Research Working Paper Series

Is High Frequency Trading beneficial to market quality?

Dr Richard Phillip
Business School
The University of Sydney

Dr Amy Kwan
Business School
The University of Sydney

Professor Graham Partington
Business School
The University of Sydney

OCTOBER 2015
WORKING PAPER NO. 083/2015 / PROJECT NO.T013

This research was supported by the Centre for International Finance and Regulation, which is a Centre of Excellence for research and education in the financial sector, funded by the Commonwealth and NSW Governments www.cifr.edu.au.
Abstract

This report discusses how high frequency trading (HFT) has changed the dynamics of the market and whether traditional academic measures of market “quality” are relevant in the new world of electronic trading. Using existing measures of market quality, which were designed over 20 years ago, much of the academic literature suggests HFT is beneficial for market quality. However, a closer examination of HFT reveals that the results may not be so beneficial and that many of these metrics are no longer applicable. This paper presents new metrics for market “quality”, which suggests that with the growth in HFT the probability of institutions getting orders filled has fallen and the time required to achieve a fill has increased. Additionally HFT trades tend to supply liquidity on the thick side of the order book, where it is not needed.
IS HIGH FREQUENCY TRADING BENEFICIAL TO MARKET QUALITY?

WHAT IS MARKET QUALITY?

When posing the question “Is HFT beneficial to market quality”, to expect a binary yes or no answer is a little naive. There are likely to be positive and negative aspects to HFT and whether on balance it is judged beneficial will depend upon the criteria used. How one defines market “quality” or if a trader’s actions are “beneficial” are open to interpretation. Accordingly, the notion of market quality and how it is to be measured must first be addressed.

Harris (2003) argues that the highest priority of financial markets is to promote the interests of utilitarian traders, or more specifically, those whose needs cause the markets to exist in the first place. A large portion of these utilitarian traders are institutional investors. Their interests should take high priority, because if they did not use the market, then the market might not exist, and then nobody would gain any benefit from the market. At the lower end of priority is the interest of traders seeking profits from trading rather than investment. In fact, Harris (2003) argues that their interests should only be supported when necessary to achieve other objectives. For example, a market maker, who adds to liquidity and contributes to price informativeness, should be supported. There is also one class of trader that Harris suggests markets should be hostile towards. These are the profit motivated traders who design trading strategies with the key purpose of exploiting other traders. Using the definition of HFT, provided in the next paragraph, it becomes apparent that high frequency trader’s fall into two possible categories; they are either profit motivated traders who support other market objectives, or they are profit motivated traders who design trading strategies to exploit other traders. Thus, the market should give HFT interests low priority, or potentially, be hostile towards them.

DEFINING HFT

It is apparent that a definition of HFT is required. Unfortunately, to date, there is no agreed upon definition for HFT in either the regulatory or academic environment. Definitions range from the very general to the very detailed (see Gomber et al. (2011) for an exhaustive list of different definitions used through time).

In our opinion, it is important to keep the definition as broad as possible, and HFT is thus defined as “a fully automated proprietary trading strategy which executes multiple intraday trades for profit”. The reason for this broad definition is that the market is constantly evolving and strategies are constantly changing. Categorising HFT by a very specific and distinct set of features is likely to prove an elusive goal. Consistent with this view is the current lack of consensus on a generally accepted definition of HFT. Markets are constantly evolving and HFT is simply the evolution of new trading strategies. A strategy that works today, may not work tomorrow, with a similar notion applying to the technology used. Furthermore, the idea that all HFTs behave in a similar manner is wrong. If all high frequency traders behaved in the same manner, then the inefficiency they were all exploiting would likely quickly disappear, and
shortly after so would the HFT firms themselves. HFTs do not all behave the same way and exploit the same inefficiencies. Brogaard (2010), for example, shows that HFTs demand (supply) liquidity in 42.7% (41.1%) of all dollar volume traded. This suggests that HFTs both take and supply liquidity in approximately equal quantities. Similarly, there are just as many contrarian HFTs as there are momentum HFTs. HFT is not one specific type of trading strategy, much of it is simply an evolution of traditional trading strategies into an automated process on a shorter time frame.

It is important to be clear about the definition of HFT as it is a loosely used term. As Vuorenmaa (2013) highlights, HFT and algorithmic trading (AT) are not the same concepts. Nonetheless, HFT is often defined as a subset of algorithmic trading (see Gomber et al. (2011) for a review of the different definitions of HFT and AT). Among the trading community however, HFT is not considered to be a subset of AT. As Vuorenmaa points out, HFT and AT are subsets of automated trading, but HFT is not a subset of AT. HFT strategies are designed to determine when a profitable trade should be made, whereas AT is about determining how to execute a large order so as minimise market impact. It is important to understand that these strategies and their objectives are different. In a following section, where a large body of literature which supports the benefits of AT is presented, it is often concluded that if HFT is a subset of AT and AT is beneficial to market quality then HFT must also contribute to market quality. This is not necessarily the case when it is realised that HFT is not a subset of AT. While they may use the same technology, their objectives, and hence the strategies behind them are very different.

THE RESULTS OF HFT

In recent years, institutional investors have been crying foul and have even begun looking for alternative investments and moving transactions off exchange due to their concerns about HFT. These concerns have also prompted regulators to conduct investigations into HFT behaviour. If HFT behaviour is driving institutional investors off exchanges, then the claim that HFT is beneficial or adding to market quality is a little unconvincing. The private interests of these institutional investors is what Harris (2003) argues to be the primary objective for market quality.

Perhaps the institutions concerns have not received more attention because the academic literature paints a positive picture of the effects of HFT on market quality. While some literature suggests that HFT has negative effects on market quality, for example Zhang (2010), a substantial majority of the research suggests HFT is beneficial to market quality. The main findings are that HFT activity reduces the bid-ask spread, increases the depth, or liquidity, in the limit order book, reduces volatility, increases price efficiency/discovery, and supplies liquidity when it is needed. A summary of the positive literature and its findings can be found in Table 1. What is clearly evident is that many papers provide similar results. However, this literature is a combination of results for both algorithmic and HFT traders. As mentioned before, HFT is not a subset of AT and any positive results found using AT data should not be automatically extended to HFT. In the final column of Table 1 an X indicates if the study used data that can identify an HFT trader with a high level of certainty, and a P indicates that a proxy for HFT was used. The method of identifying HFTs is important, as the use of proxies such as
the trade to message ratios to identify HFTs can significantly bias results. It is apparent from Table 1 that the literature which separately examines HFT is sparser than the body of literature examining both AT and HFT. Questions therefore arise about the validity of the conclusion that HFT is beneficial to market quality.

It must also be asked if the metrics used in the academic research are perhaps missing the picture. The metrics were created over 20 years ago, before the advent of extensive computer trading. These metrics were designed to measure how easy it was for somebody who wanted immediacy to execute a large order. But do institutional investors trade today in the same manner as they did in the days of open outcry pits? Equity markets have evolved significantly over the past 20 years, yet the metrics academics use have not.

MORE EFFICIENT PRICES/PRICE DISCOVERY

Based on Hasbrouck’s (1991) vector autoregression (VAR) model, studies commonly find that HFTs contribute to price efficiency. However, the VAR model has the implicit assumption that all information from the order flow is incorporated into prices - a notion which 20 years ago was reasonable. However, concern about this assumption was even raised in Hasbrouck (1991) when the model was first proposed: “The first part of the dichotomy asserts that the quote revision innovation reflects only public information. Since the quotes are provided by dealers or limit order traders, this assertion may be violated if either of these parties possesses superior information. While specialists do not typically engage in analysis of the sort that is likely to directly uncover corporate financial information, they do possess an informational advantage in their knowledge of the limit order book”. At the time this informational advantage was assumed to be insignificant, however, 20 years later it is very clear that information in the order book is a critical component for HFT trades and quotes, a notion supported by Brogaard, Hendershott and Riordan (2012). This suggests that Hasbrouck’s model and its extensions are miss-specified when using HFT data as its inherent assumptions are violated.

Even ignoring the issue of model miss-specification, conclusions from the use of the VAR model still have inherent problems when used to study HFT trading versus non HFT trading. When observing the impulse response function, it is generally interpreted that the larger the value of the response, the more permanent the price impact and hence the larger the contribution to price efficiency. However, the values computed often only use a response 10 events into the future. By construction, therefore, HFT trades will have a larger impulse response function than non HFT trades. This is because HFTs are trying to profit from short term price fluctuations; they are interested where the price will be in ten trades’ time. HFTs will be buying (selling) when they believe that prices are going to be rising (falling) in the very

---

1 Brogaard’s (2010) study shows that some HFT firms almost exclusively take liquidity, resulting in a message to trade ratio that is almost 1. On the other hand, a market maker would have an extremely high message to trade ratio. Naturally, a market maker is likely to be a lot less predatory than a firm constantly taking liquidity from the market.
<table>
<thead>
<tr>
<th>Authors/ year</th>
<th>Title</th>
<th>Supply liquidity when needed</th>
<th>Spread Tighter</th>
<th>Increased Liquidity/ Depth</th>
<th>Reduced Volatility</th>
<th>Improve Price efficiency</th>
<th>HFT Firms Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brogaard 2010</td>
<td>HFT and its impact on market quality</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Hasbrouck, Saar 2010</td>
<td>Low Latency Trading</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Hendershott Riordan 2009</td>
<td>Algorithmic Trading and Information</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Hendershott, Menkveld 2009</td>
<td>Does Algorithmic Trading Improve Liquidity</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angel, Harris, Spatt 2010</td>
<td>Equity Trading in the 21st century</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGM Advisors 2010</td>
<td>Market Efficiency and Microstructure Evolution in US Equity Markets: A High Frequency Perspective</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Credit Suisse 2012</td>
<td>Who Let the Bots Out? Market Quality in a High Frequency World</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brogaard 2011</td>
<td>High Frequency Trading and Volatility</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hendershott Riordan 2010</td>
<td>High Frequency Trading and Price Discovery</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Jarnecic Snape 2010</td>
<td>An analysis of trades by high frequency participants on the London Stock Exchange</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CME Group 2010</td>
<td>Algorithmic Trading and market dynamics</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lepone 2011</td>
<td>The impact of High Frequency Trading: International Evidence</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Riordan, Storkenmaier 2009</td>
<td>Latency, Liquidity and Price Discovery</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hendershott Moulton 2010</td>
<td>Automation, Speed and Stock Market Quality: The NYSE's Hybrid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Hendershott Riordan</td>
<td>Algorithmic trading and the market for liquidity</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 1: LITERATURE ON AT AND HFT AND SUMMARY OF RESULTS**
near future, whereas an institutional investor is investing with a long term outlook and looking to capture a long term alpha. Institutional investors have no interest in where the price will be in ten trades’ time; in fact, most brokers executing such orders simply periodically post a small component of the order into the market. If the magnitude of the impulse response function was not larger for HFT then they would be being beaten at their own game, which seems an unlikely scenario. Looking at longer term impacts, Bennett, Sias and Starks (2003) find that quarterly institutional order flow is positively correlated with future quarter returns; an implausible notion for HFT data, given that Brogaard, Hendershott and Riordan (2012) (Figure 1) show that HFT net order flow does not have significant correlation with returns beyond ten seconds.

An important question is how do HFTs know where prices are moving in the near future? Given that Cao, Hansch and Wang (2009) find that imbalances between buyers and sellers in the limit order book predict short run price movements, this is an obvious piece of information used for some HFT strategies. These strategies exploit all the order flow entering the market to extract alpha whereas institutional investors exploit fundamental information to try to extract alpha. If so, it follows that HFTs are exploiting order flow arriving from the institutional investors, but institutional investors make no exploitations of HFTs. This means that institutional investors are getting worse execution as a result of HFT. Essentially, HFTs are skimming cream off the top of an institutional investor’s fundamental analysis, but they win regardless of whether the institutional investor’s analysis was correct or not.

TIGHTER SPREADS

It is no surprise that the growth of HFT has narrowed spreads; this is simply a result from economies of scale. According to Harris (2003) the spread is composed of two components. The first is the transaction cost component and the second is the adverse selection component. The transaction cost spread component is the part of the spread that compensates the market makers for their normal costs of doing business. These normal costs of doing business are primarily fixed. An automated market making computer can monitor almost an endless number of stocks, whereas a human is limited to only a handful of stocks. As such, with automated market making, the profit per trade required to cover these fixed costs has dramatically decreased and accordingly, the spreads have narrowed to reflect this. For example, if a market maker must cover $1000 fixed costs a day and manually can only make 100 trades per day for a parcel of 1000 shares then the transaction cost spread component must be one cent or more to cover the fixed cost. But if a machine has the ability to make 500 trades per day on the same sized parcel of shares the transaction cost spread component is now reduced to only one fifth of a cent.

Tighter spreads are generally considered beneficial to market quality. However, there are trading strategies which will result in a tighter spread, but which are detrimental to the interest of institutional investors. One example is the strategy of front running or penny jumping. This strategy involves submitting a limit order one tick in front of large passive limit orders, and as a result tightening the spread if that large order is at the best bid/offer. If the front runner’s order is executed, the payoff is option like. If prices move unfavourably, the front runner’s losses are limited as passive limit orders behind the order provide price support. However, if
prices move favourably, the front runner’s order has unlimited upside. By taking liquidity that should be available to the passive institutional investors, front runners profit at their expense. This is a strategy which adds nothing to market quality, yet will improve the bid/ask spread metric.

An example of how a front running strategy can destabilise markets was evident during the mid-2000s. An anonymous trader, now famous among the trading community, known only as the Flipper, but later identified as Paul Rotter, was extremely successful by taking advantage of automated front runners. The Flipper’s strategy was simple; he would go long on a position and put an opposing ask quote several ticks above where his long position got filled, with the aim to go short there. With these steps in place he would then submit an extremely large bid near, or at, the best bid. This would trigger a whole array of HFT activity all using a front running strategy. Within a short time, the market would be driven up several ticks and his ask order would be filled, leaving him with a short position. At this stage he would cancel his large bid order and submit an extremely large ask order near the best offer. This in turn would set off all the HFT front runners trying to sell and would drive the price back down several ticks, thereby allowing the flipper to make money on the artificially created rise and fall. This is an example of how HFT strategies can significantly add to transitory volatility and pricing inefficiency.

Litzenberger, Castura, Gorelick (2012) make the important point that when discussing the spread, a comprehensive view of liquidity is important. Under certain conditions, for institutional investors a tight spread is of little concern, or sometimes even unwanted. This arises from the fact that when a large order is being executed it can be split into smaller orders that will sometimes make the spread and sometimes pay the spread.

**INCREASED LIQUIDITY**

The notion that HFT increases depth or liquidity is again not a great surprise. Using the previous example of an automated market maker versus a human one, it is evident that the automated market maker could have thousands of active limit orders near the spread across many stocks. Whereas the human market maker may only be able to manage 20 orders at once, the computer is able to monitor many more limit orders in the book than a human, and accordingly, these bots are adding more depth to the market. Once again, this seems like a positive attribute to market quality, however, if this metric is tweaked to reflect supplying liquidity when it is most important, it appears that HFT is not contributing much to market quality.

Cao, Hansch and Wang (2009) show that the imbalance between the amount of liquidity available for buying and selling predicts short term price movements; a result which arises due to a temporary imbalance between supply and demand. Following from this, supplying liquidity on the thick side of the order book is of little value, as there is already a surplus there. The real value is supplying liquidity on the thin side of the order book, where it is most needed.

---

2 A detailed analysis is given in the section considering new metrics for market quality.
The idea that HFT supplies liquidity where it is least needed is reflected in Brogaard (2010). Here he measures the impact of various sized trades if all HFT orders are removed from the limit order book, this is also repeated for non HFT traders. The results show that, depending on the size of the order, the impact when non HFT orders are removed is between 3 and 15 times larger than the impact when HFT orders are removed. This is attributed to non HFTs supplying much more depth in the order book. In 41% of the trades, HFTs supply liquidity, implying that non-HFTs supply liquidity in 59% of the trades. Using this as the ratio for HFT versus non-HFT limit orders suggests that the impact when non HTF orders are removed should only be around 1.44 (59/41) times greater than when HFT orders are removed. A plausible explanation for this large discrepancy in impact is that HFTs are more often than not supplying liquidity where it is least needed. In other words, HFT are supplying liquidity on the thick side of the order book where there is already a surplus, but less frequently supply liquidity on the thin side of the order book where liquidity is most needed.

**REDUCED VOLATILITY**

The evidence on whether or not HFT reduces volatility is inconclusive. While Brogaard (2010), Hasbrouck and Saar (2010) and Hendershott and Riordan (2012) provide evidence that HFT has no effect, or reduces volatility, the previous description of the Flipper’s trading style shows how HFT can exacerbate volatility and temporary pricing inefficiency. Evidence of the later can be found in Kirilenko, Kyle, Samadi and Tuzun (2011), who show that while HFTs did not trigger the Flash Crash of May 6, 2010, their responses to the large selling pressure exacerbated market volatility. Boehmer, Fong and Wu (2012) provide robust results showing short term volatility systematically increases when AT and/or HFT intensity increases. Zhang (2010) studies the effect of HFT on long term volatility and provides evidence that HFT activity increases long term volatility. However, Kirilenko et al. (2011) and Zhang (2010) both used proxies for HFT, thus their results should be treated with caution.

**INDUSTRY CONCERNS**

In the discussion above, some of the potential drawbacks of work based on classical academic measures of market quality were suggested. The following discussion addresses some of the concerns about HFT that are particularly relevant to the investment industry.

Trading is a zero sum game; accordingly, if HFTs are making money it must be at others’ expense. Brogaard (2010) estimates that 26 trading firms collectively made approximately $3 billion annually or roughly $115 million per firm, arguably no small amount. Often when the zero sum game concern is presented, proponents of HFT argue that this is compensation for the risk of supplying liquidity. This argument is flawed for two reasons. Primarily, as Brogaard (2010) shows, HFT supply roughly the same amount of liquidity as they take. But the second reason is that there is almost no risk. A one off event of posting a limit order carries adverse selection risk, but if thousands of strategic orders are placed all with a positive expected return then any risk is quickly diversified away.

For example if the payoff of a limit order was replaced by a coin tossing game where a head earns a dollar and a tail loses a dollar. To entice a risk-averse person to play this game it must
have a positive expected payoff. So let the coin be favourably biased with probability of heads be 53%. Playing this game just once, then there is an almost fifty percent chance of losing money. However, if this game is played 1000 times, then the probability of loss is only 3%. This is one of the great appeals of HFT as it takes advantage of the fundamental law of active portfolio management probably better than any other strategy. As the evidence in Menkveld (2013) shows, the strategy of an HFT is conservatively reported to have a Sharpe ratio of 7. To give the reader a visual interpretation of the risk-return trade off of such a strategy, Figure 1 plots the simulated cumulative percentage return of a strategy with a Sharpe ratio of 7 over a two year period. Clearly, risk is of little concern and so there should be little or no reward for bearing any perceived risk from supplying liquidity.

FIGURE 1: CUMULATIVE PROFIT FOR A TRADING STRATEGY WITH A SHARPE RATIO OF 7

The almost risk free profits that can be made by some HFT strategies, does not in itself mean all are detrimental to market quality. If a new player has entered the market and is consistently profitable, they cannot instantly be condemned because they are more successful than their contemporaries. As previously mentioned, some HFT strategies are likely to be the evolution of a strategy into a shorter time frame and across more stocks. For example, there are evolutions of the simple pairs trading strategies developed more than 30 years ago now being executed at an intraday level. However, there are certain aspects of HFT which should be raising concern.

A primary concern is the conflict of interest which arises from any broker that also has a proprietary trading book, in particular ones with dark pools. It is these market players who are best positioned to partake in HFT strategies. They already have the infrastructure set up for HFT and so would have the lowest marginal cost to trade. Furthermore, brokers compete for order flow and use executed volume as a metric to determine how they are positioned among their competitors and as a signal to the public. An HFT strategy which results a large amount of turnover can be very appealing as it will increase the broker’s volume metric.
Where a broker has a prop desk there is a concern about inherent conflict of interest. Brokers should be working to get best execution for their clients, yet no broker will advertise that they can consistently beat a day’s VWAP. Yet, an HFT strategy can consistently make money (see Menkveld (2013)), which implies it must be buying at less than VWAP and selling at higher than VWAP, thereby consistently beating VWAP. So the question arises why can a broker not consistently beat VWAP over the day for a client’s orders, yet their prop desk can? One possible explanation for this is that any potential edge discovered by their analysts goes into the prop trading where its profit potential is maximised, however this comes at the expense of a broker no longer meeting its fiduciary obligations of getting best execution for its clients. A broker with its own dark pool, or internal crossing engine, with a prop book is of particular concern to their clients. Time and price priority of the limit order book are the two pillars for maintaining market integrity, yet a dark pool has the ability to circumvent the time priority rule. Accordingly, a broker with an internal crossing engine can at any stage implement the front running strategy mentioned earlier, but now, depending on regulations, with an almost negligible price improvement.

It is not the speed of information or technology used to gather information and execute trades that is the worry, but rather the strategy used that is of concern. Strategies which are detrimental to market quality have been around since the early days of financial markets, but when the strategies can become fully automated and repeated thousands of times over then the risks of adverse effects increases.

Not all HFT is detrimental to market quality. One example is when a market is fragmented. A HFT strategy of cross market arbitrage is able to maintain the law of one price and transfer liquidity from one venue to another if required. Naturally the HFT firm will extract a fee or profit for this service. The positive benefits of this HFT strategy are further examined in Menkveld (2013). However, as with many strategies, it can become a fine line between positive and negative contributions to market quality. For example, if the HFT was using the practice of quote stuffing, such that the investor was misinformed as to which exchange had the best pricing, then the HFT strategy involves manipulation and predatory behaviour and is detrimental to market quality.

NEW METRICS FOR MARKET QUALITY

In this section possible new metrics for market quality are proposed. We first examine institutional investor’s current trading style. This facilitates an understanding of the institutions’ primary concern when their orders are being executed.

Traditionally a large institutional investor would give a large order to their broker to execute. The broker then executes it; often by breaking this single large order up into several smaller orders worked throughout the day in an attempt to minimise impact. In recent years, these orders are often no longer being executed by a person, but instead by a trading algorithm. However, the basic principle, of subdividing one large order into several small ones, remains the same.
PROBABILITY OF FILLING AN ORDER

A simplified, but not grossly unrealistic, execution strategy for an algorithm executing a large buy order during the day is subdividing this order into $n$ equal sized smaller orders. These $n$ orders are then periodically submitted as limit orders at the best bid. If, after some specified amount of time this limit order is unexecuted, then this order becomes a market order and is executed by crossing the spread.

If the limit order is executed, with probability $p$, before becoming a market order then the order outperforms the benchmark, which in this case is the midpoint at time of order entry, $t$, by half the spread, $S$. However, if this limit order does become a market order then it underperforms the benchmark a minimum of half the spread, but possibly more. This additional underperformance occurs if the market has moved away from this order since its time of submission. Under these conditions it will underperform by half the spread plus the number of ticks the market has moved away from the prior price, which is defined as $u$. Algebraically, the expected outperformance of the benchmark of this strategy is defined as:

$$ E[r] = p \frac{S}{2} - (1 - p) \left( \frac{S}{2} + u \right) $$

(1)

It is easy to see that if $p > 0.5$ and $u$ equals zero, then market participants implementing this strategy would actually prefer a wide spread as they would expect to outperform the benchmark. If $p = 0.5$ then the bid ask spread plays no effect on this execution strategy’s performance. Only under the condition that $p < 0.5$ is the spread of concern for a large investor using this execution strategy.

By observing equation (1) it is apparent that an institutional investor’s primary concern is that there is a high probability that their limit order goes unexecuted, or that $p$ is small. Accordingly, a metric to reflect this concern is developed and is defined as the probability of fill. This is computed by observing all institutional orders posted at the best bid, or best ask, then observing if this order gets at least partially filled before it is cancelled or amended. The number of orders which get filled on a given day divided by the total number of orders submitted that day gives the probability of fill for a specified day.

TIME FOR ORDER EXECUTION

What the above probability metric does not account for is the time it takes to get the partial fill, for example an order left unattended in the order book for an extended time has a much higher probability of getting filled than an order posted then quickly cancelled. As such, another metric to complement the probability of fill is also designed, this metric is named expected time to fill. To compute the expected time to fill for a day, all institutional orders which were posted at the best bid or ask and receive a partial fill are identified. With this information the average time taken from submission of the order until it is executed is computed and is the expected time to fill for a specified day. For robustness the time computed is also measured in event time, or more specifically the number of order book events between the order submission and order execution. The use of expected time to fill is to capture the notion that the market has evolved through time. Decades ago a millisecond was an insignificant amount of time on an
exchange, now it is considered a substantial interval of time. Who, even ten years ago, would have seriously considered spending $300 million to lay a high speed fibre optic cable, in order to cut the time it takes to send a message from New York to Chicago from 16 milliseconds to 13 milliseconds.

**SUPPLYING LIQUIDITY**

Does HFT supply liquidity on the side of the order book where it is most needed, or does it only add to the side of an order surplus thus exacerbating any temporary imbalance? A suitable metric to capture this is termed *top of book imbalance* and is defined as follows:

\[
Top \ of \ book \ imbalance \ t = \frac{Vol_{Bid,t-\epsilon} - Vol_{Ask,t-\epsilon}}{Vol_{Bid,t-\epsilon} + Vol_{Ask,t-\epsilon}}
\]

Where a trade occurs at time \( t \) and the volume on the best bid (ask) just prior to execution is defined as \( Vol_{Bid,t-\epsilon} \) (\( Vol_{Ask,t-\epsilon} \)). This metric is bounded between -1 and 1 and is a standardised measure of the imbalance at the top of the order book just prior to an order being executed. This metric will be examined for all possible HFT trades including buys which either take or supply liquidity and sells which take or supply liquidity. If HFT supplies liquidity on the side of the book where it is least needed then we can expect that when a HFT buys (sells) then *top of book imbalance* is positive (negative) on average.

**DATA AND RESULTS**

The data used for this research uses the trading activity on the Australian Securities Exchange (ASX) covering the period from January 1, 2009 to August 27, 2013. This data was obtained from the Securities Industry Research Centre of Asia-Pacific (SIRCA). The data contains every order entry, emendation, and cancellation along with trades and broker IDs for each order. An example of the data set can be found in Table 2 below.

<table>
<thead>
<tr>
<th>Ticker</th>
<th>Date</th>
<th>Time</th>
<th>Event Type</th>
<th>Price</th>
<th>Volume</th>
<th>BidId</th>
<th>AskId</th>
<th>Bid/Ask</th>
<th>Old Price</th>
<th>Old Volume</th>
<th>BuyerBrokerId</th>
<th>SellerBrokerId</th>
</tr>
</thead>
<tbody>
<tr>
<td>BHP</td>
<td>20130102 10:21:39.870</td>
<td>AMEND</td>
<td>37.470</td>
<td>13</td>
<td>27254</td>
<td>A 37.470</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>210</td>
<td></td>
</tr>
<tr>
<td>BHP</td>
<td>20130102 10:21:42.286</td>
<td>DELETE</td>
<td>37.460</td>
<td>14779</td>
<td>31385</td>
<td>B</td>
<td>31385</td>
<td>29679</td>
<td>29679</td>
<td>397</td>
<td>297</td>
<td></td>
</tr>
<tr>
<td>BHP</td>
<td>20130102 10:21:42.341</td>
<td>ENTER</td>
<td>37.460</td>
<td>2130</td>
<td>31385</td>
<td>29679</td>
<td>297</td>
<td></td>
<td></td>
<td></td>
<td>297</td>
<td>366</td>
</tr>
<tr>
<td>BHP</td>
<td>20130102 10:21:42.341</td>
<td>TRADE</td>
<td>37.460</td>
<td>2130</td>
<td>31385</td>
<td>29679</td>
<td>297</td>
<td></td>
<td></td>
<td></td>
<td>297</td>
<td>366</td>
</tr>
</tbody>
</table>

The broker ID is used to identify if the order originated from a broker with institutional or retail clients or alternatively came from one of the HFT firms operating in Australia who have their own exchange membership. The HFT firms are VIRTU, GETCO, Susquehanna, IMC, Optiver, and Tibra. This will not capture the full set of HFT trades, but it will capture a clean sample of the HFT population.
Using the broker ID for each order it was determined if the order was submitted by an HFT firm, institutional broker, retail broker or other. Some HFT will flow through institutional brokers and as explained below we filter these orders in an attempt to reduce misclassification of HFT trades as institutional trades. Any misclassification is likely to bias against finding negative effects from HFT. Assuming the “misclassified” HFT get execution approximately comparable with the correctly classified HFT then the “true” not HFT will do worse that our results suggest.

The data set also allows the limit order book to be completely reconstructed so that every individual order and its queue position is known. The ability to reconstruct the full order book also allows identification of which order was supplying (demanding) liquidity for every trade without the need for approximation techniques such as the Lee and Ready Algorithm.

The period studied was chosen to span almost 5 years commencing in 2009. In 2009 HFT was still in its early stages on the ASX. Accordingly, using this time span it is possible to observe how the market has evolved through time as HFT activity has increased. For illustrative purposes the results are presented for only one stock, BHP, which is one of the most actively traded stocks on the ASX and until March 2013 also had the highest market capitalisation, but the results hold true for all other actively traded stocks.

When computing the probability of fill and the expected time to fill institutional orders posted at the best bid and ask must be identified. The identification of orders originating from an institutional broker is achieved using the broker ID’s, however there is no guarantee this order has come from an institutional client as there is no certainty that the broker does not have its own HFT proprietary trading desk. To mitigate this problem a filter is applied so that only the largest 50% of orders identified as institutional are included.

**PROBABILITY OF A FILL**

The probability of an institutional buy (sell) limit order posted at the best bid (ask) getting at least partially filled is computed for every day across the sample. A time series regression defined by equation (2) is used to determine if the probability of fill has changed over time.

\[
\text{Probability of fill}_t = \beta_0 + \beta_1 t + \epsilon_t
\]  

(2)

If \( \beta_1 < 0 \) then we can conclude that over time, as HFT has become more prevalent, then the probability of an institutional investors limit order being executed has decreased. In addition to the regression model, a visual display of how the probability of fill has changed through time is also provided in Figure 2. This is achieved by using a locally weighted regression.

From Table 3 it is clear that \( \beta_1 \) is significantly less than zero, suggesting that as HFT has become more prevalent an institutional investor’s limit order becomes less likely to execute. This can also be seen visually from Figure 2 where at the beginning of the sample period an institutional investor’s limit order is expected to execute more than 50% of the time, whereas by the end of the sample, the expected probability of execution has decreased to 30%. 

The expected time of an institutional buy (sell) limit order posted at the best bid (ask) getting at least partially filled is computed for every day across the sample. A time series regression defined by equation (3) is used to determine if the expected time to fill has changed over time.

\[ \text{Expected time to fill}_t = \alpha_0 + \alpha_1 t + \epsilon_t \]  

(3)

If \( \alpha_1 > 0 \) then we can conclude that over time, as HFT has become more prevalent, then the expected time for an institutional investors limit order to be executed has increased. In addition to the regression model, a visual display using a locally weighted regression is presented in Figure 3 and 4 which displays how Expected time to fill has changed through time. Figure 3 represents Expected time to fill in wall clock time while Figure 4 represents Expected time to fill in event time.

The results of the regression in Table 4 show \( \alpha_1 \) when Expected time to fill is reported in clock time (Panel A) and also in event time (Panel B). Using either metric it is clear that \( \alpha_1 \) is significantly greater than zero. This suggests that as HFT has become more prevalent that the time an institutional investor must wait for their limit order to execute has increased. This can
also be seen visually from both Figure 2 and 3 where it can be observed as time goes forward the expected waiting time of an institutional order to execute has increased.

### TABLE 4: REGRESSION RESULTS FOR EXPECTED TIME TO FILL (EQUATION 3)

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimate</th>
<th>Std Error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\alpha_0$)</td>
<td>11.61</td>
<td>8.70E-02</td>
<td>133.289</td>
<td>&lt;2E-16 ***</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>5.48E-04</td>
<td>1.29E-04</td>
<td>4.267</td>
<td>2.14E-05***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimate</th>
<th>Std Error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\alpha_0$)</td>
<td>129.8</td>
<td>5.83E+00</td>
<td>22.28</td>
<td>&lt;2e-16 ***</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.13</td>
<td>8.60E-03</td>
<td>14.88</td>
<td>&lt;2e-16 ***</td>
</tr>
</tbody>
</table>

**FIGURE 3: EXPECTED WAITING TIME MEASURED IN SECONDS.**
TOP OF BOOK IMBALANCE

The above metrics are consistent with institutional investors being squeezed out from the limit order book by HFT traders, evidenced by the fact institutions now face longer waiting times for order execution and there is less chance of an order being executed. Perhaps this might be attributed to competition by HFT market makers who are competing to provide liquidity when it is needed. The top of book imbalance metric is intended to indicate if institutional orders are being crowded out of the order book because there is always HFT competition on both sides of the book, or, whether HFTs are providing liquidity where it is least needed by front running institutional orders and effectively crowding them out of the limit order book.

The results for this study only uses the most recent year of data as this is when HFT is at its most mature. Initially, the distribution for top of book imbalance is determined just prior to a trade for every type of market participant. It is expected that such a distribution will be symmetrical and centred on zero and by construction be bounded by -1 and 1. A histogram for top of book imbalance just prior to every trade is depicted in Figure 5. Here it is noted that it is symmetrical, with a spike in observations near -1 and 1, a result due to the non-linearity of the variable. What is most important about this distribution is that there is symmetry around zero, implying the average market participant is not systematically trading in front of the thicker side of the order book.
The top of book imbalance is computed for HFT traders and depicted in Figure 6 below. Figure 6 depicts a histogram of the aggregate of all HFT trades, however the same results hold true when each HFT firm is analysed separately. Panel A and B of Figure 6 show HFT buy orders that demand and supply liquidity respectively, while Panel C and D shows HFT sell orders that demand and supply liquidity respectively.

The results for the top of book imbalance when HFTs demand liquidity (see Panels A and C) shows that HFTs consistently use buy (sell) market orders when the volume on the bid (ask) is much larger than the volume on the ask (bid). It seems that when HFTs trade and are demanding liquidity they more often than not seek it from the thin side of the order book. This scenario is consistent with front running large limit orders by HFTs exacerbating an order imbalance. Alternatively, the argument may be put that when the HFT is buying (selling) and there are already many limit orders already at the best bid (ask), then submission of a market order by the HFT is the way to achieve quick execution. Whether it is front running, or seeking quick execution, the liquidity is supplied not where it is needed, but on the thick side of the order book.
Panel D and C of Figure 6 show some evidence that when HFTs trades supply liquidity there is a tendency for that trade to be on the thin side of the order book. The relative provision of liquidity by HFT and non-HFT traders is examined in Figure 7. Figure 7 shows Panels D and C of Figure 6, but with the histogram of *top of book imbalance* for non-HFT trades overlayed. From figure 7 it is clear that non-HFTs have a larger proportion than HFTs of their trades on the thin side of the order book, thus supplying liquidity where it is most needed.

The conclusion from the analysis of the *top of book imbalance* is that when demanding liquidity HFTs make strong demands on the thin side of the order book and when supplying liquidity HFTs provide some supply on the thin side of the order book, but non-HFTs have a greater tendency to supply liquidity on the thin side of the order book.
FIGURE 7: LIQUIDITY SUPPLY BY HFT AND NON-HFT

HFT buy supply vs non-HFT buy supply

HFT sell supply versus non HFT sell supply

CONCLUSION

This report addresses some of the differences between the investment industry and academia’s perspectives of HFT. The report highlights several of the potential pitfalls when using traditional measures of market quality, which were created before electronic trading became prevalent. Three new metrics are developed, which are more appropriate to today’s high frequency trading environment. Using these new metrics, the evidence suggests that HFT trading strategies are likely to detract from, rather than add to, market quality.

We report three key findings. First, as HFT became more prevalent, the probability that an institutional limit order would execute decreased. Second, the average time taken for institutional order execution increased with HFT activity. Taken together, these two findings suggest that institutional investors are crowded out from the order book in the presence of HFT. HFTs strongly demand liquidity on the thin side of the order book and are less generous in supplying liquidity on the side of the order book where it is most needed. These results hold for all HFT firms in aggregate and at an individual level. The results are suggestive of HFT trading strategies that front run the limit orders of other traders.

Our conclusions are subject to some caveats. First they rest on the assumption that changes in the execution of institutional trades over time are due to the growth of HFT and not to some other correlated factor that has changed over time. We think it unlikely that some other factor explains our results, but we cannot absolutely rule it out. Second, our results depend upon our classification of orders and trades, particularly whether they are classified as institutional or HFT. There is always the possibility that some orders/trades were misclassified, but given the strength and consistency of the results and their robustness across individual stocks and
individual HFT firms it seems unlikely that this was driven by misclassification. Lastly, the analysis of liquidity supply by HFT was only based on one year’s data. Despite these caveats we consider that case to be that, contrary to much of the literature, HFT is on balance detrimental to market quality. Logically, if trading is zero sum game and HFT traders are making profits then it must be at the expense of other traders.
References


Brogaard, Jonathan., Terrence Hendershott and Ryan Riordan, 2012, High Frequency Trading and Price Discovery, working paper


Gomber, Peter., Bjorn Arndt, Marco Lutat and Tim Uhle, 2011, High Frequency Trading, working paper


Harris, Larry., 2003, Trading and Exchanges Market Microstructure for Practitioners, Oxford University Press

Hasbrouck, Joel., and Gideon Saar, 2010, Low-Latency Trading, working paper


Hendershott, Terrence., Riordan, Ryan., 2009, Algorithmic Trading and Information,

Hendershott, Terrence., Riordan Ryan., 2012, Algorithmic trading and the market for liquidity

Hendershott, Terrence., Moulton, Pamela c., 2011, Automation, speed, and stock market quality: The NYSE’s Hybrid


Riordan, Ryan., and Andreas Storkenmaier, 2009, Latency, Liquidity and Price Discovery, working paper.

Van Kervel, Vincent, 2012, Liquidity: What you see is what you get?, working paper


Zhang, Frank, 2010, High Frequency Trading, Stock Volatility, and Price Discovery, working paper