2011

The Impact of a Millisecond: Measuring Latency Effects in Securities Trading

Bartholomäus Ende
E-Finance Lab, ende@wiwi.uni-frankfurt.de

Tim Uhle
Goethe University Frankfurt, tuhle@wiwi.uni-frankfurt.de

Moritz C. Weber
Goethe University Frankfurt, moweber@wiwi.uni-frankfurt.de

The Impact of a Millisecond: Measuring Latency Effects in Securities Trading

Bartholomäus Ende
E-Finance Lab
Grüneburgplatz 1
60323 Frankfurt, Germany
+49 (0)69 798 33853
ende@wiwi.uni-frankfurt.de

Tim Uhle
Goethe University Frankfurt
Grüneburgplatz 1
60323 Frankfurt, Germany
+49 (0)69 798 33864
tuhle@wiwi.uni-frankfurt.de

Moritz C. Weber
Goethe University Frankfurt
Grüneburgplatz 1
60323 Frankfurt, Germany
+49 (0)69 798 33854
moweber@wiwi.uni-frankfurt.de

ABSTRACT

In the course of technological evolution security markets offer low-latency access to their customers. Although latency figures are used as marketing instruments, only little research sheds light on the means of those figures. This paper provides a performance measure on the effect of latency in the context of the competitive advantage of IT. Based on a historical dataset of Deutsche Börse's electronic trading system Xetra an empirical analysis is applied. That way we quantify and qualify the impact of latency from a customer's point of view.

Keywords

Securities Trading, Latency, Error Rate Estimation, Performance Measurement.

1. INTRODUCTION

Competition among European exchanges has been significantly fueled: in November 2007 the *Markets in Financial Instruments Directive* (MiFID) became effective. With MiFID the European Commission aimed at fostering competition and at increasing transparency in securities trading. Before this date, trading was concentrated at national exchanges in Europe [1] which faced nearly no national competitors.

MiFID enabled the entry of new competitors for traditional exchanges. Increasing trading volumes [2] of these so called Multilateral Trading Facilities (MTF) force exchange operators to focus more on the needs of their customers (market participants): these are retail and institutional investors. Market operators aim at attracting customers on their trading systems. On top of different pricing schemes they compete through special services such as low latency access. That way they account for the fact that "*[l]atency is one of the major issues in today's trading business"* [3, p. 1].

In general trading can be defined as the act of transferring an investment decisions into actual portfolio positions. Thereby sophisticated trading plans for the slicing and timing of individual orders as well as their precise realization are imperative success factors for exchange customers [4]. On the one hand portfolio turnovers often require the simultaneous coordination of transactions in multiple instruments to minimize implementation risks. On the other hand execution performance is evaluated by benchmarks based on market prices available at the time of the investment decision or during the time span for entering or closing the targeted position. Thus a successful market participant (trader) is supposed to "*sense a market, spot pricing discrepancies, and make lightning-fast decisions"* [1, p. 60].

Concerning these requirements for fast reactions, market setups based solely on manual trading floors are restricted mainly by human traders' limited capacity of reaction and perception. For such markets latencies, i.e. the time which elapses from the emergence of a new trade opportunity and the actual order arrival at the market, correspond to multiple seconds. The reduction of this time period by employing IT is said to exhibit positive effects already since the 1980s [5].

Among other efficiency improvements triggered by IT the most notable has been the shift from floor trading to electronic trading systems [6, 7]. The electronification of market venues in Europe, i.e. exchange trading systems like Xetra (Deutsche Börse), SETS (London Stock Exchange) or NSC (Euronext France) took place in the late 1990s and enabled market participants to access electronic order books[1] via remote access without the need for physical presence on an exchange floor [1]. This so called Direct Market Access allows straight through processing for accessing securities markets which reduces the necessity of media breaks and manual human interventions [8]. Beyond these benefits it enables algorithmic trading engines which simulate order placing strategies of human traders to enter or close portfolio positions. A typical example is to reach the Volume Weighted Average Price (VWAP) when buying or selling an instrument.

[1] A list of buy and sell orders for a specific instrument sorted by price/time priority. Each update might change its structure, i.e. the included price limits and their respective volumes.

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

Deutsche Börse reports 45 % of transactions on Xetra to originate from algorithms in Q1/2009 and to be still increasing [9]. The rational for the success of algorithmic trading is plentiful: firstly, algorithms allow overall cost savings in comparison to human brokers [10]. Secondly, they break human limitations and thus allow permanent surveillance of outstanding orders. This capability allows algorithms to readjust their trading decisions "immediately" to changing market conditions – i.e. retain their unexecuted orders at best market prices (top of the book) [11]. Besides, algorithms have been proven to substantially improve market liquidity, though the effects of HFT on welfare are ambiguous [12]. I.e. they post passive limit orders and thus provide trade opportunities to potential counterparties in times when they are scarce.

Institutional investors which generate most trading volume [1] exhibit an increasing need for algorithmic trading. Therefore their trading needs became the focus of market operators which have entered an arms race for low latencies [13]. Typically they offer so called co-location or proximity services: here the latency to send orders from the clients' office location is eliminated by hosting these clients' trading algorithms on servers nearby the marketplace's system. Table 1 depicts exemplary latencies from October 2008 used in promotion by the MTF Chi-X Europe.

Table 1. Latencies for Direct Market Assess from [14]

Market Place	Average Latency [ms]
Chi-X Europe	0.4 (co-located)
London Stock Exchange	< 6
Euronext	13
Deutsche Börse	37
OMX	43

Additional to algorithmic trading, which is designed to enter or close stock position based on the decisions from portfolio management, the electronification of trading paved the way for another kind of quantitative trading strategy [15]: so called high-frequency traders (HFT) basically aim at taking advantage from short-timed market inefficiencies. In this respect HFT trades are triggered by computer systems as immediate reactions to changing market conditions. That way they perform a vast number of trades with relatively low profits. The price discrepancies HFT strategies are based on are only restricted to leave a gain over after trading costs. According to [16] HFT margins in the US are as low as 0.1 cent per share (cps) after trading costs while typical brokerage services amount to 1-5 cps [16].

Another distinctive feature of the high monetary turnovers of HFTs is their short position holding times: typically not more than hours or even just seconds. On top, over-night positions are avoided. A typical evolution of the cumulated inventory changes of a HFT acting as a market-maker or middleman at the MTF Chi-X Europe as well as Euronext simultaneously is depicted in Figure1. Similar to a classical money changer market-making is designed to earn the price difference from buy (bid) and sell (ask) price differences. Therefore a HFT following a market-making strategy will try to have a limit at the best prevailing prices on both sides of the order book.

Altogether HFT strategies have become a billion-dollar industry: in the US they account for more than 60 % of the average daily volume in equities trading [18]. Although still entering the European market, HFT strategies are already involved in one out of four trades there and are expected to reach 45 % in 2012 [19].

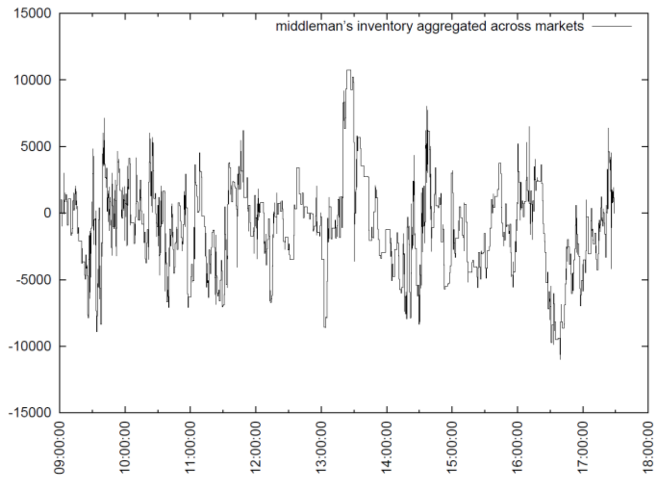

Figure 1. Inventory evolution of a market-maker from [17]

As trading is a zero-sum game profits of HFT traders correspond directly to losses of other market participants. Basically if some participants are able to react quicker to new information they can exploit limit orders of slower market participants as a kind of free-trading option [20].

From an IT business evaluation perspective therefore the following two research questions arises [21]: *What are effects of latency and do they require market participants to employ low latency technology?* To provide market participants guidance in answering this strategy dependant question, we develop a performance metric to measure the impact of latency consistently among different combinations of markets and instruments.

The paper proceeds as follows: section 2 presents a literature review. The research methodology is introduced in section 3 before section 4 describes the employed data set. Our results are depicted in section 6 and discussed in the following section 7. Finally, section 8 summarizes and concludes.

2. RELATED LITERATURE

Our research – the investigation of the impact of latency on securities trading – is related to two different disciplines: research on (i) *the general value of IT* and (ii) *literature dealing with latency in the security trading domain.*

Due to the complexity in IT valuation research different attitudes on the economic impact of IT have been discussed [22]: one major research stream takes the perspective of sustainable competitive advantage for which IT is seen as a key resource [23]. At least IT investments are valued as strategic options to safeguard from potential future losses [21]. Nevertheless IT-created value manifests itself in many ways [22] which might be intangible [24]. In the case of latency reduction technologies such intangible dimensions might be an improvement of execution quality in terms of a higher precision concerning the realization of targeted positions. Thus, our research focuses on the probability of relevant order book changes which occur before an order arrives at the market and the relation of this probability to different latency levels as well as time periods within a trading day. This constitutes a performance metric for latency.

[25] propose that *"[t]he greater the degree of competition in an industry, the greater the extent to which firms achieve efficiency gains via IT"* (p. 306). Electronic securities markets exhibit a highly competitive character and an ongoing arms race of IT. In this respect our performance metric contributes particularly to this proposition, i.e. to which extent investments in IT in this field may yield competitive advantages.

[26] states process performance to be related to business performance from various IS perspectives. Customers in our case, which are primarily institutional investors such as banks, exhibit tendencies for standardization, automation and flexibilization of IT and the supporting processes [27]. In case of the order submissions process our performance metric helps to assess the effects of automation. [8] argue that banks can yield high internal straight through processing rates, which implies the necessity of low error rates in our context, by consistent integration of all systems involved in the trading process.

Within the domain of securities trading, related literature like [5] investigate the impact of latency reductions on market quality criteria like liquidity[2] by the introduction of IT. That way [5] analyze the improved information disintermediation for off-floor traders from two minutes to 20 seconds at the New York Stock Exchange in the 1980s. Their results predict a positive effect on liquidity. Nevertheless, these results should be interpreted with care as they might be affected by other market structure changes during the investigation period. Current technology allows latencies of millisecond or sub-millisecond magnitudes. Thus different measurement starting and end points might distort results as pointed out by [3]. To overcome this problem they propose a standard benchmark methodology based on order action round trip times: it is defined as the time span from the order action initiation (i.e. order submission) and trading system response (i.e. execution confirmation) at the customer's market access point. This notion is similar to our definition of latency. Further they analyze the properties of latency based on data from Deutsche Börses Xetra trading system in 2007. That way three drivers for latency are identified: trading activity, time of day as well as the distance between customer access- and market operator host computer. Latency exhibits different levels with a similar structure for every trading day (day pattern). Basically latency increases during the day due to rising trading activity. On top a remarkable latency peak can be observed at releases of US economic data. The mean latency is reported to amount to 51.9 ms with a standard deviation of 25.2 ms. Their numbers provide a range of latencies for our analysis setup. More recent empirical work on the effects of latency on market quality measures are [28] and [20]. Unfortunately their results are ambiguous: [28] find that the latency reductions by the NYSE Hybrid upgrade cause a decrease of liquidity. In contrast the results of [20] show positive effects for a Deutsche Börse system upgrade on April 23rd, 2007 which decreases the system's roundtrip time from 50 ms down to 10 ms.

Modeling the costs of latency, the working paper of [29] is also related to our work. In a highly stylized model the development of the costs of latency in US securities markets from 1995-2005 are examined. Costs in this model only arise from limit changes whereas our perception of order book fluctuations includes limit and volume alterations. Several assumptions of the study seem critical in face of our results. Especially a constant arrival rate of impatient buyers and sellers seems unlikely considering the day patterns of order book fluctuation. However findings such as the concave effect of latency on costs are congruent with our findings.

In the field of IS literature our study contributes to the research as it provides a performance measure on the effect of latency in the context of the competitive advantage of IT. Regarding the domain of securities trading we introduce the notion of order book fluctuation as the key variable which determines the latency impact. This differs from trading activity and volatility as these do not incorporate volume changes. Because algorithms tend to rapidly place and cancel limit orders neither trading activity nor volatility is affected. Whereas order book fluctuation does increase and latency issues arise.

3. METHODOLOGY

3.1 Modeling the Impact of Latency

While conceptions of latency differ not only among research fields but even within a research area an approach to assign economic value to latency can only be undertaken with respect to the specific business (equity trading in our case) that depends on latency. As described before the need for speed in today's marketplaces raises the question who actually demands the low latency connections and what is the economic driver behind this.

To our best knowledge so far no concept has been developed that attempts to assign meaningful economic numbers (amount of cash) to latency in this context. The phenomenon that high speed accesses seem to be utterly indispensible for some trading strategies raises the question about the effects for other traders without such an access. Following the argumentation of [23] *"…a firm with a unique access to IT may be in a position to earn higher profits from that access"* (p.124), it might well be the case that HFT is an example of such a unique access. While not only the low latency connection but also the developed algorithms to exploit them would define this unique access to IT.

The following paragraphs will describe a method which aims at connecting latency to expected untruthfulness of information and deduce a metric to account for this information unreliability. In this respect differing concepts will be examined. However, the basic idea behind them is the same. Every trader, human or algorithmic, depends on latency. When submitting an order at t_1, a decision has to be made about order size and volume based on information (usually the order book, describing current bids and offers at the market) generated at time t_0. When the order reaches the market at time t_2 the situation at the market might again have changed (c.f. Figure 2).

[2] A simplistic definition of liquidity is the ability of a stock position to be established or unwind quickly without or only minimal negative price movement despite its actual size [1].

Our concepts all make use of this fact. Just based on latency figures alone no definitive predictions of the amount as to which the situation might differ between t_0 and t_2 can be made. Thus, it is impossible to conclude from a given latency whether the inherent risk of meanwhile market changes is small or large.

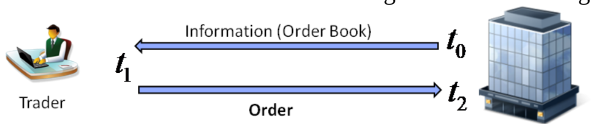

Figure 2. General dependence of a trader on latency

Since the amount of changes and the impact on ones strategy are unknown it is only possible to estimate the outcomes of the gamble which is caused by the latency lag. In the following subsections we present ideas how this can be done.

3.2 Order Book Fluctuations

Taking a closer look at the demanders of low latency trading connections exhibits that most orders of algorithmic trading especially high frequency trading concerns only the top of the order book [11]. Most orders issued by algorithms exactly match the best ask/bid price and volume and if no execution takes place orders are canceled immediately. Therefore we introduce the notion of *order book fluctuation*, which we define as the probability of a change in either the best ask or bid limit or the corresponding volumes at the top of the order book. Formally we define $p_{fluc}(x)$ as the probability of such a change in x milliseconds. This is of course a fundamentally different approach than to concentrate on volatility because order book fluctuations can occur without price changes.

For the case that no information about trading intentions is available, we cannot distinguish whether they are favorable or unfavorable. Thus in this situation we regard any change in the order book as possibly negative. In the progress of this paper we refine this measure to 4 fundamental trading strategies, where only specific changes are regarded to be relevant.

3.2.1 Global Order Book Fluctuation

As described before, without any knowledge about a strategy, any change in the order book may have negative consequences, which a trader could not predict when he submitted the order. An infrastructure provider of data warehouses for traders for example has to decide where to place his facilities in order to meet his customer's demands and she certainly has no information of the different trading strategies it will be used for.

Thus in this case, for a given latency x, the probability of a change of the order book within this time, is the probability that either the limits or the volume has changed without taking care of the direction of that change.

The following paragraphs will define relevant changes for four basic strategies. These cases are chosen rather for demonstration purposes of the methodology than to simulate a real application on a complex algorithm. However, every strategy is a combination of those four basic strategies. The institutional investor's VWAP Buy strategy and a Market Making strategy are stretched out in different directions regarding the basic components as Figure 3 depicts.

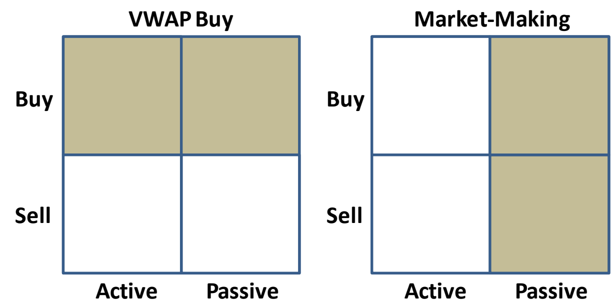

Figure 3. Characteristics of typical trading strategies

The differentiation between active and passive strategies refers to the application of marketable and non-marketable orders respectively. This is explained in more detail in the following subsections.

3.2.2 Active Strategies: Buy Active, Sell Active

We define active strategies as strategies which only uses market orders, i.e. orders that are executed immediately at the best currently available price in the order book. These orders are always executed, whenever a corresponding counterpart exists in the order book.

Thus *Buy Active* is a strategy, where a trader, who wants to build up a position, simply submits market buy orders. After the submission order book changes can occur that may lead to an unfavorable result. It can happen that the best available offers at the time of the order submission are already taken either partly or completely cleared by the time the order reaches the book. If they are taken partially the order is filled only partially at the expected price. Then we could observe a decrease in the volume at the top level in the order book at the time the order reaches the order book. If at the time of the order arrival the ask limit has increased, i.e. the orders were cleared completely, the full order will be executed at a higher price. Accordingly relevant unfavorable order book changes are ask volume decreases and ask limit decreases.

Analogously for a *Sell Active* strategy undesirable events at the bid side of the order book are of the same type. Volume decreases may lead to partial executions and inferior prices. Only here of course bid limit decreases are considered negative since the seller receives a lower price.

3.2.3 Passive Strategies: Buy Passive, Sell Passive

Passive strategies are those which only apply non marketable limit orders. A typical example could be that of a market maker who, like a classical money changer, makes profits by spread earnings from simultaneously buying and selling an asset.

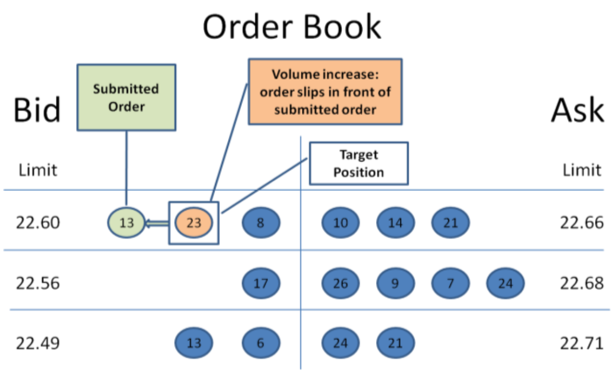

Figure 4. Effects of latency on a trader's order submission

Again we distinguish between buy and sell side in order to determine events that are unfavorable. For a *Buy Passive* strategy which aims at buying a stock by posting bid limit orders an increase in the volume of the bid side during the time of order submission and reception by the exchange would be disadvantageous as the order is further behind others according to price/time priority in open order books. Figure 4 depicts such a situation. Order volumes are written in the circles. Thus the next incoming market order of 31 shares would execute against the fist two orders of 8 and 23 respectively leaving the last order untouched.

Also any bid limit change can be regarded as a negative event. This is because the order has either been overtaken by another limit order with a higher bid or orders have been taken away leaving the order with a possibly to high limit in the order book and what is more with a high execution probability.

Accordingly for the *Sell Passive* strategy increases in volume at the top of the ask side and any limit change in the ask limit are regarded negative.

A summary of changes which are considered negative for the four basic strategies is given in Table 2.

Table 2. Unfavorable top of the book changes

Property / Side	Buy	Sell
Active	Ask Limit ↑ Ask Volume ↓	Bid Limit ↓ Bid Volume ↓
Passive	Bid Limit ↕ Bid Volume ↑	Ask Limit ↕ Ask Volume ↑

3.3 Estimation

Due to the model's simple structure finding estimators for $p_{fluc}(x)$ is straightforward. We take the relative frequency in which order book changes occurred in the past. As reasonable time spans we will take latency in the range of those reported by [3].

Estimators for limit and volume changes can be derived by taking the mean of the quoted volume and limit changes in the time span for which $p_{fluc}(x)$ is estimated.

4. DATA SET

The impact of latencies in magnitudes of milliseconds is of particular interest for algorithmic traders as even such little speed advantages can provide them a competitive edge. Algorithms require fully-electronic open central limit order books and a remote access via technologies like Direct Market Access to be applicable. Thus we choose the Xetra trading system of Deutsche Börse for our analysis. Typically algorithms are employed for instruments with high trade volumes (high liquidity). A proxy therefore is capitalization which is also utilized for index weights. Thus capitalization expresses the particular interest of investors for each instrument. The 30 most capitalized instruments in Germany are represented in the DAX. As expected this index exhibits on Xetra most algorithmic activity [3].

To allow a cross-sectional overview we choose 6 DAX constituents based on their free float market capitalization. That way a pair of two instruments is employed for three different capitalization classes: Siemens and E.ON as high; Deutsche Börse and Deutsche Post as medium and Salzgitter and Hannover Rück as low capitalized constituents (c.f. Table 3).

The employed capitalization data (c.f. Table 3) belongs to our last observation day. Nevertheless it is checked to remain stable during the whole sample period. It is made of 10 trading days starting from August 31st, 2009 and ending at September 11th. Results remain stable for the first and second week of our sample implying that the 10 selected trading days are sufficient. To obtain unbiased results we avoided periods of extreme market activity by expiry dates like so called Triple Witching Days or high market volatility. In contrast the VDAX-New, which can be interpreted as a trend indicator for the volatility of the DAX, exhibits a stable and rather low value compared to the US sub-prime crisis already since August 2009.

Our data set originates from the archives of Thomson Reuters Data Scope Tick History. For the selected instruments all order book updates are retrieved. These updates consist of the first ten quoted limits and volumes on both sides of the book, i.e. the ten highest bid and ten lowest ask limits. Each change within these limits results in an update record. For multiple changes, occurring within one millisecond, we account only for the last one, as investors with the investigated latencies of above 1 ms are not able to react pointedly to such instant changes. Finally, we restrict our analysis to the limit and volume changes of the best bid/ask as algorithmic activity can be predominately found at the top of the book [11].

The focus of our investigation is set on continuous trading where order book changes as well as trades can occur at any times. For DAX instruments continuous trading takes place from 9:00 till 13:00 o'clock in the morning and 13:02 till 17:30 in the afternoon. Accordingly order book updates for auctions are removed and validity times of the last limit updates before auctions adjusted appropriately. Unfortunately, our data lacks secured information on volatility interruptions. But as this mechanism to switch from continuous trading to an auction results in one limit change per interruption and occurs seldom, its effects are expected to be smoothed out by the multitude of order book updates observed.

Table 3 depicts the basic characteristics of the data set described above: besides the free float market capitalizations for our three classes, mean lifetimes of top of the book situations, fractions of

Table 3. DAX order book data sample characteristics

Capitalization Class	Instrument	Free Float Market Capitalization [m€]	Order Book Top					
			Lifetime [ms]		Fraction of Changes [%]		Mean Quoted Volume [€]	
			Mean	Std. Dev.	Limits	Volumes	Best Bid	Best Ask
High	E.ON	57,829	1,129	2,758	19.91	80.09	94,911	96,184
	Siemens	52,070	860	2,342	30.55	69.45	60,205	60,263
Medium	Deutsche Börse	10,902	925	3,216	35.30	64.70	30,645	28,381
	Deutsche Post	10,673	1,507	4,168	21.45	78.55	51,268	45,798
Low	Salzgitter	2,673	1,255	4,189	34.42	65.58	22,849	22,975
	Hannover Rück	1,785	4,020	10,085	33.25	66.75	23,914	25,052

limit and volume changes as well as the mean quoted volume for the best bid/ask are depicted. No general conclusion can be drawn that lower capitalized instruments' best bid/ask limits and volumes exhibit lower lifetimes. Nevertheless standard deviations of lifetimes are generally high and increase for lower capitalizations. Further, there are about twice to four times more volume than limit changes. Basically the fraction of limit changes increases with lower capitalizations. This is obvious as lower capitalizations come along with lower quoted volumes and thus induce more trades to completely remove the volume of the targeted limit level. As limit price changes come generally along with different volumes the depicted numbers reflect only such volume changes without simultaneous limit alterations.

5. MEASUREMENT AND RESULTS

5.1 Measurement

For our goal to find a universal and neutral measure for the impact of latency we try to assume as few as possible restrictions by a specific trading strategy. Consequently our measurement procedures are not based on strategy specific information such as: when an individual trader submits orders, receives executions, which kind of orders are used or how harmful unexecuted orders for her strategy might be. Instead we take a general perspective and aim at investigating the expected probability of relevant order book alterations as well as the expected magnitude of such alterations.

Further, as we expect day patterns within our data, trading days are divided into investigation intervals: the shorter these intervals the more flicker arises whereas longer interval potentially might smooth out patterns. Therefore we checked different interval lengths. Overall the found patterns remain stable. For the illustration below an exemplary interval length of 15 minutes is chosen.

For each interval of a trading day (34 for a length of 15 minutes) we calculated the probability of being hit by an order book change within a given latency delay. This is carried out for any change for the strategy independent measure and for relevant changes for the four simple strategies as described in sections 3.2.2 and 3.2.3. Besides the magnitude in volume and limit price changes at the top of the book (i.e. for the best-bid/ask limits) are calculated.

For all calculations we applied a sliding window. It compares the order book situation at a time m_i with that after an assumed latency delay x milliseconds later, i.e. at $m_i + x$. This window slides through every millisecond of an interval. In every millisecond where we can find a relevant change after x milliseconds we increase our number of relevant observations by one. At the end of each interval we divide the number of observations by the amount of milliseconds in that interval, i.e. 900,000 ms in case of 15 min interval. As an estimator for the probability of a (relevant) order book change we take the average of the ten trading days for each interval of those ratios.

Because an order could be submitted in any millisecond this ratio estimates the probability of being hit by an order book change when submitting an order at any time in the interval. Variations of the window size which simulates our latency delay are set from 5 to 100 ms in 5 ms steps to assess latency impacts over typical traders' latency experiences [3]. Additionally, latencies of 1 and 2 ms are included to focus border cases. To assess the impact of those changes we also measured the average limit and volume changes within those time spans.

Limit price changes typically come along with volume changes. Thus we only account for such volume alterations where the limit price remains unchanged to avoid overestimations of the alteration probability.

5.2 Day Pattern in Order Book Fluctuations

As expected the probability of alterations clearly shows a significant day pattern. The trend of the average probability for our four basic strategies and the overall measure of limit and volume changes for a latency of 10 ms is depicted in Figure 5. Basically one can see that all 5 lines exhibit the same form that is only shifted upwards or downwards. As the top line in the graph accounts for all kind of changes it takes the highest probabilities. The two next lines represent the passive (buy/sell) strategies and the two last with the lowest probabilities correspond to the active (buy/sell) strategies. Obviously there are no striking differences among the buy/sell pairs of active or passive strategies as the corresponding best-bid/ask limits are symmetric around the instruments midpoint. Further, the fact that passive strategies exhibit higher probabilities to be effected by order book alterations is due to the fact that they account for three kinds of changes whereas active strategies do only for two.

Concerning the overall trend all five lines share a modified U-shape which can be also observed for trading volumes [30, 31].

Thus, in the morning the probability of order book alteration is high and decreases continuously. It reaches its minimum just after the midday-auction. Then it increases again. Different to typical volume U-shapes it falls sharply again at ~14:30. Then a striking large increase occurs at approx. 15:30. This is congruent with the opening time of the US markets.

The line with the highest probabilities represents the case, where any change in the order book is viewed as disadvantageous. As stated before this is an entirely strategy independent measure which could be useful for an infrastructure provider, who does not have access to any information about the algorithms that use the infrastructure.

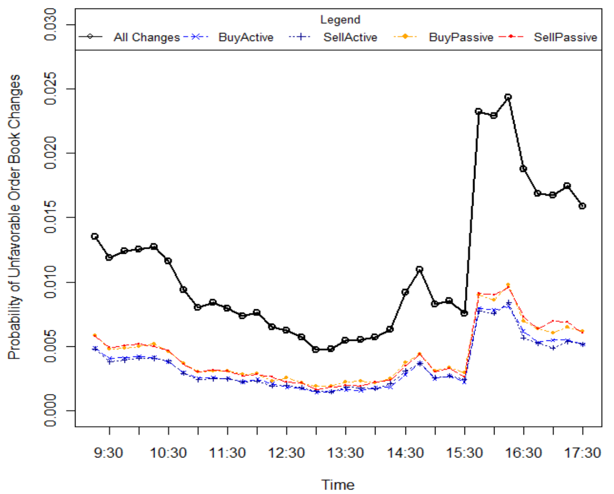

Figure 5. Order book alteration in the course of the trading day for Siemens and 10 ms latency.

5.3 Latency Impact

The length of the latency delay has of course an impact on the probability that the order book situation changes in a way that seems unfavorable for a submitted order. A first hint as to how much this influences the pattern can be seen in Figure 6.

The graph shows the day patterns for 10 to 100 ms for a Buy Active strategy in E.ON. The lowest line represents the probabilities for a 10 ms delay, the next higher line 20 ms etc. We omit the 5 ms step here for demonstration purpose. It can already be seen at this point that the day pattern is not only preserved but even amplified by the latency effect.

In consideration of this fact latency impact is examined for every 15 min interval separately. In every interval the effect of latency on the probability of unfavorable order book changes shows a typical slightly concave relation. This concave effect on the probability can be found in any interval across all stocks and for all strategies in our sample. The graph in Figure 7 depicts the average increase of probabilities for a Buy Active strategy in E.ON. The empiric values can be fitted with a log-linear regression.

From the slope of this regression we can deduce the following simple rule of thumb. A 1 % increase in latency leads to a 0.9 % increase in the probability of unfavorable order book changes.

Thus reducing latency about 1 ms has a greater effect on the probability the lower the latency already is. Due to data restrictions our study only covers latencies from 1 ms upwards. However, with more accurate data and an extension in the submillisecond area this might provide an additional explanation why high investments in relatively small improvements in latency can be found in the market.

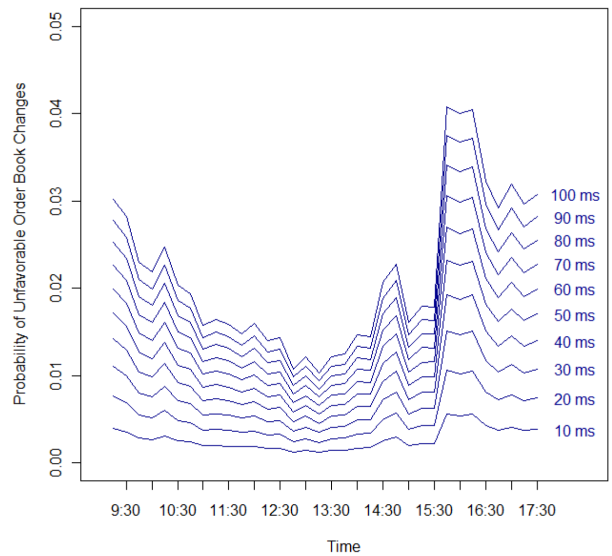

Figure 6. Day pattern for E.ON and latencies of 10 – 100 ms

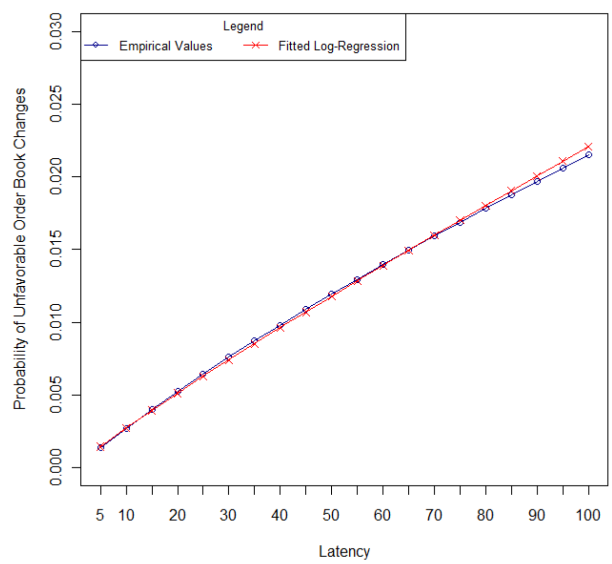

Figure 7. Scaling of hit probability due to latency

5.4 The influence of market capitalization

As depicted in the introduction heavily traded stocks will be more prone to latency risk. Since market capitalization is a fairly good proxy for the interest of traders in the stock (c.f. section 4) we expect highly capitalized stocks to exhibit a higher probability and a higher latency impact than lower capitalized ones.

Figure 8 clearly confirms this assumption. Highly capitalized stock's probability of unfavorable order book changes is on average twice as high than those of low capitalized stocks. The figure shows the day patterns of probabilities for a Buy Active strategy and a latency of 50 ms for the three classes "Low", "Medium" and "High" capitalization.

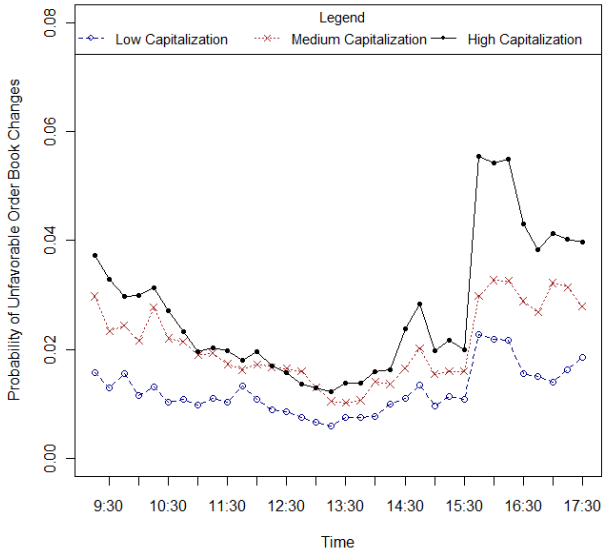

Figure 8. Hit rate for our three capitalization classes

5.5 Average Limit and Volume Changes

Though day patterns are common for limits, prices, spreads and volumes in stock trading, it remains unclear how changes of limits and volumes within latency delay evolve over time. Among others e.g. [30, 31] find typical U-shape of trading volumes. This is congruent with our results. However the risk that one faces due to latency rather depends on the amount of changes in volumes within the order book than on the overall trading volume.

To our best knowledge we do not know any study that examined the average amount by which limit and volume change. In order to combine information of those changes with the probabilities from the previous paragraphs, we use the same sliding window measurement method as before. That is, we compare the limits and volumes after an assumed latency delay and take the average after every 15 min. Limit decreases are measured relatively in basis points (1 bps = 0.01 %) to allow for comparisons among different stocks.

In case of volume changes we could not find any significant trend which is stable over all stocks and trading days, whereas limit changes show a significant decrease in a trading day. Therefore for a typical volume change one should take the average for the whole trading day as an approximation. Changes in limits tend to be higher in the morning than in the evening. As described before limit changes are higher in the morning. A typical example is shown in Figure 9.

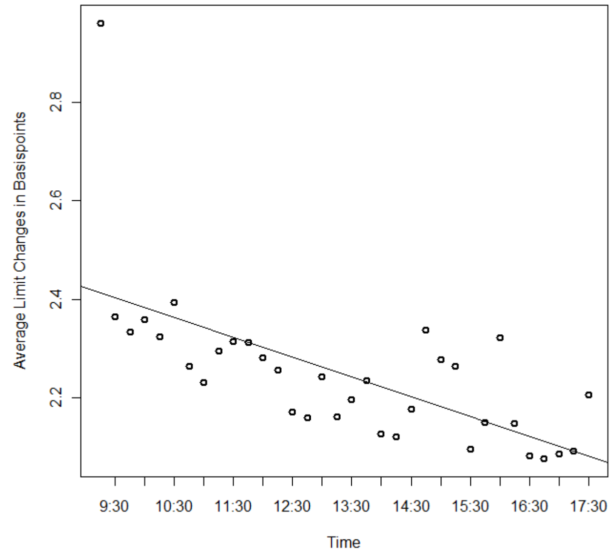

Figure 9.Evolution of limit changes in E.ON

The graph shows the sum of absolute values of changes in bid limits in basis points, averaged over ten trading days in E.ON. The line is that of the linear regression that exhibit a highly significant p-value (at the 1 % level or more) for all cases, except for Hannover Rück, where significance can only be found at the 10 % level.[3]

Interestingly this does not reflect a typical U-shaped volatility pattern. But since limit changes do not necessarily reflect price changes this does not contradict results concerning price volatility.

6. DISCUSSION

6.1 Impact on Active Strategies

For market or marketable limit orders, i.e. the means to implement an active strategy, two unfavorable situations can be encountered (c.f. Table 2): an unfavorable movement of the limit price or a decrease of its volume. To assess their impact we make use of probabilities discussed in the last section. For actual executions the assessment of limit price changes is straight forward as they can be directly converted into costs. Therefore we take the probability of such changes times the expected limit change:

$$E(LimitChangeCosts) = p_{LimitChange} \cdot E(LimitChange)$$
$$= p_{fluc} \cdot P(LimitChange \mid fluc) \cdot E(LimitChange)$$

As we have encountered significant trends within the limit changes (c.f. section 5.5) and day patterns for the probability to be hit by them (c.f. section 5.2) we calculate these figures for each interval. Again we encountered a U-shape for the expected limit change costs. An overview of their magnitudes is provided in Table 4 for an Active Buy strategy and an assumed latency of 50 ms. The latency cost impact ranges between 0.01 and 0.06

[3] Changes in ask limits reveal the same tendency. Significant decreases can be found for all stocks except for Salzgitter.

bps. Basically differences among instruments highly depend on the proportion of unfavorable limit price to volume changes. This is also the rationale behind the low figures for the highly capitalized instruments E.ON and Siemens.

Table 4. Buy executions limit change costs – 50 ms latency

Instrument	Limit Change Costs [bps]			
	Min	Max	Mean	Std. Dev.
E.ON	0.0133	0.0509	0.0275	0.0114
Siemens	0.0140	0.0617	0.0310	0.0140
Deutsche Börse	0.0188	0.0625	0.0382	0.0113
Deutsche Post	0.0124	0.0522	0.0280	0.0109
Salzgitter	0.0150	0.0453	0.0263	0.0084
Hannover Rück	0.0093	0.0363	0.0186	0.0077
Overall (Average)	0.0093	0.0625	0.0282	0.0126

Overall this part of the latency impact costs is low compared to typical implicit trading costs (i.e. market impact, timing or opportunity costs). Nevertheless for strategies yielding only low profits per trade, like those of HFTs, these figures become relevant: for example the US HFT Tradeworx [16] reports average net earnings of 0.1 cent per traded share. With an average share price of 41.84 $ within the S&P 500 this corresponds to net earnings of 0.24 bps. Hence, the sole limit change impact for an active strategy with latencies of 50 ms might diminish their profits by as much as 26 %.

While market and marketable orders face the costs described above in case of executions, it can also happen that due to latency marketable orders cannot be executed. For this situation no direct costs can be associated but a loss of immediacy. Depending on the underlying strategy cost of immediacy need to be assigned if one wants to model the limit change costs completely.

For the second component of the latency impact, i.e. decreasing volume, exact cost figures cannot be calculated without knowledge of the underlying strategy either. Nevertheless our figures show that e.g. in E.ON an average volume decrease of 29 % occurs with a probability of 1.7 – 6.7 % depending on the order submission day time. This is particularly harmful for algorithms which aim at taking advantage of promising trade opportunities as much as possible. For Xetra we know that 76.7 % of all orders that exactly match the best bid/asks and volume are submitted by algorithms [11]. Further, 17.7 % of such orders submitted by algorithms succeed in match the best bid/ask and volume.

6.2 Impact on Passive Strategies

Limit and volume changes result in wrong positioning of the submitted limit order in the order book. For an exemplary buy order a best ask limit increase the order is placed too far up the book, whereas decreases lead to a position below the top. At last the volume effect is opposite to that of the active strategies. An increase in the volume of the top of the order book puts the limit order at a more distant position regarding the price/time priority thus diminishing the execution probability. This effect has already been illustrated in Figure 4. The targeted position is taken by another order that entered the book within the latency delay. The submitted order is now behind this order. The next incoming order that triggers a trade will be matched against this order before the submitted order. It may well happen that this effect hinders submitted orders to be executed at all when marketable sell order volumes are small.

Passive strategies aim at saving or earning the spread, i.e. they seek price improvement at the cost of execution probability. The latency effect decreases the execution probability. Therefore the low latency trader can seek more price improvements than a trader who has to bear high latency. Our figures show that volume changes occur far more often than limit changes (c.f. Table 3), in our sample up to four times more often. This is not captured in volatility or other standard parameters usually reported for stocks.

In this study we calculated the probabilities of the occurrence of relevant volume and limit changes. The impact of latency can in this respect be regarded as an impact on the error rate of order submission.

Mean volume increases are about 147.7 % with a standard deviation of 73.5 %. But the maximum of 15 min average volume changes we found was (at 9:15-9:30 for Hannover Rück) 583.5 %. E.g. a trader with a latency of 50 ms has to expect for E.ON that there is a 2.9 % chance that her order will be "overtaken" by another incoming limit orders increasing the existing volume by 147.7 %.

Since it would be desirable to assign costs to these numbers, strategy independent models need to be applied to assess the impact of those effects on execution probabilities and then to convert these into trading costs. This extension is not in the scope of this study but builds an interesting field of future work.

As mentioned in data section an extension towards more instruments, other markets and the sub millisecond granularity constitutes a potential course for further research.

7. CONCLUSION

This paper examines the effects of latency in securities trading. Based on data for DAX30 instruments traded at Xetra fluctuations at the top of the order book are analyzed. These fluctuations encompass limit and volume changes. To assess their impact on securities trading four fundamental strategies are dealt with.

Concerning our first research question on the effects of latency we show that latency impact differs significantly among instruments: in general highly capitalized stocks exhibit higher probabilities to encounter unfavorable order book changes during the latency delay than lower capitalized ones. Among fluctuations volume changes occur twice to four times more often than limit alterations. Further, for all strategies a significant day pattern for the probability of unfavorable changes is found. Thereby, passive strategies based on non marketable limit orders are more often affected by order book changes than active ones. For commonly observed latencies at Xetra (1 to 100 ms) the dependence of probabilities for unfavorable events turns out to be nonlinearly increasing with latency. Nevertheless they can be fairly well approximated by a log-linear regression.

Concerning the scale of relative changes, limit alterations significantly decrease over the trading day whereas for volumes no common day trend can be found. Limit increases and decreases are symmetric. Further, volume increases are typically higher than decreases, which is obvious as decreases cannot exceed 100 %.

To answer our second research questions, whether these latency effects require market participants to employ low latency technology, we investigated four fundamental trading strategies. For these the calculation of directly associated cost is only applicable for active ones. Passive strategies cannot be associated with direct costs without further assumptions regarding the true underlying trading strategy. In this case we present average latency effects regarding the limit and volume effect market participants face. That way buy and sell strategies do not exhibit significant deviations.

From an exchange's customer perspective the following conclusions can be drawn: for each individual retail investor, who cannot make use of low latency technologies, price effects are neglectable. Also volume effects seem irrelevant as retail trade sizes are typically low compared to quoted best bid/ask volumes. For institutional investors the answer depends on their business model: basically for algorithmic traders latency effects yield low increases of error rates. For investors whose business follows long term profits this latency effects seem bearable. In contrast the lower the profits associated to each trade are the more fatal these effects become.

Future research steps should include an extension of the cost analysis to passive strategies and the volume effect of active strategies. Therefore it should aim at incorporating estimations for execution probabilities and models for the cost of immediacy.

8. REFERENCES

[1] Schwartz, R.A. and Francioni R. 2004. *Equity markets in action: The fundamentals of liquidity, market structure & trading*. Wiley, Hoboken, N.J.

[2] Fidessa Fragmentation Index 2010. http://fragmentation.fidessa.com/ (last accessed 06/13/2010).

[3] Schweickert, U. and Budimir, M. 2009. Latency in Electronic Securities Trading – A Proposal for Systematic Measurement. *Journal of Trading*, 4(3), 47-55.

[4] Kissell, R. and Glantz, M. 2003. Optimal Trading Strategies: Qunantitative Approches for Managing Market Impact and Trading Risk. *AMACOM*, New York.

[5] Easley D., Hendershott, T. and T. Ramadorai 2008. The price of latency, *Working Paper, Cornell University*.

[6] Kempf, A. and Korn, O. 1998. Trading System and Market Integration, *Journal of Financial Intermediation*, 7, 220–239.

[7] Theissen, E. 2000. Market structure, informational efficiency and liquidity: An experimental comparison of auction and dealer markets, *Journal of Financial Markets* 3, 333–363.

[8] Weitzel, T., Martin, S.V. and König W. 2003. Straight Through Processing auf XML Basis im Wertpapiergeschäft, *Wirtschaftsinformatik*, 45, 409–420.

[9] Deutsche Börse AG 2009. Interim Report Q1/2009, Germany. http://deutsche-boerse.com/dbag/dispatch/de/binary/gdb_content_pool/imported_files/public_files/10_downloads/14_investor_relations/51_presentations_NEW/presentation_conference_12may2009.pdf, (last accessed 04/10/2010).

[10] Domowitz, I. and Yegerman, H. 2005. The Cost of Algorithmic Trading. A First Look at Comparative Performance, *Algorithmic Trading*, 30–40.

[11] Gsell, M. and Gomber, P. 2009.Algorithmic trading engines versus human traders – Do they behave different in securities markets?, 17th ECIS, 98–109.

[12] Hendershott, T., Jones, C.M. and Menkveld, A.J. 2010. Does Algorithmic Trading Improve Liquidity?, *Forthcoming in Journal of Finance*.

[13] Grob, S. 2010. Battle for European liquidity rages in first half, http://www.thetradenews.com/asset-classes/equities/ 4785 (last accessed 07/20/2010).

[14] Chix-Europe 2010. Latency, http://www.chi-x.com/trading-on-chi-x/latency.asp, (last accessed 08/22/2010).

[15] Aldridge, I. 2010: High-frequency trading: a practical guide to algorithmic strategies and trading systems, *Wiley Trading Series*, New York.

[16] Narang, M. 2010. Tradeworx, Inc. Public Commentary on SEC Market Structure Concept Release, http://www.tradeworx.com/TWX-SEC-2010.pdf, (last accessed 04/27/2010).

[17] Jovanovic, B. Menkveld, A.J. 2010. Middlemen in Limit-Order Markets. *Working Paper, VU University Amsterdam*.

[18] Aite Group 2009. New World Order: The High Frequency Trading Community and Its Impact on Market Structure, http://www.aitegroup.com/Reports/ReportDetail.aspx?recordItemID=531, (last accessed 08/15/2010).

[19] Cave, T. 2010. High frequency traders to take 45 % market share. http://www.efinancialnews.com/story/2010-03-31/aite-high-frequency-trading-volumes, (last accessed 08/20/2010).

[20] Riordan, R. and Storkenmaier, A. 2009. Latency, Liquidity and Price Discovery, *16th Annual Meeting of the German Finance Association (DGF)*.

[21] Clemens E.K. 1991. Evaluation of strategic investments in information technology, *Communications of the ACM*, 34 (1), 22–36.

[22] Kohli, R and Grover V. 2008. Business value of IT: An essay on expanding research directions to keep with the times, *Journal of the Association for Information Systems*, 9 (1), 23–39.

[23] Brynjolfsson, E. and Hitt, L.M. 1996. Productivity, Profit and Consumer Welfare: Three Different Measures of Information Technology's Value, *MIS Quarterly*, 20 (2), 121-142.

[24] Bharadwaj, A.S., Bharadwaj, S.G. and Konsynski B.R. 1999. Information Technology Effects on Firm Performance as Measured by Tobin's q, *Management Science*, 45 (7), 1008–1024.

[25] Melville, N., Kraemer, K., Gurbaxani, V, 2004. Information Technology and Organizational Performance: An Integrative Model of IT Business Value, MIS Quaterly, 28 (2), 283–322.

[26] Schryen, G. 2010. Preserving Knowledge on IS Business Value, *Business&Information Systems Engineering*, 2 (4), 233-244.

[27] Braunwarth, K.S., Kaiser, M. and Müller A. 2010. Economic Evaluation and Optimization of the Degree of Automation in Insurance Processes, *Business&Information Systems Engineering*, 2 (1), 29-39.

[28] Hendershott, T., Moulton, P., 2007. The Shrinking New York Stock Exchange Floor and the Hybrid Market, *Working Paper, Haas School of Business and Fordham Graduate School of Business.*

[29] Moallemi, C.C. and Saglam, M. 2010. The Cost of Latency, *Working Paper*, Columnbia Universtity.

[30] Abhyankar, A., Ghost, D., Levin, E. and Limmack, R.J. 1997. Bid-Ask Spreads, Trading Volume and Volatility: Intra-Day Evidence From the London Stock Exchange, *Journal of Business & Accounting*, 24(3), 343–362.

[31] Stephan, J.A. and Whaley, R.E 1990. Intraday Price Change and Trading Volume Relations in the Stock and Stock Option Markets, *Journal of Finance*, 45(1), 191-220.