{T} P_t \Delta v_t \\
&\approx \frac{1}{V_T} \sum_{t=1}^{T} P_t \Delta V_t - \frac{1}{v_T} \sum_{t=1}^{T} P_t \Delta v_t.
\end{aligned}
$$

Define $y_t \overset{\text{def}}{=} v_t/v_T$ $(t = 0, 1, \cdots, T)$, and we consider $\{y_s\}$ as the controllable variable rather than $\{v_s\}$. The constraints can be rewritten as

(Constraint 1') $y_T = 1$.

(Constraint 2') $0 = y_0 \le y_1 \le y_2 \le \cdots \le y_{T-1} \le y_T = 1$.

(Constraint 3') $\{y_s\}$ is $\{\mathcal{F}_s\}$-predictable.

Note that $\{\frac{V_t}{V_T}\}$ is $\mathcal{F}_T$-measurable but not $\{\mathcal{F}_t\}$-adapted. A simple calculation shows that

$$
\begin{aligned}
\frac{1}{V_T} \sum_{t=1}^{T} P_t \Delta V_t - \frac{1}{v_T} \sum_{t=1}^{T} P_t \Delta v_t &= \sum_{t=1}^{T} \left(P_0 + \sum_{s=1}^{t} \Delta P_s \right) \left(\frac{\Delta V_t}{V_T} - \frac{\Delta v_t}{v_T} \right) \\
&= \sum_{t=1}^{T} \sum_{s=1}^{t} \Delta P_s \left(\frac{\Delta V_t}{V_T} - \frac{\Delta v_t}{v_T} \right) \\
&= \sum_{s=1}^{T} \Delta P_s \sum_{t=s}^{T} \left(\frac{\Delta V_t}{V_T} - \frac{\Delta v_t}{v_T} \right) \\
&= \sum_{s=1}^{T} \Delta P_s (y_{s-1} - \frac{V_{s-1}}{V_T}) \\
&= \sum_{s=1}^{T-1} \Delta P_{s+1} (y_s - \frac{V_s}{V_T}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathrm{E} \left((TVWAP - vwap)^2 \right) &\approx \mathrm{E} \left(\left(\sum_{s=1}^{T-1} \Delta P_{s+1}(y_s - \frac{V_s}{V_T}) \right)^2 \right) \\
&= \sum_{s=1}^{T-1} \mathrm{E} \left(\Delta P_{s+1}^2 (y_s - \frac{V_s}{V_T})^2 \right) \\
&\quad + 2 \sum_{s=2}^{T-1} \sum_{t=1}^{s-1} \mathrm{E} \left(\Delta P_{s+1} \Delta P_{t+1}(y_s - \frac{V_s}{V_T})(y_t - \frac{V_t}{V_T}) \right).
\end{aligned}
\tag{6.1}
$$

Based on empirical analyses such as Jain and Joh (1988), we make the following assumptions.

Assumption 6.2.1

1. *$\Delta P_s = \sigma_{P_s} \epsilon_{P_s}$ where ϵ_{P_s} is a sequence of independent random variables.*

2. *$\mathrm{E} \left(\epsilon_{P_s} \right) = 0, \mathrm{E} \left(\epsilon_{P_s}^2 \right) = 1 \ (s = 1, \cdots, T).$*

3. *$\{\epsilon_{P_s}\}$ is independent of $\{\frac{V_s}{V_T}\}$ and $\{\sigma_{P_s}\}$.*

Regarding the first term of (6.1), since y_s is $\mathcal{F}_{s-1}$-measurabel and independent of $\epsilon_{P_{s+1}}$,

$$
\mathrm{E} \left(\Delta P_{s+1}^2 (y_s - \frac{V_s}{V_T})^2 \right) = \mathrm{E} \left(\sigma_{P_{s+1}}^2 (y_s - \frac{V_s}{V_T})^2 \right).
$$

Regarding the second term of (6.1), for $s > t$,

$$
\mathrm{E} \left(\Delta P_{s+1} \Delta P_{t+1}(y_s - \frac{V_s}{V_T})(y_t - \frac{V_t}{V_T}) \right) = \mathrm{E} \left(\epsilon_{P_{s+1}} \right) \mathrm{E} \left(\epsilon_{P_{t+1}} \sigma_{P_{s+1}} \sigma_{P_{t+1}} (y_s - \frac{V_s}{V_T})(y_t - \frac{V_t}{V_T}) \right) = 0.
$$

Therefore, the objective function to minimize is given by

$$
\mathrm{E} \left((TVWAP - vwap)^2 \right) \approx \sum_{s=1}^{T-1} \mathrm{E} \left(\sigma_{P_{s+1}}^2 (y_s - \frac{V_s}{V_T})^2 \right).
\tag{6.2}
$$

Note that without the non-decreasing constraint for $\{y_s\}$, $y_s = \mathrm{E}\left(\left.\dfrac{V_s}{V_T}\right| \mathcal{F}_{s-1}\right)$ is the optimal execution strategy. However, in a dynamic strategy, the non-decreasing constraint plays an important role in increasing the risk of the excessive execution risk because $\dfrac{V_s}{V_T}$ may decrease after big news arrives and the market trading volume surges while $\mathrm{E}\left(\dfrac{V_s}{V_T}\right)$ is generally monotone increasing in a static strategy.

We decompose the sequence $\{\dfrac{V_s}{V_T}\}$, which is not observed accurately, into a observable part and an error part as

$$\frac{V_s}{V_T} = Y_s + \epsilon_{V_s}, \qquad Y_s \stackrel{\mathrm{def}}{=} \mathrm{E}\left(\left.\frac{V_s}{V_T}\right| \mathcal{F}_s\right), \qquad s = 1, 2, \cdots, T-1; \qquad\qquad \epsilon_{V_T} = 0, \quad a.e,$$

where we make some regular asumptions for the error part.

Assumption 6.2.2

1. $\{\epsilon_{V_s}\}$ *is a sequence of independent random variables.*

2. $\mathrm{E}\left(\epsilon_{V_s}\right) = 0$ $(s = 1, \cdots, T-1)$.

3. $\{\epsilon_{V_s}\}$ *and* $\{Y_s\}$ *are independent.*

Since $\{y_s\}$ is $\{\mathcal{F}_s\}$-predictable, $\{y_s\}$ is independent of $\{\epsilon_{V_s}\}$. Therefore,

$$
\begin{aligned}
\mathrm{E}\left(\sigma^2_{P_{s+1}}(y_s - \frac{V_s}{V_T})^2\right) &= \mathrm{E}\left(\sigma^2_{P_{s+1}}\{(y_s - Y_s) - \epsilon_{V_s}\}^2\right) \\
&= \mathrm{E}\left(\sigma^2_{P_{s+1}}(y_s - Y_s)^2\right) - 2\mathrm{E}\left(\sigma^2_{P_{s+1}}(y_s - Y_s)\right)\mathrm{E}\left(\epsilon_{V_s}\right) + \mathrm{E}\left(\sigma^2_{P_{s+1}}\epsilon^2_{V_s}\right) \\
&= \mathrm{E}\left(\sigma^2_{P_{s+1}}(y_s - Y_s)^2\right) + \mathrm{E}\left(\sigma^2_{P_{s+1}}\epsilon^2_{V_s}\right).
\end{aligned}
\tag{6.3}
$$

Since the second term of (6.3), $\mathrm{E}\left(\sigma^2_{P_{s+1}}\epsilon^2_{V_s}\right)$, is not controllable, the objective function to minimize in (6.2) is given by

$$\sum_{s=1}^{T-1} \mathrm{E}\left(\sigma^2_{P_{s+1}}(y_s - Y_s)^2\right).$$

6.3   Derivation of the Optimal Strategy

6.3.1   Uncorrelated Case

In order to formulize our optimization problem as a linear Markov decision process, we make the following assumption.

Assumption 6.3.1 $\{Y_s\}$ *is a Markov chain.*

The state is represented as the pair of the trader's execution ratio and the observed market trading volume ratio (y_s, Y_s). If we write the state at time s as $(y_s, Y_s) = (u, w)$, the execution proceeds as follows:

1. According to the value of Y_s, y_{s+1} is determined, satisfying $y_s \leq y_{s+1} \leq 1$.

2. Under $Y_s = w$, Y_{s+1} is given by the Markov chain. Also, P_{s+1} is given.

3. y_{s+1} and Y_{s+1} are executed at price P_{s+1}, and the state moves to (y_{s+1}, Y_{s+1}).

If we define

$$C_t^*(u, w) \stackrel{\text{def}}{=} \min_{u \leq y_t \leq \cdots \leq y_{T-1} \leq 1} \sum_{s=t}^{T-1} \mathrm{E}\left(\sigma_{P_{s+1}}^2 (y_s - Y_s)^2 \,\Big|\, Y_{t-1} = w \right), \qquad t = 1, \cdots, T-1$$

where the superscript $*$ denotes the optimal solution henceforth. Bellman equation is given by

$$C_t^*(u, w) = \min_{u \leq y_t \leq 1} \left\{ \mathrm{E}\left(\sigma_{P_{t+1}}^2 (y_t - Y_t)^2 \,\Big|\, Y_{t-1} = w \right) + \mathrm{E}\left(C_{t+1}^*(y_t, Y_t) \,\big|\, Y_{t-1} = w \right) \right\}, \qquad (6.4)$$
$$t = 1, 2, \cdots, T-2.$$

Since this formula is still too complicated to generate implications, further assumptions are made which are considered reasonable based on standard theories and empirical analyses. Firstly, since $Y_0 = 0$, $Y_T = 1$, and Y_t moves randomly between time 0 to T, we assume $\{Y_t\}$ is a Brownian bridge. Secondly, when we fix w, $C_{T-1}^*(u, w)$ is a quadratic function of u over the minimum, or a constant otherwise, it is natural to approximate C_t quadratically. Therefore, we made the following assumptions.

Assumption 6.3.2

1. *The process of Y_t is represented as*

$$Y_t = \bar{\eta}_t(Y_{t-1}) + \sigma_{Y_t} \epsilon_{Y_t}; \qquad Y_0 = 0, \quad Y_t = 1$$

   where $\bar{\eta}_t(\cdot)$ is a linear function, and $\{\epsilon_{Y_t}\}$ is a sequence of normal random variables that is uncorrelated with $\{\epsilon_{P_t}\}$.

2. *The market trading volume can be predicted precisely enough that σ_{Y_t} is sufficiently small.*

3. *The value function in the Bellman equation can be written as*

$$C_{t+1}^*(u, w) = \begin{cases} g_t(w), & w \geq -\frac{b_t u + c_t}{a_t}, \\ (a_t w + b_t u + c_t)^2 + g_t(w), & w < -\frac{b_t u + c_t}{a_t} \end{cases}$$

   for some coefficients $\{a_t\}$, $\{b_t\}$, $\{c_t\}$, and a function $\{g_t(w)\}$, which is independent of u.

Further, let $\{\sigma_{P_t}\}$ be uncorrelated with $\{\epsilon_{Y_t}\}$ in this subsection while the correlated case is analyzed in Subsection 6.3.2. In this case, (6.4) is reduced to

$$
\begin{aligned}
C_t^*(u, w) &= \min_{u \leq y_t \leq 1} \left\{ \mathrm{E}\left(\sigma_{P_{t+1}}^2 \right) \mathrm{E}\left((y_t - Y_t)^2 \,\big|\, Y_{t-1} = w \right) + \int_{-\infty}^{\infty} g_t(\eta) f_t(\eta - \bar{\eta}_t(w)) d\eta \right. \\
&\qquad\qquad \left. + \int_{-\infty}^{-(b_t y_t + c_t)/a_t} (a_t \eta + b_t y_t + c_t)^2 f_t(\eta - \bar{\eta}_t(w)) d\eta \right\} \\
&= \min_{u \leq y_t \leq 1} \left\{ \mathrm{E}\left(\sigma_{P_{t+1}}^2 \right) (y_t - \mathrm{E}(Y_t))^2 + I_t + const \right\}
\end{aligned}
$$

where $f_t(x) \stackrel{\text{def}}{=} \dfrac{1}{\sqrt{2\pi}\sigma_{Y_t}} \exp(-\dfrac{x^2}{2\sigma_{Y_t}^2})$,

$$I_t \stackrel{\text{def}}{=} \int_{-\infty}^{-(b_t y_t + c_t)/a_t} (a_t \eta + b_t y_t + c_t)^2 f_t(\eta - \bar{\eta}_t(w)) d\eta$$

and

$$const \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} g_t(\eta) f_t(\eta - \bar{\eta}_t(w)) d\eta$$

which is constant in y_t. Define $\zeta \equiv \eta - \bar{\eta}_t(Y_{t-1})$, $z_t \equiv -(b_t y_t + c_t)/a_t - \bar{\eta}_t(Y_{t-1})$ and $F_t(x) = \int_{-\infty}^{x} f_t(y) dy$. Then,

$$I_t = \int_{-\infty}^{z_t} a_t^2 (\zeta - z_t)^2 f_t(\zeta) d\zeta = a_t^2 \int_{-\infty}^{z_t} (\zeta^2 - 2z_t\zeta + z_t^2) f_t(\zeta) d\zeta.$$

Using the following equations,

$$\int_{-\infty}^{z_t} \zeta f_t(\zeta) d\zeta = -\sigma_{Y_t}^2 f_t(z_t),$$

$$\int_{-\infty}^{z_t} \zeta^2 f_t(\zeta) d\zeta = \sigma_{Y_t}^2 F_t(z_t) - \sigma_{Y_t}^2 z_t f_t(z_t),$$

we obtain

$$I_t = a_t^2 \sigma_{Y_t}^2 \{(1 + z_t^2) F_t(z_t) + z_t f_t(z_t)\}.$$

Since σ_{Y_t} is small, the risk of the excessive execution is expected to be low, which makes z_t small. Consequently, I_t can be approximated as

$$
\begin{aligned}
I_t &\simeq I_t(0) + I_t'(0)z_t + \frac{I_t''(0)}{2} z_t^2 \\
&\simeq \frac{a_t^2 \sigma_{Y_t}^2}{2} + \sqrt{\frac{2}{\pi}} a_t^2 \sigma_{Y_t} z_t + \frac{a_t^2 \sigma_{Y_t}^2}{2} z_t^2.
\end{aligned}
$$

Therefore, the quadratic approximation of $C_t^*(u, w)$ is given by

$$
\begin{aligned}
C_t^*(u, w) &= \min_{u \leq y_t \leq 1} \left[\left\{ \frac{b_t^2 \sigma_{Y_t}^2}{2} + \mathrm{E}\left(\sigma_{P_{t+1}}^2\right) \right\} y_t^2 \right. \\
&\quad \left. + \left\{ \left(a_t b_t \sigma_{Y_t}^2 - 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)\right) \bar{\eta}(w) + b_t c_t \sigma_{Y_t}^2 - \sqrt{\frac{2}{\pi}} a_t b_t \sigma_{Y_t} \right\} y_t + const \right] \\
&= \begin{cases} g_{t-1}(w), & w \geq -\frac{b_{t-1}u + c_{t-1}}{a_{t-1}}, \\ (a_{t-1}w + b_{t-1}u + c_{t-1})^2 + g_{t-1}(w), & w < -\frac{b_{t-1}u + c_{t-1}}{a_{t-1}} \end{cases}
\end{aligned}
$$

where the minimum is attained by

$$y_t = \max \left\{ -\frac{\left(a_t b_t \sigma_{Y_t}^2 - 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)\right) \bar{\eta}(w) + b_t c_t \sigma_{Y_t}^2 - \sqrt{\frac{2}{\pi}} a_t b_t \sigma_{Y_t}}{b_t^2 \sigma_{Y_t}^2 + 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)}, \ u \right\}.$$

Therefore, the optimal $\{y_t^*\}$ is solved iteratively starting with $y_0^* = 0$.

Specifically, if $\bar{\eta}(u) = \alpha_t + \beta_t u$, then

$$y_t^* = \max(d_t Y_{t-1} + e_t, y_{t-1}^*), \qquad t = 1, 2, \cdots, T-1$$

where

$$
\begin{aligned}
y_0^* &= 0, \\
d_t &= -\frac{\left(a_t b_t \sigma_{Y_t}^2 - 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)\right)}{b_t^2 \sigma_{Y_t}^2 + 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)}\beta_t, \\
e_t &= -\frac{\left(a_t b_t \sigma_{Y_t}^2 - 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)\right)\alpha_t + b_t c_t \sigma_{Y_t}^2 - \sqrt{\frac{2}{\pi}}a_t b_t \sigma_{Y_t}}{b_t^2 \sigma_{Y_t}^2 + 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)}, \\
a_T &= b_T = c_T = 0, \\
a_{t-1} &= \frac{a_t b_t \sigma_{Y_t}^2 - 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)}{2b_{t-1}}\beta_t, \\
b_{t-1} &= \sqrt{\frac{b_t^2 \sigma_{Y_t}^2}{2} + \mathrm{E}\left(\sigma_{P_{t+1}}^2\right)}, \\
c_{t-1} &= \frac{b_t c_t \sigma_{Y_t}^2}{2b_{t-1}}.
\end{aligned}
$$

For example, if $\{Y_t\}$ is a Brownian bridge,

$$dY_t = dB_t + \frac{1 - Y_t}{1 - t}dt$$

and therefore,

$$\bar{\eta}(Y_{t-1}) - Y_{t-1} = \frac{1 - \bar{\eta}(Y_{t-1})}{1 - t}\Delta t$$

or, equivalently,

$$\bar{\eta}(Y_{t-1}) = \frac{1 - t}{1 - t + \Delta t}Y_{t-1} + \frac{\Delta t}{1 - t + \Delta t}.$$

Substituting into the general solution, we obtain

$$y_t^* = \max(d_t Y_{t-1} + e_t, y_{t-1}^*), \qquad t = 1, 2, \cdots, T-1$$

where

$$
\begin{aligned}
y_0^* &= 0, \\
d_t &= -\frac{\left(a_t b_t \sigma_{Y_t}^2 - 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)\right)}{b_t^2 \sigma_{Y_t}^2 + 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)}\frac{1 - t}{1 - t + \Delta t}, \\
e_t &= -\frac{\left(a_t b_t \sigma_{Y_t}^2 - 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)\right)\frac{\Delta t}{1 - t + \Delta t} + b_t c_t \sigma_{Y_t}^2 - \sqrt{\frac{2}{\pi}}a_t b_t \sigma_{Y_t}}{b_t^2 \sigma_{Y_t}^2 + 2\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)}.
\end{aligned}
$$

As we have mentioned, the non-decreasing constraint plays an important role in a dynamic strategy because Y_t may decrease after some big news arrives and the market trading volume surges while $\mathrm{E}(Y_t)$ is generally monotone increasing in a static strategy.

In order to avoid excessive execution, y_t^* is expected to be smaller than Y_t, the optimal strategy without non-decreasing constraint. Also, once news arrives, and Y_t happens to decrease, y_t^* stays constant until Y_t gets back large enough. Therefore, if the market trading volume surges, the trader should hold his execution rather than follow the market trading volume.

Figure 6.1 shows Y_t and y_t^*, and Figure 6.2 shows g_t and h_t for $\sigma_{Y_t} = 1$ and all the ϵ_{Y_t} are zero. We can see that y_t^* is smaller than Y_t due to the risk that Y_t decreases, which delays trade execution. This is more significant for larger σ_{Y_t}.
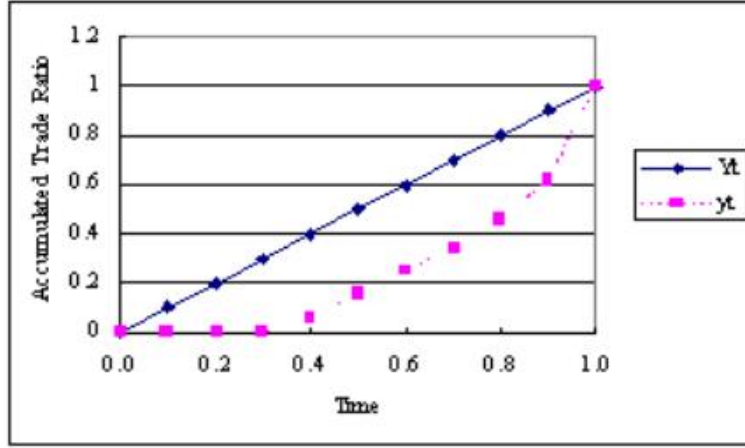

Figure 6.1: **Numerical example of Y_t and y_t^* for $\sigma_{Y_t} = 1$ and where all the ϵ_{Y_t} are zero.** We can see that y_t^* is smaller than Y_t due to the risk that Y_t decreases, which delays trade execution. This is more significant for larger σ_{Y_t}.

6.3.2 Correlated Case

If we take correlation between Y_t and σ_{P_t} into account, the following equation holds.

$$
\begin{aligned}
\sum_{t=0}^{T} \mathrm{E}\left((Y_t - y_t)^2 \sigma_{P_{t+1}}^2\right) &= \sum_{t=0}^{T} \mathrm{E}\left(\left\{\frac{\mathrm{E}\left(Y_t \sigma_{P_{t+1}}^2\right)}{\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)} - y_t\right\}^2\right) \mathrm{E}\left(\sigma_{P_{t+1}}^2\right) \\
&= \sum_{t=0}^{T} \mathrm{E}\left(\left\{\mathrm{E}(Y_t) + \frac{\mathrm{Cor}(Y_t, \sigma_{P_{t+1}}^2)\sigma_{Y_t}\mathrm{Std}(\sigma_{P_{t+1}}^2)}{\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)} - y_t\right\}^2\right) \mathrm{E}\left(\sigma_{P_{t+1}}^2\right)
\end{aligned}
$$

where Cor and Std stands for the correlation and the standard deviation, respectively. Let the superscript $**$ denote the optimal solution of the correlated case. Then, if $\mathrm{Cor}(Y_t, \sigma_{P_{t+1}}^2)$ and $\mathrm{Std}(\sigma_{P_{t+1}}^2)$ are
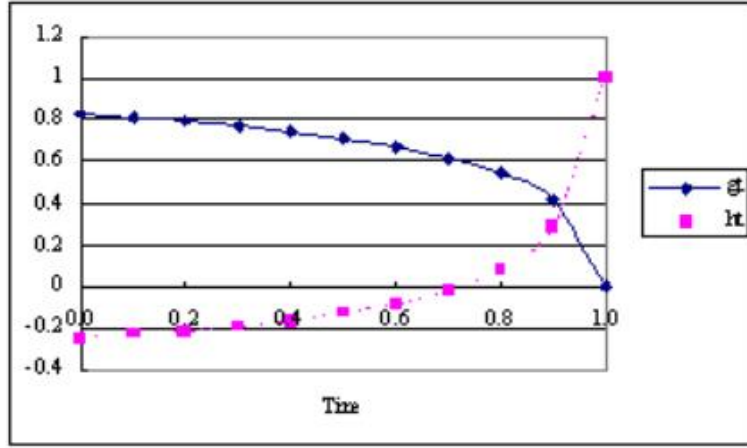
Figure 6.2: g_t **and** h_t **for** $\sigma_{Y_t} = 1$. g_t and h_t are deterministic functions that depends only on $\sigma_{Y_t} = 1$.

deterministic, we can solve y_t^{**} just as previous argument as

$$y_t^{**} = y_t^* + \frac{\mathrm{Cor}(Y_t, \sigma_{P_{t+1}}^2)\sigma_{Y_t}\mathrm{Std}(\sigma_{P_{t+1}}^2)}{\mathrm{E}\left(\sigma_{P_{t+1}}^2\right)}.$$

6.3.3   Back Tests

Next, we confirm our theoretical results empirically, using actual intraday data from the Tokyo Stock Exchange provided by Nikkei Quick Information Co. Ltd. All empirical analyses in this chapter are based on trading data from October 2000 to March 2001 on the 10 stocks with the largest market capitalization. Data from December 29, 2000 and January 4, 2001 are excluded since there was no afternoon session on these days. We chose these large stocks in order to retain statistical accuracy since their ticks are frequent and stable. We believe that results can be applied to other stocks since the relationship between price volatility and market trading volume is ubiquitous for all stocks as Chan et al. (1995) pointed out. We assume that accumulated market trading volume follows the linear difference equation as

$$\bar{\eta}(u) = \alpha_t + \beta_t u$$

and the coefficients α and β are obtained through ordinary least square regressions where realized $\frac{V_t}{V_T}$ is used as the estimate of Y_t. Executions are allocated to periods according to each strategy, and the trade is assumed to be executed at the VWAP of the corresponding period. Static strategy is performed in the exactly same manner as in Chapter 5.

   Table 6.1 shows VWAP execution errors for each stock. We can see that dynamic optimization effectively reduces VWAP execution errors. Also, we can further improve our results by taking correlation between accumulated market trading volume ratio and price volatility. In practice, we suffer further price

changes within the interval and discritization errors due to the minimum trading unit, both of which are considered to be proportional to the price volatility as in the static optimization in Chapter 5.

Also, Figure 6.3 shows the average delay of the optimal executions, $Y_t - y_t^*$ for the uncorrelated case and $Y_t - y_t^{**}$ for the correlated case. We can see that the risk of Y_t decrease and the correlation between price volatility and trading volume indeed delays the optimal trade execution.

	Static	Dynamic	Dynamic (Correlated)
Takeda	0.00102	0.00071	0.00071
Matsushita	0.00095	0.00039	0.00039
Sony	0.00093	0.00044	0.00044
Toyota	0.00100	0.00059	0.00058
Honda	0.00121	0.00083	0.00083
Seven-Eleven Japan	0.00219	0.00167	0.00167
Mizuho	0.00143	0.00092	0.00092
Tokyo-Mitsubishi	0.00133	0.00077	0.00076
Nomura	0.00142	0.00065	0.00064
NTT	0.00176	0.00058	0.00059
NTT Docomo	0.00128	0.00050	0.00050
Tokyo Electric Power	0.00065	0.00042	0.00041
Average	0.00127	0.00071	0.00070

Table 6.1: **Delay of the optimal execution.** This graph shows the delay of the optimal execution from the average trading volume (y-axis), which coincides with $Y_t - y_t^*$ in the Uncorrelated Case and $Y_t - y_t^{**}$ in the Correlated Case in each period (x-axis). We can see that the risk of Y_t decrease and the correlation between price volatility and trading volume indeed delays the optimal trade execution.

6.4 Closing Remarks

In this chapter we analyzes an optimal execution strategy of a VWAP trade by dynamic control and derives approximating solution. Non-negative constraint plays an important role in a dynamic strategy because the market trading volume ratio may decrease after big news arrives. If sell order is not allowed, the optimal execution delays in order to avoid excessive execution. Also, if the market trading volume surges, the trader should hold his execution rather than follow the market trading volume. We confirm execution error reduction by actual trading data.

For simplicity, this analysis studies only one asset execution but not multiple asset execution with diversification effect. Also, estimating the market trading volume may not be an easy task even with qualitative judgment. These issues are left for further analysis.
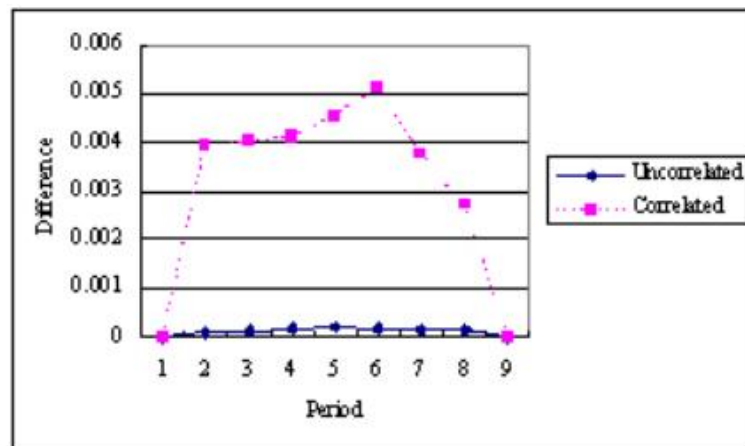
Figure 6.3: **This graph shows VWAP execution errors for each stock.**   We can see that dynamic optimization effectively reduces VWAP execution errors. Also, we can further improve our results by taking correlation between accumulated market trading volume ratio and price volatility.

Chapter 7

Closing Remarks

7.1 Introduction

As the investment business becomes increasingly competitive, it has turned out that the trade execution also significantly affects investment performance. As a result, many practitioners start recognizing trade execution as an independent task. For example, responsibilities of fund managers and traders are clearly separated in the most institutional investors, and principal trades in which security brokers take responsibility of trade execution has become popular for complicated trading needs while execution risk remains investors' responsibility in conventional agency trades. In contrast, few theoretical researches can be found about trade execution. Therefore, this study analyzes the optimal trade execution strategies that minimize trading costs for uninformed traders (traders without specific information) whose trading needs are given exogenously.

In practice, there are mainly two approaches for saving trading costs:

1. To balance market impact and volatility risk by referring a fixed price in portfolio or block trades.

2. To mitigate impact of trades by referring volume weighted average price (VWAP henceforth) in VWAP trades.

These approaches are chosen according to trade size, traders' objectives, circumstances, and so forth. In our analysis, we divide Approach 1 into long-term execution scheduling problem and short-term order placement problem since it is difficult to analyze Approach 1 as a whole. Also, in Approach 2, static optimization works well for numerous small orders while dynamic optimization is more suitable for few large orders. Therefore, the four cases below are analyzed.

1. Fixed–Price Trade

   - Execution Schedule

   - Order Placement

2. VWAP Trade

- Static Optimization

- Dynamic Optimization

Each issue is explained in the following sections.

7.2 Optimal Slice of a Block Trade

Chapter 3 of this study has studied execution scheduling of liquidation of a block trade and formulated it as a static optimization problem. Based on empirical studies on market microstructure, we have distinguished between temporary market impact and permanent market impact, and assumed that each component to be a linear function of trading volume within moderate volume and time. Also, we have defined the measure of opportunity cost as standard deviation of price movement, and minimized transaction cost, sum of execution cost and opportunity cost. Under this framework, we have obtained explicit solutions both for the single-asset case and a multiple-asset portfolio model in the form of pure exponentials. Further, we have derived an approximation formula when there is a limit to the trading volume executable instantaneously.

As a result, we have shown that the optimal execution strategy is independent of permanent impact, that optimal execution duration is an increasing function of whole trading size, and that the optimal average cost is an increasing function of whole trading size, price volatility, and temporary impact, and a decreasing function of the limit to instantaneous execution volume, which are all consistent with trading practice and empirical studies. Specifically, optimal average cost increases as whole trading size to the power of 1/3 for small orders and 1/2 for large orders. Further, it is shown that the transaction cost can be lowered by making order execution late when price movement risk of a portfolio is well diversified.

Our results are immediately applicable to existing risk management framework and provides explicit solution for minimum L-VaR. Also, although Chapter 3 of this study uses an example of liquidation of a block of common stock, results can be extended over wider problems on liquidation of any large block of assets.

While Chapter 3 of this study has focused on static optimization, decision change in accordance with new information might be handled by dynamic programming. Also, this model can reflect more reality by making volatility and market impact time dependent. Further, it is an interesting issue how to relate order slicing strategy and choice between market order and limit order, and how to control information effect of orders. These issues are left for further analysis.

7.3   Selection of Market and Limit Order

Chapter 4 of this study analyses the optimal selection of market and limit orders in a series of single price batch auctions when the expectation of the limit order book is given. We derive the analytical solution of a single limit order model for risk neutral traders, and then extend the result for risk averse traders. Regarding the execution at each period, we find that the optimal limit order size is independent of the whole trade size so as to balance the non–linearity of execution volume and price. In contrast, the limit order price and the market order size are linear functions in order to balance trading costs across trading periods. Regarding the execution throughout the trading session, market orders replace limit orders as time passes, and necessity of completing execution grows. Further, if trading volume and price volatility are larger at the opening and closing of the trading session as in the practical markets, more limit orders are tried at the opening and closing.

Although most traders know the public order book as expectations as in our model, member security brokers are allowed to monitor the public order book without time lags. So, we evaluate the value of monitoring public order books and provides criteria for selecting stocks to monitor.

Although our results can be extended to multiple limit order models, mathematical analysis is left for further research. Besides, while Chapter 4 of this study analyzes batch auctions for simplicity, continuous auction is also of interest. Also, although Chapter 4 of this study assumes randomness comes only from the number of market orders, the number of limit orders are also stochastic in practice. Further, we can extend our research to risk averse traders who uses VWAP as a reference price. Besides, while Chapter 4 of this study performs partial equilibrium type analysis in which public orders are given exogenously, general equilibrium analysis which studies how each trader's strategy interferes is left for further research.

7.4   Optimal Slice of a VWAP Trade

Chapter 5 of this study derives the static optimal execution strategy of a VWAP trade that minimizes the expected squared execution error. This method is powerful because the optimal execution strategy is determined by an iteration of a single variable optimization, rather than by a multivariable optimization. Analytical solutions are derived in some cases. The following results are obtained through our analysis. For a single-stock trade, if price volatility is independent of market trading volume, the optimal execution strategy is determined only by the expected market trading volume distribution and is independent of expectations regarding the magnitude and time dependency of price volatility. If price volatility is positively correlated with market trading volume, optimal execution times turn out to lag behind the expected market trading volume distribution. This is because once price volatility and market trading volume surge, traders have to make considerable trades to track market trading volume while execution times do not matter when these parameters remain small throughout the day. In a basket trade, execution

error can be reduced by spreading out execution times according to the correlation of price movement. Further, we examine these theoretical results with actual trading data and simulations.

Chapter 5 of this study focuses on static optimization since observed variables such as price volatility and the market trading volume of small orders with low liquidity may contain statistical errors too large to be used in forecasting, and there are some concerns that dynamic optimization is inaccurate. However, we might be able to predict the behavior of stocks with frequent trades to some extent, and dynamic optimization may reduce execution error further in such cases. Dynamic optimization is analyzed in Chapter 6.

For simplicity, Chapter 5 of this study minimizes execution error caused by price movement. In the real world, however, the direct cost of market impact and the indirect cost of information leakage are certainly significant factors in a VWAP trade. This issue is left for further analysis.

7.5 Dynamic Optimal Slice of a VWAP Trade

Chapter 6 of this study analyzes an optimal execution strategy of a VWAP trade by dynamic control and derives approximating solution. Non–negative constraint plays an important role in a dynamic strategy because the market trading volume ratio may decrease after big news arrives. If sell order is not allowed, the optimal execution delays in order to avoid over execution. Also, if the market trading volume surges, the trader should hold his execution rather than follow the market trading volume. We confirm execution error reduction by actual trading data.

For simplicity, this analysis studies only one asset execution but not multiple asset execution with diversification effect. Also, estimating the market trading volume may not be an easy task even with qualitative judgment. These issues are left for further analysis.

7.6 Closing Remarks

In practice, all securities trading are carried out through some trade execution, regardless of recognition. At execution, one of four approaches above are chosen according to assets' characteristics and traders' objectives and circumstances. In order to make appropriate choice traders have to deeply understand the nature of the alternatives.

Although few analyses have been made about the optimal trade execution based on findings of market microstructure, this study derives the optimal execution strategy in each approach, analyzes its characteristics, and provides the guidance to the selection of these approaches. Therefore, extensive range of application can be expected for any security trading.

Bibliography

Abramowitz, M., and I.A. Stegun, (1972). *Handbook of Mathematical Functions*, Dover.

Admati, A., and P. Pfleiderer, (1988). A theory of intraday patterns: volume and price variability, *Review of Financial Studies* 1, Spring, 3–40.

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

Almgren, R., and N. Chriss, Optimal execution of portfolio transactions. *Journal of Risk* 3(2), 5–39.

Andersen, T.G., (1996). Return volatility and trading volume: an information flow interpretation of stochastic volatility, *Journal of Finance* 71, 169–204.

Barclay, M. J., W.G. Christie, J. H. Harris, E. Kandel, and P. H. Schultz, (1999). The effects of market reform on the trading costs and depths of Nasdaq stocks. *Journal of Finance* 54, 1–34.

BARRA. (1997). *Market Impact Model Handbook*.

Bellman, R. (1997). *Introduction to Matrix Analysis*, SIAM.

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

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

Black, F., and R. Litterman, (1991). Asset allocation: combining investor views with market equilibrium. *Journal of Fixed Income* 1(2), 7–18.

Chakravarty, S., and C. Holden, (1995). An integrated model of market and limit orders, *Journal of Financial Intermediation* 4, 213–241.

Chan, K.C., W.G. Christie, and P.H. Schultz, (1995). Market structure and the intraday pattern of bid-Ask Spreads for NASDAQ securities, *Journal of Business* 68, 35–60.

Clark, P.K., (1973). A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 135–155.

Collins, B., and F. Fabozzi, (1991). A methodology for measuring transaction costs. *Financial Analysts Journal* 47, 27–36.

Copeland, T. C., and D. Galai, (1983). Information effects on the bid-ask spread. *Journal of Financial Economics* 7, 229–263.

Epps, T.W., and M.L. Epps, (1976). The Stochastic dependence of security price changes and transaction volumes: implications for the mixture-of-distributions hypothesis, *Econometrica* 44, 305–321.

Foster, D., and S. Viswanathan, (1990). A theory of the interday variations in volume, variance, and trading costs in securities markets, *Review of Financial Studies* 3, 593–624.

Grinold, R., and R. Kahn, (1999). *Active Portfolio Management*. McGraw-Hill Professional Publishing.

Hausman, J., A. Lo, and C. MacKinlay, (1992). An orderd probit analysis of transaction stock prices, *Journal of Financial Economics* 31, 319–379.

Harris, L., and J. Hasbrouck, (1996). Market vs. limit orders: the SuperDOT evidence on order submission strategy, *Journal of Financial and Quantitative Analysis* 31, 213–231.

Holthausen, R., R. Leftwich, and D. Mayers, (1987). The effect of large block transactions on security prices: a cross-sectional analysis. *Journal of Financial Economics* 19, 237–268.

Hisata, Y., and Y. Yamai, (2000). An analysis on practical liquidity risk measurement. Institute for Monetary and Economic Studies Discussion Paper 2000-J-3.

Jain, P.C., and G. Joh, (1988). The dependence between hourly prices and trading volume. *Journal of Financial Quantitative Analysis* 23, 269–283.

Jorion, P., (2000). *Value at Risk*. McGraw-Hill.

Karpoff, J.M., (1987). The relation between price changes and trading volume: a survey, *Journal of Financial and Quantitative Analysis* 22, 109–126.

Kawahara, J., (1994). Evaluation of execution cost and best execution (in Japanese), *Security Analysts Journal* 12, 32–47.

Kawahara, J., and Y. Murase, (1993). Analysis on intraday stock price movement. (in Japanese) *Security Analysts Journal* 11, 10–21.

Kolmogorov, A.N., and S.V. Fomin, (1970). *Introductory Real Analysis*. Dover.

Konishi, H., (2002). Optimal slice of a VWAP trade, *Journal of Financial Markets* 5(2) 197–221.

Konishi, H., and N. Makimoto, (2001). Optimal slice of a block trade, *Journal of Risk* 3(4), 33–51.

Konno, H., and A. Wijayanayake, (1998). A mean-absolute deviation portfolio optimization model under transaction costs. Dept. of IE and Management, Tokyo Institute of Technology.

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

Lehman, B.N., and D. M. Modest, (1994). Trading and liquidity on the Tokyo Stock Exchange: a bird's eye view. *Journal of Finance* 49, 951–984.

Mannen, S., and J. Uno, (2000). Basket orders and basis costs. (in Japanese) *NQI Report* 3, 19–22.

Markovitz, H., (1959). *Portfolio Selection: Efficient Diversification of Investments.* John Wiley & Sons, Inc.

McInish, T. H., and R. A. Wood, (1991). An analysis of intraday patterns in bid/ask spreads for NYSE stocks. Working Paper, Memphis State University.

Parlour, C. A., (1998). Price dynamics in limit order markets. *Review of Financial Studies* 11, 789–816.

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

Protter, P. E., (1990). *Stochastic Integration and Differential Equations: A New Approach.* Springer Verlag.

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

Sharpe, W., (1970). *Portfolio Theory and Capital Market.* McGraw-Hill.

Subrahmanyam, A., (1991). Risk aversion, market liquidity, and price efficiency. *Review of Financial Studies* 4(3), 417–442.

Tauchen, G.E., and M. Pitts, (1983). The price variability-volume relationship on speculative markets, *Econometrica* 51, 485–505.

Uno, J., and M. Yamada, (1993). Empirical study on market impact. (in Japanese) *Security Analysts Journal* 11, 22–34.