The Welfare Effects of Financial Innovation: High Frequency Trading in Equity Markets

Emiiliano S. Pagnotta (NYU Stern), Thomas Philippon (NYU Stern)

December 2010

“I wish somebody would give me some shred of evidence linking financial innovation
with a benefit to the economy.”


I Introduction

The social value of financial innovation is an issue on which economists disagree. Whereas a few innovations like the ATM and more recently online banking can clearly improve life quality and productivity, in most cases the overall assessment is controversial. Advocates of financial innovations argue that they lead to a better allocation of financial risks across individuals and firms, more liquid and efficient markets, and better access to credit. These factors can foster investment and thus economic growth. Merton and Miller (Miller (1986), Merton (1992)), for example, argue that the U.S. national mortgage market, the development of international markets for financial derivatives like currency swaps, and the growth of the mutual fund investment industries played such role.

There are reasons to believe, however, that some innovations can have negative social welfare effects. Many innovations are focused on “getting around” regulations and taxes. To the extent that (at least some) regulations serve a useful purpose, these innovations can have negative consequences (structured investment vehicles (SIV) for instance, were used to circumvent banking capital requirements). Innovations that create rents for one participant at the expense of another (zero-sum games) are also most likely to have negative social value. The so-called “Flash orders” in markets like NASDAQ, that enabled selected traders to see and act on orders for less than a second before the public is given an opportunity to trade with those orders, are one recent example. Even when some value is created, the social value might be less than the sum of costs sunk in the innovation
process (arms race). Finally, as recent events have shown, the process of innovation may also incentivize large investments in financial products whose risks are not well understood, with systemic consequences.

The lack of evidence on the value of financial innovation reflects key technical challenges researchers face in trying to assess net welfare effects. First, there is an absence of relevant data sources. Databases that are commonly used for financial research, like the bank call reports, the Center for Research in Security prices (CRSP) data files, and COMPUSTAT data files, yield no directly useful information about financial innovation. Second, it is difficult to isolate the effect of financial innovations on measures of net welfare like GDP growth from other causal effects. Third, for many innovations, like securitization, costs can become obvious when a crisis takes place while benefits are more dispersed and empirically difficult to grasp.

The analysis of HFT and related innovations in security markets provides an opportunity to get around some of these issues. As we discuss later in this document, our special data files allow us to observe the actions of the innovators (high frequency traders) and study their effect on market aggregates and other participants’ investment decisions at different points in time. Second, unlike opaque asset markets such as those organized as Over the Counter (OTC), the interactions between investors in centralized exchanges are well defined and documented in trading tapes, making it easier to distill the consequences of innovations from other pre-existing factors we know little about. Third, the high degree of digitalization in equity markets makes decision-level trading records extremely accurate and detailed.

A simple way to think about HFT is that it increases the speed at which liquidity can be provided. The benefits potentially include faster and cheaper trades, as well as more informative prices. For example, combining fast technologies and co-location, HFT can quickly arbitrage and eliminate price inefficiencies between say the price of an ETF and the underlying basket of stocks. The potential costs include the opportunity cost of human capital (highly skilled programmers) and physical capital (high speed computers and networks), the barriers to entry and information processing costs that HFT creates for other users, predatory trading and market manipulation (e.g. quote stuffing) and the potential instability in liquidity provisions (HFT-based liquidity can disappear quickly in times of stress and its speed makes it less compatible with human interventions, as shown by the May 6 flash crash).

Research on these issues is important for the following reasons. First, to understand the impact of equity trading on other spheres of economic activity (both in the U.S. and internationally). Whether HFT and algorithmic trading may affect a market’s systemic stability, for example, is central to the current public debate and to policy makers. Second, to anticipate impacts in a much larger class of assets for which HFT is or will become increasingly important, like derivative
Policy makers are generally concerned about the impact on long-term investors. These market participants provide capital investment and are willing to accept the risk of ownership for an extended period of time. Instead, we presume professional HF traders generally seek to establish and liquidate a position in a short timeframe. In this regard, HF traders may like short-term volatility to the extent it provides more trading opportunities, while long term investors do not. However, the net effect of trading strategies of multiple short-term HFT may not necessarily increase volatility and may even dampen volatility. Does the fact that sophisticated HF traders are likely always able to trade faster than long-term investors render the equity markets unfair? Does the very brief duration of most of their orders significantly detract from the quality of liquidity in the current market structure? These are key empirical questions we would like to investigate further.

The events of May 6 best illustrate the risks of overlooking the potential costs of financial innovations in equity markets. Recently, the strategies of HFT in many trading centers have largely displaced the role of market makers with affirmative and negative obligations. We presume that the absence of specific regulation regarding liquidity obligations during times of stress in this new era may be one of the factors behind such large disruptions. Some particular set of strategies, on the other hand, like the so-called momentum ignition strategies, may worsen the downward spiral. This may occur, for instance, by triggering standing stop loss orders. Whether some or all HFT firms should be subject to affirmative and negative obligations (and if so, to which ones?) is an important open question. We seek to contribute to such debate providing an explicit framework of analysis of the current market place. Overall, we expect our research to provide further guidance in the search for a new regulatory framework for equity and related markets where HFT may prevail.

II Analytical Framework

Consider the following stylized framework. Risk-neutral investors trade a risky asset in a purely electronic market. Their valuation for the asset depends on a common value component $z$, which evolves according to

$$z_t = z_0 + \sigma B_t$$

and a private value component denoted $x$ and $y$ for sellers and buyers, which is distributed according to $F$ (independently of $z$). The common value process is publicly observed, but with a latency that
may be agent specific. A potential buyer submits a single “fill-or-kill” limit order for one unit of the asset at a random time $\tau$ which is exponentially distributed with parameter $\mu$ and independent of $z$. The buy order is priced at $p^b_\tau = z_\tau + y$. In this simple single trade version the random time $\tau$ also indicates the closing of the market.

We concentrate here on the problem of sellers. In the benchmark case there are no gaps in trading latencies.

A  No latency gaps: the case of a single seller

Consider the case of a single seller with valuation $x$ that is able to submit a single order at time zero. We assume here that the seller latency of recontacting with the market, $\Delta$, is high. In other words, the event $\tau > \Delta$ is very unlikely. She then chooses a limit order price, denoted $p^a_{\Delta}$, for a single unit of the asset so as to maximize expected trading profits:

$$\max_{p^a_{\Delta}} \mathbb{E}\pi^s(p^a_{\Delta}) = \mathbb{E}[p^a_{\Delta} \mathbf{1}\{\text{trade}|p^a_{\Delta}\} + (z_\tau + x) (1 - \mathbf{1}\{\text{trade}|p^a_{\Delta}\})]$$

$$= \mathbb{E}\left[(p^a_{\Delta} - z_\tau - x) \mathbf{1}(p^b_\tau \geq p^a_{\Delta})\right] + (z_0 + x) \quad (1)$$

$$= \mathbb{E}\left[(p^a_{\Delta} - z_\tau - x) \mathbf{1}(z_\tau + y \geq z_0 + p^a_{\Delta})\right] + (z_0 + x) \quad (2)$$

where the expectation above is conditional on $x$ and $z_0$. The program (1) can then be expressed as follows

$$\max_{u^\Delta_0} \mathbb{E}\pi^s(u^\Delta_0) = \mathbb{E}\left[\left(u^\Delta_0 + z_0 - z_\tau - x\right) \mathbf{1}(z_\tau + y \geq z_0 + u^\Delta_0)\right] + (z_0 + x) \quad (3)$$

where $u^\Delta_t$ represents $p^a_{\Delta} - z_t$. Maximizing the RHS of (3), the optimal price premium can be derived.

In the particular case in which $y \sim N(1,1)$, and in the absence of latency gaps, the optimal price premium $u^*_0$ maximizes

$$\left[1 - \Phi\left(\frac{u^*_0 - 1}{\sigma^2 + \mu}/\mu\right)\right] (u^*_0 - x) - \int_{-\infty}^{\infty} \phi\left(\frac{u^*_0 - y}{\sigma^2/\mu}\right) \Phi\left(-\frac{u^*_0 - y}{\sigma^2/\mu}\right) \phi(y - 1) \, dy \quad (4)$$

where $\Phi$ and $\phi$ represent the Gaussian CDF and pdf functions. By solving this framework we will derive the market liquidity conditions in the absence of latency gaps and compute the associated expected gains from trade.
B High-Frequency Trading and Latency gap

The previous simple framework can be augmented with a second seller who derived no private value from the security but can re-contact the market much faster (with a latency $\Delta_L$ that is significantly smaller than the industry standard $\Delta$). We call such agent a high-frequency trader. The HFT presence will affects long term investors’ quoting and trading behavior. Solving the model with and without HFT will then offer (i) provide an identification schemes for hidden quantities in the data (mainly the distribution of private values) (ii) insights on the counterfactual states where investors have access to the similar trading technologies.

In the context of the previous example, suppose such seller contacts the market at time $\Delta_L > 0$. If the slow seller order is still outstanding, the HFT will compare the expected profit of the following possible actions (i) hitting that order (ii) undercutting the price (iii) posting an order at an equal or higher price. We seek to generate expressions for the value of a general latency value $\Delta_L$ by computing the associated expected profits. $V(\Delta_L, \theta)$ for a given environment $\theta = (\sigma, \mu, F, \Delta)$.

C The innovation: a “market for speed”

The innovation. Assume there are competitive markets for speed technologies $s_\Delta = 1/\Delta \in [1, \infty]$, associated with latencies in the set $[0, 1]$. The cost of each technology is given by $C(\Delta)$. Before trading, a potential investor solves a technology investment problem

$$\max_{\Delta} V(\Delta, \theta) - C(s_\Delta)$$

It is to expect that the decentralized solution to this problem will differ from a central planner solution due to inefficient business stealing. By comparing both solutions we will derive the expected welfare loss once the market for speed opens.

To the best of our knowledge, there is no existing equilibrium framework of trading at different speeds (or with endogenous entry of fast traders). In a recent working paper, Jovanovic and Menkveld (2010) consider a related market setting with private values but abstracting from speed asymmetries and assuming that regular investors do not provide liquidity in the presence of a high-frequency trader. Instead, our preliminary evidence shows that non high-frequency traders actively provide liquidity. Pagnotta (2010) studies a framework where each trader chooses an optimal speed of trading but in the absence of market contact latencies. Moallemi and Saglam (2010) introduce latencies in a trading framework but they solve a single agent optimization problem, abstracting from the strategic interactions between both type of investors. Kirilenko and Cvitanic (2010) analyze an elegant statistical model of high-frequency trading and analytically study the impact on prices.
III Empirical Analysis

For any given innovation, the overall assessment of its welfare consequences naturally reduces to whether social benefits are greater than social costs. However, “ideology” and special interests, rather than well-grounded scientific research, tend to fuel the debate. As a matter of fact, the stock of empirical evidence on either costs or benefits of financial innovation is shockingly scarce given its importance.

The goal of the empirical part of the paper is two fold. We seek to first document, from the perspective of an institutional investor, the costs of executing large orders before and after HFT became dominant market participants.

We will exploit a unique data set that contains 120 stocks from the Russell 3000 and includes complete trading records between January 2008 and February 2010. NASDAQ OMX, based on industry contacts, general knowledge of the business, and analysis of its own data, takes the lead in identifying “High Frequency Traders”. Thus, all activity will be identified as either HFT or non-HFT activity. This is a critical feature for the purposes of this study and will allow documentation of the liquidity provided by and removed from HF traders, the price impact of their actions and an indication of their trading profits. There are no other comparable datasets for US equity markets with this feature. In a study of high frequency trading, having fast snapshots of the market is naturally crucial. Since trading records are stored at maximum disaggregation level, milliseconds, we can hope to document these firms’ actions and market consequences accurately.\footnote{1}

More specifically, the data fields include: (1) all trades that took place for each stock within the sample periods. Trading records specify who, whether a HFT firm or a regular investor, took and provided liquidity for each transaction (2) NASDAQ’s best bid and ask quotes, together with the best collective bid and ask quotes prices and quote sizes from HFT firms and from non-HFT firms (3) Liquidity status of the limit order book.

A Time Variation in the Execution Cost of a Large Trade

We seek to document the cost of large trade using high-frequency equity data from two recent periods: 2002 and 2008-2010. The 2008-2010 sample is described above. The 2002 data is from a proprietary order-level data set of NYSE and NASDAQ listed stocks that traded in the NYSE. We believe the trading environment in the NYSE in 2002 is an ideal benchmark. Many of the important determinants of trading costs- such as decimalization and the ability of investors to place and interact with orders in an open electronic LOB -were present in both periods. However, high-

\footnote{1This data set is also used by [Brogaard 2010]. We believe there are serious shortcomings in his approach and methodology to support the conclusion that high-frequency trading has been unequivocally beneficial.}
frequency trading was not a market-wide significant phenomenon in 2002. We believe that a rigorous computation of such trading costs is necessary to make more informed claims on the HFT impact on market liquidity.

B The Cost and Benefits of HFT

For any given innovation, the overall assessment of its welfare consequences naturally reduces to whether social benefits are greater than social costs. However, “ideology” and special interests, rather than well-grounded scientific research, tend to fuel the debate. As a matter of fact, the stock of empirical evidence on either costs or benefits is shockingly scarce given its importance.

Our approach seeks to identify upper and lower bounds for both benefits and cost of the HFT phenomenon.

- Upper bound on benefits: assuming no negative externalities exist, the potential benefits of HFT are likely to be related to the improvement in market liquidity leading to a higher volume of non-HFT trades (measurable in our dataset). [Hollifield et al. (2006)] propose a methodology to compute the gains from trade using the order flow in an electronic limit order market. We plan to adapt such methodology to an environment of latency gaps using the insights from the theoretical framework. Intuitively, the likelihood of any given HFT taking or providing liquidity and his limit order prices will depend, ex ante, on (i) the distribution of latencies (i.e. the “latency gap”) (ii) the asset characteristics (e.g. the volatility parameter \( \sigma \)) and (iii) the distribution of private values. Hence, we will exploit the identification power of the evolution of the make-take ratio and the evolution of the quotes for HFT and non-HFT firms in our sample. The outcome of this exercise will be ratio between the realized gains from trade and the potential ones. We will contrast the results with (i) the gains from trade ratio in the pre-innovation period, 2002, and (ii) the model-based counterfactual state of liquidity when no high-frequency traders are present. A significant advantage of our own dataset is that, being able to identify HFT-to-HFT trades, we can more accurately identify trades with zero private value gain.

- Upper bound on costs: We will compute the total trading profits over the period 2008-2009. We believe this is a reasonable metric for the maximum possible technology investment during recent years.

- Lower bound on costs: We will use industry estimates on the cost of setting up a high-frequency shop to construct a lower bound on investments in the HFT industry.
We provide in this document some preliminary evidence on price and trade behavior from the special NASDAQ HFT dataset. Figures 1-5 display the intraday behavior of HFT and regular investors for five randomly selected stocks in the sample. The blue and red lines correspond in each case to the difference between the best collective quote of regular investors and HFT at the bid and offer sides, respectively. Positive values indicate that HFT are posting limit orders with the most competitive prices ($\text{ask}^H < \text{ask}^H$ for the red line). Figures 1-5 suggest that (i) there is persistency and pronounced time variation in the way HFT interact with the rest of the market place (ii) slow investors and HFT alternate in providing liquidity at the most competitive prices in intriguing ways (iii) very frequently fast investors provide the best prices on one side of the market only.

IV Final Remarks

The process of innovation in financial markets has been at the center stage of policy debates since the recent crisis hit the global economy. The issue of whether financial innovation creates net value for society underlies much of the efforts to redesign the international financial system. Despite enormous social impacts, virtually no well-grounded scientific research documenting its welfare consequences is available. In this project we aim to provide evidence of both social costs and benefits by focusing on the role of high frequency traders in equity markets.

We believe equity markets are a good first stage from which this research program can develop. Besides the ubiquitous place of equity markets in the economy, there are key technical advantages related to the quality of the data sources. In particular, our research will be supported by a unique set of up-to-date trading files from NASDAQ in which the actions of high frequency traders (the “innovators”) are precisely identified. We are confident that our analysis will provide valuable insights that will help policy makers make more informed decisions concerning policy and regulation reforms. But the benefits are likely to extend beyond equity markets. We think our work and methods can provide a foundation and stimulate more academic work on related aspects of the complex process of financial innovation.
References


Figure 1: Quoting Behavior: High Frequency Traders versus regular investors. MMM February 24 (Nasdaq)

This Figure shows the difference between the best collective quote of regular investors and HFT in the bid (blue line) and offer sides (red line), measured in dollars. Positive values indicate that HFT are posting limit orders with the most competitive prices ($ask^H < ask^H$ in the case of the red line).
Figure 2: Quoting Behavior: High Frequency Traders versus regular investors. BZ February 26 (Nasdaq)

This Figure shows the difference between the best collective quote of regular investors and HFT in the bid (blue line) and offer sides (red line), measured in dollars. Positive values indicate that HFT are posting limit orders with the most competitive prices ($ask^H < ask^H$ in the case of the red line).
This Figure shows the difference between the best collective quote of regular investors and HFT in the bid (blue line) and offer sides (red line), measured in dollars. Positive values indicate that HFT are posting limit orders with the most competitive prices ($\text{ask}_H^H < \text{ask}_H^H$ in the case of the red line).
Figure 4: Quoting Behavior: High Frequency Traders versus regular investors. DCOM February 24 (Nasdaq)

This Figure shows the difference between the best collective quote of regular investors and HFT in the bid (blue line) and offer sides (red line), measured in dollars. Positive values indicate that HFT are posting limit orders with the most competitive prices (\( \text{ask}^H < \text{ask}^H \) in the case of the red line).
Figure 5:

LANC Feb 24

ht_bid_improvement  ht_ask_improvement
Figure 6: High-frequency traders: Dollar-weighted Trade Participation

This Figure shows the daily time series fractions of trades for the sample of 120 NASDAQ stocks. The label HN indicates that a high-frequency trader initiated the trade, taking liquidity from a non HFT investor. Similarly, NH indicates that a non HFT investor took liquidity from a HFT. The sample spans every active day since January 1 2008 to December 31 2009.
Figure 7: High-frequency traders: Shares-weighted Trade Participation. This Figure shows the daily time series fractions of trades for the sample of 120 NASDAQ stocks. The label HN indicates that a high-frequency trader initiated the trade, taking liquidity from a non-HFT investor. Similarly, NH indicates that a non-HFT investor took liquidity from a HFT. The sample spans every active day since January 1 2008 to December 31 2009.
Figure 8: High-frequency traders: Trades-weighted Trade Participation. This Figure shows the daily time series fractions of trades for the sample of 120 NASDAQ stocks. The label HN indicates that a high-frequency trader initiated the trade, taking liquidity from a non HFT investor. Similarly, NH indicates that a non HFT investor took liquidity from a HFT. The sample spans every active day since January 1 2008 to December 31 2009.