Theoretical and Practical Aspects of Algorithmic Trading

Zur Erlangung des akademischen Grades eines Doktors der
Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät fuer
Wirtschaftwissenschaften
des Karlsruher Instituts fuer Technologie

genehmigte

Dissertation

von

Dipl.-Phys. Jan Fränkle

Tag der mündlichen Prüfung: 07.12.2010
Referent: ....Prof. Dr. S.T. Rachev
Korreferent: ....Prof. Dr. M. Feindt

Erklärung

Ich versichere wahrheitsgemäß, die Dissertation bis auf die in der Abhandlung angegebene Hilfe selbständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und genau kenntlich gemacht zu haben, was aus Arbeiten anderer und aus eigenen Veröffentlichungen unverändert oder mit Abänderungen entnommen wurde.

Contents

Chapter 1

Introduction

1.1 Objective

The current work is concerned with relevant aspects of algorithmic stock trading. It was first and foremost motivated by a concrete problem arisen from the order execution process of the Lupus alpha NeuroBayes® Short Term Trading Fund. This is a high frequent trading fund, invested in large cap stocks, and classifies as a hedge fund in the popular terminology. For this purpose, a quite general issue has to be solved, namely the execution of many large orders at the stock exchanges in a given predefined time. The optimal solution would of course be an optimized fully automated trading algorithm executing orders with low market impact and minimum exchange fees. This is a typical challenge for market participants such as brokers and asset managers of mutual funds, ETFs, and also hedge funds.

In this research thesis I want to investigate in more detail the most important components of the mentioned automated trading algorithm. Furthermore, analyses for the optimization of execution strategies are provided. The realization of such a trading algorithm requires a multitude of aspects which have to be taken into account. They range from regulation aspects via market microstructure to very technical aspects such as electronic market access and computer systems. Therefore, many topics will be discussed by reviewing the appropriate literature. Further research is focused on a market impact model and a model for the prediction of trading volume.

One of the most crucial points is the understanding of trading costs, especially the implicit ones, which are called “market impact”. Market impact is the interaction of a market participant’s own action on the entire market, i.e. the price of the traded security is affected by the trading action. The direction of the price impact is in the same direction as the causal trade (e.g. a buy drives the price up). Thus, the execution price of the security is in general worse than the price before the trading starts. Qualitatively, this effect is easily explainable but the quantification and the prediction of this effect is a rather complex task. In order to reduce the market impact of an order execution, the influencing factors have to be known. Hence, an empirical analysis of a large data set of homogeneous algorithmic trades is done. The result of this analysis is a model which is able to predict the market impact of an algorithmic trade. This model is used as an important input of the portfolio optimization of the above mentioned NeuroBayes Fund.

Based on this market impact model, it is shown that the so-called VWAP trading strategy is optimal. The implementation of such a VWAP execution strategy requires knowledge about the overall future trading volume of the given security. Therefore, a model for trading volume predictions is developed.

1.2 Approach

The market impact analysis is based on a proprietary data set. The main advantage of this data set as compared to publicly available ones is due to the fact that single orders of a certain market participant are identified and logically connected. Because of that, the resulting market impact of the entire transaction can be measured. Thus, the market impact of the total algorithmic execution, which is spread over an extended period of time, is observed. Otherwise, the single orders usually are assumed to be independent which is obviously not true. The underlying algorithmic trades originate from the trading activity of the NeuroBayes Fund from April 2008 to July 2010 with an overall trading volume of more than 30 billion USD equivalent. Its stock universe covers the most liquid 800 stocks of various markets (Canada, Europe, Japan, and USA). The analyzed data set contains homogeneous algorithmic executions which have several parameters in common, for example the trading period and the usage of the same execution strategy. However, they differ in some aspects such as size of the orders and different markets. Nevertheless, I think that the conclusions of this analysis are universally valid because they correspond to the characteristics of the stock market.

The distribution of the intraday trading volume of almost all stocks has the famous u-shaped pattern. This means that the trading activity is high shortly after the opening of the market in the morning and before the closing in the evening. During lunch time the market activity is at its low. Although the pattern is quite similar for all stocks over years, there are significant differences. For example, a change in this pattern is observed during the financial crisis in 2008 and there are slightly different patterns for highly liquid stocks and for stocks with lower liquidity. The trading volume model takes these aspects into account and is able to predict the intraday trading volume pattern for different stocks. To this end, market data on a minute-by-minute basis of the US stock market is used.

1.3 Outline

This thesis covers many aspects of algorithmic trading. The work is split into 5 parts. Part I provides the mathematical background of the most important methods used in the analyses. Part II comprises a detailed literature review about algorithmic trading. Several topics of market microstructure are discussed, such as the functionality of a financial market and the meaning of trading costs. Additionally, an overview is given of the existing execution strategies which are well known in the finance industry.

The market impact model is presented in part III. First of all, the data set is described in detail. Subsequently the models and the fit results are described. Additionally, the individualized linear model is introduced. Finally, the current results are compared with another empirical measurement and the market microstructure theory. It is shown that the VWAP strategy is optimal for the presented market impact models.

Part IV encompasses the trading volume model, starting with the description of the data set. The data is processed by calculating different mean values and writing them as matrices. These matrices are analyzed with the help of SVD (singular-value decomposition). The results are used to predict the future trading volume distributions.

Finally, chapter 14 recapitulates the main results of the work and gives an outlook for potential future research.

Part I

Theoretical Background

Chapter 2

Mathematical Methods

2.1 Maximum Likelihood

2.1.1 Principle of the ML Method

Maximum likelihood is a general and often used method for parameter estimation. It is a concept to determine an adequate probability distribution for a given set of observations. A set of observations is called a sample. It is usually assumed that the observations are independently drawn from the same probability density distribution, Rachev *et al.* (2005). Under appropriate conditions, the maximum likelihood estimators are consistent, asymptotically unbiased, and efficient. The maximum likelihood procedure searches for a parameter set that is most suitable for the set of observations.

The random variables $y_0, ..., y_T$ are assumed to be iid (independent identically distributed) and drawn from the probability density distribution $f(\cdot;\theta)$, where θ is the parameter vector, see Rachev *et al.* (2007). The joint probability distribution function of $Y_T = (y_0, ..., y_T)$ can be written as

$$f(Y_T;\theta) = \prod_{t=0}^{T} f(y_t;\theta) \tag{2.1}$$

The function $f(Y_T;\theta)$ may be viewed as a function of θ and thus $f(Y_T;\theta)$ is an indication of the plausibility of a particular θ for the data set Y_T. The likelihood function of $f(\cdot;\theta)$ for data set Y_T is defined as:

$$L(\theta;Y_T) = f(y_t;\theta) \tag{2.2}$$

The MLE (maximum likelihood estimator) of θ, $\hat{\theta}_{ML}$ has the property that for any other estimator $\hat{\theta}$

$$L(\hat{\theta}_{ML};Y_T) \geq L(\hat{\theta};Y_T) \tag{2.3}$$

The MLE $\hat{\theta}_{ML}$ of θ is obtained by finding the value of θ that maximizes $f(Y_T;\theta)$. Thus, the aim is to find the values of the unknown parameter set that maximizes the likelihood computed for a given set of observations $y_0, ..., y_T$ (see Rachev *et al.* (2007)).

Hence the maximum likelihood method is able to estimate parameters of a probability density distribution in contrast to the estimation of parameters of an arbitrary function as it is the case when using the χ^2 method. To estimate parameters describing a functional dependency between two (or even more) variables, the issue has to be formulated as an estimation of a probability density function. This is usually done by the estimation the probability density function of the residuals.

2.1.2 Error Estimation

The maximum likelihood method also enables the calculation of the statistical uncertainty for the estimated parameters.

In the case $T \to \infty$ the likelihood function converges to a Gaussian function and the variance $V[\hat{\theta}_{ML}] \to 0$. The negative log-likelihood function can be expanded around its minimum.

$$logL(\theta) = logL(\hat{\theta}_{ML}) + \left[\frac{\partial logL}{\partial \theta}\right]_{\theta=\hat{\theta}_{ML}} (\theta - \hat{\theta}_{ML}) + \frac{1}{2!}\left[\frac{\partial^2 logL}{\partial \theta^2}\right]_{\theta=\hat{\theta}_{ML}} (\theta - \hat{\theta}_{ML})^2 + ... \tag{2.4}$$

The negative log-likelihood function has the form of a parabola around its minimum and the second derivation is constant. The first derivation at $\theta = \hat{\theta}_{ML}$ is 0 and higher order terms are ignored. Then one get:

$$logL(\theta) = logL(\hat{\theta}_{ML}) - \frac{(\theta - \hat{\theta}_{ML})^2}{2\sigma^2_{\hat{\theta}_{ML}}} \tag{2.5}$$

or

$$logL(\hat{\theta}_{ML} \pm \sigma_{\hat{\theta}_{ML}}) = logL(\hat{\theta}_{ML}) - \frac{1}{2} \tag{2.6}$$

with

$$\sigma_{\hat{\theta}_{ML}} = \left(\left. \frac{d^2L}{d\theta^2} \right|_{\hat{\theta}_{ML}} \right)^{-\frac{1}{2}} \tag{2.7}$$

The error calculation for more parameters can be done analogously, where the minimal function $L_{min}(\theta_i)$ of the i-th parameter of $\vec{\theta}$ is formally given by $L_{min}(\theta_i) = minL(\vec{\theta})$. $L_{min}(\theta_i)$ is the minimum of $L(\vec{\theta})$ with respect to all the other parameters.

Further detailed information about the error calculation of estimated parameters can be found in Blobel and Lohrmann (1998) and Cowan (1998).

2.2 Singular-Value Decomposition

2.2.1 Theorem

The Singular-Value Decomposition (SVD) is a factorization of a rectangular real or complex matrix A of the dimension $m \times n$ (see Quarteron *et al.* (2000)):

$$A = U\Sigma V^* \tag{2.8}$$

Hereby, U is a $m \times m$ unitary matrix whose columns are the left singular vectors of A. Σ has the same dimension as A and is a diagonal matrix with nonnegative real numbers being in fact the singular values of A. Finally, the unitary matrix V^* (conjugate transpose of V) has the dimension $n \times n$ whose rows are the right singular vectors. A common convention is to order the diagonal entries $\Sigma_{i,j}$ in descending order.

2.2.2 Low-rank Approximation

A practical application of the SVD is the low-rank matrix approximation. The matrix A can be approximated by matrix $\tilde{A}$ with rank r

$$\tilde{A} = U\tilde{\Sigma}V^* \tag{2.9}$$

where $\tilde{\Sigma}$ is the same matrix as Σ except only the largest r singular values are unequal to 0 and $rank(\tilde{A}) = r$. This is known as the Eckart-Young theorem (see Gower and Dijksterhuis (2004)).

If the matrix $\tilde{A}$ is a good approximation for matrix A, the Frobenius norm of the difference $A - \tilde{A}$ has to be minimal

$$||A - \tilde{A}||_F \stackrel{!}{=} min \tag{2.10}$$

where the Frobenius norm is defined as $||X||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij}|^2}$.

Take $A = U\Sigma V^*$ and the invariance under unitary transformations of the Frobenius norm, one arrive at the following:

$$\begin{aligned} min||A - \tilde{A}||_F &= min||\Sigma - S||_F \\ &= min\sqrt{\sum_{i=1}^{n}(\sigma_i - s_i)^2} \\ &= min\sqrt{\sum_{i=1}^{r}(\sigma_i - s_i)^2 + \sum_{i=r+1}^{n}\sigma_i^2} \\ &= \sqrt{\sum_{i=r+1}^{n}\sigma_i^2} \end{aligned} \tag{2.11}$$

Note that Σ is diagonal, so $S = U^*\tilde{A}V$ also has to be diagonal on order to minimize the Frobenius norm. s_i and σ_i denote the diagonal elements of S and Σ respectively. Thus $\tilde{A}$ is a good approximation for A when $\sigma_i = s_i$ and the corresponding singular vectors are the same as those of A.

The low rank approximation of a matrix can be used for data compression and noise reduction. Writing it out, the rank 1 approximation explicitly leads to

$$\tilde{a}_{ij} = \sigma_i \cdot u_{i1} \cdot v_{j1} \tag{2.12}$$

Hence the matrix elements of $\tilde{A}$ are a product of the singular value and the according components of the left and right singular vector.

Further details about SVD as well as applications can be found numerously in literature, for example Eldén (2007) and Berry and Browne (2005).

Part II

Algorthmic Trading [1]

[1] this part of the work is published in Fränkle and Rachev (2009)

Chapter 3

Algorithmic Trading

Algorithmic trading is automated trading, i.e. a computer system is completing all work from the trading decision to the execution. Algorithmic trading has become possible with the existence of fully electronic infrastructure in stock trading systems such as (direct) market access, exchange and market data provision. The following overview gives a flavor of chances and challenges in algorithmic trading as well as an introduction to several components needed for setting up a competitive trading algorithm.

3.1 Chances and Challenges

In contrast to trading by humans, algorithmic trading has several advantages. Computer systems in general have a much shorter reaction time and reach a higher level of reliability than humans. The decisions made by a computer system rely on the underlying strategy with specified rules. This leads to a reproducibility of these very decisions. Thus back-testing and improving the strategy by varying the underlying rules is made possible. Algorithmic trading ensures objectivity in trading decisions and is not exposed to subjective influences (such as panic for example). When trading many different securities at the same time, one computer system may substitute many human traders. So both the observation and the trading of securities of a large universe becomes possible for companies without employing dozens of traders. Altogether these effects may result in a better performance of the investment strategy as well as in lower trading costs. For further information concerning algorithmic trading and artificial agents, see Boman *et al.* (2001), Kephart (2002), LeBaron (2000) and Gudjónsson and MacRitchie (2005).

Nevertheless, it is challenging to automate the whole process from coming to investment decisions to execution. System stability is crucially important. The algorithm has to be robust against numerous possible errors. The trading system is dependent on different services such as market data provision, connection to market and the exchange itself. These are technical issues which can be overcome by diligently implementing the system. Even more complex is the development of an investment strategy, i.e. deriving trading decisions and strategies implementing these decisions. This work focuses on the implementation and thus the execution strategy of externally given investment decisions. It is beyond the scope of this work to cover the process of a quantitative framework for derivation of investment decisions.

The inputs for the execution strategy are for example security names, the number of shares, and the trading direction (buy/sell). In addition there may be inputs such as aggressiveness and constraints (for example market neutrality during the execution when trading a basket), dependent on the needs of the investment strategy.

The main challenge for trading algorithms is the realization of low trading costs in preferably all market environments (independent from falling or rising markets and high or low liquidity). Another critical point which has to be taken into account is the transparency of the execution strategy for other market participants. If a structured execution strategy acts in repeating processes (for example, orders are sent in periodical iterations) other market participants may then observe patterns in market data and may take an advantage of the situation.

3.2 Components of an Automated Trading System

Trading algorithms have the advantage that the execution performance is measurable and predictable for a specified order (see part III of the current work). Hence, the profitability of algorithmic trades can be calculated in advance by taking trading costs into account. The realization of this implies additional components which are already suggested by Investment Technology Group (2007) and Kissell and Malamut (2005).

A **pre-trade analysis** component provides a preliminary estimate of transaction costs of a given algorithmic order. To this end, an econometric model based on historical trading data is used. The pre-trade analysis can be used to optimize the expected transaction costs by varying the parameters or even the trading strategy. In the pre-trade analysis also a more general optimization function can be optimized due to the trader's preferences, for example:

$$(1-\lambda) \cdot E(C) + \lambda \cdot Var(C) \rightarrow min$$

where C is the total execution cost of a trade, E(C) the expected value of C, and Var(C) is the variance of C. λ is the traders risk aversion parameter (see Investment Technology Group (2007)). The expected cost of a trade E(C) can contain opportunity costs if the trader allows the algorithm not to execute the complete position. Yang and Jiu (2006) provide an empirical approach of selecting algorithms which satisfy the traders needs best. Domowitz and Yegerman (2005) explain how to compare the performance of algorithms and specify some parameters for trading algorithm. An approach to forecast and optimize execution is also provided in the work of Coggins *et al.* (2006). The second component is the **trading algorithm** itself. It's the part of executing orders according to the underlying strategy (see 5.2). The optimal strategy has to be found with the help of the pre-trade analysis, but further improvement can be reached by adjusting parameters during the trading period. Therefore Bialkowski *et al.* (2005) and Bialkowski (2008) provide a model of decomposing trading volume and model the components to forecast the trading volume (see also part IV). This can be taken into account by the trading algorithm if it is based on trading volume such as VWAP. Obizhaeva and Wang (2005) show the relationship of supply and demand dynamics of a security in the market and the execution performance of a given order. They provide a model of the impact of supply/demands dynamics on execution costs. **Post-trade analysis** is the third component of the system. After all information of the trades are available, a performance measurement can be done and be compared to the pre-trade estimation. This is a very important information to improve pre-trade analysis for future trades. See Investment Technology Group (2007) as an example of a post-trade analysis framework. Kissell and Malamut (2005) suggest a two part post trade analysis of trading cost measurement and algorithm performance measurement. The estimation of the market impact model in part III of the current work can also be seen as a post trade analysis. However the estimated model is a very important input for the pre-trade analysis.

Chapter 4

Market Microstructure

Several definitions of market microstructure have been suggested in the literature. Two of the more notable ones are provided by O'Hara (1995) and Stoll (2001). O'Hara defines market microstructure as "the study of the process and outcomes of exchanging assets under explicit trading rules". Stoll defines market microstructure as "the study of trading costs and the impact costs resulting in the short-run behavior of security prices". As it will be shown, both definitions are very similar in their meaning. Moreover, it will be explained why trading costs are a very basic element in market microstructure. This section introduces market microstructure theory and gives a short overview of the literature.

A general overview and introduction in market microstructure theory is given by O'Hara (1995). Besides an introduction to price determination, inventory models of market makers are presented and also a theory behind bid-ask spreads. She identifies the influence of trading strategies on market microstructure and the information of trades in the price process. Harris (2002) provides a more practical view on market microstructure, explaining the background for some key elements of market microstructure and the investment objectives and activities of different market participants. Harris also presents a review of trading platforms and the role they play. Cohen *et al.* (1986) provides a detailed cross-sectional comparison of the worldwide equity markets. Stoll (2001) focuses on trading costs, market designs and the forces leading to the centralization of trading in a single market versus the forces leading to multiple markets. Madhavan (2000) provides a review of theoretical, empirical, and experimental literature on market microstructure with a focus on informational issues.

4.1 Nature of the Market

One of the principal functions of financial markets is to bring together the parties interested in trading a security. Trading platforms are the most efficient way to bring these counterparts together. Such trading platforms can be accomplished via the physical presence of brokers and traders on the floor of an exchange. But it can also be realized as an electronic platform where the physical location is unimportant and market participants are just connected electronically. The third alternative is a hybrid market wherein there is both a trading floor and an electronic platform. The best example of a hybrid market is the New York Stock Exchange (NYSE).

The trading process itself is similar for all financial markets. All market participants express their trading interest with an order which is sent to the market. An order contains the information which security to trade, the direction (buy or sell), the quantity of shares, and a limit price expressing the worst price the party is willing to accept. When the limit price is not identified as part of the order, this results in a market order and the party is willing to accept all prices.

The task of financial markets is to match compatible orders and execute them. Each market defines its trading rules to enable high liquidity and fast execution with low price volatility in order to get fair prices and an efficient trading process. A very basic idea for the trading process is the Walrasian auctioneer. Each market agent provides a demand-price function to the auctioneer who first aggregates these orders and then computes a price where demand and supply are equal, called the market-clearing price.

Walrasian auctions are discrete auctions; that is, trading takes places only at specified times during the trading day. Modern exchanges provide continuous trading, hence market participants have the opportunity to trade at any time during the trading day. But for each trading interest a counter-party has to be found, willing to trade the same position in the contrarian direction. In the limiting case of iterating Walrasian auctions with infinite frequency, continuous trading would be realized but the execution probability of a trader's order would be equal to zero. The probability of two orders reaching the same auction declines with the increase of the frequency of auctions if the trader's order is valid for exactly one auction. So there is a need for orders which are valid for more than one auction. Such orders do not satisfy investor's needs to be executed immediately, but their existence enables immediate execution of other orders. Thus besides market participants preferring immediate execution, other market participants providing liquidity are needed. Traditionally, market participants providing liquidity are called market makers. Their profit arises from the existence of the implicit liquidity premium that the counter party seeking liquidity is willing to pay. This premium is represented by the bid-ask spread of the order book. The bid-ask spread increases with the size of the trade and reflects the expected risk the market maker incurs.

4.2 Continuous Trading and Open Limit Order Book

Most stock markets provide continuous trading. Some markets have additional discrete call auctions at specified times when the uncertainty is large, for example at the open, close, and reopen after a trading halt caused by large price movements. The economic justification is that call auctions are especially helpful in uncertain times during the trading day because of the information aggregation argument, see Madhavan (2000).

Open limit order books are the core of most continuous trading systems. A limit order book contains limit orders of market participants, including the information about the limit price, quantity of shares, and trading direction (buy or sell). The content of open order books is published in contrast to closed order books where no information about the status of the market is provided as it is realized in so-called "dark pools".

The most relevant measure of order books is the bid-ask spread. It is the difference between the lowest provided sell price (ask) and the highest buy price (bid).

By definition of the order book, the ask price is always higher than the bid price. The bid-ask spread is a good measure for the liquidity of a security, i.e. in actively traded securities the spread in general is smaller than in inactive markets. Implicit trading costs arise in continuous trading through the existence of the bid-ask spread. Liquidity takers have to cross the spread for trading which can be interpreted as the premium for liquidity provision. This premium is justified by the risks and the costs the liquidity provider faces, such as inventory risk and order handling costs. The competition between liquidity providers forces the market in the direction to lower spreads. Some theoretical studies concerning liquidity provision are provided by Biais *et al.* (1995), Biais *et al.* (1999), Harris and Hasbrouck (1996), and Foucault (1998).

Trading takes place when an order arrives at the order book and matches at least one existing order book position. This is the case if the limit price of the incoming buy (sell) order is higher (lower) than the current ask (bid). Otherwise the order is inserted into the order book as an additional position. The execution price of a trade is always the limit price of the order book position which is involved. This leads to jumps in security prices from bid to ask prices depending on the direction the initiator trades, see Garman (1976) and Madhavan *et al.* (1997) for models describing time series behavior of prices and quotes.

For the best bid and ask positions, the provided volume is typically quite small compared to the entire order book volume and also small compared to typical order sizes of institutional investors. Submitting a large aggressive order to a continuous trading system leads to a sharp price movement and a rebuilding of the order book afterwards, resulting in huge implicit trading costs because of the large realized bid-ask spread. So optimal trading in continuous trading systems requires adapted strategies where large orders are split up into several smaller orders which are traded over an extended period of time. In the period between the execution of the slices, the order book can regenerate in the sense that liquidity providers narrow the spread after it has widened through a trade, see Obizhaeva and Wang (2005).

4.2.1 Resilience of the Order Book

What happens with a limit order book during and after the execution of a market order? The following section gives an overview of the interaction between the order book and a marketable order. A marketable order is an order taking liquidity from the order book, that is, all market orders but also limit orders with aggressive limits. The direct reaction of the order book on such a marketable order is quite simple, the incoming order is matched with the waiting passive orders. This results in the widening of the spread and the reduction of the provided volume in the order book. The more interesting effect is the reaction of the market after the execution. How the spread will narrow and how the provided volume in the book will re-rise. This effect is called the resilience of the limit order book.

Alfonsi *et al.* (2007) presents two approaches of modeling the resilience. An exponential recovery of the limit order book is assumed. One approach models the recovery of the limit order book inventory. The second approach models the narrowing of the bid-ask spread. For measuring this effect, a reference limit price has to be defined. This unaffected limit (best bid or ask) is modeled by a Brownian motion. A similar model together with an empirical test on TAQ data is provided by Dong and Kempf (2007). They use the following model to describe the last execution price:

$$S(t) = F(t) + Y(t) \tag{4.1}$$

where

$$F(t) = \mu + F(t-1) + \epsilon(t) \tag{4.2}$$

and

$$\Delta Y(t) = Y(t) - Y(t-1) = -\alpha Y(t-1) + \Phi(t) \tag{4.3}$$

and

$$\Phi(t) \sim N(0, \sigma_\Phi^2), \quad \epsilon(t) \sim N(0, \sigma_\epsilon^2) \tag{4.4}$$

The last execution price $S(t)$ is written as a sum of the components $F(t)$ and $Y(t)$. $F(t)$ represents a random walk with drift describing the underlying price process. The term $Y(t)$ describes the price recovery approach and $\Delta Y(t)$ is interpreted as the "pricing error" which tends towards zero because of the market forces. The resilience is denoted by the mean-reversion parameter α. This model is applied on 1-minute NYSE TAQ data using a Kalman-filter smooth estimation procedure to estimate the resilience measure α. The mean value of all the resilience estimates is $\alpha = 0.60$ which is significantly different from both zero and one. This means that the pricing error is stationary. On average around 60% of the pricing error is corrected in the following 1-minute interval.

Additionally, some description variables of the order book resilience are analyzed. The price level (inverse of tick size) has a negative effect on resilience indicating that lower tick size leads to more resilience. The number of trades is positively correlated to resilience whereas average trading size is negatively correlated and also the volatility of the stock price.

4.2.2 The Open Limit Order Book and Execution Probability

The functioning of a limit order book is described in section 4.2. The following focuses on the dynamics of limit orders in order books and thus the interaction between the order book and the order flow.

Theoretical models provided by Kyle (1985) or Glosten and Milgrom (1985) focus on market maker quotations. Glosten (1992) analyzes limit order markets by modeling the price impact of trades reflecting their informational content.

Biais *et al.* (1995) provide an empirical analysis of order book characteristics, starting with descriptive statistics. They compute the number of ticks between bid and ask quotes as well as between adjacent quotes. First, they find that the bid-ask spread is twice the difference between adjacent quotes on each side of the order book. And secondly they find a tick size dependency of these calculated differences. The median difference between neighboring limits is larger than one tick size. Additionally, the bid-ask spread and the relative spreads on each side of the book show an intraday u-shaped pattern. They also analyzed the order volume distribution in the order book dependent on the limit prices. The depth (cumulative order volume per price) increases with the distance from the best bid and ask respectively, see figure 4.1.

Besides order book characteristics, order flow is analyzed in detail by Biais *et al.*

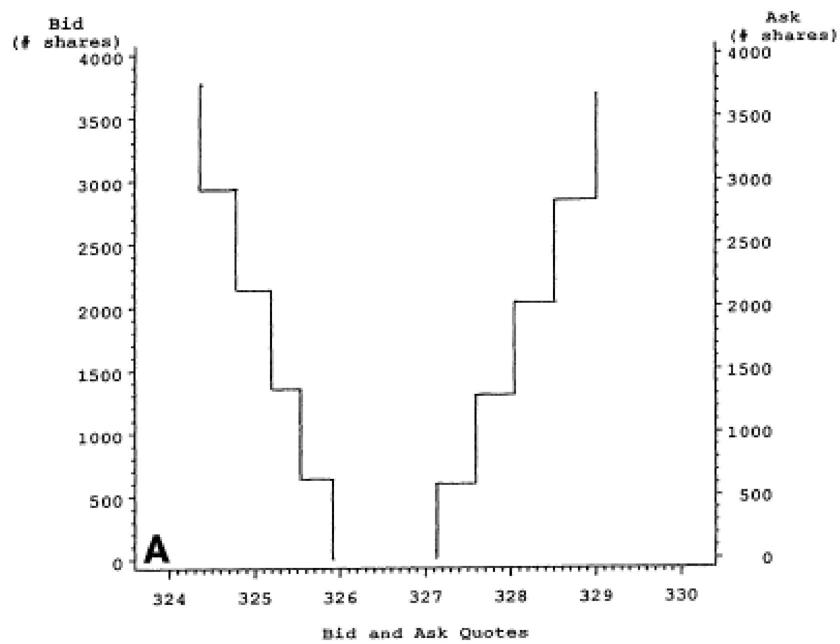

Figure 4.1: The cross-sectional average (across stocks) of the time series averages of the five best ask and bid quotes and their associated depth, Biais *et al.* (1995).

(1995). Orders can be classified according to their direction, aggressiveness and size. They cluster orders in different categories, for example a "large buy" which is an aggressive order larger than the volume behind the best ask. For each of these categories (large buy, small buy, large sell, small sell) the unconditional probabilities for the arrival of such an order in a given period of time are calculated using a data sample of stocks included in CAC 40 in 1991. In addition, the probabilities of orders and trades conditioned on the last action (order or trade) are calculated and can be written in a matrix form. This matrix shows an interesting diagonal effect, i.e. the probability of a given order or trade is higher after this event has just occurred. Furthermore, they try to connect further orders or trades with the current state of the order book by calculating order and trade probabilities conditioned by the state of the order book. Besides the probability of the occurrence of a certain event, they also provide an approach to predict the time interval between order and trade events.

The analysis (Biais *et al.* (1995)) provide very interesting empirical approaches to describe market microstructure aspects in limit order books. Knowledge about the probability of further events in the book can be used to calculate the execution probabilities of the own limit orders which can be used to optimize execution strategies.

4.3 Trading Costs

Trading costs will be discussed in more detail in part III of the current work. However, a brief overview is given here.

With its presence in the market, a buy or sell intention has an impact on the future price process. In stock markets buy and sell intentions are generally expressed by orders sent to the market. Perold (1988) introduces the implementation shortfall which is defined as the performance difference between the paper portfolio and the realized one. Implementation of investment strategies leads to friction losses. This difference in performance is dominated by three blocks of costs. One consists of fees and commissions for brokers and exchanges, the second part are market impact costs, and the third part are opportunity costs.

Market impact costs of a trade arise from the information acquisition and the demand of liquidity. It is a function of the aggressiveness of the trade, liquidity of the security and the amount of ordered shares. Market impact increases when trading large volumes in a short time span. Opportunity costs arise when less volume than originally wanted is traded because of the loss of profit. They also arise if a longer period of time is needed because of volatility risks. An investor has to find the trade-off between opportunity costs and market impact costs leading to optimal overall execution costs. See Kissell (2006) and Wagner and Edwards (1993) for further introduction in different kinds of trading costs.

Market impact is the most interesting trading cost component because of its complexity reflecting the interaction between one market participant and the market. Thus the realized execution price for a security is worse than the security price before the beginning of the trading activities of the investor. A possibility of measuring market impact is the calculation of the difference between the realized average execution price of a trade and the security price before the trading activity has begun (arrival price). The reasons for market impact are, as already mentioned, information acquisition and demand for liquidity. If an informed trader expects a higher price of the security in the future, he is willing to pay a higher price than the current one with the constraint that the price has to be lower than the expected price in the future. The investor's information is anticipated by the market resulting in market impact. The liquidity demanding component of market impact arises from the risk and costs the trading counterpart is faced with, see section 4.2. These effects differ in the sustainability of their impact, and while the information component is a permanent effect, the liquidity component is a temporary effect. Further description of market impact and the differentiation of temporary and permanent impact can be found in Kissell (2006), Kissell and Malamut (2005), Madhavan (2000) and Almgren and Chriss (1999).

4.4 Market Design

The design of the market determines the market microstructure. Thus, the market design is responsible for the quality and the success or failure of the trading venue. The microstructure influences investing strategies, patterns of trades, liquidity, and volatility. Therefore exchanges have to find their setup to attract traders. There are several studies in literature describing the impact of market designs on the market characteristics. Levecq and Weber (2002) and Stoll (2001) give a general overview of different possibilities how a market can be organized. Levecq and Weber (2002), Levecq and Weber (1995) and Barclay *et al.* (2001) focus on information technology and electronic systems in financial markets.

To evaluate the quality of trading at a certain exchange, quantities for market quality have to be defined. Madhavan (2000) mentions bid-ask spread, liquidity, and volatility. Others, for example Boehmer (2005), add availability and execution speed to the list of quality measures. The availability expresses the reliability of the exchange. The execution speed is the period of time an investor needs to get a trading decision executed depending on the size of the order. Also the reaction time is an important quality measure for some certain traders who are interested in ultra-high frequency trading, as it is described by Byrne (2007).

Market structure choices are elementary for exchanges to offer a market environment satisfying the investor's needs best in a competitive environment.

4.4.1 Market Architecture

"Market architecture refers to the set of rules governing the trading process", Madhavan (2000). These rules cover the market type including degree of continuity, choice between order-driven and quote-driven markets and also the degree of automation. Most stock markets are continuous trading systems combined with discrete auctions when the uncertainty is high. Most stock markets are organized as a mixture of order- and quote driven markets. Another aspect in market architecture is price discovery. There are several possibilities for price discovery processes. For example the already described process used in open limit order books, where the execution price is equal to the limit price of the involved limit order. Another example is the process used for example in closing auctions of most stock markets known as the Walrasian auctioneer. Some dark pools use another method where the midpoint of the bid-ask spread of the primary exchange is used as the execution price. Another important aspect is the transparency. Most stock markets provide pre-trade information such as quotes and related order sizes. Additionally, they provide post-trade information such as times and sales. This information can be used by an investor as a basis for trading decisions and execution optimization. Certain markets, such as dark pools,do not provide any market information except trading confirmations for directly involved trading parties. It is assumed that trading has less price impact if the order information is not published because other market participants cannot react on the presence of an order if it is not visible. More detailed information concerning market architecture can be found in Madhavan (2000).

Levecq and Weber (2002) focus on aspects of the market architecture of electronic trading systems. Electronic trading systems have their origin in the 1960s and 1970s with NASDAQ and Instinet. They have experienced strong growth up to now and dominate stock trading today. Two parallel evolutions occurred concerning electronic markets; there are the traditional markets such as NYSE using the electronic trading system to support their existing trading system. Automation helps to improve efficiency because it lowers trading costs. With the spread of electronic networks in the finance industry, the second type of markets have arisen. They are called ECN (electronic communication network) or MTF (multilateral trading facility). These trading platforms only concentrate on electronic trading mainly in liquid securities such as large caps and currencies. They provide very fast trading systems and charge low exchange fees. For institutional investors it is easy, and inexpensive to connect to an ECN. MTFs are established for years in the USA and cover a significant fraction of the trading volume in large caps. The ECNs in Europe currently experience a fast growth in trading volumes. Some important examples are Chi-X, BATS, and Turquoise. They have similar fee tariffs working as follows: market participants have to pay a fee for the execution of aggressive orders and they get a rebate for executed passive orders. This trading tariff concept of the ECNs attracts liquidity and therefore they do not need explicit market makers.

4.5 Fragmentation of the Market

Today we are faced with a widespread fragmentation of the stock market, in the US and also in Europe. Besides the primary markets, there are many ATSes (alternative trading systems) playing an important role. Figure 4.2 and 4.3 show the market share of the most important trading venues for DAX 30 stocks and FTSE 100 stocks. The market share of the primary market for the DAX 30 stocks is less than 70 % and for the FTSE 100 stock it is even less than 55 %. Thus, a significant fraction of the order flow goes to ATSes. This is quite similar to almost all important European stock markets. The US stock market also is fragmented where NYSE with its 3 platforms has 27.6 % and NASDAQ with its 2 venues has 24.5 % of market share. The third largest trading venue for stocks in the US is BATS with a market share of 10.6 % [1].

[1] source: businessweek March 02, 2010; http://www.businessweek.com/news/2010-03-02/bats-tops-direct-edge-to-become-third-largest-u-s-stock-market.html

This quite large fragmentation of these important stock markets is a quite new

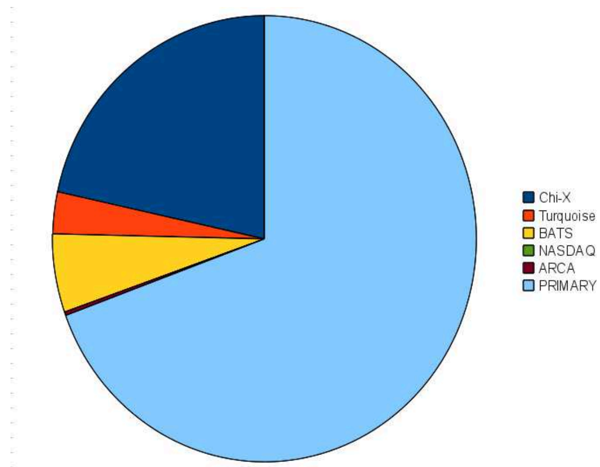

Figure 4.2: Market share of DAX 30 stocks, source: www.Chi-X.com

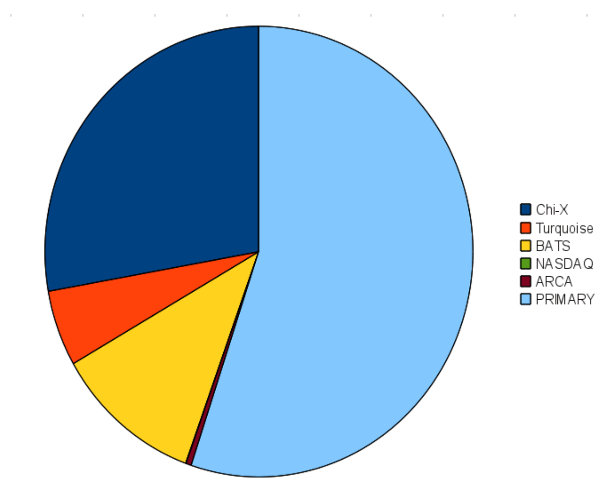

Figure 4.3: Market share of FTSE 100 stocks, source: www.Chi-X.com

development and is founded in changes of the market regulation in the US and also in Europe. The SEC[2] established the regulation of ATSes together with the Rule 3b-16 in 1998. This has implicated that price discovery is no longer a prerequisite for exchange status and has opened the market for trading venues. Additionally, the

[2]U.S. Securities and Exchange Commission

SEC requirement called NBBO (national best bid and offer) also takes the ATSes into account. Hence, marketable orders of clients have to be routed to the trading venue with the current best bid or offer to guarantee best execution. In Europe, MiFID was introduced in 2007. It regulates ATSes and encourages competition between trading venues. The preference rule and also the concentration rule for the primary markets have been replaced in many European countries. Thus the European stock markets have become harmonized and more open.

The main advantage of the "new" regulations of the stock markets and the fragmentation is the competition between the different trading venues. This has led to innovations, such as much lower exchange fees. Examples for the innovations are lower tick sizes in Europe, much faster trading systems and trading fee tariffs attracting liquidity providers.
Barclay *et al.* (2001) find that increased trading on ATSes improves most measures of overall market quality. As an explanation, they find that ECNs attract a higher fraction of informed orders reducing adverse selection costs faced by the market makers. This leads to lower spreads in competitive markets. Another effect of fragmentation is the lower level of trade disclosure. An investor trading large positions can benefit from this effect, see Madhavan (1995). In a consolidated market the effect of "front running" their own order can also be much more significant.

The main disadvantage of the stock market fragmentation is the reduced liquidity at the primary market and that it is more complex to trade large sizes in a fragmented market. There is some literature describing the effects of reducing liquidity by fragmentation of market, see Mendelson (1987), Chowdhry and Nanda (1991), Grossman (1992), Madhavan (1995) and Hendershott and Mendelson (2000). Bennett and Wei (2006) chose stocks which switched from listed on the NYSE to NASDAQ and vice versa. They measure the market quality before and after the switches and find that the NYSE has a better market quality than the NASDAQ for illiquid stocks. NYSE is one market where NASDAQ is a pool of different ECNs and exchanges, while NASDAQ is, in itself, a fragmented market.

As described above, there are opposing influences on market quality from fragmentation. Because of the interests of market participants to be well executed, there are forces in the direction of maximal market quality. Both extreme scenarios of a complete consolidation and also an extremely highly fragmented market are no optimal scenarios. In order to achieve good execution on fragmented markets, there are several ways of linking and consolidating them. One idea for quasi-consolidation is that every trade has to occur between the nationwide best bid and offer. If a marketplace does provide a worse price, the order has to be sent to another market with a better quote. Additionally, market participants can do pre-trade analysis to find out how to split the order and where to send it to have the best possible execution. Systems doing so are called "smart order routing" systems and are provided by most brokerage firms. In recent years also, many startups arise with the business idea of doing arbitrage by high frequent trading on different markets. These linkages of markets are a kind of consolidation with different impact on competition, see for example Blume (2007).

Chapter 5

Execution Strategies

The cost-efficient implementation of investment decisions is quite important for a successful realization of investment strategies. Depending on the frequency of the reallocation of the portfolio, trading costs can reduce the performance significantly. Especially large trading volumes cannot be executed instantly and the trade has to be split over a period of time. To this end, execution algorithms can be used. The current section wants to introduce some ideas behind common execution strategies.

5.1 Benchmarks

In order to measure the execution quality of execution strategies several measures can be taken into account: the executed fraction of the order, average execution price, and the execution price uncertainty. The most important measure is the execution price. This average execution price is usually compared to price benchmarks. These benchmarks can be categorized into pre-, intra-, and post-trade prices, see Kissell (2006). Very common price benchmarks are the VWAP benchmark (volume weighted average price) or the TWAP benchmark (time weighted average price) of the trading horizon. These benchmarks are so called intra-trade prices because security prices during the trading period are used for the calculation. Another benchmark such as the arrival price (price of the security before the arrival of the order) is a pre-trade price. An example for a post-trade benchmark is the day's closing price or the departure price (price of the security after the execution of the order). There are a variety of more benchmark definitions and also a spectrum of similar but slightly different definitions for each kind of benchmark, see for example Madhavan (2002) for various definitions for VWAP.

Different kinds of benchmarks have diverse characteristics, so investors have to take care by choosing their benchmarks with regard to their trading strategy and preferences. Pre-trade benchmarks are suitable for measuring market impact because they are not influenced by the price movement induced by their own trades. Measuring execution costs as part of the implementation shortfall, introduced by Perold (1988), has to be done by pre-trade benchmarks. Intra-trade benchmarks are a good indicator (see Berkowitz *et al.* (1988)) for the quality of the trading algorithm and market impact in the case of a quite passive execution strategy. Because the VWAP benchmark is heavily influenced by the trades if a market participant plays a dominant role on the market. In the limiting case of a completely dominant trader, the VWAP benchmark is equal to the average execution price, but the market impact is very high anyway. The intra-trade benchmarks have the advantage that the variance of the difference between the benchmark and the average execution price is smaller compared to many other benchmarks. That is because the intra-trade benchmarks contain the security price movement during the trading period, whereas pre-trade or post-trade benchmarks do not. The variance of the difference between intra-trade benchmarks and the execution price of a sample of algorithmic trades generally is significantly smaller than the differences between pre-trade (or post-trade) benchmarks and the execution price. Post-trade benchmarks aren't reasonable for measuring the market impact. But some investors or traders may desire the execution near the closing price for some reasons, (see Kissell (2006)).

Having a maximally objective view on the execution quality, several benchmarks should be taken into account. Only one benchmark is not able to represent execution quality as a whole.

A basic concept behind all execution benchmarks is the fact that trading is a zero sum game. The sum of all market impact costs of all market participants is zero which has to be considered by any measure of market impact costs. Otherwise the benchmark is biased and there are unexploited arbitrage opportunities, see Berkowitz *et al.* (1988).

5.2 Implementation of Execution Strategies

Domowitz and Yegerman (2005) introduce a spectrum of different execution strategies. They reach from unstructured, opportunistic liquidity searching to highly structured, precisely scheduled sequences of trading activity, generally linked to a certain benchmark. An example of a highly structured trading algorithm is the VWAP strategy. The unstructured strategies have the disadvantage that they generate in general large execution risks. Satisfying the investor's needs better and if some constraints have to be fulfilled, more sophisticated strategies are needed. This can be realized by using structured strategies and combining them with opportunistic components in order to achieve favorable prices during the constraints are fulfilled. Coggins *et al.* (2006) gives some introduction in algorithmic execution strategies, Obizhaeva and Wang (2005) provides the possibility of optimal execution, taking market dynamics into account.

5.2.1 Examples of Algorithmic Execution Strategies

Some examples of common execution strategies are presented in the following:

- The **Arrival Price** is the price of the security price at the moment before the first order is sent. The basic idea of execution strategies with this benchmark is to concentrate trading volume at the beginning of the trade, thus near the arrival price to minimize volatility risk. Minimization of volatility risk leads to fast execution and thus to high market impact. Every trader has to find his optimal point on the efficient frontier of the execution, introduced by Almgren and Chriss (1999).
  A more enhanced strategy is the adaptive arrival price strategy of Almgren and Lorenz (2007) where execution speed is updated in response to observed price motions leading to a better formulation of the mean-variance tradeoff.

- The **TWAP** execution strategy tries to reach the time weighted average price. Such a strategy divides the trading period in equally sized time slots and distributes the order volume equally over these slots. Thus, the same amount of shares is traded in each time slot.
  The TWAP benchmark is given by:

$$TWAP_i = \frac{1}{T}\sum_{t=1}^{T} p_{it} \qquad \{t\} \in T \tag{5.1}$$

  where p_{it} is the price of security i at time t.

- The **VWAP** trading strategy is very popular and is often used in the finance industry. The underlying benchmark is the volume weighted average price (VWAP) of the security i during a specified period T including all trades observed at the market with price p_{it} and size v_{it}.

$$VWAP_i = \frac{\sum_t^T v_{it} p_{it}}{\sum_t^T v_{it}} \qquad \{t\} \in T \tag{5.2}$$

  More detailed information and some variations of VWAP definitions can be found in Berkowitz *et al.* (1988) and Madhavan (2002).
  VWAP strategies work similarly to the TWAP strategy. The given time horizon where the trade ought to take place is divided in n (equal) sized time slots. In every time slot a certain fraction of the overall trading volume is executed. The executed volume per time slot divided by the overall trading volume in this security at the market should thereby be constant. Thus, how large the volume in each time slot is, depends on the historical trading volume of the special security in this time period taken as an estimation for the overall trading volume. Trading volume in equities is generally u-shaped over the trading day, i.e. in the first and in the last trading minutes, trading volume per time unit is extremely large and the minimum is around noon. A model predicting the trading volume is presented in part IV.
  Within a time slot, the algorithm may send limit orders to the market and then wait for execution at favorable prices. When the end of the time slot nears,

limits may become more aggressive and finally a market order is sent if the execution is forced.

- The **TVOL** (target volume) strategy is more opportunistic and trades a constant fraction of the actual overall trading volume in the security. Thus it is a modification of the VWAP strategy and only takes actual and not historic volume into account. There is no benchmark this strategy tries to beat. Before the beginning of the algorithmic execution, the overall trading volume and thus the duration of trading is not known.

Examples for opportunistic trading algorithms cannot easily be named because there is no industry standard. Using these algorithms is much more challenging because on the one hand they may provide lower execution costs, but on the other hand the handling of marginal constraints of the execution is more complicated.

One issue may arise especially when using schedule-driven algorithms. If the algorithm always acts very periodically, other market participants can observe patterns and take advantage of it. This leads to worse execution quality of the trading algorithm.

Comparisons between different execution strategies are available in literature. Kearns *et al.* (2004) compares one way algorithms as well as El-Yaniv *et al.* (2001). Yang and Jiu (2006) and Domowitz and Yegerman (2005) provide approaches for comparing different trading algorithms taking structure and performance into account.

Part III

Market Impact Measurement

Chapter 6

Introduction to the Market Impact Measurement of a VWAP Algorithm

The performance of mutual funds strongly depends on transaction costs. For high frequency hedge fund strategies with a large turnover, transaction costs thus play a crucial role. Very often the size of a fund is limited because too large sizes cannot be traded profitably: given the price predictions usually the market impact increases when trading volume becomes larger thus reducing the benefit of the strategy.

Transaction costs generally consist of two components: explicit costs including exchange and broker fees and also implicit costs such as market impact. Market impact is the interaction of a market participant's own activity on the market. In general the price observed at the beginning of a large trade is not equal to the actual execution price - on average the execution price is worse. This effect plus the explicit costs (fees) is also well known as implementation shortfall and discussed for example by Demsetz (1968) and Perold (1988).

The current analysis is an empirical analysis of the market impact of a homogeneous set of algorithmic trades in the stock market from April 2008 to July 2010 on Canadian, European, Japanese, and US stocks. It is done with the help of a proprietary data set originating from real trading activity of the Lupus alpha NeuroBayes®Short Term Trading Fund. This data set covers more than 2 years of trading activity on over 800 stocks in various countries (Europe, USA, Canada, and Japan) and a trade volume of over 30 billion USD equivalent. The main advantage of this data set compared to publicly available ones is given by the fact that single orders of one market participant are identified and connected. Because of that, the resulting market impact of the entire transaction can be measured. Without connecting single orders, the orders usually are assumed to be independent which is obviously not true. In that case, the trading strategy of a market participant cannot be reconstructed. So the characteristics of the proprietary data set enables me to provide rare empirical measurements to verify theoretical considerations.

The used trading algorithm is a so called VWAP (short for volume weighted average price) trading algorithm combined with a smart order router[1]. Its objective is the execution of a algorithmic order within a given time at minimal execution cost. It is realized by splitting up the size of the algorithmic order over the trading period according to the entire trading volume profile. It is shown that the VWAP trading algorithm is the strategy which produces the lowest market impact when taking the current market impact models as a basis.

The current analysis is an approach motivated by the needs of a practitioner. It provides several models describing the dependency between the market impact and some description variables. I found the participation rate as the by far most important variable to describe the market impact. As participation rate the ratio between the algorithmic order size and the entire number of traded shares in the respective period of time is defined. The first market impact model is linear in the participation rate whereas the second one is based on a power law. The explanation power of the linear model is improved with the introduction of an individualized linear regression. By doing so, the linear regression parameters (slope and intercept) are dependent on additional exogenous variables. For this purpose the linear model is taken: when this route is taken for a portfolio optimization, it is much easier to find an optimum in a multidimensional space since the model for relative transaction

[1]The purpose of the smart order router is to find the best trading venue in consideration of low exchange fees and a good execution price.

costs is linear in order size. In contrast to arbitrary functions there are very efficient and fast optimization algorithms, such as Simplex and Gauss-Newton for linear and quadratic optimization functions.

The model based on a power law is motivated by the results of the microstructure theory and by a slightly concave curve (figure 8.2) observed in the data especially for a wide range of participation rates.

The different markets behave quite similar in many aspects, but there are significant differences between the Japanese market and the remaining markets. This can partly be explained by different regulations (such as the up-tick-rule in the Japanese market and different tick size definitions).

Additionally, it can be shown that the VWAP trading strategy is the optimal execution strategy taking the results of my market impact models into account.

Chapter 7

Description of the Data Set

7.1 Trading Specifications

The underlying data sample of the current analysis contains all relevant informations within a set of about 120.000 algorithmic trades from April 2008 to July 2010. The data originates from the trading activity of the Lupus alpha NeuroBayes®Short Term Trading Fund[1]. Its stock universe consists of stock with a large market capitalization (large caps) from Europe, USA, Japan and Canada. More precisely, the universe consists of the 500 most liquid stocks in the USA. They are mainly covered by the S&P500 index. The 250 most liquid European stocks belong to the universe as well and also the 110 most liquid Japanese stocks and about 200 Canadian stocks. The investment strategy is based on statistical arbitrage on a day to day basis.

The implementation of the investment strategy is realized with the help of a trading algorithm. For trade execution, the orders are split over a given time period and are executed incrementally, since large orders cannot be executed at once at an attractive price (due to finite liquidity in the order book). The trading algorithm works on the basis of a VWAP trading strategy. This means that the trading volume of the algorithmic order is distributed over time, weighted by the entire trading volume. Further details can be found in Fränkle and Rachev (2009).

All analyzed trades have the same trading period. In the US and Canadian

[1] http://www.ise.ie, Sedol: B1HMBP7

markets, this period encompasses in the last 15 minutes of the official trading hours, i.e. from 3.45 pm to 4.00 pm ET. The trades in the Japanese market are executed in the last 25 minutes of the trading day. The trading of European stocks is entirely different in the sense that there are closing auctions with significant trading volumes. So quite a large fraction of the algorithmic orders is executed in the closing auction. However, the execution of the algorithmic orders starts during the continuous trading session about 20 minutes before the closing auction begins.

The used trading algorithm tries to execute as much as possible with the usage of passive limit orders in order to reduce the market impact and explicit transaction costs.

7.2 Market Impact Definition

Market impact is defined as the interaction of the investors own order with the market, i.e. how large is the price change influenced by this order? Therefore the quantity "market impact" is described by a price difference between a benchmark price, which should as little as possible be influenced by the own order, and a price incorporating the full impact. In the current analysis, I take the relative price change r_s between the average execution price P_{vwap} and the arrival price $P_{arrival}$

$$r_s = d \cdot \frac{P_{vwap} - P_{arrival}}{P_{arrival}} \tag{7.1}$$

where

$$P_{vwap} = \frac{\sum_i p_i \cdot q_i}{\sum_i q_i} \tag{7.2}$$

The direction d is 1 for buys and -1 for sells, p_i is the execution price of the i-th partial fill and q_i is the corresponding size. The arrival price $P_{arrival}$ is the current stock price ultimately before the order arrives at the exchange. The average execution price is the volume weighted average price (vwap) of all transactions of the specific order during the trading period.

Note, however, that this relative price difference contains also the externally triggered price movements which are not part of the market impact of the own order. So the price change r_s can be written as a sum of two components:

$$r_s = r_e + I \tag{7.3}$$

As mentioned above, one component is the stock price move r_e which is induced by external influences. The second component is the market impact I of the own trade which is analyzed here. To get rid of this effect the market impact definition is modified in equation 7.1 as the mean value of the distribution of r_s (averaged over different stocks and days). The advantage of this definition is that the mean value of the r_e distribution is 0, so the mean value of the r_s distribution is an unbiased estimator for the empirical market impact. The reason why this assumption is reasonable, is market neutrality (dollar and beta[2] neutral) of the fund's investment strategy. Therefore the trades are also market neutral and market movements do not affect $< r_s >$[3].

For these reasons the market impact of a trade can be defined as:

$$< r_s > = < r_e > + < I > = < I > \tag{7.4}$$

without having a bias in the data. Although the externally induced return does not contribute to the average impact $< I >$, it dramatically increases the variance of r_s. Therefore the width of the distribution of r_s is dominated by the external induced price movements. Hence the market impact can only be measured significantly with enough statistics.

The r_s distributions for the different markets are shown in the figures 7.1, together with the statistics in table 7.1. It is quite notable that the distribution of the transactions in the Japanese market has a strong peak at 0. This peak is explained by large tick sizes for many stocks, leading to a higher probability of unchanged stock prices. The average tick size over the execution price, measured for the European universe, is 4.7 BPS, for the US stocks 3.7 BPS and for the Japanese ones 18.0 BPS.

[2] beta factor, known from the CAPM (Capital Asset Pricing Model)

[3] $< x >$ mean value of a set of numbers $x_1, x_2, x_3, ..., x_n$

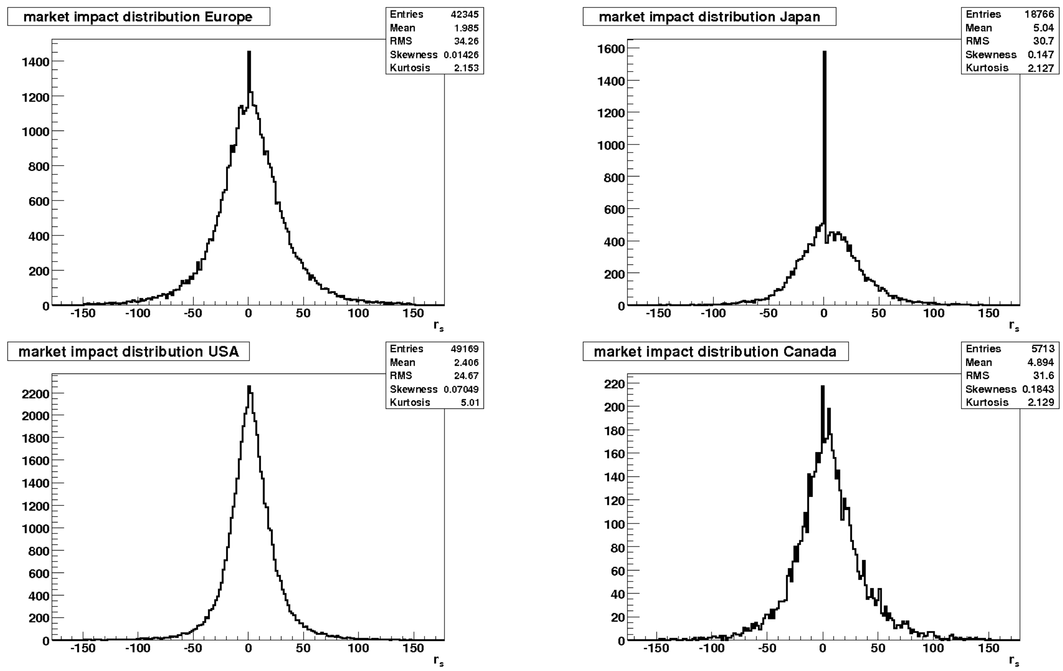

Figure 7.1: Distributions of the price change r_s for the various markets, top left to bottom down: Europe, Japan, USA, and Canada.

7.3 Comparison of Sell and Short-Sell Trades

For all sell transactions in the Japanese and US market the data provides the information whether the order was a long-sell or a short-sell. As explained in section 7.2 there is no bias in the average impact because $< r_e >= 0$ in equation 7.4. This is not the case if taking only subsets of the trades into account such as buy or sell orders. It is still reasonable to look at the difference between sell orders and short-

Table 7.1: Statistics of r_s distributions

	# trades	mean	error of mean	RMS
Europe	42345	1.99	0.166	34.26
Japan	18766	5.04	0.22	30.7
USA	49169	2.41	0.11	24.67
Canada	5713	4.89	0.42	31.60

sell orders. Although there is a nonzero $< r_e^{\text{(short-)sell}} >$ as shown in figure 7.2 the expected return of the market can be approximated to be the same for sell orders and short-sell orders:

$$< r_e^{\text{sell}} > \approx < r_e^{\text{short-sell}} > \tag{7.5}$$

Figure 7.2 shows the impact of trades in the Japanese and the US market for long-sell and short-sell orders. The US equities' impact distribution differs not statistically

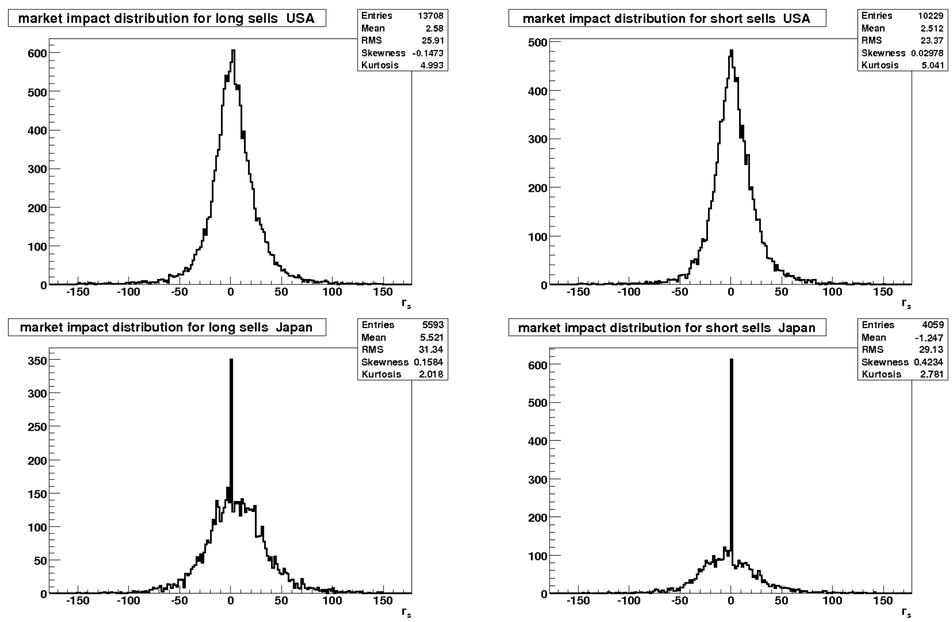

Figure 7.2: Comparison of the r_s distribution of long-sells and short-sells.

significant for long-sells and short-sells because the mean, standard deviation, skewness and kurtosis of both distributions cannot be distinguished by statistical tests. This is not the case for the trades in the Japanese market where the means of both distributions are significantly different [4]. The difference between the Japanese and the US market is due to a Japanese market rule, the so called 'up-tick rule'. For the rule to be satisfied, the execution price for short-sells must be equal or above the last traded price of the security. So there is a bias in the execution of short-sells in getting better prices for short-sells. This comes for the price of a lower execution

[4]The statistical error of the mean σ_μ is defined by $\sigma_\mu = \sigma/\sqrt{n}$ where σ is the standard deviation of the impact distribution and n is the number of observations.

probability.

Chapter 8

Market Impact Model

The dependencies between the observable description variables and the market impact is analyzed. As the most important variable, the participation rate is identified. It is defined as the ratio between the size of the algorithmic order of which the market impact is measured, and the entire number of shares of the security traded in the same time period. The profile plot[1] depicting market impact over participation rate suggests a slightly concave curve, as it is also observed by Almgren *et al.* (2005). As mentioned earlier, I provide two alternatives to explain this relation between participation rate and market impact. The first proposal is a simple linear model and the second is a power law model.

8.1 Linear Model

Due to the technical reasons, a linear model may be preferable for some applications (see for example chapter 9). Additionally, it has the advantage that it can easily be implemented in a portfolio optimization algorithm without increasing the complexity of the problem (see appendix B). It also can be motivated by the fact that it is the first term of a Taylor expansion and a good approximation for a small range of the

[1] average impact per bin of participation rate, see appendix A

participation rate. The linear model is given by the following function

$$M(v) = m \cdot v + b \tag{8.1}$$

where M is the market impact and v the participation rate. m and b are the parameters of the model.

The maximum likelihood method is used for the parameter estimation and an asymmetric Laplace-distribution for the residuals r.

$$r(x) = \frac{\tau(1-\tau)}{\sigma} \cdot e^{-\rho_\tau(\frac{x-\mu}{\sigma})} \tag{8.2}$$

where ρ_τ is given by

$$\rho_\tau(u) = \frac{|u| + (2\tau - 1)u}{2}. \tag{8.3}$$

The parameters of the linear model (equation 8.1) m and b are estimated separately for every market (Europe, Japan, USA, Canada) because it is reasonable to assume that the different market characteristics lead to different market models. This effect can be observed in the fit results (see table 8.1 and figure 8.1).

Table 8.1: Fit results of the linear model

	scale, m	intercept, b
Europe	55.924 ± 8.837	0.938 ± 0.197
Japan	123.377 ± 9.810	2.235 ± 0.394
USA	77.827 ± 6.149	1.797 ± 0.129
Canada	114.02 ± 22.890	0.997 ± 0.280

8.2 Power Law Model

The various profile plots which show market impact over participation rate, suggest a slightly concave model, which is also in line with the existing literature (see for example Almgren *et al.* (2005)). Especially for larger ranges of the participation rate, the concave model fits the observations much better. Therefore a power law

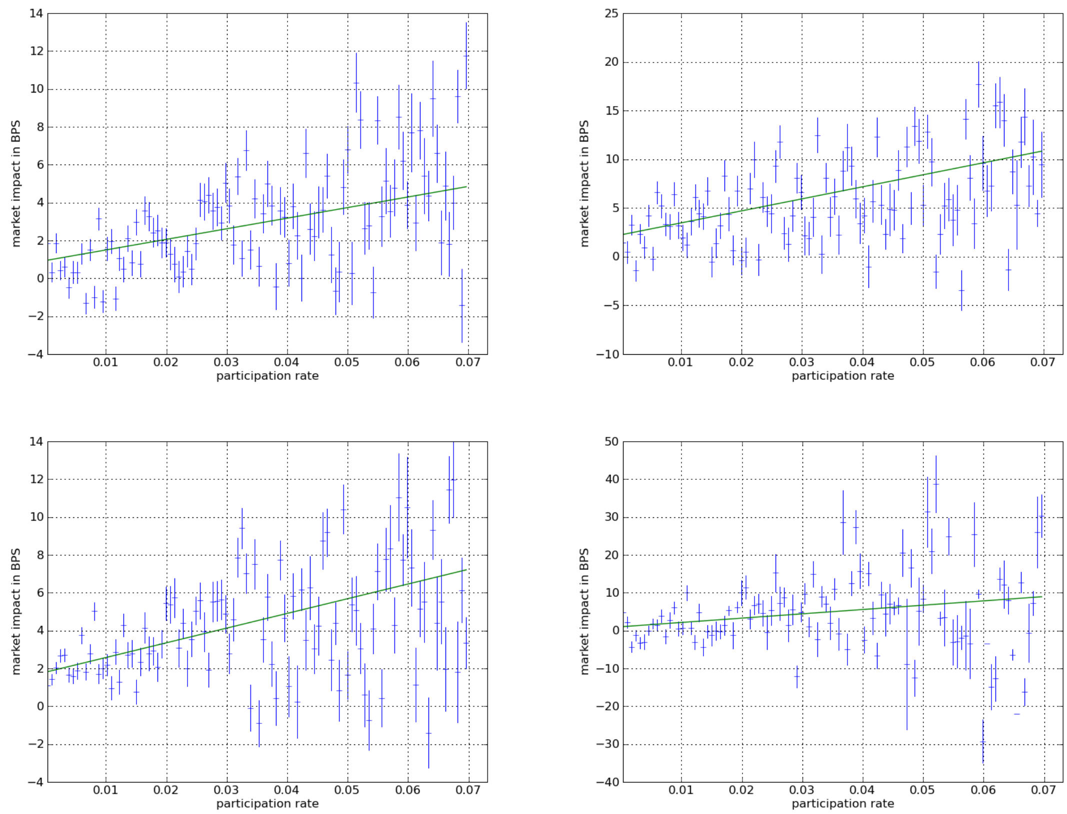

Figure 8.1: Linear market impact model, market impact over participation rate; top left to bottom right: EU, JP, USA, CN.

model is proposed, defined by:

$$M(x) = m \cdot v^a + b \tag{8.4}$$

Here, two slightly different interpretations of the power law model for the market impact are suggested for the different markets. The first approach is similar to the one of the linear model, where all parameters are estimated separately for each market. The results of this fit procedure can be found in table 8.2.

The second approach is different in the sense that the parameter m and b are estimated separately for each market, but the exponent a is estimated together for all markets. This has the advantage that the complete data sample can be used to

Table 8.2: Fit results of the power law model, each market fitted separately.

	scale, m	exponent, a	intercept, b
Europe	12.992 ± 8.362	0.437 ± 0.222	-0.247 ± 0.947
Japan	40.275 ± 24.602	0.511 ± 0.208	-1.112 ± 1.427
USA	30.320 ± 21.972	0.686 ± 0.210	1.231 ± 0.374
Canada	20.061 ± 15.685	0.423 ± 0.241	-0.221 ± 1.327

estimate the parameter, leading to lower statistical uncertainty of the fit results. The likelihood function of this parameter estimation can be written as

$$L = \prod_i L_i\left(\vec{v}_i, \vec{y}_i, \vec{p}_i\right), \tag{8.5}$$

where i represents the different markets (EU,US,JP,CN). The parameter set of market i is given by $\vec{p}_i = (m_i, a, b_i)$. L_i denotes the likelihood function for one market which can be written as

$$L_i\left(\vec{v}_i, \vec{y}_i, \vec{p}_i\right) = \prod_{j=0}^{N} r(M(v_i^j) - y_i^j) \tag{8.6}$$

where $\vec{v}_i$ is the sample of participation rates and $\vec{y}_i$ are the price changes in market i. $M(v)$ is defined by equation 8.4 and $r(x)$ is the residual distribution, see equation 8.2.

Furthermore, it seems reasonable to take the same exponent for more than one market, because the estimated exponents are quite similar in all 4 markets (see table 8.2). They do not differ significantly on a 95 % significance level, when using the error propagation on the difference of the estimated parameters with their uncertainty[2].

The results of the combined fit including all 4 markets (EU, US, JP, CN) can be

[2] The difference of the estimated parameters for different measurements is calculated. The error of the difference can be estimated with the help of error propagation, see equation 8.7, 8.8 and 8.9.

$$\mu = \mu_1 - \mu_2 \tag{8.7}$$

$$\sigma^2 = \left(\frac{d\mu}{d\mu_1} \cdot \sigma_1\right)^2 + \left(\frac{d\mu}{d\mu_2} \cdot \sigma_2\right)^2 \tag{8.8}$$

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 \tag{8.9}$$

found in table 8.3. The exponent is, by definition, the same for all 4 markets and is estimated as 0.534 ± 0.115.

It is notable that there is evidence for a negative intercept of the Japanese market. A negative intercept in this model does not make sense, because this means that very small trades create negative costs which would imply the possibility of arbitrage. This fact and the a priori knowledge that the Japanese market behaves differently than the other markets with respect to regulations (such as up-tick rule (see 7.3)) motivates me to modify the fit procedure and remove the Japanese market. The effect of the negative intercept can be explained by the up-tick rule because short-sells have an execution probability significantly lower than 1. So the up-tick rule which affects only the short-sells in the Japanese market leads to a bias which can be explained as follows: If the execution price is higher than the arrival price, a very high percentage of the short-sells should be executed and the measured market impact for these trades is negative. If the execution price is lower than the arrival price, the execution probability is worse (fewer shares are traded) and the measured market impact is large. So there exists a bias towards lower market impact in the Japanese market using the current method to estimate market impact. In spite of the knowledge about this effect, there is no obvious solution to circumvent this problem and integrate the Japanese data in the analysis. Assuming removal of the short-sells for the Japanese market, this bias would be lost, but another bias may appear: the assumptions for the negligence of the market movement in section 7.2 would be hurt.

The results of the fit with the European, US, and Canadian market can be found in table 8.4. Figure 8.2 shows the corresponding plots.

Comparing the statistical uncertainties of the parameter estimations of table 8.1

Table 8.3: Fit results of the power law model Europe, Japan, USA and Canada together

	scale, m	exponent, a	intercept, b
Europe	15.872 ± 6.017	0.534 ± 0.115	0.095 ± 0.521
Japan	42.939 ± 15.577	0.534 ± 0.115	-0.980 ± 0.680
USA	18.996 ± 7.160	0.534 ± 0.115	0.957 ± 0.344
Canada	26.816 ± 11.432	0.534 ± 0.115	0.256 ± 0.807

on the one hand and tables 8.2, 8.3, and 8.4 on the other hand, it is conspicuous that

Table 8.4: Fit results of the power law model Europe, USA and Canada together

	scale, m	exponent, a	intercept, b
Europe	16.345 ± 7.500	0.547 ± 0.143	0.131 ± 0.557
USA	19.801 ± 9.412	0.547 ± 0.143	0.982 ± 0.389
Canada	27.774 ± 13.726	0.547 ± 0.143	0.302 ± 0.860

the scale parameters of the first table are estimated with more statistical significance than in the remaining tables. This is explained by the error estimation method of the parameters and their correlations. The model parameters scale, intercept, and exponent are correlated. By slightly varying one of the model parameters, a solution for the remaining parameters can be found describing the data set almost as good as the optimal solution. The errors of the parameters which are estimated by the maximum likelihood procedure can be estimated by varying one parameter until the likelihood function rises by 0.5. During the variation of this parameter, for all other fit parameters, the maximum of the likelihood function has to be found (see Blobel and Lohrmann (1998), p. 189-191). It have been done a thorough analysis of the regression errors and have been rather conservative in the error estimate. To convince the reader of the significance of the findings two Null-Hypotheses are tested: keeping on the one hand the intercept equal to 0, the slope is significantly different from 0. And likewise on the other hand if the slope is kept equal to 0, the intercept is significantly different from 0.

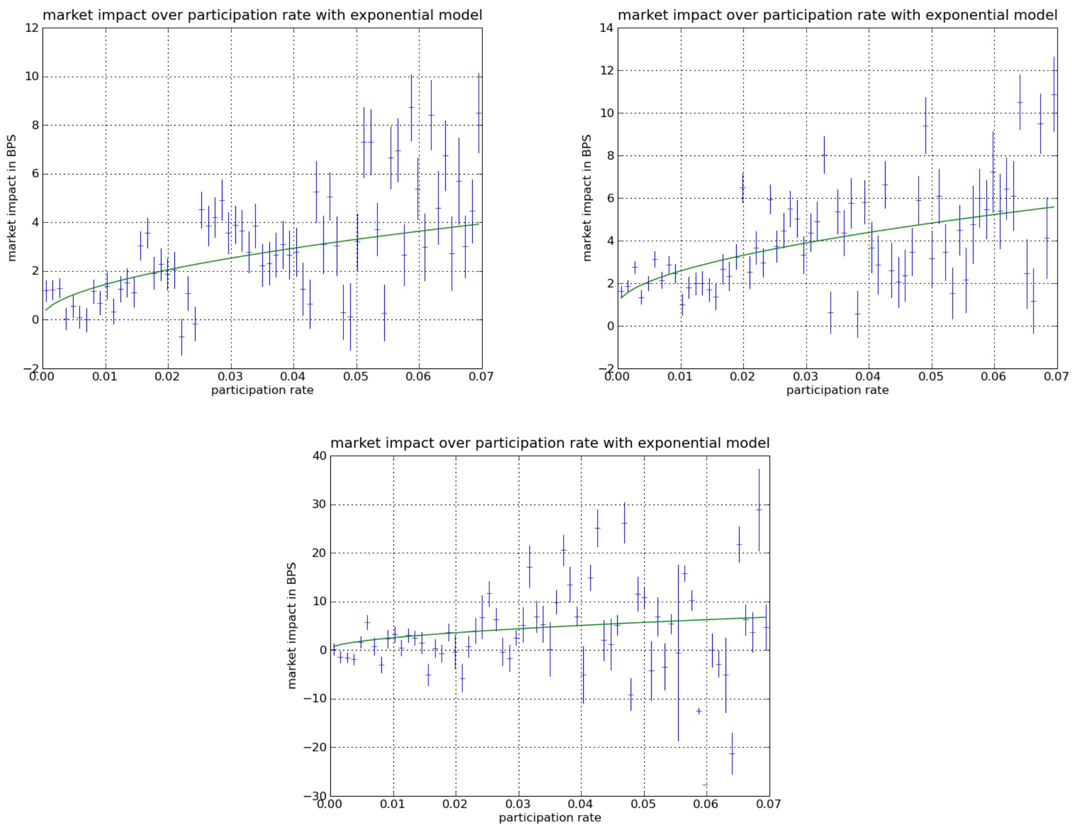

Figure 8.2: Power law market impact model, market impact over participation rate; top left to bottom right: EU, USA, CN.

Chapter 9

Individualized linear regression analysis

In this section the simple linear regression model is improved by using an algorithm described in Scherrer *et al.* (2010). The algorithm is an advanced linear regression model in which the slope and the intercept of the regression are allowed to be dependent on every single event. This model applied to the impact analysis generalizes the equation of the simple linear regression

$$t_i = mv_i + b + \epsilon_i \tag{9.1}$$

$$t_i = m(x_{1,i}, ..., x_{n,i})v_i + b(x_{1,i}, ..., x_{n,i}) + \tilde{\epsilon}_i \tag{9.2}$$

where ϵ_i is the residual of event i and v_i is the volume fraction. The slope m and the intersect b are not constant any more and can depend in a nonlinear way on the additional description variables $x_1, ..., x_n$.

This ansatz is reasonable because already the simple linear regression describes the impact quite well but it is interesting to understand the corrections to the linear model with respect to some external variables such as volatility of the specific stock, tick sizes, market capitalizations etc.

It is used the individualized linear regression instead of the power law for some practical reasons. A market impact model may be used in a portfolio optimization performed by a trader. The trader takes the predictions for the stock market returns into account, but his trades in turn will influence the stock returns. Thus, a portfolio optimization has to be fast and the solver for the optimization problem is much faster for a linear impact model.

In the next two sections the basic idea of the algorithm will be presented and also the results and improvements compared to the simple linear regression model.

9.1 The individualized linear regression algorithm

The details of the individualized regression analysis are described in Scherrer *et al.* (2010).

The first step is to transform the input variables $x_1, ..., x_n$ to be uniformly distributed. This means, by definition, a histogram of the specific input variable has the same amount of events in every bin. In the next step the input variable is divided into k bins. The parameters m and b and their errors are estimated for every bin of the input variable. In order to make the algorithm robust against statistical fluctuations a spline fit is used additionally to smoothen the dependencies of m and b on the specific input variable. This procedure is done for all input variables.

For one event i there is one prediction for m_i and b_i for each input variable. That means that there are n predictions for m_i (b_i). It is required to end up with **one** prediction only for m_i (b_i) of a certain specific event. The easiest ansatz would be to average the m's and the b's to get

$$m_i = \frac{1}{n}\sum_{j=1}^{n} m_j \qquad b_i = \frac{1}{n}\sum_{j=1}^{n} b_j \tag{9.3}$$

But this choice is not optimal. The prediction coming from a variable with a high correlation to the target t_i should have a larger weight than the prediction coming from a weakly correlated variable.

A problem could also appear if vector $\vec{x}$ is introduced in which all the components are highly correlated to each other. The algorithm should recognize such correlations and make sure that the statistical significance of the correlation between the input

Table 9.1: Input variables for the individualized linear regression analysis

input variables
market-return-arrival-close
dir-market-return-arrival-close
dir-market-return-eq
vola
rel-ts
liquidity
t
volume-fraction
market-cap

variables and the target is not increased by introducing further redundant variables which are highly correlated to the rest of the variables.

Obviously, I would like to use an algorithm which can deal with correlations among the input variables and which is able to decide if a variable has a statistically significant correlation to m (b) at all. If there is a large correlation of a variable and m (b), the weight of the estimator should be larger than the weight given to an unimportant variable. And if the input variables are correlated among each other the algorithm should treat these correlations correctly.

For this kind of problem the NeuroBayes®software [1] can be used which is described in Feindt (2004). The n predictions for m_i, the n predictions for b_i and the variable v_i as input vector (details see Scherrer *et al.* (2010)) are used. The target is defined by the execution price.

9.2 Input variables of the individualized linear regression

As there is not enough statistics for the Canadian market, there is only taken the Japanese, the European and the US market into account.

To understand the underlying dynamics of execution price and impact, some appropriate variables (see table 9.1) are introduced and the improvements to the simple linear regression model are analyzed. This analysis should describe all dependencies

[1] Developed by Phi-T® Physics Information Technologies GmbH

of the parameters m and b on the input variables. Therefore, additional variables will be used which does not include only past information, but also variables which include information on the future.

The execution price for a specific stock is firstly dependent on the return of the stock which would have taken place without the order of the market participant and secondly on the impact of the order, according to equation (7.3).

Note that the return of the stock which was traded must not be used as input variable, because the return already includes its impact. Therefore it is not reasonable to explain the impact using an input variable which already includes the impact. However, the impact of the order on the underlying stock market index (EuroStoxx 50 for Europe, S&P 500 for the USA, Nikkei for Japan) can be neglected and instead an input variable is defined using the index return.

The following variables for the model are defined:

- "market-return-arrival-close" denotes the return of the stock market index in the trading period
- "dir-market-return-arrival-close" denotes the return of the stock market index multiplied with the direction the stock was traded (+1 for a buy and −1 for a sell)
- "dir-market-return-eq" is defined by the stock market index return in the trading period multiplied with an estimator of the beta-factor[2]. The beta-factor is estimated from historical data and does not include information of the future.
- "vola" is an estimator for the volatility
- "rel-ts" is the relative tick size of the stock (tick size divided by the arrival price when the trading period begins)
- "liquidity" is defined by the traded volume in Euro (for the specific stock) at the trading day

[2]The factor β is defined for an asset i in the CAPM (see Sharpe (1964)) as $\beta_i = \frac{cov(r_i, r_M)}{\sigma^2(r_M)}$ where r_i is the return of the asset i and r_M is the return of a market portfolio.

- The variable “t” is introduced reflecting the date to account for an explicit time dependency of the parameters m and b.
- The variable “volume-fraction” includes the relative traded Euro-volume of the specific stock compared to all stocks of the current universe at that day for the market in which the stock is traded (Europe, Japan, USA).
- Finally for the market capitalization the variable “market-cap” is introduced.

9.3 Results of the individualized linear regression

My findings suggest that the variables “market-return-arrival-close”, “liquidity”, “volume-fraction”, “market-cap” and “t” do not have any significant correlations to the parameters m and b.

Due to the fact that the relative frequencies of buy and sell orders are equal, it is not surprising that this variable is not important. Much more important is the return weighted with the trading direction (“dir-market-return-arrival-close”). The basic idea why this variable has been introduced is that the return of the asset is generally correlated to the market portfolio. The influence of the market participant on the market portfolio is negligible, so this variable should be a good estimator for the return of the asset in the trading period as it contains information on the future. It is included in the analysis to understand the underlying market components which influence the execution price, but it cannot be used for prediction.

All information of the variables “liquidity”, “volume-fraction” and “market-cap” is completely absorbed in the participation rate.

The most important effect in all markets investigated is a high correlation of the parameters m and b to both the index-return weighted with the trading direction (see figure 9.1) and to the index-return weighted with the trading direction and the beta-factor (see figure 9.2). The dependency of the b-parameter on these two parameters is much more significant than the dependency of the m-parameter.

In the algorithm the last bin of the plots has a special meaning (see Scherrer *et al.* (2010)): If some input variables are not known or believed to be wrong for some events, one can activate a special flag for these events. Hence there are two

Table 9.2: Structure of the tick sizes in Europe.

price of share	tick size
0 EUR - 9.999 EUR	0.001 EUR
10 EUR - 49.995 EUR	0.005 EUR
50 EUR - 99.99 EUR	0.01 EUR
100 EUR - ∞	0.05 EUR

different possibilities:

The first possibility is to define a variable that is not known (or not correct) for all events. Consequently those events which are not known (or incorrect) are separated in the last bin. The parameters m and b are then estimated for this input variable bin and afterwards not included in the spline fit[3]. The second possibility is that there are no events in the training sample which are filled in the last bin. Then the estimator of m (b) is defined by the mean of all other bins.

If the user has adjusted the parameters of the prediction model on historical data and would like to use the results to forecast an event in which the variable is not known (or wrong), the estimator of the special bin is used.

In figure 9.3 it can be seen that the parameter b is significantly correlated to the relative tick size but only in the US and in the Japanese market. A possible reason could be that the definition of the tick size in Europe is relative to the price level (see e.g. the tick size structure at XETRA in table 9.2) whereas in the USA tick size is absolute 1 cent and constant for all stocks. The tick size definitions in the Japanese market lead to extremely large relative tick sizes (tick size over stock price) for some stocks. This is also valid for US stocks with a low absolute stock price. It is different to the European market where the tick size depends on the price of the stock (the rules are similar for all European exchanges). This leads to quite small relative tick sizes for all European stocks. Maybe the European tick size definition is responsible for the independence of tick size and execution price.

The parameters of the model are also slightly dependent on the volatility but only in Japan and in the USA (figure 9.4). While the parameter m is fairly constant

[3]It has a completely different meaning compared to the rest of the bins, so the assumption that m and b are smoothly depending on the input variable is not valid for the last bin.

Table 9.3: Comparison of the results coming from the simple linear, the individualized linear regression model, and the model based on power law.

	simple linear regression	individualized linear regression	relative improvement of ind. lin. reg. compared to simple lin. reg	power law (separate markets)	power law, markets fitted together
mad_{EU}	25.114	22.168	11.7%	25.116	25.115
σ_{EU}	34.608	31.053	10.3%	34.613	34.612
mad_{US}	16.307	14.726	9.7%	16.308	16.307
σ_{US}	23.657	21.910	7.4%	23.655	23.655
mad_{JP}	22.634	21.014	6.2%	22.534	xxx
σ_{JP}	30.529	28.513	6.6%	30.583	xxx
mad_{CN}	22.813	xxx	xxx	22.035	22.034
σ_{CN}	31.510	xxx	xxx	30.579	31.575

and independent of the volatility, the b parameter is correlated to the volatility.

As mentioned earlier the goal of the individualized linear regression applied in this chapter is to find the importance of the underlying factors which are responsible for the impact. The mean absolute deviation (mad) and the standard deviation σ, which are defined as follows, are compared:

$$mad = \frac{1}{N}\sum_{i=1}^{N} |m(x_{1,i}, ..., x_{n,i})v_i + b(x_{1,i}, ..., x_{n,i}) - r_{s,i}| \tag{9.4}$$

$$\sigma = \frac{1}{N}\sum_{i=1}^{N} (m(x_{1,i}, ..., x_{n,i})v_i + b(x_{1,i}, ..., x_{n,i}) - r_{s,i})^2 \tag{9.5}$$

In table 9.3 the results of the simple linear regression and the individualized linear regression are summarized. The last column is the relative improvement of mad and σ if the individualized linear regression is used. For the European market the simple linear model can be improved approximately by 11%, while the US market and the Japanese market are improved by 8.5% and 6.5% respectively. In the current analysis I found that this effect is mainly based on the return of the underlying stock market index as long as there is a dependency of the relative tick size and the volatility in the US and the Japanese market.

If a trader would like to use a market impact model for the portfolio optimization, he would have to estimate the volatility and the volume which will be traded during the trading period. These input parameters can be estimated quite accurately from historical data. The relative tick size is also known before the trading period.

I have thus explained that the beta weighted return of the underlying stock market index during the trading period is extremely important for the impact. This variable decreases the variance of the residuals quite dramatically. But it is a problem to estimate the return of the stock market index during the trading period a priori. A trader can either have a mathematical model for the index return and utilize this for the impact model. Alternatively he could relinquish the variable at all, which would lead to a larger variance of the distribution of the residuals.

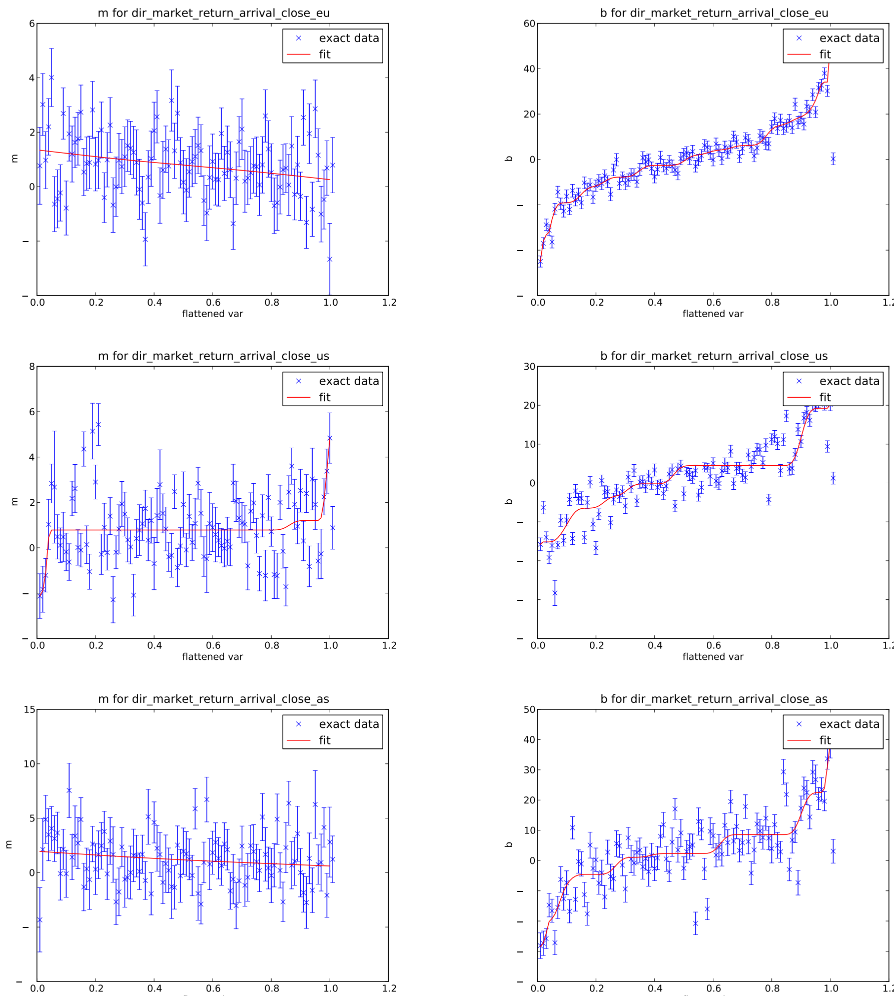

Figure 9.1: Correlations of the parameters m and b to the stock market return of the underlying index (see text) weighted with the direction of the trade.

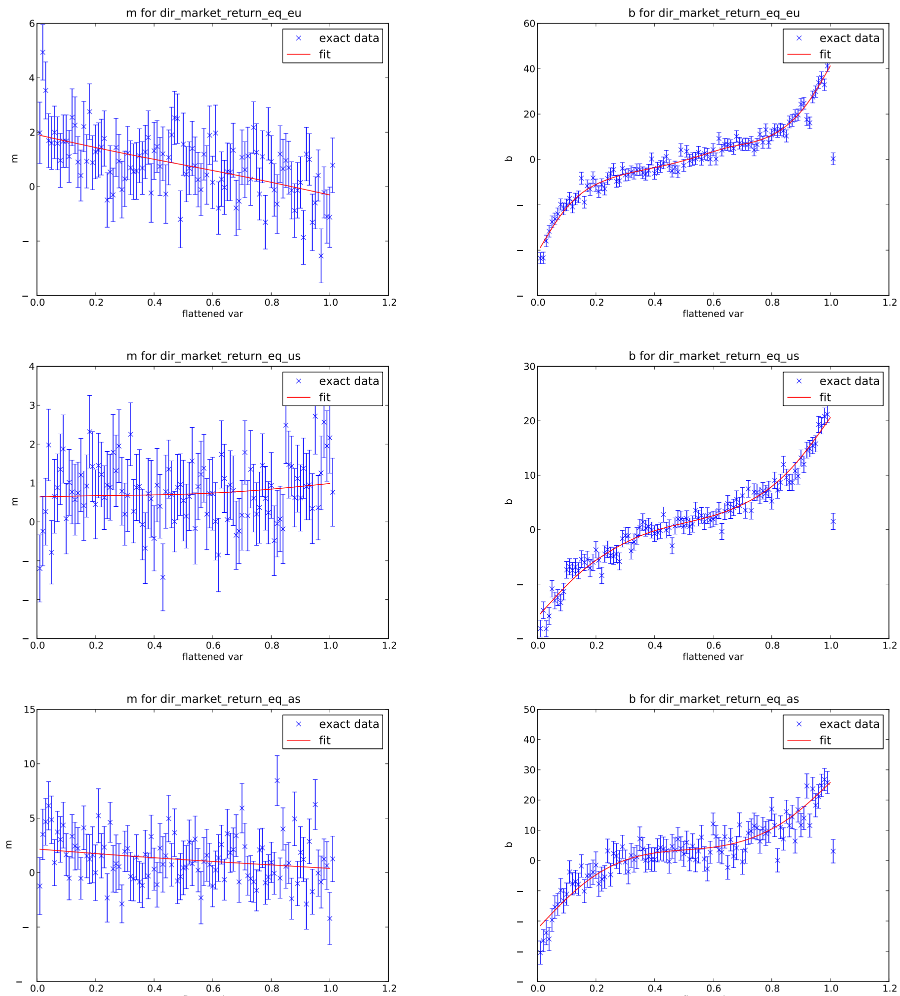

Figure 9.2: Correlations of the parameters m and b to the stock market return of the underlying index (see text) weighted with the direction of the trade and the beta-factor.

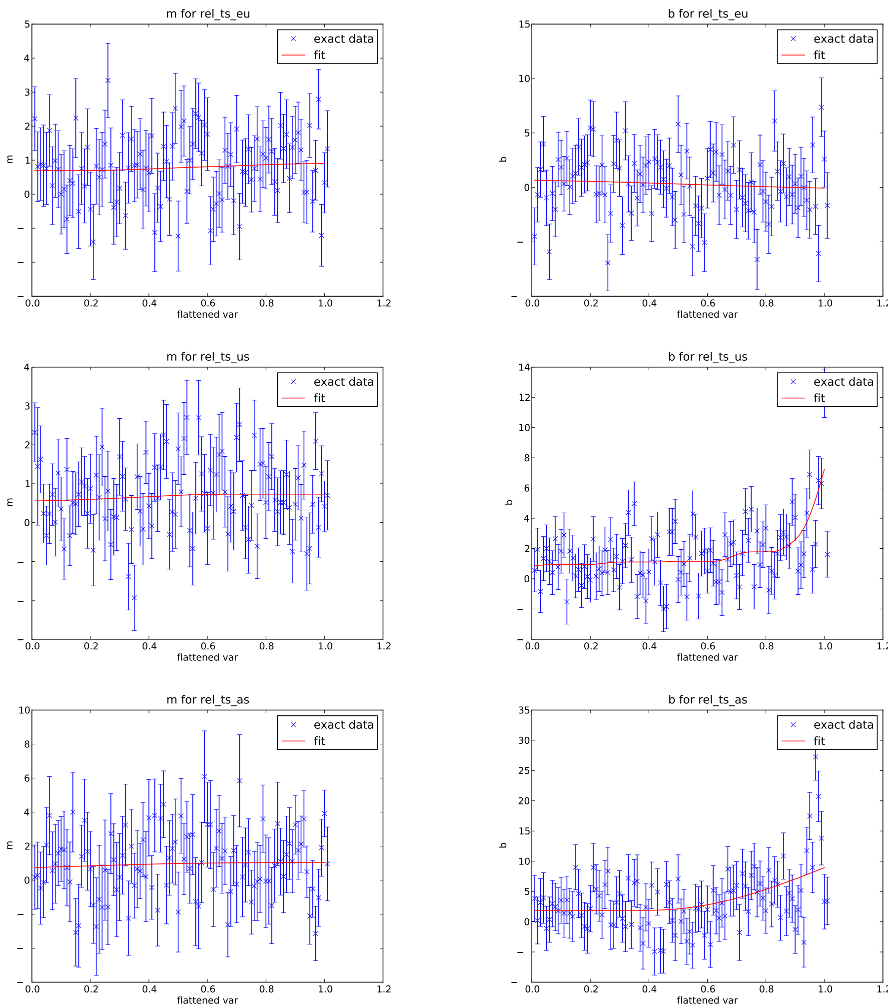

Figure 9.3: Correlations of the parameters m and b to the relative tick size of the stock.

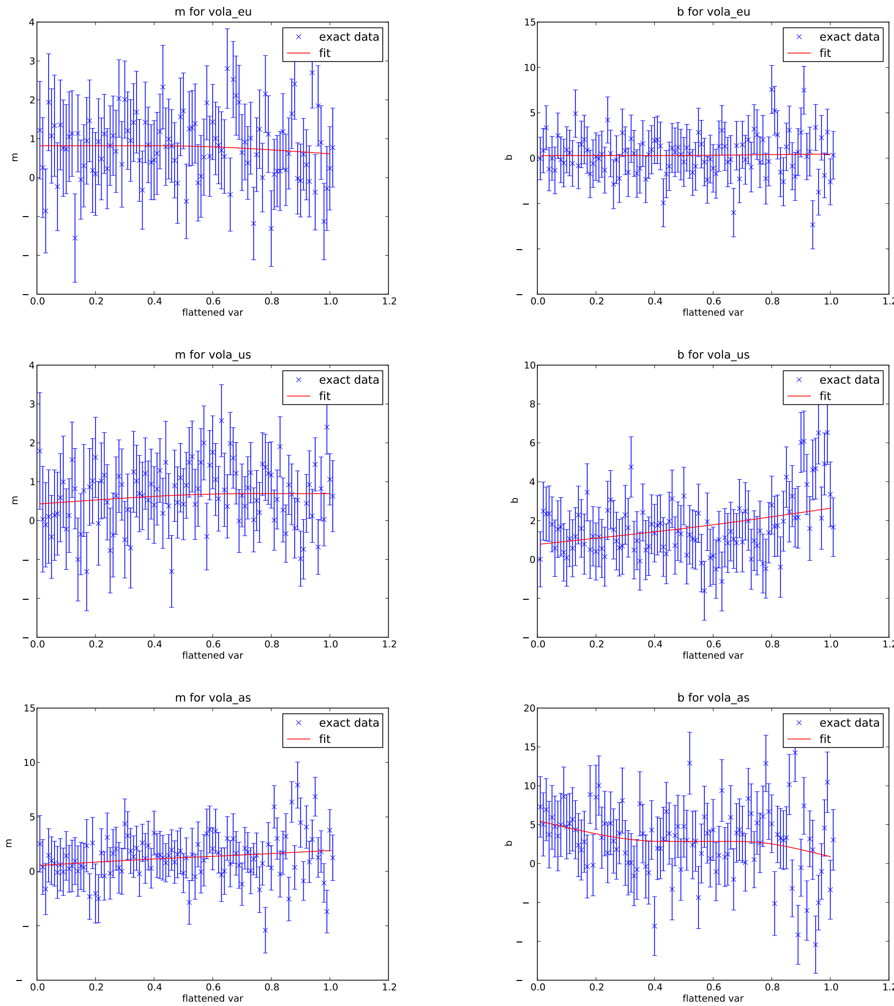

Figure 9.4: Correlations of the parameters m and b to the volatility of the traded stock.

Chapter 10

Discussion of the Results

10.1 Theoretical Background

In order to appraise the results, I have to introduce appropriate research. Huberman and Stanzl (2004) present a theoretical work of the relation between trading volume and market impact. They define the realized market impact as a sum of a temporary and a permanent component. They argue that the market impact function must have certain characteristics. It is not allowed that there is a possibility of arbitrage when trading a round turn in an asset (buy and sell the same amount of shares). This leads to the result that the permanent market impact function has to be linear in trading volume and the temporary function can have a more general form.

Figure 10.1 shows a schematic illustration of a typical average price evolution during and after an algorithmic trade. Also the temporary and permanent components of the market impact are presented.

10.2 Existing Empirical Measurements

Almgren *et al.* (2005) present an empirical analysis of the dependency between trading volume and market impact. Their data set contains algorithmic trades of the Citigroup US equity trading desks. They measure the two components of market

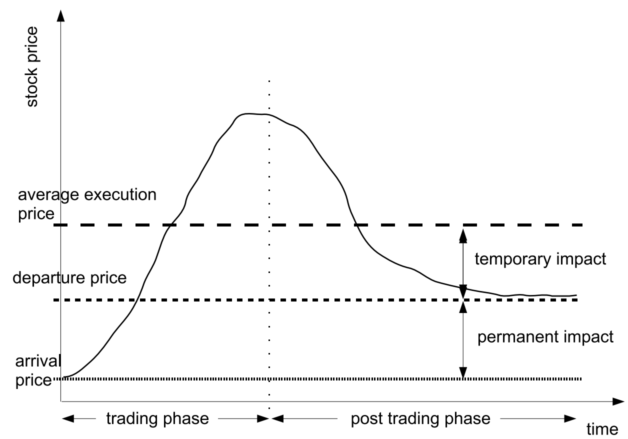

Figure 10.1: Scheme of price evolution during a algorithmic trade

impact separately (temporary and permanent). The permanent impact is defined as the relative price change between the arrival price and the price about half an hour after the trade was finished. The realized impact is the relative price change between the average execution price and the arrival price. The difference between realized and permanent impact is the temporary one.

To describe the functional form of the dependency between trading volume and market impact they choose the following power law functions:

$$g(v) = \pm\gamma|v|^{\alpha} \tag{10.1}$$

$$h(v) = \pm\eta|v|^{\beta} \tag{10.2}$$

where $g(v)$ describes the permanent impact term and $h(v)$ the temporary impact term and v is again the participation rate. The realized market impact is the sum of these two components, $g(v) + h(v)$. The fit results are:

$$\begin{aligned} \alpha &= 0.891 \pm 0.10 \\ \beta &= 0.600 \pm 0.038 \\ \gamma &= 0.314 \pm 0.041 \\ \eta &= 0.142 \pm 0.0062 \end{aligned} \tag{10.3}$$

Thus, the results are consistent with the theory in which the exponent α has to be 1 because of the absence of arbitrage. The theoretical value $\alpha = 1$ is marginally outside the 1 σ interval of the measurement and is therefore not rejected. The value of the exponent β is 0.6, so the temporary impact function is concave.

10.3 Comparison of the Results

In the current analysis the realized market impact is described without the decomposition into the temporary and permanent component. The current data set has the property that the departure prices are always the close prices so there is no price about 30 minutes after the execution. The possibility to take the open of the next day has the disadvantage that there is usually a large price change (high volatility) between the close and the open price. Additionally, there are overnight effects which are unreasonable for the analysis. Hence, just the realized market impact is modeled with the given concave function, e.g. 8.4.

Joining the two models, the one of this work and the one of Almgren *et al.* (2005), one get the more general model:

$$M(v) = p_1 \cdot v^{p_2} + p_3 \cdot v + p_4 \tag{10.4}$$

The power law term plus the intercept can be interpreted as the temporary impact and the linear term can be interpreted as the permanent impact component. In this analysis there is evidence for an intercept which is obviously ignored by Almgren *et al.* (2005). Comparing the fit functions which are used in the two analyses with the more general function 10.4, one can see that Almgren *et al.* (2005) provides a lower bound for the exponent and this analysis provides an upper bound. The lower and the upper bound respectively is the result of the withdrawal of one term from the model 10.4. If the intercept term is removed the exponent is underestimated, if the linear term is removed, the exponent is overestimated. This is shown with the help of a toy Monte-Carlo simulation, in which the data is generated with the model 10.4. A model without the intercept term is fitted and also a model without the linear term, the results are presented in table 10.1.

Thus, the two measurements provide 0.600 ± 0.038 as the lower bound and

Table 10.1: results of the Monte Carlo simulation

	p_1	p_2	p_3	p_4
original model	10	**0.5**	10	0.1
model with intercept=0	6.62	**0.41**	17.76	0
model without linear term	15.22	**0.58**	0	0.15

0.547 ± 0.143 as the upper bound. Without further information of the measurement of Almgren *et al.* (2005), it is not possible to ascertain whether the two results are compatible or not, taking the statistical uncertainties into account.

10.4 VWAP - The Optimal Trading Strategy

With the knowledge of the market impact models and the dependency of the market impact on the participation rate, it is obvious to have a look on the optimization of the trading strategy. It is assumed that the functional form between participation rate and market impact is the same for different trading periods. For the presented three types of models (the linear, the power law, and the individualized linear model) it can be shown that the VWAP strategy is the optimal execution strategy.

This is demonstrated by splitting the trading period in N sub-periods and varying the volume which is executed in each of the sub-periods. The market impact is optimal with the constraint of full execution inside the given time period. This may be written as follows:

$$f = \sum_i^N \left(m \cdot \left(\frac{v_i}{V_i}\right)^\beta \cdot v_i + b \cdot v_i \right) + \lambda \cdot (\sum_i^N v_i - v) \tag{10.5}$$

V_i is the entire traded volume of the current stock in time period i, v_i is the volume traded by the certain trading algorithm in i and m, β, as well as b are the model parameters. Parameter λ is the Lagrange multiplier of the constraint of full execution and v the size of the algorithmic order which has to be executed.

The result of the optimization of the entire market impact is:

$$v_i = \frac{V_i}{V} \cdot v \qquad \forall i \in [1, 2, ..., N] \quad and \quad \beta \neq 0 \tag{10.6}$$

and

$$\frac{v_i}{v_j} = \frac{V_i}{V_j} \tag{10.7}$$

This can be interpreted in the way that the volume of the algorithmic order should be distributed over the given period of time, proportional to the entire trading volume of the stock, which is exactly the idea of a VWAP trading algorithm.

Part IV

Trading Volume Prediction

Chapter 11

Trading Volume

The trading volume of a security in a given period of time is an important measure for the liquidity of a security. Of course, it depends heavily on the market capitalization and the free float of the company. Temporary trading volume fluctuations can be influenced e.g. by strong trading interest triggered by corporate news, changes in a stock market index composition or market movements. There are also significant intraday and interday seasonalities.

As it is shown in section 10.4, the VWAP trading strategy is the optimal strategy in order to trade large sizes in a given period of time. The idea behind the VWAP strategy is to trade a constant fraction of the overall traded volume. The market impact analysis in part III is a post trade analysis where the traded volume is already known. During the activity of a trading algorithm the trading volume of the future is not known, so it has to be predicted. The quality of the volume prediction has a direct influence on the execution quality. In contrast to stock prices, trading volume shows much stronger predictable patterns.

The following section presents a model to predict the trading volume per minute of each stock of the universe for different time horizons in the future.

11.1 Description of the Data Sample

The data set used for this analysis contains trading volume of about 500 of the most liquid US stocks which are basically all covered by the S&P 500 stock market index between January 2007 and July 2010. Trading volume is defined as the number of stocks traded in a given period of time multiplied by the execution price. Figure 11.1 shows the average daily trading volume per stock of the universe over time. One can see that the trading volume was quite high during the crisis in October 2008. In 2010 it is on a level which is about half of the level of 2007.

Table 11.1 shows the average trading volume per day for some stocks to get an

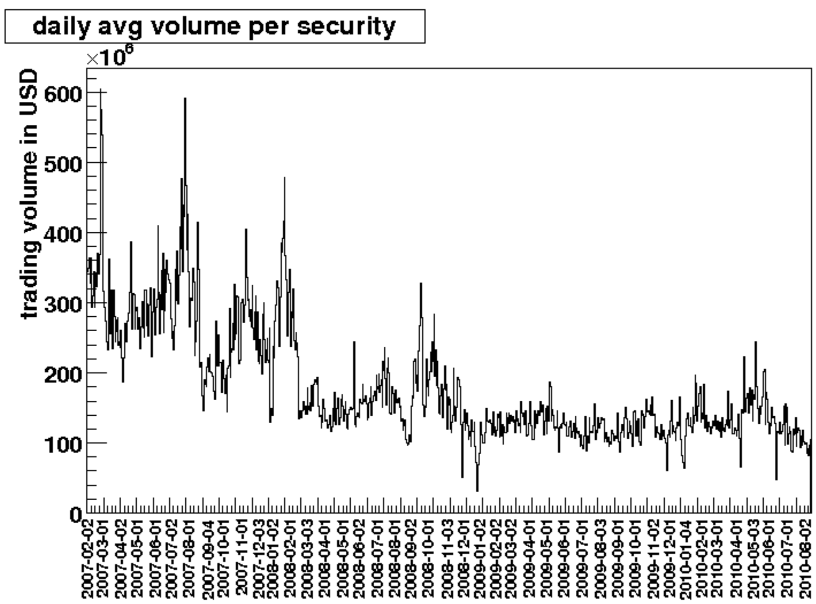

Figure 11.1: Average trading volume per security and day over time.

idea of how large the difference between the most and least liquid stocks can be. In the period of time which is covered by the current data sample, Apple Inc is by far the most traded stock in the USA.

The current data sample contains trading volume λ_{ijk} aggregated per security i, date j and trading minute k. Figure 11.2 depicts the average absolute trading volume over stock and minute $\hat{\lambda}_{ik}^{date}$. The ticker are sorted by their liquidity. $\hat{\lambda}_{ik}^{date}$ is

Table 11.1: Ranking of US stocks by daily trading volume.

rank	security	avg daily volume in Million USD
1	Apple Inc	2700
2	Bank of America Corp	1500
3	Google Inc	1400
4	Goldman Sachs Inc	1300
5	Exxon Mobil Corp	1200
6	Citigroup Inc	1200
7	Microsoft Corp	1100
8	JP Morgan & Co	1000
9	General Electric Co	900
10	Intel Corp	800
⋮	⋮	⋮
98	Dow Chemical Co	180
⋮	⋮	⋮
200	Northern Trust Corp	90
⋮	⋮	⋮
400	AmerisourceBergen Corp	40
⋮	⋮	⋮
494	Unisys Corp	9

the average volume over date per ticker and trading minute.

In order to be able to compare the volume patterns of securities with different

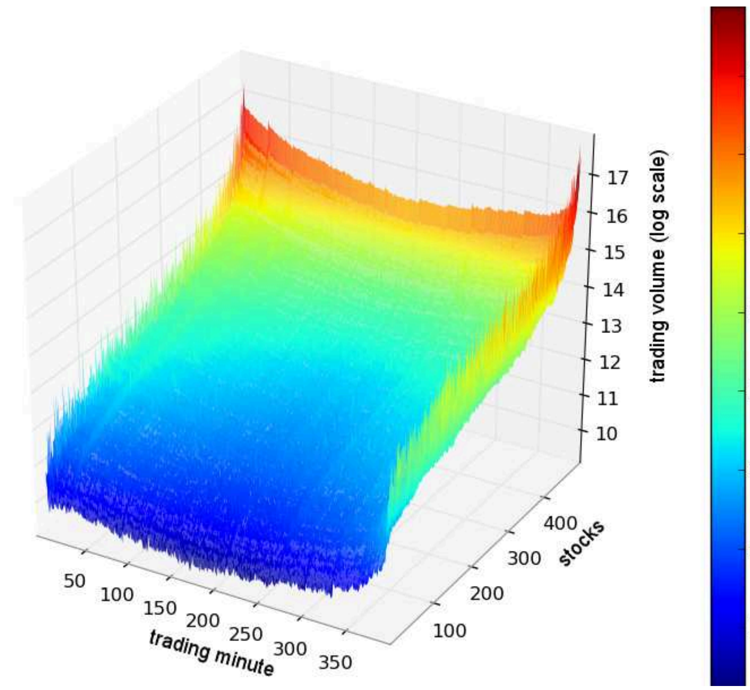

Figure 11.2: $A^{(date)}$ times average daily trading volume for each stock with log-scale.

liquidity, the trading volume λ_{ijk} of security i, day j, and trading minute k is divided by the sum of the trading volume over all trading minutes k of the current trading day j and stock i to get the relative volume v_{ijk} of each minute per stock and date

$$v_{ijk} = \frac{\lambda_{ijk}}{\sum_{k}^{N_{minutes}} \lambda_{ijk}} \tag{11.1}$$

where $N_{minutes}$ denotes the number of minutes per trading day.

A very typical characteristic for intraday trading volume is the u-shaped pattern. All stocks in the universe have high trading volume in the first trading minutes and in the last ones as well as minimum around lunch time. Figure 11.3 shows the volume profile $\hat{v}_{ik}^{(date)}$ where i is chosen to get the data of the stock of Apple Inc and $k = [1, ..., 390]$. In order to predict the trading volume of a given security in a given time interval, such typical trading volume profiles have to be considered.

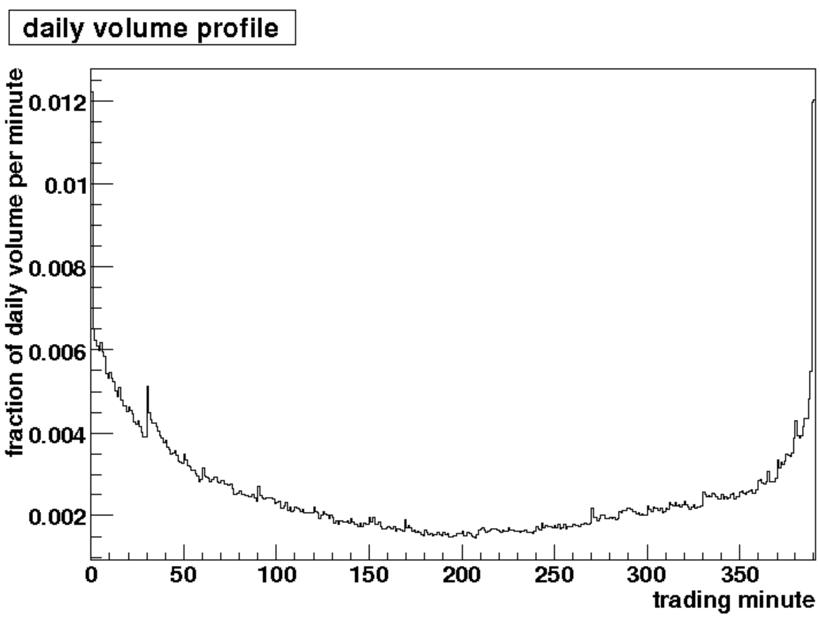

Figure 11.3: Normalized intraday trading volume profile of Apple Inc.

Chapter 12

Dynamics of Trading Volume

In order to analyze the dynamics of the trading volume, several different averages are calculated. The first one is the average over securities per date and trading minute $\hat{v}_{jk}^{(security)}$:

$$\hat{v}_{jk}^{(security)} = \frac{1}{N_{securities}} \sum_{i}^{N_{securities}} v_{ijk} \tag{12.1}$$

where v_{ijk} is the relative trading volume of security i, day j, and trading minute k. $N_{securities}$ denotes the number of securities in the data sample. The analogous proceeding is done for the average over date per security and the trading minute:

$$\hat{v}_{ik}^{(date)} = \frac{1}{N_{days}} \sum_{i}^{N_{days}} v_{ijk} \tag{12.2}$$

where N_{days} is the number of days in the data sample.

These mean values can be written in form of two matrices where the respective columns represent the trading minutes and the rows refer to the different securities or days respectively. In this form, each row of the matrix represents an average trading volume pattern per day or security. These two matrices are given by:

$$A^{(security)} = \begin{pmatrix} \hat{v}_{1,1}^{(security)} & \dots & \hat{v}_{1,N_{minutes}}^{(security)} \\ \vdots & \ddots & \vdots \\ \hat{v}_{N_{days},1}^{(security)} & \dots & \hat{v}_{N_{days},N_{minutes}}^{(security)} \end{pmatrix} \tag{12.3}$$

and

$$A^{(date)} = \begin{pmatrix} \hat{v}_{1,1}^{(date)} & \dots & \hat{v}_{1,N_{minutes}}^{(date)} \\ \vdots & \ddots & \vdots \\ \hat{v}_{N_{securities},1}^{(date)} & \dots & \hat{v}_{N_{securities},N_{minutes}}^{(date)} \end{pmatrix} \tag{12.4}$$

$N_{minutes}$ denotes the number of trading minutes per day. Figure 12.1 depicts matrix $A^{(date)}$. The stocks are ordered by their overall average daily trading volume with the inverse order of the stocks as in table 11.1. There is a dominating u-shaped pattern over the average trading day for each stock. This was already shown in figure 11.3. After having a closer look, one can observe stock dependent patterns. In the following section, these effects will be analyzed in more detail.

12.1 Singular-Value Decomposition of Volume Fraction Matrices

In order to analyze the dynamics in the matrices $A^{(security)}$ and $A^{(date)}$, a SVD (singular value decomposition) to both of them is applied. The SVD is applied for two intentions: The first aim is the decomposition of the matrices in their components. These components can be identified, interpreted, and predicted (see section 12.2 and 12.3). The second aim of the SVD is noise reduction which is realized by a low-rank matrix approximation (see section 2.2). This is done by the decomposition of the matrix, and the removal of unimportant singular values. The insignificant dynamics of the matrix is removed by that procedure. An approximation of the original matrix is computed by just using the most important r singular values and vectors.

A SVD leads to a decomposition into three matrices: the first contains the left

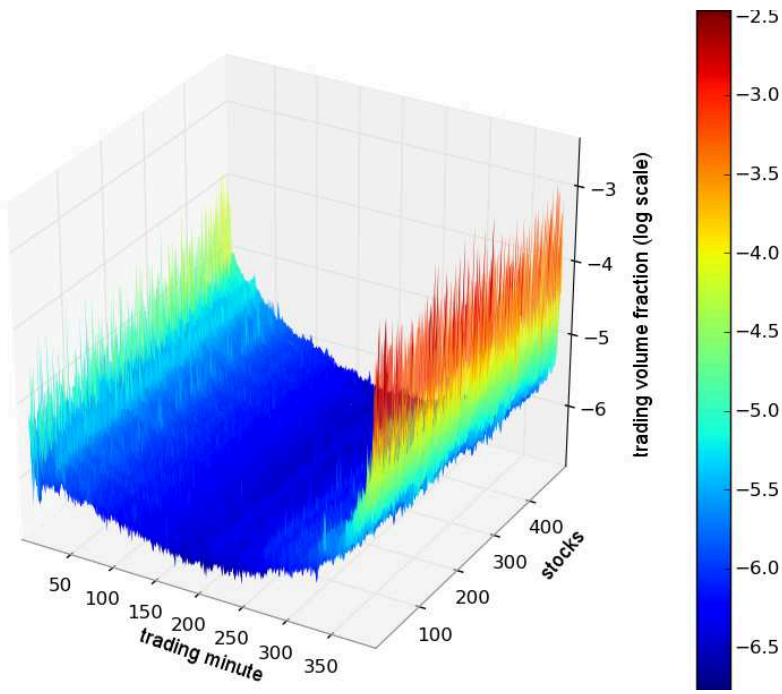

Figure 12.1: $A^{(date)}$ with log-scale.

singular vectors, the second contains the singular values and finally the third contains the right singular vectors. Each of these components are related to special characteristics of the original matrix. The right singular vectors describe the nature of the rows of the original matrix whereas the left ones can be associated with the features of the columns.

The results of the SVD of $A^{(date)}$ and $A^{(security)}$ are analyzed and interpreted below. The plots of the singular values and especially the plots of the singular vectors show very interesting effects.

12.2 Discussion of the SVD of $A^{(date)}$

Figure 12.2 shows the 10 largest singular values of $A^{(date)}$. The first singular value is by far the largest one. The relative value of a singular value, as compared to the sum

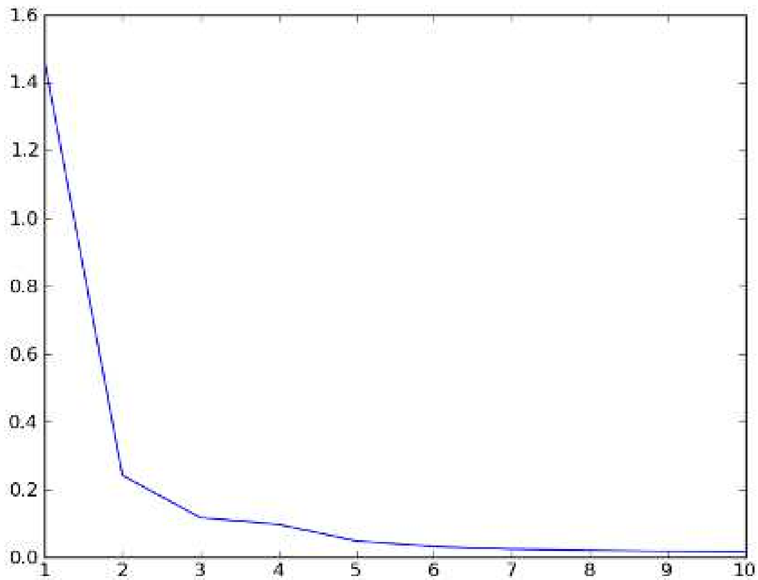

Figure 12.2: Singular values of $A^{(date)}$.

of all singular values, is a measure of the importance of that corresponding singular value and vectors, respectively. Because of the very fast decay of the singular values, the low-rank approximation with $r = 4$ seems to be reasonable. Thus, just the first 4 singular vectors have to be considered which are presented in figures 12.12 to 12.17.

The length of the left singular vectors of $A^{(date)}$ corresponds to the number of stocks, i.e. the number of rows of the matrix. According to that, the right singular vectors' length is equal to the number of trading minutes. In matrix $A^{(date)}$ the stocks are already ordered by the overall average trading volume (liquidity) (see table 11.1). Thus, the left singular vectors are also ordered in the same way as the rows of $A^{(date)}$. The value 0 on the x-axis represents the security with the lowest trading volume. A value near 500 represents one of the most traded stocks, for example the stock of Apple Inc. Analogously, the x-axis of the plots of the right singular vectors, for example figure 12.13, represents the trading minute from 9:30 AM EST (minute 0) to 16:00 AM EST (minute 390).

The first right singular vector is interpreted as the major shape of the intraday trading volume which is the familiar u-shaped pattern. The value of the left singular vector component describes the stock's specific weight of the right singular vector. Thus, the observed trends in the left singular vectors of $A^{(date)}$ lead to different volume profiles for heavily and weakly traded stocks, respectively. The second to fourth singular vectors with their corresponding singular values can be regarded as corrections for the first one.

The second order correction adjusts the volume profile for heavily traded stocks upwards in the first minutes of the trading day. And the peak in volume at the end of the day is diminished. For weakly traded stocks the second order correction works in the opposite direction because their values of the second left singular vector have the opposite sign.

12.3 Discussion of the SVD of $A^{(security)}$

The analysis for matrix $A^{(security)}$, with the help of the SVD is done in a similar way as it is done for $A^{(date)}$. Figure 12.11 shows the singular values of $A^{(security)}$. There is a sharp drop of the singular values after the third one, so most of the dynamics is described by the first 3 singular values. Thus, I decide to make a low-rank approximation with rank 3.

The singular vectors of $A^{(security)}$ have analogous meanings as those from $A^{(date)}$ with the only difference that the left singular vectors describe the weights of the days and not securities. The first right singular vector (see figure 12.12), shows again the u-shaped pattern of trading volume during the trading day. This pattern is as expected quite similar to the one which is depicted in figure 12.12 because it is the overall dominating intraday trading volume profile.

The first left singular vector, describing the trading days of the data sample from January 2007 to July 2010 (about 850 trading days) appears quite constant from day 0 to around day 250. Subsequently there is an upshift until around day 500. From day 500 to the end it seems to be again quite constant. A very similar but opposed pattern shows the second left singular vector with a downshift instead

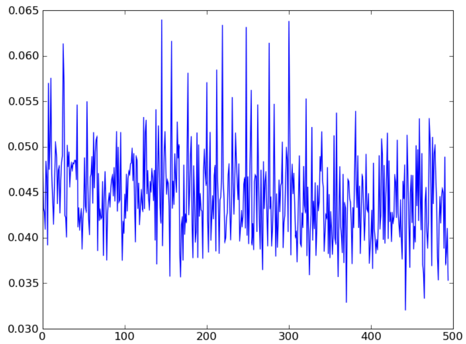

Figure 12.3: 1. left singular vector of $A^{(date)}$.

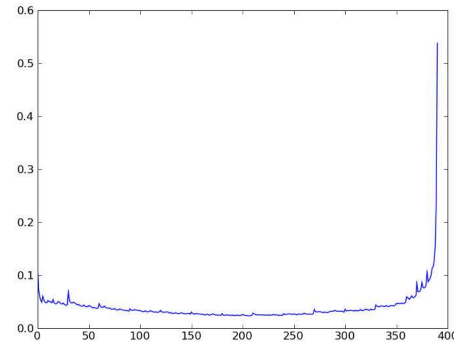

Figure 12.4: 1. right singular vector of $A^{(date)}$.

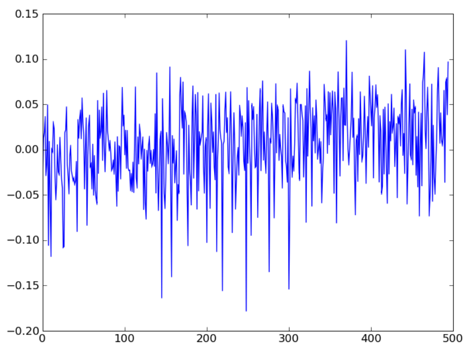

Figure 12.5: 2. left singular vector of $A^{(date)}$.

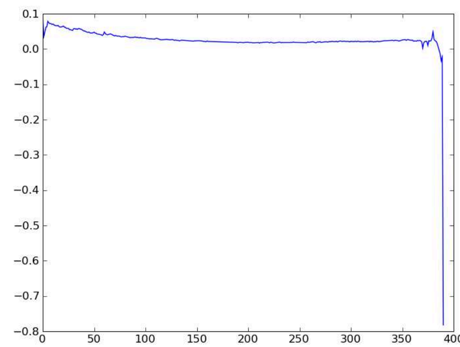

Figure 12.6: 2. right singular vector of $A^{(date)}$.

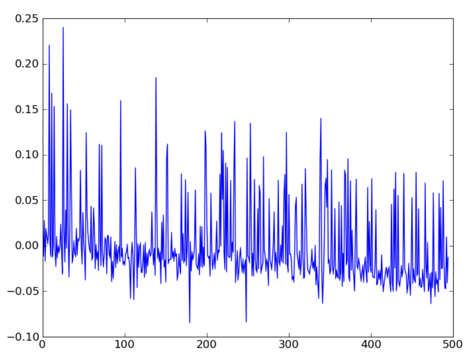

Figure 12.7: 3. left singular vector of $A^{(date)}$.

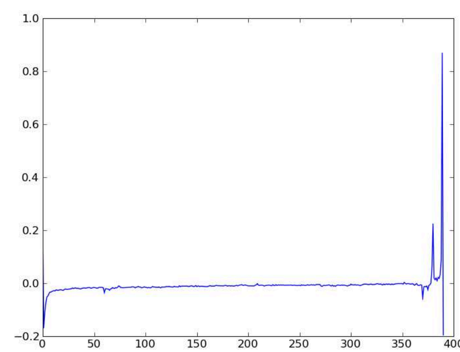

Figure 12.8: 3. right singular vector of $A^{(date)}$.

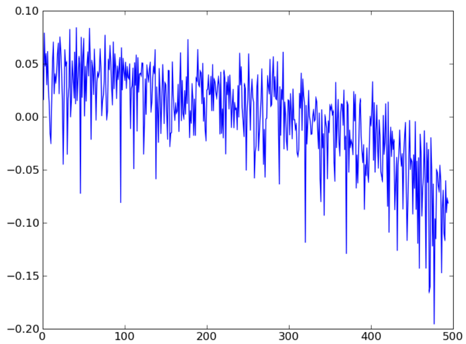

Figure 12.9: 4. left singular vector of $A^{(date)}$.

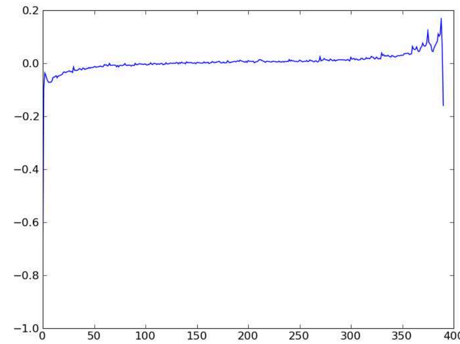

Figure 12.10: 4. right singular vector of $A^{(date)}$.

of an upshift in the same period of time. It is interesting to observe that the financial crisis in 2008 had a strong impact on the trading volume profile. The collapse of Bear Stearns on March 16th 2008 is at day 278 in the current time scale, and the collapse of Lehman on September 15th 2008 is at day 402 in the time scale. So the upshift and the downshift of the left singular vectors were during the climax of the financial crisis, respectively. By looking at the second right singular vector, it can be seen how the crisis influenced the intraday volume profile. The correction of the second singular vector is almost 0 in the time after the crisis. Before the crisis, the second left singular predominantly has positive values. Together with the negative peak in the last trading minutes of the second right singular vector, it reduces the large peak in the trading volume pattern in the last trading minutes.

A possible explanation for this effect could be the decrease in the number of intraday trading market participants, such as hedge funds or high frequency traders, who cause a large trading volume during the day. An additional effect may be the increasing number of passive investments, such as exchange traded funds (ETF) which rather trade at the close.

The third singular vector also describes a very interesting and interpretable effect. By looking at the third left singular vector, the vector components develop very smoothly over time except some very large equally spread negative peaks. The days with these large peaks are always the third Fridays in month. As it is generally

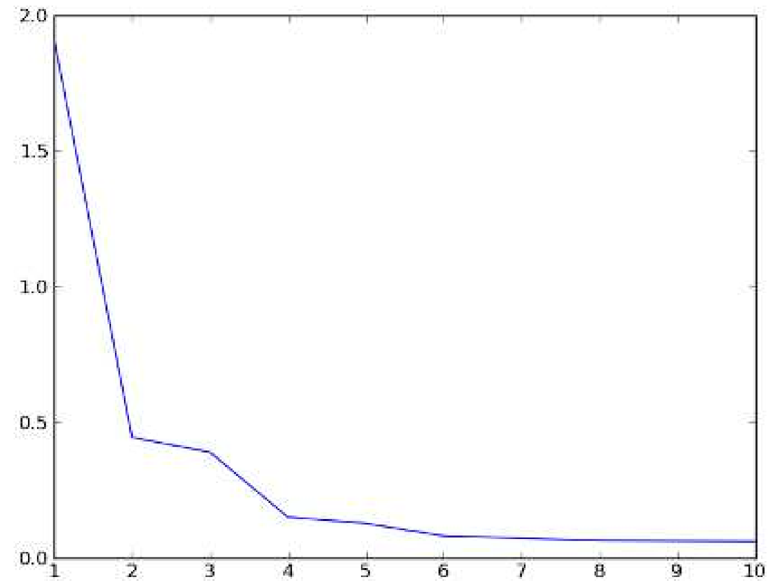

Figure 12.11: Singular values of $A^{(security})$

known, the third Friday of each month is the day of the expiry of the exchange-listed equity options in the USA. Thus it can be expected that this effect has an influence on the trading behavior of their underlyings. By having a look on the right singular vector and keeping in mind that the peaks in the left singular vectors are negative, one can see that the trading volume in the morning is higher and in the last trading minutes it is lower compared to the other days.

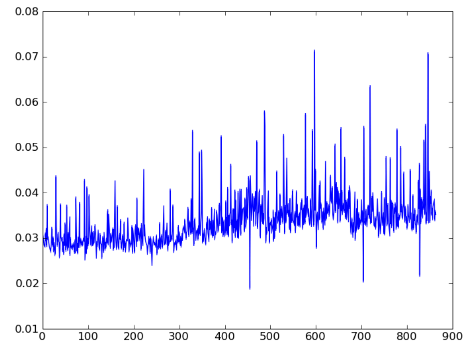

Figure 12.12: 1. left singular vector of $A^{(security)}$

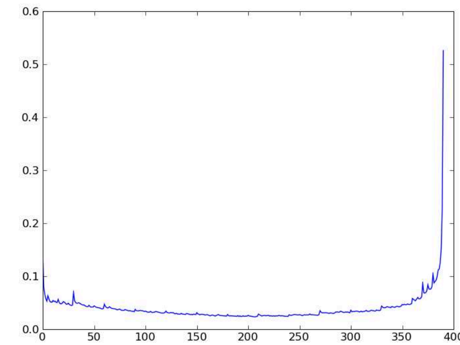

Figure 12.13: 1. right singular vector of $A^{(security)}$

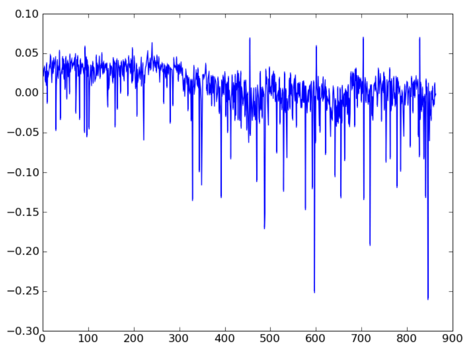

Figure 12.14: 2. left singular vector of $A^{(security)}$

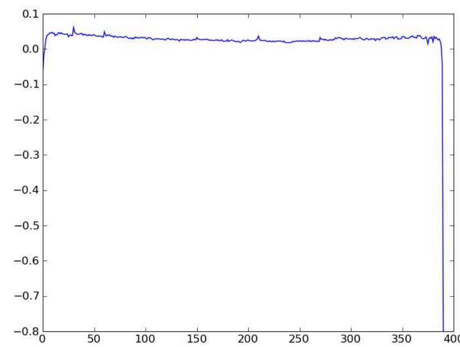

Figure 12.15: 2. right singular vector of $A^{(security)}$

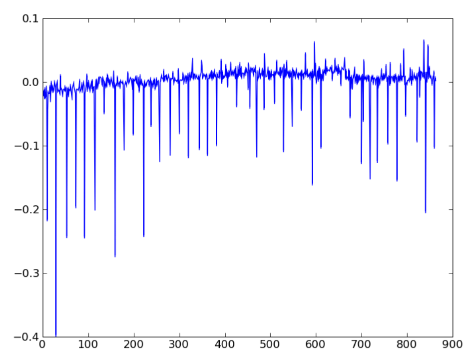

Figure 12.16: 3. left singular vector of $A^{(security)}$

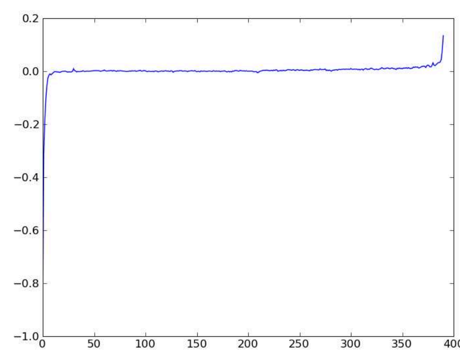

Figure 12.17: 3. right singular vector of $A^{(security)}$

Chapter 13

Prediction of the Trading Volume

In order to optimize VWAP trading strategies, it is important to have good estimations for the future trading volume. Trading volume has significant intraday seasonalities, which have to be taken into account for all kinds of predictions. Therefore, a straight forward approach to model the trading volume with standard methods such as ARMA-models, moving averages, or neural networks does not work. The approach which is presented here decomposes the intraday trading volume into its components and models them separately from each other. The predictions for each of the components are eventually put together to form one final prediction. For each of these components, common models can be used.

13.1 Absolute Trading Volume

The first step in modeling trading volume is to split up the absolute volume per day for each stock and the relative intraday distribution. The current work is focused on modeling the relative trading volume distribution and only use a simple model for the absolute trading volume.

One way to estimate the absolute volume is to use the moving average of the trading volumes of the specific stock of some days before. This method can be applied before the trading day has begun. After the beginning of the trading day another method is reasonable. The observed trading volume up to the present minute m and the expected fraction of the trading volume can be used for an estimation of the absolute volume in a future minute. Assume v_{ijk} as the volume fraction of the stock i and the current day j for minute k. Λ_{ijm} is the accumulated observed absolute volume until minute m. Now one can calculate the absolute trading volume p_{ijn} for any minute n of the current trading day as follows:

$$p_{ijn} = \Lambda_{ijm} \cdot \frac{\hat{v}_{in}}{\sum_{l=1}^{m} \hat{v}_{il}} \tag{13.1}$$

where $n > m$. The results of this method are getting better during the day when the volume of more trading minutes is observed.

13.2 Relative Intraday Trading Volume Distribution

To estimate the relative volume distribution during the trading day, the results from the SVD of the volume fraction matrices $A^{(security)}$ and $A^{(date)}$ are used (see section 12). In general both matrices can be used to estimate the volume fraction of a given minute because both matrices contain mean values of volume fractions. Later on, an approach is presented, which combines the two matrices to one single prediction matrix.

Matrix $A^{(date)}$, also after the low-rank approximation, contains stock specific mean values of trading volume fractions, which are directly estimators for future volume fractions. The matrix $A^{(security)}$ certainly contains only data from the past and has to be extended by at least one day (one row). This is done by treating the left singular vectors as a time series and make a one-day-ahead prediction with an appropriate time series model. For both matrices $A^{(date)}$ and the extended $A^{(security)}$, a low-rank approximation is applied.

The expanded matrix $A'^{(security)}$ can be written in the following form:

$$A'^{(security)} = U'^{(security)} \Sigma'^{(security)} V'^{(security)*} \tag{13.2}$$

where $A'^{(security)}$ has the dimension $m+1 \times m+1$. Consequently also $U'^{(security)}$ has to have the dimension $m+1 \times m+1$, $\Sigma'^{(security)}$ has $m+1 \times n$ and the dimension of $V'^{(security)}$ stays unchanged at $n \times n$. Hence, the matrix is unchanged in the expansion of the dimension, thus $V'^{(security)} = V^{(security)}$. The matrix $\Sigma'^{(security)}$ is expanded by an additional row with 0 and only the 3 largest singular values are different from 0. The remaining singular values are set to 0 due to the low-rank approximation.

$$\Sigma'^{(security)} = \begin{pmatrix} \alpha_1 & & & & & \\ & \alpha_2 & & & 0 & \\ & & \alpha_3 & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{pmatrix} \tag{13.3}$$

Now, matrix $U'^{(security)}$ is calculated as follows:

$$U'^{(security)} = \begin{pmatrix} & & u_{1,m+1} \\ U^{(security)} & & u_{2,m+1} \\ & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \tag{13.4}$$

where $u_{i,m+1}$ are components of the left singular vectors, which can be interpreted as a time series and are predicted with a moving average model in this case.
The matrix $A'^{(date)}$ is just the low-rank approximation of $A^{(date)}$ with the rank 4.

$$A'^{(date)} = U^{(date)} \Sigma'^{(date)} V^{(date)*} \tag{13.5}$$

with

$$\Sigma'^{(date)} = \begin{pmatrix} \alpha_1 & & & & & & \\ & \alpha_2 & & & 0 & & \\ & & \alpha_3 & & & & \\ & & & \alpha_4 & & & \\ & & & & 0 & & \\ & 0 & & & & \ddots & \\ & & & & & & 0 \end{pmatrix} \tag{13.6}$$

13.3 Verification of the Relative Volume Predictions

In order to get a prediction of the trading volume of a given stock i, day j, and minute k, the above presented components have to be combined to a single number. I propose a linear combination of $\hat{v'}_{ik}^{(date)}$, $\hat{v'}_{jk}^{(security)}$, and $\hat{v}_k^{(security,date)}$ as the prediction of the relative trading volume of security i, day j, and trading minute k:

$$v_{ijk}^{pred} = a_1 \cdot (\hat{v'}_{ik}^{(date)} - \hat{v}_k^{(security,date)}) + a_2 \cdot (\hat{v'}_{jk}^{(security)} - \hat{v}_k^{(security,date)}) + a_3 \cdot \hat{v}_k^{(security,date)} + a_4 \tag{13.7}$$

$\hat{v'}_{ik}^{(date)}$ and $\hat{v'}_{jk}^{(security)}$ are the matrix elements of $A'^{(date)}$ and $A'^{(security)}$ respectively. $\hat{v}_k^{(security,date)}$ denotes the average value of the volume fraction of minute k, averaged over all stocks and days up to the latest day.

To verify the model, an out of sample backtest is done. To this end, the data up to day $j-1$ is taken to calculate the matrices $A^{(date)}$, $A^{(security)}$, and the inclusive trading volume fraction profile $\hat{v}_k^{(security,date)}$. In the next step, a SVD is applied to the matrices. The matrix $A^{(security)}$ has to be extended by one day which is done by the prediction of the components of the left singular vectors for the next day (see 13.2). The result of the low-rank approximation of $A^{(date)}$ and of the extended $A^{(security)}$ leads to two different estimations ($\hat{v'}_{ik}^{(date)}$ and $\hat{v'}_{jk}^{(security)}$) for the trading volume fraction of the security i, day j, and minute k.

The parameters of the linear model (13.7) are estimated by fitting the model to all events observed until day $j-1$. For each day, another result for the parameters is observed. As an example, the results of the parameter estimation at June 30th

2010 are presented:

$$\begin{aligned} a_1 &= 1.061 \\ a_2 &= 0.366 \\ a_3 &= 1.148 \\ a_4 &= -0.000466 \end{aligned}$$

The above described out of sample backtest is done day by day from November 2009 to August 2010. Figure 13.1 depicts the residual distribution of the volume fractions. The residual distribution is very asymmetric because the volume fractions

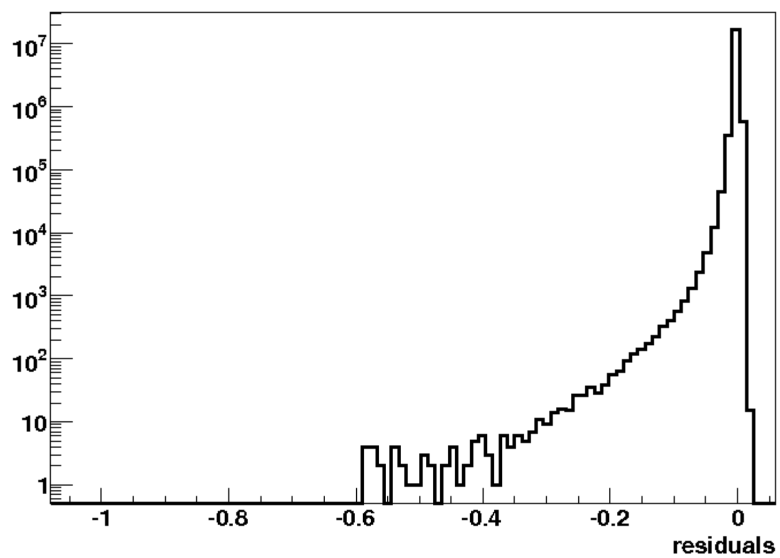

Figure 13.1: Residual distributions of the trading volume fraction estimations.

vary extremely. For some stocks, trading minutes with more than 50 % of the daily trading volume are observed. Additionally, the fraction of the trading volume per minute, of course, cannot be lower than 0. These two effects result in a strong asymmetry of the residual distribution.

In order to discover the quality of the predictions for the fractions of the trading volume, the rms (see equation 9.4) and mad (see equation 9.5) are calculated for

Table 13.1: Comparison of several estimations for the relative trading volume.

	v_{ijk}^{pred}	$\hat{v'}_{ik}^{(date)}$	$\hat{v'}_{jk}^{(security)}$	$v_{ik}^{(date)}$	$v_{j-1,k}^{(security)}$
rms	0.003187	0.003188	0.003199	0.003185	0.003292
mad	0.001594	0.001614	0.001692	0.001611	0.001696

the prediction model and also for some alternative approaches. The alternative approaches to predict the volume fraction are the appropriate matrix elements of $A'^{(date)}$, $A'^{(security)}$, $A^{(date)}$, and $A^{(security)}$.

Table 13.1 shows quite interesting results. The difference of the prediction power of $\hat{v'}_{ik}^{(date)}$ and $v_{ik}^{(date)}$ is very small, i.e. the low-rank approximation of matrix $A^{(date)}$ does not improve the prediction power. In contrast to that, the low-rank approximation and the prediction of the components of the left singular vectors of matrix $A^{(security)}$ does improve the prediction power. Another result is that the prediction of the volume fraction with v_{ijk}^{pred} leads to the lowest mad. Nevertheless, the rms of $\hat{v'}_{ik}^{(date)}$ is almost as good as the rms of v_{ijk}^{pred}. Thus, $\hat{v'}_{ik}^{(date)}$ or $v_{ik}^{(date)}$ seems to be also a good predicton. However, $\hat{v'}_{ik}^{(date)}$ (red) shows larger systematic discrepancies for trading minutes near the close. This can be observed in figure 13.2 which depicts a profile plot of the residuals over the trading minutes. v_{ijk}^{pred} is the blue curve and shows systematically good results for all trading minutes.

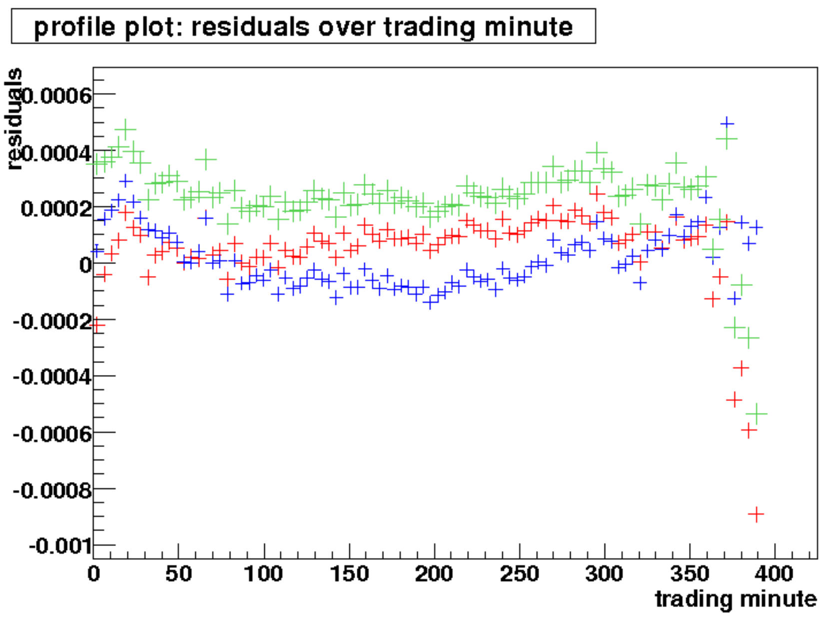

Figure 13.2: Avgerage residuals of volume fraction predictions over minutes for v_{ijk}^{pred} (blue), $\hat{v'}_{ik}^{(date)}$ (red), and $\hat{v'}_{jk}^{(security)}$ (green)

Chapter 14

Conclusion and Outlook

At today's stock markets, almost the entire transaction volume is traded electronically (more than 95% of the trading volume of the German market is traded electronically; in 2009 high frequent trading is responsible for more than 60% of the stocks traded in the US[1]). The important stock exchanges in the world provide electronic trading platforms, the floor-trading is dying out or even already abolished. However, not only the exchanges, but also most of the market participants, like brokers or hedge funds, execute their order flow automatically with the help of trading algorithms.

The current thesis originates from the needs of a smooth order execution system for the Lupus alpha NeuroBayes® Short Term Trading Fund. Therefore a trading algorithm has to be developed in order to trade the investment decisions fully automatically, especially in view of the need to transact over 500 stocks within a short time period. In order to overcome this challenge, it was rather helpful to review the appropriate literature about market microstructure and existing execution strategies.

Another important aspect of the current stock markets in Europe and the US is fragmentation. By this, the fact is meant that there are several trading venues where the same stocks can be traded. Because of different exchange fees, it may be attractive to send the orders to other trading platforms and not to the primary market although the liquidity is still higher there. On the one hand the competition between

[1]source: http://www.nasdaqomx.com/whatwedo/markettechnology/marketview/marketview_3_2010/moving_closer/

the different trading venues leads to several advantages for the market participants such as low fees and high quality exchange systems. On the other hand, it is much more challenging to find liquidity which is often done by systems known as smart order routers.

One of the main topics of the current work is the measurement and the analysis of the market impact of transactions performed by the trading algorithm. The variables are investigated which describe the market impact of orders at stock markets executed by a trading algorithm. The findings indicate that the by far most important variable is the participation rate. Altogether, order executions in four markets (Canada, Europe, Japan, and USA) are examined. The Japanese market behaves differently compared to the other markets due to the up-tick rule for short-sells. The market impact is biased towards lower values because the execution probability of short-sells is small in bear markets if the up-tick rule is valid.

We provide a pragmatical approach of this measurement which is rarely done in the literature. The reason for the latter may be that most publicly available data sets cannot be utilized because the observed orders cannot be related to the market participants and its consequential impact. Only if the set of (small) sub-orders belonging to a (large) algorithmic order can be identified as a whole and be followed up, the market impact of large transactions can be measured.

A linear model describing the dependence between the participation rate and the market impact is provided for all four markets. The linear model as input for a portfolio optimization has the advantage that the optimization function is quadratic in the order size which is much faster to solve. In order to improve the linear model on the one hand and keep it linear in order size on the other hand, an individualized linear regression algorithm is introduced. This algorithm allows to handle additional description variables in the model. The stock market index movement during the trading period of the algorithmic order is a quite important variable to describe the market impact. It is not possible to predict the market impact with the help of the market movement because this would contain future information, but it helps to better understand the dynamics. The market movement of the trading period is of course not known when trading starts. The volatility of the stock price is also slightly correlated to the market impact. For the US and the Japanese markets there is a tick size dependency. This is not true for the European market. Thus, the tick size definitions of the European markets are more efficient in that sense.

A comparison of the different models shows that there is no large difference in the reduction of the *mad* and σ of the residuals. Except the individualized linear regression reduces the width of the residuals due to the stock market movement as an input variable. The linear models work well for a narrow range of the participation rate. Having a wide range of participation rates, it becomes obvious that the functional dependence between participation rate and market impact is not linear anymore but concave. Therefore, a power law is proposed as it is already done in literature. For the combined measurement (EU,USA,Canada), the value of the exponent of the model is estimated to be 0.547 ± 0.143. The result is in line with other similar measurements. It can be utilized to verify the considerations of the market microstructure theory.

Additionally, it has been demonstrated that the VWAP execution strategy is the optimal execution strategy for all the discussed market impact models.

The presented analysis of the market impact does not use all the information available on the data set. All sub-trades of the algorithmic trades are cumulated per stock, day, and trading direction. The evolution of the market impact in time is not taken into account. Thus, it may be an interesting topic for future research to do a "timescale decomposition" analysis, i.e. to investigate a fine time scale.

The quality of a VWAP execution strategy strongly depends on the quality of the overall trading volume predictions. These predictions are necessary because the strategy works optimal if a constant participation rate is realized. The participation rate has to be calculated before the execution starts and therefore the overall trading volume per time interval has to be known. Thus, I developed a model to predict the trading volume.

As it is seen here, the average intraday stock trading volume distribution has a significant u-shaped pattern. The trading activity is high shortly after the opening and before the closing of the market. During lunchtime the average trading volume has its intraday low. For intraday trading volume predictions of a certain time period, this pattern has to be taken into account. In order to analyze the data for all stocks of the US universe together, the minute-by-minute trading volume is normalized by the daily volume of each stock. Then one receive the relative trading volumes per stock, date, and minute. The average values of the relative trading volumes over security and date is calculated. These average values can be written in the form of two matrices. The two matrices are analyzed with the help of a singular value decomposition. This method has two main advantages. On the one hand it is able to reduce noise and statistical fluctuations with the help of the low-rank approximation. On the other hand the resulting components (left and right singular vectors) can be identified, interpreted, and predicted. It can be observed, for example that the intraday trading pattern has changed during the financial crisis in 2008. Since then, the average fraction of the daily trading volume in the last minutes of the day has increased significantly. Another interesting observed effect is the change in the pattern on the third Friday in every month. The third Friday of each month is the day of the expiry of exchange-listed equity options. In these certain days the fraction of the trading volume in the first minutes is higher and the fraction of the last minutes of the trading day is lower. Additionally, the dependency of the intraday pattern from the liquidity of the stock is observed. Stocks with low liquidity show a larger fraction of the trading volume in the last minutes of the trading day.

The provided method to model and predict the intraday trading volume of the about 500 most liquid stocks in the US market is a reasonable approach. It is able to handle the very significant intraday trading pattern whereas other methods, such as ARMA models or neural networks cannot be adapted easily.

Future research on the basis of the current model may focus on the improvement of the prediction of the components resulting from the SVD. An ARIMA model can be used to predict the additional components for one day ahead of the left singular vectors. With the help of neural networks, the combination of the results of the SVD may be improved.

The current work focuses on the prediction of the intraday trading pattern and provides a quite simple approach to predict the absolute trading volume of the future trading day. Another interesting topic for future research is the prediction of the intraday pattern when the trading day has already started. After observing trading volume of the first minutes of the day, there is additional information which is currently not used for the update of the intraday pattern.

Algorithmic trading as it is known today has become very popular during the last few years. There are high frequent trading strategies whose purpose is to make profit with arbitrage, i.e. they buy a security at exchange A and sell it at the same time at exchange B. They also try to detect and trade imbalances on a very short time scale. These kinds of algorithmic trading strategies make markets efficient and ensure fair prices at all trading venues. Their profit results from the bid-offer spread if they provide liquidity and from inefficiencies caused by other market participants.

The dominance of algorithmic trading leads to smaller execution sizes at the exchanges worldwide. Thus, it has become more difficult to execute a large order at once. That is the reason why many investors execute their orders with the help of execution algorithms.

Altogether, stock markets have become more liquid and are currently more efficient than in former times. The spreads narrowed with the rise of algorithmic trading activity[2]. Now it is up to the regulation and the trading platforms to make the systems more reliable and robust to avoid undesirable market behavior, such as the so called "flash crash" in May 2010[3].

[2] source: http://exchanges.nyse.com/archives/2009/08/hft.php

[3] http://www.sec.gov/news/studies/2010/marketevents-report.pdf
WSJ, October 6, 2010; http://online.wsj.com/article/SB10001424052748704689804575536513798579500.html

Appendix A

Profile Histogram

Profile histograms are used to depict the inter-relation of two measured variables X and Y. To this end, the mean value of Y and its statistical error is plotted for each bin in X. These errors are calculated as root mean square (RMS) divided by $\sqrt{N}$. N is the number of events in the particular bin.

Profile histograms are a very useful alternative for two-dimensional histograms or scatter-plots. If Y is an unknown (single-valued) approximate function of X, this function is displayed by a profile histogram with much better precision as compared to a scatter-plot (see http://root.cern.ch/root/html/TProfile.html). This is illustrated with the help of figures A.1 and A.2. Figure A.1 shows a scatter plot of the variable Y (market impact) over the variable X (participation rate). It shows that there are more events with a small participation rate than events with a large one. Nevertheless, this plot clearly shows the possible functional dependency of X and Y. Figure A.2 depicts the profile plot of the same data set. For each of the bins on the X-axis, the mean value of Y is calculated and displayed. The vertical bars represent the errors of the mean values. These errors are smaller in the bins of smaller X values because the number of events per bin N is larger. The profile plot clearly suggests a functional dependency of Y and X.

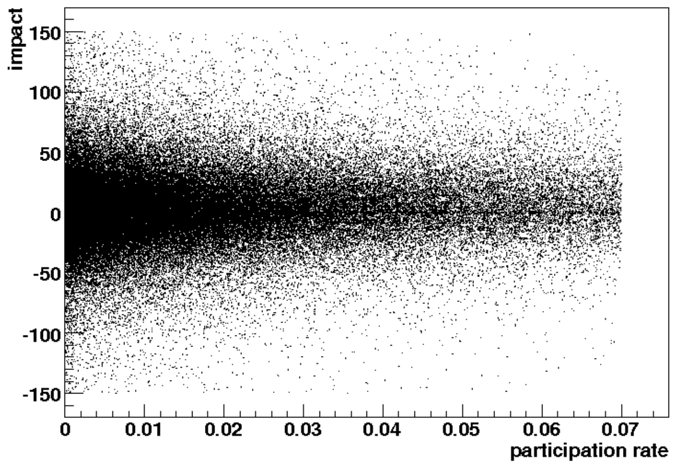

Figure A.1: Scatter Plot.

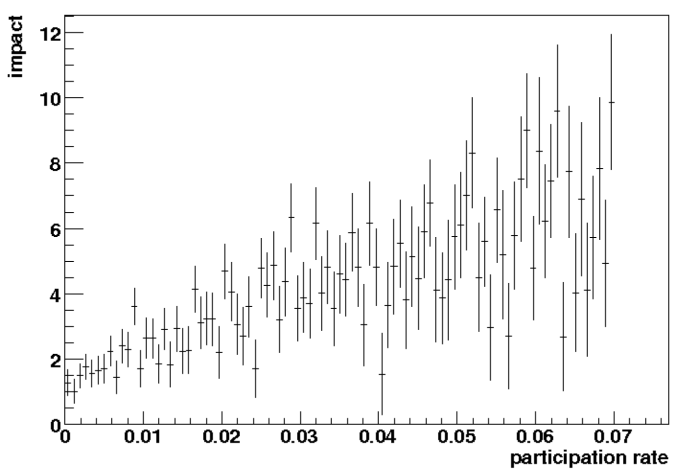

Figure A.2: Profile Plot.

Appendix B

Portfolio Optimization

The general objective of a portfolio optimization is the maximization of the expected profit while keeping risk constant. Having predictions and the estimated risk, an optimization software has to find the best allocation. The actual asset allocation before the optimization is started is denoted by $\vec{a^{(0)}}$, whereas $\vec{a}$ is the optimal asset allocation to be determined. The expected earnings are given by

$$f(\vec{a}) = \sum_{i=1}^{n} \left(a_i \mu_i - I(a_i - a_i^{(0)}) \right) \tag{B.1}$$

where μ_i is the expected return of asset i and I is the impact depending on the traded volume of the asset.

The market impact model describes the relative impact of an order. The calculation of the expected profit takes into account the absolute costs of the impact. Altogether the relative impact, coming from the impact model, is multiplied by the traded volume leading to:

$$f(\vec{a}) = \sum_{i=1}^{n} \left(a_i \mu_i - (a_i - a_i^{(0)})(m(a_i - a_i^{(0)}) + b) \right) \tag{B.2}$$

Using the linear impact model, the optimization problem itself remains quadratic (QP). Using the power law, it ends up with a general nonlinear problem. Optimization algorithms solve a linear (LP) or a quadratic problem much faster than general nonlinear problems. In a high frequency trading set-up, the optimization algorithm has to be fast. Therefore the linear impact model has the great advantage of not increasing the complexity of the optimization problem.

Bibliography

Alfonsi, A., Schied, A., and Schulz, A. (2007). Optimal execution strategies in limit order books with general shape functions.

Almgren, R. and Chriss, N. (1999). Optimal execution of portfolio transactions. *The Journal of Risk*, *3*.

Almgren, R. and Lorenz, J. (2007). Adaptive arrival price. *Algorithmic Trading III: Precision, Control, Execution*.

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

Barclay, M. J., Hendershott, T., and McCormick, D. T. (2001). Electronic communications networks and market quality. *SSRN eLibrary*.

Bennett, P. and Wei, L. (2006). Market structure, fragmentation, and market quality. *Journal of Financial Markets*, *9*(1), 49–78.

Berkowitz, S. A., Logue, D. E., and Noser, E. A. (1988). The total cost of transactions on the nyse. *The Journal of Finance*, *43*(1), 97–112.

Berry, M. W. and Browne, M. (2005). Understanding search engines: mathematical modeling and text retrieval. *Siam*.

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

Biais, B., Hillion, P., and Spatt, C. (1999). Price discovery and learning during the preopening period in the paris bourse. *The Journal of Political Economy*, *107*(6), 1218–1248.

Bialkowski, J., Darolles, S., and Le Fol, G. (2005). Decomposing volume for vwap strategies. *CREST INSEE*, *16*.

Bialkowski, J. P. (2008). Improving vwap strategies: A dynamic volume approach. *Journal of Banking & Finance*, *32*(9), 1709–1722.

Blobel, V. and Lohrmann, E. (1998). *Statistische und numerische Methoden der Datenanalyse*. Teubner Studienbuecher.

Blume, M. E. (2007). Competition and fragmentation in the equity markets: The effect of regulation nms. *Working Paper Series*.

Boehmer, E. (2005). Dimensions of execution quality: Recent evidence for us equity markets. *Journal of Financial Economics*, *78*(3), 553–582.

Boman, M., Brouwers, L., Hansson, K., Jansson, C. G., Kummeneje, J., and Verhagen, H. (2001). Artificial agent action in markets. *Electronic Commerce Research*, *1*(1), 159–168.

Byrne, J. A. (2007). Hooked on speed. *Alpha Magazine*.

Chowdhry, B. and Nanda, V. (1991). Multimarket trading and market liquidity. *Review of Financial Studies*, *4*(3), 483–511.

Coggins, R., Lim, M., and Lo, K. (2006). Algorithmic trade execution and market impact. *IWIF 1, Melbourne*, 518–547.

Cohen, K., Maier, S., Schwartz, R., and Whitcomb, D. (1986). *The Microstructure of Securities Markets*. Prentice-Hall.

Cowan, G. (1998). *Statistical Data Analysis (Oxford Science Publications)*. Oxford University Press, USA.

Demsetz, H. (1968). The cost of transacting. *The Quarterly Journal of Economics*, *82*(1), 33–53.

Domowitz, I. and Yegerman, H. (2005). Measuring and interpreting the performance of broker algorithms. *Tempelhove*.

Dong, J. and Kempf, A. (2007). Resiliency, the neglected dimension of market liquidity: Empirical evidence from the new york stock exchange. *SSRN*.

El-Yaniv, R., Fiat, A., Karp, R. M., and Turpin, G. (2001). Optimal search and one-way trading online algorithms. *Algorithmica*, *30*(1), 101–139.

Eldén, L. (2007). *Matrix methods in data mining and pattern recognition*. Society for Industrial and Applied Mathematics.

Feindt, M. (2004). A neural bayesian estimator for conditional probability densities. *arxiv*.

Foucault, T. (1998). *Order Flow Composition and Trading Costs in Dynamic Limit Order Markets*. No. 1817 in CEPR Discussion Papers. C.E.P.R. Discussion Papers.

Fränkle, J. and Rachev, S. T. (2009). Review: Algorithmic trading. *Investment Management and Financial Innovations*, (1), 7–20.

Glosten, L. and Milgrom, P. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, *14*(1).

Glosten, L. R. (1992). *Equilibrium in an Electronic Open Limit Order Book*. No. 92-39 in Papers. Columbia - Graduate School of Business.

Gower, J. C. and Dijksterhuis, G. B. (2004). *Procrustes Problems (Oxford Statistical Science Series)*. Oxford University Press, USA, illustrated edition ed.

Grossman, S. J. (1992). The informational role of upstairs and downstairs trading. *Journal of Business*, *65*(4), 509–528.

Gudjónsson, J. G. and MacRitchie, G. A. (2005). Agent brokerage – a human-centred approach to automated stock trading.

Harris, L. (2002). *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford University Press.

Harris, L. and Hasbrouck, J. (1996). Market vs. limit orders: The superdot evidence on order submission strategy. *The Journal of Financial and Quantitative Analysis*, *31*(2), 213–231.

Hendershott, T. and Mendelson, H. (2000). Crossing networks and dealer markets: Competition and performance. *Journal of Finance*, *55*(5), 2071–2115.

Huberman, G. and Stanzl, W. (2004). Price manipulation and quasi-arbitrage. *Econometrica*, *72*(4), 1247–1275.

Investment Technology Group, I. (2007). Itg ace? – agency cost estimator: A model description.

Kearns, M., Kakade, S. M., Mansour, Y., and Ortiz, L. E. (2004). Competitive algorithms for vwap and limit order trading. *EC '04: Proceedings of the 5th ACM conference on Electronic commerce*, 189–198.

Kephart, J. O. (2002). Software agents and the route to the information economy. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(10), 7207–7213.

Kissell, R. (2006). The expanded implementation shortfall: Understanding transaction cost components. *Journal of Trading*, 6–16.

Kissell, R. and Malamut, R. (2005). Understanding the profit and loss distribution of trading algorithms.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, *53*(6), 1315–1335.

LeBaron, B. (2000). Agent-based computational finance: Suggested readings and early research. *Journal of Economic Dynamics and Control*, *24*(5-7), 679–702.

Levecq, H. and Weber, B. W. (1995). Electronic trading systems: Strategic implications of market design choices. *SSRN eLibrary*, *IS-95-19*.

Levecq, H. and Weber, B. W. (2002). Electronic trading systems: Strategic implications of market design choices. *Journal of Organizational Computing and Electronic Commerce*, *12*(1), 85–103.

Madhavan, A. (1995). Consolidation, fragmentation, and the disclosure of trading information. *Rev. Financ. Stud.*, *8*(3), 579–603.

Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets*, 205–258.

Madhavan, A. (2002). Vwap strategies. *Journal of Portfolio Management*.

Madhavan, A., Richardson, M., and Roomans, M. (1997). Why do security prices change? a transaction-level analysis of nyse stocks. *The Review of Financial Studies*, *10*(4), 1035–1064.

Mendelson, H. (1987). Consolidation, fragmentation, and market performance. *Journal of Financial and Quantitative Analysis*, *22*, 187–207.

Obizhaeva, A. and Wang, J. (2005). Optimal trading strategy and supply/demand dynamics. *Working Paper Series*.

O'Hara, M. (1995). *Market microstructure theory*. Blackwell.

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

Quarteron, A., Sacco, R., and Saleri, F. (2000). *Numerical Mathematics (Texts in Applied Mathematics)*. Springer.

Rachev, S. T., Fabozzi, F. J., and Menn, C. (2005). *Fat-Tailed and Skewed Asset Return Distributions : Implications for Risk Management, Portfolio Selection, and Option Pricing*. Wiley.

Rachev, S. T., Mittnik, S., Fabozzi, F. J., Focardi, M., and Jasic, T. (2007). *Financial Econometrics From Basics to Advanced Modeling Techniques*. The Frank J. Fabozzi Series. Wiley.

Scherrer, C., Rachev, S. T., Feindt, M., and Fabozzi, F. (2010). From the simple linear regression to the individualized linear regression.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*, 425–442+.

Stoll, H. R. (2001). Market fragmentation. *Financial Analysts Journal*, *57*.

Wagner, W. H. and Edwards, M. (1993). Best execution. *Financial Analysts Journal*, *49*(1), 65–71.

Yang, J. and Jiu, B. (2006). Algorithm selection: A quantitative approach.