18th May 2015

Briefing notes principally on the 6th May 2010 Flash Crash and its Implications for Non-Live Testing

Nicholas Hallam and Nick Idelson, TraderServe Limited, 12-16 Laystall Street, London, EC1R 4PF,
e: research@traderserve.com, w: www.traderserve.com, t: 02071935717.

On Tuesday 20th April Navinder Singh Sarao was arrested for market manipulation that was deemed "significantly responsible" for the 6th May 2010 flash crash of the US equity and index futures indices.

Clear evidence of "spoofing" can indeed be seen in the June 2010 S&P 500 e Mini futures on that day (see Fig.1). By "spoofing" we refer to the practice of placing large orders near the top of the order book – i.e. just under the best bid for buy orders, just over the best offer for sell orders – with the intention of pressuring the market but without the intention actually to trade with these orders. Sarao's spoofer was more complex than many, placing very large sell orders at a number of levels above the best offer, and moving them up or down as the price moved. This is called "layering" and requires an algorithm to manage the multiple orders to prevent their being filled. Having caused the market to move some way down, he bought contracts and cancelled the sell orders. Given the size of his layering orders it is highly likely that he did cause the market to move down – at times his orders constituted more than 50% of the selling interest in the entire market – but there has to be some doubt whether he should be blamed entirely for the whole flash crash. The reason for this is that he stopped "layering" a few minutes before the most precipitous declines occurred (see Fig.2).

In their Sep 2010 report on the incident - https://www.sec.gov/news/studies/2010/marketevents-report.pdf - the SEC and CFTC had blamed a large institutional sell order for 75,000 e Minis ($4.1 billion) as the principal cause. This order was fed into the market by an execution algorithm which scaled the amount it traded in response to changes in recent volume. As there was already a lot of activity in the market (though not a lot of liquidity) the algorithm was very heavy and forced the market sharply down. There were three main problems with attaching principal blame to the institutional order: (1) the algorithm did not trade when the market was falling most steeply ; (2) the algorithm was a liquidity provider trading only passively – i.e. leaving resting orders in the market above the best offer - rather than a liquidity consumer trading aggressively – i.e. crossing the spread to trade instantly with resting buy orders; and, consequently (3) the heaviest trading it did was **after** the flash crash when the market was recovering. So while the large institutional order may well have contributed to the general disorder there were clearly other factors involved.

The report also mentioned very heavy aggressive selling by High Frequency Traders (HFTs). The algorithms employed by HFTs normally do not allow large exposures to develop, but during the early falls many of them accumulated very large long positions. To reduce these exposures they traded aggressively, consuming liquidity on the buy side of the market, often taking all available liquidity from several levels of the order book. The report saw this as contributing to the problem but not its main cause.

A further source of disruption at the time was identified in a Nanex research report a month after the crash - http://www.nanex.net/20100506/FlashCrashAnalysis_Intro.html. This was the

presence of “quote stuffing” in the equity markets on a prodigious scale. Quote stuffing is the frequent adding and removing of resting orders some way from the best bid and offer. In the Nanex analysis of the 6th May data they found hundreds of examples where the number of quote updates for a single stock on a single exchange was greater than 1000 in one second. This activity was almost certainly responsible for lags in the data from some exchanges which contributed to the mayhem. The rationale behind the activity is to increase the processing requirements of participants and to cause latency, both of which benefit the participants with the fastest systems. Despite the clear evidence that Nanex produce regularly on the practice the very existence of quote stuffing is denied by many in the industry.

To recap: there was a layering algorithm operating until a few minutes before the crash, an execution algorithm trading too heavily in low liquidity conditions, HFT activity which involved excessive liquidity consumption and on the equity exchanges, it seems, quote stuffing contributing to delays which exacerbated the disorder.

None of this algorithmic activity would have been present in the live markets had suitable non-live testing been mandatory at the exchanges in question as European regulators are proposing for European trading venues as part of the MiFID II consultation. Such testing would be in a created market with microstructure that is as realistic as possible and with the facility of producing stressed conditions.

Research at TraderServe has shown that that both “Spoofing” and “Quote Stuffing” are detectable in non-live testing. It also shows how an insufficiently tested execution algorithm could contribute to market disorder when exposed to extreme market conditions and that this could also be induced in non-live testing where a realistic, responsive market is created and algorithmic behaviour examined in stressed conditions of high volatility or dramatically reduced liquidity. The research is reported in "*Trading Algorithms, Disorderly Markets and Non-Live Testing A study of emergent behaviours supporting the case for non-live testing regulations* " - http://www.traderserve.com/download.php?file=publicdomainresearch/Trading%20Algorithms-Disorderly%20Markets-Non-Live%20Testing-20141202-researchpaper_final.pdf.

The HFT activity merits closer attention. The presence in the markets of High Frequency Traders is controversial. Although most claim to be liquidity providers aiding efficient price discovery and reducing the effective costs for other market participants, it is noteworthy that in crash conditions like those obtaining on 6th May 2010 they have been seen to contribute to excessive liquidity *consumption* by “dumping” large accumulated positions in marketable orders that break through many price levels. It is this activity that endangers market stability whatever the merits of HFT in other circumstances. This behaviour could also be induced in non-live testing and the liquidity consumption rates be used by the exchange as a basis for barring unsuitable algorithms from its live markets.

One further point: it is very possible that none of the algorithms would on its own have caused the flash crash. It is almost certain that it was the combined activity of more than one algorithm that caused the damage. It is very important, therefore, that non-live testing should explore the behaviour of multiple algorithms together and their disposition to create emergent market disorder.

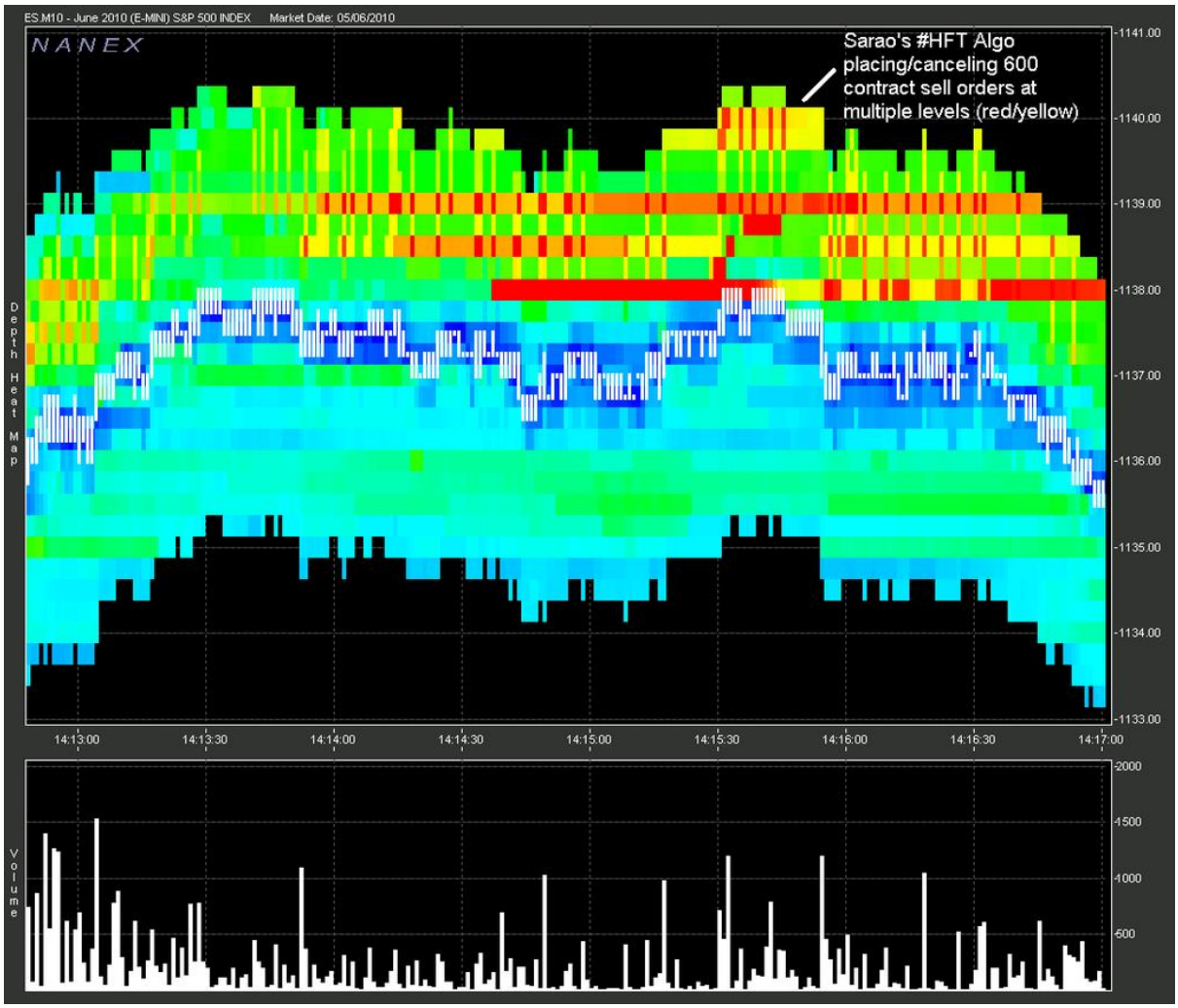
ES.M10 - June 2010 (E-MINI) S&P 500 INDEX Market Date: 05/06/2010
NANEX
Sarao's #HFT Algo placing/canceling 600 contract sell orders at multiple levels (red/yellow)
Depth Heat Map
Volume

Fig.1

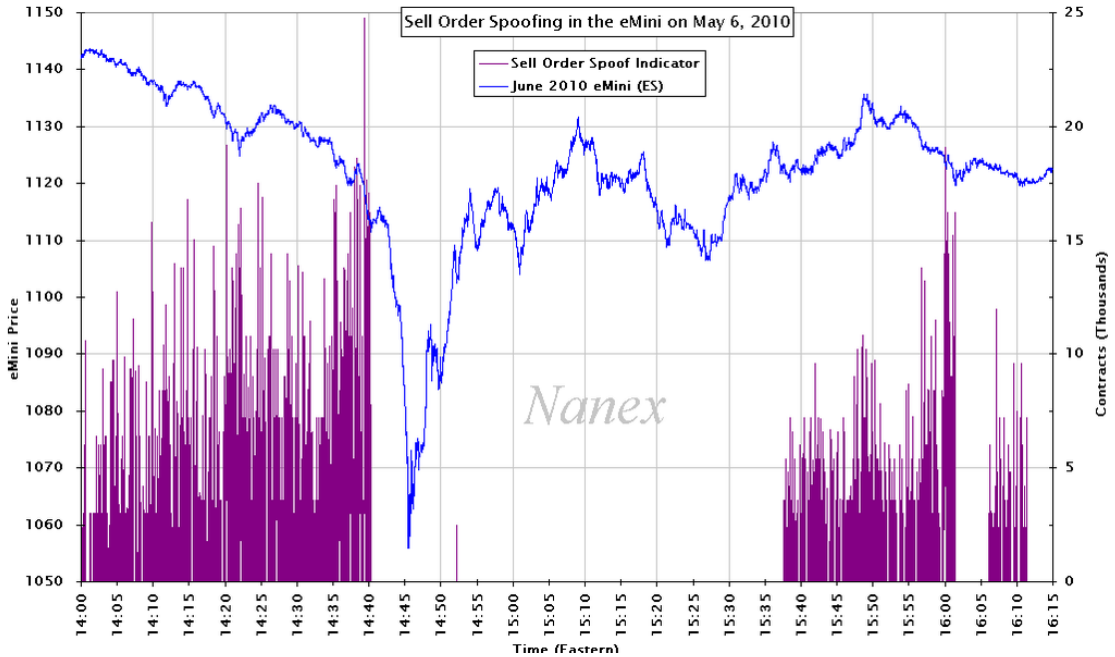
Sell Order Spoofing in the eMini on May 6, 2010
Sell Order Spoof Indicator
June 2010 eMini (ES)
eMini Price
Contracts (Thousands)
Nanex
Time (Eastern)

Fig.2

The Knight Capital Fiasco and its Implications for Non-Live Testing

For an account of the multiple disasters that happened at Knight Capital on 1st August 2012 we suggest this article: http://www.futuresmag.com/2014/10/31/how-power-peg-brought-down-knight.

The Knight saga is a catalogue of failures, including:

- Operations Failure: new code was released to only 7 of 8 servers
- QA/Development Failure: a flag in the code was repurposed
- Audit/Operations Failure: code which had been dormant for 9 years was left on live servers
- Risk Management Failure: an unrecognised error message was ignored
- Audit/Risk Management Failure: there were no limits on the error account
- Emergency Procedures/Operations Failure: improper reversion of new code - repurposed flag left on

If any one of these failures had been averted the disaster would not have happened (or at least not have been so severe). But note that all these failures have one thing in common: **they are failures of the participant**. Non-live testing at the exchange (as proposed in the current draft MiFID II RTS see appendix below) offers an **independent**, orthogonal firewall to failures in participant procedures. This protects investment firms from their own shortcomings while, more importantly, protecting the venue itself from the sort of market disorder which can disrupt other trading unfairly and undermine investor confidence.