Foresight

Market-making obligations and algorithmic trading systems

A feasibility assessment of the March 2012 draft of MiFID2 Article 17(3)

Economic Impact Assessment EIA19

Contents

Market-making obligations and algorithmic trading systems: a feasibility assessment of the March 2012 draft of MiFID2 Article 17(3)

Prof. Dave Cliff, University of Bristol, Bristol BS8 1UB, U.K.

Director, UK Large-Scale Complex IT Systems Initiative (www.lscits.org)

dc@cs.bris.ac.uk

12 May 2011

Summary

This brief report presents an analysis and discussion of the current proposals in the revised, second iteration, of the EU Markets in Financial Instruments Directive (widely known as MiFID2), that oblige certain automated algorithmic and high-frequency trading (HFT) systems to act in a market-making capacity, always posting competitive prices regardless of prevailing market conditions, as laid out in Article 17 Clause 3 of the MiFID2 Draft released in October 2011:

> 17(3) An algorithmic trading strategy shall be in continuous operation during the trading hours of the trading venue to which it sends orders or through the systems of which it executes transactions. The trading parameters or limits of an algorithmic trading strategy shall ensure that the strategy posts firm quotes at competitive prices with the result of providing liquidity on a regular and ongoing basis to these trading venues at all times, regardless of prevailing market conditions.

At the time of writing this report, it is clear that as a result of a European Parliament consultation process, the October 2011 draft phrasing of 17(3) is likely to be subject to significant revisions by mid-2012. The proposed 2012 revisions to 17(3) are introduced and discussed in detail here.

In summary, my opinion is that in both the October 2011 version and the proposed 2012 revision, the text of Article 17(3) is deeply flawed and appears to have been written with good intention but with insufficient thought having been given to the meaning of the words used.

In the course of analysing Clause 17(3) in both its 2011 and 2012 drafts, it becomes clear that the tasks of attempting a detailed interpretation of its aims, and of formulating a firm assessment of its likely costs and potential benefits, are rendered largely meaningless because of the severe lack of specificity or precision in the clause's phrasing. Much of the content is wide open to interpretation, and some of the specific requirements imposed by the clause can very easily be avoided.

For those reasons then, the narrative structure of this report is as follows: first, problems in the non-specific phrasing of the 2011 draft version of Clause 17(3) are discussed; then, the 2012 revisions are introduced and analysed. The ease with which even this, the most recent form of 17(3), can be circumvented or "gamed" is then discussed, with particular reference to the "flash crash" kind of market perturbations that seem likely to have been a major motivating factor behind the introduction and redrafting of 17(3).

The final conclusion of this report is that the current version of 17(3) is in need not only of re-writing, but also of re-thinking. The problems with 17(3) identified here seem to be rooted in a wider, systemic issue: the opacity of the regulation-writing process. There is no formal record of why regulations such as 17(3) are introduced or are subsequently altered: all that is seen in the public domain is the end-result of that process, the successive drafts of the regulations themselves. What is missing from the public domain is a record of the rationale for the regulations, and discussion of the background evidence-base and the reasoning about that evidence. Put simply, the problem is that there is no requirement on regulators to say *why* they

are introducing a specific regulation, nor to argue *how* the regulations will solve whatever problems or address whatever issues it is that they are intended for.

The danger then is that, in fulfilling the desire to be seen to be doing something, regulations are drafted that are tokenistic, difficult to interpret coherently, and largely ineffective in practice because of the ease with which they can be avoided. Unfortunately, it would seem that 17(3) suffers to significant extents from all three of these failings.

Note

In a previous report (Cliff, 2012)[1], I discussed at length the background to, and probable motivation for, the introduction of MiFID2 Article 17, albeit with specific reference to Article 17(2). The "Risk Assessment" section of the earlier report discusses at length recent adverse market events that are likely to have motivated the introduction of Article 17, and for brevity that discussion is not repeated here.

My previous report was finished in January 2012 and discussed the October 2011 draft version of MiFID2. In the text that follows, where it is necessary to distinguish between the October 2011 version of MiFID2 and the European Parliament's proposed revisions to MiFID2 widely circulated in March 2012, I refer to MiFID2-2011 and MiFID2-2012 respectively. Where there are no material differences between the October 2011 and March 2012 phrasings, I refer simply to MiFID2.

1. Objective

The overall objective of MiFID2 Article 17 is not stated explicitly in the text of the clauses that make up the Article, but nevertheless its intentions can reasonably be inferred from MiFID2-2012 Article 51 (on "systems resilience, circuit breakers, and electronic trading") Clause 3, which seems to give the clearest indication for the motivation underlying the introduction of Article 17:

> [51.3] Member States shall require a regulated market to have in place effective systems, procedures and arrangements to ensure that algorithmic or high-frequency trading systems cannot create or contribute to disorderly trading conditions on the market including systems to limit the ratio of unexecuted orders to transactions that may be entered into the system by a member or participant, to be able to slow down the flow of orders if there is a risk of its systems capacity being reached and to limit the minimum tick size that may be executed on the market.

That is, there is a desire to ensure orderly markets, or at least to prevent algorithmic trading systems from contributing to market disorder. As discussed in greater detail in (Cliff, 2012), it seems very likely that events such as the "Flash Crash" in US markets on the afternoon of May 6th 2010, when equity markets went into a sudden free-fall, shedding more than $850billion of value in a less than 15 minutes and yet then regaining most of that loss in the following half an hour, are exactly the type of market disorder that Article 17 is aimed at limiting or preventing.

---

[1] D. Cliff (2012), *Regulatory Scrutiny of Algorithmic Trading Systems: An assessment of the feasibility and potential economic impact.* Report for the UK Government Office for Science Foresight Project on the Future of Computer Trading in the Financial Markets. To be published by the Government Office for Science in late 2012.

While May 6th 2010 was the most extreme such event thus far, it is important to be aware that it is not an isolated instance: as discussed in (Cliff, 2012), in May, June, and July 2011 there were similar events, commonly dubbed "mini flash-crashes" in markets for gold, silver, oil, and natural gas. As this report was being written, a remarkable mini flash crash occurred in the share-price of the American trading-venue operator BATS; made all the more painful because it occurred exactly as shares in BATS were first made available to trade via the company's initial public offering (IPO). As BATS shares opened for trading, their price plummeted from the IPO value of $15.25 to under $0.29 in less than one second (the time taken was around 900ms); and within 1.5 seconds of the IPO, the price quoted for BATS was $0.0002. In less than two seconds, 567 trades had been executed, and trade was then halted; BATS subsequently cancelled its IPO. Following this, there was open speculation that the dramatic crash in BATS share price was not due to a software glitch or "bug", as claimed by the BATS company itself,[2] but rather to a deliberate malicious attack via an HFT system that had been explicitly programmed to cause this failure: the fact that the orders which pushed the price of BATS down seem to have originated from the Nasdaq exchange (i.e., not BATS' own trading-venue), and employed a specific type of order known as an "intermarket sweep order", further fuelling the speculation.[3]

There is widespread concern that the presence of *algorithmic trading* systems (a phrase used to encompass a wide range of different computerised, automated trading systems) in modern financial markets may increase the likelihood of such disorderly events, and/or may accelerate the speed at which such events take place, and/or may increase the amplitude of price swings. Specific concerns are often voiced about *high-frequency trading* (HFT) systems, which employ algorithmic trading methods to operate at speeds that no human trader could reasonably be expected to work at, sometimes issuing hundreds or thousands of orders in a few seconds, and possibly waiting only a second or less between buying and selling some number of units of a financial instrument such as a stock or share or bond, or block of currency or commodity. While all HFT systems are algorithmic, not all algorithmic systems are instances of HFT.

Understanding major market-disorder events such as the Flash Crash or the subsequent mini flash crashes in specific markets is very difficult. There are two reasons for this. First, there is a general culture of secrecy, of the specifics of trading strategies and algorithms and the computer systems that implement them being closely-guarded commercial secrets, so it is very hard to get the complete picture of *who was doing what and when, and why*. Second, national regulatory authorities seem to have been caught somewhat by surprise, and may not have the resources or skills to perform the necessary analysis: in the aftermath of the Flash Crash, the USA's Securities and Exchange Commission (SEC) and Commodities and Futures Trading Commission (CFTC) jointly launched an official inquiry into the events of that afternoon. The

---

[2] See e.g. T. Demos (2012), "BATS pulls its own IPO over 'technical issues'", *The Financial Times*, March 24th, 2012. http://www.ft.com/cms/s/0/073ae952-750c-11e1-90d1-00144feab49a.html; and T. Demos (2012), "IPO software behind BATS' failure". *The Financial Times*, March 26th, 2012. http://www.ft.com/cms/s/0/1f4ccaaa-774c-11e1-93cb-00144feab49a.html.

[3] See e.g. T. Durden (2012), "SkyNet Wars: How A Nasdaq Algo Destroyed BATS", *ZeroHedge*, March 27th 2012. http://www.zerohedge.com/news/skynet-wars-how-nasdaq-algo-destroyed-bats; I. Schmerken (2012), "Did a Rogue Algo Cause BATS IPO Crash?" *Advanced Trading*, April 4th 2012, http://advancedtrading.com/algorithms/232800284; and C. Westfall (2012) "BATS IPO: Accidental Death or Murder?" *The Street*, 6th April 2012; http://www.thestreet.com/story/11486165/1/bats-ipo-accidental-death-or-murder.html.

final joint CFTC/SEC report[4] was not produced until five months after the event, and in the immediate aftermath of the report's publication it became clear that the CFTC/SEC version of events was at best contested and at worst directly contradicted by the "tape" data showing the sequence of events on an order-by-order millisecond-by-millisecond basis. I am not aware of either the SEC or the CFTC having formally responded to the criticisms of their analysis: their position seems to be that the best way of dealing with such criticisms is to ignore them.

The contested issue in the CFTC/SEC Report on the Flash Crash concerns the root cause of the events that day: exactly what triggered the crash? It is beyond the scope of this document to explore that issue in any depth here. Nevertheless, the story that the Report tells of the sequence of events that followed, after whatever the trigger-event was, i.e. the chain of actions that then drove prices further down and subsequently also gave rise to the remarkable recovery, is less contested. The Report claims that algorithmic trading systems (commonly referred to in the industry as "algos") interacted with one another and unexpectedly engaged in a "hot-potato" pass-the-parcel dynamic where blocks of shares were passed from algo to algo, giving rise to transaction-volumes that gave the superficial impression that the market was still liquid when in fact much of the real liquidity in the market had disappeared or "dried up" because both algorithmic and human traders, confused and alarmed by the unprecedented events, had withdrawn from the market.

Although the text of MiFID2 does not provide any narrative on why it is written as it is, intuitively it seems very likely indeed that Article 17 is a response to such recent events. Article 17(2), already discussed at length in (Cliff, 2012), seems directly intended to address the issue of regulators not knowing who was doing what and when and why. On the other hand, Article 17(3) seems likely to have been motivated by the commonly held belief that in the drying up of liquidity was a major problem in the Flash Crash.

2. Background

2.1. MiFID2-2011: The October 2011 Draft

MiFID2-2011, Article 17 carries the title "Algorithmic Trading". For completeness, and to give some context, I reproduce here the entire text of the first three clauses, 17(1) to 17(3), although it is only 17(3) that will be examined in depth below:

17(1) An investment firm that engages in algorithmic trading shall have in place effective systems and risk controls to ensure that its trading systems are resilient and have sufficient capacity, are subject to appropriate trading thresholds and limits and prevent the sending of erroneous orders or the system otherwise functioning in a way that may create or contribute to a disorderly market. Such a firm shall also have in place effective systems and risk controls to ensure the trading systems cannot be used for any purpose that is contrary to Regulation (EU)No.(MAR) *[i.e., Market Abuse Regulation]* or to the rules of a trading venue to which it is connected. The firm shall have in place effective continuity business arrangements to deal with any unforeseen failure of its trading systems and

---

[4] CFTC & SEC (2010b). *Findings Regarding the Market Events of May 6th, 2010.* Report of the staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory issues. September 30th, 2010. http://www.sec.gov/news/studies/2010/marketevents-report.pdf

shall ensure its systems are fully tested and properly monitored to ensure they meet the requirements in this paragraph.

17(2) An investment firm that engages in algorithmic trading shall at least annually provide to its home Competent Authority a description of the nature of its algorithmic trading strategies, details of the trading parameters or limits to which the system is subject, the key compliance and risk controls that it has in place to ensure the conditions in Paragraph 1are satisfied and details of the testing of its systems. A competent authority may at any time request further information from an investment firm about its algorithmic trading and the systems used for that trading.

17(3) An algorithmic trading strategy shall be in continuous operation during the trading hours of the trading venue to which it sends orders or through the systems of which it executes transactions. The trading parameters or limits of an algorithmic trading strategy shall ensure that the strategy posts firm quotes at competitive prices with the result of providing liquidity on a regular and ongoing basis to these trading venues at all times, regardless of prevailing market conditions.

Clause 17(3) has become known colloquially as the "market-maker obligation", because of its apparent implication that under MiFID2, algorithmic trading systems will be required to provide liquidity in much the same way as firms or individuals known as "market-makers" do (market-makers are also known as "specialists" in the New York markets). Market-makers hold an inventory of some financial instrument and are expected to always quote a price for buying (the bid-price) or for selling (the ask-price or the offer-price) that instrument, making their profits by the difference (or "spread") between the bid and offer prices. To people outside the finance industry, exposure to market-maker businesses comes most frequently in high-street foreign-exchange retail outlets: these shops hold a stock of various currencies, and may for example sell US$1.00 for €0.80 but would then buy that one dollar back for only €0.70; so their profit comes from the bid-offer spread of €0.10.

It is important to note that Clause 17(3) represents a significant shift from the status quo. Currently there is no *requirement* on algorithmic trading systems to act as market-makers. For sure, if a company or individual wishes to create an algorithmic market-maker trading system, they are free to do so, but they are also free to create algorithmic trading systems that operate in other ways, ways that would not be described as market-making. For instance, some algorithmic systems may exist purely to silently monitor a market ("lurk") and then move quickly to enact trades ("snipe"), buying units of some financial instrument that the algorithm judges to be temporarily underpriced (relative to a longer-term statistical average), and then selling them when the price anomaly reverts to its average. Other algorithms exist purely to automate the execution of large-quantity orders, for instance '"slicing" a single buy-order for 100,000 shares in a specific company into 1,000 orders each for 100 shares, to be sent to the exchange over some specified period of time, the small size of each order making it easier to execute without causing an unfavourable shift in the price. Under any reasonable interpretation, it seems that 17(3) will prevent the future deployment of, or at least require radical changes in, existing algorithmic systems such as the "lurk-and-snipe" or "slicer" examples.

It is worth examining the phrasing of Clause 17(3) in some depth, attempting to understand its motivation, exploring any implicit assumptions, and identifying any potential difficulties that may arise in attempting to actually implement the regulation and to enforce it. In the following sections, we explore the key phrases in 17(3) and discuss each in turn.

2.1.1 An algorithmic trading strategy shall be in continuous operation...

Algorithmic trading systems are almost always implemented as high-availability computer systems: algorithmic trading firms do not want a power cut or component failure to suddenly put a halt to their automated trading activities midway through a trading period, with the manifest potential for incurring significant losses (or reduced profits) from missed trading opportunities.

But the phrasing of 17(3) is not addressed at the *system* level. Rather, it is specific to individual *strategies*. It seems to be saying that an individual strategy such as "lurk-and-snipe" or "slicer" examples (introduced above) are required to be in continuous operation, for as long as the market in which it operates is open. It seems likely that this is an attempt to prevent algorithmic trading systems withdrawing from the market in times of turmoil, as happened in the Flash Crash.

In practice, unless an algorithm's specification requires it to continuously post prices (such as if it was a market-maker), it may be impossible to detect whether an algorithm is indeed in continuous operation or not. Consider the "lurk-and-snipe" strategy: it may in principle run continuously from midnight to midnight, watching the market 24 hours per day seven days per week, but it could be that the statistical anomalies that it seeks as trading opportunities only ever manifest themselves for a few seconds at market-open, at the start of each new business day. In that case, if the operators of such a strategy decided to save electricity by only running the strategy for 5 minutes per day, its externally observable behaviour would be identical. So the fact that it is not "in continuous operation" would simply not be detectable.

Indeed, in principle any algorithmic strategy could be programmed to suspend its trading activities in times of market stress or uncertainty or high volatility or whatever the algorithm's designers deem to be an unfavourable situation, and still not violate this aspect of 17(3). If, while the strategy is in its suspended state, it remains operational by passively monitoring the market so that it can identify when the unfavourable situation has ended, and thereafter returns to active trading in the market, then it could reasonably be argued that the strategy has nevertheless been in continuous operation throughout the unfavourable period, and so does not violate 17(3).

Furthermore, it is common practice for firms that operate algorithmic and high frequency trading systems to actually deploy a number of different strategies simultaneously, each of which generates signals to buy or sell various instruments with various constraints on what constitutes an acceptable price, and acceptable quantity, and an acceptable deadline by which the transaction should be executed. But, crucially, the trading signals from all the firm's different strategies will be aggregated, reconciled, and optimised by an order management system (OMS). So, for example, if the firm's strategy S1 generates a signal to buy 100 units of shares in XYZ, and at the same the firm's strategy S2 generates a signal to sell 150 XYZ shares, then the firm's OMS should internally transfer 100 units of XYZ from S2's portfolio to S1's and only actually sell the unwanted 50 units, thereby saving transaction costs at the trading venue. The phrasing of Clause 17(3) seems to be entirely ignorant of this: the wording is in terms of individual *strategies*, not trading *systems* or *platforms*; the problem in implementing 17.3 will be that the actions of individual strategies may simply not be observable from a viewpoint external to the firms that are operating those strategies, because so much of the activity of individual strategies is masked by each firm's OMS.

2.1.2 ...the strategy posts firm quotes...

The desire for "firm quotes" is presumably targeted at specific HFT practices that are viewed by many as of highly questionable value, and by some as potentially constituting market abuse. Two examples of such practices are "flash quoting", and "noise quoting".

Flash quoting is where an HFT algorithm sends a quote (an order to buy or sell) to a market, followed very shortly later by a cancellation of that order – the time that the order is valid on the market may last only a few milliseconds. Flash quotes are used as "teasers" or "bait", because they can prompt other traders in the market to reveal their intentions, and thereby give the HFT system that issued the flash quote a significant indication of the market's underlying levels of supply and demand, with little danger of the flash quote actually being accepted by a counterparty because it is so short-lived. In this way, the HFT system issuing the flash quote has the possibility of gaining useful market information. In a situation where there are multiple HFT systems issuing flash quotes into a market, the job of forming an understanding of what is currently a fair price for a transaction rapidly increases in complexity: the true signals of market intent are masked by the flickering "noise" of the flash quotes.

This observation has led to another tactic that is also viewed by many as market abuse: the deliberate and systematic flooding of a market with "noise" quotes that are not intended to do anything other than obscure the genuine signals of intent to buy or sell. If an HFT system is programmed to produce hundreds or thousands of orders per second, none of which are actually intended to result in a transaction, and which systematically vary in some fashion, then by the time those quotes appear on the publically-viewable market-data feeds, they are anonymized and mixed in with other traders' genuine quotes (which have also been anonymized). Because the noise-quoting HFT system introduced a systematic variation in the noise quotes, those quotes can easily be identified and stripped out of the market data by the one HFT system that generated them. So that noise-quoting HFT system can form a clear picture of the true supply and demand in the market, while the view of the market formed by all other traders is obscured by the "smokescreen" of noise quotes that the HFT system produced, the sheer volume of which may be overwhelming for some traders.

While the desire to limit or eliminate practices such as flash-quoting and noise-quoting is understandable, again there is difficulty in attempting to implement this aspect of the Directive, because so much will depend on what is meant by a "firm quote". To enforce the requirement of posting firm quotes, there must be some firmness criteria by which quotes can be judged, to determine whether they are firm or not. Exactly how does one measure firmness? One possibility would be to impose *minimum resting times*, i.e. to require that orders, once issued, are required to stay active in the market and cannot be cancelled until they have "rested" in the market for some minimum period of time. The process of deciding how best to implement minimum resting times as a mechanism for ensuring firmness of orders, and how to determine what the minimum time-period should actually be, is not discussed in MiFID2. Yet clearly if the time-period is set too low (one femtosecond, say: a femtosecond is one millionth of one billionth of a second) it will have no effect, and if it is set too high (a year, say) it could have obvious negative impact on market dynamics and market efficiency. I expect any market practitioner to agree that one femtosecond is too short and one year is too long, but I also expect widespread disagreement and protracted debate over exactly where, between those two extremes, the line should be drawn.

2.1.3 …at competitive prices…

There is no definition within MiFID2 of what it means for a price to be "competitive". Without precise specification of how to decide whether a price is competitive or not, the potential impact of this requirement is simply impossible to judge. In a similar fashion to the problem of agreeing a minimum resting time somewhere along the spectrum from one femtosecond to one year, the problem of deciding whether a quote price is competitive will most likely be reduced to whether the difference between the quote price, and the current best bid/offer price, is greater or lesser than some threshold value, possible expressed as a percentage of the best bid/offer price.

A precedent for this exists in the American equity markets, where in 2010 the SEC approved new rules that banned the use of "stub quotes", or resting price orders, in automated trading systems.[5] The ban was implemented by insisting that traders only issue quotes that are within a specified threshold percentage (8%, 20%, or 30% depending on various factors, and with each threshold being allowed to "drift" by up to 1.5% in practice) of the national best bid and offer (NBBO) price. Prior to the SEC's ban on stub quotes, operators of algorithmic systems that had been designed to meet a requirement to always post prices for bids and offers in an instrument would use extreme values, way outside the normal price range for units of that instruments, to signal no interest in transacting while still meeting the strict requirement of posting two prices. So, for example, in equity markets where a single stock would usually trade in the range $10-$1000, the stub quotes might be a bid price of $0.01 or an offer price of $100,000.00. In the Flash Crash of May 6th 2010, so much liquidity disappeared from the market that such stub quotes became exposed as the best prices in the market, and algorithmic systems then briefly executed transactions at these prices, prices that would have been recognised as nonsensical by a human trader, with the consequence that shares in major blue chip companies that had previously traded for a few tens of dollars suddenly dropped to a single cent, or rocketed to £100k.

The SEC's stub-quote rule has been criticised by some market practitioners and commentators, who have observed that since its introduction, there has been no evidence of it having actually been enforced – and an unenforced rule is of absolutely no use in preventing a recurrence of the type of event it was designed to prevent. For instance, Nanex Research report[6] that in a single period of four consecutive trading days in August 2011 there were an average of more than one million quotes successfully posted each day that violated the stub quote ban, and that as far as Nanex can discern, these four days were not exceptional; rather, Nanex imply, they represent the norm.

Also it should be noted that while a competitive-pricing rule that limits quotes to within a specified percentage of the NBBO will (if adequately enforced) prevent the kind of stub-quote phenomena witnessed in the 2010 Flash Crash, such rules cannot prevent precipitous drops in the price of an instrument. For example, if the current best bid price is $10.00 and the competitive-price percentage threshold is set to 20%, then a bid of $8.00 is permissible. If that $8.00 bid becomes the best bid, then a bid 20% below that, i.e. of $6.40 is then permissible. In

---

[5] See e.g. T. Demos & J. Grant (2010), "SEC passes 'stub quote' rules", *The Financial Times*, 8th November 2010. http://www.ft.com/cms/s/0/fab7c5be-eb8d-11df-bbb5-00144feab49a.html; see also SEC (2010), *SEC Approves New Rules Prohibiting Market Maker Stub Quotes*, press release, http://www.sec.gov/news/press/2010/2010-216.htm.

[6] See http://www.nanex.net/Research/StubRuleViolations/StubViolations.html.

a short sequence of trades, the price can decay logarithmically and rapidly reduce to pennies: the sequence in this example would continue from $6.40 with prices of $5.12, $4.10, $3.28, $2.62 and so on, down to $0.04 after a sequence of only 25 transactions. According to research published by Nanex,[7] one of the truly extraordinary facets of the March 2012 mini flash crash in BATS' IPO stock mentioned in Section 1 is that the price crash followed exactly this style of progressive percentage reduction; there, a sequence of 567 transactions took the price from $15.25 to $0.0002 in an almost perfect exponential decay lasting roughly two seconds. Even if it had been enforced with maximum rigour, the SEC's stub quote rule would have had absolutely no effect whatsoever on slowing the BATS crash.

2.1.4 …with the result of providing liquidity on a regular and ongoing basis…

There is an implicit assumption here that posting firm quotes at competitive prices is sufficient to provide liquidity. Now it may be that the intention of Clause 17(3) is to ensure that liquidity is provided on a regular and ongoing basis, but the current phrasing of the Clause is seems to be wide open to "gaming". That is, HFT systems can readily be constructed that obey the letter of the law, but not its spirit. The simplest way of gaming this aspect of 17(3) would be for the HFT system to continuously post firm quotes at competitive prices, thereby complying with 17(3), but for the HFT system to greatly reduce the sizes of its competitively-priced orders in times of market stress.

For instance, an algorithmic trading system could be programmed to show a competitive price but at a very low volume, say 1% of the systems holding, while at the same time showing a different, non-competitive price for the remaining 99% of the volume; that system would, it seems, not be in violation of Clause 17(3). Such practice is already well known in today's markets: algorithmic and HFT systems can readily be programmed to issue so-called "compliance quotes" when necessary. In this example, the quote showing a competitive price for 1% of the holding is a compliance quote: that is, it is issued solely to ensure that the trading system complies with the regulations.

In the limit, an HFT system can continuously post firm quotes at competitive prices for only a single share, or a single dollar's worth of currency, say, and (because Clause 17(3) says nothing about the sizes of orders) that HFT system would be compliant with Clause 17(3). If each HFT trader is quoting a firm and competitive price for a size of one, then that really is not a liquid market at all. In its current phrasing, Clause 17(3) can deliver exactly this. It seems highly unlikely that this is what the authors of 17(3) desired or intended.

2.1.5 …at all times, regardless of prevailing market conditions.

This phrase in Clause 17.3 is curious, as a strict interpretation of it would require that algorithmic trading systems are capable of operating at all times, posting firm quotes at competitive prices, even in the most extreme conditions, and could only cease their operations if the markets themselves are closed. It seems to prevent the owners of an algorithmic trading system from ever switching it off or taking it out of commission even if only for a period of scheduled maintenance.

---

[7] See http://www.nanex.net/aqck/2970.html.

Algorithmic trading systems are commonly constructed with a "kill-switch" or "scram" facility, activation of which immediately puts the system into a shutdown sequence, possibly liquidating as much of the algorithmic system's portfolio in the process. This is a basic safety feature: if a problem with an algorithmic trading system manifests itself and it starts to generate major losses, or otherwise contributes to market disorder, the owners of that system need to be able to close it down as quickly as possible, as an act of self-preservation and also possibly for the good of other market participants.

There is no discussion of the potential penalties for failing to operate at all times regardless of prevailing conditions, but the costs and risks of producing an algorithmic system that can be assured to operate at all times regardless of prevailing conditions have the potential to be truly prohibitive.

2.1.6 Summary

This discussion of the 2011 draft of Clause 17(3) has hopefully demonstrated that several devils lie in the details of its 76 words. It suffers from key words and concepts being used without any care having been taken in defining those words or concepts, and from failing to consider how those words and concepts relate back to actual industry practice. As we shall see in the next section, the 2012 revisions proposed by the European Parliament go some way to rectifying this, articulating a particular definition of what constitutes high-frequency trading. Nevertheless, the 2012 revisions are only a partial remedy to the problems identified here.

2.2. MiFID2-2102: The European Parliament's Amendments, March 2012

In mid-March 2012, The European Parliament's Committee on Economic and Monetary Affairs produced a draft report on its proposed amendments to MiFID2-2011. The report carries the name of the designated rapporteur for the Parliament's consultation process, Markus Ferber MEP, and the report is already being called the "Ferber Report" by industry practitioners.

The pages of the substantive parts of the Ferber Report are typeset in two columns: the left-hand column shows text quoted from MiFID2-2011, with italicized text indicating areas that may need correction; the right-hand column shows the Parliament's amendment to the draft, with the new text highlighted in bold italics. This layout enables direct comparison of the original and amended versions of the Directive. The amended text is what is referred to in this report as the MiFID2-2012 draft.

As with the discussion of MiFID2-2011 in Section 2.1, in the text that follows we will discuss the salient changes in sequence.

2.2.1 Article 17, Title & Clause 17(1): addition of HFT

In MiFID2-2012, Article 17 is renamed from "Algorithmic Trading" to "Algorithmic and High-Frequency Trading", and correspondingly the phrase "high-frequency trading" is added to the text of Clause 17(1) but is otherwise not substantively altered.

2.2.2 Clause 17(2): no change

No amendments whatsoever are proposed to the text of Clause 17(2).

2.2.3 Clause 17(3): restricted to a particular style of HFT

Significant changes have been made to Clause 17(3), primarily to the opening sentence. The text of MiFID2-2012 reads as follows (with the original emphasis):

> 17(3) ***Where an investment firm engages in a high-frequency*** trading strategy ***which meets the conditions of Article 4(30b)(vi) that strategy*** shall be in continuous operation during the trading hours of the trading venue to which it sends orders or through the systems of which it executes transactions. The trading parameters or limits of ***such a high-frequency*** trading strategy shall ensure that the strategy posts firm quotes at competitive prices with the result of providing liquidity on a regular and ongoing basis to these trading venues at all times, regardless of prevailing market conditions.

From this it is clear that the breadth of applicability of 17(3) has been greatly reduced: in MiFID2-2011, Clause 17(3) was phrased to apply to *any* and *all* algorithmic trading systems; now it applies only to HFT systems meeting a specific set of conditions laid out in MiFID2 Clause 4(30b)(vi), which is a new clause, one that was not present in MiFID2-2011. Other than that, all of the key phrases that were discussed above in Sections 2.1.1 to 2.1.5 remain unchanged: the European Parliament have not noticed any of the problems highlighted above.

Before moving on to explore the definition of HFT laid out in MiFID2 4(30b)(vi), it is useful to note that in MiFID2-2012, the lack of any substantive changes to Clauses 17(1) and 17(2) means that ***all*** automated, algorithmic, and HFT systems are expected to comply with 17(1) and 17(2), but 17(3) is now intended ***only*** to apply to the subset of HFT systems that meet the precise definition of HFT introduced in MiFID2-2012 Clause 4(3) and the particular conditions specified in MiFID2-2012 Clause 4(30b)(vi).

2.2.4 Clause 4(30): defining HFT; defining the scope of 17(3).

The full text of MiFID2-2012 Clause 4(30b), all of which is entirely new text (and hence typeset in bold italics in the Ferber Report) relies on a definition of HFT that is given in Clause 4(30a), which is also entirely new text. The text for 30a and 30b is as follows:

> 30(a) 'High frequency trading' means trading in financial instruments at speeds where the physical latency of the mechanism for transmitting, cancelling, or modifying orders becomes the determining factor in the time taken to communicate the instruction to a trading venue or to execute a transaction.

> 30(b) 'High frequency trading strategy' means a trading strategy for dealing on own-account in a financial instrument which involves high-frequency trading and has at least four for the following characteristics:

> (i) it uses co-location facilities;

> (ii) it relates to a daily portfolio turnover of at least 50%;

> (iii) the ratio of orders to trades exceeds 4:1;

> (iv) the proportion of orders cancelled exceeds 20%;

(v) the majority of positions taken are unwound within the same day;

(vi) over 50% of the orders or transactions are made on trading venues offering discounts or rebates to orders which provide liquidity are eligible for such rebates.

Clause 4(30b)(vi) seems to have been written or translated in something of a hurry, as strictly it is not grammatical: the closing phrase "are eligible for such rebates" needs at least to be preceded with a conjunctive such as "and" or "or". It seems reasonable to assume that authors' intention was that the text here should actually be something like this:

(vi) over 50% of the orders or transactions are made on trading venues offering discounts or rebates to orders which provide liquidity, **and those 50% or more of its orders or transactions** are eligible for such rebates.

The wording of 4(30b)(vi) may be difficult to follow for anyone who is not familiar with the way in which trading venues levy fees to their customers, the traders who seek to make a transaction. In a fairly recent development, a number of major exchanges and trading venues have introduced what is known as a "maker-taker" pricing system where all customers are charged a fee per order or per transaction, but where the fee that a customer pays is dependent on whether the customer's order or transaction involved that customer acting as a provider of liquidity to the market (a "maker") or as a consumer of the available liquidity (a "taker"). To encourage liquid markets, maker orders are charged less, either by being levied a discounted transaction fee, or by paying the full fee (as is also paid by taker orders) but then being given a rebate. In essence, maker orders and transactions are ones where the customer quotes a specific price at which a specific quantity of some instrument will be bought or sold, whereas a taker order or transaction is one that executes against a price already quoted. Traditionally, the role of "specialists" and "market-makers" was to provide a steady stream of maker orders to the markets in which they operated. Maker-taker pricing is discussed in depth by Angel, Harris, & Spratt (2010).[8]

With that context established, Clause 4(30b)(iv) could be rephrased as:

(vi) over 50% of the orders or transactions are made on **maker-taker** trading venues, and those 50% or more of its orders or transactions are eligible for such rebates.

But an HFT system's orders or transactions are only "eligible for such rebates" if the HFT system is operating as a liquidity-provider, i.e. as a market-maker. So 4(30b)(iv) could be rephrased as:

(vi) over 50% of the orders or transactions are made on maker-taker trading venues, and **they are market-making** orders or transactions.

Which, with a little polishing, can be rewritten as:

---

[8] J. Angel, L. Harris, & C. Spratt (2010). *Trading in the 21st Century.* Unpublished manuscript. http://www.sec.gov/comments/s7-02-10/s70210-54.pdf.

(vi) the majority of its orders or transactions are market-making, made as liquidity providers on maker-taker trading venues.

Now recall that the reason why Clause 4(30b)(iv) is so significant is that the "market-making obligation" in Clause 17(3) *only* applies to HFT systems that exhibit the characteristic specified in 4(30b)(iv). So, in principle we can substitute the revised, abbreviated text of 4(30b)(iv) directly into the text for Clause 17(3), giving:

17(3) Where an investment firm engages in a high-frequency trading strategy for which ***the majority of its orders or transactions are market-making, made as liquidity-providers on maker-taker trading venues,*** that strategy shall be in continuous operation during the trading hours of the trading venue to which it sends orders or through the systems of which it executes transactions. The trading parameters or limits of *such a high-frequency* trading strategy shall ensure that the strategy posts firm quotes at competitive prices with the result of providing liquidity on a regular and ongoing basis to these trading venues at all times, regardless of prevailing market conditions.

So, in essence, the revised MiFID2-2012 Clause 17(3) says that *whenever* an investment firm operates a market-making HFT strategy, that strategy will be obliged to *continuously* operate as a market-maker, posting firm quotes at competitive prices, providing liquidity at all times regardless of market conditions.

This is greatly reduced in scope from the MiFID2-2011 original phrasing of Clause 17(3): the 2011 version, interpreted literally, applied the market-making obligations to *all* algorithmic trading systems. The MiFID2-2012 revision applies the market-making obligations *only to HFT systems that are operating as market-makers*.

This may appear to be a circular requirement, but it is not: what it says is that if an HFT system ever spends the majority of its time acting as a market-maker (and receiving the discounts or rebates that market-makers enjoy) then it is required to act as a market-maker all the time, and cannot suddenly stop being a market-maker if things start going badly. It seems reasonable to assume that this is intended by the authors of MiFID2-2012 as an attempt to prevent the occurrence of events such as the withdrawal of real liquidity that took place during the Flash Crash of 2010.

As with any such regulation, it effectiveness will be significantly determined by the ease with which the regulation can be avoided. Avoidance of Clause 17(3) is the topic of the next section.

2.3. Summary

The March 2012 amendments to MiFID2 take care to precisely define what is meant by HFT and to greatly reduce the scope of Clause 17(3), limiting it only to those HFT systems that operate as market-makers on trading venues with maker-taker pricing systems. Although the amended MiFID2 is therefore, in that specific respect, more readily implementable than the original, the deeper problems arising from the phrasing of MiFID2-2011 that were highlighted in Sections 2.1.1 to 2.1.5 all remain in the MiFID2-2012 draft.

3. Ease of avoidance

If the MiFID2-2012 version of Clause 17(3) is enacted, then that clause's dependence on the specific HFT characteristic introduced in Clause 4(30b)(iv) means that Clause 17(3) affects *only* HFT strategies for which 50% or more of the orders or transactions are subject to liquidity-making discounts or rebates from the trading venue to which the orders are sent or on which the transactions occur.

The simplest way of avoiding the requirement of 17(3) then, is to write an HFT strategy that never issues more than 50% of its orders or transactions as qualifying for maker discounts or rebates on a trading venue with a maker-taker pricing structure. In principle, an HFT strategy could be constructed to keep a count of how many such orders it has issued, and to act in such a way that it keeps the proportion of its total orders that are liquidity-makers to under 50%. Three ways in which this could be done are:

- By also issuing sufficiently many "taker" orders on the same trading venue that the HFT strategy has always issued more taker orders than maker orders.

- By also issuing sufficiently many orders to a different, non maker-taker, exchange, that the HFT strategy's total proportion of orders qualifying for maker discount is always less than 50%.

- By simply issuing no more orders once the proportion of maker orders has reached 49.9%, orders and then ceasing to operate.

The last of these three options is probably the most readily practicable. An investment firm runs an HFT strategy, and when that strategy's proportion of market-making orders reaches 49.9%, the strategy simply ceases to operate, closing down its operation and handing any holdings of financial instruments that it has built up while it was operating back to the HFT firm's overall portfolio, the firm's pooled holdings from all its trading strategies.

Some time later, the firm can then start operating another HFT strategy, and give it some initial allocation of financial instruments, a starting portfolio. That HFT strategy can then be in continuous operation until its proportion of maker orders or trades reaches 49.9%, and it then shuts itself down and "hands back" its holdings to the firm's overall pool.

Of course, for high frequency trading systems, "some time later" does not necessarily have to be much time at all. Say, for the sake of argument, that the pause between the cessation of the first HFT strategy, and the commencement of the second, was one second. For a whole second (which in HFT terms is actually quite a long period of time) the firm is operating no HFT strategy, but otherwise it always has a strategy running. Furthermore, say, for the sake of argument, that actually the second HFT strategy that the firm operates is an identical copy of the first one. Technically, they could be argued to be different strategies (e.g., they would have different process-identifier numbers in the computer operating system that is running the HFT algorithms). This means that the firm can run a sustained sequence of "clone" instances of the same HFT strategy, running each instance until its proportion of maker orders or trades reaches 49.9%, then closing that instance down, pausing for a brief rest period, and then re-starting a new instance of the same strategy. In essence, the firm is continually operating the same strategy, but with brief one-second breaks. Note that the one-second duration of a break

is just a number plucked from the air for the sake of this discussion: it could just as easily be a millisecond, one thousandth of a second. From the external perspective of a human observer of, or participant in, the market, this HFT system would be in continuous operation. And yet, because the firm's HFT strategies never exceed the 50% threshold specified in Clause 4(30b)(iv), they are never subject to the requirements of Clause 17(3), and so this firm's HFT strategies are not required to post firm prices, nor competitive prices, and they can exit market at any time the firm wishes them to.

There are of course other ways in which MiFID2-2012 Clause 17(3) could be avoided: to qualify for regulation by Clause 17(3), an HFT system needs to not only meet the market-maker condition spelled out in Clause 4(30b)(iv), but *also* has to meet at least three of the conditions specified in Clauses 4(30b)(i) to 4(30b)(v). Each of Clauses 4(30b)(ii), (iii), and (iv) seem similarly vulnerable to being avoided or "gamed" in much the same way as I have here shown Clause 4(30b)(iv) to be.

4. Conclusion

The MiFID2-2012 version of Clause 17(3) is greatly restricted in scope in comparison to the original 2011 draft. The restrictions are sensible, but reduce Clause 17(3) to requiring those HFT systems that *ever* predominantly engage in market-making to *always* do so. The lack of any supporting rationale or narrative for these clauses in MiFID2 means that it is necessary to make informed guesses or estimates about what the new regulations are intended to achieve: it seems likely that Clause 17(3) is intended to ensure that market-making HFT systems do not withdraw their liquidity in times of crisis. Nevertheless, the current phrasing of the regulations is such that programming self-monitoring HFT strategies that can avoid the regulation is relatively straightforward, and hence this clause in Article 17 is unlikely to have the desired effect.

The largely unenforceable nature of the current versions of 17(3) means that, in the absence of further significant revision, it is likely to have very little impact. If the authors of MiFID2 had been more explicit about what they intended to achieve with the introduction of 17(3), it would be easier to recommend a more precise phrasing, preferably one that introduces measures whose effectiveness can be demonstrated by appeals to relevant empirical evidence and/or by logical reasoning. But, in the absence of a statement of intent from the authors of MiFID2-17(3), further assessment of the risks or impact of modified versions of that clause would be purely speculative.

Certainly, if the intention really is to introduce legislation that prevent the flight of liquidity in times of market stress, then the authors of MiFID2 seem to not be mindful of the extent to which such drying up of liquidity is a phenomenon that long pre-dates the introduction of algorithmic and high-frequency trading. In earlier market crashes, traders simply stopped answering their phones; 'twas ever thus.

The medieval ruler King Cnut is popularly known for setting his throne on the sea-shore and commanding the tide not to rise, yet he ended up with wet feet. The authors of MiFID2 should be mindful that even the most well-intentioned of efforts are likely to be viewed with scepticism by others who recognise the futility of that effort.

If a market's legislation forces traders to provide liquidity even in circumstances where no sane trader would want to do so, traders are likely to be discouraged from entering that market and

would most likely go elsewhere. Thus, in attempting to ensure liquidity is provided in occasional times of extreme stress, badly-written rules can mean that overall liquidity is reduced even in the good times. Sometimes, the best course of action is to do nothing. Wholesale deletion of 17(3) from MiFID2 would certainly save everyone a lot of time and effort.

Acknowledgements