

Chapter 12

The Role of High-Frequency Trading in Modern Financial Markets

Wolfgang Eholzer and Randolph Roth

In the past decades, financial markets have undergone a profound change driven by a combination of technological advancement and fierce competition between both market participants and marketplaces. The technological advancement has moved market venues from floor to electronic venues and in that process generated substantial benefits to the investing public as cost of market entry has been drastically reduced with respect to:

- Provision of low-cost Internet-based direct electronic market access versus traditional telephone-based high-touch market access.
- Ease of access to information and therefore transparency on traded prices, available bids and offers as well as fundamentals of financial instruments have increased substantially through Internet and other electronic means.
- Markets have become more liquid in terms of displaying tight bid-offer spreads across the globe reducing the implicit transaction cost.
- Competition between trading venues and central infrastructures has reduced their fees significantly.

The largest part of this impact has materialized in the last decade, i.e., from 2000 onwards when most of the traditional floor-based markets have moved electronic and new electronic market venues such as MTFs and ECNs developed. Since the financial crisis HFT is seen by an increasing part of the public but also political and regulatory stakeholders as a source of trouble. At a first glance this seems unjustified as the reasons for the financial crisis in the USA as well in Europe had nothing to do with HFT and not a single tax Euro or USD has spent on those firms.

The purpose of this chapter is to have a more in-depth look at the role of HFT in modern electronic markets. This includes a discussion of the concerns and reservations

W. Eholzer (✉) R. Roth
Eurex Frankfurt AG, Mergenthalerallee 61, 65760 Eschborn, Germany
e-mail: wolfgang.eholzer@deutsche-boerse.com; randolf.roth@eurexchange.com

towards HFT. The chapter is structured as follows: Part 1 discusses the definition of HFT and differentiates it from algorithmic trading in general. Part 2 discusses the typical trading strategies of HFT and the concerns related to those. Part 3 addresses the question why speed is so important in (modern) financial markets and finally part 4 provides empirical evidence on the behavior of HFT. This is based on Eurex data, the largest European derivatives marketplace.

12.1 Definition of HFT and Differentiation to Algorithmic Trading

Despite the attention HFT has received by the public, law makers, and regulators, a common definition of HFT is yet to be found. This irony is best explained by starting with a look in the history.

As long as financial markets exist, the speed of receiving information and executing the resulting actions out of this information has always been a critical success factor for a significant part of the market ecosystem. Early examples include the use of pigeons by Paul Julius Reuters to transmit important stock news from the Paris Stock Exchange to Brussels or Aachen in 1850. Using carrier pigeons to relate messages between the two cities, he bridged the missing telegraph route between the terminal points of the German and the French-Belgian telegraph lines. His idea saved hours. Another example is the Chappe Telegraph. This system of communication relays, a precursor to the modern telegraph, was designed by a Frenchman named Claude Chappe. Each line consisted of signal towers built every 10–20 miles and operators in each tower kept their eye on the adjacent towers through a telescope. Using semaphore signals, they could send messages at what was then considered a staggering speed. Furthermore, the pneumatic tube system of the New York Stock Exchange (NYSE) was launched around 1930. With this, the NYSE went to a great length to ensure that speed location differences within the building did not matter, a principle which still applies in modern financial markets. As a matter of fact, this contends that the speed-sensitive exchange participants try to be in the co-location center of the exchange (comparable to the NYSE building in the 1930s) and the exchanges make sure through the same cable length and identical gear that everybody in the center is treated equally, as the pneumatic tube system did. Obviously participants who do not use the co-location center have a speed disadvantage but this affected also those who were unable to get an office in the NYSE building. The only difference is that today the space in the co-location center is virtually unlimited while in the old days, the space on the floor and in the exchange building was obviously strictly limited.

Against that background, HFT seems to be a natural evolution, resulting from two forces: fierce competition between market participants and technology advances which is used as a competitive element. A prerequisite for HFT and algorithmic trading in general has therefore been the exchanges and marketplaces being transformed into

electronic venues. The term “HFT” has been introduced around 2006, but there has been no single event which could be seen as a starting point of HFT. For example, on Eurex and its predecessors DTB and Soffex, in the 1990s, most proprietary futures trading was done manually while in options the market makers were forced from the beginning to be competitive on speed through electronic means. Today these strategies are seen as HFT, while in those days the term did not exist yet.

12.1.1 General HFT Definition

In general HFT is a technology used to implement a wide variety of trading strategies; most of these existed for many years. There are two principal ways to define HFT further; both are discussed here: a qualitative-descriptive definition as well as a mathematical-technical definition. However, as HFT is a technology, it is not possible to have a 100% clear definition of what activities or trading desks should be considered HFT and which not.

To start with the qualitative/descriptive approach, HFT is obviously a type of algorithmic trading but it needs to be differentiated from the algorithmic trading executed by institutional investors and brokers/banks acting on behalf of these participants.

Common factor of all algorithmic trading is that in general a computer generates orders without human interaction as it implements predefined and pre-parameterized trading strategies. Algorithms employed by institutional investors typically have the intention to minimize market impact of large orders by working those over time and various venues. The resulting positions are held for a relatively long period, i.e., weeks, months, or even years.

In contrast, HFT is typically characterized by trading for their own account; the ability to add, modify, and delete orders within very short time periods (Milliseconds); and the holding of positions for short (intraday) time periods. The HFT activity will be based on a latency-minimizing trading infrastructure.

The problem with the qualitative/descriptive definitions is that it is not possible to define it in a way, whereby HFT firms are characterized by criteria which all of them fulfill and at the same time these criteria do not apply to others. To make a simple example: A hedge-fund dealing on behalf of their funds trading global macro will be a very good fit to the institutional investor algorithm criteria, but from the moment that it starts to also do short-term arbitrage, it will fulfill most of the HFT criteria.

Looking from an exchange point of view at this definition problem, the exchange has the “know-your-customer” advantage. Therefore exchanges are able to do a judgment call based on customer relationships but also the analysis of the activity of the members. This in itself is obviously not a sound basis for any academic or objective research, but this “know-your-customer” information can be used to validate objective HFT definitions.

As a consequence Eurex tried to develop an objective mathematical-technical approach on how to measure HFT behavior and validated the results against the available “know-your-customer” information.

12.1.2 A Mathematical-Technical HFT Definition

The cornerstone of this approach is the assumption that all HFT firms have a high dependency between their profitability and their latency pattern when executing orders at exchanges.

In contrast to following the beaten path in trying to define HFT behavior using criteria like number of trades, overnight positions as close to flat as possible, mean reversion of positions, numerous short-lived orders with follow-up cancellations, etc. this approach tries to use the unbiased measurable latency sensitivity of exchange participants using their transactions arriving at the exchange trading system level in comparison to competitors.

In theory, transaction arrival at the exchange level can be predicted as long as it is uncorrelated. The probability of such transactions arriving with a time difference of t is given by the following formula, where μ is defined as the mean arrival rate of incoming transactions with t being a time interval:

$$f(t) = 1 / \mu e^{[-t/\mu]}$$

Using this formula we can project the expected theoretical inter-arrival distribution and compare the result with the observation from reality for each and every trading participant.

Pretending that transaction arrival at the exchange is uncorrelated would result in the same distribution of intervals for all members (Fig. 12.1).

To simulate this, we generate a large amount of random numbers between zero and a billion. Assuming those to be time stamps of transactions arriving in our trading system we sort them in ascending order and calculate the difference between two consecutive transactions. Next, we count the occurrences of inter-arrival time interval and plot the respective chart (Fig. 12.2).

The chart depicts the number of observation for any given time interval between two consecutive messages. Taking the log of this function will put the context into a linear relation (Fig. 12.3).

Looking at the transaction arrival data from our trading system, we understand that trading is indeed correlated since we notice a massive burst of transactions around specific points of time (Fig. 12.4).

This effect can easily be explained, as all participants with a latency-sensitive trading pattern will react on the same given signal and thereby increase the number of observations of short time intervals. It is important to emphasize that we do not focus on the event itself but only on the reaction of the participants in relation to each other (Figs. 12.5 and 12.6).

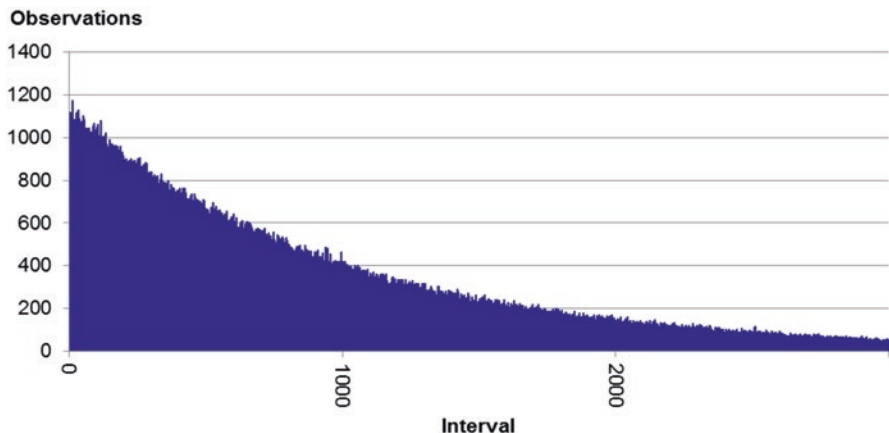


Fig. 12.1 Stylized uncorrelated transactions

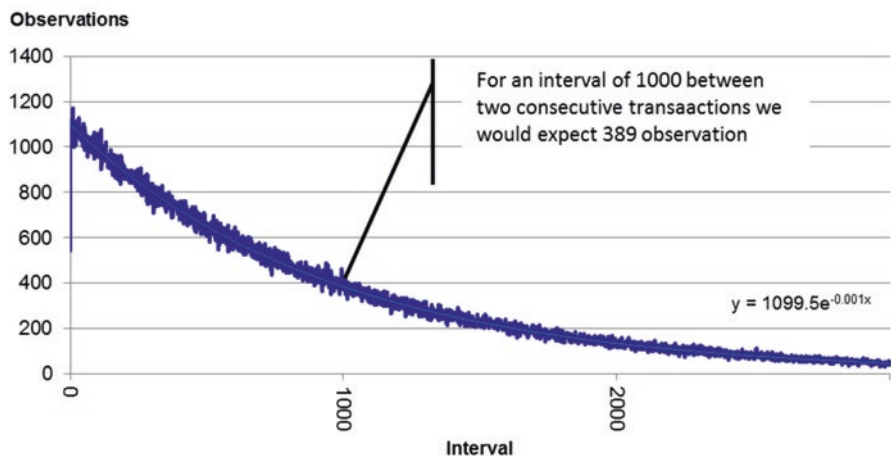


Fig. 12.2 Frequency distribution of intervals

To gain insight, we identify and analyze correlated transaction arrival in our trading system in order to define the latency sensitivity of our members. This can even be enhanced to a more granular view based on the technical connection used to send transactions or on the person responsible for the transaction, the trader.

For consistency in this approach we omit multiple consecutive transactions from the same trading participant and take those ones into consideration which come from different trading participants. As a result of this theoretical approach combined with the actual observations from our production data, we expected that the frequency of lower intervals is higher than the theoretical values for a random distribution just because of the correlated transactions.

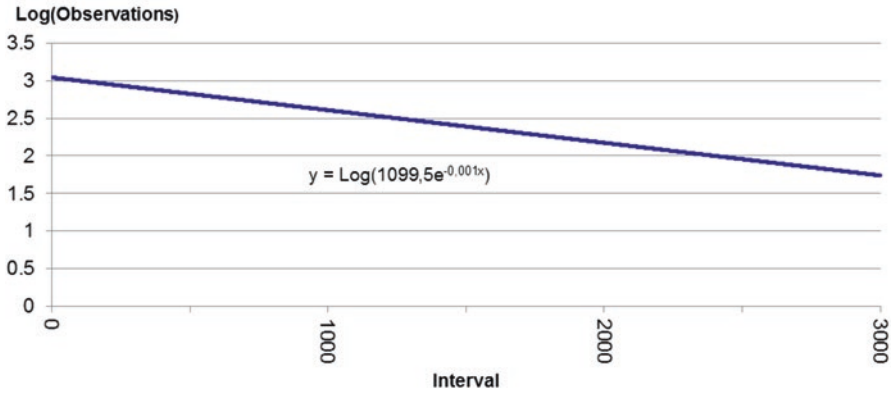


Fig. 12.3 Example of linear relation

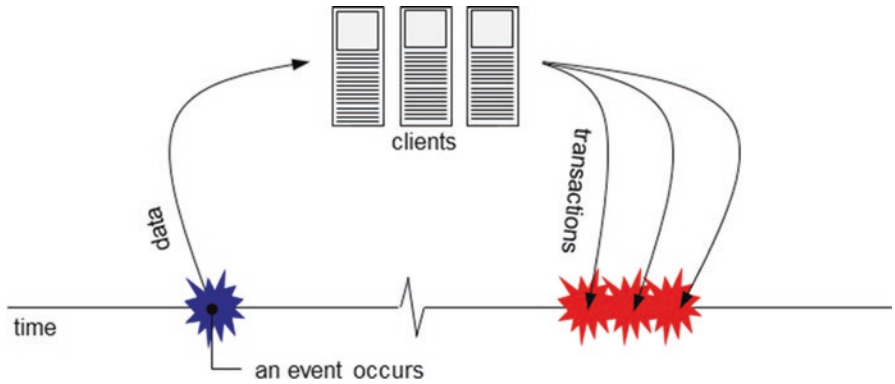


Fig. 12.4 Transaction burst

Micro bursts increase the number of observations with low inter arrival times.

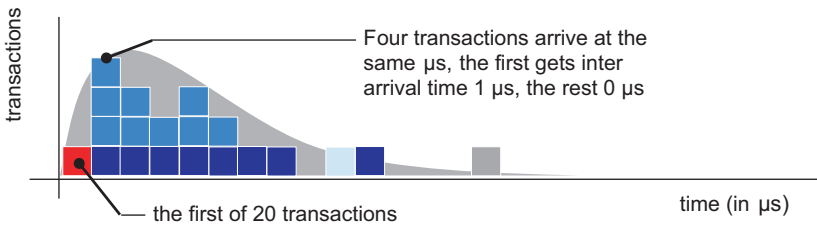


Fig. 12.5 Micro burst

Fig. 12.6 Micro burst frequency distribution

| t | E(X) | X |
|---|------|---|
| 0 | 3 | 9 |
| 1 | 2 | 8 |
| 2 | 1 | 1 |
| 3 | 0 | 0 |
| 4 | 1 | 1 |

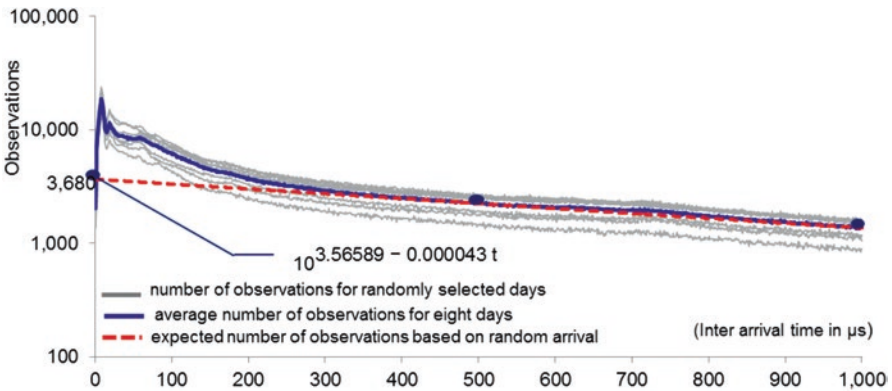


Fig. 12.7 Inter-arrival time frequency distributions

The graph below depicts the actual inter-arrival time frequency distributions for eight randomly selected days in November and December 2012. Due to the log scale of the observation axis, the relationship seems linear (Fig. 12.7).

Focusing on the interval above 500 μs, we observed the expected exponential relationship. Based on our described random process which provides a very good fit for the higher intervals, we notice up to four times (at 8 μs) more observations than expected for random arrival for lower inter-arrival times (<500 μs). Of all observations within the first millisecond above the predicted level from our random process, approximately 86% occur in the <200 μs area and approximately 67% occur in the <100 μs area. We therefore presume that most of the nonrandom arrivals are near-simultaneous reactions from strategies using HFT techniques to market data (Fig. 12.8).

Using this methodology to determine the nonrandom part of participants' transactions gives us an indication about their latency sensitivity, and hence their HFT-ness. As the measured dimension is not unswayable, this methodology only contains a decision that is subjective: whether the borderline between excess and conformity is expected with respect to the values from random distribution.

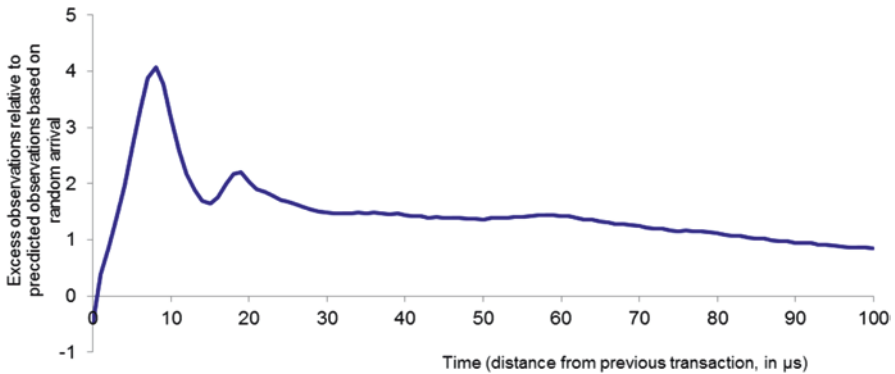


Fig. 12.8 Excess observations relative to random arrival

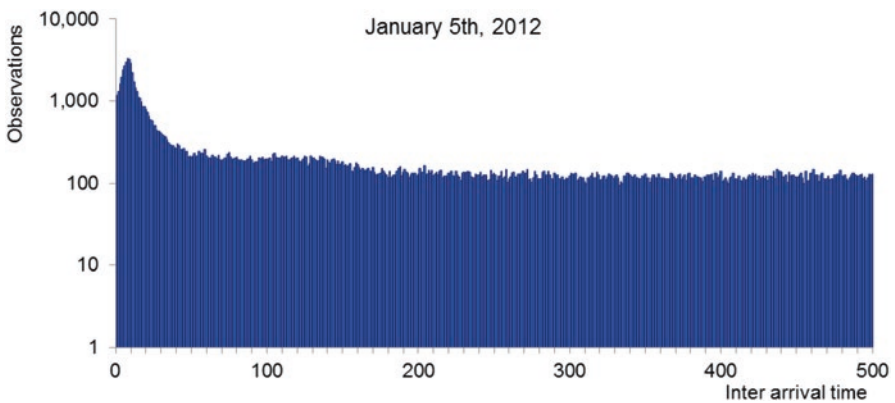


Fig. 12.9 Example accounting for a latency-sensitive participant

Here is an example accounting for a latency-sensitive participant (Fig. 12.9):

Whereas this data accounts for a clearly latency-insensitive participant (Fig. 12.10):

Using this method on actual trading data, we deliver a list of members with excess in the short intervals (see Fig. 12.11), the constituents of which are our well-known HFT participants that could also be on our list by just using the know-your-customer principle. The methodology proved to be an easy flash test to find new latency-sensitive participants.

The table only shows the relation between the intervals of 0–10 and 0–1000 μs . Typically we would expect approx. 1% to be normal; everything above is “excess” and is deemed to be a hint on HFT activity.

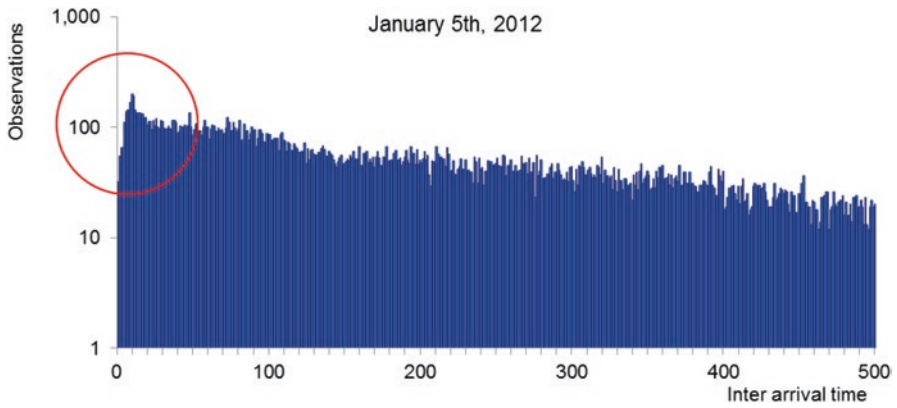


Fig. 12.10 Example accounting for a latency-insensitive participant

| Inter arrival time pattern comparison - 5 January 2012 | | | | |
|--|--------|------------------------|--------------------------|-------|
| # | Member | 0 μ s - 10 μ s | 0 μ s - 1000 μ s | Ratio |
| 1 | AAA | 13.374 | 79.517 | 17% |
| 2 | BBB | 20.491 | 125.464 | 16% |
| 3 | CCC | 24.111 | 166.402 | 14% |
| 4 | DDD | 3.325 | 25.863 | 13% |
| 5 | EEE | 1.230 | 11.307 | 11% |
| 6 | FFF | 8.081 | 78.635 | 10% |
| 7 | GGG | 1.371 | 19.704 | 7% |
| 8 | HHH | 1.842 | 29.208 | 6% |
| 9 | III | 1.660 | 26.607 | 6% |
| 10 | JJJ | 1.496 | 24.383 | 6% |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 32 | MMM | 1.143 | 43.665 | 3% |
| 33 | NNN | 35 | 1.416 | 2% |
| 34 | OOO | 163 | 6.597 | 2% |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

Fig. 12.11 Inter-arrival time pattern comparison (example)

12.2 HFT Trading Activities

Trading strategies using HFT techniques can generally be grouped into four categories.

12.2.1 *Market Making/Liquidity Provision*

A liquidity provider typically contributes two-sided orders, i.e., quotes to markets in order to earn money from the implied bid-offer spread. Typically the provider does not have a preference for one side of the order book/market.

No efficient continuously trading financial market can exist without some participants acting as liquidity provider/market makers. That means even highly liquid futures, where exchanges have usually no dedicated market making schemes, can only be robust to shocks if there are participants who act as liquidity providers.

In modern electronic markets, it is effectively not possible, to provide liquidity, without utilizing HFT technology. The reason is that liquidity providers generate quotes on a certain time-sensitive information basis. These quotes are passive, i.e., can be traded against by everybody. If the underlying information for the active quotes changes, the liquidity provider quotes are still in the market even though they are outdated. Accordingly, the liquidity provider needs to update its quote as swiftly as possible in order to avoid to be taken advantage of at its outdated prices (adverse selection).

In summary, one can implement HFT technology on non-liquidity-providing strategies (see below) but it is highly unlikely to be a liquidity provider in modern electronic markets without being seen as a low-latency trader conducting HFT. As a result, a significant portion of HFT activity is related to liquidity provision. Based on the criteria stipulated by BaFin for qualifying as HFT, we can see approximately 95% of all transactions coming from HFT participants in the liquidity provision area. The attached chart shows the development over time (Fig. 12.12).

12.2.2 *Arbitrage*

Arbitrage strategies seek to monetize price differences between identical or related instruments. Those price differences are usually short lived as they are removed by such arbitrage strategies. A classical arbitrage example would be that a stock trades at different venues at different prices. Those price differences are swiftly removed by the arbitrageurs selling the expensive stock and buying the cheap stock. An example of arbitrage between nonidentical instruments is statistical arbitrage whereby there is a statistical mispricing of one or more assets based on the expected value of these assets; that is, assets should stay, based on statistical analysis, in a certain price relationship.

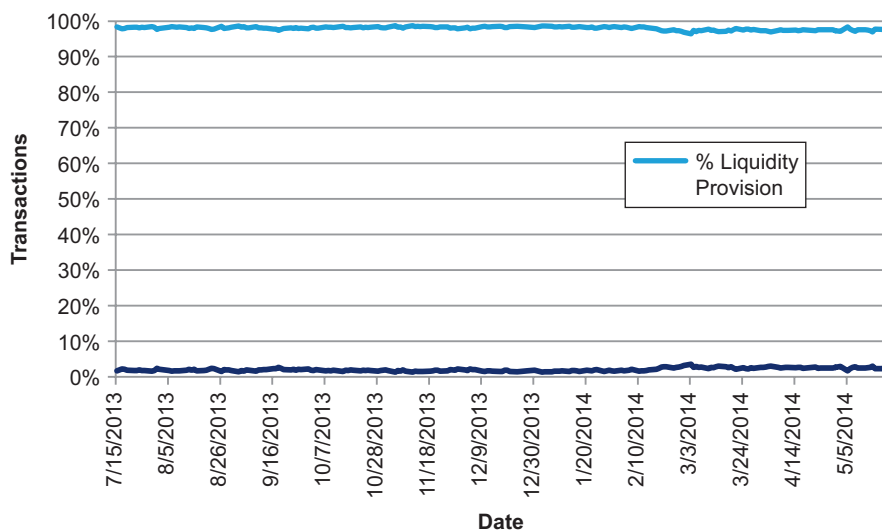


Fig. 12.12 Distribution of HFT members (BaFin criteria)

As the arbitrage opportunities are only short lived in modern financial markets, taking advantage of this requires HFT technology. Arbitrage is beneficial to the market as the prices across different venues and/or related products are on a highly competitive arbitrage-free level reducing the need to compare prices at different venues for the investing public.

12.2.3 News Trading

Unexpected news typically cause prices to move. The relative small and specialized HFT news trading community seeks to benefit from this by being the first market participant to digest and respond to usually prescheduled news. Traders who want to benefit from being able to react to news first not only need to be able to act very fast but also need to be able to understand market sentiment. For example, if an unemployment-related figure is newly reported to be worse the market reaction will depend on the ex ante expectation of the market. So even in case of a worse unemployment figure, the market might go up due to the fact that the market expected an even worse number.

12.2.4 Liquidity Detection Strategies

Liquidity detection strategies are thought to be controversial by the public and also by some institutional investors. To fully comprehend the issue, one needs to dive a little bit into the market structure.

Besides news, the driver for price shifts is large institutional orders; for example, a mutual fund decides to build a position in a certain stock whose additional demand will increase the stock price. The mutual funds' decision will have a market impact leading to a conflict between the liquidity provider (now HFT) and the institutional investor who will carry the market impact cost. This conflict is as old as there is a market price-building mechanism. To make a concrete example, we assume that the fair value of a stock is ten monetary units. The liquidity providers' offers account for 10.05. The new fair value after the institutional investor placed its large demand is at 10.20. Obviously, the institutional investor would like to get all his or her stock at 10.05. He tries to do that by taking advantage of the liquidity provided across various market venues. If the institutional investor succeeds, it still will push the fair value to 10.20, as the liquidity providers need to buy the stock back after acquiring a large short position from the institutional investor. In this case the market impact of 0.15 per stock would be completely carried by the liquidity providers, i.e., HFT, thus creating a loss for them.

Therefore, liquidity providers use statistical models to detect patterns in order to make likelihood-based calls, on where the institutional flow is going. In our example that means when the institutional investor starts buying at 10.05, it is likely that at other markets the offers will change towards 10.10 when the buying pressure continues towards 10.20 or even 10.25. The result is that the institutional investor needs to carry a large portion of its market impact on its own. This is normal but it generates frustration when the institutional investor sees at the beginning a much bigger quantity displayed at the 10.05 offers across markets than he or she is able to get at that price.

Consequently, most liquidity providers use liquidity detection strategies as a defensive measure. However, in competitive electronic markets, there are also HFT firms which used the liquidity detection strategies to parallelly run with the institutional flow, i.e., buying when their statistical models indicate that there is institutional buying pressure. As a result the fair price moves short term not only to 10.20 but also to 10.25 or even 10.30 due to the additional buying power.

12.3 The Importance of Speed in Modern Financial Markets

As outlined above, electronic financial markets cannot operate without liquidity providers and those firms need to employ HFT technology. The biggest risk of liquidity providers is that they are not able to update their quotes or orders swiftly enough when new information arrives. Updating the quote/order has three components:

- Receive the new information
- Calculate the new prices/order parameters
- Replace the outdated quotes at the market venue

Accordingly, liquidity providers need to invest in technology/speed in all of those three dimensions: The faster the liquidity provider, the smaller the risk, and the higher its liquidity contribution.

A minimum order resting time will inhibit the liquidity provider in the third dimension. It will create a situation in which the liquidity provider is unable to remove its outdated prices while aggressing strategies can take advantage of the outdated prices. As an immediate consequence, the liquidity provider will either massively increase its spread or leave the market completely.

The long-term consequence is always the same: The market gets uncompetitive and will move in a different jurisdiction.

An internal Eurex study completed in 2012, when the discussions around order resting times (ORT) started, can give insight into this effect and predict the increase of spreads caused by a potential order resting time based on empirical data.

For analyzing the reluctance of liquidity providers to take volatility-based risk we took several days from August 2012 and calculated the average spread quoted for the front month of the DAX future (FDAX).

In order to predict the impact on spreads caused by a potential ORT we needed to analyze the relation between the probability that prices move before participants can update their orders (volatility) and the inherent compensation for risks taken when providing the visible liquidity (spread). To get the most granular insight into the risk aversion we calculated the standard deviation of the mid-price between the best available bid and ask prices from 1 ms to the next (Fig. 12.13).

The graph shows ten daily observations of the spread and volatility on a 1 ms basis. The green dot represents a high-volatility day, August 12th, 2012. We can derive from the chart that the spread increases with the standard deviation by a factor of ten and changes in volatility explain 60% of changes in spreads. To gather information about the volatility component we measure the standard deviation for several fixed time frames; 50, 100, 250, 500, and 1000 ms.

Fig. 12.14 depicts how volatility increases on a micro level dependent on the measured time frame.

Each line in the chart represents a day, whereas the dots represent the time frames where volatility was measured.

Pretending that the actual time frame it takes to add or delete an order in our system (approx. 1 ms) is an ORT itself and is a crucial part of the above-shown figures in combination with the assumption that the risk aversion of liquidity providers stays unchanged, we are able to predict spreads for different order resting times (Fig. 12.15).

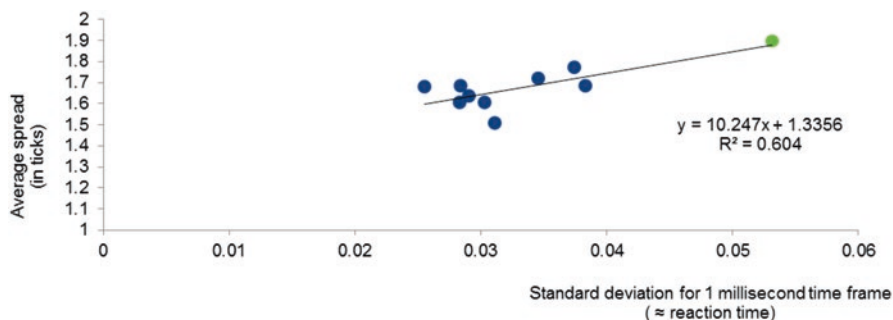


Fig. 12.13 Overview of daily observations of spread and volatility on millisecond basis

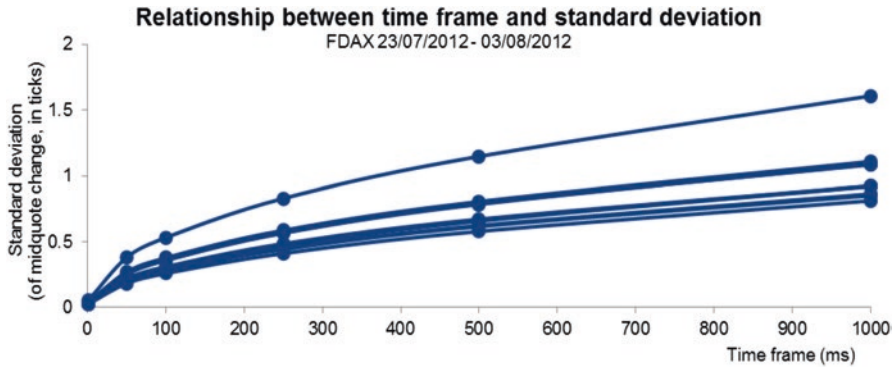


Fig. 12.14 Relationship between time frame and standard deviation

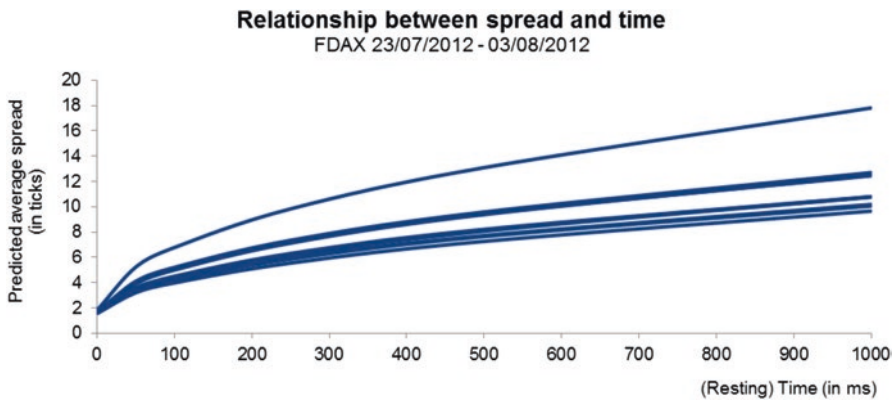


Fig. 12.15 Relationship between spread and time

The graph shows the estimated spread for 10 different days (lines) and several resting times. Eye-catching is the concave positive relationship between minimum order resting time and spread.

Reading the chart, a resting time of, e.g., 500 ms on a normal trading day would lead to a spread widening of 400 %, from 1.5 ticks to 9 ticks. On a very volatile day, a spread widening of 600 %, from 2 ticks to 12 ticks, can be expected.

So in summary, modern electronic markets require participants to provide liquidity. A prerequisite of liquidity provision is to be competitive on speed, as speed is and always has been a predominant competitive element in trading. In other words, no matter how often or seldom the information related to a specific instrument change, if it changes, it takes the fastest and best technology available to be a competitive liquidity provider; these days, using this technology is called HFT.

12.4 Empirical Evidence on the Behavior of HFT

Market quality as such can be defined as a function of spread width and book depth. The Eurex Liquidity Measure (ELM) is ideal to get a first indication about those components. The ELM measures the round-trip market impact cost of somebody executing a €10 million market order against the public order book. It consists of two components: The liquidity premium (LP) measures the spread cost of a simple 1-lot round-trip market order, and the advanced price movement (APM), which measures the additional market impact cost when a €10 million market order is executed in a round trip via market orders, in the DAX future (FDAX).

It is therefore mirroring the displayed size in the order book: The larger the price impact, as measured by the ELM, the smaller the available size in the order book. At times of crisis, the market impact cost increases, as participants scale down their risk profile, implying somewhat wider spreads and significantly reduced sizes. In general the liquidity readily available in the order book is slightly worse in Q3 2012 than in 2005 (Fig. 12.16).

This is an effect of a change in market behavior. A major driver for this is the use of execution algorithms by the buy side, which has vastly reduced the placement of large resting orders by the buy side in the transparent order book.

HFT adds significant liquidity, but their order sizes are typically smaller even though the orders are faster compared to other market participants. This ensures that participants get optimal execution even on a microsecond scale.

The ELM is because of the change in market behavior, a suboptimal indicator, and we will focus on spread resilience instead, i.e., how fast the spread between bid and ask recovers after a large trade hit the order book. For a more precise view on the topic we separated the analysis into two parts:

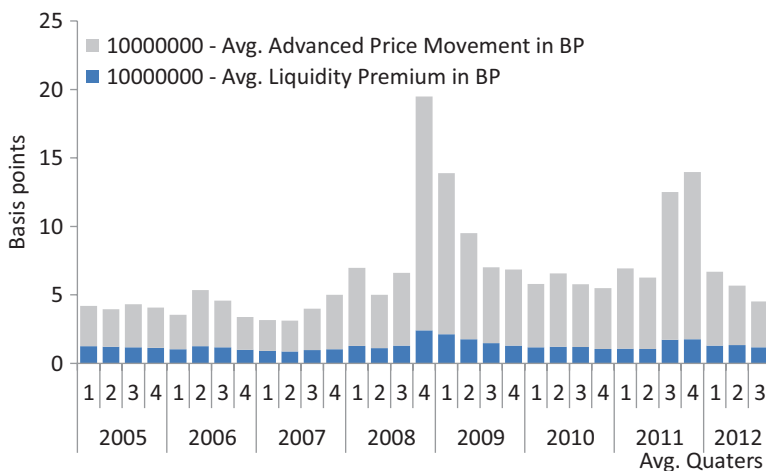


Fig. 12.16 Example of the Eurex liquidity measure

Order book liquidity share of low-latency participants at the “best bid offer” (BBO) during average-day trading

At first, we have to define how we quantify the resilience. Therefore we calculate the average traded size of a product (here: FDAX front month). We assume that trade sizes of at least ten times the median will have enough impact to move bid or ask; on the other hand we ensure using that number that we have enough samples to check the quality of our results. Fig. 12.17 shows a stylized picture of the expected market behavior before and after a large buy order hit the order book.

Focusing on timescales, where a human interaction is nearly impossible (<200 ms), and using data from 2010 to 2012, the result shows a much faster return to former levels in the spread in 2012 compared to 2010 after a hit (Fig. 12.18).

Compared to 2010, the liquidity in the DAX futures (FDAX) became much more resilient. The averages of 2010 and 2012 converge around 500 ms after a big trade.

As we see a lot of trading activity on the back of large trades, the faster return to smaller spreads increases the quality of the executions of those related trades. Taking into consideration that nearly no human interaction can take place in such short timescales, it seems justifiable to attribute that positive aspect to low-latency participants.

Order book liquidity share of low-latency participants at BBO in crisis times

Even if the public perception is that low-latency participants might provide liquidity during normal market circumstances, the majority highly doubt that those participants are also providing liquidity during crisis times. Market turbulence in combination with exceptional high volatility levels often puts low-latency trading participants in the spotlight. It is publicly assumed that price volatility would be significantly reduced if high-speed trading did not exist.

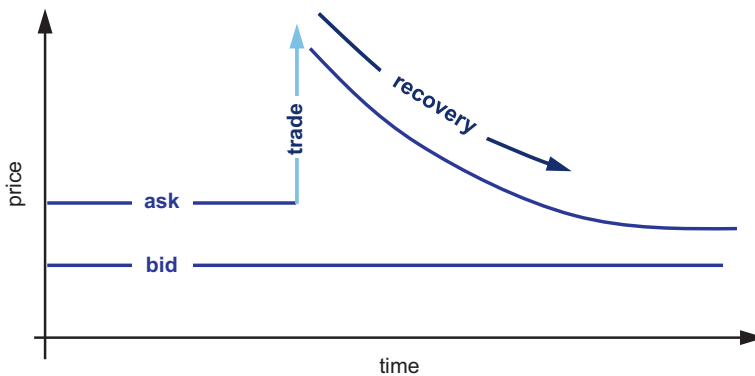


Fig. 12.17 Example for expected market behavior before and after a large buy order

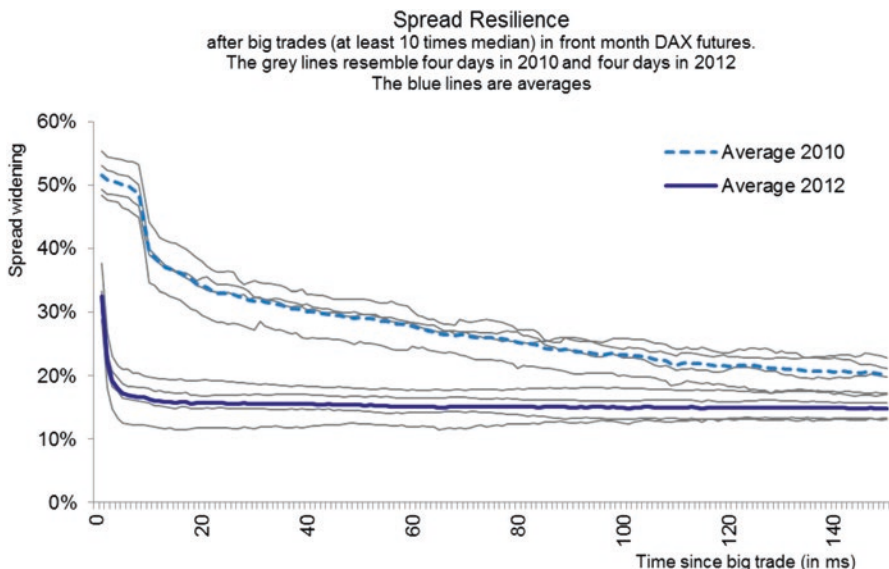


Fig. 12.18 Example for spread resilience after big trades

To verify or falsify this allegation, Eurex investigated some significant market situations where the products in scope moved a huge percentage and especially examined the participation of low-latency participants during the event and in the immediate aftermath of such an event.

12.4.1 August 25th, 2011 Futures on DAX (FDAX)

In the afternoon of August 25th, 2011, the FDAX lost more than 4% of its value within 17 min, only to reverse this move by 2% within minutes. The decline was caused by a big institutional order, which was sliced and diced by algorithms into a large number of smaller sell orders flooding the market during that relevant period of time. The total amount of the sliced and diced orders sent was 6000 contracts.

At the starting point, the order book was highly liquid with an average volume per minute of slightly above the monthly traded minute average of 300 contracts and around 60 members active on the bid and ask side of the order book.

During the peak minute at 16:02 a high number of small orders were processed with only small price increments, causing a peak turnover of 4700 contracts during that particular minute. The number of participating members during that minute doubled (Fig. 12.19).

The high number of members involved on both sides of the market during this event shows the high variety of trading interests in our markets, a key driver for liquidity and quality.

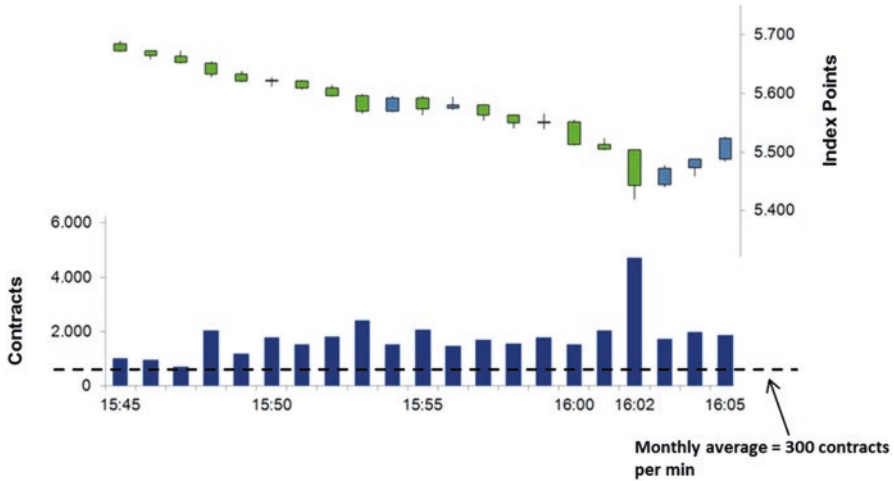


Fig. 12.19 Development of Futures on DAX (FDAX) on August 25th, 2011

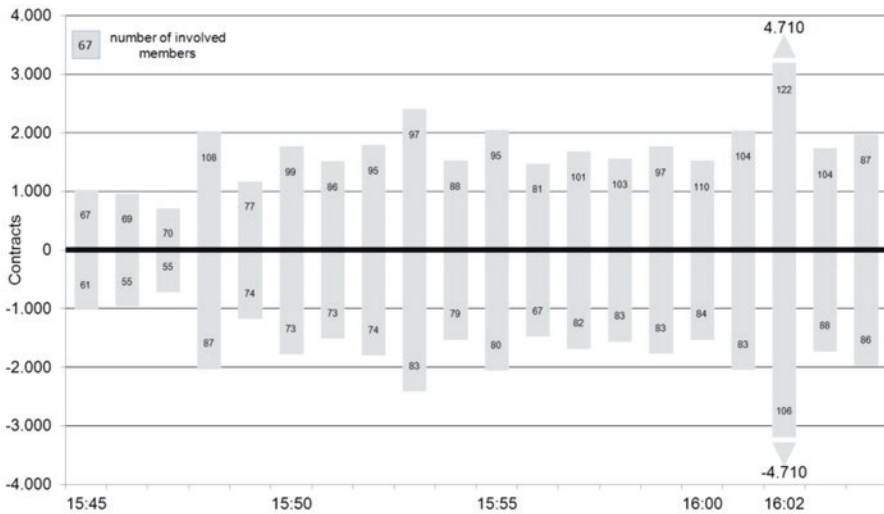


Fig. 12.20 Involved members in Futures on DAX (FDAX) on August 25th, 2011

The total of around 200 different trading members acted during the time slice in scope as buyers in a falling market, including but not limited to low-latency participants (Fig. 12.20).

A large junk of the enormous liquidity was provided by low-latency participants applying liquidity provision and arbitrage strategies. The often-heard allegation that the strong movements are accelerated by computer-based trading strategies of low-latency participants cashing in by simply using their speed could not be observed. For more details and an insight on market activity during volatile periods

of trading see our homepage at http://www.eurexchange.com/exchange-en/technology/high-frequency_trading/ with videos on low-latency trading activity.

12.4.2 April 6th, 2014, Futures on DAX (FDAX)

On 6 February 2014 at 13:45 CET the ECB was scheduled to publish its announcement on interest rates. The publication of the rate decision followed the standard ECB protocol and was in line with market expectations. As scheduled ECB decisions always have a potential to move prices/markets; typically the order books and trading as such tend to get thinner the closer the deadline for the announcements gets. This is in anticipation of the potentially market-moving information and is typically adjusted back to normal as soon as the information is released. On the particular day in scope, just 4 s and 403 ms after the ECB’s announcement on interest rates was made public, a strong selling pressure emerged in the FDAX in form of sell orders. In the following 414 ms those orders started to push prices sharply lower while a total of 49 sellers and 82 buyers traded 1488 contracts (Fig. 12.21).

Such a situation can arise from the fact that one or more participants are placing one or some large orders in the order book to adjust their respective position in relation to the just published information. In cases where the size of the orders surmounts the available, still decreased liquidity in the order book, such an order entry, can cause a move in prices. Where under normal circumstances this happens only very seldom, it still might happen at any time during trading. Accordingly, the Eurex T7 trading system has built-in functionality to safeguard and manage the

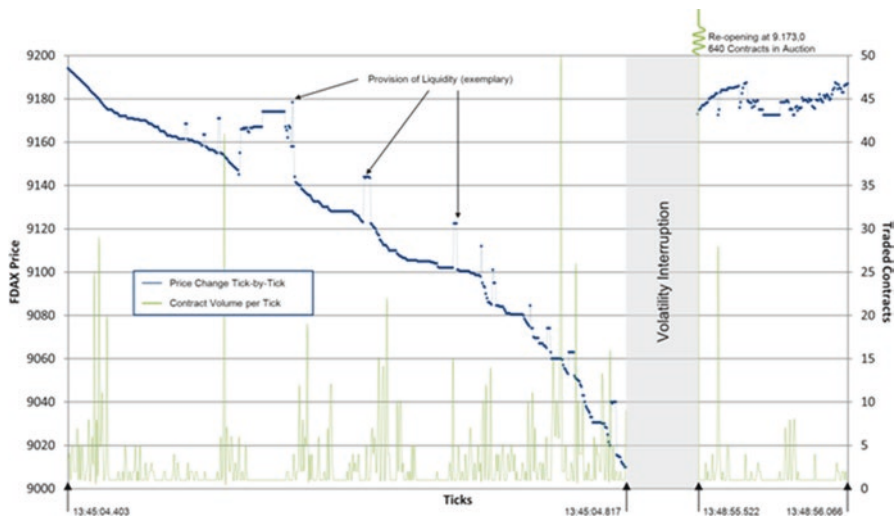


Fig. 12.21 Price development and traded volume in FDAX on February 6th, 2014

impact of such events. Also in this event those safeguards worked out and halted the market to guarantee fair and orderly market conditions and executions.

The up spikes in the chart show clearly that new orders, coming into the order book, provided new liquidity at a better level and market orders resting in the book, because of our market order matching range that halted the execution of those due to the missing price references, were executed at better levels than it would have been possible before.

To stress the fact that we talk about a time slice of 414 ms, we tend to attribute at least the newly provided liquidity again to the low-latency participants, as a human reaction in the time interval is at least very unlikely.

12.4.3 Effects of Low-Latency Participants Engaging in New Products

As already stated, low-latency participants play an important role when it drills down to liquidity provision and market depth (order book elasticity). Their ability to digest news and market information at high speed allows them to be in or back in the book faster than anyone else. This increases the quality of executions after the release of news or big orders hit the book and caused price moves.

Especially in new products, where the liquidity is still growing, low-latency participants can be extremely helpful and important. In order to gain insight into the importance and the effects of low-latency participants engaging in new products we researched the Futures on Italian (FBTP) and French Government bonds (FOAT) traded at Eurex. The products were introduced on Sept 14th, 2009 (FBTP), and April 16th, 2012 (FOAT).

Contrary to general market development, OAT and BTP futures performed pretty well in 2012 and gained even more market acceptance. During the month of August 2012, the spread in both products experienced a stellar improvement.

The development did not seem gradual, and all signs were that a structural break around August 17th occurred. This break is most noticeable in the development of the spread quality depicted as the one tick spread percentage of the day in Figs. 12.22 and 12.23.

As an additional hint for a structural break, the size available at the improved spread can be taken, which also increased starting August 17th, 2012 (Fig. 12.24), for the FBTP. While there is a clear signal in the FBTP data, there is no clear picture in the FOAT (Fig. 12.25).

When looking at the average spread in the two products we again got confirmation of a structural break in the FBTP and no signal for such a break in FOAT. Even though the average spread in FOAT is also decreasing over the period in scope it is not clearly attributable to a particular date (Figs. 12.26 and 12.27).

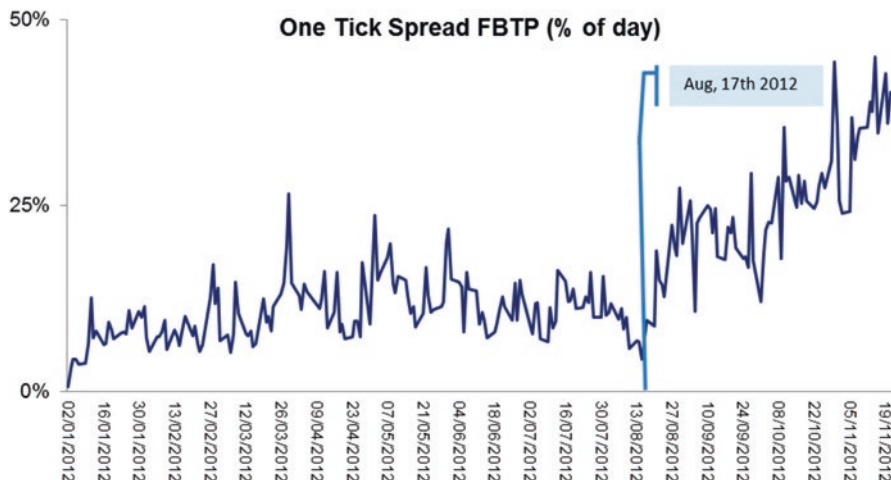


Fig. 12.22 One tick spread in FBTP

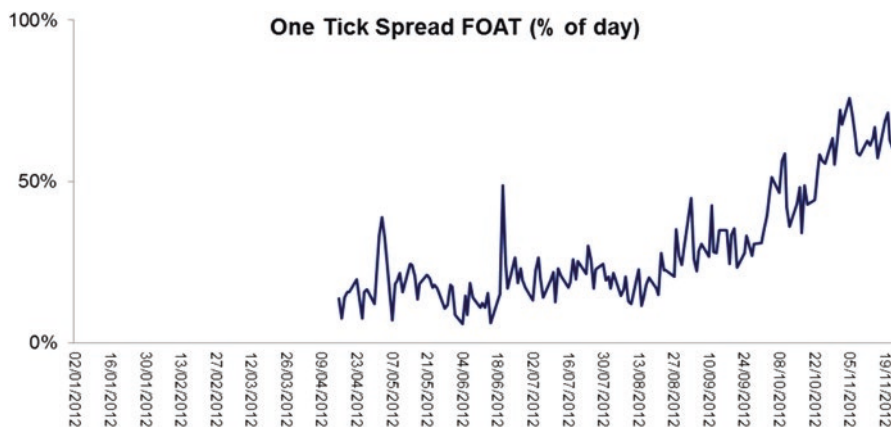


Fig. 12.23 One tick spread in FOAT

A major, identifiable market structural development that took place on July 23rd might at least partially explain the visible change in FBTP markets starting August 17th, 2012.

Here is what happened: up to July 20th, one particular low-latency participant (AAA) provided close to 10% of the BBO on a daily basis. On the next trading day (July 23rd, 2012) another low-latency participant (BBB) joined in providing the BBO and took over the position of AAA on Aug 17th, 2012, by providing a

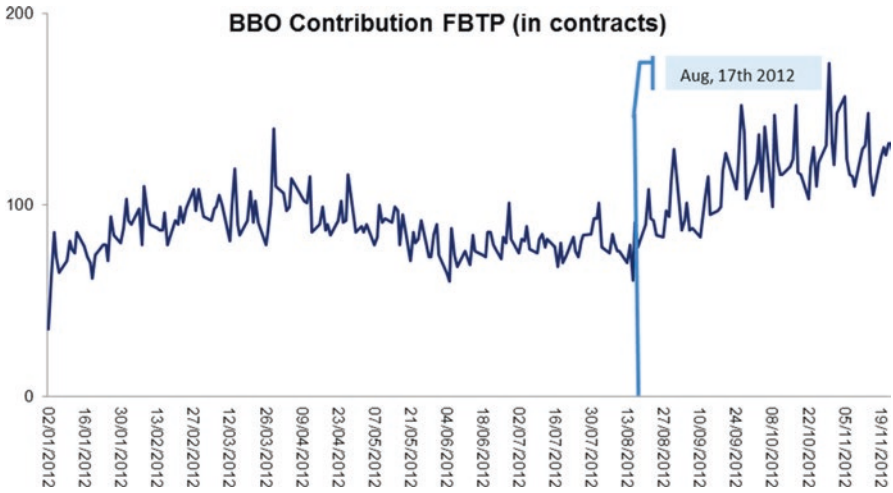


Fig. 12.24 Size available at improved spread for FBTP

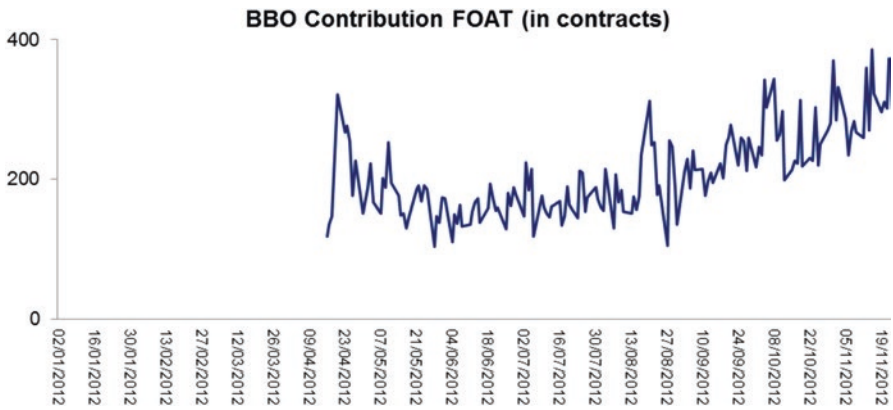


Fig. 12.25 Size available at improved spread for FOAT

larger share of the BBO on a daily basis, while participant AAA left the market over the following 5 days.

Our thesis is that participant BBB’s business model is very similar to AAA’s, but faster in the execution.

When AAA realized that another participant is running in parallel, but faster in taking decisions and sending orders, AAA specialized in areas of competence with less competition.

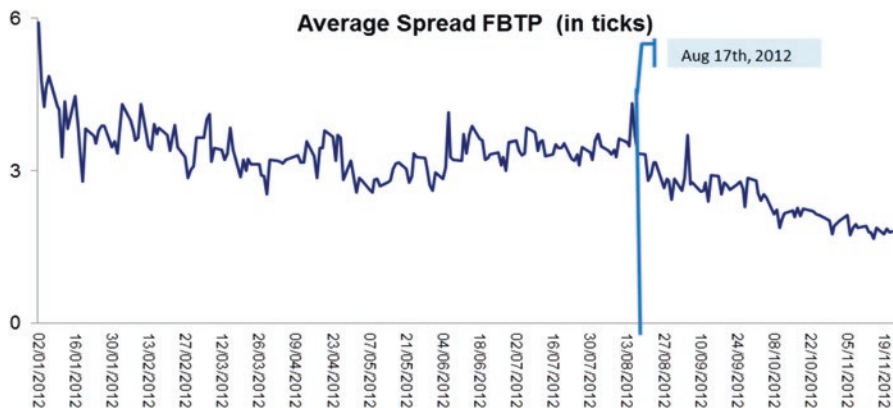


Fig. 12.26 Average spread in FBTP

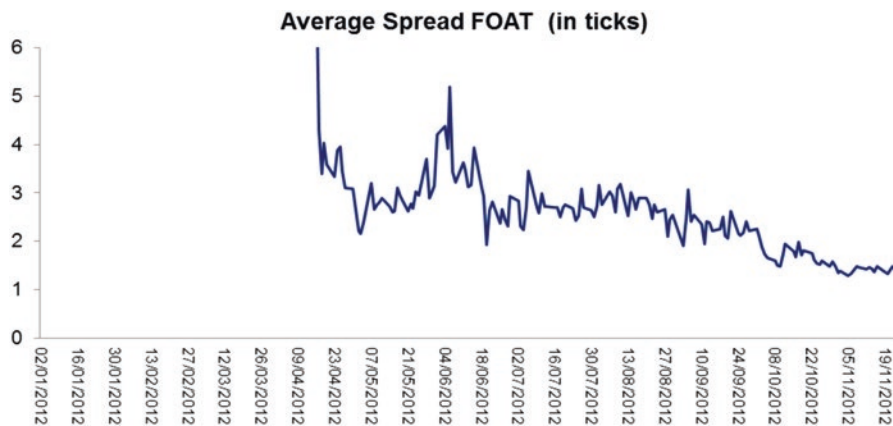


Fig. 12.27 Average spread in FOAT

One piece of evidence in favor of this thesis is the improvement of spread resilience following trades, which is to the benefit of all participants as the executions following relatively large trades would otherwise be suboptimal (at worse prices). A remarkable improvement (red) took place on the date BBB entered the market. This improvement did not take place in OAT futures, where BBB did not enter the market. It shows clearly that the new entry of a low-latency participant into a product can significantly increase the quality of executions (Figs. 12.28 and 12.29).

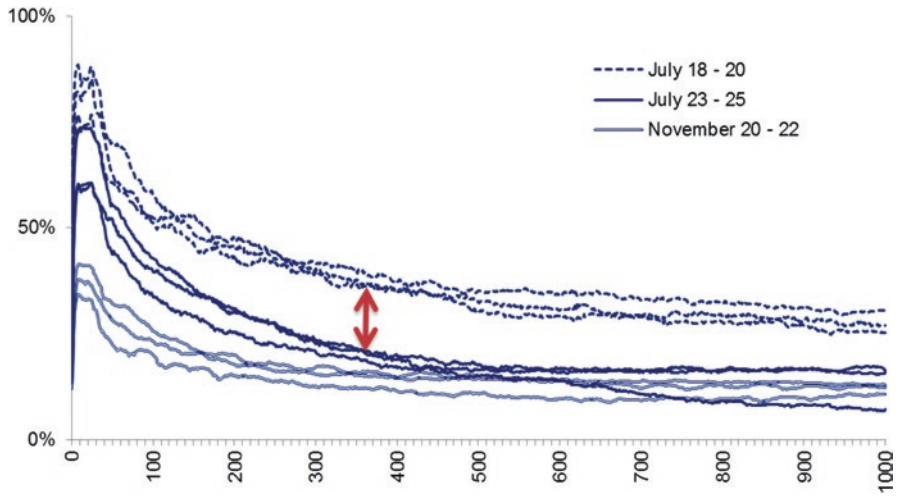


Fig. 12.28 Resilience in FBTP

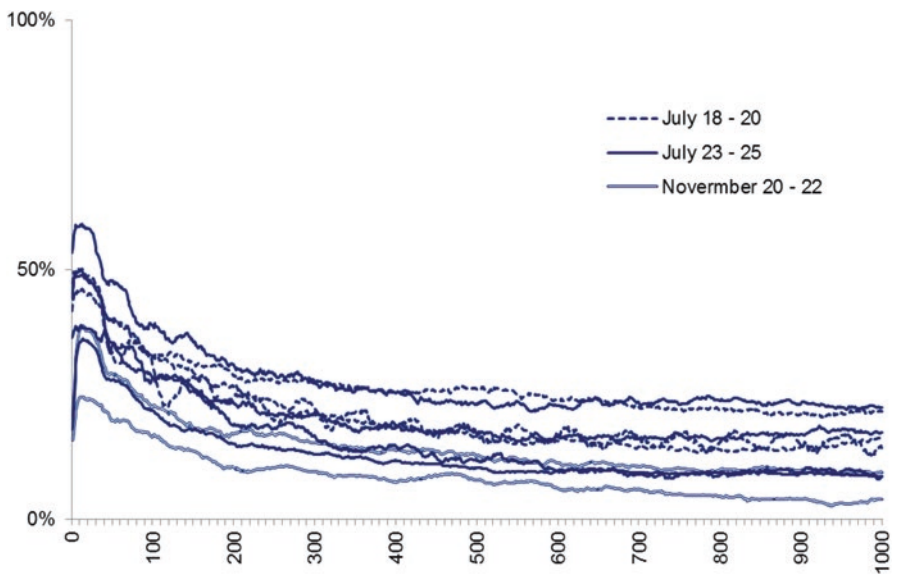


Fig. 12.29 Resilience in FOAT

Bibliography

- Bloomberg, Published Article, Apr 2nd, 2014. 'High-Speed Trading Used to Mean Carrier Pigeons'. <http://www.bloombergtv.com/articles/2014-04-02/high-speed-trading-used-to-be-for-the-birds>
- Hagströmer B. & Norden L. 'The Diversity of High Frequency'. Stockholm University of Business. September 27, 2012.
- Lee H. G. & Clark T. H. 'Impacts of the Electronic Marketplace on Transaction Cost and Market Structure' International Journal of Electronic Commerce volume 1, Number 1, fall, 1996, pp. 127–149.
- Markham J. W. 'A Financial History of the United States. From JP Morgan to the Institutional Investor (1900–1970)'. Volume II, 2002 M.E. Sharp Inc.
- Oxera, May 2011 'Monitoring prices, costs and volumes of trading & post trading services'. Report prepared for the European Commission, DG Internal Market and Services
- Reuter, Paul Julius, Baron de. Encyclopedia Britannica (11th ed.). 1911
- U.S. Securities and Exchange Commission. March 18, 2014. 'Equity Market Structure Literature Review'. Part II : High Frequency Trading. Staff on the Division of Trading and Markets. HFT Overview, Metrics for Defining HFT, Page 4. Statistics Web F/E