An apparent paradox may have emerged in market making: bid-ask spreads and aggregate profits from market making (after accounting for higher trading volumes) have declined but aggregate profits from high-frequency trading (HFT), through which market making is now conducted, have increased. There are divergent views as to the aggregate profits of HFT, but some estimate that they are higher by perhaps as much as a factor of 10 than the profits from traditional market making in the pre-HFT era. This article describes how the technological evolution of the US equity market’s microstructure and competitive forces may have combined to produce this paradox, the implications for market performance, and the dilemma this poses for regulators. With high-frequency trading now exceeding 50 percent of the US trading volume for listed equities, observers worry whether or how the US Securities & Exchange Commission (SEC) will strike a balance on HFT regulation.

When Congress mandated the creation of a National Market System through the Securities Acts Amendments of 1975, stock trading was dominated by the manual trading floor. The SEC sought to promote the “linking of all markets for qualified
securities through communication and data processing facilities”, with the objective of promoting efficient execution of transactions, best execution of investor orders, price transparency, and the opportunity for investor orders to interact with each other. Incremental advances in the National Market System over the last 20 years have resulted in radical changes in market structure. Manual equity trading intermediated by regulated specialists or semi-automated market makers has been replaced by machine trading, which is highly distributed, disintermediated and fully automated by largely unregulated and highly interconnected trading firms.

Despite these technological changes, four trading or execution based strategies continue to dominate. First, specialists that make active quote markets and benefit from the bid-ask spread have been replaced by firms that engage in passive market making. In the past, central location access provided adequate compensation for specialists’ market making services. Now, liquidity provision is compensated directly through liquidity rebates for providing resting orders. Whereas specialists used to hold both the order book and inventory to ensure liquidity, the new passive trading system of liquidity is based upon an enormous volume of small order layering and cancelling that is taking place over very short periods of time without inventory. Second, arbitrage continues to be important. This strategy still involves using information about pricing differences across similar products that only differ in their combinations (e.g., ETFs v. the basket of stocks) and locations (stock in Chicago or NY). But whereas latency (the time taken to execute a trade) was measured in months centuries ago (e.g., the time it took to cross the Atlantic), in the modern era latency has fallen to milli- or even microseconds. Arbitrage has always used liquidity but now the amounts of liquidity soaked up by HFTs are enormous. Third, structural differences have arisen whereby those with fresh price quotes based upon structural advantages (faster computing or better locations) trade against those with stale prices. Fourth, directional strategies based upon information about order flow that affects the momentum and direction of price changes (e.g., order anticipation strategies that predict large orders) are empowered with special order types not generally available to all investors. A concern with directional or momentum trading is that traders will attempt to manipulate the market to their advantage.

In aggregate, the difference between the National Market System now and 40 years ago when the basic regulatory and structural concepts were formulated is that the effects of technology, speed, decentralisation and lack of transparency on the source of profits has created an enormously complex system of algorithmic bot traders. The aggregate economic incentives of the bot traders are of concern to regulators and market participants. Importantly, while prices are transparent to regulators, liquidity is not.

High frequency trading has reduced the cost of market making (e.g., through automation and creating economies of scope across markets) and competition has delivered tighter spreads but generally only for smaller orders and over short time periods. Aggressive HFT inventory management counters the benefits to non HFT-market participants of tighter spreads, but is necessary for HFT firms to limit exposures from market making: on average holding times are a few minutes, and HFT firms are mainly flat overnight. Aggressive HFT inventory management can also take the form of sudden withdrawal from the market and even demanding liquidity rather than supplying it in
times of market stress. This can make an order book fragile, a fact that was witnessed during the Flash Crash on 6 May 2010. This episode showed that when market conditions are disrupted, HFT market makers operate with small inventory, high trading volume, and take more liquidity than they provide, according to Kirilenko, Kyle, Samadi and Tuzun (2010, revised 2014). A possible explanation for this is that market making is more risky in the HFT era. In their characterisation of the market, a handful of the most technologically-advanced HFTs are able to impose an “immediacy-absorption cost” (i.e., by demanding liquidity from others following a liquidity shock) on all non-HFT market participants, and on HFT market makers that are not quite as fast. “Thus, high frequency trading can make it both costlier and riskier for market makers to maintain continuous market presence,” they note. At the same time, the few most advanced HFT firms have become the “main beneficiaries” of the competitive expansion of the provision of immediacy (reduction in latency) by HFT firms. The apparent paradox is explained: HFT market making is more competitive, but the most advanced HFT firms appear to be earning rents.

The dilemma for the SEC (as for the CFTC in the case of the E-Mini forwards market that experienced the Flash Crash) is that a superficially desirable outcome of technology innovation and competition may be tighter spreads, but this comes with greater fragility and lower quality market making. Furthermore, increased market competition and dramatic market fragmentation have “effectively eliminated much of the profitability of the registered market maker function and therefore, eliminated the ability for the Exchanges to impose significant quoting or trading obligations” (Recommendations Regarding Regulatory Responses to the Market Events of May 6, 2010, Summary Report of the Joint CFTC-SEC Advisory Committee on Emerging Regulatory Issues, 8 February 2011) that historically were used to regulate quality.

So what is the true value of additional innovation and competition between HFT firms focused on improving immediacy (i.e., reducing latency)? All market makers now compete and invest in speed improvements of dubious social value because the issue of efficient price discovery for listed equities has long been solved. It appears that an arms race is underway with the undesirable effect of the winner takes all, or at least a race that needlessly wastes resources to no end.

A core question is whether this activity is purely rent seeking, which simply changes the distribution of resources among different sorts of traders, as opposed to profit-seeking, which creates wealth through mutually beneficial trades with no negative spillovers to others. The creation of ever larger dark pools where buyers and sellers can interact independent of the influences of bot traders suggests that HFT is creating undesirable effects for institutions that manage fiduciary portfolios for millions of individual long term investors.

The SEC faces a challenge to adapt rules and oversight to balance its objectives of improving pricing for investors while ensuring that dealers are rewarded for intermediation (Putting Technology and Competition to Work for Investors, Mary Jo White, Chair, US Securities and Exchange Commission, on 24 June 2014). However, a first order problem that arises from effectively deregulating intermediation, through interconnection of trading centres, payments for order flow, market access and price priority rules, has been the creation of a system of bots so complex that humans may have lost their regulatory power. The SEC once had the ability to impose significant quoting or trading obligations
on market makers in return for a transparent and privileged oligopoly position. This loss of regulatory power is reflected by the fact that the SEC has repeatedly extended its 2010 study of the equity market structure without resolution. It has provided some measures of the system’s effectiveness in fostering price competition, studied the Flash Crash and gathered policy recommendations, but has not recognised the deeper problems with the quality of liquidity provision. These problems, underemphasised by policymakers, arise from the current market structure, which fosters greater competition between HFT market markers, but lowers the quality of intermediation. Likewise, the instinctive regulatory desire to increase the transparency of dark pools and to regulate order execution therein, poses the risk of decreasing intermediation options for investors with no obvious offsetting benefits.

There are several dimensions of the quality of liquidity provision or intermediation: quote size, the lifetime of any resting bid or offer, and the time-in-inventory of positions taken to supply liquidity. These quality measures affect order execution of block trades that exceed the size of the quoted best bid or offer. Even if the bid offer spread narrows for small orders as a result of competition between HFT market makers, there is no guarantee that long term investors and retail investors’ trade executions will improve. Indeed, a possible explanation for higher aggregate HFT profitability could be deterioration in order execution performance of long term investors resulting from the effect of aggressive HFT trading strategies that exploit superior speed and special order types to anticipate order flow.

SEC Chair Mary Jo White rejects the contention that Reg NMS is responsible for market fragility associated with growth in HFT and in aggressive trading strategies. She emphasises that the Flash Crash in 2010, in which HFT activity played a role, occurred in the E-Mini S&P 500 futures market that the SEC does not regulate and so it was not subject to Reg NMS. However, a principal question posed by the agency when it launched its study, of whether the market structure “serves the interests of long-term investors who are willing to accept the risk of equity ownership over time and are essential for capital formation”, is yet to be answered. The 2010 Concept Release focused on whether new regulatory measures were needed in relation to HFT, co-location and dark pools.

Order anticipation strategies were a notable topic, described as involving “misappropriation of information, or other misconduct” but whether the current market structure enabled these strategies and demanded regulatory action was posed as an open question. Analysis of the Flash Crash has established that passive HFT market makers’ demand more liquidity than they supply in a crisis. This has led to a recommendation for an anti-disruptive trading rule “tailored to apply to active proprietary traders in short time periods when liquidity is most vulnerable and the risk of price disruption caused by aggressive short-term trading strategies is highest” (Enhancing Our Equity Market Structure, Mary Jo White, Chair, US Securities and Exchange Commission, on 10 June 2014). No such concern has yet been expressed for order anticipation strategies.

Critics of the SEC’s focus on promoting price competition between deregulated HFT market makers at the cost of ignoring systemic risk need look no further than the band-aid solutions to systemic risk better known as circuit breakers. For instance, in recent Congressional testimony, Jeff Sprecker, CEO of ICE, provided a critique and made recommendations: “While NMS achieved its goal of
increasing competition among markets, the pendulum has swung too far at the cost of less competition among orders... We should eliminate and ban maker-taker pricing schemes at trading venues... all routing practices should be disclosed by those trading centers and brokers who touch customer orders”. Other critics make a connection between the two tier system that encourages legalised front running with the fragility and systemic risk that cannot be corrected with circuit breakers. Their proposals are to re-establish broker dealer obligations and define momentum ignition and order anticipation strategies as manipulation.

SEC Chair Mary Jo White announced the formation of the Market Structure Advisory Committee of experts to review specific initiatives and rule proposals on 5 June 2014. Her speech provides support for the proposition that order execution costs have decreased as a result of the market structure developments of recent years. We are concerned that the trend towards a fragile system reliant on algorithmic market making which institutional traders are avoiding through ever greater use of dark pools should not be ignored by appealing to selected facts about increased market quality that do not capture overall the implications of this systemic shift.