ON THE IMPACT AND FUTURE OF HFT

STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

WHITE PAPER

Khaldoun Khashanah
Ionut Florescu
Steve Yang

Financial Engineering Division
School of Systems and Enterprises
Stevens Institute of Technology

May 10, 2014

This work was supported by the Investor Responsibility Research Center Institute.

ABSTRACT

This white paper informs on the state of high frequency trading (HFT) mainly in the U.S. The paper addresses three major issues: First, it addresses HFT as it is seen from various market agents' perspectives, traders, institutional investors, regulators, academicians, and the public, collectively referred to as stakeholders. The paper establishes a survey to get information on aspects of HFT. An examination of a HFT dataset verifies known trends and claims of HFT volume, price efficiency and liquidity. Second, the paper examines the imminent problems and risks seen by various stakeholders from their vantage point. An assessment of sources of risk posed by HFT to institutional investors and other components of the financial system reveals two types of risks to be examined more carefully: the first is HFT-driven systematic risk and the other is a potential HFT systemic risk. Third the paper examines possible solutions to existing issues of HFT along with recent claims. We find that there are two classes of claims of unfair practices facing HFT: one is the insider information through asymmetric access to information flows and the other is price manipulation claim. The paper introduces the concepts of information transmission distance and systemic latency. We propose a new solution based on information transmission zoning concept, which requires minimum financial information flow re-architecting and no major changes in regulation NMS.

Keywords: high frequency trading, institutional investors, data, finance, financial regulations, systemic risk, information transmission distance, systemic latency, insider information.

1 - INTRODUCTION AND LITERATURE REVIEW

High-quality trading markets promote capital formation and allocation by establishing prices for securities and by enabling investors to enter and exit their positions in securities wherever and whenever they wish to do so. The one important feature of all types of algorithmic trading strategies is to discover the underlying persistent tradable phenomena and generate trading opportunities. These trading opportunities include microsecond price movements that allow a trader to benefit from market-making trades, several minute-long strategies that trade on momentum forecasted by market microstructure theories, and several hour-long market movements that surround recurring events and deviations from statistical relationship (Aldridge (2010)). Algorithmic traders then design their trading algorithms and systems with the aim of generating signals that result in consistent positive outcomes under different market conditions. Different strategies may target different frequencies, and the profitability of a trading strategy is often measured by a certain return metric.

In particular, there is a subgroup within the algorithmic trading strategies called High Frequency Trading (HFT) strategies that have attracted a lot of attention from investors, regulators, policy makers, and academics broadly. According to the U.S. Securities and Exchange Commission, high-frequency traders are "professional traders acting in a proprietary capacity that engage in strategies that generate a large number trades on daily basis." (The SEC Concept Release on Equity Market Structure, 75 Fed. Reg. 3603, January 21, 2010). The SEC characterized HFT as (1) the use of extraordinary high-speed and sophisticated computer programs for generating, routing, and executing orders; (2) use of co-location services and individual data feeds offered by exchanges and others to minimize network and other types of latencies; (3) very short timeframes for establishing and liquidating positions; (4) the submission of numerous orders that are canceled shortly after submission; and (5) ending the trading day in as close to a zero position as possible (that is, not carrying significant, under-hedged positions over night). Although many HFT strategies exist today and they are largely unknown to the public, researchers have shed lights on their general characteristics recently. Several illustrative HFT strategies include: (1) acting as an informal or formal market-maker, (2) high-frequency relative-value trading, and (3) directional trading on news releases, order flow, or other high-frequency signals (Jones (2012)).

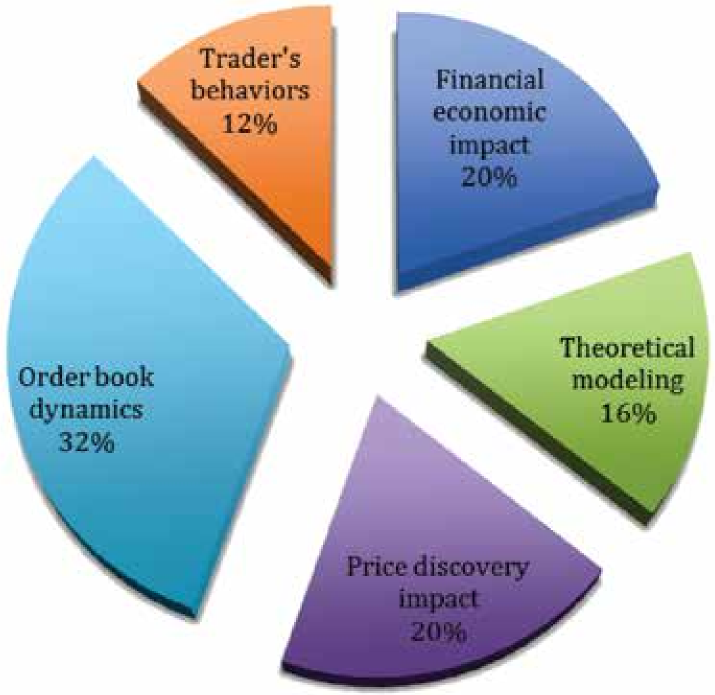

Figure 1. Academic Research Papers on Algorithmic and High Frequency Trading Practices

In the past few years, there have been a number of studies of HFT and algorithmic trading more generally. In this white paper, we surveyed the 56 academic research papers, which had significant impact on our understanding of the algorithmic trading and HFT trading. These papers cover five primary topics concerning financial economic impact, theoretical modeling, price discovery impact, limit order book dynamic modeling, and traders' behavior study of algorithmic and HFT trading practices. Figure 1 shows the distribution of these academic papers on this subject. Among these five areas are three topics, which offer direct answer to the question whether algorithmic and HFT trading provides positive or negative value to the market overall quality. The remaining 44% of the research papers look into the trading mechanics and behavior of these participants in the market. These papers lay the foundation for others to answer direct questions. Next we then need to dive into each of the four clusters of academic findings and provide a thorough review on their results.

FINANCIAL ECONOMIC RESEARCH

The most influential topic regarding algorithmic and HFT addresses the financial and economic perspective. Their primary objective is to understand the financial economic impact of these algorithmic trading practices to the market quality including liquidity, price discovery process, trading costs, etc. On the empirical side, some researchers have been able to identify a specific HFT in data, and others are able to identify whether a trade is from algorithmic traders. Given the amount of information provided by exchanges and data vendors, it is possible to describe patterns in algorithmic order submission, order cancellation, and trading behavior. It is also possible to see whether algorithmic or HFT activities are correlated with bid-ask spreads, temporary and/or permanent volatility, trading volume, and other market activity and quality measures. Hendershott et al. (2011) study the implementation of an automated quote at the New York Exchange. They conclude that the implementation of auto-quote is associated with an increase in electronic message traffic and an improvement in market quality including narrowed effective spreads, reduced adverse selection, and increase price discovery. These effects are concentrated in large-cap stocks, and there is little effect in small-cap stocks. Menkveld (2012) studies the July 2007 entry of a high-frequency market-maker into the trading of Dutch stocks. He argues that competition between trading venues facilitated the arrival of this high-frequency market-maker and HFT more generally, and he shows that high-frequency market-maker entry is associated with 23% less adverse selection. The volatility measured using 20 minutes realized volatility is unaffected by the entry of the high-frequency market-maker. Riordan et al. (2012) examine the effect of a technological upgrade on the market quality of 98 actively traded German stocks. They conclude that the ability to update quotes faster helps liquidity providers minimize their losses to liquidity demanders, and more price discovery takes place. Boehmer et al. (2012) examine international evidence on electronic message traffic and market quality across 39 stock exchanges over the 2001-2009 period. They add that co-location increases algorithmic trading and HFT, and that the introduction of colocation improves liquidity and the information efficiency of prices. However, they claim volatility does not decline as much as it would be based on the observed narrower bid-ask spreads. Gai et al. (2012) study the effect of two recent 2010 Nasdaq technology upgrades that reduce the minimum time between messages from 950 nanoseconds to 200 nanoseconds. These technological changes lead to substantial increase in the number of canceled orders without much change in overall trading volume. There is so little change in bid-ask spreads and depths. Overall, these studies have focused on empirical evidence that an increase in algorithmic trading has positive influence on market quality in general.

FINANCIAL THEORETICAL MODELING RESEARCH

The second topic focuses on the theoretical modeling of the algorithmic and HFT trading practices. There are a number of models developed to understand the economic impact of these algorithmic trading practices. Biais et al. (2012) conclude HFT can trade on new information more quickly, generating adverse selection costs. In addition, HFT requires significant fixed investments in technology. Their model shows that only sufficiently large institutions are likely to make these fixed investments. Smaller firms and investors are left to bear the adverse selection costs from HFT. Finally, they model the arms race feature of HFT. Iovanovic et al. (2010) show that HFT can avoid some adverse selection, and can provide some benefit to uninformed investors who need to trade. Their model shows that HFT can update limit orders quickly based on new information. As a result, HFT can avoid some adverse selection, and HFT can provide some of that benefit to uninformed investors who need to trade. Some of these trades might not have occurred otherwise, in which case HFT can improve welfare. Martinez et al. (2012) conclude from their model that HFT obtains and trades on information an instant before it is available to others, and it imposes adverse selection on market-makers. Therefore liquidity is worse and prices are no longer efficient. They focus on HFTs that demand liquidity, and suggest that HFT makes market prices extremely efficient by incorporating information as soon as it becomes available. Markets are not destabilized, as long as there is a population of market makers standing ready to provide liquidity at competitive prices. Foucault, Hombert, and Rosu (2012) show that HFT obtains and trades on information an instant before it is available to others. This imposes adverse selection on market-makers, so liquidity is worse, and prices are no more efficient. Pagnotta et al. (2012) focus on the investment in speed made by exchanges in order to attract trading volume from speed sensitive investors. Moallemi et al. (2012) argue that a reduction in latency allows limit order submitters to update their orders more quickly, thereby reducing the value of the trading option that a limit order grants to a liquidity demander. The common theme in these models is that HFT may increase adverse selection, and it is harmful for liquidity. However, the ability to intermediate traders who arrived at different times is generally good for liquidity.

ORDER BOOK DYNAMICS MODELING STUDIES

The third topic area is concerned with modeling limit order book dynamics. Although these papers do not provide direct interpretation of influences of the algorithmic and HFT trading practices, they nevertheless offer great insight for researchers to understand the mechanics of these automated trading practices. Albert J. Menkveld (2007) looks to extend the Chowdhry and Nanda (1991) model to detect the presence of order-splitting traders across real world markets, in hopes of understanding the effects of trading in the fragmented markets. He observes that in the last few decades, it has become common for firms to cross-list their shares on different foreign exchange markets, which has proved to benefit firms by reducing the cost of capital and enhancing the liquidity of the stock. He concludes that it is the arrival of large liquidity trader volume and the lower profits of informed traders that make the market more liquid in the overlap. Through empirical data, the paper finds that order-splitting as order imbalance is positively correlated across markets in the overlap and in the cross-section of British stocks, it significantly increases with NYSE small liquidity trading. John Y. Campbell et al. (2005) look at high-frequency trading information on equity transactions and quarterly information on institutional equity holdings to draw conclusions about institutional equity ownership. Changes in institutional ownership and order flow were then used to show short-term covariance between institutional flows and equity returns across a broad selection of stocks during the years of 1999-2000. They created a new method that gives results such that smaller buy volume is associated with decreasing institutional ownership and large buy volume is associated with increasing institutional ownership. Extremely small buys also predict increasing institutional ownership which suggests that institutions use the trades to test the liquidity of the market, to round small positions up or down, or to hide their activity. David Easley et al. (2012) present a new method of estimating flow toxicity based on volume imbalance and trade intensity (VPIN). They assert that order flow is toxic when it adversely selects market makers, who may be providing liquidity at a loss unknowingly. They suggest that VPIN can be a valuable risk management tool. Results shows that high levels of VPIN signify a high risk of subsequent large price movements, deriving from the effects of toxicity on liquidity provision. Boyan Jovanovic and Albert J. Menkveld (2012) study how high frequency trading might affect investor welfare in standard limit-order markets both theoretically and empirically. They document that a competitive sector of middlemen (high frequency traders) might reduce the informational friction, and therefore improve welfare, as information technology is at the heart of what they do. Their model also implies that regulations or fee structures that include HFTs to shift from producing price quotes to consuming them could result in substantial welfare losses. Joel Hasbrouck (2012) studies variance on time scales of as small as fifty milliseconds for the National Best Bid and Offer in the US equity market. He shows that the highest quoted volatilities occurred during the 2004-2006 time period, which ultimately corresponds to the transition to electronic trading in the markets. Based on empirical evidence, he concludes sub-second high frequency variance for the National Best Bid and Offer are in excess of that would be expected when compared to random-walk volatility over longer interval. These changes in volatility may be attributed to the change in market environment and the change to electronic trading. Joel Hasbrouck and Gideon Saar (2013) propose a new measure of low-latency activity in order to discover the impact of high frequency trading. This new measure is used to study how low-latency activity affects market quality during normal market conditions and times of economic uncertainty. They conclude that increased low-latency activity improves market quality in the area of liquidity and short-term volatility. This type of behavior is true for both normal market activity and declining prices.

TRADING STRATEGIES STUDIES

The forth topic addresses price discovery process with respect to algorithmic and high frequency trading practice and their impact. As it is commonly acknowledged, price discovery is a way to measure efficiency of the market. Frank Zhang (2010) examines the implication of high frequency trading for stock price volatility and price discovery. He documents that HFT has become a dominant driver of trading volume in the U.S. capital market, and HFT strategies are agnostic to a stock's price level and have no intrinsic interest in the fate of companies, leaving little room for a firm's fundamentals to play a direct role in its trading strategies. He finds that HFT is positively correlated with stock price volatility after controlling for firm fundamental volatility and other exogenous determinants of volatility. He also finds that HFT is negatively related to the market's ability to incorporate information about firm fundamentals into asset prices, and stock prices tend to overreact to fundamental news when HFT trading is at the high volume. Ryan Riodan and Andreas Storkenmaier (2011) document that the speed of trading is an important factor in modern security markets, although relatively little is know about the effect of speed on liquidity and price discovery, two important aspects of market quality. Their results show that decreasing the latency in a market leads to increased liquidity, mostly in small and medium sized stocks; the efficiency of prices clearly improves post upgrade, as does the relative contribution of quotes to price discovery. Their results also highlight a lack of competition between liquidity suppliers, as the realized spread increases fourfold. This translates into an increase in liquidity supplier revenues of roughly 185 million euros for the entire sample. Terrance Hendershott and Ryan Riordan (2011) examine the role of high-frequency traders in price discovery. They conclude that HFT plays a positive role in price efficiency by trading in the direction of permanent price changes and in opposite direction of transitory pricing errors on average days and the highest volatility days. HFT passive non-marketable orders are adversely selected in terms of the permanent and transitory components of these traders are in the direction opposite to permanent price changes and in the same direction as the pricing errors. They conclude that there is no evidence to say that HFT contribute to market instability in prices, in contrast, HFT overall trades in the direction of reducing transitory pricing errors both on average days and on the most volatile days. David Easley et al. (2013) examine the impact on stock and market after a major upgrade that happened to the New York Stock Exchange in 1980 to improve its environment. This increase in transparency and reduction in transaction latency allowed off-floor traders to condition their orders on more up-to-date information and reduced the free trading option that their limit orders provide. They also conclude that the competition enhancing upgrades also generated relatively greater turnover and relatively lower transaction costs. The results of their study indicate that the latency that the traders experience is important for market participants and exchange alike. The results also suggest that leveling the playing field between the public and intermediaries leads to higher liquidity and higher prices. In our own study Bozdog et.al. (2011), we have discovered that mini market crashes are a much more often occurrence than previously known. We have created an algorithm to detect these mini-crashes, which we call rare events and we show that they are related to pressure in the market and a lack of liquidity existing in the market at the time of those events.

HF TRADERS BEHAVIORAL STUDIES

Moreover, there have been a number of studies focused on algorithmic traders' behaviors. These studies examine the trading activities of different types of traders and try to distinguish their behavioral differences. Hendershott et al. (2012) use exchange classifications o distinguish algorithmic traders from orders managed by humans. They document that algorithmic traders concentrate in smaller trade sizes, while large block trades of 5,000 shares or more are predominantly originated by human traders. Algorithmic traders consume liquidity when bid-ask spreads are relatively narrow, and they supply liquidity when bid-ask spreads are relatively wide. This suggests that algorithmic traders provide a more consistent level of liquidity through time. Brogaard (2012) and Hendershott et al. (2011) work with Nasdaq data and show whether trades involve HFT. Hendershott et al. (2011) find that HFT accounts for about 42% of (double-counted) Nasdaq volume in large-cap stocks but only about 17% of volume in small-cap stocks. They estimate a state-space model that decomposes price changes into permanent and temporary components, and measures the contribution of HFT and non-HFT liquidity supply and liquidity demand to each of these price change components. They find that when HFTs initiate trades, they trade in the opposite direction to the transitory component of prices. Thus, HFTs contribute to price discovery and contribute to efficient stock prices. Brogaard (2012) similarly finds that 68% of trades have an HFT on at least one side of the transaction, and he also finds that HFT participation rates are higher for stocks with high share prices, large market caps, narrow bid-ask spreads, or low stock-specific volatility. He estimates a vector autoregressive permanent price impact model and finds that HFT liquidity suppliers face less adverse selection than non-HFT liquidity suppliers, suggesting that they are somewhat judicious in supplying liquidity. Kirilenko et al. (2011) use account-level tick-by-tick data on the E-Mini S&P 500 futures contract, and they classify traders into various categories, including HFTs, opportunistic traders, fundamental traders and noise traders. Benos et al. (2012) conduct a similar analysis using UK equity data. These different datasets provide considerable insight into overall HFT trading behavior.

One of the goals of this study is to provide a comprehensive overview of the current academic research in HFT, so that investment community and the public in general will be well informed of our current understanding of HFT and their influences related to such important economic issues as multiple characterizations of price formation processes, market liquidity, and order flow, etc. We assert that enhanced understanding of the economic implication of these different algorithmic and HFT trading strategies will yield quantitative evidence of value to market policy makers and regulators seeking to maintain transparency, fairness and overall health in the financial markets. Overall, although there are still differences in opinion with regard to HFT and their impact to the market quality, a general consensus suggests that HFT provides liquidity and on average improves market quality, with more discernible positive effects in large-cap stocks. However, under distressed market conditions such as the 2010 Flash Crash, HFTs reportedly played a very different role. Kirilenko, Kyle, Samadi, and Tuzun (2011) study HFT in the E-Mini S&P 500 futures market during the Flash Crash. Using audit trail data for nearly 15,000 accounts traded the E-Mini that day, and they find that HFT did not trigger the Flash Crash, but their responses to the unusually large selling pressure on that day exacerbated the decline and worsened market volatility. In particular, as a large number of aggressive sell orders arrived, HFT initially provided liquidity. Within a few minutes, possibly because they were overwhelmed by selling pressure, HFT's reversed course and aggressively liquidated their long positions, and thereby contributing to the price decline. The SEC, the national exchanges, and FINRA have since then agreed to and adopted single-stock circuit-breakers, which assuaged investor fears about the wholesale disappearance of liquidity over a short period of time. Though most observers believe that these single-stock circuit breakers have generally worked well, they are sometimes triggered by a single erroneous trade on one trading venue, at a time when the market in that stock was operating in an orderly fashion on all other venues.

The literature review only provides a survey of academic research findings on HFT and its role to the overall financial market health. Due to the limited data that academic society can access, the answers to questions regarding HFTs' economic merit and regulation surrounding HFT behaviors are far from being definitive. In the next section, we will use online surveys and interviews to poll a broader range of interest groups in an effort to bring more knowledge about HFT to light.

2 - HFT SURVEY

SURVEY DESIGN

The survey as designed has a total of 18 questions. The survey is anonymous but the surveyed individuals can declare their name and email as answers to a non-mandatory question. The survey questions are divided in four major categories. This is done for two reasons; firstly it allows the survey respondents, which are considered to be informed agents in HFT, to understand the purpose of the specific questions asked. Secondly, it allowed us when designing the survey questions to concentrate on what each question asked in an attempt to eliminate unnecessary questions which will only make the statistical analysis of the survey results harder to perform.

The four categories covered are:

I. Demographic information about the survey taker
II. Assessment of characteristic behavior of high frequency traders
III. Assessment of impact of High frequency trading to the market behavior
IV. Assessment of need for regulating HFT in the future

A copy of the survey and the results is included at the end of this document in the Appendix.
The survey has been given to the participants in the 5th Annual Modeling High Frequency Data in Finance (Oct 24-26, 2013) held at Stevens Institute of Technology. The survey is still available and gathering answers . The survey has additionally been distributed via email to over 200 specialists working with data sampled with high frequency.

The population of the survey was intended to reach three distinct groups. These groups are academics, financial industry people working in the area, and regulators. One serious drawback we have encountered in the distribution of the survey is that the industry and regulators do not want to take the survey even though it is completely anonymous. As a consequence there is less representation of industry opinion and even less of the regulatory opinion. To counterbalance this drawback, which we did not anticipate initially, we perform analysis of data ourselves and try to obtain objective answers supported by data to answer some of the survey questions. We intend to keep the survey accessible for the foreseeable future and collect opinions on the subject yearly. The survey results are shown in the appendix A.

INTERPRETATION OF SURVEY RESULTS

Rather than going through each survey question (which we do in the Appendix A), here we want to state and interpret the survey's results as of March 31, 2014. It is very clear from the answers received that there is a distinct duality in the answers we received. Most of the answers from academia on one side and from industry on the other side seem to converge only on a few questions. In the section about characterizing HFT the answers from academia overall seemed less informed than the answers from industry. In the section on the HFT impact on the market both categories agreed, and with about the same ratios, that HFT provides liquidity to the market. However, as expected the academia is much more reserved when asked questions such as "does HFT obscure price discovery" and "does HFT increase market volatility". In the section on regulating HFT there is again a dichotomy in the answers. Industry disagrees with the need of more regulations while the academia agrees with the need. However, when faced with the question which regulations should be imposed (Q16), academia selects random answers (the percentages for the 4 questions are close to 25%). Industry on the other hand does not want to limit the rate at which quote messages are sent instead the least disliked option was to limit the order cancelation rate. The most interesting question for us was the last one which asked about investing in HFT. The distribution of answers to this question is remarkably similar for both categories and the majority of answers (48.15%) selected: "I will invest in smarter algorithms for HFT because regulation is coming that will limit the frequency of the trades thus the need on relying on smarter rather than faster algorithms".

3 - HFT IMPACT

MECHANICAL IMPACT ON MARKET

The mechanical impact on market can be measured from samples of data wherein HF trades can be separated from non-HF trades. Once that is achieved, several quantitative measures can be developed. Normally, access to this data is not allowed to researchers due to the sensitive nature of the information. However, we have obtained a "benchmark" sample of HFT data provided by the NASDAQ to HFT researchers. Analyzing this data has produced a number of interesting results; however, we see that this white paper is not the place to go through them in details. Only a few comments on the results are included here.

Data

The NASDAQ dataset contains trading and quoting activities of 26 HFT firms in 120 stocks on the NASDAQ exchange. In our analysis, we mainly use trade reports, of which the sample period covers all of 2008, 2009 and one week in 2010. Specifically, trade reports contain a field with the following codes: HH, HN, NH, or NN. H refers to a HFT firm and N refers to a non-HFT firm. The first term in the pair classifies the liquidity seeking side, and the second term classifies the liquidity supplier. For example, HN indicates that an HFT firm took liquidity from a non-HFT firm. Obviously, HH is not very informative since both HFT firms are labeled as H in the sample.

Indices

The volume index is the number most mentioned in the literature related to HFT trading and it refers to the percentage of the total trades which is attributed to HFT. Table 1 shows the percentages for years 2008, 2009 and two weeks in 2010. Indeed, these ratios confirm the number most circulated in literature of 70% of the trades having an HFT counterparty.

Table 1. HF percentage volume in the sample

	Percent of trades where at least one counterparty is HFT
Year 2008	0.713452891
Year 2009	0.681901682
Year 2010	0.744922944

However, this number is deceiving. The number is calculated as (HH+NH+HN)/(HH+NH+HN+NN). Clearly the number is not an accurate measure of liquidity. Furthermore, when looking at the actual percentages it became apparent to us that the behavior of HFT is very different depending on the type of stock they are trading in (large average daily volume vs. low average daily volume).

Thus, we decided to introduce two easy to understand measures: the index of cross-liquidity (from an HFT unit H to a non-HFT unit N), INH, and the index of auto-liquidity, IHH. The first measure, the cross-liquidity index is calculated as and it calculates the percentage of volume exchanged between HFT and Non-HFT where HFT provided liquidity to Non-HFT market participants. The second measure, which we call auto-liquidity is calculated as , and represents the percentage of volume where HFT firms exchange shares between themselves from the total volume where the same category of traders exchange shares. The respective complimentary liquidity indices are IHN and INN, and can be easily calculated as one minus the primary indices. The numbers obtained are quite different for each stock but one interesting feature emerged. Please consult Figure 1. In this figure we first color the stocks based on the Average Daily Volume (ADV) of shares traded. We label blue chip stocks in blue and in decreasing order in orange and red. We then sort the stocks by the cross-liquidity index INH.

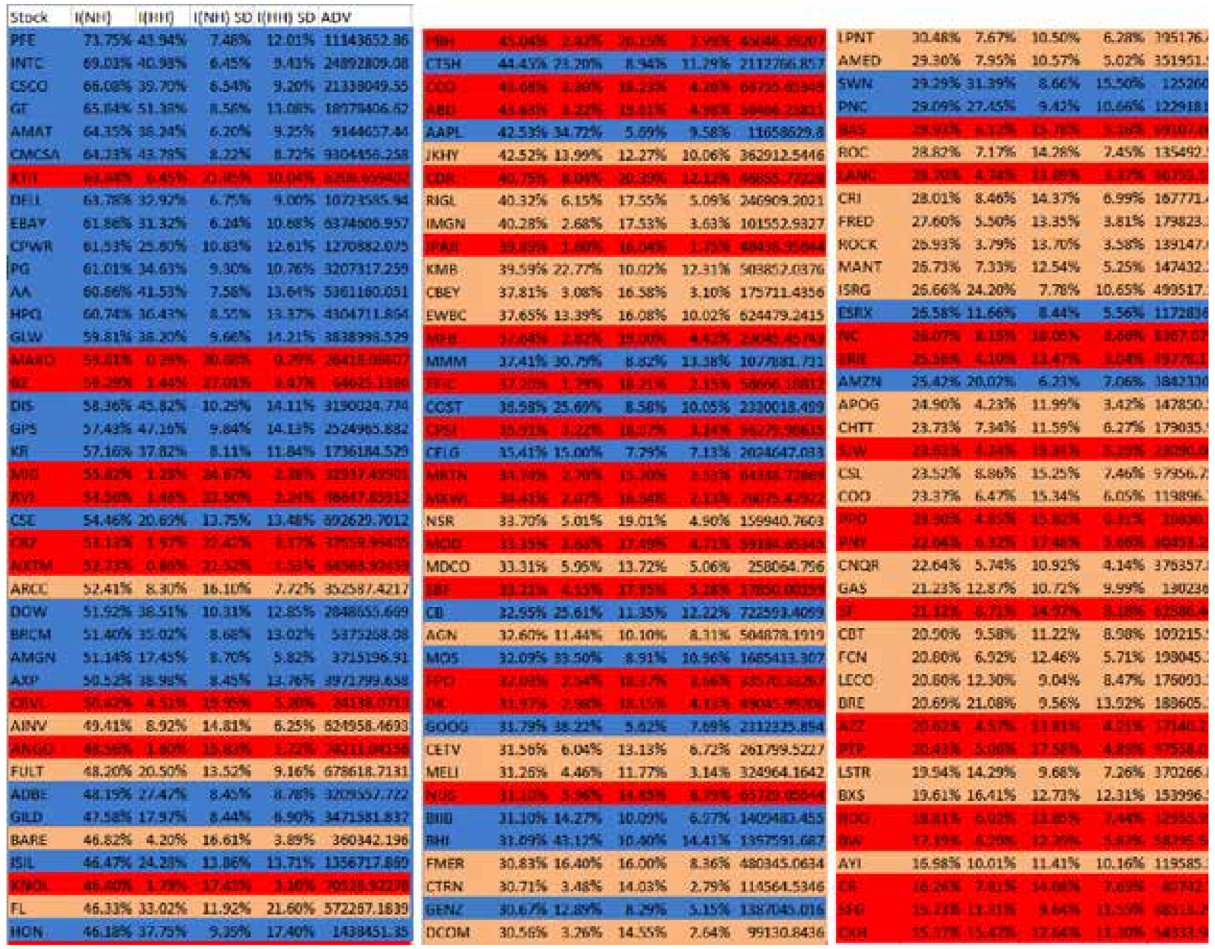

Figure 2. Stocks in the sample ordered by the cross-liquidity index (largest to smallest). Colors are denoting large ADV (blue), mid ADV (orange), low ADV (red). Picture provided for 2008.

We can clearly see from this picture that the HFT provides liquidity primarily in large-cap stocks while in mid and small stocks only in a small percentage of shares traded between HFT and Non-HFT they actually provided liquidity. In fact if we look carefully to the isolated red lines in the blue majority and the isolated blue in the red majority we will see that both indices are important to determine the behavior (first two column numbers are totally different than surrounding ones). When we look at the daily variability of these indices the picture is even more striking. Figure 2 presents the stocks ordered by the standard deviation of the daily cross-liquidity index.

Figure 3. Stocks in the sample ordered by the standard deviation of the cross-liquidity index from smallest to largest. Colors are denoting large ADV (blue), mid ADV (orange), low ADV (red). Picture provided for 2008.

We can see that the colors almost mimic the ADV categorization. This picture tells us that the liquidity providing behavior of HFT in stocks highly traded is much more consistent from day to day than it is in stocks which are not traded all this much. All of this points to different algorithmic behavior in stocks highly traded versus stocks which are traded infrequently. HFT tends to place limit orders and thus provide liquidity in large stocks while it plays a much more opportunistic role in small-cap stocks. These measures and others will be investigated in subsequent work. However, it is worth mentioning one important observation. Providing ONE number to characterize HFT behavior is misleading and impossible. This remark is even more obvious in the following image (Figure 4). In this image we present a histogram of daily average profit and loss (P&L) for 2008 for all the HFT's in the sample. Each observation is a particular stock from the sample of 120 stocks.

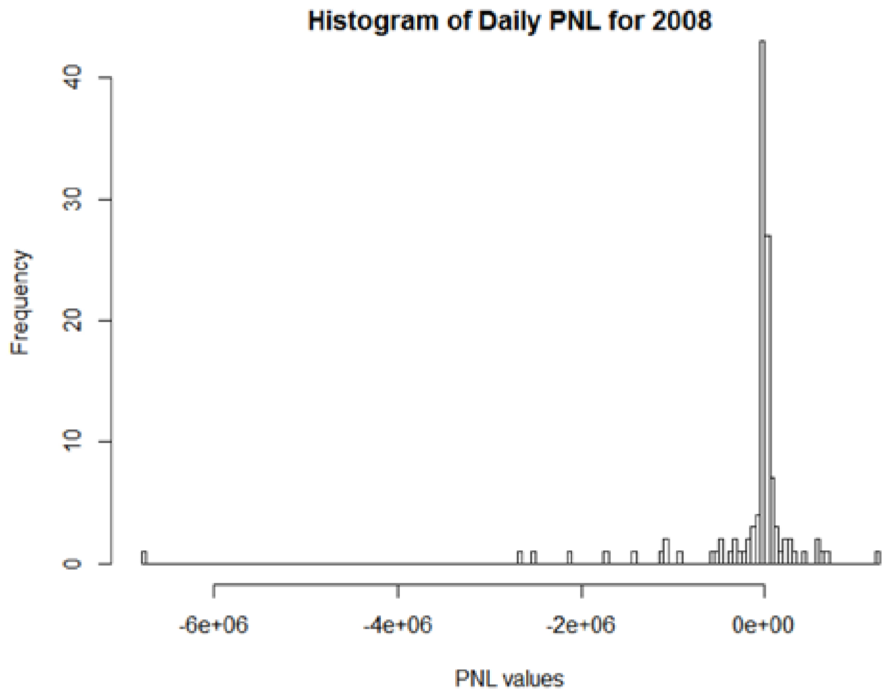

Figure 4. The histogram of the daily average P&L for ALL the HFT units for the year 2008. Overall, the HFT made money in some stocks and lost a lot in some others. Here the notation e+06 means one million dollars.

We can see that this is a histogram skewed to the left. Therefore the average profit per stock would be a really bad measure and one that would not scale to the entire universe of the market. As we learn in any statistical course, the mean of a sample is heavily influenced by outlying observations. A better measure is the median and the five number summary. However as mentioned above it is very hard to describe the HFT P&L with one number per stock per day (as most researchers try to do).

It is also important to note that the data does not contain information about transaction cost and "rebate". The rebate idea is structured differently in different exchanges but in principle it basically relies on the exchange collecting an amount $b1 per 1000 shares from liquidity takers and rewards $b0 per 1000 shares to liquidity providers thereby netting $b1-b0 per 1000 shares (usually in the order of $0.001/share.) This reward structure was reversed by the CBSX exchange after the Spread Networks completed its connection between Chicago and New Jersey .

Another dark sopt in analysis of high frequency finance is the issue of dark pools for which the reader is referred to the book by Scott Patterson . Furthermore, notwithstanding our gratitude to NASDAQ for providing the data on which this research is conducted, there are several criticisms regarding this type of studies conducted based on samples provided by exchanges to extract intelligence about HFT. The first observation is that we are, as most researchers, able to extract useful stylized characteristics from the sample about the stocks in the sample. However, to our knowledge, there has been no scientific sampling of the markets that justifies the extrapolation of results of the sample to the stock population it is supposed to represent. Therefore, most claims in this domain should be viewed as valid only for the sample at hand. The second observation is that the 26 HFT firms are aggregated as one entity labeled H in the sample. We understand the rationale behind this due to the liability that will be produced by labeling the individual firms; however, the results can only inform on aggregate positions, P&L, volume percentages and liquidity of all the HFT body. In other words, the results of such investigations are limited to the mechanical aspects of HFT and non-HFT interactions post-executions while in reality the issues raised for and against HFT can only be addressed with instantaneous observation of the process of price discovery as it forms with the depth of the book on record and the ability to fill an order as observed in real time not order execution ex-post. Thus, in this paper, we put more emphasis on financial information flow architecture to arrive at a better system.

IMPACT ON INSTITUTIONAL INVESTORS

There is an important point to note in discussing HFT impact on institutional investing. To explain it, we examine the concept of disproportionality of capital at risk of HFT versus institutional investments capita at risk. In this paper, capital at risk at time t refers to the total amount of capital that an entity or a collection of entities deploys in all of its market positions in all of its portfolios. It is well documented that an HFT unit does not deploy large capital at risk at any point in time because of the "round trip" executions in a very short time with small-volume orders. While it is also well known that that HFT accounts for about 65%-70% of volume in equities, its capital at risk makes a negligible percentage of total market capitalization. For example, let us suppose that there are about 400 HFT firms, which on any given day at any given time cannot deploy on average more than $10 million each in diverse markets. The $10 million per HFT unit is a postulated upper estimate of the average capital at risk at a single point in time. This puts the deployable capital at risk at a given time at a maximum of $4 billion deployed in various positions. The actual deployed capital at risk in a specified moment in time is a fraction of the HFT deployable capital at risk. On the other hand, it is estimated that institutional investments made up upwards of 64% of market ownership at the end of 2009. Let us assume for the sake of argument that the universe of markets in which both institutional investors and HFTs coexist has a capitalization at $10 trillion. Then the average relative equity of HFT to investments at any time t is equal to $4billion/$6.4trillion = 0.000625. The anomaly with this picture lies in the ability of a small percentage of minority ownership to have a greater influence on instantaneous price dynamics than the majority ownership while realizing the paralysis of the majority ownership to prevent sizable price dislocation under some scenarios. From the perspective of portfolios, if we combine the HFT entities together into one portfolio and combine the institutional investments into one portfolio, the smaller and transitory HFT portfolio fluctuations determine the institutional investments fluctuating values.

We find that the impact of HFT on institutional investors can be divided into two components: systematic impact and systemic impact. The systematic impact refers to the impact of HFT on institutional investors through the adjustment of market risk. Most HFT affects the price locally with respect to the expected fundamental value. Many HFT tactics are mean-reverting tactics of the statistical arbitrage type, which classify them as pure Alpha. Pure Alpha tactics play the idiosyncratic risk that is particular to the equity. The repeated applications of directional tactics and statistical arbitrage may lead to price dislocation causing disturbances in beta-based strategies in the case when a sufficient number of equities are affected, hence the HFT systematic impact on investment portfolios. The systematic impact affects all investment portfolios simultaneously including pension funds, insurance, savings, and foundations.

As for financial stability as understood by the charge of Financial Stability Oversight Council established by Title I of the Dodd Frank ACT, we see that the HFT systemic impact refers to the conditional probability that HFT may destabilize the markets through a phenomenon analogous to the Butterfly effect in highly connected and nonlinear systems. So far there is no definitive scientific assessment for such an event. The Flash Crash of May 6, 2010, even in the presence of partial evidence that HFT caused the exasperated decline in markets in a short time, cannot-by itself as a singular event- constitute an argument for HFT as being an imminent source of systemic risk. HFT becomes a source of systemic risk when there are repeated episodes of events similar the Flash Crash that threaten markets' stability at large and that can be shown to be at least caused by HFT in the sense of Granger causality. Such sequence of events would be a threat to financial stability. An assessment of the probability of such sequence of events taking place in the U.S. markets is needed.

4 - RECENT HFT DEVELOPMENTS

There are many voices that advocate slowing, curbing or abolishing the HFT practices by many methods. We believe that many of those proposals that fall into the category of banning HFT are not realistic or essentially violate free-market principles. Other proposals for creating friction or discretizing trading into frequent auctions are worthy of examination. For example, the University of Chicago economists and the University of Maryland (Budish, Cramton and Shim) or BCS proposed that stock exchanges process orders in batches as a solution to the HFT practices. BCS believe that converting the market design from a serial process to a batch process with an optimal tick time subinterval for auctions would solve the problem of racing to continuous finance. The frequent batch auctions are sealed-bid, uniform-price, double auction at discrete times. Orders during the submission stage are not displayed, which technically does not conflict with Regulation NMS.

On March 18, 2014, the New York State's attorney general, Eric Schneiderman, said that "the U.S. stock exchanges and alternative trading platforms provide high-frequency traders with unfair technological advantages that give them early access to key data" . The claim rests on 1) stock exchanges allowing colocation of servers within trading venues; 2) HFT units having extra bandwidth and high speed switches; 3) asymmetric information is obtained based on asymmetric technological capabilities. There were no comments by the exchanges on those claims. Schneiderman endorsed the frequent batch auction solution proposed by BCS.

The recent claims in Michael Lewis's "Flash Boys", released while writing this paper, that the market is rigged are addressed in the context of our proposed solution to market information transmission flow architecture. The story of RBC, Brad, Ronan, John and how Thor came about in Flash Boys is a remarkable one. The idea that a solution to the HFT lies in the formation of a new dark pool, the IEX , is quite interesting and warrants further examination. The IEX, formed essentially by the heroes of the Flash Boys, is a trading platform with a matching engine wherein latency advantages are neutralized. There are numerous articles on the need to regulate HFT without really saying what exactly to regulate in a system whose information packets and signals are moving simultaneously at the speed of light. In this paper we offer a framework for a better design of financial information transmission architecture.

5 - DISCUSSION OF HFT

Arguments for or against HFT are mixing four issues to the extent that no clear understanding of the subject could emerge. There are four distinct characteristics of HFT arguments

- Technology as an enabler
- Location and time-scale
- Fair practices using algorithmic trading strategies independently of time-scale
- Unfair advantages through asymmetric insider information and quote manipulation

We now argue those points. The first three bullets are part of any evolving complex socio-technical system. Technology edge, location and time-scale, and algorithmic trading strategies cannot and should not be the subject of this debate. The forces of technology are not stoppable and, in this context, for example, Michael Lewis mentions that technology can drive up volatility, which can be ture but not particular to HFT and asset prices in finance. New technologies enable new possibilities, which lead to new volatilities associated with valuation uncertainties of innovations. Those innovations include new financial products and methods of modeling. This is particularly true at the advent of a disruptive technology that result in new complexities . The phenomenon is not particular to the domain of finance but a characteristic of complex adaptive socio-technical systems. Technology risk is the subject where this type of assessment can be made.

The issue of regulating location and time-scale of private enterprises is also not useful and cannot purely stand on rational arguments in free markets. Under existing regulations partly shaping the financial ecosystem, all investment firms (small, large, or individuals) seek competitive advantages with respect to trades, investments, commissions, tax laws and the like and part of this competitive advantage is location. Imagine someone arguing that the colocation of large low-frequency investment firms gives those firms unfair advantages by being in New York City or London while small investors cannot afford to be in the proximity of vital information flows and high visibility spotlights. There is more than a physical address to colocation in as much as it provides insider informational proximity as a function of time-scale; however, as argued in the discussion section of this paper, this is a system's information flow design problem not an agent problem of real estate.

As for time-scale, it is a non-issue as well when it comes to trading practices. A form of risk-reward proportionality in an informationally equitable ecosystem is basic to free markets while the time-scale at which this exchange of risk and reward happens is not specifiable in free markets unless it becomes a source of instability. Discounting intent, a longer-term investor, from the perspective of trading, is a lower-frequency (LF) trader. By stretching the time scale, an investor shares the same objectives of taking a risk based on manual decisions, algorithmic analysis, technical analysis or fundamental analysis or any proprietary analysis as those who operate at higher frequencies. The LF trader opts to operate on a time-scale that is, say a billion times, slower than the HF trader and as such invests in an information cycle that is proportionate to the duration of deployment of capital based say on fundamental analysis. On the other extreme side, the HF trader opts to operate on a time-scale that is proportionate to the market microstructure by exchanging local-in-time risks and rewards without awareness of the longer information lifecycle of the asset. The LF trader trades the fundamental value based on fundamental corporate information and market information while the HF trader trades price noise generated by local corporate and market information fluctuation and superposed on the fundamental price. In other words, longer-term investors or LF traders buy and sell time-bulk risk while HF traders buy and sell time-retail quantas of risk. In between those two categories of LF traders and HF traders there is a spectrum of traders who operate based on a multitude of tactics, strategies or behavioral impulses. This white paper is not the place for a philosophical debate, however, it is hard to find a moral or legal basis for the distinction between similar objectives and actions to achieve returns based on space or time-scale arguments of such actions. There can be distinctions based on the intent of markets and why financial intermediations came about, which we do not go into in this paper.

We come to the fourth bullet of unfair practices, which is in fact the issue to be proved or disputed. We emphasize that fairness becomes an issue whether it is violated at high speed or at low speed and regardless of location. The set of practices in question that leads to unfair advantages are associated with HFT insider information, as one classification of violation of principles of fairness. The second set of claims against HFT falls under manipulation of prices via quote stuffing and other fancy localized price skewing mechanisms, which may act on insider information at the HFT time-scale. The rest of the factors like colocation and time-scale are natural adaptive alignments with presented opportunities in the presence of smart people.

Our view is that the arguments afforded by the HFT economic value to market liquidity cannot be used as a justification for violating principles of fairness in free markets--once those violations are proved scientifically not in a court of public opinion. It also matters none who is affected by such violations be it big investors or mom-and-pop folks. In this direction, the reader is referred to the experiment by Canada's stock market regulators limiting HFT activities in April 2012. In that experiment, the regulator increased HFT messaging friction, which led to a 30% drop in order submission and cancellation and 9% average increase in bid-ask spread on the Toronto Stock Exchange. The decreased HFT participation led to lower liquidity and higher transaction costs. Institutional investors performed better while small investors performed worse in the limited HFT activity mode .

It has been reported by some HFT firms that there was a one-day loss in more than one thousand days of HFT trading. Such a return pattern agrees more with a broker fee structure rather than a trading strategy return. The question becomes completely different and can be perhaps rephrased based on a functional argument. In other words, does an HFT unit want to be viewed as a trader or an electronic specialist (e-specialist) liquidity provider? The classification is important since the classification as trader implies that the return comes from applications of competitive algorithms in fair financial information order flows with no systematic information advantage while the classification as a liquidity provider implies that the business of the HFT unit is that of an e-specialist, which earns its returns based on fees collected for providing liquidity and making the market.

6 - SOLUTIONS: THE HFT ISSUES ARE INFORMATION TRANSMISSION ZONING PROBLEMS

The ideas we present in this section are new in their formulation. The HFT issues are not issues of financial mechanics but issues of financial information flow architecture that complies or does not comply with the intent of the Regulation NMS. First we mention the proposal of the Chicago Booth School of Business, the BCS paper , in which BCS proposed a model to solve the HFT related issues. The proposal is a good attempt and commendable effort. However, since the proposed solution is mechanical in nature, it may only transfer the problem from one place of the system to another. The description of the solution as "mechanical" refers to the idea of replacing current price dynamics with the frequent batch auctions (FBA) that are sealed-bid, uniform-price, double auction at discrete times. The idea is that, by discretizing the time step size, the race to higher speed will be rendered of no competitive value.

The concerns about the solution are summarized in some points. The first concern is that under the frequent batch auction (FBA) regime, it is not clear if HFT insider information impacts sealed bids inside the batch frame under the current information flow architecture. The second concern is that under the FBA, what happens at the peak of order flows resulting from news with sizable information content and with high-volume equities? Is there a model that can anticipate the batch performance in the optimal time tick size? The third concern is what happens to the options market associated with the underlying equities? Do they trade at the same synchronized clock for each batch frame? Does the option market have to become also a frequent batch option synchronized with the underlying asset? The fourth concern is what is the estimated cost of re-architecting the information system to perform FBA? Who pays for the IT and software to support the re-architecture? The fifth concern is that market adaptation will create a new market with a new exchange/product as follows: the price of a batch at time tick t becomes the basis for an option on the underlying for time ticks that follow. That option will HF-trade continuously and will impact the sealed bids even if they are not displayed inside the batch frame. The speculation transfers from what is streaming as exchange orders to spaculation on what sealed bids have already streamed inside the batch. We emphasize that there may be counter-arguments by the authors for each of these concerns.
We adopt a different philosophy in addressing the HFT issues. In the information age, the concept of insider information has to be reformulated in information metrics and the financial system architecture should be designed to support those metrics including requirements. The metrics should apply to high or low frequency as a function of time and space in as much as they affect information transmission.

It comes out that all claims against HFT practices can be understood with the introduction of the concept of information transmission distance zoning. In order to explain this point, we formalize a concept of information transmission distance between two points A and B as the average time it takes for information packets to travel between A and B and in most cases it coincides with the familiar idea of latency including throughput. The distance is defined in terms of "average time"-not space-to account for evolving speed of information transmission as a function of time. The information transmission distance accounts for the possibility of having two points that are farther apart in physical space distance to be closer in information transmission. It also allows for ordering of agent's access to actionable information that is not just a function of location but transmission capabilities conditional on location and technology. For example, under certain conditions, it is possible for an agent who is farther away in physical location than market participants to be closer in information transmission distance to an information source (an exchange) if the technology and the transmission protocols are superior. This particularly occurs at the emergence of a disruptive technology. In that case, systematic information asymmetry in favor of the agent can be achieved. In the HFT context, it is only material to insider information when the information transmission distance of the agent to the exclusive source becomes systematically smaller than the distance of the SIP subscribers to the same exclusive source. The SIP is where the National Best Bid and Offer or NBBO is calculated in compliance with Regulation NMS.

In order to understand aspects of the information transmission distance in terms of time, we need to briefly know the differences between protocols for Internet information traffic, namely TCP and UDP. The first stands for Transmission Control Protocol and the second for User Datagram Protocol. TCP is used when quality and reliability are more important than transmission time while UDP is suitable for fast applications (games for example) as it does not perform error-checking for streaming packets and there is no packet-handshaking and no acknowledgment. The header size for TCP is 20 bytes and 8 bytes for UDP. In that context, the reader is encouraged to see the simplified animation of Nanex . The securities information processor (SIP) uses TCP while most HFT units use UDP. The information transmission distance between the exchange and the SIP is greater than that of the HFT units to the same exchange. The differences in protocols and sizes of the "information transmission pipe", and even operating systems position the HFT units at an information transmission distance from the exclusive source, the exchange, that is smaller than all SIP subscribers. This means that under the current infromation flow architecture, HFT units participating with the UDP information transmission "super-highway" have a systematic information advantage with respect to all participating members of the securities information processor, SIP.

In Regulation NMS, the "Adopted Rule 603(a) establishes uniform standards for distribution of both quotations and trades. The standards require an exclusive processor, or a broker or dealer with respect to information for which it is the exclusive source, that distributes quotation and transaction information in an NMS stock to a securities information processor ("SIP") to do so on terms that are fair and reasonable. In addition, those SROs, brokers, or dealers that distribute such information to a SIP, broker, dealer, or other persons are required to do so on terms that are not unreasonably discriminatory."

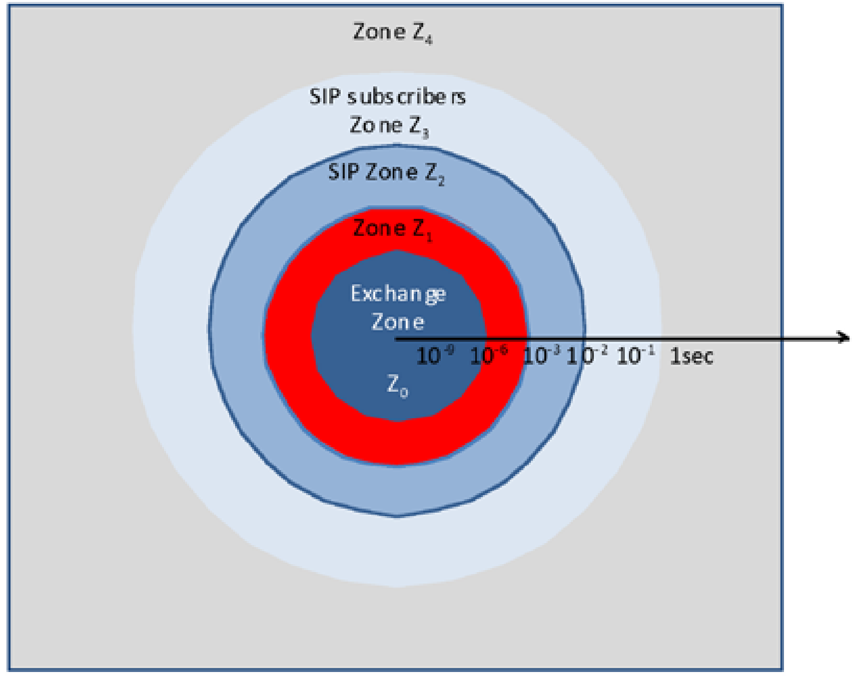

Figure 5. Information transmission distance stratification centered at an exclusive source.

We find the interpretation of the regulation in expressing the requirement as "not unreasonably discriminatory" to be not unreasonably opaque. Therefore, we propose the idea of information transmission zoning, which is not dependent on subjective interpretation. In Figure 5, each circle radius determines the information transmission distance from the center. The center of the concentric circles represents the information source that is understood in the sense of the Regulation NMS as the "exclusive source", which is the exchange in the HFT case. The exchange itself occupies zone Z0 indicating near-zero latency zone. The next latency zone is Z1, termed "the red zone". The red zone is the closest in information transmission distance than any other zone including the SIP, designated zone Z2. The SIP subscribers occupy zone Z3 and the rest of the slower information transmission agents, depending on transmission layers, occupy zone Z4. Therefore, in terms of information transmission distance, for agents in those zones, we simply have |z0 |< |z1|< |z2|< |z3| < |z4|, with zone Zi defined as the region Zi = {z: |zi |< |z| <|zi+1|}, i=0,1,2,3,4, where |z|= information transmission distance and the zi is the ith interval cutoff, which is a function of location and technology at a given time. Currently those zones can be thought of as identified with time intervals Z0 [10-12,10-9], Z1 [10-9, 10-6], Z2 [10-6,10-3], Z3 [10-3,10-1], Z4 [10-1,∞). We now propose a model for computing the information transmission distance and the associted zones. We also provide an insider information transmission criterion for a specified exclusive source.

MODELING INFORMATION TRANSMISSION DISTANCE

We give a method for calculating the information transmission zoning cutoffs. Some terminology is needed in order to express the ideas. First the concept of *intrinsic latency* is introduced. The theoretical latency that is often used is what we will call the *absolute intrinsic latency*, which assumes that information packets travel at the speed of light without restrictions on capacity. In that case latency is only a function of physical location. In reality there is a difference between absolute intrinsic latency and real network latency as a function of transmission in a medium other than vacuum. Suppose that the physical distance between the point A and the information source (exchange) is $x > 0$ miles. The *intrinsic latency* is

$$L_0(x) = \frac{x}{\kappa c}$$

where c is the light speed in vacuum given at 186,282.40 mi/second. In reality the speed of information using light is scaled down from ideal by a coefficient κ that measures the medium transmission efficiency. The coefficient κ is a constant between 0 and 1. For example, in normal fiber optics traveling in silica glass, the coefficient of transmission of the medium is κ=69%. New reports of hollow fiber raise the coefficient κ to 97% but for only short distances so far.

COMPUTING INFORMATION TRANSMISSION DISTANCE

In addition to the intrinsic latency, there is *technology latency*, which accounts for servers, protocols and bandwidth. We find the decoupling between intrinsic latency and technology latency to be a useful concept. The information transmission distance between the point A in a network and the information source E denoted by $D_E(A)$ can be decomposed into intrinsic latency and technological latency and written as

$$D_E(A) = L_0(x) + T(A),$$

where the intrinsic latency $L_0(x)$ accounts for the universal "physics time-tax on information transmission" represented at least by the limit of the speed of light in vacuum as an upper bound for information travel—courtesy of Einstein. The excess transmission distance $T(A)$ is a function of network technology connecting point A and source E and server transmission-receiving technology to achieve throughput. The technology latency is usually expressed in terms of protocols, block size, and connection speed. See for example Mathew Mathis et. al. (1997) formula for such calculations .

For another point B at distance y from the information source in the network and using the same intrinsic latency,

$$D_E(B) = L_0(y) + T(B).$$

Then the information transmission distance between those two points A and B is defined as

$$d_E(A,B) = |D_E(A) - D_E(B)| = |L_0(x) - L_0(y) + T(A) - T(B)| = \left|\frac{x-y}{\kappa c} + T(A) - T(B)\right|$$

where $|.|$ stands for the absolute value function. For example, if the technologies connecting A and B to the information source are identical, then $T(A) = T(B)$ and the information transmission distance is purely a function of physical distance and it is exactly equal to the intrinsic latency $\frac{x-y}{\kappa c}$. On the other hand, if two points have the same physical distance from the information source and the same fiber optics, then the information transmission distance is purely a function of technology connecting A and B to the information source, i.e., it is equal to the technology latency $|T(A) - T(B)|$. In that case if A has a superior technology, A is closer to the information source than B in information transmission distance and vice versa. To explain this point, we consider some cases subsequently.

ASYMMETRIC INTRINSIC LATENCY WITH SYMMETRIC TECHNOLOGY

Suppose that the SIP is located 50 miles away from the exchange and the HFT units are colocated at 1 mile away from the exchange and both are using the same x, the exact technological capabilities, the same Internet Protocols, the same packet size, and the same types of servers, then the advantage is purely physical location distance. The identical protocols, say TCP, require an information handshake and validation, which doubles the distance for both the SIP and the HFT. Suppose further that the speed of light is as it is in vacuum ($\kappa = 1$) given at 186,282.40 mi/second. Then the information transmission distance of the SIP from the exchange is 100/186,282.40= 0.000536819 second while the information transmission distance of the SIP from the exchange is 2/186,282.40= 0.0000107364 second. The advantage that HFT would have on a generic SIP message is the difference between the two distances, which is 0.00052 second or half a millisecond. Changing the TCP to UDP for HFT affects the latency difference approximately 1 microsecond but it allows for maximum flow, which is necessary for HFT to accommodate streaming orders and cancellations. In terms of zones, the HFTs in this example have |z1|=0.0000107364 while the SIP has |z2|=0.000536819 in their ideal cases. In this example since |z1|<|z2|, the HFT unit is in the red zone.

ASYMMETRIC TECHNOLOGY VS. ASYMMETRIC INTRINSIC LATENCY

For simplicity we assume that the network is using the same transmission coefficient κ. Suppose that A is closer to the information source than B in physical distance, i.e., $y - x > 0$. Then A has an inrinsic latency advantage. Furthermore, suppose that B possesses superior technology connecting it to the information source than A, which means that $T(A) - T(B) > 0$. The intrinsic latency based on physical distance difference is given by $\frac{y-x}{\kappa c} = \delta > 0$ and the technology latency is given by $T(A) - T(B) = \tau > 0$. If $\tau > \delta$, then B is closer than A in information transmission distance to the information source despite the assumption that B is farther away than A in physical distance to the same information source. In other words, theoretically one can make up for intrinsic latency deficiency by having sufficient technology advantage. This is why, in principle, we reject arguments of asymmetric information based solely on physical distance represented by the issue of colocation.
On the other hand, if $\tau > \delta$, then B cannot make up for the intrinsic latency advantage. More importantly, we point out that the colocation in the case of HFT is associated with up-to-date superior technology with respect to the SIP so that both the intrinsic latency is an advantage and the technology latency is also an advantage. In that case, If A represents HFT units colocated at physical distance x_{HFT} from the exclusive source E and B represents the SIP at a distance y_{SIP}, then $y_{SIP} > x_{HFT}$ means that $L_0(HFT) < L_0(SIP)$ and $T(HFT) < T(SIP)$. The information transmission distance between the colocated HFT units and the SIP is given in total by

$$d^*(HFT,SIP) = \frac{y_{SIP} - x_{HFT}}{\kappa c} + T(SIP) - T(HFT)$$

where the unit of measurement is in seconds. This information transmission distance provides a systematic advantage, which is the high frequency insider information. The HFT units have sufficient time advantage to react on this information and to convert it dynamically, inside the micro-time frame of the market order flows, into cash flows for the HFT units and the exchange.

INSIDER INFORMATION TRANSMISSION CRITERION

What makes the actionable information obtained below the SIP information transmission distance classify as insider information is the fact that the SIP provides the NBBO by Regulation NMS. If an HFT unit places itself between the exclusive source and the SIP in the sense information transmission distance, it gains systematic insider information. We now state the insider information transmission criterion for HFT as:

Given an exclusive source E, SIP, and HFT unit, then the HFT unit has insider information transmission access if and only if . The possibility of converting positively or negatively on the insider information transmission is irrelevant to the criterion or the designation. It is also irrelevant to the question of insider information whether the HFT unit provides liquidity or takes liquidity.

In the case where there is insider information transmission, we say that the HFT unit resides in the information transmission red zone (Z1) as in Figure 5. On the other hand, if $d^*(HFT,SIP) < 0$, then the HFT or the algorithmic unit resides in zones, Z2, Z3 , or Z4 with no insider information transmission access. Furthermore, we define systemic latency of the exclusive source (the exchange) to mean the average information transmission distance from the exclusive source to the SIP, which is given by the formula

$$D_E(SIP) = L_0(x_{SIP}) + T(SIP),$$

with x_{SIP} denoting the physical distance between the SIP and the information source E and $T(SIP)$ the technology latency of the SIP to the same information source E.

Systemic latency associated with an exclusive source E defines the minimum information transmission distance separating all market participants from the exclusive source. Any access given to agents, including HFT units, not classified as an e-specialist that is part of the exclusive source, below the systemic latency constitutes a systematic arbitrage that results from HF insider information residing in the red zone. The systemic latency as a lower bound on minimum information transmission induces a natural upper bound on the frequency of HF trading beyond which there is no asymmetric utility of information with respect to other market participants, i.e., all SIP subscribers have a reasonable and fair access including HF traders, which complies with the intent of the Regulation NMS. As the technology of the SIP is upgraded, the systemic latency $D_E(SIP)$ becomes smaller and the HFT "natural frequency" of the system is increased. All associated activities such as fronting trades, skewing the microstructure price discovery by order posting and cancellation or quote stuffing and other practices would be of random competitive advantage when all electronic trading units are operating at or above the systemic latency.

In Figure 5, the cutoffs between zones are shown here for clarification purposes only and they can only shrink orders of magnitude with the advancement of processing and transmission; however, the ordering principle remains the same as long as we use an information transmission distance. The SEC definition of a specialist says that it is a member of the stock exchange, whose role is to facilitate trading in certain stocks and to maintain a "fair and orderly market" in the stocks they trade. The rules of the exchange prohibit specialists from trading ahead of investors who have placed orders to buy or sell a security at the same price . In 1935, there was a study by The Twentieth Century Fund that concluded that "specialists, as well as other exchange members, should be permitted to function either as traders or as brokers, but not both." In a somewhat similar manner, an HFT unit that functions primarily as a liquidity provider is closer in classification to being an e-specialist broker not a trader. On the other hand, an HFT unit that makes its returns from frequent trades based on directional price movements and pure algorithmic mechanisms should classify as an HF trader and should not care whether it is providing or taking liquidity. The two functions should be decoupled at the high-frequency scale so that the privileges of a broker are not shared within the same HFT unit functioning as a trader. The HFT broker can only be allowed to be in the red zone if it becomes an e-specialist as part of the exchange. The rest of the HFT units should be permitted to compete outside the red zone with all technological and algorithmic advantages.

7 - CONCLUSIONS

In light of existing regulation and in terms of zoning, the red zone in Figure 5 can be occupied by HF e-specialists not an ordinary HF trader while HF traders, without the designation of e-specialists, should be moved from the red zone to at least zone Z2 in compliance with the insider information transmission criterion for HFT as stated in this paper. Furthermore, an HFT unit that is not in the red zone should not be concerned or involved in any of those issues since it does not satisfy the insider information transmission criterion. HF traders residing outside the red zone should be able to trade with superior technology, transmission networks, algorithms, and computational capabilities that minimize the latency in decision support systems inside and outside the firm as long as their information transmission distance is greater than the systemic latency.

The philosophy for HFT reform should not be aiming at stopping the natural adaptation of the financial system to emerging technology. Purely mechanical solutions will create financial plumbing problems in other parts of the system or will generally transfer the problem to another point in the information chain. A successful proposal allows for innovation complexity to appear and and enables the system to contain it and benefit from it. The philosophy should be to build an adaptive transparent financial information flow architecture that complies with regulation, achieves markets objectives, and maintains credibility among stakeholders.

ACKNOWLEDGMENT

The authors acknowledge the NASDAQ OMX Group and especially Frank Hatheway, Chief Economist of NASDAQ for graciously providing us with a sample of 120 stocks with various levels of market capitalization listed on NYSE and NASDAQ where HFT traders were globally labeled under designation H.

Several graduate students in the Financial Engineering Division at Stevens Institute of Technology helped us with literature review, data, and computation. The authors want to acknowledge Ilya Bezdetko, Anthony DeFilippis, Mitchel Epperly, Alexandra Middleton, and Alex Moreno for their contributions to the literature review and we thank Chenjie Shao for computations involving the analysis of the NASDAQ data. The authors claim all the errors.

REFERENCES

Baron, Matthew, Jonathan Brogaard, and Andrei Kirilenko (2012), "The trading profits of highfrequency traders," November 2012 working paper, University of Washington.

Boguslaw Blawat (2012), "Optimal Order Execution Problem in the Framework Model of High Frequency Trading" SSRN Working paper

Benos, Evangelos and Satchit Sagade (2012), "High-frequency trading behaviour and its impact on market quality: evidence from the UK equity market," Bank of England Working Paper No. 469, December 2012.

Boehmer, Ekkehart, Kingsley Fong, and Julie Wu (2012), "International evidence on algorithmic trading," working paper, EDHEC.

Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang (2012), "Shackling short sellers: the 2008 shorting ban," working paper, EDHEC.

Biais, Bruno, Thierry Foucault, and Sophie Moinas, "Equilibrium high-frequency trading," October 14, 2011 working paper, available at http://ssrn.com/abstract=2024360, retrieved May 31, 2012.

Bozdog D, I. Florescu, K. Khashanah and J. Wang, "Rare Events Analysis of High-Frequency Equity Data", Wilmott Journal, Volume 2011, Issue 54, Pages 74-81, July 2011

Brogaard, Jonathan (2011a), "The activity of high frequency traders," December 2011 working paper, available at http://ssrn.com/abstract=1938769, retrieved May 29, 2012.

Brogaard, Jonathan (2011b), "High frequency trading and market quality," December 2011 working paper, available at http://ssrn.com/abstract=1970072, retrieved May 29, 2012.

Brogaard, Jonathan (2012), "High frequency trading and volatility," January 2, 2012 working paper, available at http://ssrn.com/abstract=1641387, retrieved May 29, 2012.

Domowitz, Ian (2010), "Take heed the lessons from the 1962 flash crash", June 21, 2010, available at http://www.advancedtrading.com/exchanges/225700888, retrieved May 29, 2012.

Easley, David, Terrence Hendershott, and Tarun Ramadorai (2009), "Levelling the trading field," November 19, 2009 working paper, UC Berkeley.

Easley, David, Marcos M. Lopez de Prado, and Maureen O'Hara (2011), "The microstructure of the 'flash crash': flow toxicity, liquidity crashes and the probability of informed trading," Journal of Portfolio Management 37(2):118-128.

Easley, David, Marcos M. Lopez de Prado, and Maureen O'Hara (2012), "Flow toxicity and liquidity in a high frequency world," Review of Financial Studies 25(5):1457-1493.

Egginton, Jared F., Bonnie F. Van Ness, and Robert A. Van Ness, "Quote stuffing," March 15, 2012 working paper, available at http://ssrn.com/abstract=1958281, retrieved May 30, 2012.

Foucault, Thierry, Johan Hombert, and Ioanid Rosu (2012), "News trading and speed," working paper, HEC.

Gai, Jiading, Chen Yao, and Mao Ye (2012), "The externalities of high-frequency trading," November 16, 2012 working paper, University of Illinois.

Glosten, Lawrence R. and Paul R. Milgrom (1985), "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," Journal of Financial Economics 14:71-100.

Hasbrouck, Joel and Gideon Saar (2012), "Low latency trading," working paper, New York University.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld (2011), "Does algorithmic trading improve liquidity?" Journal of Finance 66(1):1-33.

Hendershott, Terrence and Ryan Riordan (2011), "High frequency trading and price discovery," working paper, UC Berkeley.

Hendershott, Terrence and Ryan Riordan (2012), "Algorithmic trading and the market for liquidity," forthcoming, Journal of Financial and Quantitative Analysis.

Jovanovic, Boyan and Albert J. Menkveld (2011), "Middlemen in limit-order markets," October 24, 2011 working paper, available at http://ssrn.com/abstract=1624329, retrieved May 28, 2012.

Jones, Charles M. and Paul J. Seguin (1997), "Transaction costs and price volatility: evidence from commission deregulation," American Economic Review 87(4):728-737.

Kirilenko, Andrei A., Kyle, Albert S., Samadi, Mehrdad and Tuzun, Tugkan (2011), "The flash crash: the impact of high frequency trading on an electronic market," May 26, 2011 working paper, available at http://ssrn.com/abstract=1686004, retrieved May 29, 2012.

Lawrence E. Harris and Ethan Namvar (2011), "The Economics of Flash Orders and Trading", SSRN Working paper.

Madhavan, Ananth (2012), "Exchange-traded funds, market structure and the flash crash," January 13, 2012 working paper, available at http://ssrn.com/abstract=1932925, retrieved May 29, 2012.

Martinez, Victor Hugo and Ioanid Rosu (2011), "High frequency traders, news and volatility," December 29, 2011 working paper, available at http://ssrn.com/abstract=1859265, retrieved May 31, 2012.

Menkveld, Albert J. (2012), "High frequency trading and the new-market makers," February 6, 2012 working paper, available at http://ssrn.com/abstract=1722924, retrieved May 28, 2012.

Moallemi, Ciamac C. and Mehmet Saglam (2011), "The cost of latency", May 27, 2011 working paper, available at http://ssrn.com/abstract=1571935, retrieved May 30, 2012.

Pagnotta, Emiliano and Thomas Philippon (2012), "Competing on speed," April 27, 2012 working paper, available at http://ssrn.com/abstract=1967156, retrieved May 31, 2012.

Riordan, Ryan and Andreas Storkenmaier (2012), "Latency, liquidity and price discovery," forthcoming, Journal of Financial Markets.

Dorian M. Noel (2011), "The Application of SAS Hash Object to Ultra-high Frequency Financial Data: A Case Study in Limit Order Book Reconstruction" NESUG 2011 Conference Proceedings

United States Commodities and Futures Trading Commission and Securities and Exchange Commission (2010), "Findings regarding the market events of May 6, 2010," Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues, September 30, 2010.

APPENDIX A:

SURVEY STATISTICS

The results are based on a sample of 40 respondents as of Jan 22, 2014. In this appendix we detail the questions and the results obtained when analyzing these responses. The survey is entirely anonymous.

Survey Part 1: Demographics
The first part of the survey tries to categorize the type of people participating in the survey and by extent the population under study.

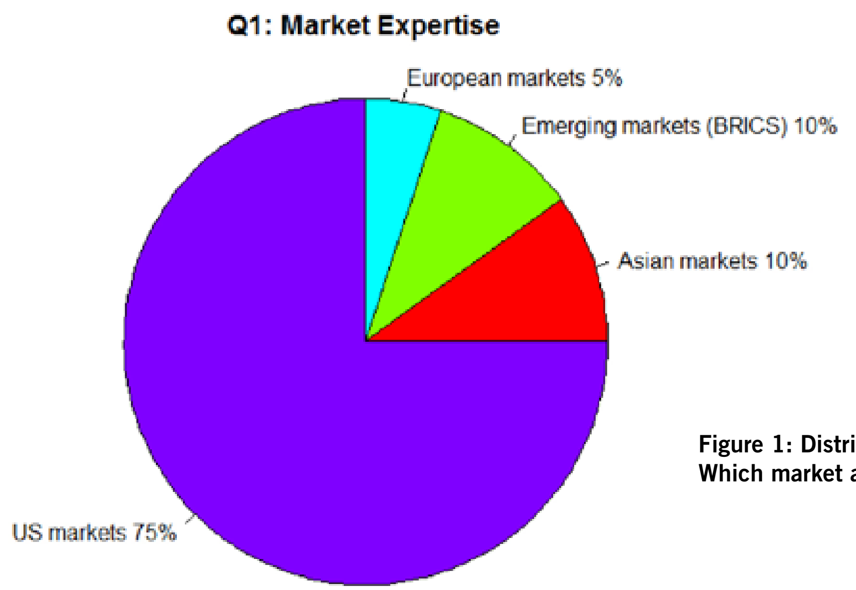

Figure 1: Distribution of answers for Question 1: Which market are you more familiar with?

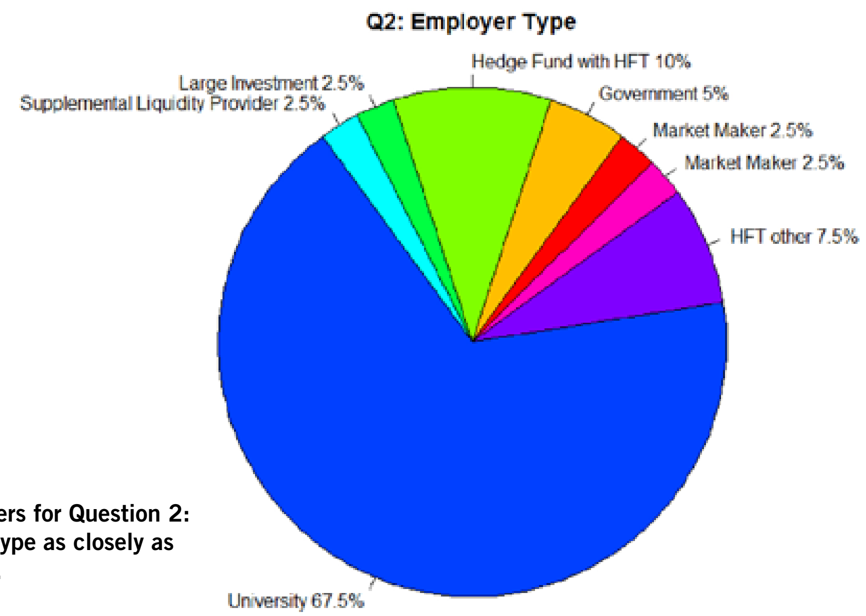

Figure 2: Distribution of answers for Question 2: Please choose your employer type as closely as possible (circle all that apply).

As we may see from the answers of the two questions the expertise is largely US markets and the majority of the people surveyed are academics. This is easy to understand since they are most likely to answer to this categories but it is encouraging that the population is very diverse and with expertise outside US.

Based on this demographics and the distribution of the population we will see that the results are much more informative if we separate the results obtained when surveying academia and industry& government as two separate entities.

The next two questions are specifically directed to industry respondents and thus we only present those results.
In Figure 3 (right) we show the distribution of the number of employees working specifically on HFT algorithms. We believe that the extreme observation (500 employees) is not an outlier since it actually corresponds to a supplementary liquidity provider and that category of companies typically employs multiple accounts and trading algorithms running simultaneously.

In question 4 respondents could make multiple selections. Furthermore, this was a required question and a large segment of answers from academia selected: "Others". This is why we only present the results from industry respondents. Furthermore, by looking at the number of choices each responder makes we can create a histogram of the number of markets they are active in. The results are presented in Figure 4.

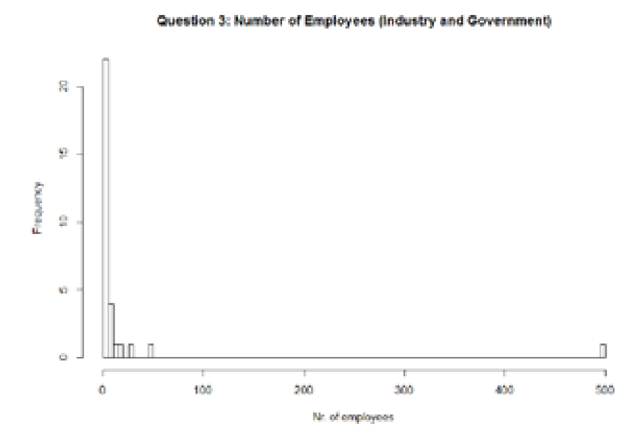

Figure 3: Histogram for Question 3: The number of employees involved in HF trading operations in my company is:

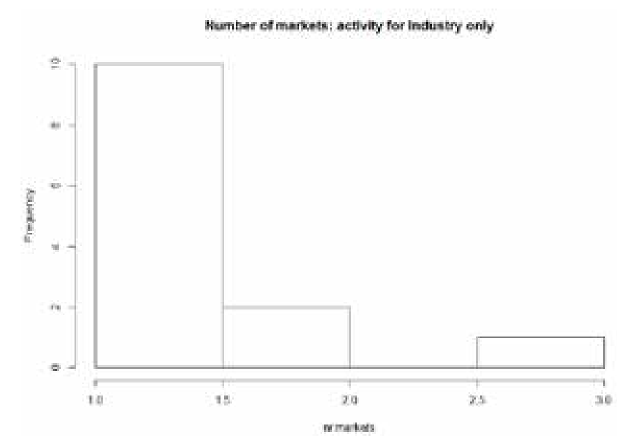

Figure 4: Answers for Q4: If you are a HFT firm what type of markets are you actively participating in?

SURVEY PART 2: CHARACTERISTICS OF THE HIGH FREQUENCY TRADING ACTIVITY

In this part of the survey we wanted to gather the informed opinions of respondents on what makes a trader a high frequency trader. The results obtained are presented in the next figures. We will separate the results obtained by industry participants and government participants. By doing so we believe we can illustrate that the perception of HFT is different in each of the two categories. Furthermore, we believe that for this section the industry answers may be perhaps more informative than the academicians' answers.

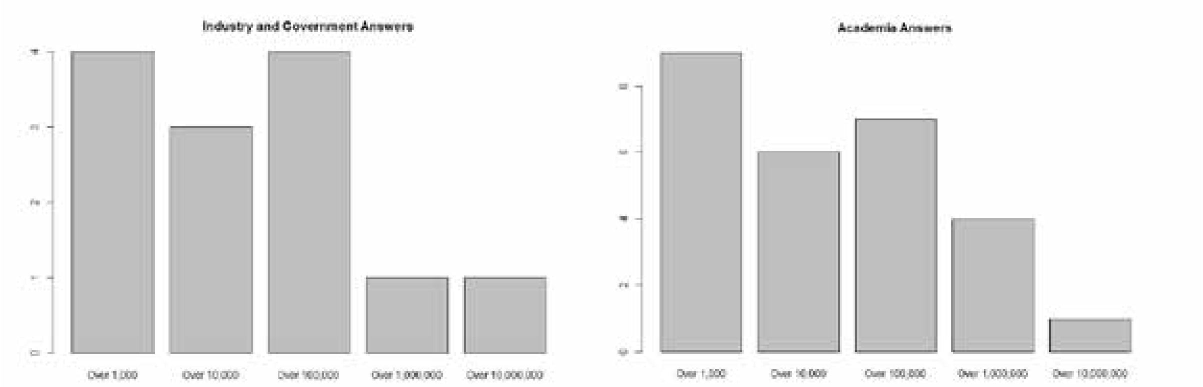

Figure 5: Answers for Q5: In you assessment, a trading entity becomes a high frequency trading entity if the number of total orders placed per day over any time interval during the day (executed, canceled, and still open) is:

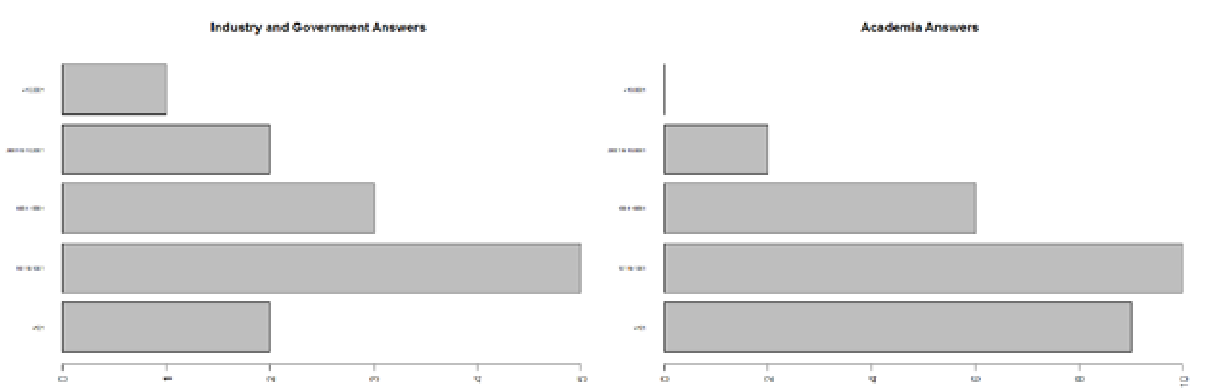

Figure 6: Answers for Q6: "In your assessment a HF trader has an average ratio of canceled orders to trades:" Possible choices from bottom to top: Less than 10 to 1, Between 10:1 and 100:1, Between 100:1 and 1000:1, Between 1000:1 and 10000:1, More than 10000.

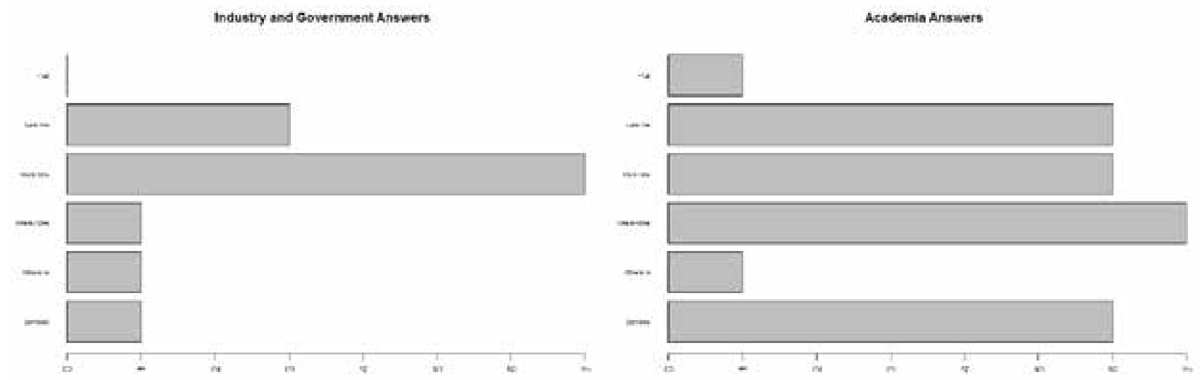

Figure 7: Answers for Q7 "In your assessment, the latency for a HF trader is of the order:" Possible choices from top to bottom: Less than 1μs, Between 1μs and 1ms, Between 1ms and 10ms, Between 10ms and 100ms, Between 100ms and 1s, I don't know.

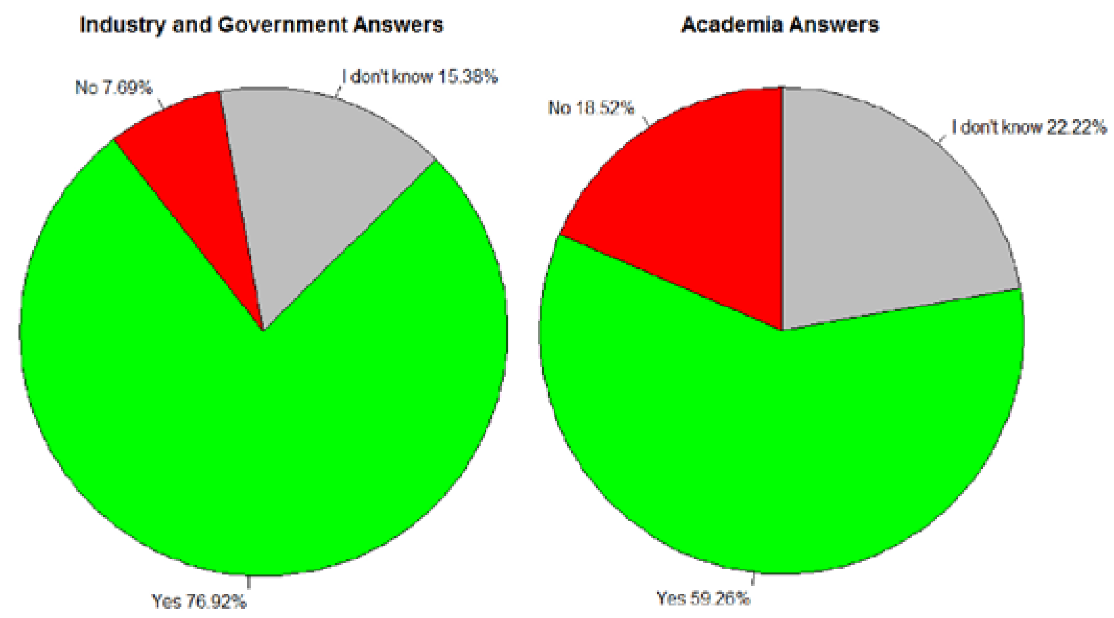

Figure 8: Answers for Q8 "In your assessment, a HF trader requires colocation to be competitive".

As we can see from the plots presented in some cases there are divergences when assessing the characteristics of the HFT. We can see that the answers from industry are more precise. We can observe this fact by noticing that the answers from industry are further apart from a uniform (random answers) distribution.

Interpreting the results provides us with the following insights about the perception of HFT. In question 5 about the total number of orders there is no clear defining factor. The randomness of the answers seems to indicate that as long as 1000 orders are placed every day you are qualified as doing HFT. In this respect it seems to us that the concept of HFT is being mixed with the concept of algorithmic trader.

In question 6 about the ratio of order canceled versus orders executed, the industry seem to think the ratio is somewhere between 10 to 1 and 100 to 1. In question 7 about the latency of messages, again the industry gives a clear answer: between 1ms and 10ms. The answers obtained for academia are completely random. Finally, a high frequency trader requires collocation which is evidenced more clearly in the industry answers. Collocation is the HFT practice of placing the trading algorithms on a machine which is hosted in a datacenter with high messaging speed performance between the machine and the trading exchange.

SURVEY PART 3: THE IMPACT OF HFT ON THE MARKET BEHAVIOR

In the next section of the survey we assess the perception about the impact the HFT has on the market behavior. Recall, that HFT was only possible since 2005 and clearly the market behavior in 2014 is much different than what it was at the turn of the century. All of the questions have a Yes/No/Don't know format and we again separate the answers into industry&government and academia.

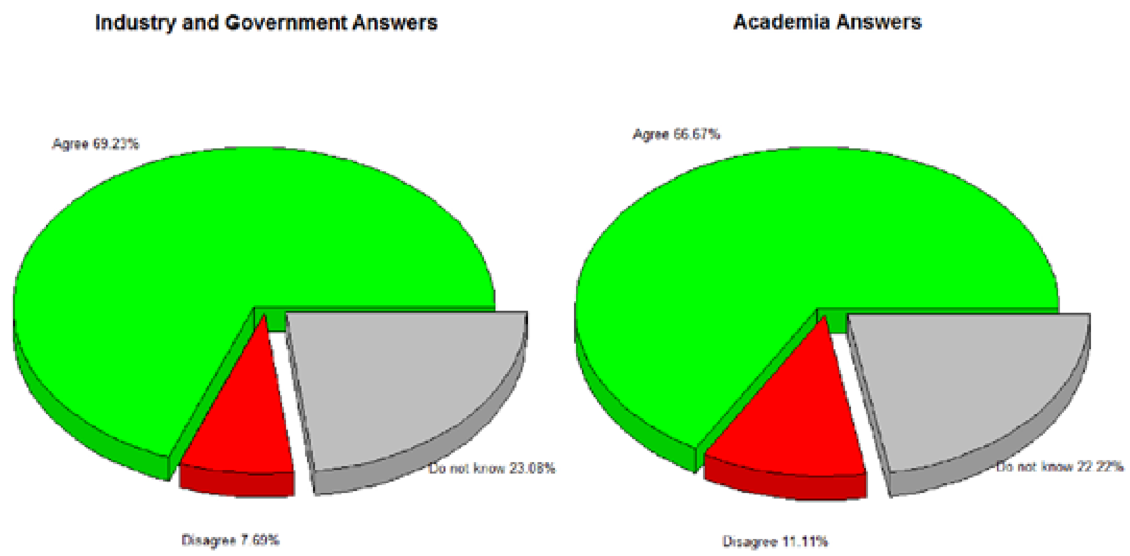

Figure 9: Answers for Q9 "In your assessment, does HFT increase liquidity in the markets on average?"

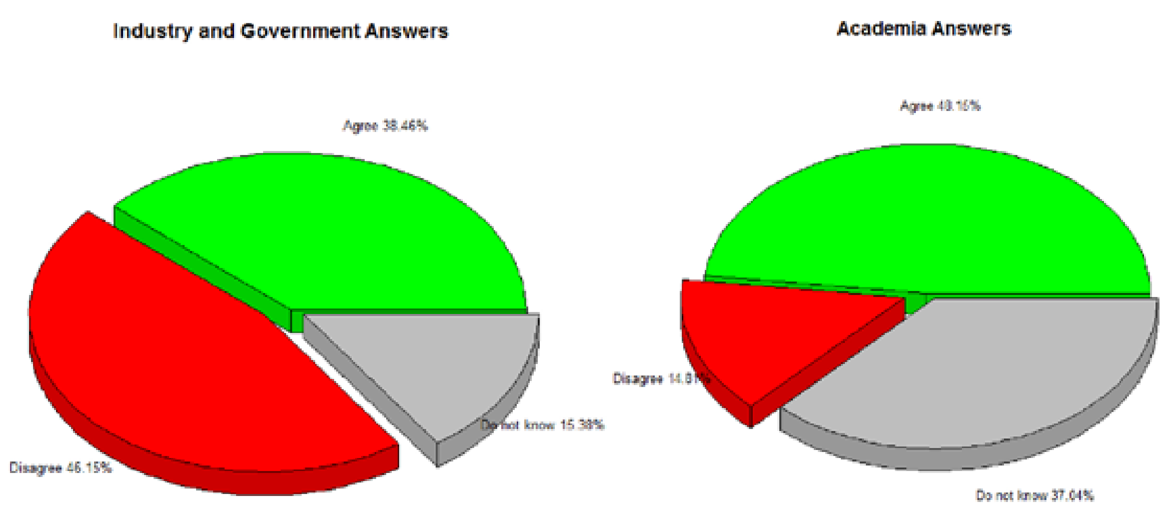

Figure 10: Answers for Q10 "In your assessment, does HFT obscure price discovery in markets?"

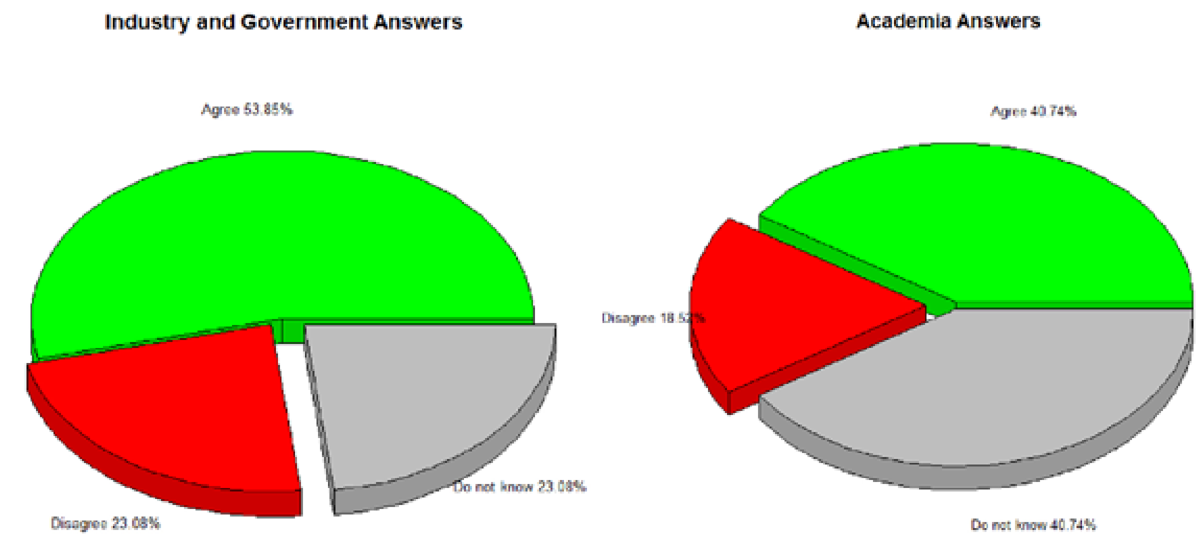

Figure 11: Answers for Q11 "In your assessment, does HFT increase the market volatility?"

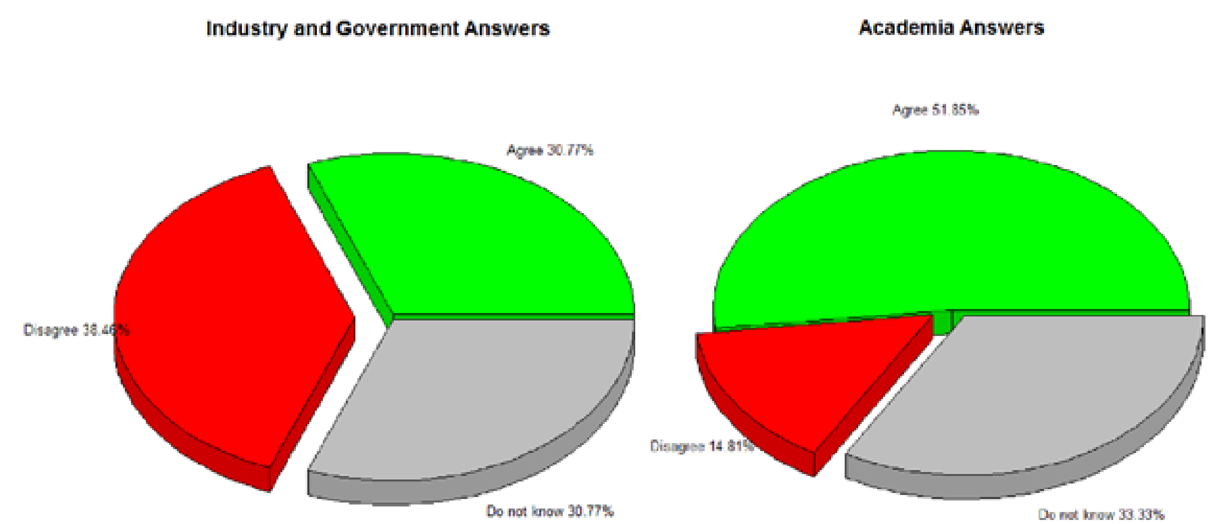

Figure 12: Answers for Q12 "In your assessment, does HFT increase the frequency of market crashes?"

Analyzing the results to the questions related to the impact of HFT we see a difference of assessment from the two sides industry and academia. However, everybody agrees that HFT increases market liquidity which is a pretty straightforward observation.

The next three questions show a difference of opinion. In regards to the question about increasing the frequency of market crashes, a much larger percentage disagree in industry while the academia is much more reserved about this question. About the question HFT increasing market volatility both parties agree that it does but once again academia is more reserved and the percentage of people without a definite answer is much larger. Finally in question 12 about the frequency of market crashes as expected industry disagrees that HFT increases this frequency while academia agrees.

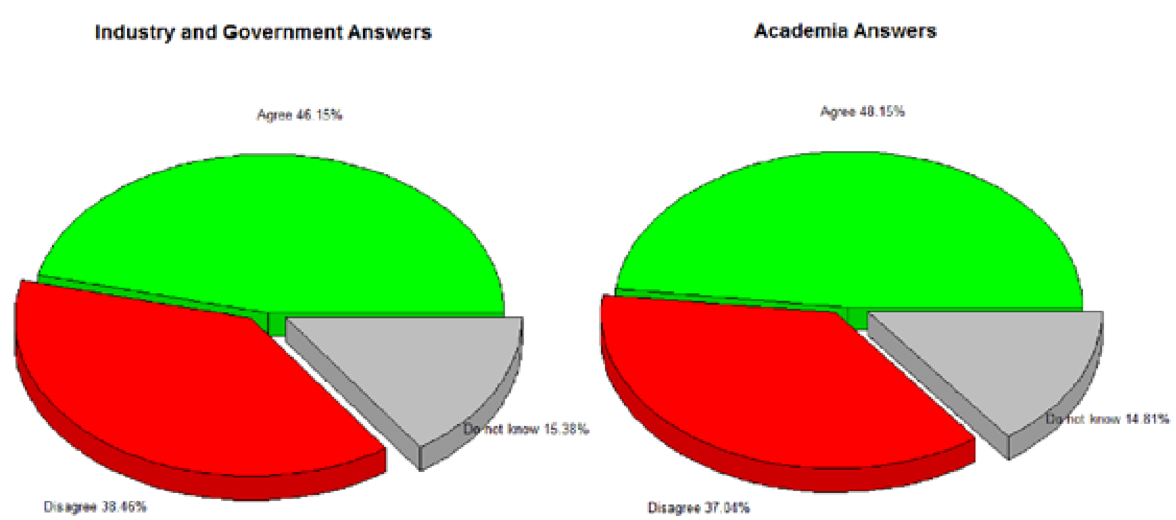

Figure 13: Answers for Q13 "In your assessment, do HF traders have an unfair advantage over other market participants who do not practice HFT?"

Question 13 about HFT possessing an unfair advantage over the market participants contains the most interesting results in this part. We expected the answers to show a dichotomy of opinions a la question 12. However, the results for the two groups are strikingly similar and in particular it is very surprising that 46% of the industry considers HFT as having an unfair advantage. However, given the small difference in opinions (46% agreeing and 38% disagreeing) and the importance of the question it is clear that more studies are needed to give a definite answer to it.

SURVEY PART 4: THE NEED FOR REGULATING HFT

This section is trying to gather the opinion of the population on a very important question about the future of HFT. Clearly if more regulations are coming than the HFT will need to modify its profile and algorithms and this is typically not desired by them. However, from the perspective of both large investors and government regulators this is an important question.

Not surprisingly, industry and academia answers differ when asked about the need to regulations addressing the amount of messaging traffic/unit of time (question 14).

Both parties feel that there is a need for more regulations in HFT and one interesting result is the academia opinion that banning the HFT is a ridiculous idea and not one single choice of this option was made by academia (Figure 15).

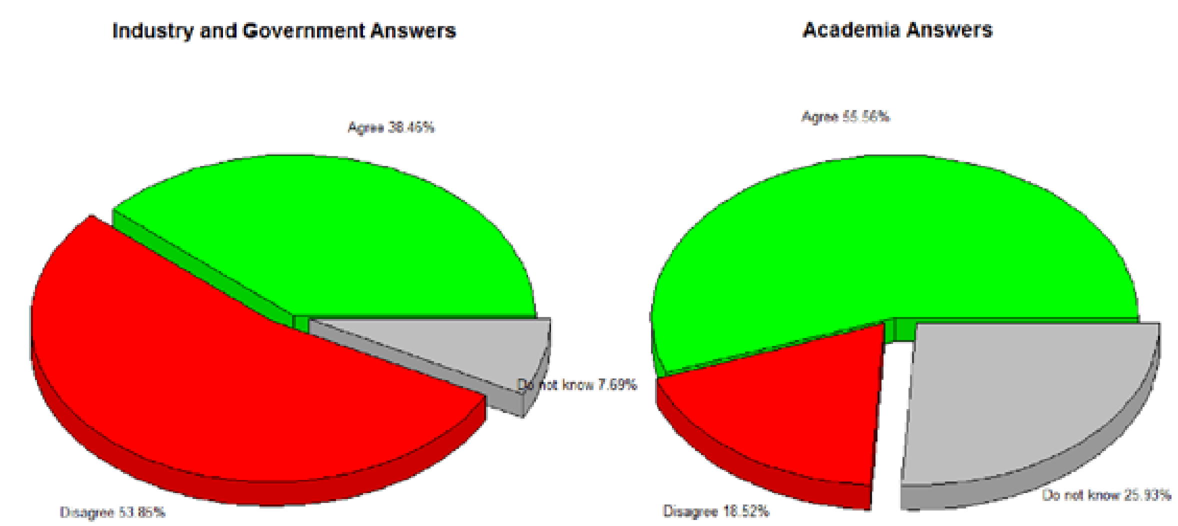

Figure 14: Answers for Q14 "In your assessment, do markets need more regulation restricting the amount of quotes/trading assets per unit of time?"

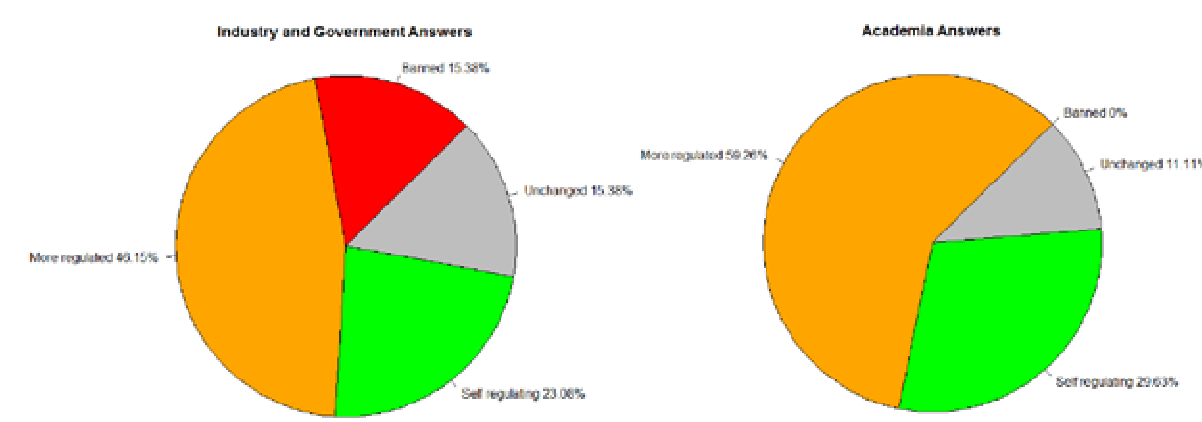

Figure 15: Answers for Q15 "HFT should be:"

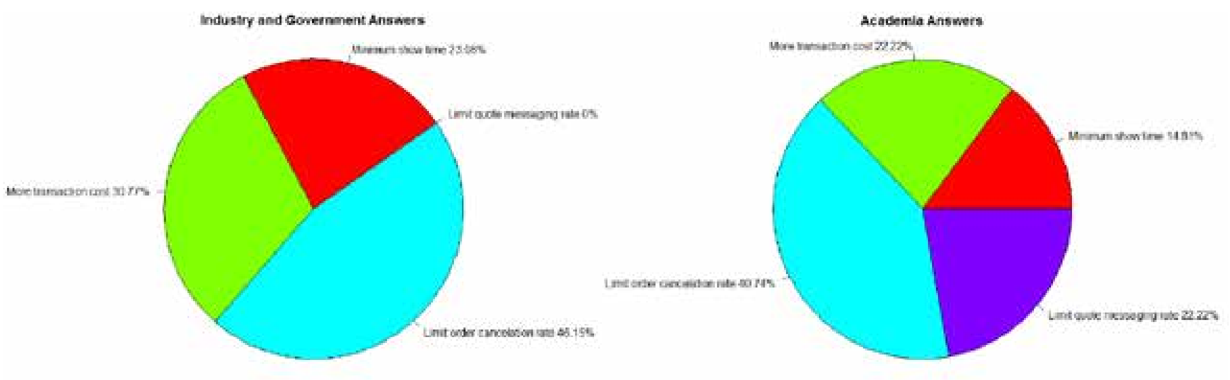

Figure 16: Answers for Q16 "Which of the following regulations should be implemented?"

Question 16 asked the respondents to choose one of the following 4 options

- Imposing more transaction tax
- Limiting quote messaging rate
- Imposing minimum order show time
- Limiting order cancelation ratio

It is worth noting that a selection (and only one selection) was required as answer to this question. Looking at the answers summarized in Figure 16 it is worth noting that academia is indifferent to the type of regulation imposed. The majority of respondents in both categories selected to limit the order cancelation rate (since they had to choose one selection). It is also easy to interpret the industry being adamant against regulation limiting the quote messaging rate since the entire HFT industry is built upon the capability to read fast and react faster to market changes than the rest of market participants.

SURVEY PART 5: HFT FROM THE PERSPECTIVE OF LARGE INVESTORS

The last part was not labeled in a particular way but we believe it contains two questions important for large investors. Question 18 asked the participants whether or not to invest in HFT and what is the reason. One needs to remember that the population under study is made of people connected with HFT either trading or researching it. Thus the reason why to invest is more important than the answer to whether or not to invest.

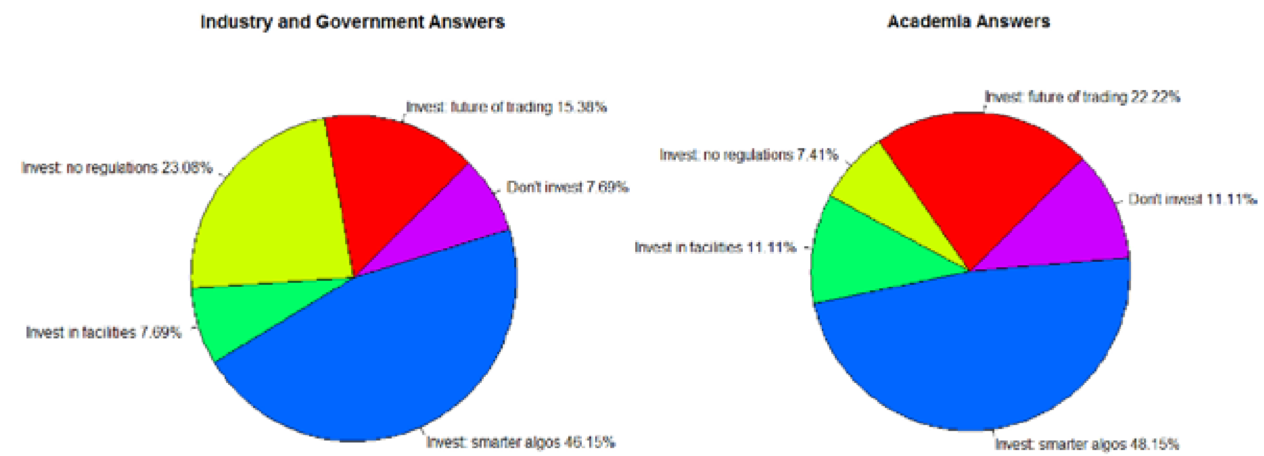

Figure 17: Answers for Q17 "Assume that you have a large sum for investing in HFT firms and/or designing an infrastructure for HFT. Please select the statement that most closely approaches your thinking:"
Possible choices were:

- **I will rather invest in a rather different area because the future of the HFT is uncertain**
- **I will definitely invest in HFT because this is where the future of trading is**
- **I will invest in HFT facilities just in case it becomes the norm**
- **I will invest in HFT algorithms and implement them because no regulation is coming and HFT has an advantage over everyone else**
- **I will invest in smarter algorithms for HFT because regulation is coming that will limit the frequency of the trades thus the need on relying on smarter rather than faster algorithms**

For the most part both parties agree with their respective answers. Furthermore, the majority in either group seems to believe that if one is to invest in HFT, it would make sense to invest in developing smarter algorithms rather than faster because some kind of new regulations are coming.

The last question asked the respondents about the return of investment for a HFT unit in percent per year. This was an open ended question and the answers varied from a range to numbers to "Most HFT strategies lose money over sufficiently large time interval, 2-3 years. Stable and robust algorithms, capable of placing above $50 mln. produce 15-25% net of commissions/rebates, dividends and financing.". To obtain numerical values we deleted the uncertain numerical values.

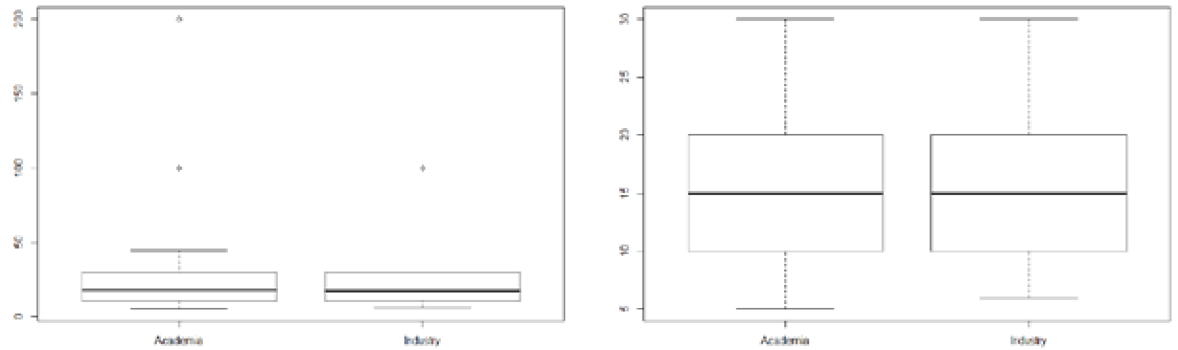

Figure 18: Boxplots for numerical values of yearly returns represented using side by side boxplots. The plot at right contains the same boxplots with outliers removed.

If we remove the outliers (values over 100% ROI) identified using the IQR criterion the two resulting distributions are remarkably similar. The median is somewhere around 15%.