Schroders

Know Your Counterparty: The New Paradigm of Equity Market Microstructure and The Impact to Institutional Investors.

Will Psomadelis, Head of Trading, Australia, Schroders
Stuart Baden Powell, Head of European Electronic Trading, RBC Capital Markets

Abstract

The global debate about High Frequency Trading (HFT) being beneficial or detrimental for markets is ongoing. The concern for institutional investors should not be about the legitimacy of HFT, rather it should be about identifying, measuring and minimising transaction costs no matter the source. With the explosion of HFT volumes over the past five years, it is impossible for the buy side to accurately measure transaction costs without properly understanding the nature of the counterparty, to what is now a significant portion of their trades. During this paper we examine some wider relationships between dark trading and HFT covering several topics as background before moving on to provide data from conducted primary research.

The paper decomposes HFT into two distinct subsets, those being “microstructure arbitrage” and “high frequency quant” to identify how both sources of cost can be measured and handled. Empirical evidence of systemic losses in certain Australian dark pools is provided along with a study of lit market order book cancellations and the impact on prices immediately after a cancellation event. A framework for the measurement of market impact including trade reversion after the order has completed is also provided.

Introduction

Defining HFT has been a material problem with many arguing that it is in fact impossible to define. HFT lobby groups scored a win in this regard by forcing some to take on the notion that HFT was a multi-faceted “tool” not a strategy that defined a firm; the result was that it pushed any attempt to regulate into the long grass; it is very difficult to regulate what cannot be defined. That said the recent German HFT Act does attempt to use a definition of HFT in an attempt to bring it under the direct control of regulator Bafin which may act as something of a precedence for others.

However, given there is no universally agreed definition, we split HFT into various subsets – in fact into two distinct strategy umbrellas, one being microstructure arbitrage and the other high frequency quant, this is for the purpose of allowing appropriate measurement of the impact of each format on institutional order flow. Whilst both subsets can individually increase transaction costs of the institutional investor, when both strategies are deployed concurrently, it becomes clearer why traditional institutional trading techniques are no longer best practice. This new paradigm of liquidity leads to a deterioration of portfolio returns through transaction costs, no matter your investment horizon, and needs to be understood by the buy-side\(^1\) before trading strategies can be properly designed to counteract it.

\(^1\) The buy side is the informal name given to institutional investors, mutual funds and hedge funds amongst others. The Sell Side is the informal name given to investment banks and brokers.
For professional advisers only

The point about the mismatch in investment horizon between buy side and HFT is one often made but in reality it can be seen as a moot point. Stating that investment horizons for a buy side are long and for HFT are short and therefore the buy side should have no concern is not actually running the right comparative. By this way of thinking the buy side dealing desk is merged into the investment decision which, whilst in other facets of organisational structure may be important, in this particular area, it diminishes the value of the dealers on the desk and could lead to questions about their existence in tight economic times. Although a portfolio manager makes a long term decision and may for example buy a stock slowly to then hold for the long term of five years plus, the trading desk will trade each child order in a short term fashion often directly via algorithms or indirectly via cash desks who in turn use algorithms\(^2\). HFT, also operating in a short term fashion, can frequently be the counterpart to each trade. The role and the importance of the buy side dealing desk is central to the implementation of longer term ideas yet in reality each child order is often very short term. If costs are lost to a counterparty on say 50%\(^3\) of child orders by default the total parent order will incur weaker performance. As such, the “investment horizon” of the buy side does not represent its practical “dealing horizon” – the two are separate and the dealing desks can add substantial value to the investment process by monitoring and executing its short term implementation of the longer term view. During this paper we also look into what else the buy side can do at the dealer level.

The composition and sophistication of participants in electronic markets has changed dramatically over the past decade; yet many institutional traders continue to use benchmark strategies where the core logic has not evolved at the same pace as the market. The central flaw of adhering to an execution schedule has resulted in these benchmark algorithms being gamed by providing a signal that is discernible by, amongst others, genetic algorithms using Bayesian probabilities – for example those adopted by HFT. These signals lead to material cost increases for counterparts who revisit the market several times to complete one parent order\(^4\).

The investigation of systemic losses in dark pools uncovers some interesting results, crucially we measured returns immediately after an execution. Given the diversification of orders, order types, anti-gaming rules and counterparty selection, we expected a normal distribution of returns – however there were persistent losses in certain venues more consistent with arbitrage. A framework for the measurement of signal detection through price deviation during and post order completion is suggested.

Discussions about order cancellations have been increasing in frequency as the education around the order-to-trade ratio has led to increased cynicism about its cause and effect. A piece of research introduced and evaluated in this paper looks at order cancellations in the Australian market in the context of subsequent price moves. Whilst it is usually the objective of investors to actually buy or sell stock when a bid or offer is placed, the discovery of orders being cancelled as they are about to be fully executed is interesting.

By identifying two different catalysts for order cancellations, post cancellation price moves show that prices tend to move in the direction of the cancellation and can be a function of cancellation size. This may be a window into how the decisions of algorithms are made in terms of order aggressiveness based on any order book imbalance; however it raises an equally important question as to whether the buy-side should co-locate their algorithms to take advantage of this low latency phenomenon.

\(^2\) This excludes certain types of trading for example on risk, natural crossing etc.

\(^3\) This is an arbitrary number and not a reflection of HFT being approximately 50% of the US market volume. Transaction cost increases are not restricted to HFT as a cause, but may include other sell side algorithms, inappropriate trading strategies or algorithm selection and market impact to name a few other possibilities.

\(^4\) This excludes the “mom and pop / Mrs Watanabe” retail investor, typical retail investor order size could match HFT trade size meaning that in fact the retail investor may just simply achieve better execution. This is a very different story for the institutional investor however who represents a far larger portion of retail investors through collective investment schemes or pension funds.
HFT Decomposition and Institutional Impacts

With HFT being such a broad term and with no specific industry accepted definition, most trading firms can be differentiated by their average holding duration, trading techniques and strength of understanding of the nuances of market microstructure.

Treynor (Bagehot, 1971) describes the essence of market making as being where the “gains from liquidity motivated transactors (e.g. buyside cashflow trades or retail) must exceed his/her losses to information motivated transactions”. In this sense most executions are either liquidity motivated, have information or a combination of the two, however since the date of that quote, inefficiencies in microstructure have developed resulting in a complex web of opportunity in financial markets today. Abnormal returns can be generated by exploiting these manufactured arbitrage opportunities through simply intermediating trades that otherwise would not need intermediation whilst increasing the probability of transacting against liquidity traders and avoiding informed traders through early detection.

Because of this complexity and the way order information is disseminated, information about potential executions that may occur on lit markets can be released to selected counterparties whilst the parent order itself may remain within the broker “firewall”. This inefficiency has tilted the playing field away from the traditional liquidity providers to HFT where infrastructure advantage vies with risk management as a dominant factor in profitability.

Ultimately the spotlight is on the shorter end of the trade duration horizon where latency differentials, infrastructure arbitrage, statistical and near deterministic signalling tend to be main drivers of return. Simplistically these are firms that engage in microstructure arbitrage and ultra-short term quantitative high frequency trading processes. There is however, a distinct difference between the two and as such, they need to be analysed separately; it is this which we now review.

Microstructure Arbitrage

Microstructure arbitrage relies on differing levels of access, latencies and order types to generate returns. As an example, Haim Bodek, a former HFT employee, recently released a series of papers (and a book) that reviewed several order types used by some of the exchanges largest customers. Some order types allowed the user to gain queue priority. Although not new to well versed market structure practitioners, Bodek’s work is an important step in socialising issues to the wider trading community from someone “on the inside”. Some non-HFT practitioners have claimed the opposite to Bodek – that order types do not advantage certain market participants; however the area of concern is on both the understanding of the order type in question and any differences between the marketing of an order type and the reality of its function.

Being top of the order book is the primary objective of most HFT passive orders whilst the ability to be the fastest to market to aggressively unwind a position is equally vital. Speed plays a key role in order creation, event reaction and order cancellation therefore giving HFT a more balanced approach to trading rather than purely a passive or aggressive form in isolation. ASIC themselves provided objective evidence of this HFT “mixed” behaviour in the Australian market. The balance of HFT trading is vital when orders are deployed to test factors such as order book elasticity, and is contrary to the simply “passive” role that is often subjectively promoted by various HFT firms, commercially conflicted firms or lobby groups.

The ability of firms to gain the top of book advantage differs regionally and depends on both the structure of the market and its regulation. In the US, order protection and the trade through rule has played a role in the proliferation of order types used beyond the usual market, limit and hidden. Whilst the media has provided some information to the wider public on a number of these order types that can circumvent time and book

---

6 ASIC has quoted that HFT trades more actively than passively in stocks outside the S&P/ASX 200 and behaviour in the S&P/ASX 20 was roughly 50% active and 50% passive. (ASIC; March 2013, Dark liquidity and High frequency Trading, REP331, pp.81).
8 The Trade through Rule is designed to ensure that investors always receive the best explicit price no matter the venue that they send their order to placing an obligation on the exchange to onward route orders if better prices are available.
priority, in some cases it is order types such as the Intermarket Sweep Order (ISO)\textsuperscript{9} that have been shown to be the cause of some of the microstructure failures seen often through the end result of micro flash crashes. For example, Golub, Keane and Poon (Nov 2012) show that "mini flash crashes are the result of regulatory framework and market fragmentation; in particular due to the aggressive use of Intermarket Sweep Orders and Regulation NMS protecting only the “top of book”.

Most other countries have not suffered order type proliferation and complexity to this scale, and queue priority is generally achieved through more traditional latency differentials. That said Europe has also seen some interesting new order types come into play from alternative venues called MTF's\textsuperscript{10} and we expect more will arise in due course as volumes continue to stay low and order types increasingly become differentiators for venues. Therefore, it is critical to understand the details of new order types as well as study order type approaches from both user and crucially the trading counterpart perspective. As in the US there can often be more to the order type than first meets the eye and it is important to glean transparency from the providing venue and also your options for customization within the order parameters.

Perhaps related to the above, higher than market wide win ratios of HFT trading calls into question the theory of the random walk at the micro level especially when dark markets, client tiering and order routing are introduced into the market ecosystem. In some markets considering the predictive and reactive abilities of algorithms coupled with ultra low latency, an order that is top of book can easily be cancelled and a new order created to execute against the latent liquidity that was behind it at the same price. Such strategies allow firms to maintain increased levels of “riskless trades” across specific intervals whilst informed investors suffer lower fill rates leading to higher market impact for their orders. Concepts such as rebating firms for executions add economic viability to HFT cost control as well as profit generation.

Recent press on this topic (for example, latency hops in NYSE to specific feeds for which NYSE were fined 5million USD\textsuperscript{11}) are an indicator of some of the publicly reported practices that have raised eyebrows around the true meaning of fair access to markets. These issues are perhaps related to the increased level of complexity and subsequent opacity at the broker and exchange level. Related examples include issues such as Pipeline, Level and Liquidnet in the US which have also raised questions as did Chi-X Europe sending client ID tags on trades in 2010.

Opacity is by no means new when it comes to execution; however if we look to the US where orders can be subjected to thousands of routes per trade as pools are cross aggregated and the trade through rule is operational, the potential case for arbitrage is compelling. Although the trade through rule in Europe and Asia does not exist, the opportunities for information leakage are still profound – the concept of dark aggregators for example is an area that has yet to see significant research, particularly where those aggregating the flow are other brokers who are unlikely to gain the same tiered access to bank internal pools as buy side firms and could incur significant information leakage to counterparts. More Transaction Cost Analysis (TCA) and practitioner research should be conducted in this area for performance metrics as well as wider academic research if possible.

**High Frequency Quant**

High frequency quant processes can refer to short term statistical processes that trade off calculated probabilities of price reversion, momentum or demand/supply imbalances to name a few. The adoption of genetic algorithms (via import from other industries e.g. the Baum Welch speech recognition algorithm and particle filters amongst others) has meant that the signals being released by institutional orders are more readily identifiable. Whilst this branch of trading is not seminal today, these processes become a major factor in execution costs. Institutional investors may not be aware that their signal has value and is more easily identifiable to trading counterparties.

Easley, De Prado and O'Hara (March 2012) illustrate that the reluctance of investors to move away from wall clock time and into volume clock paradigm allows for a signal release via order imbalances at predictable

\textsuperscript{9} An ISO order is an order type that allows the participant to send an order directly to a venue even when another venue is publishing a better price. The participant must also fulfill their Reg-NMS obligations by sending concurrent orders to the venues with protected quotes. This order type can create a trade-through event where the executed prices pierce the protected quotes at the top of the consolidated order book.

\textsuperscript{10} Multilateral Trading Facility.

\textsuperscript{11} Patterson, S and Strasburg, J; Sep 14, 2013. *Nyse is fined in data probe*, The Wall Street Journal http://online.wsj.com/article/SB10000872396390443524904577651450485707824.html
time periods. Some investors utilise scheduling algorithms such as VWAP in an attempt to extend order duration and theoretically minimise impact (at the expense of timing risk). QSG\textsuperscript{12} counters this more traditional market impact theory by showing that VWAP can be a costlier strategy than Implementation Shortfall, despite a lower order duration, due to its statistical signal and increased interactions with the market\textsuperscript{13}. It is such repetition and strict adherence to arbitrary benchmarks such as open, close, VWAP\textsuperscript{14}, PWP and even to some extent IS that is providing the necessary signal for HFT to profit and is detrimental to portfolio returns. However, the benchmark and the algorithm deployed to achieve the actual benchmark are two separate components and it is the use of algorithmic schedules that causes the problem, not the benchmark in isolation. The benchmark and the algorithm used to achieve that benchmark are not always of the corresponding name. For example, a liquidity seeker with a name created for marketing may be used to achieve or beat the VWAP benchmark; however, despite sales gloss, as a general rule a schedule based algorithm or even many liquidity seekers have an inherent schedule that can be monitored and gamed.

Much of the core logic from this form of trading can be derived from Bayes Theory. The theory provides a solution to inverse probability and provides key insights into conditional probabilities; in the process of creation it changed the mathematics of probability from a gamblers measure of frequency into a measure of informed beliefs.

Bayes itself has been used throughout history by Napoleon for calibrating his cannons, by the brilliant British statistician/computer scientist Alan Turing at Bletchley Park to crack the German Engima machine during WWII, to helping find the missing USAF nuclear bombs\textsuperscript{15} in the Mediterranean. In trading Bayes is used extensively by HFT firms with posterior probabilities assisting in allowing algorithms to forecast where the market/stock will move next. In its simplest form the equation for Bayes is below:

\[
P(A | B) = \frac{P(B | A)P(A)}{P(B)}
\]

The probability of \( A \) occurring given \( B \) has occurred is equal to the probability of \( B \) given \( A \) multiplied by the probability of \( A \) and divided by the probability of \( B \).

The best expectation of a future price is given by the current price; the expected value of an asset can then be broadly priced as \( t+1 \). Whilst the exact future price may not be known the use of sophisticated stochastics allows a set of probabilities to create outcomes with a percentage chance of occurrence attached. The current price of stock thus acts as what Bayes would call a prior “belief”. As continuous market data and (non) execution information feeds into the algorithm evidence for or against the belief is accounted for. As a result, this form of genetic algorithm is evidence based and can adapt to today’s complex and dynamic marketplace. Activity in dark pools where quote to trade ratios are thought to be much higher than the lit market acts as evidence (positive or negative) that can re-enforce belief and provide more confidence to an algorithm looking to uncover buy side order parameters. Despite its extensive use it is still not well understood by those actually using algorithms which deploy Bayes as well as the consequences of trading with more advanced counterparts that do.

A further area of insight is from Chaos Theory\textsuperscript{16}. Chaos is often thought to mean that no underlying form or systematic arrangement exists – however that is not the case. Chaos in fact is a state where a process appears to proceed by chance but is in fact determined by a set of precise laws; for example climate change.

\textsuperscript{12} QSG, formally known as Quantitative Services Group has recently been purchased by data provider, Markit and been re-named Markit TCA.

\textsuperscript{13} QSG (now Markit). Beware the VWAP Trap , November 2009

\textsuperscript{14} VWAP=Volume weighted Average Price, PWP=Participation Weighted Price, IS = Implementation Shortfall.

\textsuperscript{15} Palomares, 1966.

is deemed “chaotic”. The random walk that underpins the Efficient Market Hypothesis (EMH) deems that the market is indeed both random and chaotic and so it appears so. Randomness can equate to a broader absence of determinism and yet chaos is itself almost deterministic. If the market is chaotic and yet appears random, then chaos theory can be deployed to unpack the deterministic reality of the chaos from the appearance of randomness. We shall not go into detail on this now but it is important to understand the concept of signal generation and signal recognition in dynamic, chaotic markets.

Institutional Impacts

Examining market microstructure is essential to dig deeper into the causes of return slippage through excess transaction costs for parent and child orders. The myriad of (some) academic and (often) broker supported research tends to argue that things have never been better for the investor. Whilst this may be true from a technological standpoint leading to buy side trader empowerment for their orders, identifying the risks in execution has changed from human information leakage, to something far more computational in nature. Such is the structural change that although many years of experience prior to algorithmic trading is valuable, it is no surprise that some of the larger asset managers are now employing market structure experts to take on newly defined roles either in market structure itself or running electronic trading within the centralised dealing desk and reporting to the head of trading to counter any possible skill gap.

Part of that role is the measurement of market impact, not arbitrary benchmark comparisons, which can be vital when measuring frictional cost. This process consists of isolating the effect of the trader being in the market, which can be undertaken when the effect of trade reversion is analysed after each order is completed and therefore when the execution signal vanishes. A simplistic example is the correct use of effective and realised spreads to discount the reversion. Predatory trading interacts at the child order level where buy side hit rates are reduced due to quote cancellations on various venues leaving an imbalanced footprint. This is coupled with the accumulated trade level where poorly constructed algorithms that generally tend to force outcomes, generate patterns that are relatively simple for genetic algorithms to interpret. Both of these child and accumulated level impacts should be analysed by the buy side, preferably where the TCA and algo provider are separate firms.

Review of Global Insight

Adding problems to one of the central troubles for buy side traders today is the need for resources, primarily in order to genuinely understand what is and is not happening in the world in which they trade. For many firms analytical skills remain scarce and in some firms there remains a question as to whether the buy side dealing desk is a cost centre or an alpha saver/generator supportive of the investment process thus creating more of a troublesome equation for future specialised headcount. Despite its importance to execution costs obtaining transparency on market structure issues often becomes a secondary or even tertiary research operation with firms relying on brokers and to some extent consultants, vendors and even academics for more macro insights. This in itself can sometimes represent a problem.

The computational world of HFT requires a more numerical, academic and objective research driven approach as opposed to a compliance or high touch background or even an “electronic” trading skill set. In economics a form of unemployment exists called “structural”, part of structural unemployment revolves around a mismatch between the employment demands of the industry and the skill sets of the available workforce; technological change is often cited as a key cause. Electronic and algorithmic trading is part way through something of a related pattern where the buy side struggle to consume reliable, valid information from client facing staff with appropriate skill sets for modern day algorithmic trading. Although in flux, many deployed skills still revolve around political skills, subjectivity of opinion versus objectivity and fact, and pure “salespersonship” and personality. Whilst these are a short term positive to the sell side, long term this leads to increased challenges for the buy side. With this in mind the ability for example to academically critique, both positive and negative is essential, however it is a skill in short supply; the signal to noise ratio is weak in this area and with high levels of noise and opinion there are few valuable, reliable signals for the buy side to build on. This structural “employment” deficiency can, over time, weaken the competitive positioning of a buy side dealing desk.
Aside from these more macro personnel issues another area within the topic that causes concern is what happens to an order once it leaves the firewall of the buy side firm. Whilst there has been very little specific academic or regulatory based research on this topic an insightful report by ASIC\textsuperscript{17} was released recently that looked into this area.

The task force found some genuinely alarming findings. Whilst many in the industry have talked about some of the potential nuances and pitfalls of dark trading, for a regulator to come out with such a series of statements should be a wake-up call to the buy side to improve monitoring, due diligence and research of the sell side and market structure.

One of the most common general selling points of a dark pool particularly a crossing system is that it is “natural liquidity” which is often taken to mean buy side to buy side flow – the reality is somewhat different. The UK recently had a “horse meat scandal” where burgers labelled as “beef” burgers were sold that contained percentages of horse meat – some in the trading industry have noted, perhaps somewhat provocatively, that this event contains parallels on disclosure to the world of dark trading. As an insight into possible disclosure problems ASIC noted the below in figure 2:

![Figure 2 ASIC 331 Report on Dark, Q1 2013 - Quotes](image)

"Many crossing system operators have described their crossing system (s) to fund managers and to us (ASIC) as providing "natural liquidity" or as having no high frequency trading….however some crossing systems allow, or have previously allowed access to their crossing systems by clients that the industry widely considers to be high-frequency traders while maintaining there is no high frequency trading in their crossing system" (paragraph 226a, pg 59).

"Seven crossing system operators simultaneously display on a lit exchange market at least some element of orders within the crossing system" (paragraph 174, pg 49).

"It appears that one or more crossing system operators may be offering specific order types to an exclusive subset of their clients and advising these clients how to benefit from these order types" (paragraph 230, pg 60).

These findings should alarm buy side firms. It remains to be seen how firms will react to this report from ASIC; it is intriguing that ASIC stopped short of fining the firms (at least there is no mention in the report or to date) instead choosing to comment and pass new market integrity guidelines post the uncovering. This perhaps implies a gap in the regulation where regulators are “playing catch up”, albeit in ASIC’s case with some success; it is an area that should be monitored in the future.

The second quote in figure 2 refers to a technique called “shadowing” – where firms take orders that are meant to be “dark” and map them onto lit venues presumably looking for a faster fill. Given a large part of the rational for trading in the dark is the reduction in signal risk as part of a wider term called information leakage this offering should be fully disclosed to clients. ASIC’s concern here appears to be that clients were not informed about the existence of shadowing as opposed to the actual technique itself.

No regulator to date has produced such a similar level set of results from such a study into the practical aspects of dark and we would encourage regulators to undertake similar investigations in other jurisdictions around the world and publicize accordingly.

Away from the ASIC research and more generally speaking, probably the single biggest issue around trading behaviours whether it be in the dark, HFT or HFT in the dark or indeed any questionable behaviour from participants is the problem around transparency, data quality and data scope. In Europe this area is compounded with several national regulators in charge of surveillance struggling to have a genuinely pan European view due to, \textit{inter alia}, stocks issued in one regulatory jurisdiction being traded in others, and this is aside from the transparency issues of dark\textsuperscript{18}.

\textsuperscript{17} Australian Securities and Investment Commission, 2013. \textit{REP 331 Dark Liquidity and High-frequency trading}. Sydney: ASIC

\textsuperscript{18} A separate discussion is occurring in Europe as to whether broker BCN’s should report with a defined MIC (Market Identifier Code) code instead of as a blanket OTC. MTF Dark do report with an identifier –however the true state of total dark (MTF+S+OTC(BCN)) in Europe is reliant on broker
A further key point in the discussion is the size characteristics of dark trading. Some have talked about a threshold of 10% being appropriate for dark trading whereby any breach of that threshold is considered detrimental to the price formation process. Indeed both Europe and Australia have discussed thresholds for dark and potentially minimum size requirements. It is also widely acknowledged that the Irish proposal might prevent the so-called “reference price waiver” from being removed in the MiFID 2 process which would have a highly material impact on the majority of dark venues, MTF\textsuperscript{19}, SI or BCN (OTC) by not allowing those books to trade in small size. Size in the dark and the relationship with HFT is an area where very limited academic research has taken place in large part due to the (somewhat predictable) troubles of data quality on both HFT and dark. Many in Europe focus on the notion that the word “dark” has become all encompassing of a variety of constituent types of dark trading which has led to an assumed inflated figure of dark. In 2011 AFME\textsuperscript{20} made an attempt to cover this area by seeking to differentiate between executable and non-executable liquidity; although well publicized throughout the industry its impact on decision makers in Brussels seemed muted. However, there is a side to the discussion not yet fully covered.

The closest work we have to plugging this lacuna is from the University of Melbourne by Carole Comerton-Forde who sensibly splits dark trade into two, block and non-block. This provides a filter on how to view dark trading. The paper carried over well known features of block trading namely that block and above in size does not harm price formation on the lit markets and should be encouraged. However, Professor Comerton-Forde looked more deeply into the role and rise of non-block trading in the Australian market. She finds that the 10% number has validity where if a stock is traded at 10% of its market share or above in non-block dark trades this is harmful to price discovery.

Clearly, simply being defined as a non-block trade has next to no correlation to being an HFT trade and that should not be assumed; although Professor Comerton-Forde did not seek to focus on this relationship this is where the research could be continued further to look into the relationship between the role of HFT in the dark and price formation in the lit as well as the role of non-block broker algorithmic flow in the dark and the relationship to lit price formation. This latter point is an important distinction as although many in the market have agreed that HFT is a subset of the wider “algorithmic trading”, the results of some papers still get interchanged between the two terms. Any discovered problems of HFT can thus be attached to all algorithmic trading and vice versa where the benefits of algorithmic trading (and more simple “electronic”) can be attached to HFT. Clearly, neither are appropriate and it is important to watch for the interplay of these words in research reports.

This brings us into the UK Foresight report which attempted to look into a range of areas of algorithmic trading including HFT and a less than clear term, computer based trading (CBT). Unfortunately, as with the above, the interchanging of these three terms in the final report gives cause for concern particularly when definitions are not clear cut (including the appendix). Whilst not the fault of the academics involved in the underlying Driver Reviews (DR’s) and Economic Impact Assessments (EIA’s) the data used across the 31 DR’s and 22 EIA’s was not completely sufficient. For example, papers struggled to contain population sets with appropriately time stamped trade data across the full spectrum of primary exchange, MTF and all types of dark activity (BCN (OTC), SI and MTF); there were also problems with a split between direct and indirect tracking of HFT i.e. not tracking trades directly from HFT which implied certain assumptions that may not fully hold. The unfortunate incomplete nature of the underlying data opened questions as to whether the work, whilst helpful to the discussion through creating debate is, when taken on balance, actually of appropriate quality to be used for setting policy use even though parts may be useful as a guide to some.

\textsuperscript{19} MTF = Multilateral Trading Facility, SI = Systematic Internaliser, BCN = Broker Crossing Network.

\textsuperscript{20} The Association for Financial Markets in Europe.
That said, one of the more interesting diagrams out of the DR’s\(^2\) shows the execution cost\(^2\) in FTSE 250 names from 2003 to 2011 (page 12 of the DR21) complete with moving average on a yearly basis – see Figure 3 below. Many in the industry would agree that commission costs have decreased during this time frame, however questions remain over market impact costs which are the materially larger component of trading costs. These questions are in part due to the complexity of market fragmentation, dark transparency and HFT trading that have made many pre-existing TCA benchmarks highly dubious in their ability to capture the full picture of cost assessment and alpha loss\(^3\). Perhaps counter-intuitively to some, Figure 3 below shows that one of the largest declines in execution cost over the eight year data period actually occurred in March 2003 to March 2004 a period not normally associated with high levels of HFT activity. In fact a bar chart earlier in the DR emphasises that point (page 14 of the study and Figure 4 below) showing the upward trend of HFT activity but only starting in 2006, most would agree HFT, in 2003, is below the 2006 level of approximately 4%.

It is also challenging to say if execution costs have dropped over the whole period of study, whilst the last reading is below the first and the moving average shows a downward trend if we look closer a more interesting pattern emerges. The author has removed the credit crises years (see figure 3 right hand side graph below) which would show an uptick in execution costs (HFT activity also rose in this period according to Figure 4), unfortunately by doing this it could be seen as a rather rushed approach and we would argue that more diligence should have been applied to strip out this and other exogenous factors from the data set rather than deploying a method which does not show the full picture as a result. Given the removal of data showing an upside move in execution costs (which smooths the average downwards) it may have been more sensible to also remove the aggressive spike in the data that also pushed the cost trend downwards in Q12006. By removing data that pushed the costs up and thus the average down and keeping spike data that also pushed the average down, the data is not as balanced as we would prefer. No reason was given as to why this (or indeed other spikes) pushing the trend downward remained in the study. That aside, if we maintain use of the “stripped” version of the graph (right hand side Figure 3) we can see that if we bring the start of the study forward by just four data points and bring the end of the study backwards by just six data points the execution cost actually increases over the period.

![Figure 3](image)

![Figure 4](image)

---


It is also worth noting that with HFT at ~4% of the market in 2006, the year of the large drop it seems that on balance there are exogenous factors impacting execution cost alteration and the relationship with HFT is not strong – this supports the notion that significantly more research should be conducted in this area.

Some would note that HFT might not be widely present within the full range of the FTSE 250 during the study period but are rather focused on the most liquid names. Figure 5 shows the breakdown by market cap across the FTSE 100 names split by the category on a numerical ranking, 1-10 being the largest market cap and 51-100 being the smallest.

Figure 5 Impact of HFT on Execution Cost According to Foresight Driver Review 21 for FTSE 100 Split by Market Cap Category, Largest =1.

The pattern with the overall FTSE 250 is broadly repeated here for the FTSE 100 with general cost inconsistency versus HFT growth throughout. In the highest cap category, FTSE 1-10, there are material drops in execution costs between March 2003 and March 2004 as well as mid 2005 to March 2007. A simple cross reference to figure 4 shows that HFT in this time was extremely small, approximately 4% in 2006 and 8% in 2007 – the relationship between HFT growth and decreasing transaction cost is not clear cut in the top market cap names. Post March 2007 costs are more volatile moving both up and down in large swings again distorting the relationship with HFT, although this period is more likely to be caused by the wider financial crises, there is a large upward move at March 2011 where HFT decreases only marginally by approximately 0.5% to ~36% (according to Figure 4) supporting the notion that other factors may be in play.

The 11-30 range also shows large drops in the “pre HFT” environment (2003 and 2004), and is broadly flat if slightly rising between the mid of 2005 and 2008, a period where HFT is rising, however again this remains insufficient to draw conclusive results.

Both the 31-50 and 51-100 remain largely flat throughout the period indicating next to no conclusive impact by HFT.

The authors of the research do attempt a two stage least square regression (2SLS) to the data and apply an instrument variable. Unfortunately, the first stage of HFT activity is locked at the day level and material percentages of HFT activity were not captured in the study. 2SLS has problems in its use and there are assumptions to the technique that require caution when viewing the results. To corroborate the coverage of HFT within the data set the authors compare four years of HFT data provided by the UK FSA (now the UK FCA) to just 30 days of data in 2010 offered by The LSE, Chi-X and BATS. There is no rational provided as to why such a short time frame was used for the corroboration data, which days were used or why they were chosen. To co unfortunate that of HFT activity is locked at the day level and barely approximately 30% of HFT activity was not captured.
Overall, as can be seen from the evidence provided in the Foresight paper, it is not clear cut that HFT reduces transaction costs for the institutional investor. This is supported by the authors when they note “we find no clear evidence...these results should be interpreted with some caution” (page 25). This is of no surprise given the difficulties on the data and methodology. Relating this study directly to dark trading would be a logical next step and one we look forward to reading, albeit preferably with an improved data set.

Outside of the study many HFT firms have consistently argued that HFT as a group have reduced execution costs however whilst we acknowledge that say Getco24, a now large HFT, was live in Europe (according to the FSA register) in mid December 200325 it is unlikely that Getco could (even combined with other HFT firms) have played a material role in total market wide execution cost decline in the interval of March 2003 to March 2004 as per Figure 3 – furthermore, if the above relationship was accurate at strong confidence levels, we would expect to see a clearer correlation to increased HFT and declining execution costs than we do in the referenced paper.

HFT as firms sit the other side of the trade for the buy side and are not widely providing algorithms to the buy side (during the time period of study) as such any decrease in execution costs from HFT’s role could come from the market impact side of the total cost, an area where answers are far from clear cut. What is clearer though is that, over the last decade, electronic trading as a whole, the creation of the FIX protocol and the, now simple, use of the algorithms by the buy side to slice large orders into smaller orders to reduce market impact has played a substantial role in reducing certain execution costs, both direct and indirect.

ASIC provides some insight into perhaps one of the structural trends in trading that is often overlooked. “Our analysis suggests that high-frequency trading DOES NOT26 have a significant effect on price formation, liquidity and execution costs, and that systemic factors, including the wider adoption of automated trading technology are more significant”27. Although relevant to a period of study later than the aforementioned Foresight work, it seems plausible that given the wider adoption of electronic trading in the early to mid 2000’s the decline in execution costs could in part be related to that impact as opposed to HFT. It is important to remember that investment banks and brokers chose to provide algorithms to institutional investors cutting commission rates and market impact in the process, it was not a process widely recognised as being driven by HFT firms nor did HFT firms play a direct role in providing competition to reduce broker commission costs as HFT firms, at that time, traded off their proprietary capital and had no clients. ASIC raise a point that should not be forgotten when looking at historical trends and noting any possible interrelationships between variables: clearly more macro non market structure impacts exist and these need to be accounted for more accurately before passing judgement for policy use. Perhaps former SEC Chairman Mary Schapiro said it best when she recently noted in an interview with Australian Financial Review “we lack really solid analysis and data that would tell us whether the quality of our markets have been negatively impacted by high frequency trading” – we would add that the industry lacks solid data and analysis establishing whether our markets have also been positively impacted by high frequency trading – the problem of data and methodology is present in both directions.

Identification of Market Impact and Evidence of Adverse Selection

Attempts to detect microstructure arbitrage involves the analysis of execution venues where identifying the skew of the distribution of returns around the “child” order can isolate venues that provide poorer fill qualities. Whilst the reasons for skew may vary, execution analysis at this level allows the buy side trader to attach a probability of adverse selection to a pool, and the trader can then (de)select the most (un)successful venue. If the institutional trader is more informed, adverse selection risk is asymmetric and counterparty selection reduces in importance.

---

24 Getco is now part of KCG Holdings Group due its merger Knight Capital Group.


26 Author added emphasis.

As more trading moves into dark pools\(^{28}\), such venue analysis that attempts to identify adverse selection is especially useful. Arbitrage thrives in dark pools where the opacity of the matching process and routing strategies (inbound IOI\(^{29}\) reactions and outbound IOI/blind IOC solicitations for example) to varying types of counterparties can open up participants to adverse selection risk. Obtaining information with regards to timestamps and latency hops is difficult, especially in a labyrinth of broker cross aggregation; as a result often analysing outputs is the most effective way of inferring embedded risks in a venue\(^{30}\).

It is often said amongst buy side traders that “dark is really lit” indicating that dark pools and internalisation engines that were designed to protect clients from information leakage, seem to be compromised by the level of information gathered by firms canvassing order flow across venues. HFT tends to look for validation of a lit market signal with information in dark pools to determine demand and supply imbalances and vice versa. The ability to post two sided markets in multiple venues is a huge advantage that the buy side does not have.

Latency arbitrage profits are possible where there exists a latency differential between the matching prices, market data feeds or size of orders at various venues. This is not a new phenomenon however fragmentation of liquidity has expanded opportunities from this form of trading. For the purpose of this paper, we will focus on dark pools hence measuring possible partial profitability of latency between the reference price, being the exchange, and the matching price at the dark pool.

Through the set of analytics designed by Schroders Investment Management Australia Limited (referred to as “SIMAL” for the remainder of the paper) to measure adverse selection risk, the footprint of arbitrage is observed on dark pools in the Australian market place. The model was constructed to analyse all dark pool executions through the specific dark only algorithm also designed by SIMAL. The results showed repeated systemic losses in some specific pools despite homogenous routing logic; however, a range of outcomes were also observed.

Considering the number of observations, the diversity of orders with varying alpha characteristics and the extremely short duration of analysis around individual executions, we expect to achieve a normal, or at least a symmetrical distribution. The SIMAL dark algorithm utilised mid-point limit orders for all but one pool, minimum acceptable quantity (MAQ) where available and also utilised counterparty selection when offered by removing any principle, market maker or “HFT”\(^{31}\) flow from the counterparty list. Given the dataset was dated to before the release of the Australian price improvement rule; we expect these additional conditions to have a positive effect on the quality of the distribution versus a dataset of dark orders posted at market. We have included a distribution of a venue where market orders were used for comparison.

By measurement of the executed price against the Volume Weighted Average Price (VWAP)\(^{32}\) of trades immediately following the execution and normalising for the spread for every child order execution, we can measure if there are any systemic losses incurred\(^{33}\).

\(^{28}\) Subject to any future regulatory based restrictions.

\(^{29}\) An Indication of Interest (IOI) from one participant to another whilst an Immediate or Cancel (IOC) order is a firm order for a quantity at a price that is immediate in nature and often wont “rest” on the venue. It is usually used to “ping” a dark venue to test it for the existence of orders. Outward blind IOC is often an order type received by HFT run dark pools.

\(^{30}\) Having said that it is helpful that the FPL Investment Management Working Group are coordinating globally to work on a project to track where an order has been routed before it sends back the TAG 30 last market execution field.

\(^{31}\) There are definitional differences of HFT across brokers. Some brokers do not include their own internal market maker unit as HFT which introduces the need for the buy-side to understand exactly what to exclude.

\(^{32}\) VWAP is defined by \[
\frac{\sum_{j=1}^{n+1} P_j Q_j}{\sum_{j=1}^{n+1} Q_j} \]
where \(n\) is defined by the number of trades used in the analysis after trade \(j\).

\(^{33}\) Just as most academic research is unable to identify HFT in a data set, there are no identifiers in the data set used that flags an HFT counterparty. The study is simply measuring the systemic losses whether those losses are to HFT or not is not clear.
The formula can be written as:

\[
\text{Child Order Performance} = \frac{(P_f)}{\text{VWAP}_{i+1}} - 1
\]

Where \( P \) is the price of child order execution \( j \), \( \text{VWAP} \) is measured over the five immediate trades following the execution and \( S \) is the spread at the time of the execution.

The x-axis on the distributions below is defined by the loss or gain of an individual child order as a ratio of the average spread. The y-axis is defined by the number of observations in each 0.01 spread bucket.

The results varied significantly by dark book, which continuously leads to a recalibration of the dark routing logic.

Figure 6 below is similar to what we would expect from a distribution of returns over a large sample set. Whilst a noisy yet symmetrical distribution, most executions are at a price close to the immediate observed trades after that execution.

Figure 6  Expected Midpoint order Display

Figure 7 shows what happens when we engage a specific venue whilst utilising dark orders at market prices. Whilst we do capture the spread in some trades, there are more executions that incur full spread costs than any other increment, as well as a larger number of smaller losses between half a spread and 0. It is not surprising that our number of executions is higher in this venue.

Figure 7  Expected Market Orders Display

\[34 \text{ When displaying the distributions, we exclude all executions at 0 bps performance to focus on executions that differ from expected.}\]

\[35 \text{ Psomadelis. W.; Schroders proprietary research on dark pool venue toxicity, 2013.}\]
Figure 8 is surprising. This distribution exhibits a somewhat systemic underperformance of the executed price to the immediate post execution trades by a quarter to half a spread indicating we are generally trading at inferior prices\textsuperscript{36}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure8.png}
\caption{Deviation Venue on Midpoint Orders}
\end{figure}

Latency is not standardised by regulation or by formal industry agreement, these results highlight both the potential increase in frictional costs when blindly resting liquidity in broker dark pools that should be monitored. This research shows that measurement is necessary when utilising dark venues in the overall buy side strategy.

Identifying the accumulated impact on a parent order allows us visibility into the impact of statistical signalling. The use of relative performance measurement at the individual stock level is vital to isolate idiosyncratic risk and once this has been determined, the reversion number over various time periods is a proxy for the amount of leakage that may otherwise have not incurred.

Mathematically this is given by the equation below and graphically by figure 9.

\[ K(t) = f(t) - g(t) \] \hspace{1cm} \text{\textsuperscript{37}}.

Where market impact is the disturbance \( K \) in the stock price given \( f \) (the price move with the order) and \( g \) (the price move without the order).

Figure 9 shows an example of temporary market impact where a stock with upwards price momentum displays significant price disturbance from Order Start \( S \) to Completion \( C \) as evidenced by a reversion back to the original trajectory after order completion. The original expected trajectory is mapped by the dotted black line “\( g \)” whilst the actual price move is the blue line “\( f \)”.

Point \( P \) denotes the separation of temporary and permanent impact, which is usually set as either a defined time period from order completion or as a percentage of order volume traded post completion for more sophisticated analysis.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure9.png}
\caption{Stock Price Highlighting Market Impact and Reversion}
\end{figure}

\textsuperscript{36} An interesting observation is that when these same distributions were run for trades after the implementation of the price improvement rule and after the ASIC paper on dark trading, the distributions look slightly closer to symmetrical for some, but not all, venues. There are not enough observations in the, revised post “price improvement” data set as yet to produce clear conclusions, so we have excluded them for this paper.

\textsuperscript{37} Kissel & Glantz, Optimal Trading Strategies, 2003.
The study of the effect of microstructure on trading is focused on identifying the amount of disturbance due to trading forms such as microstructure arbitrage and high frequency quant. Obviously, the two types of HFT discussed aren’t responsible for all forms of market impact, however understanding these concepts would form part of the required preservation/recapture of alpha within the institutional trading space.

The introduction and adoption of electronic trading by institutional buy side desks has been a major contributor to the reduction of execution costs for institutions over the past ten years and by proper reversion analysis, the impact of excessive market impact can be isolated and further reductions realized.

**Australian Lit Market Order Cancellation Study**

In December 2012, SIMAL and CIMB Securities Asia implemented a model to attempt to understand the nature and possible effects of order book cancellations. High frequency data using full order book information was captured and two models were created to identify the two distinct cancellation events and their subsequent impact. Those two models were the Partial Execution then Cancel Method (PEC) and the Trade Cancel to New order Ratio method (TCN).

**PEC Study**

Criteria for PEC trades included orders that were cancelled within a second of passive partial execution. To measure slippage, we compare the cancelled price to the side adjusted VWAP of a pre-determined number of ticks. The cancellation quantities are normalised for average trade sizes in the stock on that given day.

To the investors reading this paper, having a passive order executed is the ultimate goal of posting liquidity in the first place so cancelling an order as it is being executed seems counter-intuitive. Two possible reasons for systemic cancellations are:

- The firm responsible for the order cancellation was in receipt of hard information that created a brief period of information asymmetry and was therefore mitigating their adverse selection risk from more informed traders.

OR

- The orders were placed without the intention of executing, but for the intention of gathering an information signal.

Figure 10 represents the findings of the deviation between the price of the cancelled order and the side adjusted VWAP of trades immediately after the cancellation (x-axis), against the number of observations per slippage bucket (y-axis). A positive slippage number in this chart represents a market move in the direction of the cancellation, e.g. a price fall when a bid was cancelled after being partially executed. Whilst the authors of the research agree that the data is noisy, there exists a clear skew towards positive returns in the density function (mean 1.1 bps with an interquartile range of 5.6 bps) indicating that observing order cancellations can be a predictor of short-term price moves. Whilst testing suggests that this is not statistically significant, the result considering the amount of noise in the data is interesting and as long as the trader can observe the cancellation sequence and respond fast enough, it appears a genuine trading signal is attached to cancellation events.

---

38 Model implementation was collaboration between Schroders Australia and CIMB Securities Asia. The ongoing analysis, maintenance and reporting is undertaken by CIMB Asia with thanks to Richard Knight, Andrew Freyre-Sanders and Catherine Turley.

39 Huh; Machines vs Machines,. *High frequency Trading and Hard Information*. Dec 2012

TCN Study

The SIMAL and CIMB Securities Asia model also investigated the Trade, Cancellation and New (TCN) ratio. TCN events consider the aggregate ratio of cancelled orders to new order placements around trades on the passive side of the book (we measure bid cancellations if a trade was sell initiated for example). New order placements and slippage are measured in two distinct periods. The model determines the half-life as the period where the market was reacting to the trade and thus the period in which the TCN is calculated (Time $T_0 \rightarrow T_{1/2}$). Slippage is calculated on the residual period ($T_{1/2} \rightarrow T_{end}$). For example, BHP.AX experiences a very fast reaction up to 100ms, where we calculate the TCN ratio with a fade of activity out to 500ms where the returns are measured.

Figure 11 shows that there is a positive relationship between the normalised TCN ratio ($x$-axis) and slippage ($y$-axis) indicating that identification of high order cancellation periods might be a valid predictor of short-term price moves. It is worthwhile to note that this effect is not uniform across all stocks. It is also worth noting that the maximum of 2 bps slippage means that the decision to co-locate to take advantage of this phenomenon is one that must be subject to significant analysis.

Figure 12 shows a curious existence of asymmetry between favourable and adverse price movements. The $x$-axis represents the TCN ratio for each data set, whilst the $y$-axis represents the number of observations per TCN bucket. The aggregate TCN results would be surprising to the traditional investor who might expect that a favourable temporary price move might induce the provision of more liquidity, or at least not be met with increased order book cancellations. The fact the opposite occurs is surprising.\(^{41}\)

\(^{41}\) The authors of this paper are not in favour of a minimum resting period for orders and in no way does the SIMAL/CIMB research present any case for or against it. The research paper does not study short duration orders in isolation and therefore is more focussed on the impact of cancelled orders, on short term prices. The SIMAL/CIMB research shows that there is a positive relationship between the size and number of cancelled orders.
Conclusion

The central points of this paper revolve around the obvious need for buy side firms to improve monitoring of trade performance and understanding of market microstructure through in-house research. The complexities of the modern trading landscape are profound with structural skill requirements in flux in all parts of the industry. Although some regulation to slow complexity may be probable in the future, by default as the market fragments complexities increase. Many practitioners call for choice and yet some of those same people complain about market complexity – perhaps ironically, the opening of choice will nearly always lead to more complexity – improved research whilst costly will help solve the complexity of the market whilst still allowing the competitive forces of choice to exist. The equation between cost and reward is one that the buy side needs to assess and the calculation of that equation will no doubt differ between firms.

Today's dark trading as a whole has rarely been looked into in academic circles and order book dynamics are obviously very difficult to ascertain giving the opaqueness and non-disclosure of information and data. The inability of the buy side and academia to track HFT directly is a clear impediment to HFT related research, which in turn is compounded in the dark pool context. Studies such as those in this paper that observe actual execution outcomes from primary data offer an insight to systemic losses in dark pools and order book changes based on low latency effects.

Consultants and brokers will also struggle to provide a comprehensive view on dark given their lack of access to data often relying on “contributed” data from the brokers who run the pool rather than actual raw trade data. As such, perhaps those best to review are the buy side themselves who have access to most broker dark pools; the question though is do they, in fact can they or indeed should they…that answer is not one for us to make but is probably a case of “each to their own”.

Disclaimer

Opinions, estimates and projections in this article constitute the current judgement of the author as of the date of this article. They do not necessarily reflect the opinions of Schroder Investment Management Australia Limited, ABN 22 000 443 274, AFS Licence 226473 (“Schroders”) or any member of the Schroders Group and are subject to change without notice. In preparing this document, we have relied upon and assumed, without independent verification, the accuracy and completeness of all information available from public sources or which was otherwise reviewed by us. Schroders does not give any warranty as to the accuracy, reliability or completeness of information which is contained in this article. Except insofar as liability under any statute cannot be excluded, Schroders and its directors, employees, consultants or any company in the Schroders Group do not accept any liability (whether arising in contract, in tort or negligence or otherwise) for any error or omission in this article or for any resulting loss or damage (whether direct, indirect, consequential or otherwise) suffered by the recipient of this article or any other person. This document does not contain, and should not be relied on as containing any investment, accounting, legal or tax advice.

and price movements for the dataset studied and the authors of this paper are supportive of ASIC’s efforts with regard cost recovery programme where market surveillance fees are a function of both trades and messages.