Government Office for Science

Foresight

Maker/taker pricing and high frequency trading

Economic Impact Assessment EIA12

Foresight, Government Office for Science

Contents

Maker/taker pricing and high frequency trading

Ryan Riordan

(Karlsruhe Institute of Technology)

Andreas Park

(University of Toronto)

14 February 2012

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.

1. Objective

What are the costs, benefits and risks of implementing maker-taker fees in the presence of high frequency traders? We present a view on the interaction of maker/taker fees and high frequency trading on financial markets. High frequency traders engaging in both market making and active trading strategies are increasingly driving trading volume in financial markets. Trading platforms, in turn, are offering innovative fee structures to attract high frequency and other trading volume. The most common new fee structures employ maker/taker fees that charge asymmetric fees to participants who demand and supply liquidity. Typically, markets offer a rebate to those participants' who supply liquidity and charge a fee to those who demand liquidity. These new fee structures have lead to a concern that high frequency traders are the prime beneficiaries of these fee rebates due to their ability to (almost) always place their orders at the top of the queue.

2. Background

Over the last decade, the equity trading landscape has changed fundamentally. Stocks that could once be traded only on a national, primary listing exchange, can now be traded on several international exchanges as well as on a variety of new, upstart alternative trading venues. As trading becomes fragmented across different markets, trading platforms must find ways to compete in order to attract volume. One of the key areas of competition has been the fees that venues charge traders for executing transactions on their platform.

Today, public trading is almost exclusively organized via so-called electronic limit order books. The underlying trading mechanism collects the passive limit orders that cannot execute immediately and sorts them by price and then by time. These orders form the order book and give other traders the opportunity to initiate a transaction in the sense that traders can send an "active", marketable order to be executed against orders in the book. The passive orders provide liquidity to the market, active orders remove liquidity from the market. One important difference between trading in a limit order book and old-fashioned human-touch market making is that electronic limit order books rely on the *voluntary* provision of liquidity. In competing for volume, trading venues must entice traders to supply liquidity on their platforms. To attract liquidity, many trading platforms, in particular the new, upstart alternative trading systems, now directly subsidize the provision of liquidity by paying passive orders, that are hit by active orders, a fee rebate (and subsequently charging the active order a positive fee). Since the liquidity providing passive trader is a *maker* of liquidity and the active trader is a *taker* of liquidity, this practice of charging asymmetric fees is commonly referred to as *maker/taker pricing*.

The structural move of trading away from human interaction to electronic trading facilitated another major change: with electronic limit order books, traders can implement their trading strategies using computer algorithms. Computerized trading now vastly dominates trading on public venues and makes up close to 50% of dollar trading volume on public equity exchanges (Securities and Exchange Commission, January 14, 2010).

At first sight, algorithmic trading and maker/taker pricing appear to be separate issues. The two are, however, linked, conditional on other factors, because maker rebates have helped to establish a new business model, by which computer algorithms act as de facto market makers by posting limit orders on both sides of the market at very high frequency with the dominant purpose of capturing maker rebates. For instance, in its concept release on market structure the SEC writes

> “Highly automated exchange systems and liquidity rebates have helped establish a business model for a new type of professional liquidity provider that is distinct from the more traditional exchange specialist and over-the-counter (“OTC”) market maker.” (S.E.C. “Concept Release on Equity Market Structure”, Release No. 34-61358; File No. S7-02-10.)

Some trading venues have, in fact, designed their fee schedules specifically to attract high frequency market-making traders (e.g. the Toronto Stock Exchange has a special rebate program for “electronic liquidity providers”).

Modern securities markets are now often organized as electronic limit order markets. In limit order markets executions occurs when submitted orders “cross” (active) the price of an order sitting passively in the book. The submitted order that crosses is typically referred to as a market, or a marketable order, and is thought to consume the liquidity provided by the passive order resting in the book. The market order is the market taker and the passive limit order is referred to as the market maker. Markets typically enforce price-time priority, however, time priority has lost importance in fragmented markets.

The dominance of computerized and, in particular, high frequency trading and market making (henceforth *HF trading* and *HF market making*) is a hotly debated issue among professional traders, regulators, academics, and the general public. As maker/taker pricing is seen as a major facilitating factor to the emergence of high frequency trading, trading fees have become an integral part of the debate.

There are two main concerns that are commonly voiced with regard to high frequency market making. First, “HFTs crowd out traditional traders and harm aggregate welfare”. More specifically, the concern is the following. HFTs are systematically faster at posting limit orders in the book, especially when it is beneficial to do so. Slow, traditional investors then find it difficult to obtain executions through passive limit orders. Further, when they obtain an execution these executions are on economically unfavorable terms (loosely, they execute only when HFTs disappear, indicating that trading at that point is particularly costly, for instance, because of an upcoming information event). To obtain executions, slow investors may thus be forced to trade with active (market) orders for which they have to pay (a) the bid-ask spread and (b) the taker fee. Another way to put this concern is to say that high frequency traders “squeeze themselves between” two “natural” traders to establish themselves as (unwanted) intermediaries. While the advent of the high frequency market making business model may have created some losers, the aggregate welfare implications of this development are not necessarily negative. More generally, economists usually don’t worry if a party established itself in a market and, by superior technology, is able to benefit from less nimble parties. However, if the slow investors choose to stay out of a market, possible gains from trade are no longer realized and these foregone gains from trade lead to a loss in aggregate welfare. The complaint with regard to maker/taker fees in this context is that without the fee rebates, passive, liquidity providing HF strategies would not be profitable so that, allegedly, the existence of fee rebates has lead to lowered welfare.

The second concern raised with regard to high frequency market making is that “HFTs supply liquidity only in calm markets when it is cheap, that they disappear during periods of market stress, and that, by disappearing suddenly, they exacerbate market turbulences (or they may even create artificial market crashes”. Arguably, if, in markets that rely heavily on HF liquidity provision, too many HFTs reduce their intermediation activities simultaneously, large-scale price movements are possible. Even though HF activities are not solely to blame,[1] the May 6, 2010 “Flash Crash” marks one example where liquidity providers disappeared suddenly and in large numbers. The concern is thus that if the continuous provision of liquidity cannot be guaranteed, we may see many more mini (flash) price crashes or jumps that are related purely to sudden changes in the demand and supply of liquidity and not to fundamental information. The question that arises is whether maker/taker fees may play a positive role in this context.

In the following we will synthesize the existing, though sparse evidence on the impact of maker/taker pricing on trading in general and on high frequency trading in particular.

3. Existing make/take fee structure

In general, trading platforms earn revenues by charging investors a fee to trade on their platform. Markets are interested in maximizing revenues and thus seek to attract volume to their platform. There are two ways to organize fee structures. Fees can be:

1. Symmetric in the sense that the maker and taker pay the same fee per transaction
2. Asymmetric in the sense that the maker and taker pay different fees.

The London Stock Exchange and TMX Select (the Toronto Stock Exchange’s own Alternative Trading System) are examples of platforms that charges symmetric fees. Asymmetric fees are charged on most public U.S. and European alternative trading systems, e.g., NASDAQ, Chi-X, BATS, etc. The most common format charges takers a positive fee and gives makers a rebate. There are also “inverse” schemes under which makers are charged a positive fee and takers receive a rebate (e.g., Chi-X Canada employs such pricing for TSX Venture securities).

In addition to the fee structure market operators price transactions based on

1. the number of shares traded in the transaction, or
2. the dollar value of the transaction.

Some market operators provide further rebates or reduced rates based on the total value or volume traded. Finally, some platforms have special programs for computerized trading; examples are Deutsche Börse’s Automated Trader Program, NYSE Euronext’s Pack Epsilon, or the Toronto Stock Exchange’s Electronic Liquidity Provider Program.

3.1. European (MiFID) vs. U.S. (Reg-NMS) Securities Regulation

The current regulatory environments in U.S. and Europe are different with respect to order-protection and best-execution rules, both of which are important when studying maker taker fees. Order-protection and best execution rules go hand in hand. The U.S. regulatory system provides “top of book” protection. Orders that are not explicitly routed are routed to the market with the best available price, regardless of the number of shares and number of shares available at that price. In the U.S. algorithms take this into account when routing their orders

[1] SEC-CFTC findings regarding the market events of May 6th, 2010 released September 10th, 2010 available at http://www.sec.gov/news/studies/2010/ marketevents-report.pdf.

and optimally split their orders to reduce their execution costs. Retail investors are for the most part left to their own devices.

Under MiFID brokers are under the obligations to enforce a best-execution policy on overall basis. Brokers are required to produce and publish a set of rules by which they abide when routing client orders. Best-execution is not enforced on an order-by-order basis, as in the U.S. under Reg-NMS but is evaluated yearly. Another key difference between Reg-NMS and MiFID is that in the U.S. exchanges are responsible for ensuring order-protection and therefore best execution. In Europe brokers are responsible for ensuring that orders are routed to the best venues. This leads to the third key differentiating factors, under Reg-NMS best execution is one-dimensional; price. Under MiFID best execution is multi-dimensional and includes price, speed of execution, reliability and other qualitative factors. These differences make comparison difficult with respect to the impact of maker/taker fees on the various stakeholders.

3.2. Theoretical research on maker/taker pricing

Theoretically, it is not clear why the breakdown of the total exchange fee between makers and takers of liquidity should have real economic consequences. Instead one can make an argument that the breakdown should be irrelevant. Imagine that an exchange, which used to charge a symmetric fee, switches to provide makers with a rebate, but holds constant the total fee (taker fee minus maker rebate). One key feature of a limit order book is that traders can decide whether they want to post a limit order (and thus supply liquidity) or whether they want to trade with a market order (and thus demand liquidity). Ceteris paribus, limit orders become cheaper and some traders will switch from market to limit orders. Yet, this lowers the execution probability of limit orders and thus traders will post better prices to attract a market order. If there are no frictions, so that the price can adjust infinitely finely, liquidity providing traders will improve prices to the point where the benefit from the rebates is exactly consumed by the narrowed bid-ask spread. In other words, absent frictions, the introduction of an asymmetric fee should leave the true economic costs of a market order unaffected. Moreover, there should be no effect on trader behavior or on liquidity provider entry, assuming that the total exchange fee, i.e. the difference between the taker fee and the maker rebate, remains unchanged. This point was also made in a survey by, Harris, and Spatt (2010).

Colliard and Foucault (2011) provide a theoretical model that builds on this argument. They study a limit order market in which investors can be buyers or sellers and choose either limit orders (liquidity supply) or market orders (liquidity demand) to execute their trading strategies. Traders meet on a trading platform that charges a fee per transaction, split asymmetrically between makers and takers. The platform selects the total fee and its maker/taker breakdown to maximize its expected profit. In their analysis, Colliard and Foucault carefully distinguish between the *raw* bid-ask spread and *cum fee* bid-ask spreads. The raw bid-ask spread is the difference between the lowest posted ask price and the highest posted bid price. The cum fee bid-ask spread adjusts the raw spread by the taker fees. Practitioners sometimes refer to the cum-fee spread as the “economic” spread as it reflects the true transaction cost of an active trader. Assuming a zero tick size, Colliard and Foucault develop a number of empirical predictions; the key one relating to this survey is that the fee breakdown does not affect traders’ ex ante welfare in the sense that the breakdown does not affect the cum fee bid-ask spread (as long at the total fee stays constant), even though it does affect the raw spread.

Colliard and Foucault thus prove an irrelevance result with regards to the impact of maker/taker pricing on cum-fee bid ask spreads. Importantly for the analysis of fragmented markets, they also show that their irrelevance result extends when investors can choose between multiple trading platforms, because investors would always route their orders to the venue with the tightest cum-fee spreads. Moreover, as long as they charge the same total fee, platforms with different make/take breakdowns can coexist.

Foucault, Kadan and Kandel (2010), henceforth FKK, provide further theoretical insights, for instance when there is a minimum tick size. In their framework, makers and takers of liquidity assume fixed roles but differ by the intensity with which they monitor the market. FKK show that trading platforms can use maker/taker fees to increase the rate at which buyers and sellers meet in the market. As trades involve realizations of gains from trade, FKK's results thus imply that maker/taker fees, when used appropriately, can lead to an increase in aggregate welfare. The minimum tick size, in fact, plays an important role, as FKK show that, when the tick size vanishes, the same irrelevance result as in Colliard and Foucault applies. FKK thus deliver two important insights for regulation: first, make/take fees can increase welfare, and, second, abolishing minimum tick sizes may prevent the positive impact of make/take fees.

The intuition for FKK results is as follows. They interpret the liquidity supplying side as proprietary-trading firms specialized in high frequency market making and the liquidity demanding side as brokers using smart routers to execute market orders when the bid-ask spread is narrow. In their model the liquidity makers and takers react at different frequencies to changes in the limit order book. These differences can lead to an imbalance in the supply and demand for liquidity, in that the speed at which transient decreases in liquidity are replenished and the speed at which liquidity takers hit competitive quotes. These imbalances cause gains from trade to be unrealized. FKK then show that maker/taker fees can be used to generate more trades. For instance, if the level of competition among makers is low compared to that among takers or if makers have higher monitoring costs than takers, then a maker rebate can motivate makers to be more vigilant in their market monitoring activities. With a rebate, they would replenish the order book faster after liquidity was removed, and with improved liquidity in the book, the trading rate increases.

Colliard and Foucault's (2011) findings provide an important theoretical benchmark. Yet real markets have a number of frictions. For instance, many markets have a minimum tick size so that the traders may not be able to adjust the cum-fee spread to exactly offset changes in the make/take fee breakdown (as long as the tick size isn't always binding, one could imagine that their model may be extended to allow traders to generated a non-discrete tick-size across time by varying the bid-ask spread). Furthermore, there may be differences in how the brokers, who pay the trading platform in the first place, pass on the fees to their customers. Brokers may also receive payments for specific order flows, which may affect the choice of venue where they route an order. Finally, current market regulations create frictions by how they affect order routing. So-called order protection rules (as in place in the U.S. or in Canada) require that orders are routed to the venue that displays the best prices – but these best prices are raw, and not cum fee.

Overall, these two theoretical papers provide important positive and normative benchmarks that need to be explored empirically.

3.3. Empirical evidence on the impact of maker/taker pricing

Malinova and Park (2011) provide empirical evidence on impact of maker/taker fees, and their detailed dataset allows them to examine a number of questions and contentious issues in policy debates. Namely, they study how the introduction of maker rebates affected market quality, the level of competition among liquidity providers, the trading costs of retail traders, and the degree of intermediation. They further address whether after the introduction of make/take fees, some parties trade on the active side more often.

Malinova and Park study the introduction of liquidity rebates on the Toronto Stock Exchange in 2005. At the time, only a subset of the listed securities switched to a fee structure that gave makers of a liquidity a rebate of $2.75 for a 1,000 share trade and that charged the taker a fee of $4 for a 1,000 share trade. The majority of stocks remained at the old regime (though there was a minor decline in fees), according to which only the taker was charged of 1/55% of the value of the transaction and the maker paid no fee. The 2005 (partial) introduction is thus a clean experiment that allows Malinova and Park to employ a difference-in-differences event study analysis.

Overall they find that the introduction of liquidity rebates led to decreased raw spreads, increased depth, increased volume, and an increase in the competition for liquidity supply. Moreover, identifying a subset of stocks for which the total fee is (almost) constant they show that, consistent with Colliard and Foucault, the cum-fee spreads did not change even though the make/take fee breakdown changed strongly. On the other hand, they also observe that the benefit to liquidity providers, measured by the realized spread increased, caused by the combination of the maker rebate and a reduced price impact. Thus even though the breakdown did not affect the cum fee spread, it does have an impact on other important trading parameters.

Regulators are particularly concerned about retail investors. Malinova and Park's data details trades to the level of trading desks, and they thus identify traders that manage retail flow by trading characteristics. Malinova and Park show that while the group of retail traders is on the active (taker) side of the market more often after the fee change, they experienced no change in their net costs of trading. These net costs were computed as the active cum fee spread costs minus passive cum rebate benefits. Finally, identifying those traders who provide sufficient liquidity so that they can be counted as intermediaries, they show that there is evidence for an increase in intermediation.

Although there was computerized trading in Canada in 2005, naturally there was much less such trading compared to today's markets. Still, it is arguably possible that market participants programmed some algorithms to take advantage of the new passive fee rebates. And while Malinova and Park's analysis shows that, at the time, the introduction of fee rebates led to an improvement of many aspects of market quality, the increased active to passive ratio for retail traders and the increase in intermediation raise some concerns and indicate that there may be some developments that require monitoring in future research.[2]

Battalio, Shkilko, and Van Ness (2011) (hereafter BSVN) study maker/taker pricing in the U.S. options markets. They compare trading costs for two different market structures. One, in which maker/taker fees are used to finance liquidity provision, and another, in which liquidity is supplied by market makers who make profits by skimming retail orders via a practice known as "Payment for Order Flow" (PFOF). BSVN find that raw trading costs are about 80 basis points higher in the PFOF structure. Nonetheless, cum-fee costs are actually higher in the marker-taker structure, by about 74 basis points. When BSVN account for the amount of payment for order flow (that brokers likely disburse to their customers), trading costs under PFOF appear

[2] Biais, Foucault and Moinas (2011) provide a theoretical framework to formalize some of these concerns. They provide a model with slow non-HF and fast HF traders, and they study the decision of market participants to become HF. In equilibrium, there is overinvestment in HF technology, and for a non-negligible portion of the parameter set, non-HF traders abstain from trading. These traders would generally benefit from trading, and there is thus a loss of welfare.

lower by 249 basis points. BSVN findings thus highlight the importance of assessing execution quality including all fees. BSVN however caution that the lower cum-fee liquidity costs in the PFOF structure should not be taken as direct evidence of the superiority of this structure over the maker/taker structure. Rather the lower cum-fee costs are likely the result of cream skimming of retail orders by the PFOF market makers. It is important to note that HFT is not as pervasive in the options markets as it is in the equity markets (Chapman, 2011). Therefore, BSVN results may be more attributable to the effects of the market structure and less so to the effects of HFT.

There are concerns that asymmetric maker/takers fees, particularly rebates, are of particular benefit to HFT and not benefitting the market overall in terms of more and better liquidity. Menkveld (2011) describes the activities of *one* HF market maker that operates as a market maker on Chi-X and the Euronext. Euronext uses a standard symmetric pricing scheme, whereas Chi-X uses an asymmetric maker/taker fee structure. In both markets the HFT operates primarily as a market maker: roughly 78% of their trades are passive in both markets. The gross spreads earned by the HFT are €2.09 on the Euronext and €2.52 on Chi-X. The average Chi-X rebate for passive trades is €0.31 and the average Euronext passive charge is €0.68. The fee difference is €0.99 and is quite large considering the size of the gross spread, suggesting that fees are of first-order importance for market makers. When one includes the clearing and settlement fees differences between Euronext and Chi-X, the cost/benefit difference becomes even larger.

Table 1. Taken from Menkveld (2011). This table decomposes the net spread for a high frequency trader on the Euronext, Chi-X and on both markets.

variable	large stocks	small stocks	all stocks
panel A: high-frequency trader in both markets			
entrant market (Chi-X) trade share (%)	49.8 (43.7, 62.8)	56.5 (51.6, 63.6)	50.8 (43.7, 63.6)
net spread per trade (e)	1.68 (0.76, 2.15)	0.80 (0.25, 1.64)	1.55 (0.25, 2.15)
panel B: high-frequency trader in incumbent market (Euronext)			
#trades per day	770 (216, 1189)	180 (48, 276)	684 (48, 1189)
fraction of passive trades (%)	79.5 (70.5, 82.5)	70.0 (58.7, 81.6)	78.1 (58.7, 82.5)
net spread per trade (e)	0.72 (0.09, 1.27)	−0.07 (−0.44, 1.01)	0.61 (−0.44, 1.27)
net spread per trade, passive orders (e)	1.26 (0.31, 2.03)	0.23 (0.05, 1.50)	1.11 (0.05, 2.03)
gross spread per trade, passive orders (e)	2.25 (1.25, 2.99)	1.17 (0.97, 2.44)	2.09 (0.97, 2.99)
exchange fee per trade, passive orders (e)	−0.68 (−0.71, −0.65)	−0.64 (−0.66, −0.62)	−0.68 (−0.71, −0.62)
clearing fee per trade, passive orders (e)	−0.30 (−0.32, −0.29)	−0.30 (−0.31, −0.29)	−0.30 (−0.32, −0.29)
net spread per trade, aggressive orders (e)	−1.35 (−2.21, −0.80)	−0.75 (−1.12, −0.23)	−1.26 (−2.21, −0.23)
gross spread per trade, aggressive trades (e)	−0.39 (−1.28, 0.17)	0.16 (−0.23, 0.66)	−0.31 (−1.28, 0.66)
exchange fee per trade, aggressive orders (e)	−0.67 (−0.70, −0.63)	−0.63 (−0.64, −0.62)	−0.67 (−0.70, −0.62)
clearing fee per trade, aggressive orders (e)	−0.29 (−0.30, −0.26)	−0.28 (−0.29, −0.27)	−0.29 (−0.30, −0.26)
panel C: high-frequency trader in entrant market (Chi-X)			
#trades per day	812 (128, 1269)	135 (45, 183)	713 (45, 1269)
fraction of passive trades (%)	77.1 (71.4, 81.8)	83.3 (79.0, 90.7)	78.0 (71.4, 90.7)
net spread per trade (e)	2.63 (1.88, 3.17)	1.92 (1.46, 3.05)	2.52 (1.46, 3.17)
net spread per trade, passive orders (e)	2.63 (1.97, 3.15)	1.87 (1.46, 3.14)	2.52 (1.46, 3.15)
gross spread per trade, passive orders (e)	2.46 (1.97, 3.05)	1.90 (1.49, 3.17)	2.38 (1.49, 3.17)
exchange fee per trade, passive orders (e)	0.34 (0.18, 0.45)	0.16 (0.11, 0.21)	0.31 (0.11, 0.45)
clearing fee per trade, passive orders (e)	−0.16 (−0.18, −0.14)	−0.19 (−0.22, −0.17)	−0.17 (−0.22, −0.14)
net spread per trade, aggressive orders (e)	2.61 (1.51, 3.36)	2.21 (1.43, 3.78)	2.55 (1.43, 3.78)
gross spread per trade, aggressive trades (e)	3.30 (1.91, 4.11)	2.65 (1.83, 4.18)	3.21 (1.83, 4.18)
exchange fee per trade, aggressive orders (e)	−0.48 (−0.61, −0.18)	−0.22 (−0.28, −0.18)	−0.45 (−0.61, −0.18)
clearing fee per trade, aggressive orders (e)	−0.21 (−0.22, −0.19)	−0.21 (−0.23, −0.20)	−0.21 (−0.23, −0.19)

The net spread for HF market maker trades is €2.52 on Chi-X and €1.11 on the Euronext. This result suggests that HFTs do not necessarily pass rebates on to liquidity demanding participants in the form of tighter gross or cum-fee spreads. Menkveld (2011) studies the introductory phase of Chi-X and it is an open question whether spreads and competition between liquidity suppliers reached a different equilibrium over time.

Hendershott and Riordan (2011) provide evidence on, among other things, the profitability of HFT strategies with and without exchange fees and rebates. Their sample stretches across 2008 and 2009 and encompasses 120 stocks listed on the NASDAQ and NYSE. The sample is split up into three market capitalization groups; (1) Large, (2) Medium, and (3) Small. During the sample period, HFTs are very active, counting for close to 50% of total dollar volume on NASDAQ. HF trading is more active in large stocks than in medium and small stocks. The sample descriptive statistics are as below:

Table 2: This table is taken from Hendershott and Riordan (2011).

Summary Statistics	Units	Source	Large	Medium	Small	All
Market Capitalization	$ Billion	Compustat	$52.47	$1.82	$0.41	$18.23
Price	$	TAQ	$56.71	$30.03	$17.93	$34.95
Daily Midquote Return Volatility	bps.	TAQ	16.5	25.8	42.9	30.3
Bid-Ask Spread	$	NASDAQ	$0.03	$0.04	$0.09	$0.05
Relative Bid-Ask Spread	bps.	TAQ	5.29	13.32	50.20	15.73
NASDAQ Trading Volume	$ Million	NASDAQ	$179.01	$6.35	$1.11	$62.28
HFT^{Init} **Trading Volume**	$ Million	NASDAQ	$77.06	$2.38	$0.26	$26.96
HFT^{Pass} **Trading Volume**	$ Million	NASDAQ	$75.86	$1.18	$0.11	$26.18
HFT^{All} **Trading Volume**	$ Million	NASDAQ	$152.92	$3.55	$0.37	$53.13

NASDAQ employs an asymmetric maker/taker pricing scheme. For most of the sample periods, NASDAQ charged market takers and provided rebates to market markers. The net fees range from -0.0001 and 0.00015 $US per share traded. NASDAQ's fees are not constant throughout the sample period; the maker rebate and taker fees changes at least 6 times during the sample period. NASDAQ seemingly views pricing as a relevant factor in their competition for order-flow, as shown by the number of price changes during the sample period.

Table 3 provides an overview of the pricing and changes between January 1st, 2008 and December 31st, 2009 on the NASDAQ for Tape A, B and C securities. Tape A securities are listed on the New York Stock Exchange, Tape B securities on regional exchanges, and Tape C securities are NASDAQ-listed.

Table 3: This table reports the maker, taker and net fees for Tape A, B, and C securities.

	Tape A		
Date	**Maker**	**Taker**	**Net**
01012008	0.0027	0.0028	0.0001
01022008	0.0027	0.00285	0.00015
01052008	0.0028	0.00295	0.00015
01062008	0.0028	0.00295	0.00015
01032009	0.0028	0.00295	0.00015
01042009	0.0028	0.00295	0.00015
	Tape B		
Date	**Maker**	**Taker**	**Net**
01012008	0.0026	0.0025	-0.0001
01022008	0.0025	0.00265	0.00015
01052008	0.0028	0.00295	0.00015

	Tape A		
01062008	0.0031	0.00295	-0.00015
01032009	0.0028	0.00295	0.00015
01042009	0.0028	0.0029	0.0001
	Tape C		
Date	**Maker**	**Taker**	**Net**
01012008	0.0026	0.0025	-0.0001
01022008	0.0025	0.00265	0.00015
01052008	0.0028	0.00295	0.00015
01062008	0.0028	0.00295	0.00015
01032009	0.0028	0.00295	0.00015
01042009	0.0028	0.00295	0.00015

During the sample period, NASDAQ engages in asymmetric only pricing. NASDAQ predominantly subsidized market makers during the sample period but subsidized takers during a brief period in January of 2008 for Tape B and C securities. NASDAQ applies rebates and fees on a per share, rather than dollar traded, basis. In contrast to some rebate programs (e.g., Deutsche Boerse – ATP) NASDAQ provides cash payments for participants with a positive net fee balance.

To capture the revenues that accrue to HFT strategies, Hendershott and Riordan (2011), following Sofianos (1995) and Menkveld (2011), calculate trading revenues for $\mathrm{HFT}^{\mathrm{Init}}$, $\mathrm{HFT}^{\mathrm{Pass}}$, and $\mathrm{HFT}^{\mathrm{All}}$. $\mathrm{HFT}^{\mathrm{Init}}$ captures HF trading, $\mathrm{HFT}^{\mathrm{Pass}}$ captures HF market making and $\mathrm{HFT}^{\mathrm{All}}$ captures summed trading ($\mathrm{HFT}^{\mathrm{Init}} + \mathrm{HFT}^{\mathrm{Pass}}$). They assume that HFT hold no overnight positions. For each stock and each day in the sample, they set the initial inventory to zero, cumulate revenues throughout the trading day and value the end-of-day inventory at the closing midquote for that stock and day.

Table 4 is taken directly from Hendershott and Riordan (2011). Panels A and B present the revenues that accrue to HFT for their active, passive and all of their trading. The most important point is that $\mathrm{HFT}^{\mathrm{Pass}}$ revenue is negative and (mostly) significantly different from zero before fees, and positive and significantly different from zero cum fees. They use an average rebate of 0.00275 per share traded to account for NASDAQ rebates. HF market makers ($\mathrm{HFT}^{\mathrm{Pass}}$) appear willing to supply liquidity to informed traders, thereby incurring adverse selection costs and recover these costs with fee rebates. This suggests that HFT, on average, pass some of the fee rebates on to liquidity demanders.

Table 4. This table is taken from Hendershott and Riordan (2011) and reports the revenue per stock and day broken down by HFT^{Init}, HFT^{Pass}, and HFT^{All}.

Panel A: HFT Revenue per stock day

	Large	Medium	Small	All
HFT^{Init}	\$7,464.64	\$428.88	\$59.91	\$2,681.16
(t-stat)	(6.89)	(5.38)	(3.77)	(7.30)
HFT^{Pass}	\$-1,911.30	\$-46.43	\$-0.04	\$-660.12-
(t-stat)	(-2.19)	(-0.92)	(0.00)	(-2.22)
HFT^{All}	\$5,553.33	\$382.45	\$59.95	\$2,021.04
(t-stat)	(4.03)	(5.01)	(2.94)	(4.32)

Panel B: HFT Revenue per stock day after fees

	Large	Medium	Small	All
HFT^{Init}	\$2,433.43	\$144.52	\$16.18	\$874.54
(t-stat)	(2.27)	(1.83)	(1.02)	(2.41)
HFT^{Pass}	\$4,209.15	\$148.91	\$21.62	\$1,476.56
(t-stat)	(4.76)	(2.93)	(1.43)	(4.90)
HFT^{All}	\$6,642.58	\$293.44	\$37.81	\$2,351.11
(t-stat)	(4.70)	(3.88)	(1.85)	(4.99)

The results in Table 4 also show that HFT are both profitable overall and when demanding liquidity.

Table 5, also taken from Hendershott and Riordan (2011) report descriptive statistics for the highest 10% of permanent volatility days. The highest permanent volatility days are identified using a state space estimation (SSM) (see Durbin and Koopman 2001). The SSM model decomposes price movements into their permanent and transitory components. Permanent movements are interpreted as being information related changes. Transitory movements are interpreted as pricing errors that are the result of the trading process. The top 10% of information days represent times at which information plays a large role in price movements and is generally thought to be exogenous to the trading process.

Table 5. Taken from Hendershott and Riordan (2011). This table reports the sample summary descriptive statistics on the top 10% of permanent volatility days.

	Units	Source	Large	Medium	Small	All
Daily Midquote Return Volatility	bps.	TAQ	30.99	47.05	72.50	49.96
Bid-Ask Spread	\$	TAQ	0.04	0.07	0.14	0.08
Relative Bid-Ask Spread	bps.	TAQ	9.14	27.89	82.87	27.81
NASDAQ Trading Volume	\$ Million	NASDAQ	\$271.00	\$8.77	\$1.30	\$94.70
HFT^{Init} **Trading Volume**	\$ Million	NASDAQ	\$121.74	\$3.31	\$0.33	\$42.19
HFT^{Pass} **Trading Volume**	\$ Million	NASDAQ	\$119.58	\$1.49	\$0.15	\$40.79
HFT^{All} **Trading Volume**	\$ Million	NASDAQ	\$241.33	\$4.80	\$0.48	\$82.99

The main point to be made is that HFT increase both their liquidity demanding and supplying activities equally and that the total amount of liquidity supplied and demanded remains the same.

Table 6, also taken from Hendershott and Riordan (2011) captures the revenue of trading strategies in the 10% of permanent volatility days; these days also happen to be days on which volatility, measured by the midquote return volatility, is high overall. By capturing the profitability of HFTs on these days they can say something about the willingness of HFTs to incur trading losses when adverse selection risks are the highest. Table 4 shows that before fee rebates HFTs incur adverse selection costs in large stocks, small stocks, and overall. These revenues are, however, not significantly different than the revenues on other days (the remaining 90% of the sample). Overall, the evidence suggests that HF market makers do not withdraw liquidity when markets are volatile and when liquidity supply is important to facilitate the discovery of the efficient price.

Table 6. Taken from Hendershott and Riordan (2011) and reports the revenue per stock and day broken down by HFT^{Init}, HFT^{Pass}, and HFT^{All} for the top 10% of permanent volatility days.

Panel C: HFT Revenue per high-permanent volatility stock day

	Large	Medium	Small	All
HFT^{Init}	$13,090.92	$1,228.87	$190.82	$4,883.63
(t-stat)	(1.13)	(1.43)	(1.27)	(1.28)
HFT^{Pass}	$(3,706.79)	$155.53	$(16.78)	$(1,201.14)
(t-stat)	(-0.53)	(0.56)	(-0.17)	(-0.47)
HFT^{All}	$9,384.13	$1,384.41	$174.04	$3,682.48
(t-stat)	(0.61)	(2.03)	(1.07)	(0.76)

Panel D: HFT Revenue per high-permanent volatility days after fees

	Large	Medium	Small	All
HFT^{Init}	$6,688.54	$890.09	$140.32	$2,597.46
(t-stat)	(0.88)	(1.35)	(1.21)	(1.03)
HFT^{Pass}	$3,960.65	$380.33	$9.13	$1,464.54
(t-stat)	(-0.07)	(0.64)	(-0.12)	(-0.01)
HFT^{All}	$10,649.20	$1,270.43	$149.45	$4,062.01
(t-stat)	(0.63)	(1.99)	(1.05)	(0.78)

This suggests that even on days during which adverse selection risks are high that HFT supply liquidity and continue to pass some of the fee rebates on to liquidity demanders. After Fees HFT^{Pass} has positive revenues that are not statistically different than on other days.

In their review Angel, Harris, and Spatt (2010) (AHS) provide an overview of the issues involved in 21st century equity trading. Particularly relevant to this impact assessment is their treatment of the trade-off between order-preferencing (pre-arranged order routing agreements) and order internalization. They summarize the issue as follows:

> “Many broker-dealers internalize their retail orders for the same reasons that brokers may preference the orders to certain dealers. Acting as dealers, these broker-dealers often provide price improvement to their customers. Trading this informed order flow can

produce excess dealing profits, especially if the NBBO reflects the costs of dealing to many well-informed traders. " (p. 29, Angel, Harris, and Spatt 2010)

AHS provide an overview of a possible problem generated by the current U.S. regulatory system, Reg-NMS, by which orders must be executed at the best available price (the "no trade-through rule"). The best available price often reflects adverse selection costs that retail orders are forced to pay when executing at the NBBO. One way to avoid paying these costs is to use a broker that has price improvement agreements with execution venues or that internalizes order flow at better than available public prices. Problems arise when the best available prices are only available at exchanges with access fees and/or maker taker fees. Retail investors are then forced not only to pay the spread but must also bear these costs. Conversely, the spread at venues with aggressive maker rebates may be artificially tight, forcing internalizers to effectively absorb the subsidies. AHS suggest that some of these problems could be solved by mandating that access fees, takers fees and maker rebates are passed on to end customers. This would lead to a situation under which brokers route orders to venues that provide the most benefits to their customers and not for themselves. We discuss these issues below.

4. Risk assessment

In this survey, we address the risks of asymmetric maker/taker fees on market liquidity and stability. We pay particular attention to the risks associated with computerized and high frequency trading in the current environment. There are a number of risks to be addressed.

1. Do the current fees (symmetric and asymmetric depending on execution venue) support the provision of liquidity?

The current environment is characterized by static fees that are either symmetric or asymmetric, depending on the execution venue. There exists a risk that HF market makers are taking advantage of the current structure to the detriment of retail and slower investors, but the empirical evidence is not conclusive. Specifically, Malinova and Park (2011) find that retail traders trade on the active side more often after the introduction of maker-taker fees, but they find no evidence that retail traders face higher costs. Their sample period is, however, from 2005 when HF trading was not as pronounced as it is now. They also find that in the maker-taker fee environment, which, loosely, encourages HF trading, quoted liquidity generally improves.

1. Is the supplied (if any) liquidity stable?

   i) Is the liquidity dependent on market conditions?

   ii) Is the liquidity "fleeting"?

   iii) Is HFT the cause of (in)stability?

An important risk in liquidity supply is the *stability*, which is defined as the ability to trade comparable sizes at comparable prices throughout the trading day. Particularly the periodic disappearance of liquidity is a concern to institutional and retail traders who cannot continuously monitor or contact the market, without incurring substantial costs. The most obvious example of liquidity instability was the *flash crash (May 6th, 2010)*, during which liquidity conditions deteriorated significantly within a short period of time. Although not the main culprits, HFT are thought to have exacerbated the crash by reducing their intermediation activities during the crash. Namely, Kirilenko, Kyle, Samadi, and Tuzun (2011) make three important observations: (i) HF trading behaviour did not change substantially during the crash; (ii) HFTs did not cause the crash; (iii) HFTs did contribute to the price declines in the later stages of the crash as they tried to unload their inventories when prices kept falling.

1. Are HF traders the prime beneficiaries of asymmetric pricing fee rebates?

   i) Is HF trading detrimental to the market overall

   ii) Is HF trading detrimental to certain participants?

HF market makers are pivotal drivers of modern financial markets. A non-trivial risk exists that their trading, and market making could reduce the efficiency of prices, increase transitory price movements, and reduce the overall market quality. Asymmetric pricing scheme could exacerbate these problems if they lead HF market makers to increase intermediation when it is needed least and to reduce their intermediation when it is needed most. This could lead to situations under which slower investors can enter into equity positions inexpensively when markets are quiet and at significant costs when markets are turbulent. Hendershott and Riordan (2011) find that this is not the case. HF market makers supply slightly more liquidity when markets are turbulent and generally trade to reduce transitory deviations in price. Taken together their evidence suggests that HF market makers take advantage of market maker rebates but that they are equally likely to supply liquidity during high and low volatility and thus do not appear to exacerbate periodic volatility.

5. Options

A number of regulatory options are available, have been discussed, and are presented below:

1. Can we develop a dynamic maker/taker fee model that balances liquidity demand and supply contemporaneously and/or intertemporally?

   i) Can we design field or laboratory experiments to identify pricing regimes that work better than the current ones?

2. Would periodic call auctions solve some of the problems associated with imbalances in liquidity demand and supply?
3. Should exchanges be forced to post prices cum fee?
4. What is the value of the current order protection rules (compared to best execution rules) and how can order-protection and best execution rules be adjusted to reflect, for instance, cum fee spreads and depth available?

The costs, risks and benefits of each are described below.

6. Costs, risks, and benefits

There are a number of costs, risks and benefits associated with each of the previously introduced options, discussed in the following.

6.1. Dynamic maker/taker fees

Dynamic, flexible liquidity rebates could compensate liquidity providers well when risk is high may ensure the existence of a continuous market. The introduction of dynamic fees may be costly: one would need to develop a theoretical model for optimal fees that would balance demand and supply. FKK (2011) develop a framework that enhances our understanding of the mechanisms, albeit in a stylized manner. It will be a costly endeavor to bring these ideas to market reality. Assuming that an optimal fee structure is known, a dynamic fee structure requires vigilant monitoring by trading venues and may trigger high monitoring costs for both regulators and venues. Moreover, liquidity or lack thereof is a market-wide phenomenon and since trading occurs on multiple venues, it is not clear which platform should be tasked with the implementation of dynamic fees. Furthermore, there is the risk that the dynamic fee structure can be gamed causing artificial drops in liquidity.

6.2. Periodic call auctions

Periodic call auctions (PCA) have been suggested as a solution to problems associated with liquidity demand and supply. The technology for PCAs is available and in operation at exchanges that offer opening calls auctions for their markets. As we do not observe call auctions alongside the continuous market, it is clear that private enterprise does not believe that PCAs have merit as long as there is a continuous market (some dark pools, e.g., ITG's POSIT, do have periodic matching of orders, albeit without price-finding).

A more radical approach would be to introduce PCAs as the only means of trading. PCA would then concentrate liquidity but limit the ability of market participants to manage risk throughout the trading day. PCAs could then cause large intraday price jumps uncommon in the current market. Alternatively, it is imaginable that between the PCA there will be a shadow market of when-traded/repo contracts for the next auction. Such a derivatives market would, at least for some time, be lightly regulated and unlikely to be transparent and it would be an inferior replacement for the existing continuous market. PCA would further require a fundamental change in the way that financial institutions invest and manage risk, involving substantial implementation/adoption costs for financial institutions. Finally, as many securities are cross-listed on international exchanges, abandoning continuous trading for PCAs may simply shift trading and liquidity from one country to another.

6.3. Cum fee pricing

Cum-fee pricing involves the posting and dissemination of buy and sell prices for securities that include any and all associated fees and rebates. This could be extended to include broker fees and rebates that may also be dependent on order routing decisions. The main costs would be in the development of a system(s) that capture fees that can be dependent on participant, auction types, execution venues, and overall venue trading volume. For instance, exchanges may offer fee discounts to repeat customers, differing by active and passive volume, and it may be difficult to impossible to include the discount system into public quotation systems. The costs to implement such as system are high and may be overly restrictive for trading venues business model. The primary risk is that the system doesn't capture the above dependencies properly, that it confuses market participants, and that prices are displayed incorrectly leading less-sophisticated investors to make erroneous and costly routing decisions. The prime benefit would be that investors are able to condition routing decisions on the total cost, and not solely on the posted prices for equities. Sophisticated investors and brokers already use algorithms and smart-order routers to determine the best execution venue conditional on fees and other parameters, and as such would receive no benefit.

6.4. New order protection/best execution rules

A new set of best execution rules in the E.U. would be costly to enforce, but could be included under the current broker obligations to provide for client best-execution on an overall basis. The costs of doing so would therefore be concentrated in the retail broking industry. The benefits are likely to be negligible as rebates are currently not passed on the retail investors on an order-by-order basis, rather they are reflected in the overall brokerage fees. Furthermore, as Battalio, Shkilko, and Van Ness (2011) argue, brokers are likely spreading costs and benefits across multiple customers in a way that may increase aggregate welfare of their customers.

7. Future

The existing evidence suggests little negative impact of HF trading/market making and make/take fees. Given the lack of empirical evidence and suitable natural experiments this lack of evidence is not startling. Some theoretical work suggests that there may be situations under which HF trading and market making is harmful (socially wasteful), these situations could be exacerbated by maker/taker fees. Empirically HF market makers appear to be passing-on some of the fee rebates to active traders. Whether this will remain to be the case in the future and whether or not this is true for all market participants in all markets is an open question. There are a number of further important issues that have not yet been addressed with regards to maker/taker fees and HFT:

1. Are non-HFTs forced into becoming active traders due to the constant presence of HF market makers at the best bid/ask?
2. Have trading costs decreased for all groups of investors when we take into account their switch from the passive to the active side?
3. What happens if HFT simultaneously pull out of the market and cease their liquidity provision activities?

Although important, answering these questions is beyond the scope of this review. To answer questions 1 and 2 researchers require detailed participant and order level data over long time horizons to capture the actual total trading costs paid by, for instance, retail, institutional and HF traders.

An important first step to improve the practice of maker-taker fees would to develop, test, and implement a fee pricing model that dynamically balances liquidity demand and supply (possibly based on the insights from Foucault, Kadan and Kandel (2011)).

An often-heard suggestion to help equate liquidity demand and supply is the use of periodic call auctions throughout the trading day to replace the current continuous markets. These call auctions would replace continuous trading and serve to focus the attention of market participants on fewer but centralized trading events. This may serve to increase liquidity and to decrease the likelihood of having short but dramatic liquidity deteriorations, but may have the side-effect of creating undesirable intra-day trading and price patterns. It may be worth exploring whether there is demand for such a mechanism alongside the continuous market. However, the mechanism itself is well-established in the form of opening and closing auctions, and the lack of an offering of the mechanism by private enterprise suggests a lack of interest by market participants.

The final two questions address problems associated with the current regulatory environment. Order protection regulation (as practiced, for instance, in the U.S. and Canada) distorts the competition between marketplaces by forcing investors to execute on venues that may offer the best raw price but a poor cum-fee price. U.S. regulation further forces investors to execute on the venue that offers the best raw price for the first share but worse prices for any additional shares. One approach would be to force exchanges to post prices cum fees and a second would be to change regulations to require only best execution. Each raises new questions and concerns, but a careful analysis of these issues may contribute to both better executions for customer orders and an increase in the efficiency of the market.

8. Summary and recommendation

There is little evidence in the academic literature to support the conjecture that high frequency traders are "abusing" the current fee structure. Hendershott and Riordan (2011) find evidence that high frequency market makers pass some of the rebates along to active traders as their negative pre rebate market making revenues suggest. Malinova and Park (2011) find that overall the bid-ask spread adjusted in the face of maker/taker fees to reflect the fee breakdown. Overall we find little evidence to support the conjecture that market/taker fees are the source of, or silver-bullet solution to, market quality problems.

In our section on future issues we highlights the gaps in our current knowledge and ways to close these gaps. The most important realization is that the current regulatory framework, particularly in the U.S., may be exacerbating any potential problems associated with maker/taker fees and HFT. We recommend leaving the regulatory environment, with respect to asymmetric fees, as is. We see little evidence to support the conjecture that HFTs are abusing the current system. We find no support for the mandatory implementation of pricing schemes, dynamic, symmetric, asymmetric, or otherwise. We support the current flexible best-execution policy in Europe that avoid the problems associated with static order-routing obligations.

We support providing investors with more information on execution quality. This information could come in the form of pre-trade cum-fee prices and with post-trade execution and other available price information. Such information puts investors in the position of being able to measure the quality of their routing decisions, or their brokers' routing decisions, after the fact and with the ability to make better routing decisions before future order submission decisions. We also support regulation that supports fee rebates being passed on to the end customer. Whether or not HFTs are the prime beneficiaries of fee rebates appears to be of secondary importance.

References

Angel, J. J., Harris, L. E. and Spatt, C. S. (2010), “Equity trading in the 21 century”, Manuscript, Carnegie Mellon.

Biais, B., T. Foucault, and S. Moinas, (2011), “Equilibrium High Frequency Trading”, Working Paper, Toulouse School of Economics.

Battalio, R., A. Shkilko and R. Van Ness (2011) “The Impact of Taker Fees and Order Flow Inducements on Trading Costs in U.S. Options Markets”, Working Paper.

Colliard, J. and T. Foucault (2011) “Trading Fee and Efficiency in Limit Order Markets”, Working Paper.

Durbin, J. and S. J. Koopman (2001) “Time Series Analysis by State Space Models. Oxford”, Oxford University Press.

Foucault, T., O. Kadan and E. Kandel (2011) “Liquidity Cycles and Make/Take Fees in Electronic Markets”, Working Paper.

Hendershott, T. and R. Riordan (2011) “High Frequency Trading and Price Discovery”, Working Paper.

Kirilenko, A., A. Kyle, M. Samadi, and T. Tuzun (2011) “The Flash Crash: The Impact of High Frequency Trading on an Electronic Market”, Working Paper, Commodity Futures Trading Commission.

Malinova, K. and A. Park (2011) “Subsidizing Liquidity: The Impact of Make/Take Fees on Market Quality”, Working Paper.

Menkveld, A, J. (2011) “High Frequency Trading and The New-Market Makers”, Working Paper.

Sofianos, G. (1995). Specialist gross trading revenues at the New York Stock Exchange. Manuscript 95-01, New York Stock Exchange.

Foresight
1 Victoria Street
London SW1H 0ET
www.foresight.gov.uk

URN: 12/1073