



LUND UNIVERSITY
School of Economics and Management

HIGH-FREQUENCY TRADING
HOW MONEY FLOW AFFECTS STOCK RETURNS
MASTER THESIS

ERIC OLAISON
eric.olaison@gmail.com

SPRING SEMESTER, 2010

Supervisor: Professor Hans Byström

ABSTRACT

- Title:** High-Frequency Trading: How Money Flow Affects Stock Returns
- Seminar date:** 2010-06-01
- Course:** NEKM01, Master Thesis, Economics, Master level, 15 ECTS credits.
- Author:** Eric Olaison
- Supervisor:** Professor Hans Byström
- Key words:** High-frequency trading, money flow, tape reading, intra-day volume.
- Purpose:** The main purpose with this thesis is to study tick-data in order to see if intra-day volume can predict short-term market movements.
- Methodology:** This thesis a quantitative study, using a deductive method and with an exploratory research design.
- Theoretical framework:** In the theoretical framework, important theories for this thesis are presented, such as tape reading, high-frequency trading, and previous findings about money flow and measures of floor trading.
- Empirical foundations:** The data sample consists of approximately 2.2 million trades during January 1, 2005 to April 22, 2010. The data was retrieved from Six Telekurs database.
- Conclusions:** The high-frequency measures used in this thesis showed a significant relationship between the measures and the stock return at a 1% significance level. The accumulated money flow was highly positively correlated with the stock return until late 2008, and since then it became negatively correlated. Steadily increasing activity by high-frequency trading algorithms, as well as the tick-size changes that occurred during 2009, might be an explanation. The trading model, which used the three measures studied, had almost twice as high risk-adjusted return as the stock itself.

TABLE OF CONTENTS

ABSTRACT	2
ACKNOWLEDGMENTS	5
LIST OF FIGURES	6
LIST OF TABLES	7
ABBREVIATIONS	8
1. INTRODUCTION	9
1.1 BACKGROUND	9
1.2 BACKGROUND DISCUSSION	10
1.3 PURPOSE AND OBJECTIVES	11
1.4 FOCUS AND LIMITATIONS	11
1.5 OUTLINE OF THE THESIS	12
2 METHODOLOGY	13
2.1 CHOICE OF SCIENTIFIC APPROACH	13
2.2 RESEARCH DESIGN AND METHODS	14
2.3 WORKING PROCEDURE	15
2.3.1 LITERATURE STUDIES	15
2.3.2 THE DATA SAMPLE	15
2.3.3 ANALYSIS OF DATA.....	17
2.3.4 STATISTICAL METHODS	19
2.3.5 CREDIBILITY.....	19
2.3.5.1 Reliability.....	20
2.3.5.2 Validity.....	20
2.3.5.3 Objectivity.....	21
2.4 CRITICISM TO CHOSEN METHODS	21
3 THEORETICAL FRAMEWORK	23
3.1 TAPE READING	23
3.2 HIGH-FREQUENCY TRADING	24
3.2.1 ORDERS, TRADERS AND LIQUIDITY	25
3.2.2 ALGORITHMIC TRADING	26
3.2.2.1 Volume-Weighted Average Price (VWAP).....	27
3.2.2.2 Time-Weighted Average Price (TWAP).....	27
3.2.2.3 Implementation Shortfall	28
3.2.2.4 The New Generation of Execution Algorithms.....	28
3.3 PREVIOUS STUDIES	29
3.3.1 FINDINGS OF BENNETT AND SIAS (2001)	29
3.3.2 FINDINGS OF STEIDLMEYER AND HAWKINS (2003).....	30

4	<u>ANALYSIS</u>	<u>33</u>
4.1	SUMMARY OF MEASURES.....	33
4.2	SUMMARY STATISTICS	35
4.3	HYPOTHESIS TESTING.....	35
4.4	TESTING A TRADING MODEL WITH THE MEASURES STUDIED.....	44
5	<u>CONCLUSION.....</u>	<u>49</u>
5.1	DISCUSSION OF FINDINGS.....	49
5.2	FUTURE RESEARCH SUGGESTIONS.....	51
	<u>REFERENCES</u>	<u>53</u>
A.	PRINTED SOURCES.....	53
B.	ELECTRONIC SOURCES.....	55
	<u>APPENDICES</u>	<u>56</u>
	APPENDIX I.....	56
	APPENDIX II.....	57
	APPENDIX III.....	59
	APPENDIX IV	62
	APPENDIX V.....	63
	<u>INDEX.....</u>	<u>66</u>

ACKNOWLEDGMENTS

Firstly, I would like to thank my supervisor Hans Byström, Professor at the School of Economics and Management, department of Economics, Lund University, for his patience and valuable remarks when reviewing my thesis, and for the invaluable time he has put in to it. Secondly, I also would like to show my deep gratitude to Amanda Edström and Jan Olaison for reading, commenting, and helping me with the readability in this thesis.

LIST OF FIGURES

Figure 2.1 The Process of Deduction.	14
Figure 4.1 Regression Plot of the Normalized Net Volume.	38
Figure 4.2 Regression Plot of the Net Money Flow.	38
Figure 4.3 Regression Plot of the Normalized OFI.	39
Figure 4.4 Regression Plot of the VWAP Skew.	39
Figure 4.5 Regression Plot of Accumulated Money Flow.	43
Figure 4.6 Comparison between Accumulated Money Flow and Closing Price in ABB Ltd.	43
Figure 4.7 The Return of the Trading Model vs. the Long-only Portfolio.	45
Figure 5.1 Volume Distribution for ABB Ltd.	51
Figure A.1 Plot of the Frequency Distribution of the Daily Returns.	63
Figure A.2 Plot of the Frequency Distribution of the Normalized Net Volume.	63
Figure A.3 Plot of the Frequency Distribution of the Net Money Flow.	64
Figure A.4 Plot of the Frequency Distribution of the Normalized OFI.	64
Figure A.5 Plot of the Frequency Distribution of the VWAP Skew.	65

LIST OF TABLES

Table 1.1 High-Frequency Trading Strategies.....	9
Table 1.2 Outline of the Thesis.....	12
Table 2.1 The Difference between Qualitative and Quantitative Studies.	13
Table 2.2 Number of Trades Distributed over the Month.	16
Table 2.3 SSE Closed due to Holidays that Occurred during the Sample Period.....	16
Table 2.4 An Example of the Ticker Tape in ABB Ltd.	17
Table 2.5 Daily Trade Summary.....	18
Table 2.6 Continuation of Previous Table, Daily Trade Summary.....	18
Table 2.7 Continuation of Previous Table, Daily Trade Summary.....	19
Table 2.8 Type I and Type II Errors.....	20
Table 3.1 Limit Orders versus Market Orders.	25
Table 3.2 An Interpretation of Different OFI Numbers.	31
Table 3.3 Volume @ Time Definitions of Price Ranges.....	32
Table 4.1 Definition of Buy and Sell Volume, Respectively.	33
Table 4.2 Summary Statistics.....	35
Table 4.3 Regression Statistics for the Two Variable Models.....	36
Table 4.4 Regression Statistics for the Normalized Net Volume.	36
Table 4.5 Regression Statistics for the Net Money Flow.....	36
Table 4.6 Regression Statistics for the Normalized OFI.....	37
Table 4.7 Regression Statistics for the VWAP Skew.....	37
Table 4.8 Confidence Intervals for the Two-Variable Models.	37
Table 4.9 Confidence Intervals for the Four-Variable Model.	41
Table 4.10 Regression Statistics.....	41
Table 4.11 ANOVA Table in Terms of R-squared.	41
Table 4.12 ANOVA Table for the Estimated Multiple Regression Model.	42
Table 4.13 Regression Statistics for the Accumulated Money Flow.	42
Table 4.14 The Regression Coefficients and its Confidence Intervals.	42
Table 4.15 Definition of the Dummy Variable used in the Trading Model.	44
Table 4.16 The Swedish Repo Rate.	46
Table 4.17 Trading Model vs. Long-only Portfolio.	46
Table 4.18 Distribution of Forecast Hits and Misses.	47
Table 4.19 The Fifteen Most Extreme Returns for Hits and Misses.....	47
Table 4.20 The 30 Largest Gains and Losses.	48
Table 4.21 Number of Correct Predictions for each Measure.	48
Table A.0.1 Trades Excluded from the Data Sample due to Prices outside the Daily Trading Range.	56

ABBREVIATIONS

▪ A.M.F.	Accumulated Money Flow
▪ ANOVA	Analysis of Variance
▪ CBOT	Chicago Board of Trade
▪ CME	Chicago Mercantile Exchange
▪ D.F.	Degrees of Freedom
▪ DMA	Direct Market Access
▪ ECN	Electronic Communications Network
▪ ESS	Explained Sum of Squares
▪ LIBOR	London Interbank Offered Rate
▪ MSS	Mean Sum of Squares
▪ NASDAQ	National Association of Securities Dealers Automated Quotations
▪ N.M.F.	Net Money Flow
▪ N.V.	Net Volume
▪ Norm. N.V.	Normalized Net Volume
▪ Norm. OFI	Normalized OFI
▪ NYSE	New York Stock Exchange
▪ OFI	On Floor Information
▪ OLS	Ordinary Least Squares
▪ OTC	Over-the-Counter
▪ Reg ATS	Regulation of Alternative Trading Systems
▪ Repo	Repurchase Agreement
▪ RSS	Residual Sum of Squares
▪ S.D.	Standard Deviation
▪ S.E.	Standard Error
▪ SEC	Securities and Exchange Commission
▪ SEK	Swedish Krona (singular) / Swedish Kronor (plural)
▪ SS	Sum of Squares
▪ SSE	Stockholm Stock Exchange
▪ TSA	Trading Strategy Accuracy
▪ TSS	Total Sum of Squares
▪ TWAP	Time-Weighted Average Price
▪ V@T	Volume @ Time
▪ VWAP	Volume-Weighted Average Price

1. INTRODUCTION

In this first chapter I will strive to give a thorough background of the new era of high-frequency trading, and I hope I can shed some light on the important areas that is vital to this thesis. The background discussion provides a general discussion on the subject, and leads to a problem specification. Finally, I will present the foundation for the thesis' purpose, limitations, and outline.

1.1 Background

According to Fadiman and Klein (2004), Wall Street started to use computer power in the 1980s, with algorithms that a cheap Dell computer could toss off today in no time, but back then it was in the forefront. Code words like “rocket scientist”, “proprietary systems” and “black-box trading” caught the excitement on Wall Street during the 1980s. Brown (2010, p.8) defines black-box trading as a quantitative investment strategy in which decisions are entirely defined by mathematical formulas. The short holding period for these strategies has evolved into high-frequency trading.

High-frequency trading has become particularly popular during the last decade. Mainly, there are four classes of trading strategies for high-frequency trading: automated liquidity provision, market microstructure trading, event trading, and statistical arbitrage. The typical holding period differs between the strategies mentioned above, and for a more specific classification see Table 1.1 below. High-frequency strategies provide numerous benefits to society as well. Markets will become more efficient, and added liquidity will stabilize the market system. The demand for new solutions to relieve Internet communication bottlenecks, will encourage innovation in computer technology (Aldridge, 2010, pp.2-4).

Table 1.1 High-Frequency Trading Strategies.

STRATEGY	DESCRIPTION	HOLDING PERIOD
Automated Liquidity Provision	Market-making positions with optimal pricing and execution through quantitative algorithms.	Less than 1 minute
Market Microstructure Trading	Identifying order flow through reversed engineering of observed quotes.	Less than 10 minutes
Event Trading	Short-term trading on macro economic events.	Less than 1 hour
Statistical Arbitrage	Statistical arbitrage of deviations from equilibrium.	Less than 1 day

(Aldridge, 2010, p.4)

This thesis will focus on a statistical arbitrage strategy, but first I will explore the trade signals validity. The focus will lie on volume and money flow to see if it can lead to statistical arbitrage opportunities in the market. Link (2003, pp.79, 161-162) claims that volume is of great importance, and it is often used in trading to confirm price movements. Breakouts from a trading range on strong volume tend to break through with momentum, while low-volume breakouts show lack of conviction by market participants. This thesis will focus on volume and money flow with a trader's perspective.

1.2 Background Discussion

Market making has been a lucrative business for New York Stock Exchange (NYSE) specialists and National Association of Securities Dealers Automated Quotations (NASDAQ) market makers for decades. Hedge funds soon realized there was money to be made in providing liquidity, thanks to the advent of electronic trading. The turning point came in 1997, when Securities and Exchange Commission (SEC) introduced new order-handling rules in its Regulation of Alternative Trading Systems (Reg ATS). The purpose with these reforms was to improve the efficiency of the market place, and the Electronic Communications Networks (ECNs) were a byproduct of these rules. ECNs match buyers and sellers automatically, unlike NASDAQ and NYSE, which needs an intermediary to connect buyers and sellers (Brown, 2010, pp.2, 29-35).

Hedge funds eagerly awaited the development of electronic trading technology, because the ability to transact in the equity markets without an intermediary presented significant profit opportunities. Computerized traders have blossomed since the adoption of Reg ATS in 1997. These computerized traders have started to compete with traditional participants such as floor traders in the Chicago Mercantile Exchange (CME), NYSE specialists, or NASDAQ market makers, and often reaping their livelihood (Brown, 2010, pp.31-35).

To survive in the market place, one faces the inevitable in a dynamic market, i.e. change, and to survive one needs to adapt. This study will shed some light upon some of the black-box trading strategies, and if intra-day volume can predict price movements. Heires (2009, p.15) writes in an article about FTEN Inc., a provider of execution and pre-risk management systems for prime brokers and high-frequency traders, who claims that if one is not a high-frequency trader, one will run the risk of being extinct. As I said before, to be able to capitalize on the prevailing market conditions, one needs to adapt and this is my way of doing so.

Brown (2010, p.63) claims that the stock price movements are strongly influenced of money flow trends, and I will investigate if these trends create enough momentum, and if so, how one can exploit it.

1.3 Purpose and Objectives

The main purpose with this thesis is to study tick-data in order to see if intra-day volume can predict short-term market movements. Put in other words, this thesis will focus solely on money flow, and tape reading techniques to analyze the data as opposed to the traditional analysis of stock prices.

The objective is to dissect tick-data into sub-groups to see whether the aggressiveness of the buyers and sellers, the money flow, or the distribution of volume affect stock prices.

1.4 Focus and Limitations

This thesis will only study tick-data from the Swiss-Swedish company, ABB Ltd. I have been working full-time for 6 years as a market maker in that stock, and that is why I chose to focus on its tick-data. There are mainly three topics that I am going to study throughout this thesis, and they are:

- Does money flow affect stock prices, and is it a good predictor of future price movements?
- Is there a relationship between high-frequency trading strategies and the observed return in the specific stock?
- Is the aggressiveness of buyers versus sellers affecting the return, and is it a good predictor of future price movements?

Even though ABB Ltd is traded at three stock exchanges around the world (Six Swiss Exchange, Stockholm Stock Exchange (SSE), and NYSE), only tick-data from SSE will be studied. Since algorithmic trading is widespread on these exchanges, there only exists miniscule arbitrage opportunities which the algorithms exploit within a second. Hence, the prices observed on the ticker tape will be within the same range on every exchange (if the prices are converted to the same currency). If there is a sell order on one exchange that deviates to much, an arbitrage algorithm will buy that order and execute a sell order on another exchange to lock in the profit. My point is, if one exchange receives excessive buy or sell pressure, it will instantaneous transmit that pressure to the other exchanges as well.

Therefore should the result of my analysis be quite similar if I had studied the trades from the Six Swiss Exchange or NYSE instead.

1.5 Outline of the Thesis

To give a clear view of how this thesis will be conducted, Table 1.2 will present each chapter's content and objective.

Table 1.2 Outline of the Thesis

CHAPTER	NAME	CONTENTS	OBJECTIVES
1	Introduction	Background and background discussion, with the thesis main purpose and its focus and limitations.	To define the framework of this thesis.
2	Methodology	Scientific approach and research design, as well as the working procedure, and criticism to chosen methods.	To discuss this thesis methodology and how to approach the problem.
3	Theoretical Framework	Tape reading, high-frequency trading, and previous studies.	To present previous research.
4	Analysis	Summary statistics, testing the hypotheses, and try the best measures in a trading model.	To summarize the findings, and to see if the findings are relevant as predictors of future price movements.
5	Conclusion	Discussion of findings, and future research suggestions.	To discuss the findings in this thesis, and possible further studies.

2 METHODOLOGY

This part describes the chosen methodology for this thesis, and the discussion about the scientific approach and the approach of this thesis as a whole. Because the data is of major importance in my calculations, I will also provide descriptions of the data collected and from where it originates, and how it has been processed. I will also present some criticism to the chosen methods and the literature studied.

2.1 Choice of Scientific Approach

The main difference between qualitative and quantitative research is that the latter can reveal statistically significant differences in the study in question (Coldwell & Herbst, 2004, p.13).

When you write a thesis, it is of great importance to choose the right research strategy. There are two common approaches to choose from. A quantitative study emphasizes the collection of data, which is then analyzed and calculated mathematically. A qualitative approach, on the other hand, focuses instead on a non-numerical data collection and it cannot reach any general conclusions, only informative guesses (Bryman & Bell, 2007, p.28). A research is considered qualitative if the information cannot be analyzed by mathematical techniques (Coldwell & Herbst, 2004, p.13). The differences between quantitative and qualitative studies are presented in Table 2.1 below.

Table 2.1 The Difference between Qualitative and Quantitative Studies.

QUALITATIVE STUDIES	QUANTITATIVE STUDIES
Focus on understanding informant's point of view.	Focus on facts and/or reasons for social events.
Emphasize the importance of understanding.	Emphasize the importance on testing and verification.
Interpretation and rational approach	Logical and critical approach
Observations and measurements in natural settings.	Controlled measurements.
Subjective "insider view" and closeness to data.	Objective "outsider view" and distant from data.
Explorative orientation.	Hypothetical-deductive; focus on hypothesis testing.
Process oriented.	Result oriented.

(Ghuri & Grønhaug, 2002, p.86)

Even though most researchers prefer one research method to another, qualitative and quantitative methods can be combined (Ghuri & Grønhaug, 2002, p.86). Because mixed

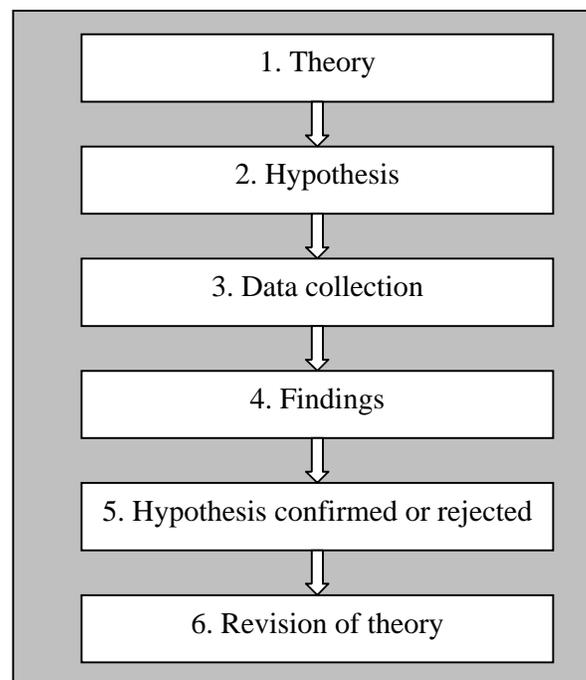
research methods is relatively new in social and human sciences, Creswell (2003, pp.209-210) recommends one to present a basic definition and description of the approach. Based on the reasoning mentioned above, I have decided to do a quantitative study.

2.2 Research Design and Methods

To answer the research problem in the best possible way, Ghauri and Grønhaug (2002, pp.47-50) claim that choosing the correct research design is critical. The quality of the empirical research is greatly influenced by the underlying research design. According to Kumar (2005, p.84) a research design is a procedural plan with the objective to answer questions validly, objectively, and accurately.

Ghauri and Grønhaug (2002, pp.13-15) describe inductive research as when one draws conclusions from empirical observations. One can never be 100 percent sure that the inductive conclusions drawn from empirical observations are accurate. Deductive research, on the other hand, either rejects or accepts the conclusions drawn from theories and hypotheses. Bryman and Bell (2007, pp.11-15) describe the deductive theory as the most common view of the nature of the relationship between theory and research. The process of deduction is illustrated in Figure 2.1 below.

Figure 2.1 The Process of Deduction.



(Bryman & Bell, 2007)

Since the deductive method involves moving from theory to empiricism, I find it appropriate to use this method in the thesis.

According to Ghauri and Grønhaug (2002, p.48) there are in general three main classes of research design; descriptive, exploratory, and casual research. If the research problem is badly understood, an exploratory research design is adequate. Ghauri and Grønhaug (2002, pp.49-50) state if the problem is well understood a descriptive research is suitable. Descriptive research is characterized by a structured problem, with precise rules and procedures. The main difference between descriptive and causal research is that in causal research the researcher also faces “cause-and-effect” problems.

Based on the research methods explained above, I think a deductive method and an exploratory research design is suitable for this thesis.

2.3 Working Procedure

In this section the working procedure of the thesis will be presented. I started out with literature studies, and after I found a subject that was intriguing, I collected the data needed. In this section I will give a general discussion about the data sample and how I will process the data.

2.3.1 Literature Studies

To be able to make this study possible, extensive literature studies were needed. The main purpose of reviewing the literature is to identify relevant concepts, frame the problem under scrutiny, and what the intended contribution of the study is supposed to be (Ghauri & Grønhaug, 2002, pp.44-45).

The literature studied consists mainly of books about trading, market risk and trading behavior. Books about research methods, scientific articles, journals and magazines have also been used. Reviewing past literature makes it easier to see whether the thesis can add to something new, and if similar research has been conducted before.

2.3.2 The Data Sample

The data sample is collected from Six Telekurs database in their trading application Six Edge 3.3. The tick-data collected is from the trading that occurred on the SSE, hence all prices are in Swedish Kronor (SEK), and when referred to money flow it implicitly means money flow in SEK.

The data sample consists of 1,331 days of tick-data (January 1, 2005 to April 22, 2010) with approximately 2.2 million trades. Table 2.2 below presents how the number of trades is distributed on a monthly basis. The data sample includes all trades made during the day; hence it includes Over-the-Counter (OTC) trades as well.

Table 2.2 Number of Trades Distributed over the Month.

	2010	2009	2008	2007	2006	2005
January	62,301	55,356	54,543	26,941	30,229	13,575
February	46,852	49,412	49,040	27,273	24,133	14,914
March	53,220	53,808	36,714	33,066	26,921	15,047
April	34,810	59,740	38,056	22,007	17,818	12,160
May	-	53,495	32,400	30,301	28,359	12,974
June	-	46,164	31,569	33,491	24,303	16,583
July	-	45,586	33,400	30,710	16,047	11,449
August	-	40,360	31,148	47,898	18,521	9,982
September	-	42,871	52,239	33,733	15,548	10,625
October	-	44,423	61,288	35,079	23,521	10,373
November	-	63,246	66,791	60,025	21,745	17,285
December	-	53,041	56,918	28,613	19,360	14,513
Total	197,183	607,502	544,106	409,137	266,505	159,480

One can clearly see a significant growth in the number of trades per year since 2005. There are several explanations for this, but the major reason is the advent of high-frequency trading. Another reason is that predetermined lot sizes at the SSE disappeared in October 13, 2008.

Table 2.3 SSE Closed due to Holidays that Occurred during the Sample Period.

HOLIDAY	DATE (YY-MM-DD)
New Year's Day	10-01-01, 09-01-01, 08-01-01, 07-01-01
Epiphany	10-01-06, 09-01-06, 06-01-06, 05-01-06
Good Friday	10-04-02, 09-04-10, 08-03-21, 07-04-06, 06-04-14, 05-03-25
Easter Monday	10-04-05, 09-04-13, 08-03-24, 07-04-09, 06-04-17, 05-03-28
International Workers' Day	09-05-01, 08-05-01, 07-05-01, 06-05-01
Ascension Day	09-05-21, 08-05-01, 07-05-17, 05-05-25, 05-05-05
National Day of Sweden	08-06-06, 07-06-06, 06-06-06, 05-06-06
Midsummer's Day	09-06-19, 07-06-22, 06-06-23, 05-06-24

Christmas Eve ¹	09-12-24, 08-12-24, 07-12-24
Christmas Day	09-12-25, 08-12-25, 07-12-25, 06-12-25
Boxing Day	08-12-26, 07-12-26, 06-12-26, 05-12-26

Out of the 2.2 million trades only 22 trades were excluded due to fact that the prices were either less than the days low or greater than the days high. These trades can be found in Table A.1 in Appendix 1. Due to technical problems, the opening had to be postponed for two days, thereof missing trades during the downtime period. SSE opened at 10:01:20 a.m. instead of 9:00:00 a.m. in April 20, 2006 and at 2:31:17 p.m. in June 3, 2008.

2.3.3 Analysis of Data

Dacorogna, Gençay, Müller, Olsen and Pictet (2001, p.6) state that the high-frequency data leads to new levels of significance.

According to Taleb (2004, p.112) there is a nonlinear relationship between the confidence level in a statistical method and the number of observations. If one increases the sample size by n times the confidence level will not increase by n times, it is merely multiplied by the square root of n .

The highest level of high-frequency data refers to tick data, where the data arrives at nonlinear time intervals. Gavridis (1998, p.5) claims; the more global the financial markets have become, the finer the frequency at which the analysis is conducted. He also states that the growing complexity and automation of markets require greater use of analytical tools to maintain real-time risk control.

Table 2.4 An Example of the Ticker Tape in ABB Ltd.

TIME	PRICE	BUYER	SELLER	VOLUME	TRADE NO.
17:24:15	134.25	CDG	EPB	10,400	2,516
17:23:45	134.50	BPP	HQB	10,000	2,515
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
09:01:19	137.25	SHB	MSI	1,693	2
09:01:17	137.25	NEO	MSI	2,927	1

¹ The day before an official holiday is in most cases treated as a de facto holiday, Christmas Eve is one of them.

To be able to analyze approximately 2.2 million trades, I needed to write several algorithms to crunch all these numbers. These algorithms summarized all trades made during each day into a trade summary, with following data presented in Table 2.5 – 2.7.

Table 2.5 Daily Trade Summary.

DATE	VWAP	OPEN	CLOSE	HIGH	LOW	MID	C2C
2005-01-03	37.94	37.50	38.00	38.40	37.40	37.90	1.88%
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2010-04-22	148.38	150.30	147.20	151.20	144.20	147.70	-7.42%

The daily trade summary in Table 2.5 presents the trading range for each day, i.e. the opening and closing price, as well as the high, low and mid price of the day. The Volume-Weighted Average Price (VWAP) is also presented, and C2C is short for Close-to-Close in percent. For a more detailed presentation of what VWAP is, I refer to 3.2.2.1.

Table 2.6 Continuation of Previous Table, Daily Trade Summary.

NO. BUYERS	NO. SELLERS	OFI	BUY VOL.	SELL VOL.	NET VOL.
393	392	0.99	1 381 335	809 551	571 784
⋮	⋮	⋮	⋮	⋮	⋮
7 437	5 666	0.76	6 256 334	7 013 850	- 757 516

Table 2.6 is a continuation of the daily trade summary, which displays number of buyers and sellers during the day, and the buy and sell volume, respectively. The On Floor Information (OFI) is the ratio between sellers and buyers, but for a more detailed explanation I refer to the findings of Steidlmayer and Hawkins (2003) in section 3.3.2. There is never an exact way to measure the number of buyers and sellers; one reason is because buyers (sellers) can buy (sell) just milliseconds apart. So if two traders using the same broker, buy (sell) within the same second, it would look like only one buyer (seller) which is not true. In one of my algorithms used (see Appendix II) I defined a buyer (seller) as all consecutive trades bought (sold) with the same broker. For example, take a look at Table 2.4 above, where Morgan Stanley International (MSI) first sold 2,927 shares and two seconds later sold another 1,693 shares. It is very likely that the two trades were executed by the same seller, thus the two first trades had only one seller, but two buyers.

Table 2.7 Continuation of Previous Table, Daily Trade Summary.

NORM. N.V.	NET MONEY FLOW	NORM. OFI	VWAP SKEW
26.10%	21 670 613.60	-1.01%	0.12%
⋮	⋮	⋮	⋮
-5.71%	- 111 885 113.20	- 23.81%	0.46%

Table 2.7 continues to present the daily trade summary, where the four main measures used in this thesis are displayed. The two first measures, Normalized Net Volume (Norm. N.V.) and Net Money Flow, are closely related, and they both measure if buyers or sellers are in control. The third measure, Norm. OFI, is just the normalized On Floor Information from Table 2.6 and the last measure describes how the VWAP is skewed compared to the mid price of the day. All these measures are presented in the beginning of chapter 4.

2.3.4 Statistical Methods

To evaluate the validity of a point forecast is to run a regression of realized values from historical data against the model's forecasts (Aldridge, 2010, p.220). Equation (1) describes the general linear regression model, hence the regression is linear in the parameters, β_0 and β_i , not necessarily in the explanatory variables (Gujarati, 2006, p.144).

$$Y_t = \beta_0 + \beta_t \cdot \mathbf{X}_t + \varepsilon_t \quad (1)$$

The classical regression model, equation (1), is used to describe the static relationship between a dependent variable, Y_t , and i independent variables $\mathbf{X}_t = (X_{1,t}, \dots, X_{i,t})^T$ and with unknown parameters $\beta_t = (\beta_{1,t}, \dots, \beta_{i,t})^T$ (Madsen, 2008, pp.31-33). I will use both simple ($i = 2$) and multiple regression ($i \geq 3$) models to test my null hypothesis, but a more thorough explanation of the regression models used will be presented in chapter 4.

According to Aldridge (2010, p.222) it is a greater challenge to test directional forecasts accuracy, than regular point forecasts.

2.3.5 Credibility

The credibility of the findings of this thesis is essential. Bryman & Bell (2007, p.162) argues that there are three criteria that are extremely important to take into account the assessment of economic research; reliability, validity and replication.

I have striven to meet the requirements necessary for a proper master thesis in Economics. I believe that the study is reliable, and that any differences or random errors have been minimized. I also believe that the study is credible and reliable and that theory, and empirical findings are related. Furthermore, both the method, approach and the statistical calculations used in this study is explained carefully, so that the test made can be replicated.

2.3.5.1 Reliability

Taleb (2008, p.185) concludes that there are severe problems with projecting from the past, the past data can confirm ones theory but it can also do the opposite. Not only can the past be misleading, but there are also numerous degrees of freedom in interpreting past events.

Bryman and Bell (2007, pp.40-41) describe reliability as the degree of whether the results of a study are repeatable or not. Concerns about the reliability in a study, are especially an issue for quantitative research. It refers to the stability of the measure (Ghauri & Grønhaug, 2002, p.68).

Mainly two errors can be made in hypothesis testing, it is either a Type I error or Type II. Making a Type I error is when one rejects the null hypothesis when it is actually true. A Type II error is the opposite, one does not reject the null hypothesis when it is in fact false. Table 2.5 illustrates the differences (Gujarati, 2006, p.116).

Table 2.8 Type I and Type II Errors.

	REJECT H_0	DO NOT REJECT H_0
H_0 is true	Type I error	Correct decision
H_0 is false	Correct decision	Type II error

(Gujarati, 2006, p.116).

Gujarati (2006, p.116) states that the only way to decrease type a Type II error without increasing a Type I error is to increase the sample size. The probability of committing a Type I error is equal to the significance level.

2.3.5.2 Validity

Validity is an important key to effective research, because if the research is invalid then it is useless (Cohen et al. 2007, p.133). According to Bryman and Bell (2007, pp.41-43) one of the most important criterions in research is its validity. A series of corroborative facts is not enough to be absolutely certain. On the other hand, one observation that falsifies the whole statement is enough. As Taleb (2008, p.56) states; seeing a white swan does not confirm the

hypothesis of non-existence of black swans, but seeing a black swan confirms the statement that all swans are not white.

The tick-data collected is tested with an independent data provider (Infront AS), to guarantee that the data sample is valid. The test compares all trades during 2005 to late April 2010, and ambiguities in the data are corrected. For some unknown reason, tick-data for two days are missing, October 18, 2007 and February 9, 2009, respectively. This validity test diminishes the risk of invalid data.

2.3.5.3 Objectivity

Humans have a hunger for rules to help us reduce the dimension of matters. With all our brain cells it is not the storage capacity that limits us; it is rather the complexity of things. Humans need to simplify in order to store complex matters in our minds. Instead of remembering vast amount of data, we seek for patterns in the data. Once a pattern is found, we remember this pattern or rule because it is much more compact than the raw data. As Taleb (2008, p.69) describes it:

“[...] the same condition that makes us simplify pushes us to think that the world is less random than it actually is.”

The thesis has been conducted in a way that is as objective as possible. I have presented criticism of chosen methods (see 2.4 below), and clarified the reasons behind different choices made during the working procedure.

2.4 Criticism to Chosen Methods

Taleb (2008) formulated one of his theories about different kinds of randomness. There are two types of randomness, either type 1 or type 2. Taleb (2008) declares type 1 randomness to be matters that seem to belong to what he calls *Mediocristan*, and type 2 randomness belongs to *Extremistan*. He defines *Mediocristan* as matters that are subjected to gravity, where the most typical member is mediocre. Extreme events are rare and the total can not be dictated by a small number of events. In *Extremistan* on the other hand, there are no physical boundaries, and a few extreme events can alter the total. Randomness from the latter is extremely wild and it is hard to predict from past information. Randomness from *Mediocristan* is easy to predict from past data, and events are distributed according to the Gaussian bell curve or its variations (Taleb, 2008, pp.32-37).

To illustrate the differences, imagine a sample with physical boundaries, e.g. the height or weight of the human body. If one randomly picks 1000 people and measure their weight (or

height), the average weight will not be significantly affected if one adds the most extreme event possible, i.e. the heaviest person alive. The average before and after will be

$$\frac{x_1 + \dots + x_n}{n} \approx \frac{x_1 + \dots + x_n + x_{n+1}}{n+1} \quad (2)$$

and as can be seen the equation above holds even though $x_{n+1} \gg x_n$. For a sample from Extremistan on the other hand, i.e. a sample without physical boundaries e.g. wealth, the equation above will not hold. If one randomly picks 1000 people and measure their wealth, the average will be significantly affected if one adds the most extreme event possible, i.e. the richest man in the world.

The reason why I presented Taleb's ideas above is to emphasize the importance of knowing that financial data is from Extremistan, and one should be aware of it when conducting the analysis. It is also important to highlight the criticisms can be raised against selected sources. Ghauri and Grønhaug (2002, pp.78) believe it is important to be critical when using different types of secondary sources. One may consider where the data comes from, who collected the information, and how credible the source is. This thesis has primarily used books about trading and the data is retrieved from SIX Telekurs. The literature comes from well-known and reputable publishers, which should be quite credible.

Other sources and references used in the study, such as websites and articles, have also been examined to ensure that they are reliable, serious and useful.

3 THEORETICAL FRAMEWORK

In this chapter I will present the theoretical framework that is necessary for this thesis. An introduction of tape reading will be presented in section 3.1. This is because tape reading is the underlying foundation that made this thesis possible. Since markets around the world have been flooded with high-frequency algorithms, lately, this thesis will try to examine if algorithmic strategies gives a hint of future price movements. Previous studies are discussed to enable the reader some insight of the measures of money flow and floor activity.

3.1 Tape Reading

The ticker tape, henceforth the tape, consists of the market action, i.e. the buying and selling of securities. Every trade that has taken place is recorded on the tape (Neill 1931/1970, p.29). The tape records money transactions and the specific information can differ between exchanges, but time, price, and volume are always displayed. Some exchanges also display the identity of the buyer and seller on the tape.

Rollo Tape (1910/2003, p.10) a.k.a. Richard D. Wyckoff describes tape reading as a method of forecasting what appears on the tape now, what is likely to appear in the near future. It is also about how to determine whether stocks are marked up or down, being accumulated or distributed, or neglected by the large interests. Tape Reading gauges the monetary supply and demand in a particular stock or in the market as a whole.

The advantage with tape reading is that the tape tells news minutes, hours and days before the news tickers and newspapers (Tape, 1910/2003, p.11). Today these numbers are not quite correct, it has shrunk to seconds and minutes rather than hours and days.

Neill (1931/1970) states that the tape interpretation depends upon consideration of the action of the volume, and not the price. It is not the price but the volume that best measures the supply and demand in a stock. It makes great difference if the buyer is willing to buy 100,000 shares or just 100 shares at any given price. The demand is greater in the first case, as is the supply. For every stock bought there is a seller or sellers. A tape reader's job is to determine whether the demand is greater than the supply, in which case the price advances. The reverse is true if the supply is greater than the demand (Neill, 1931/1970, p.41).

The price and volume on the tape represent the general consensus of opinion, the supply and demand and the aggregated effect of manipulation in the market (Tape, 1910/2003, p.74). Lukeman (2000, p.209) states that volume is the second dimension of technical analysis, and it indicates the interest in a stock at any given time.

For the untrained eye, following the tape can be incomprehensible, but with a little knowledge of a few stock symbols and a lot of persistent training, one will get in sync with the market (Foster, 1965/2005, p.67).

To detect if the volume indicates an advance in a particular stock, one can study the transactions recorded on the tape. The degree of persistency in the advance will depend on what kind of market participants that is active in the stock. If the rally mainly consists of short sellers covering their positions rather than long buyers, Neill (1931/1970, p.48) states it is likely to be brief. In a rising market a short seller needs to be more aggressive than a long buyer because the former is losing money, while the latter is just missing an opportunity to make money.

Graifer and Schumacher (2004, p.17) present their idea of tape reading as to differentiate between the smart money, i.e. well-informed traders, institutions, and insiders, from the public. The aim is to identify the smart money's footprints on the tape, and positioning oneself on the winning side. This endeavor has become harder as the smart money try to disguise themselves as the public, using different algorithmic strategies. Nowadays, many buy-side customers such as institutional fund managers use algorithmic trading strategies to improve transaction privacy and to cut costs (Collins, 2008).

Don Worden claims the term tape reading was probably the original term for what is now generally known as Technical Analysis. Originally the knowledge abstracted from the ticker tape is referred to as tape reading, which includes price and/or volume analysis (Foster, 1965/2005, p.81). During 1963, Worden published his ideas in an article in the *Encyclopedia of Stock Market Techniques*, where he stated the importance of volume and large blocks (Hale, 1995, p.13).

3.2 High-Frequency Trading

The main differences between high-frequency and low-frequency trading are the high turnover of capital, the lower average gain per trade, the computer-driven responses to changing market conditions, and the high number of trades. The globalization of capital markets favors high-frequency, because there is little if any overnight risk. Besides the riskiness in holding overnight positions, it can also be quite costly. The overnight carry rate, i.e. the interest rate paid to cover overnight positions on margin, is usually slightly above LIBOR (London Interbank Offered Rate). The cost of carry can become increasingly

expensive with volatility in LIBOR and hyperinflation around the corner (Aldridge, 2010, pp.1-2).

According to Aldridge (2010, p.1) the majority of the high-frequency managers delivered positive returns in 2008, whereas 70% of the low-frequency practitioners lost money. Renaissance Technologies Corp. has long been one of the most prominent practitioners of high-frequency strategies. The founder of Renaissance Technologies Corp., Jim Simons, was the highest earning investment manager of 2008, where Dr. Simons earned \$2.5 billion in 2008 alone.

3.2.1 Orders, Traders and Liquidity

To be able to analyze all the tick-data, and to understand how it is created, I will give a brief introduction to what kind of orders and traders there are that create liquidity in the market. For every data-point (i.e. every trade made), a market order is crossed with a limit order, either on the bid or the offer.

The two most common order types are market orders and limit orders. There are pros and cons with both order types, the main difference is where the certainty lies. Aldridge (2010, p.63) presented the main differences between limit orders and market orders, which can be seen in Table 3.1 below.

Table 3.1 Limit Orders versus Market Orders.

	LIMIT ORDERS	MARKET ORDERS
Time to Execution	Uncertain	Instantaneous
Order Execution	Uncertain	Certain
Transaction Costs	Low	High
Order Resubmission	Infinite prior to execution	None
Execution Price	Certain	Uncertain

As presented in Table 3.1 above, one can see that the order execution is certain for market orders but not for limit orders. On the other hand is the execution price certain for limit orders but uncertain for market orders. As Borsellino (2001, p.34) and Harris (2003, p.77) state, there is no guarantee that ones limit order will be executed.

According to Harris (2003, p.75) a trader who submits a limit order, writes an option to other market participants. If a trader posts a buy limit order, it is equal of writing a put option where

the options strike price equals the limit price. Posting a sell limit order equals writing a call option.

Harris (2003, p.222) generalize the trading society into two groups, informed traders and uninformed traders. The informed traders include arbitrageurs, news traders, value traders, and informative-oriented technical traders. These traders buy when prices are undervalued and sell when prices are overvalued relative to their estimates of fundamental value. The uninformed traders, on the other hand, includes liquidity traders (market makers) who aim to profit from providing liquidity and following short-term price momentum.

Aldridge (2010, pp.62-63) considers that liquidity can generally be measured by the limit order book. The market depth is classified as the total volume of limit orders available at a specific price, whereas the market breadth is number of different prices where limit orders exists. O'Hara (2002) has a more general view of liquidity, which refers to the matching of buyers and sellers.

In the limit order book, time and price priority are strictly enforced. Orders are generally executed in the sequence they were received (time priority), given two orders with the same limit price. If two orders have different limit prices, the one with highest (lowest) price will be executed first (price priority) for buy (sell) orders (Harris & Hasbrouck, 1996, p.219).

3.2.2 Algorithmic Trading

According to Brown (2010, p.27), electronic trading is commonly referred to as Direct Market Access (DMA), this because an investor can directly execute the order electronically without any intermediary, such as a broker or market maker. Schubert (2009, p.3) reports that the volume associated with program trading has increased significantly during the last decades. On the NYSE, program trading has grown from 4.95% in 1989 to 34.07 % in 2007 of the total volume. Program trading helps the buy-side to have the ability to evaluate risk and volatility of a basket, instead of each individual stock. Program trading was one of the first forms of algorithmic trading, and a more in-depth introduction to the subject will be presented below.

According to Brown (2010, p.10) is algorithmic trading the brokerage firm's contribution to black-box trading. Fabozzi, Kolm, Pachamano and Focardi (2007, p.459) state that algorithmic trading offers great benefits by increasing productivity and managing trades in a more cost-effective way. Trading algorithms need to be extremely flexible to be able to adapt to changing market conditions. According to Pole (2007, pp.183-185) was Morgan Stanley at the forefront of the creation of algorithmic trading. It started around 25 years ago, when

Morgan Stanley created a program to buy and sell stocks in pairs, a.k.a. pairs trading. During the first decade of the 21st century, algorithmic trading tools have been developed and marketed by Morgan Stanley and its competitors, e.g. Goldman Sachs, Credit Suisse First Boston, Bank of America, and others.

Narang (2009, pp.100-105) explains order execution algorithms as real-time decision makers, that decides whether the trade is urgent, the size of the trade and how to conceal the order for other market participants. The urgency dictates whether the execution algorithm should be passive or aggressive, using limit orders or market orders, respectively. To reduce transaction costs for large orders, the execution algorithm divides the order into smaller blocks. A more thorough explanation of these strategies can be found below, see section 3.2.2.1 – 3.2.2.4.

3.2.2.1 Volume-Weighted Average Price (VWAP)

Schubert (2009, p.41) states that the main benchmark for algorithmic trading is the Volume-Weighted Average Price (VWAP). Kim (2007, p.56) defines VWAP as the dollar amount traded for every transaction during a certain period divided by the total shares traded during the same period.

Kakade, Kearns, Mansour and Ortiz (2004, pp.2-3) assume that the intraday trading activity in a given stock can be summarized by a discrete sequence of price and volume pairs (p_t, v_t) for $t = 1, \dots, T$. The day's market open corresponds to $t = 0$, and the close to $t = T$. The pair (p_t, v_t) represents the total of v_t shares traded at an average price per share p_t during time $t - 1$ and t . For an intraday trading sequence $S = (p_1, v_1), \dots, (p_t, v_t)$ the market VWAP will be

$$VWAP_M(S) = \frac{\sum_{t=1}^T p_t \times v_t}{V} \quad (3)$$

where V is the total daily volume, i.e. $V = \sum_{t=1}^T v_t$.

3.2.2.2 Time-Weighted Average Price (TWAP)

Schubert (2009, p.40) describes Time-Weighted Average Price (TWAP) as a strategy that tries to capture the time-weighted average of the day's price range. In contrast to VWAP, which depends on the market volume, TWAP will use the same volume throughout the day. For large block traders, using a TWAP strategy is a more efficient way to distribute a large number of shares to the market. TWAP aims to better reflect the true market price, than to buy

or sell a large block all at once. It reduces liquidity risk and limits the market impact. For an intraday trading sequence $S = (p_1, v_1), \dots, (p_t, v_t)$ a TWAP order can be calculated as

$$TWAP(S) = \frac{\sum_{t=1}^T p_t \times v}{V} \quad (4)$$

where V is the total volume of the order, and v is one n^{th} of the total volume, hence $V = v \times n$. As can be seen from the equation above, a TWAP order allows the trader to “time-slice” a trade over a certain period of time (Kim, 2007, p.60).

3.2.2.3 Implementation Shortfall

Kim (2007, p.54) states that the optimal trading strategy needs an accurate measure of trading costs and how well the strategy is implemented. In paper trading, i.e. trading without real money at stake, one does not need to measure trading costs and implementation shortfall. The main difference between paper trading and real trading is in the real world slippage, commissions, and market impact has to be taken into consideration. Dr Elder (2002, pp.28-30) describes slippage as the difference between the price observed in the market when the order is placed and the actual price when the order is filled. A limit order is slippage free in the sense that the price is fixed; however there is a risk the order will not get filled. A market order on the other hand will get filled with certainty, but the price is uncertain.

In an ideal portfolio slippage, commission and market impact does not exist, but in the actual portfolio it does. The difference between these two is called implementation shortfall. Perold (1998, pp.106-111) proposes a way to assess the drag on performance caused by the problems associated with implementing ones portfolio decisions. By monitoring the implementation shortfall one will be able to better understand what affects the trading performance. Implementation shortfall enables one to separate bad research from poor implementation and execution.

3.2.2.4 The New Generation of Execution Algorithms

The IQx algorithm is one of the new generation execution algorithms, with the purpose to be a more effective alternative than a market order. The trading horizon will be up to 15 minutes but the order may be completed faster depending on market conditions (Schubert, 2009, p.43). BNY ConvergeEx Group (2009) explains that their IQx strategy engine is comprised of a broad array of advanced trading tactics. The IQx strategy engine tries to maximize alpha through predicting whether if the execution should be passive or aggressive.

According to NYSE Euronext (n.d.) peg orders are only accepted in continuous trading, i.e. not during the opening and closing call, respectively. Schubert (2009, p.44) explains peg orders as a limit order that moves in tandem with the market. For a buy peg order with a limit price of \$11 and the current bid-ask spread being \$10.00 – \$10.20, the peg order will be posted at \$10.00. If a new best bid comes in at \$10.10, the peg order will be modified with a new limit price of \$10.10. If the security rises above \$11 the peg order will remain in the market with a limit order at \$11, even though the best bid is \$11.25. For a thorough explanation see NYSE Euronext (n.d.).

3.3 Previous studies

Despite little published research regarding money flow, it has become a popular technical indicator. Studies made by Bennet and Sias (2001) and Steidlmayer and Hawkins (2003) are the only research I found about money flow, and how it affects stock market prices. A brief introduction of their work is presented below.

3.3.1 Findings of Bennett and Sias (2001)

Bennet and Sias (2001) conducted a vast study of money flow in 1,622 companies during a 12-month period (252 trading days) from July 1997 through June 1998.

They focused on four questions:

- Does money flow provide useful information to investors?
- Is money flow's ability to predict returns sensitive to the definition of money flow?
- Does the ability of money flow to predict returns differ for different forecast horizons?
- Does money flow's predictive ability arise from its ability to predict common factor returns or company-specific returns?

Bennet and Sias (2001) expressed money flow on a given day, summed over the n trades that occurred at different prices from the previous trade, as

$$\text{Money flow} = \sum_{t=1}^n \text{Volume}_t \cdot P_t \cdot \frac{P_t - P_{t-1}}{|P_t - P_{t-1}|} \quad (5)$$

Rearranging the equation (5) yields

$$\text{Money flow} = \sum_{t=1}^n \left(\frac{P_t \cdot P_{t-1}}{|P_t - P_{t-1}|} \right) \cdot \text{Volume}_t \cdot \frac{P_t - P_{t-1}}{P_{t-1}} = \sum_{t=1}^n \left(\frac{P_t \cdot P_{t-1}}{|P_t - P_{t-1}|} \right) \cdot \text{Volume}_t \cdot R_t \quad (6)$$

where $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$, i.e. the return from period $t-1$ to t . Equation (6) consists of three parts, the first part will always be positive, thus $P_t > 0: \{t\} \in Z^+$, as well will the second part because $volume_t \geq 1: \{t\} \in Z^+$. The last part, however, will change sign depending on the return during the period observed. Note the difference between money flow from one trade to another, and money flow observed over a greater period of time. Money flow from one trade to another will always affect the cumulative money flow in the same direction as the trade's contribution to the cumulative return. Up-tick (down-tick) trades increase (decrease) the money flow as well as the return. Over a greater period of time, however, the money flow and return do often differ in sign (Bennet & Sias, 2001, pp.65-66).

Bennet and Sias (2001) concluded that the intuitive interpretation of money flow as a technical indicator is consistent with the hypothesis that money flow does measure excess demand and supply. Their findings documented strong positive correlation between money flow and return measured over the concurrent period. Bennet and Sias (2001) also concluded that money flow exhibits strong positive serial correlation, i.e. high money flow today, forecasts high money flow in the future. Bennet and Sias (2001) most important finding is that money flow appears to provide investors with information regarding future returns. They stated that the ability of money flow to explain cross-sectional variation in future returns depends on both the measure used and the time horizon.

3.3.2 Findings of Steidlmayer and Hawkins (2003)

Steidlmayer and Hawkins (2003, p.131) state that the real purpose and economic function of any market is to distribute a product as efficiently as possible. Even though people associate the term distribution with markets, it has never been fully examined until Steidlmayer presented his theory of the Market Profile. Steidlmayer developed the Market Profile during 1981 to 1983 while he served on the board of directors at the Chicago Board of Trade (CBOT) (Bukey, 2005, p.38).

On Floor Information (OFI) has only been available for CBOT contracts. At the end of the trading day the OFI number is calculated. It consists of the average size of the buy order and the average size of the sell order. The ratio between the average size of the buy order and the sell order yields the OFI number (Steidlmayer & Hawkins, 2003, p.101). The OFI number can be calculated as

$$OFI = \frac{\frac{Volume_{buy}}{buyers}}{\frac{Volume_{sell}}{sellers}} = \frac{sellers}{buyers} \quad (7)$$

where $Volume_{buy} = Volume_{sell}$. The equation above has been slightly modified with the one presented by Steidlmayer and Hawkins (2003). They presented their theory of the OFI number with the following assumption $Volume_{buy} \neq Volume_{sell}$. The only explanation I have for this faulty assumption is that the OFI numbers are only generated by floor trade only. Since this thesis will study tick-data from one of the equities listed on SSE where no floor trading exists, the volume from buyers and sellers will be equal ($Volume_{buy} = Volume_{sell}$). How to interpret the OFI is presented in Table 3.2 below.

Table 3.2 An Interpretation of Different OFI Numbers.

OFI NUMBER	INTERPRETATION
$OFI > 1$	The average sell order is greater than the average buy order.
$OFI = 1$	The average buy order is equal to the average sell order.
$OFI < 1$	The average buy order is greater than the average sell order.

The OFI number may not be as useful today as it was in the 1980s and 1990s, this because VWAP and TWAP strategies has become vastly used. Kim (2007, p.60) claims that these strategies discourage block trading. The block order is sliced into smaller blocks dependent on the volume for VWAP orders or equally spaced in time for TWAP orders. To prevent other traders to front run a large order, money managers are now trying to disguise their order as order flow from small investors.

Steidlmayer Software developed the Volume @ Time (V@T) in 1998, and it measures the money flow. To measure the money flow the V@T compares the price range between the two most recent half-hours. Comparing the half-hour's range to the previous, one can determine whether it rotated up or down, or if it was an inside or outside half-hour. An upward rotation indicates that the last half-hour's trading took out the previous half-hour's high but not the low, the opposite is true for a downward rotation (Steidlmayer & Hawkins, 2003, p.106). The definitions for the price ranges are presented in Table 3.3 below.

Table 3.3 Volume @ Time Definitions of Price Ranges.

	HIGH	LOW
Upward rotation	$H_t > H_{t-1}$	$L_t \geq L_{t-1}$
Downward rotation	$H_t \leq H_{t-1}$	$L_t < L_{t-1}$
Inside half-hour	$H_t \leq H_{t-1}$	$L_t \geq L_{t-1}$
Outside half-hour	$H_t > H_{t-1}$	$L_t < L_{t-1}$

For inside and outside half-hour bars, volume is not calculated because market direction is not assignable. During half-hours with upward rotation the volume traded is added, and during half-hours with downward rotation the volume is subtracted. The net money flow is defined as the sum of all half-hour rotations up or down with their corresponding volume. The V@T can be calculated as

$$V @ T = \sum_{t=1}^T D_t \times V_t \quad (8)$$

where D_t is a dummy variable which is either 1 or -1, and V_t is the volume during half-hour t . According to Steidlmayer and Hawkins (2003, p.107) their V@T measures the market correctly because on average at the end of the day, when there is positive net money flow, the price is higher than the previous day.

4 ANALYSIS

In this chapter I will present the results from my calculations. The statistical measures used in the thesis will be discussed, followed by tables and graphs to explain my findings.

4.1 Summary of Measures

The four main measures used in this thesis to analyze the tick-data are; Normalized Net Volume, Net Money Flow, Normalized OFI and VWAP Skew.

Norm. N.V. is a measure inspired by Worden's theories as well as V@T presented by Steidlmayer and Hawkins, see section 3.1 and 3.3.2, respectively. Even though the amount of shares that buyers buy is exactly the same as the sellers sell, one need to differentiate between buy volume and sell volume. This differentiation is arbitrary, and in my analysis I have defined buy and sell volume, respectively, in Table 4.1.

Table 4.1 Definition of Buy and Sell Volume, Respectively.

BUY VOLUME	SELL VOLUME
All shares traded on the ask price.	All shares traded on the bid price.

By my definition above, I let market orders define if it is buy or sell volume, respectively. The bid and ask spread is composed by limit orders, recall section 3.21, and a trade is executed when a market order is crossed with a limit order. If someone who is placing an order to buy at the market, thinks that the execution is urgent and hence it is buy initiated. The opposite is true for sell market orders.

My definition of N.M.F. differs from the one Bennet and Sias (2001) defined in section 3.3.1. If one recalls equation (6) the return from period $t-1$ to t is one of the factors. If one calculates the money flow with equation (6) using tick-data, most of the volume will be ignored, because trades tends to cluster at either the bid or ask and after a while bounce the other way. If ten trades are traded at the bid the $R_t = 0$ for nine trades, given the next trade will be on the ask. This results in neglecting many trades, and consequently a lot of money flow.

Since stock prices fluctuate through time, the data had to be normalized so trades that occurred five years ago can be compared with trades today. Instead of measuring Net Volume (N.V.) during a specific day, which has no boundaries, normalizing with the total volume during the day, sets a lower and upper boundary, -100% and 100% respectively. This

normalized measure can be compared from one day to another. The Norm. N.V. on day t is presented in the equation below.

$$\text{Normalized Net Volume}_t = \frac{\text{Buy Volume}_t - \text{Sell Volume}_t}{\text{Buy Volume}_t + \text{Sell Volume}_t} \quad (9)$$

To be able to measure the N.M.F., i.e. the money flowing in or out of the stock at a specific point in time, one simply multiplies the N.V. with the stock price at that time. Choosing the best stock price during the day is arbitrary, there are several prices to choose from, e.g. the opening price, closing price, VWAP price, as well as the high and low of the day. Since the opening and closing price tend to lie close to the days high or low, the mid price gives a more stable and consistent measure. The mid price for day t is calculated as

$$\text{Mid Price}_t = \frac{P_{H,t} - P_{L,t}}{2} = P_{mid,t} \quad (10)$$

and the N.M.F. is calculated with the equation below.

$$\text{Net Money Flow}_t = P_{mid} \cdot (\text{Buy Volume}_t - \text{Sell Volume}_t) \quad (11)$$

As I describe in section 3.3.2, I have used Steidlmayer and Hawkins' OFI-ratio as a measure of what the professionals are up to. The OFI number has been normalized with the equation below, and a Norm. OFI value that is larger than 0 describes a trading day with more sellers than buyers, and a value less than 0 describes the opposite.

$$\begin{aligned} \text{Normalized OFI}_t &= \frac{\frac{\text{Volume}_{buy,t}}{\text{buyers}_t}}{\frac{\text{Volume}_{sell,t}}{\text{sellers}_t}} - \frac{\frac{\text{Volume}_{sell,t}}{\text{sellers}_t}}{\frac{\text{Volume}_{sell,t}}{\text{sellers}_t}} = \frac{\frac{\text{Volume}_{buy,t}}{\text{buyers}_t}}{\frac{\text{Volume}_{sell,t}}{\text{sellers}_t}} - \frac{\text{Volume}_{sell,t}}{\text{sellers}_t} \\ &= \frac{\text{sellers}_t - \text{buyers}_t}{\text{buyers}_t} = \frac{\text{sellers}_t}{\text{buyers}_t} - 1 \end{aligned} \quad (12)$$

The VWAP Skew measures if the trading volume occurred at a higher or lower price than the mid price during the day's trading. For a more detailed description of VWAP, see section 3.2.2.1. The VWAP Skew for day t can be calculated with the formula below.

$$\text{VWAP Skew}_t = \frac{P_{VWAP,t} - P_{mid,t}}{P_{mid,t}} \quad (13)$$

4.2 Summary Statistics

Table 4.2 presents a summary of essential statistics for the stock return on a close-to-close basis, the Normalized Net Volume, the Net Money Flow, the Normalized OFI, and the VWAP Skew, respectively. For a more detailed presentation of how the numbers were calculated, I refer to Appendix IV.

Table 4.2 Summary Statistics.

	Return	Norm. N.V.	Net Money Flow	Norm. OFI	VWAP Skew
Mean	0.001294	0.009275	1 000 437.48	- 0.003036	0.000184
S.E. ²	0.000626	0.005184	2 426 422.03	0.004852	0.000074
Median	0.001332	- 0.003255	- 1 270 862.63	- 0.029830	0.000147
S.D. ³	0.022834	0.189133	88 522 846.04	0.177015	0.002686
Variance	0.000521	0.035771	7 836 294 270 280 840	0.031334	0.000007
Kurtosis ⁴	8.222115	0.219424	5.337729	0.854012	5.662475
Skewness	- 0.333295	0.213642	0.236838	0.720170	0.277726
Range	0.350136	1.255040	1 019 273 384.50	1.207778	0.037633
Minimum	- 0.197802	- 0.555115	- 530 508 280	- 0.430000	- 0.016870
Maximum	0.152334	0.699924	488 765 105	0.777778	0.020763
Sum	1.722605	12.344869	1 331 582 281	- 4.040694	0.245297
Sample Size	1 331	1 331	1 331	1 331	1 331

As one can see from Table 4.2 above, the return, N.M.F. and the VWAP Skew are all leptokurtic, while the Norm. N.V. and Norm. OFI. are almost mesokurtic. All measures show a positive skewness, while the return, which is the dependent variable, shows negative skewness. Plots of the frequency distributions as well as the cumulative distributions for the dependent and independent variables are presented in Appendix V.

4.3 Hypothesis Testing

To see if there are statistically significant relationships between the dependent variable (daily stock return) and the independent variables (Norm. N.V., N.M.F., Norm. OFI and VWAP Skew) a linear regression will be made. Firstly, I will test each independent variable against daily stock returns, with the null hypothesis $H_0 : \beta_i = 0, \{i = 1, \dots, 4\}$ against an alternative hypothesis $H_1 : \beta_i \neq 0, \{i = 1, \dots, 4\}$. Secondly, I will do a multiple regression to test the joint hypothesis $H_0 : \beta_5, \beta_6, \beta_7 = 0$, which is the same as $H_0 : R^2 = 0$ against the alternative hypothesis $H_1 : R^2 \neq 0$.

² S.E. = Standard Error.

³ S.D. = Standard Deviation.

⁴ The kurtosis displayed in Table 4.1 is the excess kurtosis, i.e. kurtosis – 3, see Appendix IV for an exact definition.

To see if there is a linear relationship between the dependent variable (daily stock return) and the independent variables (Norm. N.V., N.M.F., Norm. OFI and VWAP Skew) a regression analysis for each independent variable will be presented below, see equation (14) to (17).

$$\hat{R}_t = \beta_0 + \beta_1 \cdot \text{Norm. N.V.}_{1,t} \quad (14)$$

$$\hat{R}_t = \beta_0 + \beta_2 \cdot \text{N.M.F.}_{1,t} \quad (15)$$

$$\hat{R}_t = \beta_0 + \beta_3 \cdot \text{Norm. OFI}_{1,t} \quad (16)$$

$$\hat{R}_t = \beta_0 + \beta_4 \cdot \text{VWAP Skew}_{1,t} \quad (17)$$

The results of the regressions are illustrated in the regression plots, see Figure 4.1 – 4.4. The regression statistics are presented in Table 4.3.

Table 4.3 Regression Statistics for the Two Variable Models.

	Norm. N.V.	Net Money Flow	Norm. OFI	VWAP Skew
Multiple R	0.396948	0.451620	0.400740	0.336346
R-Squared	0.157568	0.203960	0.160592	0.113129
Adjusted R-Squared	0.156934	0.203361	0.159961	0.112461
Standard Error	0.020966	0.020381	0.020929	0.021512
Observations	1331	1331	1331	1331

In Table 4.3 above, one can see that the goodness of fit, R^2 , is best for the regression model using N.M.F. as an explanatory variable. Table 4.4 – 4.7 display the regression statistics which summarizes the coefficients, standard error, t-stat, and p-value, respectively.

Table 4.4 Regression Statistics for the Normalized Net Volume.

	Coefficients	Standard Error	t Stat	P-Value
β_0	0.000849726	0.000575377	1.476817991	0.139961341
β_1	0.047924315	0.003039672	15.76627685	1.80938E-51

Table 4.5 Regression Statistics for the Net Money Flow.

	Coefficients	Standard Error	t Stat	P-Value
β_0	0.001178	0.000559	2.107982	0.035219
β_2	1.1649E-10	6.3130E-12	1.8453E+01	7.1651E-68

Table 4.6 Regression Statistics for the Normalized OFI.

	Coefficients	Standard Error	t Stat	P-Value
β_0	0.001451154	0.000573737	2.529299341	0.011543896
β_3	0.051694123	0.003241923	15.94551064	1.64269E-52

Table 4.7 Regression Statistics for the VWAP Skew.

	Coefficients	Standard Error	t Stat	P-Value
β_0	0.000767312	0.000591035	1.298250049	0.194426669
β_4	2.859042491	0.219584562	13.0202345	1.45923E-36

Table 4.8 displays the upper and lower confidence intervals for the two-variable models. There is only one intercept for the two-variable models that is statistically significant on a 5% level, and it is the N.M.F. model. The slope coefficients for the four models are all statistically significant on a 1% level.

Table 4.8 Confidence Intervals for the Two-Variable Models.

	Norm. N.V.		Net Money Flow		Norm. OFI		VWAP Skew	
	β_0	β_1	β_0	β_2	β_0	β_3	β_0	β_4
<i>Lower 95%</i>	-0.00028	0.04196	8.2E-05	1.0411E-10	0.00033	0.04533	-0.0004	2.4283
<i>Upper 95%</i>	0.00198	0.05389	0.00227	1.2888E-10	0.00258	0.05805	0.0019	3.2898
<i>Lower 99%</i>	-0.00063	0.04008	-0.00026	1.0021E-10	-0.00003	0.04333	-0.0008	2.2926
<i>Upper 99%</i>	0.00233	0.05577	0.00262	1.3278E-10	0.00293	0.06006	0.0023	3.4255

Figure 4.1 – 4.4 on the next two pages present the regression plot for the four regression models; see equation (14) to (17).

Figure 4.1 Regression Plot of the Normalized Net Volume.

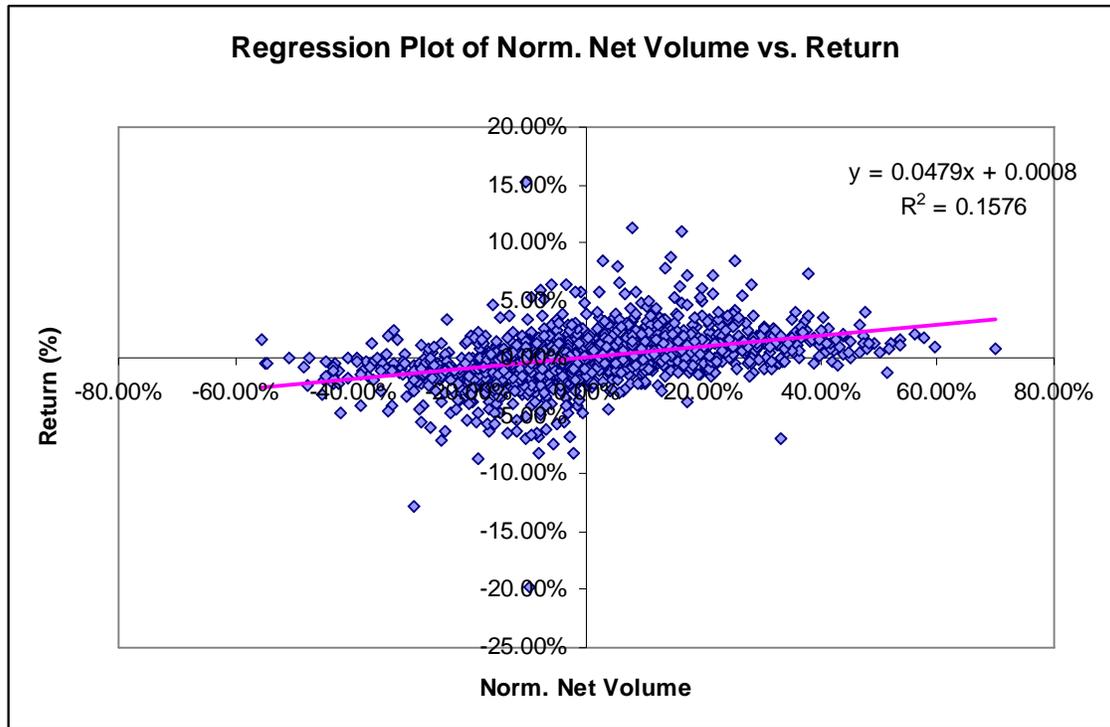


Figure 4.2 Regression Plot of the Net Money Flow.

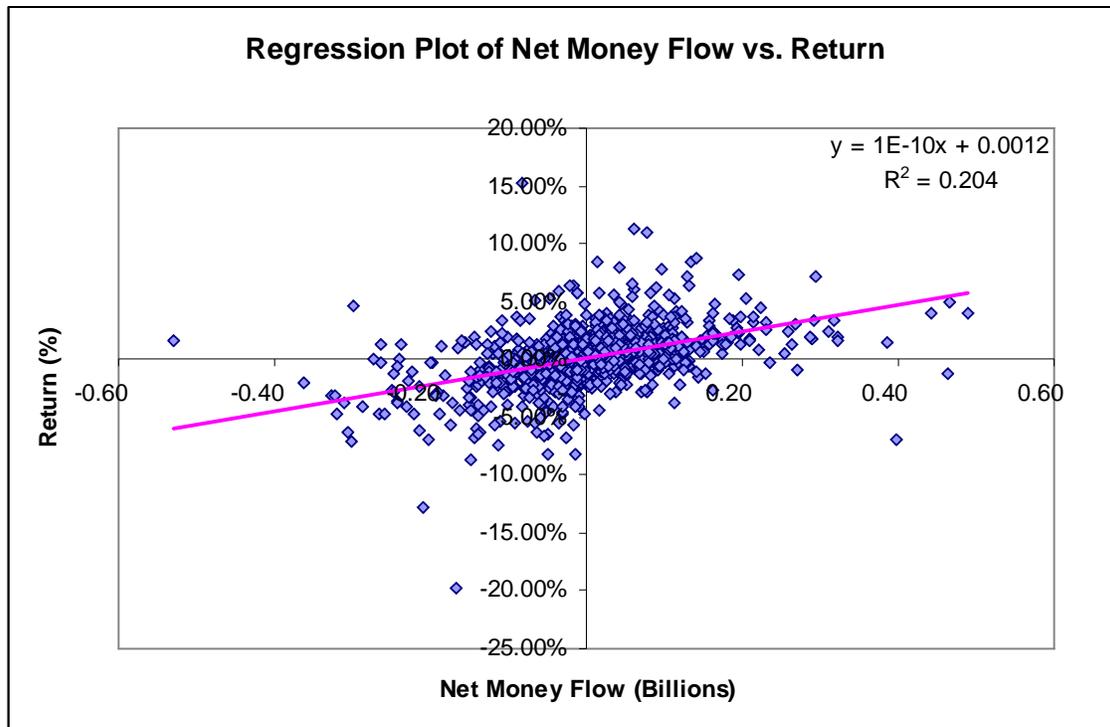


Figure 4.3 Regression Plot of the Normalized OFI.

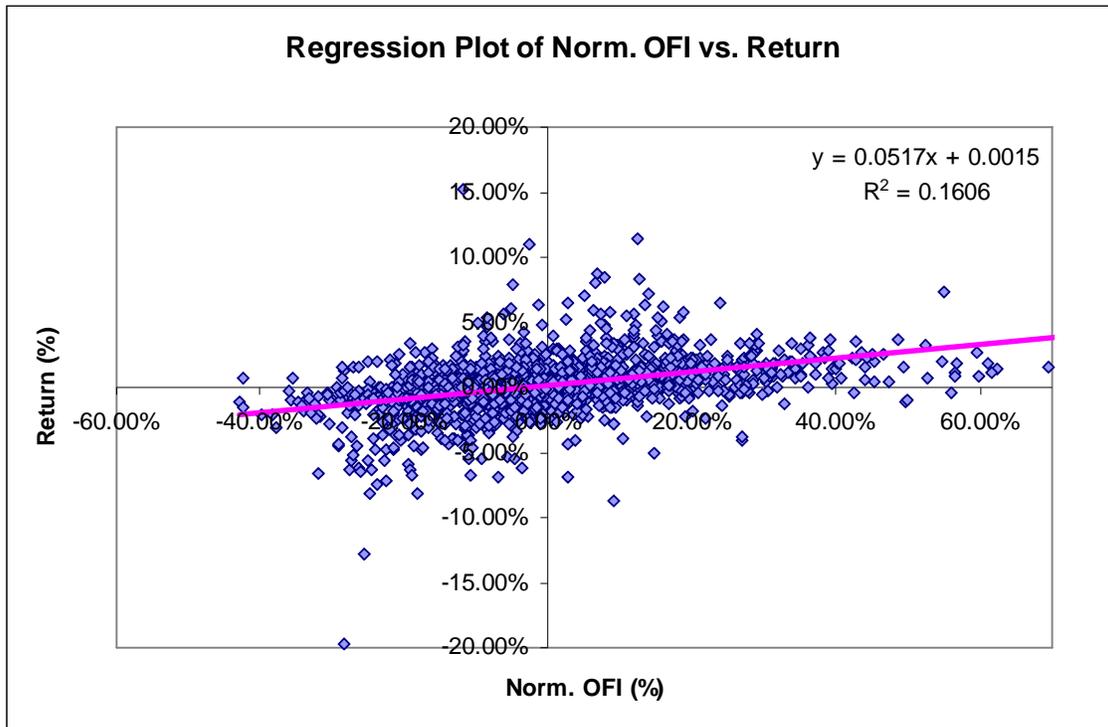
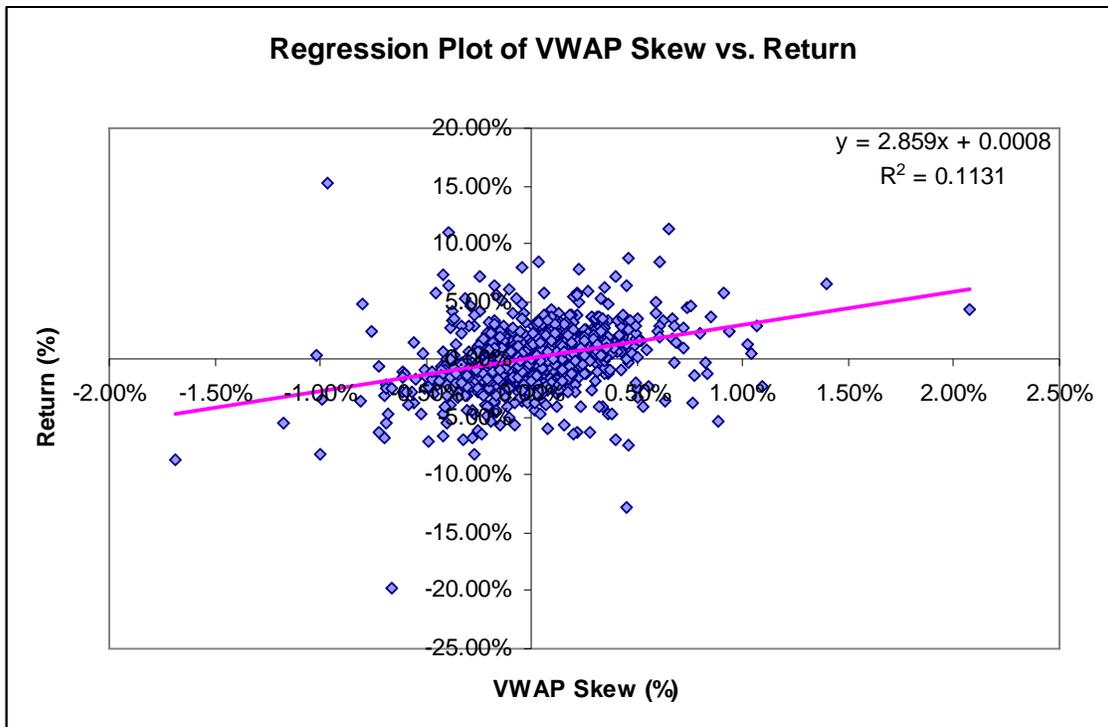


Figure 4.4 Regression Plot of the VWAP Skew.



Since the two variables, Norm. N.V. and N.M.F., are collinear, i.e. there is an exact linear relationship between them ($X_1 = \lambda \cdot X_2$), only one of them can be used in the multiple regression. The collinearity between the two explanatory variables is described by the equation below.

$$\text{Normalized Net Volume}_t = \frac{1}{P_{mid} \cdot (\text{Buy Volume}_t + \text{Sell Volume}_t)} \cdot \text{Net Money Flow}_t \quad (18)$$

In the multiple regression model, Norm. N.V. will be discarded as an explanatory variable because N.M.F. explains the dependent variable better. Another reason why N.M.F. probably is a better explanatory variable than Norm. N.V. is because the former takes into account the price level a stock is trading at. The number of shares traded is not as important as the money flowing in and out of the stock. A low-priced stock can have a high turnover measured in shares traded, but not necessarily if it is measured in money (share price \times volume). Hence, the four-variable regression model used is described by equation (19).

$$\hat{R}_t = \beta_0 + \beta_5 \cdot X_{1,t} + \beta_6 \cdot X_{2,t} + \beta_7 \cdot X_{3,t} \quad (19)$$

where

$X_{1,t}$ = Net Money Flow for day t .

$X_{2,t}$ = Normalized OFI for day t .

$X_{3,t}$ = VWAP Skew for day t .

To find the parameters ($\beta_0, \beta_5, \beta_6, \beta_7$) they are estimated with the method of Ordinary Least Squares (OLS), i.e. *Minimize* $\sum e_t^2 = \sum (R_t - \hat{R}_t)^2$. The estimated regression space is displayed below, and all parameters are significant at a 1% significance level, except for the intercept which is only significant at a 7% level. The confidence intervals for the four-variable model are presented in Table 4.5.

$$\begin{aligned} \hat{R}_t &= 0.000966108 + 7.92619E-11 \cdot X_{1,t} + 0.028259563 \cdot X_{2,t} + 1.815595905 \cdot X_{3,t} \\ \text{se} &= (0.000526407) + (6.64943E-12) + (0.003322883) + (0.203321302) \\ t &= (1.835287453) \quad (11.92010182) \quad (8.504531074) \quad (8.929688547) \end{aligned}$$

The confidence intervals for the four-variable regression model show that the three slope coefficients are all significantly different from 0 with 1% significance level, but the intercept is however not (not even with a 5% level). See Table 4.9 below.

Table 4.9 Confidence Intervals for the Four-Variable Model.

	β_0	β_5	β_6	β_7
Lower 95%	-6.6572E-05	6.62173E-11	0.021741	1.416730
Upper 95%	1.9988E-03	9.23064E-11	0.034778	2.214462
Lower 99%	-3.9178E-04	6.21094E-11	0.019688	1.291119
Upper 99%	2.3240E-03	9.64144E-11	0.036831	2.340072

Table 4.10 presents the regression statistics for the four-variable model, with a higher R^2 -value than the two-variable models.

Table 4.10 Regression Statistics.

<i>Regression Statistics</i>	
Multiple R	0.546264
R-Squared	0.298404
Adjusted R-Squared	0.296818
Standard Error	0.019148
Observations	1331

To test the joint hypothesis $H_0 : \beta_5, \beta_6, \beta_7 = 0$, I will use a technique known as Analysis of Variance (ANOVA). This study decomposes the total sum of squares (TSS) into two components; explained sum of squares (ESS) and residual sum of squares (RSS). Table 4.11 presents the calculations used, and Table 4.12 shows the analysis of the variance for the estimated regression model.

Table 4.11 ANOVA Table in Terms of R-squared.

	<i>d.f.</i>	<i>Sum of Squares (SS)</i>	<i>Mean Sum of Squares (MSS)</i>
Regression (ESS)	$k - 1$	$R^2(\sum y_i^2)$	$R^2(\sum y_i^2)/(k - 1)$
Residual (RSS)	$n - k$	$(1 - R^2)(\sum y_i^2)$	$(1 - R^2)(\sum y_i^2)/(n - k)$
Total (TSS)	$n - 1$	$\sum y_i^2$	

Table 4.12 shows the analysis of variance for the multiple regression model, and since the F-value is extremely high, it is the evidence against the null hypothesis that the three explanatory variables (β_5, β_6 and β_7) have no effect on the return. The variables are not only individually statistically significant, but collectively as well.

Table 4.12 ANOVA Table for the Estimated Multiple Regression Model.

	d.f.	SS	MSS	F	Significance F
Regression	3	0.206935	0.068978	188.134309	1.2034E-101
Residual	1327	0.486537	0.000367		
Total	1330	0.693471			

The accumulated money flow has been calculated as the sum of all net money flows from t_0 till time t , see the equation below.

$$Accumulated\ Money\ Flow_t = \sum_{i=0}^t Net\ Money\ Flow_i \quad (20)$$

The regression model for the Accumulated Money Flow (A.M.F.) is presented in the equation below.

$$\hat{R}_t = \beta_0 + \beta_8 \cdot A.M.F._t \quad (21)$$

I will test the null hypothesis $H_0 : \beta_8 = 0$ against an alternative hypothesis $H_0 : \beta_8 \neq 0$. The regression statistics for the accumulated money flow is displayed in Table 4.13 below, and note the high R^2 -value.

Table 4.13 Regression Statistics for the Accumulated Money Flow.

Regression Statistics	
Multiple R	0.760104
R-Squared	0.577758
Adjusted R-Squared	0.577441
Standard Error	26.656217
Observations	1331

Both coefficients in equation (21) are significantly different from 0 with a 1% significance level. The regression coefficients and its confidence intervals are presented in Table 4.14 below. It is quite obvious that the intercept should be different from 0, the stock price is always greater than 0 (given the company is not bankrupt).

Table 4.14 The Regression Coefficients and its Confidence Intervals.

	Coefficients	Standard Error	t Stat	P-Value	Lower 95%	Upper 95%	Lower 99%	Upper 99%
β_0	64.8298	1.4114	45.9333	1.20E-276	62.0610	67.5986	61.1891	68.4706
β_8	1.2503E-08	2.9319E-10	42.6437	4.41E-251	1.1928E-08	1.3078E-08	1.1746E-08	1.3259E-08

The regression plot of the accumulated money flow is presented in Figure 4.5.

Figure 4.5 Regression Plot of Accumulated Money Flow.

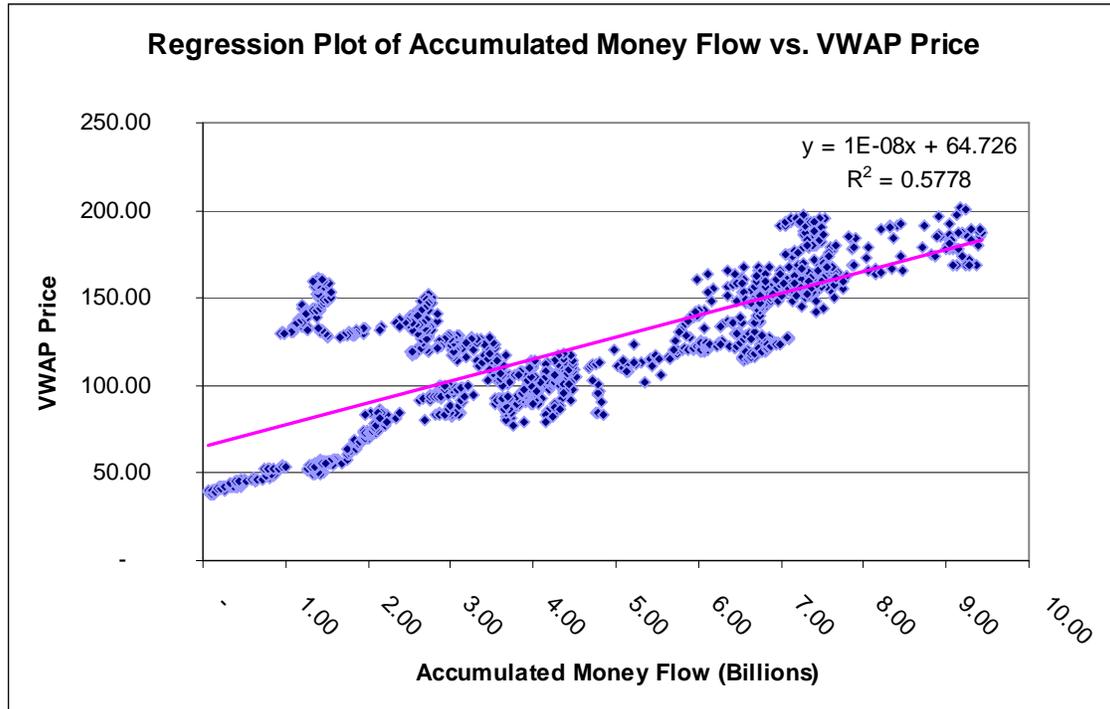
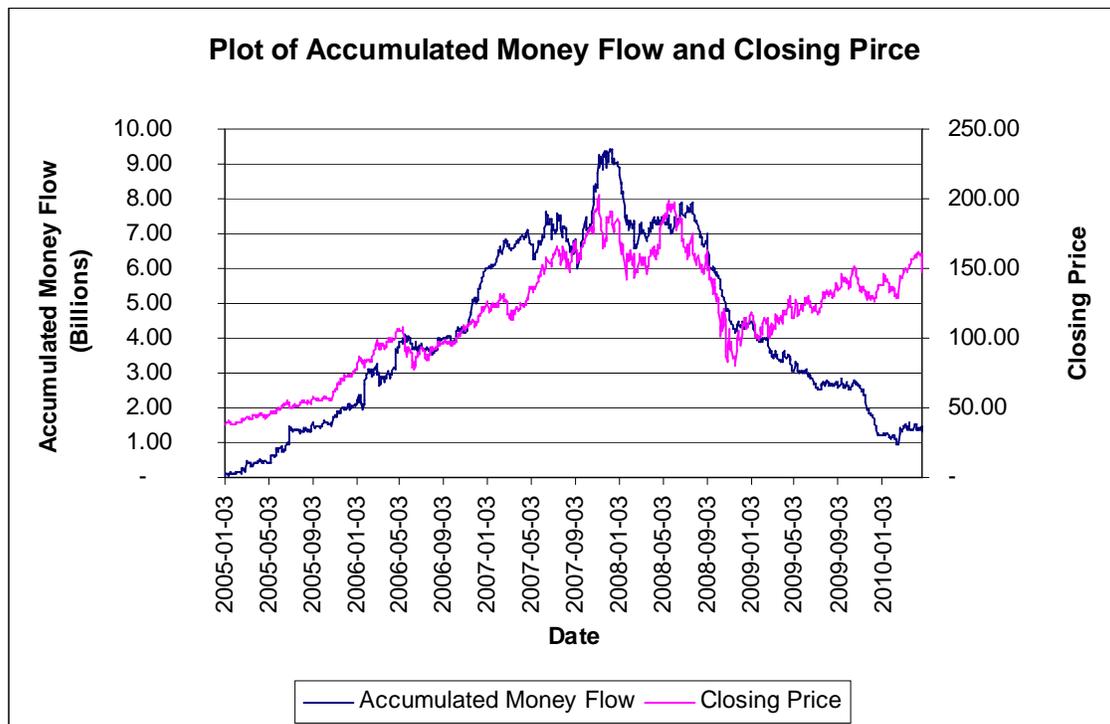


Figure 4.6 Comparison between Accumulated Money Flow and Closing Price in ABB Ltd.



As can be seen by Figure 4.6 above, there was a strong correlation between accumulated money flow and the stock price until late 2008. After that point in time, the stock price has risen, while the accumulated money flow has continued to decrease. There may be many explanations for this behavior; one explanation could be that the advent of high-frequency trading algorithms, where a tactic known as “pinging the book” can create misleading money flow. Brown (2010, p.113) describes “pinging the book” as a tactical strategy used by high-frequency trading algorithms to gather information about key price levels at which others are willing to trade. The algorithm submits orders to an ECN and if they are not filled within 60-80 milliseconds, they will be cancelled. Another reason why the accumulated money flow and the stock price started to diverge may be due to the tick-size changes made during 2009.

4.4 Testing a Trading Model with the Measures Studied

The trading model uses the three measures as indicators if one should be long or short in the stock. If all three measures, N.M.F., Norm. OFI, and VWAP Skew, are all positive (negative) the trading model will go long (sell short) at the close of the day, and sell (buy) at the end of the next day. For a mathematical definition, see equation (22) and Table 4.15 below.

$$\text{Trading Signal} = \sum_{i=1}^3 D_i \quad (22)$$

where D_i is a dummy variable and the values the dummy variable takes on for each i is display in Table 4.11 below.

Table 4.15 Definition of the Dummy Variable used in the Trading Model.

	MEASURE	MEASURE > 0	MEASURE < 0
1	Net Money Flow	1	-1
2	Normalized OFI	1	-1
3	VWAP Skew	1	-1

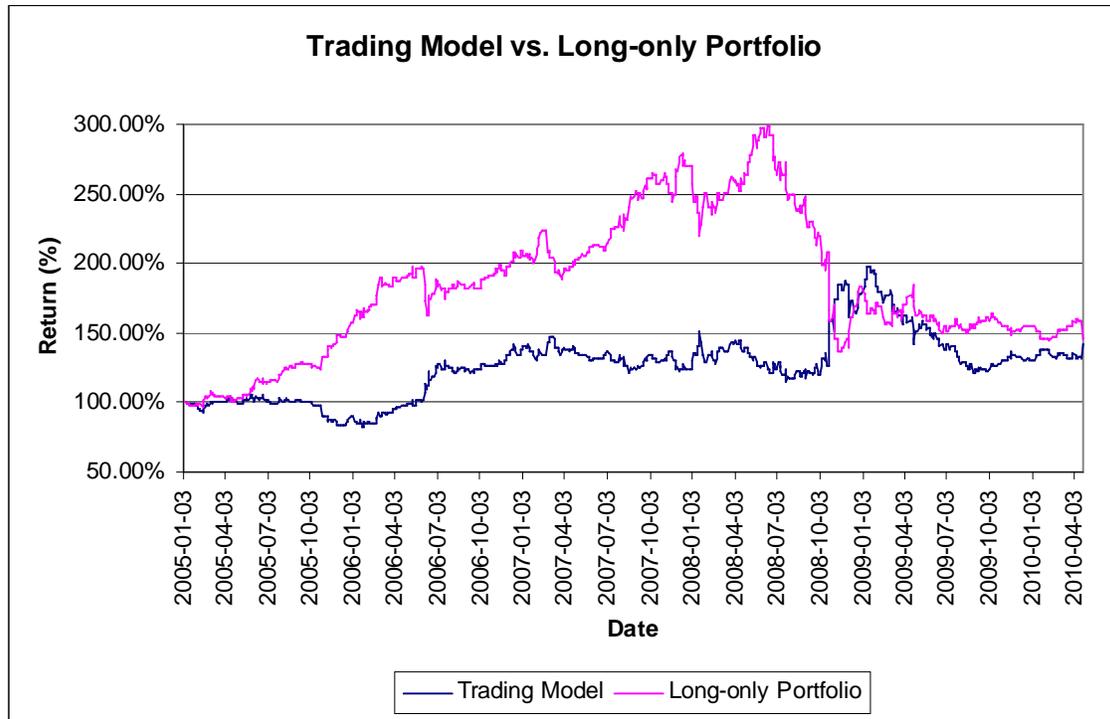
The trading model will take a long position if $\sum_{i=1}^3 D_i = 3$, and a short position if $\sum_{i=1}^3 D_i = -3$.

If the sum is either 2 or -2, the signal will be neglected, and the model will wait for the next signal, for this reason the model neglected 751 signals during the sample period. Out of the 1,330⁵ trading days in the sample, the model received 579 trading signals, i.e. when all

⁵ One trading day is lost due to the one-step prediction.

indicators were fully synchronized. The trading model was therefore only trading 43.53% of the time, with 273 long positions and 306 short positions. There were a total of 651 winning trading opportunities, thus the trading model took advantage of 88.94% of them. The return of the trading model is displayed in Figure 4.7 below, and one would be slightly better off if one had invested in the long-only portfolio (100% invested in ABB Ltd).

Figure 4.7 The Return of the Trading Model vs. the Long-only Portfolio.



Invested in the long-only portfolio during the sample period would have yielded a return of 45.88% compared to the trading model's 41.68%. One way to measure the risk-adjusted return is with the Sharpe ratio, see equation (23) on the next page. The Sharpe ratio for the trading model, however, is almost twice as good as the long-only portfolio, with a ratio of 1.25 compared to 0.67. There is another thing to take into account, i.e. the average daily return, which is significantly lower for the long-only portfolio with a daily average return (during days of risk-taking)⁶ of 0.028% compared to 0.060%. Since the trading model is only invested 43.53% of the time, the rest of the time it is invested in the risk-free rate.

As the risk-free interest rate, I have used the Swedish repurchase agreement (repo) rate, with a compounded interest rate of approximately 12.83%. See Table 4.16 to see the yearly repo rate

⁶ Daily Average Return = $\sqrt[n]{\prod_{t=1}^n (1 + R_t)}$ during n days of risk-taking.

in Sweden, however note that the last row only displays the repo rate for the first quarter of 2010 (Riksbank, 2010).

Table 4.16 The Swedish Repo Rate.

Year	Mean	Min	Max
2005	1.7312%	1.50%	2.00%
2006	2.2052%	1.50%	3.00%
2007	3.4590%	3.00%	4.00%
2008	4.1429%	2.00%	4.75%
2009	0.6534%	0.25%	2.00%
2010	0.2500%	0.25%	0.25%

(Riksbank, 2010)

Table 4.17 displays all factors affecting the Sharpe ratio.

Table 4.17 Trading Model vs. Long-only Portfolio.

	Trading Model	Long-only Portfolio
Return	41.68%	45.88%
Risk-Free Rate	12.83%	12.83%
Standard Deviation	23.04%	48.99%
Sharpe Ratio	1.2521	0.6746

The Sharpe ratio in Table 4.17 has been calculated using the equation below, where $E[R]$ is the expected return, R_f is the risk-free rate, and σ is the standard deviation of the return.

$$\text{Sharpe Ratio} = \frac{E[R] - R_f}{\sigma} \quad (23)$$

Since the trading model always trades at the close of the day, slippage will not exist in the same way as when on trades during the day. This is because of the closing-auction, where one can see how the market absorbs one's volume, and the volume can be sized accordingly to minimize the market impact. One way to measure the implementation shortfall for this trading strategy is to see how its hits and misses are distributed. To measure the trading model's ability to exploit the buy and sell signals, Aldridge (2010, p.222) proposes a method called Trading Strategy Accuracy (TSA). The method not only evaluates the trading opportunities realized by the trading model, but the trading opportunities that the model missed. A hit is defined as a correct forecast (resulting in a gain) either on the long-side or the short-side. A miss is defined as an incorrect forecast (resulting in a loss) either on the long-side or the short-side. Out of the 579 trading signals received, 304 were hits, i.e. the forecast was accurate 52.50% of the time. Table 4.18 displays the TSA of the trading model, and one can conclude that the bullish forecasts are better than the bearish.

Table 4.18 Distribution of Forecast Hits and Misses.

	Hits	Misses	Total	% Hits	% Misses
Long Positions	153	120	273	56.04%	43.96%
Short Positions	151	155	306	49.35%	50.65%
Total	304	275	579	52.50%	47.50%

To see characteristics of the tail distribution among the hits and misses, Table 4.19 presents the five most extreme returns. The best accurate trade had a return of 8.36% on the long-side and 19.78% on the short-side, respectively. On the other hand, the worst trade lost 7.12% on the long-side and 8.40% on the short-side, respectively.

Table 4.19 The Fifteen Most Extreme Returns for Hits and Misses.

	Hits		Misses	
	Up	Down	Up	Down
1	8.36%	-19.78%	-7.12%	8.40%
2	6.54%	-12.83%	-6.50%	7.27%
3	4.20%	-7.42%	-4.51%	6.42%
4	4.02%	-6.89%	-4.38%	6.31%
5	3.71%	-6.75%	-4.15%	5.78%
6	3.65%	-6.56%	-4.02%	5.68%
7	3.62%	-6.23%	-4.01%	5.66%
8	3.56%	-5.60%	-3.63%	5.29%
9	3.47%	-5.51%	-3.50%	4.89%
10	3.45%	-4.76%	-3.33%	4.58%
11	3.43%	-4.76%	-3.20%	4.40%
12	3.41%	-4.71%	-2.99%	4.37%
13	3.32%	-4.67%	-2.86%	4.33%
14	3.24%	-4.67%	-2.82%	3.86%
15	3.22%	-4.66%	-2.82%	3.85%

Dowd (2005, p.35) states that the expected shortfall is a good candidate to measure tail-risk. The expected shortfall takes the average of the N^{th} largest losses, and for the 30 largest gains and losses the average gain and loss is 5.57% and -4.70%, respectively. Table 4.20 displays the 97.74 percentile (gains) and 2.26 percentile (losses), respectively.

Table 4.20 The 30 Largest Gains and Losses.

	Gain	Loss		Gain	Loss
1	19.78%	-8.40%	16	4.67%	-4.37%
2	12.83%	-7.27%	17	4.66%	-4.33%
3	8.36%	-7.12%	18	4.20%	-4.15%
4	7.42%	-6.50%	19	4.02%	-4.02%
5	6.89%	-6.42%	20	3.71%	-4.01%
6	6.75%	-6.31%	21	3.65%	-3.86%
7	6.56%	-5.78%	22	3.62%	-3.85%
8	6.54%	-5.68%	23	3.56%	-3.63%
9	6.23%	-5.66%	24	3.47%	-3.50%
10	5.60%	-5.29%	25	3.45%	-3.33%
11	5.51%	-4.89%	26	3.43%	-3.20%
12	4.76%	-4.58%	27	3.41%	-2.99%
13	4.76%	-4.51%	28	3.32%	-2.86%
14	4.71%	-4.40%	29	3.24%	-2.82%
15	4.67%	-4.38%	30	3.22%	-2.82%
Average			5.57% -4.70%		

Hence, the trading model has an edge and it could (with some refinements) be a successful trading strategy. If the trading model could anticipate the sign of the indicators from the first hours of trading and then trade accordingly, the model would have had a significantly better performance. Instead of 52.50% hits it would have had 87.22% hits, this because the measures describe the concurrent day's return way better the next coming. The measures will converge to its final value as the time passes by, and I strongly believe that studying the first hours of trading and calculate the measures with the available data, one could get a reasonable good idea of if it would be a buy or a sell for the day. The trading model will off course not grasp the full extent of the daily return, but hopefully a great deal. Table 4.21 presents how good each measure is to predict the direction of the next day.

Table 4.21 Number of Correct Predictions for each Measure.

	Today		Tomorrow	
Net Money Flow	937	70.40%	793	59.62%
Normalized OFI	867	65.14%	651	48.95%
VWAP Skew	884	66.42%	676	50.83%

Note that the second and third column describe how many days each measure were synchronized with the same day's return. These predictions are not used, because they are ex post predictions. The fourth and fifth column describe the actual predictions for the measures individually, but remember I only received a trading signal when all measure predicted the same scenario.

5 CONCLUSION

In this chapter, I will summarize the result from the previous chapter and draw conclusions regarding the findings. Finally, I will present suggestions for further studies.

5.1 Discussion of Findings

In the beginning of this thesis I presented my main question and they were:

- Does money flow affect stock prices, and is it a good predictor of future price movements?
- Is there a relationship between high-frequency trading strategies and the observed return in the specific stock?
- Is the aggressiveness of buyers versus sellers affecting the return, and is it a good predictor of future price movements?

To start with the first question, money flow showed a significant relationship with the return during the same period. Out of the four measures presented in chapter 4, net money flow described the return the best. The accumulated money flow showed a significant relationship with the stock price, with a correlation coefficient of 0.76 and this even though it was strongly negatively correlated the last 1½ year. The correlation between accumulated money flow and stock price was 0.9652 during 2005 to the end of 2008, and from 2009 and onwards the correlation became -0.7542. This drastic change in correlation must imply that something essential has changed. The accumulated money flow described the stock price as good in the bull market as in the bear market, so it should not be a seasonal effect. I presented some reasonable explanations in chapter 4; one of them was the new era of high-frequency trading algorithms. These algorithms have an arsenal of strategies, attempting to learn important information about other market participants. My findings are consistent with Bennet and Sias (2001) conclusion, that money flow does measure excess demand and supply, and thus highly correlated to stock return.

To answer the second question, one of the high-frequency trading strategies (presented in section 3.2.2), the VWAP Skew showed a significant relationship with concurrent return on a 1 % significance level. The VWAP Skew did not, however, explain the return as good as the other measures studied. Out of the three measures, the Norm. OFI was the worst predictor with only 48.95% correct directional predictions. The VWAP Skew did slightly better than a random prediction, with 50.83% correct directional predictions. The Net Money Flow was the

best predictor; with an accuracy of 59.62% (almost 6 out of 10 trades would be winners). The trading model with the combined measures had an accuracy of 52.50%, which is better than a random forecast. It is not only the accuracy of the predictor that is vital to succeed with trading in the markets, the risk-reward ratio plays an essential role as well. Since this trading model does not use stop-loss orders and profit targets but exits the position on the close of the next coming day, the risk-reward ratio is impossible to calculate ex-ante. The tail-distribution for the 30 most extreme returns for winners and losers, presented in Table 4.20, shows that the average gain and loss is 5.13% and -4.70%, respectively. I think the expected shortfall is a good estimate for the risk-reward, which in this case is approximately $\frac{4.70}{5.57} = 0.8438$. A risk-reward ratio that is less than 1 and an accuracy of more than 50% is a very good start for a successful trading model.

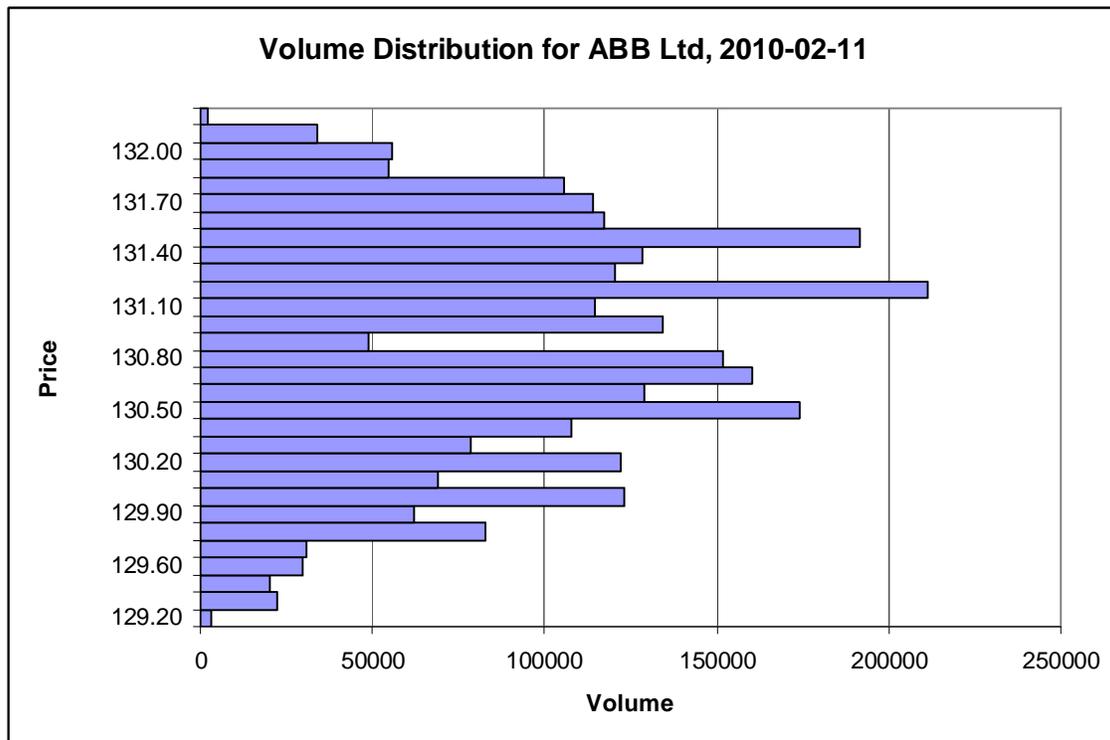
The last question I asked myself was if these measures could predict future price movements. Since all three measures proved to be highly significant, I decided to test a trading model based on the conclusion reached by Bennet and Sias (2001). Their conclusion was that money flow exhibits strong positive serial correlation, i.e. high money flow today, forecasts high money flow in the future. The trading model combines the three measures studied in this thesis as indicators to predict the next coming daily return. The model is not as good to predict the next day's return, but it is better than a random prediction. One reason is because there are so many external factors affecting the following day's return, e.g. how the U.S. and Asia are trading, and news during non-market hours. The five year return for the trading model was almost the same as for the long-only portfolio. The risk-adjusted return, however, was approximately twice as good for the trading model as for the long-only portfolio. Even though this trading model was fairly simple, I believe small adjustments could enhance the performance.

The trading model shows that it has an edge in the market, and thus the market is not as efficient as academics believe. I totally agree with Grossman and Stiglitz (1980) when they claim it is impossible for markets to stay efficient for a sustainable period of time. If markets were perfectly efficient, there would be no incentives for market participants to gather information and build trading models. Since information is costly, prices cannot perfectly reflect all available information, because if it did, there would be no motivation for those who spent resources to obtain it.

5.2 Future Research Suggestions

There is no official research, as far as I know, about daily volume distributions, i.e. how volume is distributed over different prices during the day. Figure 5.1 displays how the volume distribution looked like on February 11, 2010. The distribution is skewed slightly on the upside, which may imply that the buyers are willing to pay a higher price for the stock. One might wonder why the buyers are so eager to pay a high price for the stock. One reason might be that they know something that is not priced in to the stock price (information asymmetry); hence they believe it is undervalued.

Figure 5.1 Volume Distribution for ABB Ltd.



Another research suggestion is to develop a model that can predict the return for the day using e.g. the Norm. OFI, N.M.F. and the VWAP Skew for the first two hours of trading. I believe the relationship for the first hours and the return of the day could be stronger than the overnight return studied in this thesis. 24-hour predictions are more vulnerable to exogenous shocks, i.e. market moves which are not explainable by models using market data. These exogenous shocks are usually driven by information that is not internal to the market. Avoiding overnight exposure can be the difference between a winning strategy and a losing one.

Another suggestion is to analyze the TWAP-price, and investigate if it is skewed in the same direction as the VWAP-price, and if so, is it possible to predict price movements. Is the TVWAP-price highly correlated to the VWAP-price?

REFERENCES

A. Printed Sources

- [1] Aldridge, I. (2010). *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. Hoboken, NJ: John Wiley & Sons, Inc.
- [2] Bennett, J. A. & Sias, R. W. (2001). Can Money Flows Predict Stock Returns?. *Financial Analysts Journal*. **57**:64-78.
- [3] Borsellino, L. J. (2001). *The Day Trader's Course*. Hoboken, NJ: John Wiley & Sons, Inc.
- [4] Brown, B. R. (2010). *Chasing the Same Signals: How Black-Box Trading Influences Stock Markets from Wall Street to Shanghai*. Clementi Loop, Singapore: John Wiley & Sons (Asia) Pte. Ltd.
- [5] Bryman, A. & Bell, E. (2007). *Business Research Methods* (2nd edn.). New York, NY: Oxford University Press, Inc.
- [6] Chan, E. P. (2009). *Quantitative Trading: How to Build Your Own Algorithmic Trading Business*. Hoboken, NJ: John Wiley & Sons, Inc.
- [7] Cohen, L., Manion, L. & Morrison, K. R. B. (2007). *Research Methods in Education*. New York, NY: Routledge.
- [8] Coldwell, D. & Herbst, F. (2004). *Business Research*. Cape Town, South Africa: Juta & Co Ltd.
- [9] Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (2nd edn.). Thousand Oaks, CA: SAGE Publications, Inc.
- [10] Dacorogna, M. M., Gençay, R., Müller, U., Olsen, R. B. & Pictet, O. V. (2001). *An Introduction to High-Frequency Finance*. San Diego, CA: Academic Press.
- [11] Dowd, K. (2005). *Measuring Market Risk* (2nd edn.). Chichester, England: John Wiley & Sons, Ltd.
- [12] Elder, Dr. A (2002). *Come Into My Trading Room: A Complete Guide to Trading*. New York, NY: John Wiley & Sons, Inc.
- [13] Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A. & Focardi, S. M. (2007). *Robust Portfolio Optimization and Management*. Hoboken, NJ: John Wiley & Sons, Inc.
- [14] Fadiman, M. & Klein, J. (2004). *The Evolution of Trading: How Technology and Governance are Changing Finance In the 21st Century*. New York, NY: Technology & Marketing Ventures, Inc.
- [15] Foster, O. D. (1965/2005). *Ticker Technique: The Art of Tape Reading*. Greenville, SC: Traders Press, Inc.
- [16] Gavridis, M. (1998). Modelling with High Frequency Data: A Growing Interest for Financial Economists and Fund Managers. In *Nonlinear Modelling of High Frequency Financial Time Series*. Edited by Dunis, C. & Zhou, B. Chichester, England: John Wiley & Sons, Ltd.
- [17] Ghauri, P. & Grønhaug, K. (2002). *Research Methods in Business Studies: A Practical Guide* (2nd edn.). Harlow, England: Pearson Education Limited.
- [18] Graifer, V. & Schumacher, C. (2004). *Techniques of Tape Reading*. New York, NY: McGraw-Hill Companies, Inc.

- [19] Grossman, S. J. & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*. **June 70**:393-408.
- [20] Gujarati, D. N. (2006). *Essentials of Econometrics* (3rd edn.). New York, NY: McGraw-Hill Companies, Inc.
- [21] Harris, L. & Hasbrouck, J. (1996). Market vs. Limit Orders: The SuperDOT Evidence on Order Submission Strategy. *Journal of Financial and Quantitative Analysis*. **31**:213-231.
- [22] Harris, L. (2003). *Trading and Exchanges: Market Microstructure for Practitioners*. New York, NY: Oxford University Press, Inc.
- [23] Heires, K. (2009). Use Machines To Make Extreme Profits From Extreme Frequency. *Securities Industry News*. **July 6**:15.
- [24] Kim, K. (2007). *Electronic and Algorithmic Trading Technology: The Complete Guide*. Burlington, MA: Academic Press.
- [25] Kissell, R. & Glantz, M. (2003). *Optimal Trading Strategies: Quantitative Approaches for Managing Market Impact and Trading Risk*. New York, NY: AMACOM.
- [26] Kumar, R. (2005). *Research Methodology* (2nd edn.). Thousand Oaks, CA: SAGE Publications, Inc.
- [27] Link, M. (2003). *High Probability Trading*. New York, NY: McGraw-Hill Companies, Inc.
- [28] Lukeman, J. (2000). *The Market Maker's Edge*. New York, NY: McGraw-Hill.
- [29] Madsen, H. (2008). *Time Series Analysis*. Boca Ration, FL: Chapman & Hall/CRC.
- [30] Narang, R. K. (2009). *Inside the Black Box: The Simple Truth about Quantitative Trading*. Hoboken, NJ: John Wiley & Sons, Inc.
- [31] Neill, H. B. (1931/1970). *Tape Reading & Market Tactics*. New York, NY: B.C. Forbes Publishing Company.
- [32] Perold, A. F. (1998). The Implementation Shortfall: Paper versus Reality. In *Streetwise: the Best of the Journal of Portfolio Management*. Edited by Bernstein, P. L. & Fabozzi, F. J. Princeton, NJ: Princeton University Press.
- [33] Pole, A. (2007). *Statistical Arbitrage: Algorithmic Trading Insights and Techniques*. Hoboken, NJ: John Wiley & Sons, Inc.
- [34] Schubert, B. (2009). *Evolution of Algorithmic Trading and Challenges of the Future*. Saarbrücken, Germany: VDM Verlag Dr. Müller Aktiengesellschaft & Co. KG.
- [35] Steidlmayer, J. P. & Hawkins, S. B. (2003). *Steidlmayer on Markets: Trading with Market Profile* (2nd edn.). Hoboken, NJ: John Wiley & Sons, Inc.
- [36] Taleb, N. N. (2004). *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets*. New York, NY: Random House, Inc.
- [37] Taleb, N. N. (2008). *The Black Swan: The Impact of the Highly Improbable*. New York, NY: Random House, Inc.
- [38] Tape, R. (1910/2003). *Studies in Tape Reading: A 1910 Classic on Tape Reading & Stock Market Tactics* (5th edn.). New York, NY: Traders Press, Inc.

B. Electronic Sources

- [1] BNY ConvergeEx Group, LLC (2009). *IQx: Dynamic Trade Optimizer*. Available from:
www.bnyconverge.com/ASSETS/.../IQx%20Overview.pdf (Accessed 20 February, 2010).
- [2] Bukey, D. (2005). J. Peter Steidlmayer: In Step with the Markets. *Active Trader*, 9:38-43. Available from:
<http://www.elitetrader.com/vb/attachment.php?s=70d9b4ed5dd202fbe14f769af0f43c03&postid=921848> (Accessed 27 February, 2010).
- [3] Collins, G. (2008). *Fund Managers Switching to Algorithmic Trading*. Available from:
<http://kb.cnblogs.com/a/1111906/> (Accessed 13 March, 2010).
- [4] Hale, S. H. (1995). The Ticker Tape: Yesterday, Today and Tomorrow. *MTA Journal*, 45:10-22. New York, NY. Available from:
<http://www.mta.org/eweb/docs/Issues/45%20-%201995%20Winter.pdf> (Accessed 23 April, 2010).
- [5] Kakade, S. M., Kearns, M., Mansour, Y. & Ortiz, L. E. (2004). Competitive Algorithms for VWAP and Limit Order Trading. *EC'04*, May 17-20. New York, NY. Available from:
<http://www.cis.upenn.edu/~mkearns/papers/vwap.pdf> (Accessed 6 June, 2009).
- [6] NYSE Euronext, Inc. (n.d.). *The Peg Orders*. Available from:
<http://www.euronext.com/fic/000/041/609/416094.pdf> (Accessed 21 March, 2010).
- [7] O'Hara, M. (2002). *Liquidity and Price Discovery*. Available from:
<http://www.afajof.org/afa/all/presadd2003.pdf> (Accessed 6 June, 2009).
- [8] Riksbank (2010). *Repo rate, deposit and lending rate*. Available from:
<http://www.riksbank.com/templates/stat.aspx?id=17184> (Accessed 20 May, 2010).

APPENDICES

Appendix I

Table A.1 below displays the trades excluded from the data sample to prevent misleading results. All trades below are all outside the daily trading range, i.e. the high-low range.

Table A.0.1 Trades Excluded from the Data Sample due to Prices outside the Daily Trading Range.

TIME	PRICE	BUYER	SELLER	VOLUME	TRADE NO.
2005-06-30 09:13:09	165.00	NDS	NDS	18,000	725
2006-01-10 10:07:10	73.75	DRS	DRS	22,000	554
2006-02-20 09:01:49	80.00	FIP	FIP	8,000	81
2006-03-30 10:21:45	96.50	CAR	UBS	84443	450
2006-04-03 15:44:51	98.13	SHB	SHB	69,630	1,546
2006-05-11 11:03:03	180.00	DDB	DDB	30,000	325
2006-11-09 10:51:23	110.25	DDB	DDB	700	317
2006-12-22 15:19:43	27.85	SHB	SHB	25,000	444
2007-02-19 09:53:40	120.00	GLI	GLI	15,000	259
2007-03-06 09:06:38	116.50	DDB	DDB	2,000	144
2007-09-05 15:03:56	28.84	HBC	HBC	75,000	2,465
2007-10-23 11:47:11	173.00	MSI	MSI	300	782
2007-11-22 10:36:28	68.50	ENS	ENS	50,000	747
2008-03-10 09:09:00	157.50	SHB	SHB	92,370	103
2008-03-25 12:04:02	102.00	ENS	ENS	69,400	898
2008-09-09 09:25:12	92.00	SWB	SWB	10,000	313
2008-10-10 09:41:10	107.50	SHB	SHB	414,912	867
2008-10-13 09:33:32	101.75	SHB	SHB	414,720	652
2008-10-13 14:57:26	66.75	SWB	SWB	200	2,634
2008-10-30 11:27:03	93.10	SHB	SHB	283,575	1,597
2009-05-15 10:51:19	120.00	SHB	SHB	1,000	847
2009-07-21 15:29:47	29.75	ENS	ENS	20,000	1,174

Appendix II

This algorithm goes through every trade during the whole sample, and summarize the data into following group; date, daily VWAP, close, high, low, number of buyers and sellers, respectively. The code below is written in VBA.

```
Sub Trade_Summary()  
Workbooks("Summary").Worksheets("Summary III").Activate  
n = 1  
Application.ScreenUpdating = False  
For d = 2 To 7  
    For b = 2 To 13  
        my_sheet = Cells(b, 1).Value  
        my_book = Cells(1, d).Value  
        Workbooks("Summary").Worksheets("Summary II").Activate  
        n = Range(Cells(2000, 1), Cells(2000, 1)).End(xlUp).Row  
        n = n + 1  
        Workbooks(my_book).Worksheets(my_sheet).Activate  
        high = 0  
        low = 1000  
        day_close = 0  
        vol = 0  
        totvalue = 0  
        daily_vwap = 0  
        buycount = 0  
        sellcount = 0  
        lastcol = Range("GA1").End(xlToLeft).Column  
        For c = 1 To lastcol  
            Workbooks(my_book).Worksheets(my_sheet).Activate  
            day_date = Left(Cells(2, c + 5), 10)  
            lastrow = Range(Cells(50000, c), Cells(50000, c)).End(xlUp).Row  
            For r = 2 To lastrow  
                If Cells(r, c).Value > high Then  
                    high = Cells(r, c).Value  
                End If  
                If Cells(r, c).Value < low Then  
                    low = Cells(r, c).Value  
                End If  
                If Not Cells(r, c + 1).Value = Cells(r + 1, c + 1).Value Then  
                    buycount = buycount + 1  
                End If  
                If Not Cells(r, c + 2).Value = Cells(r + 1, c + 2).Value Then  
                    sellcount = sellcount + 1  
                End If  
                vol = vol + Cells(r, c + 3).Value  
                totvalue = totvalue + Cells(r, c).Value * Cells(r, c + 3).Value  
            Next r  
            daily_vwap = totvalue / vol  
            buycount = buycount - 1  
            sellcount = sellcount - 1  
        End For  
    End For  
End Sub
```

```

day_close = Cells(2, c).Value
Workbooks("Summary").Worksheets("Summary II").Activate
Sheets("Summary II").Cells(n, 1) = day_date
Sheets("Summary II").Cells(n, 2) = daily_vwap
Sheets("Summary II").Cells(n, 4) = day_close
Sheets("Summary II").Cells(n, 5) = high
Sheets("Summary II").Cells(n, 6) = low
Sheets("Summary II").Cells(n, 8) = buycount
Sheets("Summary II").Cells(n, 9) = sellcount
Workbooks(my_book).Worksheets(my_sheet).Activate
buycount = 0
sellcount = 0
n = n + 1
c = c + 6
low = 1000
high = 0
vol = 0
totvalue = 0
Next c
Workbooks("Summary").Worksheets("Summary III").Activate
Next b
Next d
Application.ScreenUpdating = True
End Sub

```

Appendix III

This algorithm calculates the opening price, the buy volume as well as the sell volume. The reason why I do not use the first trade of the day as the opening price is because a couple of trades from the day before tend to be reported before the opening bell. These prices may deviate significantly from the real opening price, and that is why the algorithm makes sure it takes the first trade that was executed 9 a.m.

```
Sub calc_buy_sell()
Workbooks("Summary").Worksheets("Summary III").Activate
n = 1
For f = 2 To 7
  For h = 2 To 13
    Workbooks("Summary").Worksheets("Summary III").Activate
    my_sheet = Cells(h, 1).Value
    my_book = Cells(1, f).Value
    Workbooks("Summary").Worksheets("Summary II").Activate
    n = Range(Cells(2000, 3), Cells(2000, 3)).End(xlUp).Row
    n = n + 1
    Workbooks(my_book).Worksheets(my_sheet).Activate
    a = 0
    b = 1
    c = 1
    d = 1
    e = 1
    bid = 0
    ask = 0
    buyvol = 0
    sellvol = 0
    orow = 0
    opening = 0
    Dim MyHour
    Dim MyTime
    lastcol = Range("GA1").End(xlToLeft).Column
    For c = 1 To lastcol
      Workbooks(my_book).Worksheets(my_sheet).Activate
      lastrow = Range(Cells(50000, c), Cells(50000, c)).End(xlUp).Row
      MyTime = Cells(lastrow, c + 5).Value
      MyHour = Hour(MyTime)
      If MyHour = 9 Then
        opening = Cells(lastrow, c).Value
        orow = Cells(lastrow, c).Row
      End If
      If Not MyHour = 9 Then
        Do Until MyHour = 9
          MyTime = Cells(lastrow - e, c + 5).Value
          MyHour = Hour(MyTime)
          opening = Cells(lastrow - e, c).Value
          orow = Cells(lastrow - e, c).Row
        Loop
      End If
    Next c
  Next h
Next f
End Sub
```

```

    e = e + 1
  Loop
End If
trade = Cells(orow, c).Value
lasttrade = Cells(orow - d, c).Value
If trade = lasttrade Then
  Do Until trade <> lasttrade
    lasttrade = Cells(orow - d, c).Value
    traderow = Cells(orow - d, c).Row
    d = d + 1
  Loop
  lasttrade = Cells(traderow, c).Value
ElseIf trade < lasttrade Then
  buyvol = Cells(orow, c + 3).Value
  bid = 0
  ask = 1
ElseIf trade > lasttrade Then
  sellvol = Cells(orow, c + 3).Value
  bid = 1
  ask = 0
End If
For r = -orow To -3
  trade = Cells(orow - a, c).Value
  lasttrade = Cells(orow - b, c).Value
  If trade = lasttrade Then
    If ask = 1 Then
      buyvol = buyvol + Cells(orow - b, c + 3).Value
      bid = 0
      ask = 1
    ElseIf bid = 1 Then
      sellvol = sellvol + Cells(orow - b, c + 3).Value
      bid = 1
      ask = 0
    End If
  ElseIf trade < lasttrade Then
    buyvol = buyvol + Cells(orow - b, c + 3).Value
    bid = 0
    ask = 1
  ElseIf trade > lasttrade Then
    sellvol = sellvol + Cells(orow - b, c + 3).Value
    bid = 1
    ask = 0
  End If
  a = a + 1
  b = b + 1
Next r
Workbooks("Summary").Worksheets("Summary II").Activate
Worksheets("Summary II").Cells(n, 3) = opening
Worksheets("Summary II").Cells(n, 11) = buyvol
Worksheets("Summary II").Cells(n, 12) = sellvol

```

```
Workbooks(my_book).Worksheets(my_sheet).Activate
c = c + 6
n = n + 1
buyvol = 0
sellvol = 0
a = 0
b = 1
d = 1
e = 1
Next c
Workbooks("Summary").Worksheets("Summary III").Activate
Next h
Next f
End Sub
```

Appendix IV

The data presented in Table 4.2 has been calculated using the equations below.

$$\text{Sample Mean} = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\text{Sample Standard Deviation} = s_N = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

$$\text{Standard Error} = \frac{s_N}{\sqrt{N}}$$

$$\text{Sample Kurtosis} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{\left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^2} - 3$$

$$\text{Sample Skewness} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3}{\left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^{3/2}}$$

$$\text{Range} = \text{Maximum} - \text{Minimum}$$

$$\text{Sum} = \sum_{i=1}^N x_i$$

Appendix V

Figure A.1 Plot of the Frequency Distribution of the Daily Returns.

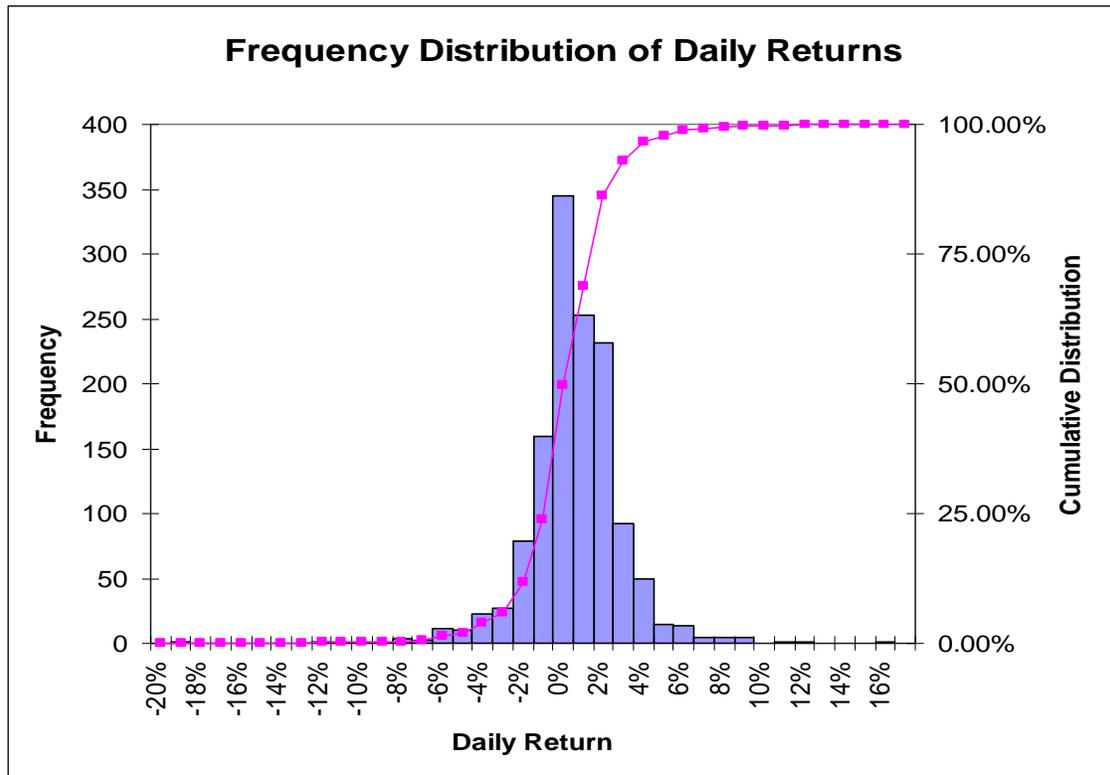


Figure A.2 Plot of the Frequency Distribution of the Normalized Net Volume.

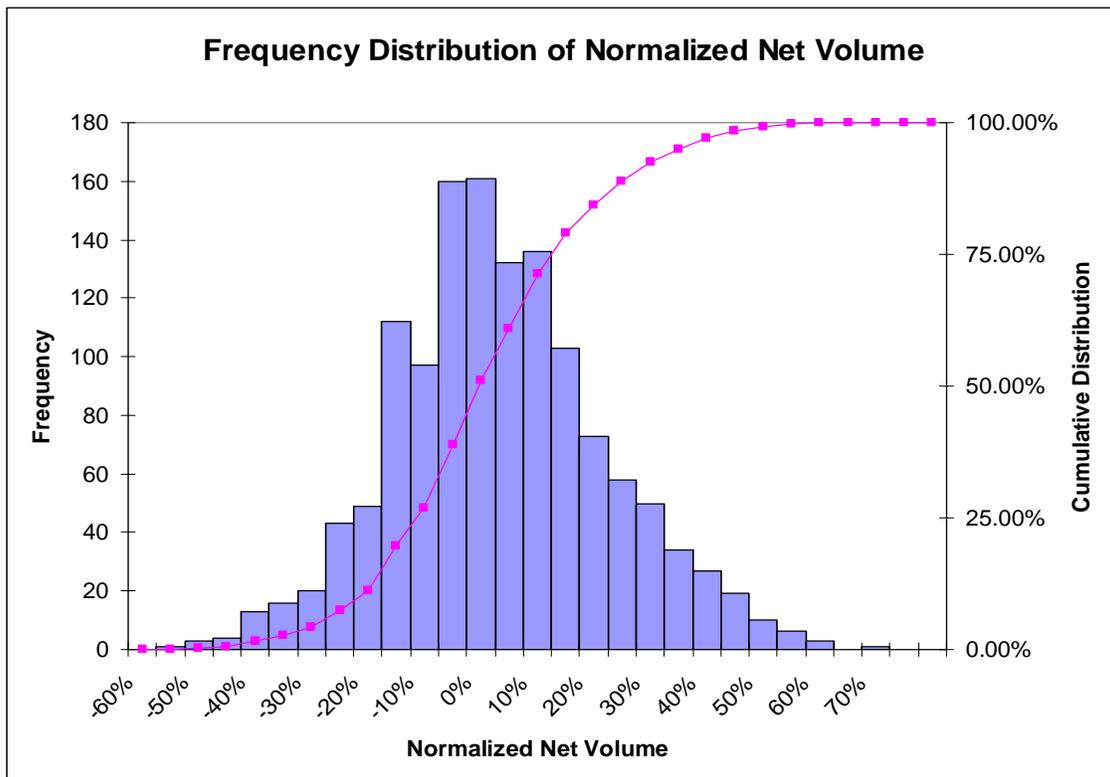


Figure A.3 Plot of the Frequency Distribution of the Net Money Flow.

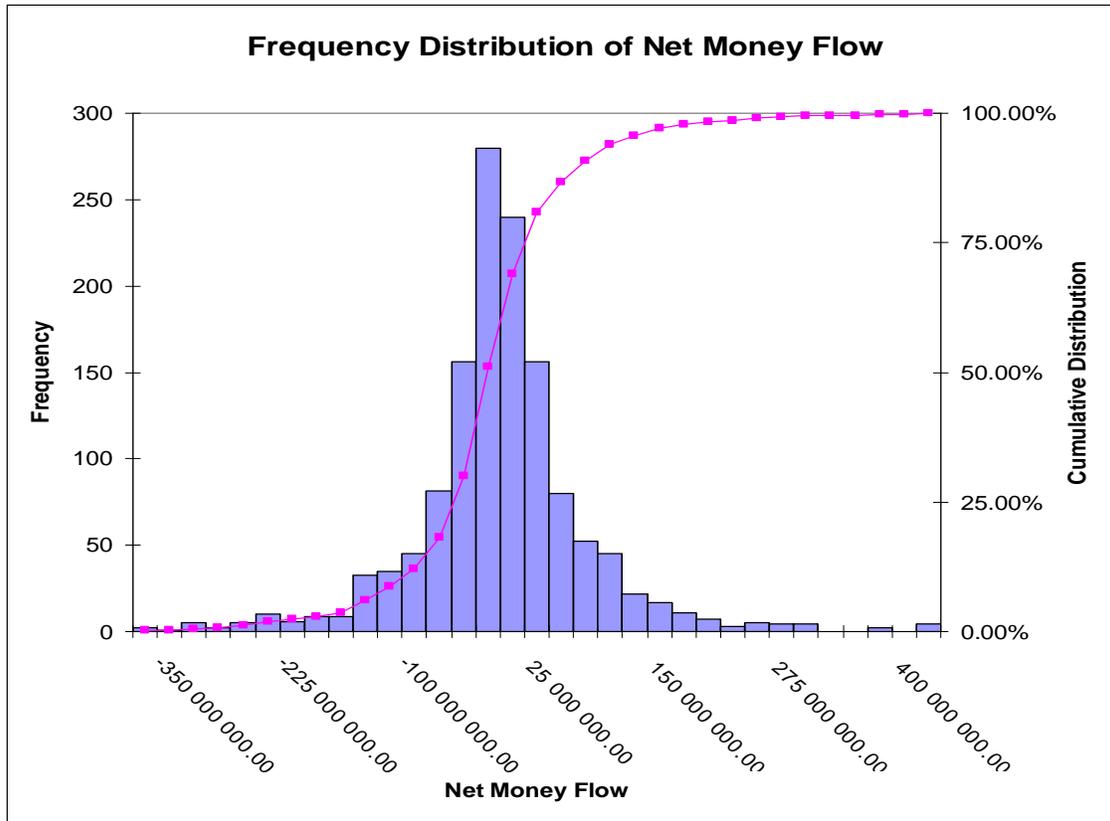


Figure A.4 Plot of the Frequency Distribution of the Normalized OFI.

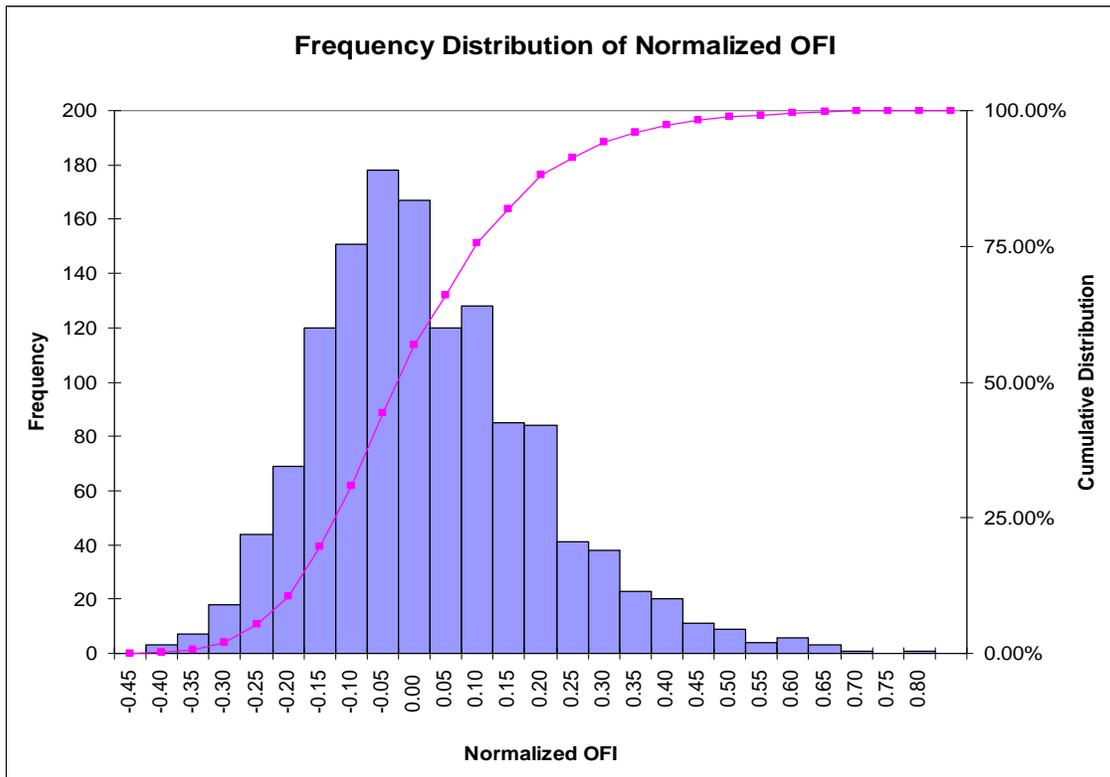
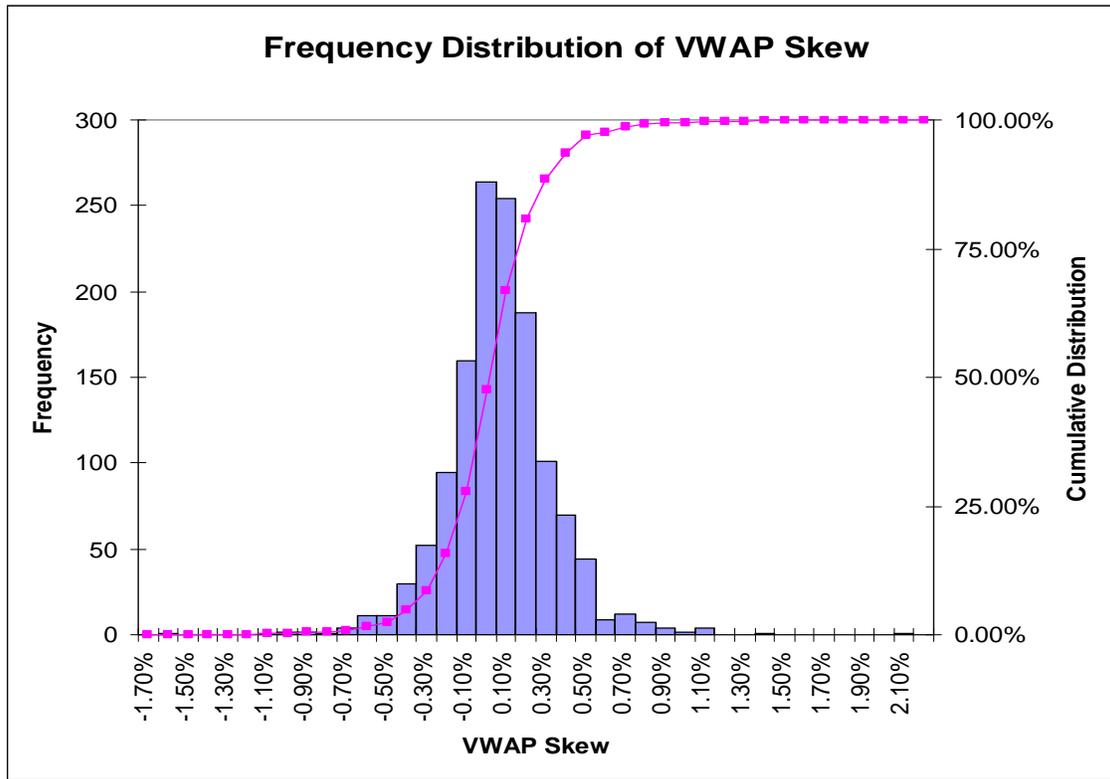


Figure A.5 Plot of the Frequency Distribution of the VWAP Skew.



INDEX

A

A.M.F. 8, 42
ABB Ltd 11, 17, 43, 45, 51
Accumulated Money Flow 8, 42, 43
Aldridge, I. 9, 19, 25, 26, 46, 53
Analysis of Variance 8, See ANOVA
ANOVA 8, 41, 42

B

Bank of America 27
Bell, E. 13, 14, 19, 20, 53
Bennett, J. 29, 53
Black-box trading 9, 10, 26
BNY ConvergEx Group 28, 55
Borsellino, L. 25, 53
Brown, B. 9, 10, 11, 26, 44, 53
Bryman, A. 13, 14, 19, 20, 53
Bukey, D. 30, 55

C

CBOT 8, 30
Chan, E. 53
Chicago Board of Trade 8, See CBOT
Chicago Mercantile Exchange 8, See CME
CME 8, 10
Cohen, L. 20, 53
Coldwell, D. 13, 53
Collins, G. 24, 55
Credit Suisse First Boston 27
Creswell, J. 14, 53

D

Dacorogna, M. 17, 53
Direct Market Access 8, See DMA
DMA 8, 26
Dowd, K. 53
Dunis, C. 53

E

Elder, Dr. A. 28, 53

F

Fabozzi, F. 26, 53, 54
Fadiman, M. 9, 53
Focardi, S. 26, 53
Foster, O. 24, 53
FTEN Inc. 10

G

Gavridis, M. 17, 53
Gençay, R. 17, 53
Ghauri, P. 13, 14, 15, 20, 22, 53
Glantz, M. 54
Goldman Sachs 27
Graifer, V. 24, 53

Grossman, S. 50, 54
Grønhaug, K. 13, 14, 15, 20, 22, 53
Gujarati, D. 19, 20, 54

H

Hale, S. 24, 55
Harris, L. 25, 26, 54
Hasbrouck, J. 26, 54
Hawkins, S. 18, 29, 30, 31, 32, 33, 34, 54
Heires, K. 10, 54
Herbst, F. 13, 53
high-frequency 9, 10, 11, 16, 17, 23, 24, 25, 49, See
High-frequency Trading
high-frequency trading 9, 11, 12, 16, 44, 49

I

Implementation shortfall 28, 46
Infront AS 21
IQx algorithm See BNY ConvergEx Group

K

Kakade, S. 27, 55
Kearns, M. 27, 55
Kim, K. 27, 28, 31, 54
Kissell, R. 54
Klein, J. 9, 53
Kolm, P. 26, 53
Kumar, R. 14, 54

L

Limit orders 25, 26, 27, 33
Link, M. 10, 54
Lukeman, J. 23, 54

M

Madsen, H. 19, 54
Manion, L. 53
Mansour, Y. 27, 55
Market makers 10, 26
Market orders 25, 27, 33
Money flow 29, 30
Morrison, K. 53
Müller, U. 17, 53, 54

N

N.M.F. 8, 33, 34, 35, 36, 40, 44, 51
N.V. 8, 19, 33, 34, 40
Narang, R. 27, 54
NASDAQ 8, 10
National Association of Securities Dealers
Automated Quotations 8, See NASDAQ
Neill, H. 23, 24, 54
Net Money Flow ... 8, 19, 33, 35, 36, 37, 38, 44, See
N.M.F.
Net Volume 8, 33
New York Stock Exchange 8, See NYSE

Norm. N.V.8, 19, 33, 34, 35, 36, 37, 40
 Norm. OFI.....8, 19, 34, 35, 36, 37, 44, 51
 Normalized Net Volume8, 19, 33, 35, 38, See
 Norm. N.V.
 NYSE8, 10, 11, 26, 29, 55
 NYSE Euronext..... 29, 55

O

O'Hara, M. 26, 55
 OFI.....8, 18, 30, 31, 33, 34, 35, 39, 44, 64
 OLS 8, 40
 Olsen, R..... 17, 53
 On Floor Information 8, 19, 30, See OFI
 Ordinary Least Squares 8, See OLS
 Ortiz, L. 27, 55
 OTC 8, 16
 Over-the-Counter8, See OTC

P

Pachamanova, D..... 26, 53
 Perold, A. 28, 54
 Pictet, O..... 17, 53
 Pole, A..... 26, 54

R

Reg ATS..... 8, 10
 Regulation of Alternative Trading Systems8, See
 Reg ATS
 Renaissance Technologies Corp See Simons, J., See
 Simons, J.
 Riksbank 46, 55

S

Schubert, B.....26, 27, 28, 29, 54

Schumacher, C.24, 53
 SEC 8, 10
 Securities and Exchange Commission ... 8, See SEC
 Sharpe ratio45, 46
 Sias, R.29, 30, 33, 49, 50, 53
 Simons, J. 25
 Six Swiss Exchange 11
 Six Telekurs 15
 SSE8, 11, 15, 16, 17, 31
 Steidlmayer, J..... 18, 29, 30, 31, 32, 33, 34, 54, 55
 Stiglitz, J.50, 54
 Stockholm Stock Exchange 8, See SSE

T

Taleb, N. 17, 20, 21, 54
 Tape11, 23, 24
 Tape reading11, 23, 24
 Tape, R.....12, 17, 23, 53, 54, 55
 Ticker tape See Tape
 Time-Weighted Average Price.....8, 27, See TWAP
 TWAP8, 27, 28, 31, 52

V,W

V@T8, 31, 32, 33
 Volume @ Time 8, 31, See V@T
 Volume-Weighted Average Price 8, 18, 27, See
 VWAP
 Worden, D.....24, 33
 VWAP.. 8, 18, 19, 27, 31, 33, 34, 35, 36, 37, 39, 44,
 49, 51, 52, 55, 57, 65
 Wyckoff, R..... 23

Z

Zhou, B. 53