Oct 15, 2014 Flash Crash in the US Treasury Markets & 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.

Summary

The July 30th, 2015 Joint Staff Report into the Oct 15, 2014 Flash Crash in the US Treasury Markets[1] merits close attention not least for the light it throws on the importance of non-live testing of algorithms to prevent disorderly trading.

In this briefing note we review relevant parts of the report and extend its conclusions. From this perspective the key observations are:

- Excessive liquidity consumption during the flash crash by High Frequency Traders initiating new positions (Momentum Ignition)
- Large single trades massively disproportionate to the available liquidity in the minutes preceding the flash crash
- Cancellation by one High Frequency Trading firm of 6,000 orders outside the visible order book in one second just after the start of the price surge– sufficient to introduce latency arbitrage opportunities and add to the general disorder

Had the algorithms identified as involved in causing and exacerbating the Treasury Flash Crash been subject to suitable non-live testing, such as that provided by the TraderServe AlgoGuard platform, we strongly suggest that they would have failed and not have been deployed to live trading. We make recommendations to extend globally the MiFID II requirement for non-live testing for propensity to cause or contribute to market disorder and to use efficacious algorithm stability stress testing rather than market replay back-testing which as we explain is not suited to this purpose.

Background

On the day of the crash there was a rapid surge in bond prices across cash and futures markets followed by a similarly rapid retracement in a twelve minute window lasting from 9:33 to 9:45 ET. Although the size of the move was not unprecedented, it was highly disproportionate to changes in exogenous information – in this case, principally, a slightly weaker than expected retail sales figure announced an hour earlier.

The Joint Staff Report examines the trading behaviour of groups of market participants – e.g. bank-dealers, hedge funds – and concludes that, while there was no single cause of the crash, the chief responsibility for the excessive volatility in the event window itself lay with the aggressive trading of what they refer to as Principal Trading Firms or PTFs (predominantly High Frequency Traders as the very high order rates and low fill rates indicate). The report contains detailed analysis of the cash and futures markets. Here, reproducing charts and data appearing in the report, we draw out a number of considerations of relevance to the MiFID II regulations on the non-live testing of algorithms to avoid disorderly trading, with particular reference to the TraderServe AlgoGuard platform which is designed to offer investment firms compliance with these regulations.

---

[1] Joint Staff Report: The U.S. Treasury Market on October 2014 https://www.treasury.gov/press-center/press-releases/Documents/Joint_Staff_Report_Treasury_10-15-2015.pdf

Aggressive Trading by PTFs in the event window

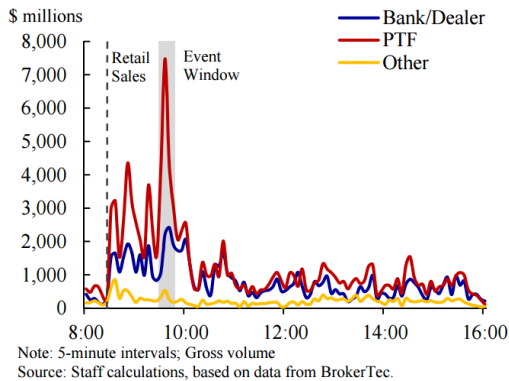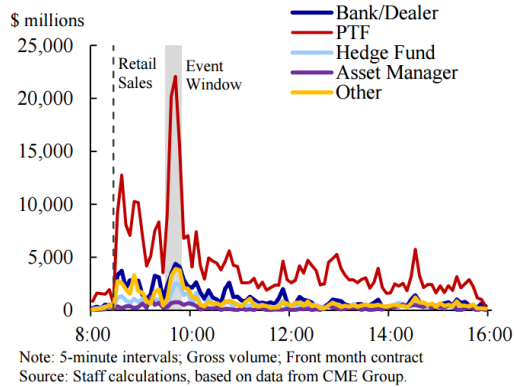**Figure 3.1: 10-Year Intraday Volume by Firm Type (Cash)**

Figure 3.2: 10-Year Intraday Volume by Firm Type (Futures)

Note: 5-minute intervals; Gross volume
Source: Staff calculations, based on data from BrokerTec.

Note: 5-minute intervals; Gross volume; Front month contract
Source: Staff calculations, based on data from CME Group.

The charts above on 10-year cash and bond futures demonstrate clearly that the key spike in traded volume in the event window came from PTFs. In just 5 minutes these participants traded more than $7.5b in the cash markets and $20b in the futures. It is also reported that 74% of all cash trades and 68% of all futures trades in the event window were made by PFTs.

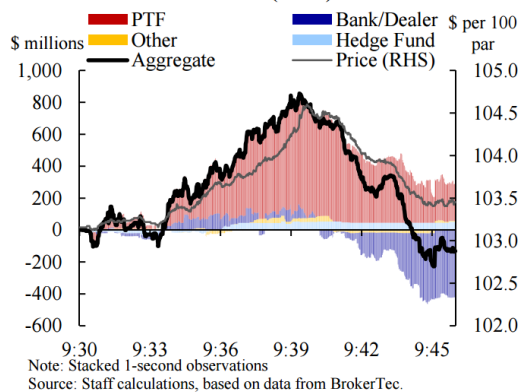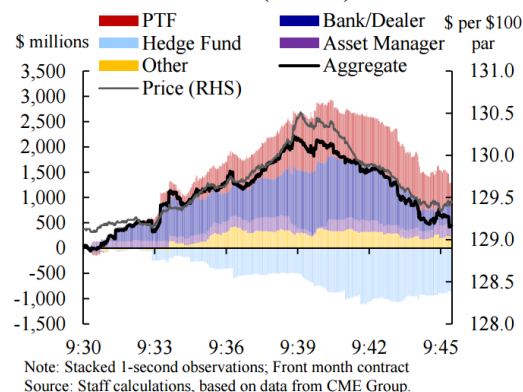**Figure 3.5: 10-Year Cumulative Net Aggressive Volume (Cash)**

Figure 3.6: 10-Year Cumulative Net Aggressive Volume (Futures)

Note: Stacked 1-second observations
Source: Staff calculations, based on data from BrokerTec.

Note: Stacked 1-second observations; Front month contract
Source: Staff calculations, based on data from CME Group.

Figs 3.5 and 3.6 show how the price movement in the event window correlates with net aggressive cash flows of the PTFs. This is particularly striking in the case of the Cash market.

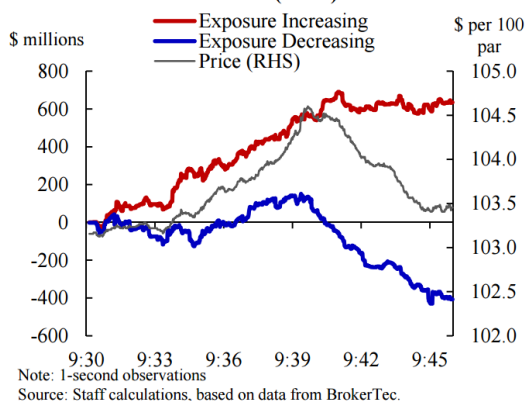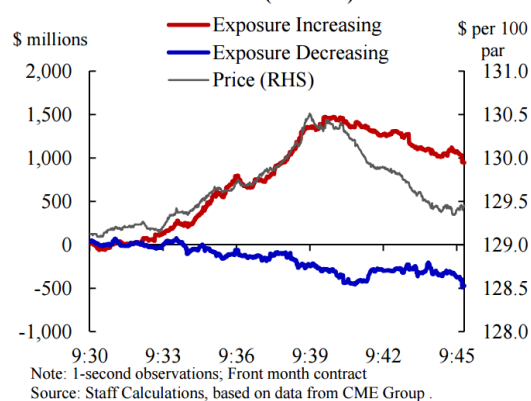**Figure 3.9: 10-Year PTF Cumulative Net Aggressive Volume (Cash)**

Figure 3.10: 10-Year PTF Cumulative Net Aggressive Volume (Futures)

Note: 1-second observations
Source: Staff calculations, based on data from BrokerTec.

Note: 1-second observations; Front month contract
Source: Staff Calculations, based on data from CME Group .

Figs 3.9 and 3.10 show furthermore that the move up coincided with *exposure increasing* aggressive trading by PTFs. In other words, they were not crossing the spread to the offer side of the market to neutralise their positions but to initiate new ones. Algorithms that behave in this way have been called Momentum Ignition and are of concern to regulators in consequence of their disposition to damage fair and orderly market conditions. They have been used abusively by some market participants, particularly in thin markets where it is cheap to penetrate several levels of the order book, to hit stops and to inject momentum into the market in the hope that other participants will extend the move.

The report's authors are silent as to how many algorithms were behaving in this way but there is enough information in the report to allow inferences about individual PTF behaviours.

Attending for now just to the cash market and assuming that the percentage of aggressive PTF trading in the event window was 42% (the median figure in the table on page 52 of the report), we estimate that $3.15b of the $7.5b traded in the 5 minute spike in Fig 3.1 was traded aggressively. Fig 3.5 shows net aggressive cumulative flows from PTFs growing by about 800m long over short in that period. It follows that about $1.975b of the $3.15b was bought aggressively by PTFs. Now Table 3.7 shows 94% of PTF volume is by the top 10 names so we estimate that $1.86b was bought aggressively by just 10 PTFs. Even if this were evenly distributed between the 10 firms it would mean that they were *EACH consuming about 93%[2] of the liquidity in the top 10 levels of the cash market during that period*. This assumes an average 10 level depth in the 10 year bond during the event of $200m in the cash, which seems reasonable given Chart 3.13 below. Similar excessive liquidity consumption can be shown on the Futures market.

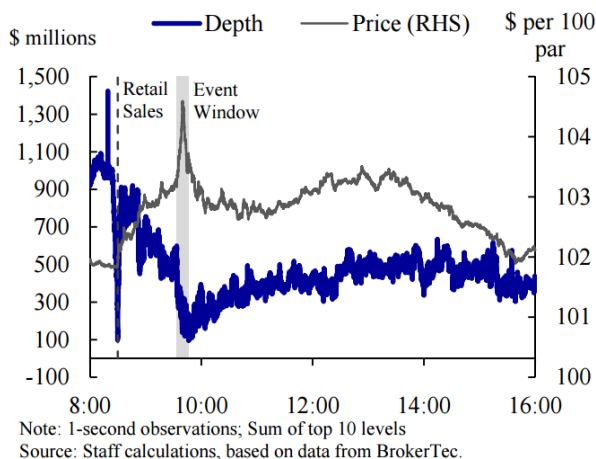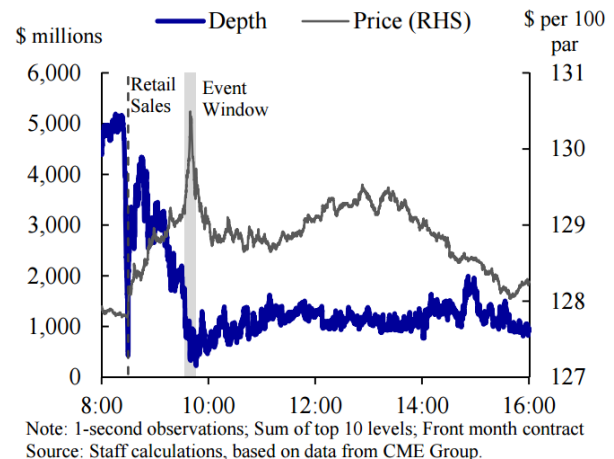
Figure 3.13: 10-Level Market Depth (Cash)

Note: 1-second observations; Sum of top 10 levels
Source: Staff calculations, based on data from BrokerTec.

Figure 3.14: 10-Level Market Depth (Futures)

Note: 1-second observations; Sum of top 10 levels; Front month contract
Source: Staff calculations, based on data from CME Group.

TraderServe AlgoGuard has pass/fail measures of the disposition of aggressive trading to provoke market disorder. These would fail any single algorithm that engaged in liquidity consumption on a scale of that estimated. MiFID II requires each investment firm to set suitable tests and criteria to avoid creating or contributing to market disorder, 93% is far beyond any reasonable threshold as any sensible non-live testing market calibration will show.

---

[2] The amount of liquidity consumed can exceed 100% for two reasons: first, the percentage is of liquidity in the top 10 levels and as this is consumed and price moves new levels come into play; secondly, liquidity in those levels is replaced over time by new orders

Large Trades in the Hour between Retail Sales Figure and the Flash Crash Event

Figure 3.25: 10-Year Timeline of Large Aggressive Trades (Futures)

Note: Trades > 2000 contracts; $200 million notional; Top 10 levels; Front month contract
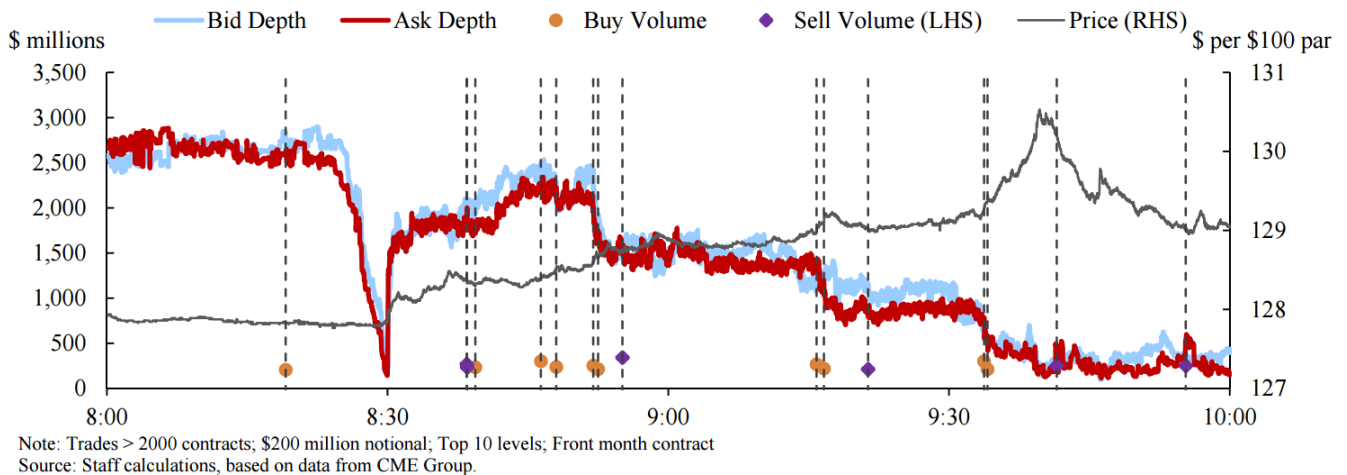Source: Staff calculations, based on data from CME Group.

Fig 3.25 above shows the price in grey, total volume on the top 10 levels of the order book (blue on the bid and red on the offer) and large aggressive trades (gold for buys, purple for sells) on the T-note futures between 8 am and 10 am.

There is, as usual, a very steep decline in liquidity prior to the 8:30 sales figure. But, unusually, after a partial recovery, from 8:45 until the event window there is a steady loss of liquidity in the top 10 levels as big orders consume liquidity without its being replaced. The Joint Staff Report points out that aggressive orders of this size are not uncommon at the same time of day and that on this occasion they seem not to have had much impact on price even if they did do on the erosion of liquidity. To say the least, this erosion of liquidity was remarkable in the case of three pairs of large long trades: the first at about 8.45, the second just after 9.15 and, especially, the third at what appears to be the beginning of the event window. *In this last case the volume on the top 10 levels falls by about 50%* from $1b to $0.5b. Even on the assumption that each trade in the pair is at the minimum 2,000 contracts required to be represented in the chart, this gives a notional value of $200 million on a single trade. Now aggressive trading on this scale in one hit might be perfectly safe in normal conditions but while the top 10 levels in the order book contains less than $1b it counts as reckless. If this sort of behaviour were manifested by an algorithm in AlgoGuard's non-live testing, it would fail on the grounds of excessive liquidity consumption. Were these large trades algorithmic? The report does not say but the very high volumes shown in Figs 3.1 and 3.2 for PTF trading in the period between the retail sales figure and the event window is suggestive.

On this occasion, though the Joint Staff Report does not so conclude, it appears likely to us *that the very much reduced liquidity produced by large single trades in an already thin market triggered the momentum ignition episode which followed and contributed to the causes of the flash crash[3].*

---

[3] Momentum Ignition is particularly potent in very low liquidity conditions where penetrating multiple levels is relatively cheap because low volumes need to be traded to do so. Stops are then easily run creating momentum which is then often added to by the reaction of short-term momentum traders

A Massive Series of Cancellations Outside the Visible Order Book

One other point of interest raised in the report is the very high number of cancellations outside the visible futures order book at 9:34:03 shown in Fig 3.27 below

Figure 3.27: Messages Processed by Matching Engine Outside Visible 10 Levels (Futures)
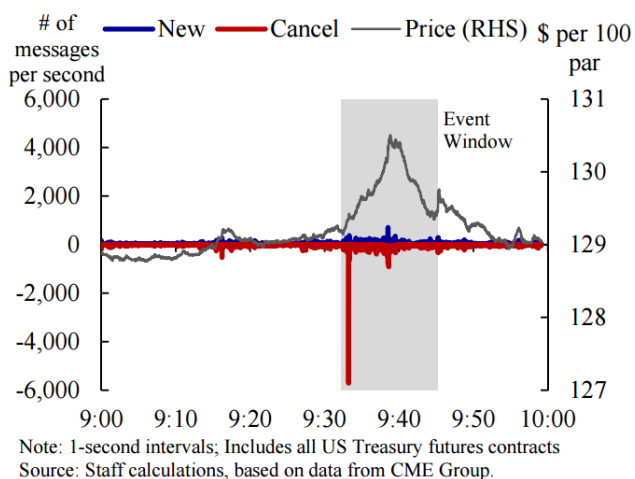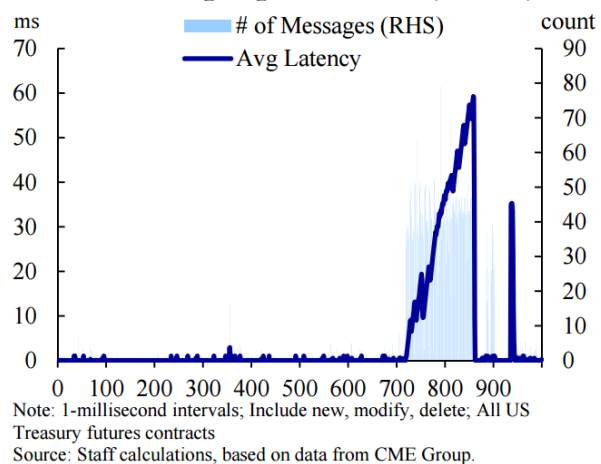
Note: 1-second intervals; Includes all US Treasury futures contracts
Source: Staff calculations, based on data from CME Group.

Figure 3.29: Latency in US Treasury Futures Matching Engine at 9:34:03 (Futures)

Note: 1-millisecond intervals; Include new, modify, delete; All US Treasury futures contracts
Source: Staff calculations, based on data from CME Group.

In one second 6,000 orders were removed from outside the top 10 levels of the order book. Fig 3.29 above shows that this induces a latency of up to 60 ms which is enough to produce substantial latency arbitrage opportunities between cash and futures markets, especially when prices are changing rapidly, as they were at the time (one minute into the event window) so there is every reason to believe that this activity also contributed to the market disorder.

The report states that the cancellations were all made by one PTF and that the firm in question only traded lightly on the day. If so, its fill rates must have been infinitesimal.

AlgoGuard has pass/fail measures of disorder provocation which apply to passive orders as well as to aggressive ones. The former are designed amongst other things to detect algorithms such as quote stuffers that cancel excessively in such a way as to threaten latency. As amply demonstrated in the 2010 Equities flash crash and again here in the 2014 Treasuries flash crash algorithms of this kind are damaging to the efficient functioning of the markets. Fortunately, non-live testing offers a way of keeping them out of the live markets and this algorithm would have been failed by AlgoGuard's passive disorder provocation measures.

Conclusions

We have analysed here a number of issues arising from the Joint Staff Report into the Oct 15, 2014 Flash Crash in the US Treasury Markets that impact on the issue of non-live testing. Other findings such as the extraordinary amount of self-trading, especially by PTFs must clearly be of regulatory concern but their remedy lies elsewhere (i.e. not with non-live testing). As with the S&P minis flash crash[4] no single cause to the Treasury Flash Crash can be identified but reckless and possibly abusive algorithmic behaviour from more than one source is evident.

Of particular note are the following:

- Excessive liquidity consumption during the event window by PTFs initiating new positions (Momentum Ignition) contributing to market disorder via quantifiable and extreme aggressive disorder provocation. *The top 10 PTFs consumed about 930% of the liquidity in the top 10 levels of the cash market during that period* (see Note 2 on Page 3).

- Large single trades massively disproportionate to the available liquidity in the minutes preceding the event window. This was reckless trading insensitive to the increasingly liquidity starved market contributing to market disorder via quantifiable aggressive disorder provocation and a likely trigger for the following Momentum Ignition. *In the case of one pair of trades the volume on the top 10 levels of the futures market falls by about 50%*

- Cancellation by one PTF of 6,000 orders *outside the visible order book* in one second just after the start of the event window– sufficient to introduce latency arbitrage opportunities of 60ms and add to the general disorder. This constituted quantifiable passive disorder provocation.

This episode illustrates the value of TraderServe's notion of Emergent Market Disorder[5] arising from interplay of algorithms[6] and strengthens the already solid case for requiring that algorithms be subject to stability testing in realistic and stressed conditions before being allowed into live trading. Had the algos identified as involved in causing and exacerbating the Treasury Flash Crash been subject to suitable testing, such as that provided by the TraderServe AlgoGuard platform, based upon our analysis we are confident that they would have failed and not have been deployed to live trading. This is a similar conclusion to our analysis[4] of the 2010 S&P minis flash crash even though most of the algorithms identified as majorly responsible for that event were different.

---

[4] "Briefing notes principally on the 6th May 2010 Flash Crash and its Implications for Non-Live Testing 18 may 2015" http://www.traderserve.com/pdf/TS-briefingnote0n6thMay2010FlashCrashandNon-LiveTesting.pdf

[5] Emergent Market Disorder: the joint staff report states that no single cause for the flash crash was detectable. A number of different algorithms appear to have been involved. This is similar to our analysis of the S&P minis flash crash linked in the previous footnote. We coined the phrase Emergent Market Disorder to cover market disorder caused by the interplay between algorithms in our 2014 research report linked in the next footnote. We have argued that its existence shows the need to stability test algos in the company of other "antagonist" algos to uncover such propensity to contribute to market disorder.

[6] "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/Non-Live%20Testing-whytradingvenuesneedtotestforEmergentMarketDisorderandstressedconditions20150126-1.pdf

Recommendations

In view of the substantial problems raised by the Joint Staff Report it is obvious why all regulators are looking to strengthen protections against misbehaving algorithms. MiFID II provides an excellent approach with its requirement for non-live testing for propensity to cause or contribute to market disorder, and over time we anticipate that similar requirements will be extended to cover most or all asset markets where there is a significant algo presence. We believe that this will be of great benefit to global market stability.

We recommend algorithm stability testing such as that offered by AlgoGuard, with pass/fail stress testing of the key attributes of disorder provocation, as a very effective way to prevent risky algorithms from reaching production. Unlike back-testing, algorithm stability testing includes the interaction of the algorithm under test with a realistic market microstructure as well as antagonist algorithms designed to cause stress. This type of testing can detect propensity to cause or contribute to market disorder even when that arises from interplay of algorithms (Emergent Market Disorder) and, by preventing vulnerable and abusive algorithms from reaching live markets in the first place, it is an especially effective means of avoiding disorderly live trading such as that seen in the 2010 S&P Minis flash crash and the 2014 Treasury flash crash.