Government Office for Science

Foresight

High frequency trading and price efficiency

The Future of Computer Trading in Financial Markets - Foresight Driver Review – DR 12

Contents

High frequency trading and price efficiency

Terrence Hendershott
3 August, 2011

This review has been commissioned as part of the UK Government's Foresight Project, The Future of Computer Trading in Financial Markets. The views expressed do not represent the policy of any Government or organisation.

Executive summary

This review examines high frequency trading (HFT) and price efficiency. Technological advances typically replace the most repetitive human activity with automation. For short-run speculation such as HFT it involves processing as much data as quickly as possible to find predictability in future price changes and then to capitalize quickly by submitting orders. As technology improves the trend of substituting technology for labor will continue in the investment process. This review focuses on how greater investment in computing and communications technologies impact HFT and price efficiency.

HFT strategies are discussed along with their differing impacts on price efficiency. The relatively sparse empirical literature on HFT and price efficiency is examined together with its conclusion that HFT appears to improve price efficiency. In the context of the future impact of HFT on price efficiency four broad questions and their impact are discussed along with the italicized conclusions below:

1) Is HFT competitive?
Standardizing market access and data and restricting the benefits of very small increases in speed enhances competition.
2) Will HFT-like services be offered to low-frequency investors?
Algorithms for non-HFT already offer tools incorporating HFT approaches; this trend will continue, especially if HFT is competitive.
3) Are automated markets with automated trading less stable, e.g., prices occasionally become quite inefficient, than markets with more human intervention?
While no systematic evidence that HFT reduced stability, the interactions between algorithms may cause instability different from interactions between human.
4) If HFT begins to reduce price efficiency will market structures evolve to limit HFT?
Rules requiring trading in markets with HFT should be re-examined.

The review concludes with the importance of expanding the availability of data on HFT across markets and asset classes. Well structured policy experiments and data availability will enable deeper understanding of HFT and its roles in financial market performance.

Market and price efficiency

More efficient prices in financial markets contribute to better allocation of resources in the economy and more informed financing and investment decisions by firms. Fama (1970) describes an efficient financial market as "one in which prices always fully reflect available information" and defines efficiency on a three-level scale, described in more detail below. Fama's taxonomy introduces the term "information set," which simply means the set of information available to investors when they optimize their portfolios and make trading decisions. The idea of an information set is a key concept when discussing market efficiency and to appreciate how and why trading firms employ technology to maximize and process data to produce that set of information. Fama's three-tier taxonomy is as follows:

1. Weak-Form Efficiency: In weak-form efficiency "the information set is just historical prices" and efficiency is often interpreted as security prices, instantly and fully reflecting all information of the past prices. This implies that past prices cannot be used to predict future price movements.
2. Semi-Strong Efficiency: Under this definition "prices efficiently adjust to other information that is obviously public", which adds other information such as stocks splits, dividends, and public news to the information set, so that asset efficiency implies prices fully reflect all publicly available information.
3. Strong-Form Efficiency: In strong-form efficiency "no investors or groups have monopolist access to any information relevant for price formation." In this case asset prices fully reflect all information, both public and private, meaning that no investor can predict price movements.

In a paper titled "On the Impossibility of Informationally Efficient Markets," Grossman and Stiglitz (1980) demonstrate the inherent contradiction in strong-form efficiency: if obtaining information and trading on it is costly, then no one will gather it if all the information is already in prices. On the other hand, if no one gathers information, then how can it be in prices? Based on this paper and subsequent work few economists believe in strong-form efficiency.

Market efficiency is closely related to the concept of arbitrage, which refers to the practice of taking advantage of price differences between related securities. Initially, most financial economists accepted that financial markets possessed what Tobin (1984) called "information arbitrage efficiency." This concept maintained that markets were informationally efficient in the weak and semi-strong sense: no investor can systematically make money trading on the basis of generally available public information. This implies an absence of arbitrage.

Subsequently, even weak-form and semi-strong-form efficiency were called into question by numerous systematic patterns found in stock returns, initially thought of as anomalies" Some of the first anomalies included the January effect (Keim (1983)) and predictability (positive autocorrelation) in the returns of portfolios of stocks (Lo and MacKinlay (1988)). In particular, Lo and MacKinlay show that while predicting future returns (percentage price changes) of individual securities is difficult, the returns of portfolios of securities are predictable. This predictability arises from information about common factors that affects many securities, for example news about cash flows in a particular industry or changes in discount rates which would affect all securities, being incorporated into some securities before others. Chan (1993) provides an insightful discussion of this phenomenon with a mathematical model.

While financial markets are no longer considered as efficient as once believed, it is hard to understand why investors would not choose to arbitrage the predictable patterns that can be identified. There are many possible reasons, but a natural starting point is the risk/return trade-off that is a cornerstone of finance. In the capital asset pricing model (CAPM), securities with higher returns involve more risk, as no security provides returns above what is warranted by its risk. Risk is referred to as beta (β), the coefficient capturing a security's return relation with market-wide fluctuations. Over time, factors beyond the market-wide risks were added. A security or strategy's return above and beyond its relationship to known risk factors is typically referred to as alpha (α), which is arises from any variable, pattern, or signal that enables the construction of a trading strategy that yields profits/returns beyond its measurable risk.

The risk-adjusted profitability of strategies based on return predictability is still being understood, as it is impossible to be sure that all sources of risk can be identified and measured. For example, strategies designed to arbitrage the cross-stock predictability found by Lo and MacKinlay exhibited little variability and risk for at least a decade after publication (Khandani and Lo (2007)). However, in August 2007, the returns of these strategies exhibited unprecedented huge negative returns for several consecutive days (Khandani and Lo (2011)). The meltdown of Long Term Capital Management provides another example of historically profitable arbitrage strategies becoming suddenly risky when market conditions change. These examples highlight what is often referred to as the limits to arbitrage. It is generally believed that the more investors who choose to follow a particular arbitrage strategy, then less profitable and more risky, the strategy becomes.

If it takes time to identify and capitalize on well-known patterns by building infrastructure and systems to support the necessary trading strategies, then the initial idea may remain profitable for some time even. Khandani and Lo (2007) discuss such implementation issues by describing statistical arbitrage as "highly technical short-term mean-reversion strategies involving large numbers of securities (hundreds to thousands, depending on the amount of risk capital), very short holding periods (measured in days to seconds), and substantial computational, trading, and IT infrastructure." The resemblance between these requirements and HFT will become clear once HFT is defined below. In fact, many HFT strategies are forms of statistical arbitrage.

Beyond risk there are a number of financial market frictions that limit the profitability of known inefficiencies based only on public information. One is simply the cost of trading immediately, often proxied for with commissions and the bid-ask spread, which presents the difference in the prices at which a security can be immediately bought and sold. Other costs include financing and borrowing costs for strategies that involve leverage or short selling. The costs of immediacy and financing can both vary over time because in times of market stress, immediacy may become more costly and lenders make securities for short selling and funds for leverage harder to borrow. These time variations then increase the risk of seemingly very profitable trading strategies. The return patterns during the week of August 6, 2007 are consistent with a number of funds following this strategy having taking significant negative returns and being forced to liquidate their positions at significant losses. Following the realization of those losses the profitability of the strategy reversed on Friday August 10th and ended the week as profitable (Khandani and Lo (2007)). HFT typically involves relatively little capital being invested at any moment in time, so a meltdown like August 2007 is unlikely to arise from HFT. However, it may take years or decades to know if other such events and risks are present in HFT strategies.

The theory behind market efficiency and arbitrage based trading strategies often employs the unobservable fictitious public information set to draw distinctions between public and private information. However, as Hasbrouck (1991, p. 190) writes “the distinction between public and private information is more clearly visible in formal models than in practice.” The implicit assumption is often that all traders are constantly monitoring the market, processing all relevant data, and optimally adjusting their portfolios and orders. When there are thousands of securities whose prices, quotes, and orders books constantly change it is hard to imagine any human keeping up. Those who are better and faster at gathering and processing public data and turning it into information can profitably trade upon information not yet incorporated by other investors. Naturally investors have invested substantial effort in information technology to avoid being taken advantage of in this way. Through this investment more information is incorporated into prices. HFT epitomizes the race to gather, process, and act on as much information as quickly as possible.

High frequency trading

High frequency trading developed recently enough that a common definition remains elusive. The Securities and Exchange Commission’s (SEC’s) 2010 Concept Release on Equity Market Structure (SEC (2010)) describes HFT as employing technology and algorithms to capitalize on very short lived information gleaned from publically available date using sophisticated statistical, econometric, machine learning, and other quantitative techniques. On page 45 the SEC writes:

> One of the most significant market structure developments in recent years is high frequency trading (‘HFT’). The term is relatively new and is not yet clearly defined. It typically is used to refer to professional traders acting in a proprietary capacity that engages in strategies that generate a large number of trades on a daily basis… Other characteristics often attributed to proprietary firms engaged in HFT are: (1) the use of extraordinarily high-speed and sophisticated computer programs for generating, routing, and executing orders; (2) use of co-location services and individual data feeds offered by exchanges and others to minimize network and other types of latencies; (3) very short time-frames for establishing and liquidating positions; (4) the submission of numerous orders that are cancelled shortly after submission; and (5) ending the trading day in as close to a flat position as possible (that is, not carrying significant, unhedged positions over-night).

Generally, the above describes trading that uses substantial technology to actively monitor markets, trade, and manage risk so that large amounts of trading is carried out without incurring significant risk on a HFT firm’s portfolio at any moment in time. The SEC reports that HFT is thought to encompass close to one half of all trading in US equity markets. The SEC continues, stating:

> this release next will briefly discuss four broad types of trading strategies that often are associated with proprietary firms – passive market marking, arbitrage, structural, and directional. The discussion of directional strategies will focus on two directional strategies that may pose particular problems for long-term investors – order anticipation and momentum ignition. The Commission notes that many of the trading strategies discussed below are not new. What is new is the technology that allows proprietary firms to better identify and execute trading strategies.

The SEC does not state what new technology allows proprietary firms to better identify and execute trading strategies. Developments in computing and communications technology over the past 20-30 years have been largely evolutionary with relatively linear rates of percentage improvements. The compounded effect of these improvements is not to fundamentally change the way technology is used, but to lower the barriers to entry and costs of pursuing quantitative trading strategies.

The SEC then goes on to discuss each of the four strategies. On page 48 the SEC describes passive market making:

> [Passive market making] primarily involves the submission of non-marketable resting orders (bids and offers) that provide liquidity to the marketplace at specified prices… the primary sources of profits are from earning the spread by buying at the bid and selling at the offer …

How an investor/trader following a passive marketing making strategy should place these bids and offers has been widely studied in finance in an area referred to as market microstructure. A survey article by Biais, Glosten, and Spatt (2005) provides a general theoretical framework and model that parsimoniously characterizes passive market making. The key decisions for the market maker are what quantities to quote at different prices: the number of shares a market maker is willing to post to firmly commitment to buy or sell at various prices. Mathematically, this is formulated as an optimization problem with the market maker offering a liquidity supply price schedule $q_t(p)$ at time t:

$$\underset{q_t(p)}{\text{Max}}\, EU(C_t + I_t v_t + (v_t - p)q_t(p) \mid H_t),$$

where $U(x)$ is the market maker's utility function, C_t is cash, I_t is the inventory of the security, v_t is the expected fundamental value, and H_t is the information set available to the market maker. The liquidity supply schedule could represent a number of orders/shares submitted to a stock market at prices such that they would not execute immediately. Such orders are referred to as nonmarketable limit orders. The above equation highlights that uncertainty about the value of the security is the primary source of risk. Part of this risk is driven by any information relevant to the value of that security that is not in the market maker's information set. While Biais, Glosten, and Spatt (2005) provide a simple one security model, the basic approach can be extended to a portfolio of securities (Ho and Stoll (1983)). More generally, as with most economic models, the Biais, Glosten, and Spatt model captures economic forces and how they affect decision making without intending to directly correspond to the real world.

High frequency traders use algorithms to lower the cost of calculating and managing this price schedule (quotes) in the presence of constant changes in the information set and to incorporate the amount of risk associated with each security in their portfolio. Any time an announcement, trade, quote change, submission, or order cancellation occurs in any security in any market, the information set of a market maker changes and his prices and quantities must be revised. On page 49 the SEC describes this process:

> If the proprietary firm is layering the book with multiple bids and offers at different prices and sizes, this strategy can generate an enormous volume of orders and high cancellation rates of 90% of more. The orders also may have an extremely short duration before they are cancelled if not executed, often of a second or less.

If a high frequency trader is too slow in updating its information set and, a faster trader who has more up to date information can pick off the market maker's not yet updated quote. Such quotes are referred to as stale quotes. For example, if the S&P 500 index moves upwards or is expected to move upwards then a market maker likely wants to revise its quotes in every security upwards. If another trader can execute against the market maker's stale quotes prior to revision, then that trader profits and the market maker loses. That concern about stale quotes along with a rapid decline of the S&P 500 index prices, as happened on May 6, 2010, may lead liquidity providers to quickly reprice or cancel their orders, causing sudden illiquidity. Difficulties on May 6, 2010 with market data, uncertainties about Reg NMS linkages, and lack of cross-market coordination of stabilization mechanisms such as trading halts make it impossible to determine the source of the illiquidity that day.

On page 51 the SEC moves on to the second HFT strategy of arbitrage:

> An arbitrage strategy seeks to capture pricing inefficiencies between related products or markets. For example, the strategy may seek to identify discrepancies between the price of an ETF and the underlying basket of stocks and buy (sell) the ETF and simultaneously sell (buy) the underlying basket to capture the price difference. Many of the trades necessary to execute an arbitrage strategy are likely to involve taking liquidity, in contrast to the passive market making strategy that primarily involves providing liquidity. In this respect, it is quite possible for a proprietary firm using an arbitrage strategy to trade with a proprietary firm using a passive market making strategy, and for both firms to end up profiting from the trade. Arbitrage strategies also generally will involve positions that are substantially hedged across different products or markets, though the hedged positions may last for several days or more.

This arbitrage strategy shares many components of a passive market making strategy: both must carefully manage risk and monitor market data to constantly update their information and revise their orders and portfolios. If HFT passive market making strategies only trade with HFT arbitrage strategies in exactly the same assets at the same time any profit for one strategy must represent a loss for the other strategy. Therefore, whatever information HFT arbitrage strategies attempt to capitalize on, HFT market making strategies must also try to identify to avoid losing money.

A main difference between HFT market making and arbitrage strategies is how they manage risk. Arbitrage strategies typically attempt to hedge risk immediately by transacting in related securities simultaneously: the exchange traded fund (ETF) and underlying securities in the ETF in the above SEC example. Pure market making strategies are passive and cannot control when they trade so they manage risk by adjusting their bid and offer prices to attempt to mean revert their position of a security to its desired level. For example, if a market maker owns 100,000 shares of Intel and wants its position to be zero shares, the market maker would induce other investors to buy from it by adjusting both its bid and offer prices downward. When other investors buy from the market maker its position moves closer to zero and is less risky. As this happens the market maker adjusts its quotes upwards so they more evenly straddle the expected value of the security.

From a regulatory standpoint the distinction between market making and arbitrage can be important, but most HFT firms do not make such a formal distinction nor are there clear economic differences in the strategies. HFT firms construct systems to maximize their information set to as precisely as possible estimate the true value of a security and to calculate and manage the risk of their entire portfolio. Once these two elements are in place, strategies using them can be constructed that trade actively or passively. Some strategies may even combine market making and arbitrage. For example, an HFT firm could place bid or ask quotes for an ETF based on the prices in the underlying securities. If the quote in the ETF is executed against, the strategy will then immediately hedge the risk by buying or selling some or all of the securities underlying the ETF.

On page 52 the SEC discusses structural strategies:

> Some proprietary firm strategies may exploit structural vulnerabilities in the market or in certain market participants. For example, by obtaining the fastest delivery of market data through co-location arrangements and individual trading centre data feeds (discussed below in section IV.B.2.), proprietary firms theoretically could profit by identifying market participants who are offering executions at stale prices.

This is a simple classic risk-free arbitrage strategy for trading a single asset in multiple markets at different prices simultaneously. On page 53 the SEC discusses directional strategies:

> Neither passive market making nor arbitrage strategies generally involve a proprietary firm taking a significant, unhedged position based on an anticipation of an intra-day price movement of a particular direction. There may, however, be a wide variety of short-term strategies that anticipate such a movement in prices. Some 'directional' strategies may be as straightforward as concluding that a stock price temporarily has moved away from its 'fundamental value' and establishing a position in anticipation that the price will return to such value. These speculative strategies often may contribute to the quality of price discovery in a stock.

Many arbitrage strategies involve the simultaneous buying and selling of related securities. An HFT firm does not care whether both securities are temporarily mispriced or whether one of them is correctly priced and the other is mispriced. If the HFT could identify which of the securities is mispriced, then it may be better to only buy or sell the mispriced security. By only trading one of the two securities the strategy involves more risk as the position is not hedged, but the HFT avoids the cost of the hedging transaction. Depending on the expectation of how quickly the temporary mispricing will disappear and the cost of risk, a directional strategy may be preferable to an arbitrage strategy. Arbitrage and directional strategies are simply approaches to profiting from HFT's ability to identify temporarily mispriced securities.

On page 54 the SEC discusses additional directional strategies:

The Commission requests comment on two types of directional strategies that may present serious problems in today's market structure – order anticipation and momentum ignition.

Order anticipation is part of the strategic interaction between all investors and traders. Every market participant attempts to guess what other market participants are doing and respond in the best way. HFT attempt to do this as do all other traders. Momentum ignition would violate anti-manipulation regulations. However, enforcement of such rules in a fully automated environment is not well developed.

In summary, the four HFT strategies discussed by the SEC -— passive market marking, arbitrage, structural, and directional — are different ways to capitalize on HFTs' investments in being able to access and process vast amounts of market data to turn it into valuable information.

At an HFT firm there is a near infinite amount of financial market data arriving continuously. To get an estimate of the magnitude of data arriving daily in 2010 I asked NASDAQ's Economic Research department how many orders NASDAQ receives per day. The answer is between 500 million and one billion. NASDAQ represents roughly 20% of overall U.S. equity trading. If markets receive orders in proportion to their trading volume (perhaps too high as dark pools have fewer public orders), then there are at least 2.5 billion orders in stocks each day. Normal market hours are from 9:30am to 4pm. 2.5 billion orders arriving over 23,400 seconds implies more than 100,000 events per second. Note that this is an estimate for only U.S equity markets. Incorporating foreign markets and derivative markets such as futures and options markets increases the estimate many times. Deciding what is relevant in this flood of data and what it means is extremely difficult.

Market microstructure provides a number of ideas for what important events may be occurring in the data. Liquidity demand (transacting immediately) can cause prices to temporarily deviate from their fundamental value and private information may be slowly incorporated into prices through the trading of informed traders (O'Hara (1995)). Market microstructure employs stylized theoretical models designed to be solvable and to provide clear economic intuition. While these do not pretend to be directly applicable to real market data, many statistical arbitrage strategies could stem from these two economic sources.

The importance of investors' demand for immediacy and private information to profitable trading strategies can be illustrated with the common statistical arbitrage strategy known as pairs trading (see Gatev, Goetzmann, and Rouwenhorst (2006) for a discussion of the history and profitability of pairs trading). The simple concept of pairs trading is to find two stocks whose prices historically move together. Then when their prices diverge, short the winner and buy the loser. If prices behave as they have historically, prices will converge and the arbitrageur profits. If markets are even weak-form efficient at all times, risk-adjusted returns from pairs trading should not be positive. However, Gatev, Goetzmann, and Rouwenhorst (2006) find this simple strategy based only on past price dynamics and simple contrarian principles is indeed profitable, although their Table 8 suggests its profitability declined as it became well known.

Nothing in the pairs trading example relies on why prices may temporarily diverge. The statistical arbitrageur, possibly an HFT firm, does not care why prices diverge, only that prices will later converge (on average). Demand for immediacy and private information offer two natural explanations for why the prices of two assets may diverge. The demand for immediacy means an investor needs to sell one asset for reasons unrelated to information, say to make a down payment on a house, buy a car, or pay a college tuition bill. The sale causes the price of that stock, call it stock A, to fall slightly; this decline following a sale is often referred to as the temporary price impact of a trade and a rise after a purchase would be similar. The other stock, call it B, in the pair's price does not change. Therefore, buying stock A and selling stock B is profitable once the temporary impact of the sale in stock A passes. The pairs trading arbitrageur's trading may be partly responsible for the convergence of stocks A and B's prices.

In the private information case an investor sells stock A because they have information suggesting it is overvalued. The sale causes the price of stock A to fall. In contrast to the demand for immediacy example above, this price impact of a trade is permanent because the trade is based on information. If the reason why stock A is overvalued also applies to stock B and the investor does not simultaneously sell B due to short selling or other constraints that may not apply to arbitrageurs, then stock B should subsequently decline in price when others discover the information. If the pairs trading arbitrageur's buys stock A and sells stock B then when stock B's price declines the strategy is profitable.

Market microstructure posits that trading results from a mixture of the demand for immediacy and informational motives. The pairs trading example illustrates an appealing aspect of statistical arbitrage: the strategy can be profitable regardless of the economic source of the divergence in stock A's and B's prices. If a HFT arbitrageur can divine whether price changes are due to temporary demand for immediacy or information, then, as discussed above, a directional HFT strategy would be more profitable.

More sophisticated statistical arbitrage can incorporate many more than two assets and employs sophisticated techniques to estimate the relationships among those assets. However, the basic principle remains to buy what appears to be temporarily cheap and sell what appears to be temporarily expensive, and when prices converge close out the position at a profit. While statistical arbitrage has been shown to be highly profitable, the events of August 2007 (Khandani and Lo (2007)) and Long-Term Capital Management suggest that they involve risk and that the use of leverage can result in years of profits being wiped out quite quickly. When market prices behave in ways inconsistent with prior data, e.g., the rapid decline of the S&P 500 index prices, as happened on May 6, 2010, the risk of statistical arbitrage potentially increases dramatically, potentially leading to the withdrawal of statistical arbitrage strategies. Such withdrawal is most severe when there is uncertainty about the quality of market data and uncertainty about any possible regulatory or market intervention, e.g., cancelling trades.

High frequency trading and price efficiency

As noted above, price efficiency is important for resource allocation in the economy and investment decisions by firms. In addition, one source of noise in prices is price pressure due to liquidity demand by long-term investors (see, for example, Hendershott and Menkveld (2010)). This transitory volatility represents an implicit trading cost to those investors. If HFT trades against this transitory volatility it can be viewed as providing liquidity and reducing long-term investors trading costs. If HFT trades in the direction of transitory volatility (pricing errors) it can be viewed as increasing the costs to those investors.

The four HFT strategies described by the SEC, passive market marking, arbitrage, structural, and directional, have differing impacts on the transitory volatility. Passive marketing making by HFT should reduce transitory pricing errors if HFT provide liquidity more efficiently than non-HFT market makers, if competition with HFT forces non-HFT liquidity provides to improve, or if HFT increase the supply of capital to the market making sector. An HFT structural strategy which takes advantage of stale non-HFT orders could lead to non-HFT liquidity providers withdrawing from the market, potentially increasing transitory pricing errors. However, generally this sort of competition leads to more efficient markets. HFT multi-asset mean reversion strategies can be viewed as efficient ways to provide liquidity cross assets and incorporate cross-asset information quickly into prices; although the profitability of these strategies could also lead to the withdrawal of non-HFT orders. Directional HFT strategies based on identifying and trading against transitory pricing errors leads to more efficient prices. In contrast, momentum ignition would increase transitory volatility.

Overall, HFT's role is similar to that of many other speculators in the market. Speculators can improve price efficiency by getting more information into prices and decreasing pricing errors, but some speculative strategies can directly decrease price efficiency or indirectly decrease it by leading other market participants to withdraw. Which role of HFT dominates is an empirical question.

A large body of research is devoted to testing market/price efficiency. Consistent with Fama's weak-form efficiency, efficiency is typically measured as degree to which stock prices follow a random walk. Initial work in this area by Lo and MacKinlay (1988) employs variance ratio tests to show deviations from a random walk, suggesting market inefficiency.

Variance ratios in market efficiency tests are formed by calculating the variance of stock returns at differing horizons. If stock prices follow a random walk then the variance of stock returns should increase linearly with the time horizon over which variances are calculated, e.g., the five-minute variance should be 5 times the one-minute variance. Therefore, a ratio of two variances scaled by the ratio of the two time horizons equals one in an efficient market.

While the variance ratios approach for testing for price efficiency is standard, testing for HFT's impact presents additional challenges. A common approach to test for causality is to identify exogenous variation in the independent variables of interest, in this case HFT activity, and see how this variation correlates with the dependent variable, here changes in price efficiency. For example, Hendershott, Jones, and Menkveld (2011) use the staggered introduction of a technological change on the New York Stock Exchange (NYSE) together with a measure of algorithmic trading to identify the causal effect of algorithmic trading (of which HFT is a subset) on measures of liquidity and price efficiency. Hendershott, Jones, and Menkveld find that more algorithmic trading leads to higher liquidity and more efficient price quotes.

One attempt to identify a quasi-natural experiment for HFT is found in Castura, Litzenberger, and Gorelick (2010). Prior to the NYSE 2006-7 automation, referred to as the Hybrid market introduction, latency for trades was in the four to ten seconds range (Hendershott and Moulton (2011)). Latency of this magnitude effectively precluded many HFT strategies. No such latency existed for Nasdaq stocks. Hence, to control for changes in overall market conditions Castura, Litzenberger, and Gorelick examine price efficiency as measured by variance ratios for NYSE and Nasdaq stocks from 2006 through 2009. The implicit assumption is that HFT was used in Nasdaq stocks before the end of 2006 and was only introduced on the NYSE around the beginning of 2007. Supportive of this is that trading in NYSE securities began to resemble trading in Nasdaq securities subsequent to 2006. Specifically there were substantial increases in NYSE stocks in the trading volume, the number of quotes to trades, and the amount of trading occurring in markets other than the primary listing market (off NYSE trading).

Figures 1-3 are taken from Castura, Litzenberger, and Gorelick (2010). These plot variance ratios each quarter for NYSE and Nasdaq stocks in the Russell 1000 (the 1000 highest market capitalization U.S. stocks) and the Russell 2000 (stocks with market capitalizations between 1,001 and 3,000 in the U.S.) at one second, 10 seconds, one minute, and 10 minutes. All variance ratios are less than one, indicating inefficiencies due to negative autocorrelation in returns. In 2006 the ratios for NYSE stocks are substantially lower than those for Nasdaq stocks. Starting in 2007 the gap between NYSE and Nasdaq stocks closes and the two markets are roughly similar in 2008 and 2009. These patterns suggest that changes on the NYSE led to more efficient prices on the NYSE. The improvement in price efficiency coincides with an increase in HFT on the NYSE. The variance ratio is a measure of signal plus noise to signal where the signal is the amount of information being incorporated into price. Therefore, an increase in the variance ratio could arise from either more information or less noise, or both.

The Castura, Litzenberger, and Gorelick (2010) study finds evidence that an increase in HFT improves price efficiency. As with all event type studies it is possible that some other change during the time period is responsible for the results. The introduction of Reg NMS is one possibility. However, this regulatory change was implemented simultaneously for both NYSE and Nasdaq stock, which makes the results based on the NYSE-Nasdaq difference in price efficiency less likely to be affected. Another possibility is that the Hybrid market brought additional changes to the NYSE. The most plausible explanation is that the Hybrid also reduced the advantage of the human traders on the floor. For the question of HFT's impact on price efficiency the ideal experiment would have been for the NYSE to first eliminate the floor's advantages while holding the market latency constant and then to reduce latency.

Hendershott and Riordan (2011) use a state space approach to examine the role HFT plays in the efficient and transitory components of prices. They use a data set from Nasdaq that identifies the participation of a large group of HFT in each transaction. Caveats with the data include Nasdaq only representing 20-40% of trading volume in these stocks and that not all HFT is identified, in particular HFT by large integrated firms, e.g., investment banks, is not identified. Hendershott and Riordan find that the direction of HFT trading, defined as order flow which is the net buying volume minus the net selling volume, overall positively correlates with changes in the efficient price and negatively correlates with the transitory pricing error. This suggests that the Castura, Litzenberger, and Gorelick (2010) results in increasing price efficiency may arise from HFT both increasing the information getting into prices and reducing the noise in prices. However, without a controlled experiment or some exogenous variation in HFT, it is difficult to demonstrate causality and to show that the information associated with HFT is "new" information that would not have gotten into prices without HFT.

HFT is a relatively new phenomenon and the empirical literature examining the role of HFT in price efficiency is still nascent. The evidence thus far does not suggest that HFT has harmed market efficiency. If anything HFT appears to have made price more efficient by both getting more information into prices and by reducing pricing errors.

The future of high frequency trading and price efficiency

Technological advances typically replace the most repetitive human activity with automation. For trading by long-term investors this includes breaking large orders into smaller pieces and determining the price, size, and timing of the individual orders. For short-run speculation such as HFT it involves processing as much data as quickly as possible to find predictability in future price changes and then to capitalize quickly by submitting orders. The algorithms implementing these tasks are programmed by humans and typically not completely automated as they leave some parameters to be set by humans, e.g., how quickly to accumulate a large position or which securities' data should be analyzed. As technology improves the trend of substituting technology for labour will continue in the investment process.

The remainder of this review focuses on how the continuing trends of greater investment in computing and communications technologies impact HFT and price efficiency. Four broad questions and their impact are discussed:

1) Is HFT competitive?
2) Will HFT-like services be offered to low-frequency investors?
3) Are automated markets with automated trading less stable, e.g., prices occasionally become quite inefficient, than markets with more human intervention?
4) If HFT begins to reduce price efficiency will market structures evolve to limit HFT?

Competition in HFT.

If HFT becomes an even more competitive industry price efficiency should improve. To the extent that HFT profit from price inefficiency, competition should reduce HFT profitability and, therefore, price inefficiency. Price efficiency improving HFT strategies, e.g., statistical arbitrage and passive market making, should further improve price efficiency. Price efficiency reducing HFT strategies, e.g., manipulative directional strategies, should be more difficult to implement as one HFT firm attempting to distort prices away from fundamentals presents a profit opportunity for other HFT firms. Keeping data costs low and ensuring equal access at reasonable prices are natural requirements to keeping barriers to entry low and increasing competition.

One of the greatest potential barriers to entry in competition in HFT could be the high fixed technological development costs. Lowering these costs is crucial for competition. If the technology used for HFT is standardized, then the costs are limited. If HFT firms compete on developing custom hardware and software, then entry costs will be significantly higher.

Limiting the value of small improvements in speed is a potential approach to lowering technology costs. This could be accomplished by slowing down the speed of trading or reducing the advantages of speed, e.g., reducing the value of time priority. However, changes to the market mechanism must be carefully considered as there is substantial scope for unintended consequences such as reducing incentives for investors to innovate and increasing the risk to limit order submitters as hedging is more difficult. Furthermore, it is possible to build markets now that trade more slowly and have priority rules which reward speed less. The fact that these markets are not widespread suggests that their disadvantages outweigh their advantages. More discussion is below under potential market responses if HFT reduces price efficiency.

HFT tools for non-HFT.

While there is no systematic evidence that HFT have reduced liquidity for institutional investors put to this point, continued study is important. If over time HFT profits come largely at the expense of long-term investors, HFT may reduce market participation and trading. Price efficiency can suffer if this results in less information production about longer-term fundamentals or higher costs for market makers due to less efficient risk management arising from lower trading volume. HFT profiting from soon to be public information imposes costs on slower traders. If lower-frequency traders can incorporate that same information set into their strategies, then HFT should not reduce market participation by long-term investors and price efficiency reducing HFT strategies should be more difficult to implement. There are no barriers to lower-frequency traders utilizing the same information as HFT and brokers already offer institutions algorithms designed to combat HFT by limiting the predictability generated by the execution of large orders. If these algorithms further incorporate HFT tools to finding other sources of predictability, then participation and price efficiency should improve. If HFT becomes competitive and less profitable, programmers and traders with HFT knowledge have increased incentive to profit from their skills and knowledge by selling these to brokers for incorporation into their algorithms. This demonstrates an indirect benefit of ensuring competition in HFT.

HFT and market stability.

Questions about market stability are typically difficult to answer as instability is often defined as rarely occurring large price movements. Theoretically, algorithms to trade written by humans could be inferior in keeping markets stable compared to humans in trading because the algorithms' decision rules must be precisely coded. If market conditions exceed the circumstances for which the algorithms were designed then they could exhibit unexpected behaviour. For this reason HFT firms may cease trading in turbulent market conditions, which could exacerbate volatility. Human emotions can also exacerbate volatility by causing overreaction to new and uncertain market conditions, making it unclear whether algorithms and HFT inherently leads to less stability.

Empirically, the events on May 6, 2010 in the U.S. have been used as an example of HFT increasing instability. For example, Kirilenko, Kyle, Samadi, and Tuzun (2011) suggest that HFT exacerbated market volatility that day. However, their conclusion requires implicit untestable assumptions about how the market would have behaved absent HFT. In addition, May 6, 2010 was quite mild compared to financial market behaviour on October 19-22, 1987. In 1987 the ex post temporary price dislocations lasted for days and were transmitted around the world. Claims can be made that May 6, 2010 would have never occurred without HFT. Similarly, one can assert that October 1987 would have been a brief blip if HFT had been prevalent. Two events from two different periods in history are unlikely to offer sufficient evidence to support either assertion.

The impact on stability of non-HFT automated trading strategies requires careful consideration. HFT systems are typically designed to tightly limit the positions taken. Automated non-HFT trading are more of a concern for stability as at times non-HFT attempt to build or liquidate large positions quickly, something that occurred on May 6, 2010.

While market-wide instability may not have increased, it is claimed that individual stock mini-crashes have become much more prevalent. It is difficult to know the source of these momentary disappearances in liquidity. The current regulatory approach employing short-horizon, e.g., five minute, price limits or trading halts in individual stocks prevents short-lived mini-crashes, but also induces inefficiency when new information arrives. The mini-crashes could be due to HFT, to related market structure changes such as SEC Regulation NMS, to trading fragmenting across many interlinked markets, or to the disappearance of designed market makers with obligations to prevent the momentary drying up of liquidity. Regulation imposing obligations on HFT may be worth rigorously considering. A major concern with such obligations is that they could increase the barriers to entry for potential HFT firms and make HFT less competitive. The reduction in competition in HFT would take time to materialize, making evaluation of market makers obligations difficult to empirically evaluate.

Potential market responses if HFT reduces price efficiency.

While there is no evidence so far that HFT reduces market efficiency, this review suggests ways in which it could occur in the future. Before exploring these possible outcomes it is worth noting that if HFT increases market and price efficiency then the need for regulatory intervention is limited, if anything leaning towards facilitating HFT. If long-term investors prefer more efficient markets and HFT begins to have measurable negative effects on price efficiency is regulatory intervention necessary? At least three barriers could prevent market forces from providing market structures that limits HFT. First, the development costs of a new market could be prohibitively large. There is little evidence to support this as one of the most successful new markets, Chi-X, was essentially written by one programmer in three months. Second, a new more efficient market could be unable to attract trading due to liquidity externalities. Liquidity externalities involve traders wanting to trade in markets where other traders are already trading. The proliferation of new markets suggests that these externalities may not be as strong as once believed. However, some success of new markets can be attributed to HFT (Menkveld (2011)), raising the possibility that markets designed for long-term investors might face difficulty growing. Finally, equal-access regulations are often designed to make it difficult for certain groups of traders to trade only amongst them themselves. For example, trade-through rules such as Reg NMS can force long-term investors to trade with HFT.

Needs for continued HFT study.

The literature on HFT is quite small with the component relating to price efficiency only one component. Future research should follow many approaches with all of them requiring better data than is currently available. Well structured policy experiments, e.g., the Reg SHO, can provide significant insight, but only if data is made widely available. For studies of HFT at a minimum this requires orders and trades be marked as HFT or not. However, HFT firms are typically highly secretive and reluctant to have any data revealing anything about their trading strategies made public. Just as with disclosures by institutional investors and hedge funds a balance must be found that protects HFT firms' intellectual property while still enabling study. It is important that data on HFT be coordinated across markets and asset classes so future research can better understand the cross asset and cross market impacts of HFT. Similar data on and study of automated non-HFT trading strategies is also needed.

References

Amihud, Yakov, and Haim Mendelson, 1996, A New Approach to the Regulation of Trading Across Securities Markets, New York University Law Review 71, 1411-146.

Biais, Bruno, Lawrence Glosten, and Chester Spatt, 2005 Market Microstructure: A Survey of Microfoundations, Empirical Results and Policy Implications, Journal of Financial Markets 8, 217-264.

Jeff Castura, Robert Litzenberger, and Richard Gorelick, 2010, Market Efficiency and Microstructure Evolution in U.S. Equity Markets: A High-Frequency Perspective, working paper RGM Advisors. Attached to a SEC comment letter posted at http://www.sec.gov/comments/s7-02-10/s70210-155.pdf.

Chan, Kalok, 1993, Imperfect Information and Cross-Autocorrelation among Stock Prices, Journal of Finance 48, 1211-1230.

Fama, Eugene, 1970, Efficient Capital Markets: A Review of Theory and Empirical Work, Journal of Finance 25, 383-417

Gatev, Evan, William Goetzmann, and Geert Rouwenhorst, 2006, Pairs Trading: Performance of a Relative-Value Arbitrage Rule, Review of Financial Studies 19, 797-827.

Grossman, Sanford, and Joseph Stiglitz, 1980, On the Impossibility of Informationally Efficient Markets, American Economic Review 70, 393–408.

Hasbrouck, Joel, 1991, the Information Content of Stock Trades, Journal of Finance 46, 179-207.

Hendershott, Terrence, Charles Jones, and Albert Menkveld, 2011, Does Algorithmic Trading Improve Liquidity? Journal of Finance 66, 1-33.

Hendershott, Terrence, and Albert Menkveld, 2010, Price Pressures, Working paper, UC Berkeley.

Hendershott, Terrence, and Pamela Moulton, 2011, Automation, Speed, and Stock Market Quality: The NYSE's Hybrid, Journal of Financial Markets, forthcoming.

Hendershott, Terrence, and Ryan Riordan, 2011, High Frequency Trading and Price Discovery, Working paper, UC Berkeley.

Ho, Thomas, and Hans Stoll, 1983, The Dynamics of Dealer Markets under Competition, Journal of Finance 38, 1053–1074.

Keim, Donald, 1983, Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence, Journal of Financial Economics 12, 13-32.

Khandani, Amir, and Andrew Lo, 2007, What Happened To the Quants in August 2007? Journal of Investment Management 5, 29–78.

Khandani, Amir, and Andrew Lo, 2011, What Happened To the Quants In August 2007? Evidence from Factors and Transactions Data, Journal of Financial Markets 14, 1-46.

Kirilenko, Andrei, Albert Kyle, Mehrdad Samadi, and Tugkan Tuzun, 2011, The Flash Crash: The Impact of High Frequency Trading on an Electronic Market, CFTC working Paper.

Lo, Andrew, and Craig MacKinlay, 1988, Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test, Review of Financial Studies 1, 41–66.

Menkveld, Albert, 2011, High-Frequency Trading and The New-Market Makers, Working paper.

O'Hara, Maureen, 1995, Market Microstructure Theory, Blackwell, Oxford.

Securities and Exchange Commission, 2010, Concept Release on Equity Market Structure, Release No. 34-61358; File No. S7-02-10.

Tobin, James, 1984, On the Efficiency of the Financial System, Fred Hirsch Memorial Lecture, New York, Lloyds Bank Review, No. 153, July, 1-15, reprinted in Policies for Prosperity; Essays in a Keynesian Mode, Edited by Peter M. Jackson (1987), Wheatsheaf Books, Brighton, Sussex, 282-296.

Foresight
1 Victoria Street
London SW1H 0ET
www.foresight.gov.uk
URN 11/1231