Modelling and mitigation of Flash Crashes

Fry, John and Serbera, Jean-Philippe

School of Computing Mathematics and Digital Technology, Manchester Metropolitan University, Sheffield Business School, Sheffield Hallam University

12 September 2017

Modelling and mitigation of Flash Crashes

September 2017

Abstract

The algorithmic trading revolution has had a dramatic effect upon markets. Trading has become faster, and in some ways more efficient, though potentially at the cost higher volatility and increased uncertainty. Stories of predatory trading and flash crashes constitute a new financial reality. Worryingly, highly capitalised stocks may be particularly vulnerable to flash crashes. Amid fears of high-risk technology failures in the global financial system we develop a model for flash crashes. Though associated with extreme forms of illiquidity and market concentration flash crashes appear to be unpredictable in advance. Several measures may mitigate flash crash risk such as reducing the market impact of individual trades and limiting the profitability of high-frequency and predatory trading strategies.

JEL Classification: C1 F3 G1 K2
Keywords: Flash Crashes; Flash Rallies; Econophysics; Regulation

1 Introduction

Algorithmic Trading (AT) incorporates a range of high speed and predatory strategies termed Flash Trading Strategies by Lewis (2014) in recognition of the sheer speed involved – orders of magnitude faster than the blink of a human eye. Searching for a competitive edge the need for speed even led to the construction of new optic fibre tunnels to shorten the network distance between New York and Chicago to save crucial milliseconds. This was later rendered obsolete by the construction of special towers to enable the transmission of orders between New York and Chicago via microwaves. Algorithmic Traders (ATs) now dominate financial markets and are currently thought to constitute between 40-55% of the trading volume on European and US equity markets respectively. For futures markets these figures could be as high as 80% (Miller and Shorter, 2016). As speed and latency continue to develop AT has profound implications for global financial markets.

Flash crashes first came to prominence during the Flash Crash of May 6th 2010. Around 2.30-3.00pm EST saw dramatic upheaval on US future and equity markets with the Dow Jones Industrial Average (DJIA) losing around 10% of its value before recovering. Official reports cite over-activity by ATs as being responsible for a liquidity crisis that caused the crash (CFTC and the SEC, 2010). Potentially only the timing of the event, away from the market close, averted Armageddon (Cliff and Northrop, 2012). Despite recent legislation fears over flash crashes persist. For instance, SEC Rule 201 (known as the Alternative Uptick Rule) of Regulation SHO (for short-selling) was introduced to limit short-selling of National Markets System securities that suffer a 10% intraday decline. However, academic research suggests that this has not reduced short-selling activity (Jain et al., 2012) amid fears that the rule would not prevent future flash crashes.

Though significant intraday volatility has long been a feature of global financial markets (Aldridge and Manciw, 2017) flash crashes are a modern phenomenon. However, flash crashes are reminiscent of a more extreme version of the intra-day price falls during the 1987 stock market crash that were exacerbated by automated sales orders triggered by portfolio insurance strategies. Perhaps finance has always needed time to adjust to new technology. The first major flash crash identified in the literature occurred on the USD/JPY currency pair on August 16th 2007 (Chaboud et al., 2014). On that day the U.S. Dollar dropped sharply (approximately 5 percent) against the Japanese yen between 6 a.m. and 12 p.m. EST.

Selected major flash crash events are shown below in Table 1. The incidence of flash crashes is alarming and has accelerated in recent years. As Table 2 shows flash crashes have affected many of the major stocks listed on the NYSE and the Nasdaq. A logistic regression model (Bingham and Fry, 2010; Chapter 7) gives formal evidence of a relationship between market capitalisation and flash crash events (p=0.000). Worryingly, the implication is that AT makes highly capitalised stocks particularly vulnerable to flash crashes (see Figure 1). This tallies with the observation in Brogaard et al. (2013) that the profitability of AT increases in line with the capitalisation of the stocks in question. With a diverse range of asset classes affected (including stocks, bonds, currencies, futures, commodities, derivatives and bonds) recent warnings from the Bank of England to "brace for future crashes" appears increasingly timely.

The importance of our contribution is threefold. Firstly, we contribute to the literature modelling flash crashes (Filimonov and Sornette, 2012; Hardiman et al., 2013). Inter alia our model builds on previous works that have used methods originating from the physical and engineering sciences to understand high-frequency financial systems (Cliff and Northrup, 2012; Shaw and Schofield, 2015). Secondly, our model leads to a new way of characterising flash crashes and flash rallies tied to the incentive structures facing ATs. However, though associated with extreme forms of illiquidity and market concentration flash crashes appear to be unpredictable in advance. Thirdly, we contribute to on-going debates regarding regulation of high-frequency financial systems. Though the merits of regulation are often hotly debated (Aït-Sahalia and Sağlam, 2017; Aldridge and Krawciw, 2017) there is some suggestion that flash crash risk may be mitigated by limiting the market impact of individual trades and the profitability of algorithmic and predatory trading strategies (see Section 4).

The layout of this paper is as follows. An overview of flash crashes and algorithmic trading is given in Section 2. Section 3 introduces the model used. Section 4 highlights policy im-

Date	Security	Description
15/10/14	US Treasury futures	Fall of 33 basis points to 1.86% before rising to 2.13%
15/01/15	EUR/CHF spot	Frankenshock following the decision of the Swiss National Bank to abandon its floor value for the Euro. The Euro lost more than 20% of its value between 4.30 and 4.33am EST before stabilizing at 5.10am EST.
18/03/15	EUR/USD spot and futures	The Euro lost more than 3% of its value in less than 3 minutes before recovering.
29/04/15	S&P 500 index (spot)	The S&P 500 fell more than 5% in early trading.
07/10/16	GBP/USD	The GBP fell 6.1% in two minutes during the Asian trading session before recovering.
27/06/17	Gold futures	Gold fell in value by 1.6% before recovering in the wake of a massive surge in trading volume.

Table 1: Selected recent high-profile recent flash crashes.

Market Capitalisation	Average capitalisation	No. of stocks affected	No. of stocks in category	Percentage of stocks affected
Mega cap. (>$200 billion)	$321.84 billion	8	19	42.11%
Large cap. ($10-200 billion)	$36.44 billion	60	589	10.19%
Medium cap. ($2-10 billion)	$4.54 billion	51	1057	4.82%
Small cap. ($0.3-2 billion)	$0.86 billion	55	1774	3.10%
Micro cap. ($0.05-0.3 billion)	$0.15 billion	16	1360	1.18%
Nano cap. (<$0.05 billion)	$0.01 billion	2	994	0.20%

Table 2: Flash crash events for stocks listed on the NYSE and Nasdaq (March 2011-June 2014) by market capitalisation. (Market capitalisation correct as of 29/7/2016).

plications. An empirical application to the flash crash of May 6th 2010 is outlined in Section 5. Section 6 concludes and discusses future work. Widely documented predatory algorithmic trading strategies are outlined in an Appendix at the end of the paper.

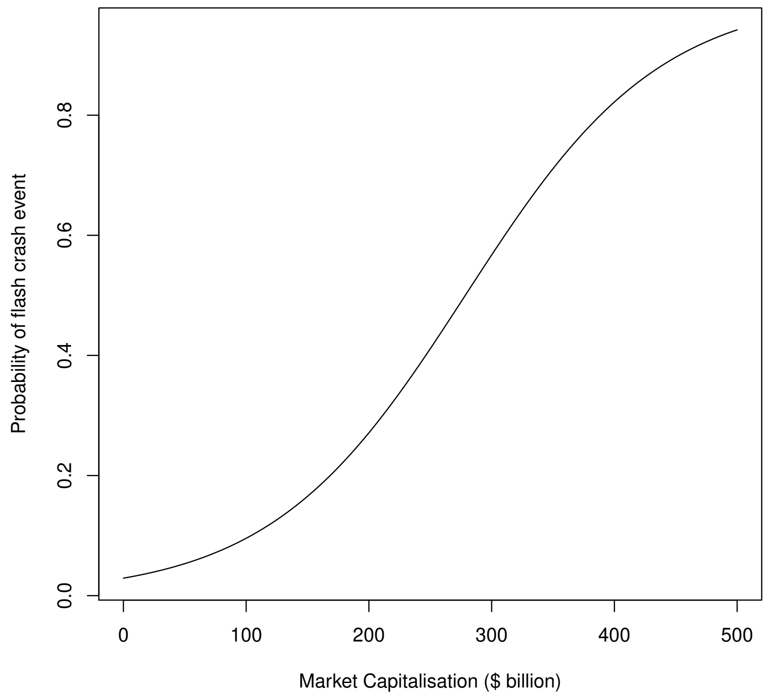

Figure 1: Probability of a flash crash event by market capitalisation for NYSE and Nasdaq listed stocks.

2 Background

2.1 Flash Crashes and algorithmic trading

ATs play a sometimes controversial role in markets. ATs are clearly part of the modern financial reality (see above) and can play a valid financial function via such things as execution of trades, market-making, statistical arbitrage and liquidity rebates. The SECs own regulations allow for a sense of financial fair play so that ATs can reasonably profit from their technological edge so long as this is restricted to publicly available information. As an illustration the profits from latency arbitrage, for example, can be substantial. Such a highly competitive environment may place intense evolutionary pressures upon human traders (Serbera and Paumard, 2016) but is not in itself unfair. Controversy arises when ATs use their super-human capacities (latency, high speed, data processing power etc.) to distort markets. Such predatory strategies even have their own popularly used names e.g. algorithm sniffing, spoofing, quote stuffing, liquidity anticipation, latency arbitrage, liquidity re-direction, marking the close etc. A description of widely documented predatory algorithmic trading strategies is given in the Appendix at the end of this paper.

The recent growth in algorithmic trading has dramatically affected market microstructure (Menkveld, 2014). Much of AT performs valid economic functions e.g. market-making and arbitrage (Serbera and Paumard, 2016). In Hagstromer and Norden (2013) market-making is attributed to 65-71% of all high-frequency trading activity. However, there is increasing concern that market-making ATs may withdraw in times of market stress when liquidity is sparse – thus amplifying shocks throughout the system (Schlepper, 2016). Other aspects of AT (including predatory strategies) are more controversial. In particular, predatory trading has intensified competition between algorithmic traders, reducing overall levels of profitability (Serbera and Paumard, 2016).

The relative merits of HFT are open to debate. Perceived benefits of HFT include reduced spreads (Bershova and Rakhlin, 2013) and increased efficiency (Nagata and Inui, 2014). Many view HFT as a trade off between enhanced price discovery and excess volatility (Benos and Sagade, 2012). This trade off can be particularly severe in times of market stress (Schlepper, 2016). Others doubt that HFT reduces transaction costs (Malinova et al., 2013; Brogaard et al., 2012). Tokic (2015) even finds a link between HFT and increased transaction costs.

Flash Crashes emerge as part of wider debates around HFT (Boultana et al., 2014; Lewis, 2014). Hudson et al. (2015) find evidence of an association between increased use of algorithmic trading and Flash Crash risk. The market regime is also hugely significant. Ordinarily HFT facilitates market discovery but in times of market stress can be a catalyst for Flash Crashes (Tokic, 2015). A common criticism is that HFTs supply liquidity when volatility is low but often withdraw from markets when volatility rises. Precursory factors associated with Flash Crashes include unusual options activity (Boultana et al., 2014) abnormally-high levels of inter-market sweep orders (McInish et al., 2014), extreme illiquidity (Easley et al., 2011) and market concentration (Bethel et al., 2011). Competition between algorithmic traders may also lead to reduced liquidity under adverse market conditions (Madhavan, 2012). Interactions between low-frequency and high-frequency traders can also generate volatility that can lead to Flash Crashes (Leal et al., 2016).

2.2 Flash Crashes and regulation

The regulation of flash crashes poses major technical and administrative challenges. During the flash crash of May 6th 2010 circuit breakers were not triggered in the DJIA because a 10% price drop specified by Rule 80b was narrowly avoided. However, during the flash crash the price of securities fell by more than 60%. In response, the SEC proposed new rules allowing stock and options exchanges and Algorithmic Trading Systems to implement decentralised circuit breakers for individual securities. However, since the initial rules were proposed market disruptions have continued to occur. In response a new set of formal rules has been promulgated by the SEC to try to improve the resilience of trading systems.

At the level of individual market participants automated trading is more regulated. In the US, the CFTC Regulation Automated Trading (Regulation AT) Rule proposes pre-trade risk controls. It requires an identification of ATs on the basis of a daily volume threshold (20,000 traded contracts), a registration as a floor trader and checks using anti-evasion measures. In the UK under the Market in Financial Instruments Directive (MiFID) all high-frequency firms need to be authorised by the financial conduct authority. Secondly, there will also be a limit placed on the number of order messages that a market participant will be able to send relative to the number of transactions they undertake (www.fca.org.uk). In Japan on the Tokyo Stock Exchange, the Financial Services Agency recently added at-trade and post-trade checks to existing pre-trade risk-control requirements. The at-trade checks occur at the moment an order is submitted and include profit and loss limits, speed and order rejection limits, open orders and positions. Post-trade checks are part of the back-office risk control and verify the overall position limits including the entire portfolio at the firm level.

To complement the efforts of regulatory agencies private initiatives from the stock exchanges themselves can increase the level of risk control. For example the Investor Exchange (IEX) advertises on the implementation of a"speed bump" of 350-microsecond delay on orders (Lewis, 2014). This delay is sufficient to counter the speed advantage of HFT over retail investors. In May 2017, the NYSE received a regulatory approval to implement an identical delay strategy in its small and mid-cap trading venue (NYSE American). According to Manahov (2016) the introduction of batch auctions once every 30 milliseconds might also be helpful in reducing the speed advantage of ATs. However, there is some debate as to whether imposing time delays, as above, may have unintended consequences (Aldridge and Krawkiw, 2017). Measures proposed by Aït-Sahalia and Sağlam (2017), thought to have only a limited chance of success, include a transaction tax, setting minimum time-limits before quotes can be cancelled, taxing the cancellations of limit orders and replacing time priority with a pro rata or random allocation.

3 The model

Let P_t denote the price of an asset at time t and let $X_t = \log\ P_t$. Based on theoretical models in Fry (2012) and Aldridge (2014) the set up of the model is as follows:

Assumption 1 (Intrinsic Rate of Return) *The intrinsic rate of return is assumed constant and equal to μ:*

$$E[X_{t+\Delta} - X_t|X_t] = \mu\Delta + o(\Delta). \tag{1}$$

Assumption 2 (Intrinsic Level of Risk) *The intrinsic level of risk is assumed constant and equal to σ^2:*

$$Var[X_{t+\Delta} - X_t|X_t] = \sigma^2\Delta + o(\Delta). \tag{2}$$

Adapting the original approach in Johansen et al. (2000) we postulate the following model

$$P(t) = e^{X(t)}(1+\kappa)^{j(t)}, \tag{3}$$

$j(t)$ is a jump process describing the post Flash Crash market recovery satisfying

$$j(t) = \begin{cases} 0 & \text{before the recovery} \\ 1 & \text{after the recovery.} \end{cases} \tag{4}$$

During the Flash Crash X_t satisfies the Stochastic Differential Equation

$$dX_t = b[dB_t - dS_t] + vdj(t), \tag{5}$$

where $v = \ln[1+\kappa] \geq 0$, b describes the market impact of individual trades, B_t and S_t denote the number of buyers and numbers of sellers respectively and are independent inhomogeneous Poisson processes with rates $\lambda_B(t)$ and $\lambda_S(t)$ respectively. Inter alia this formulation reflects the conjecture in Luckock (2003) that in a market of rational and well-informed traders the two sides of the order book should be independent. Equations (3-5) thus describe a modified version of the negative bubble model described in Fry and Cheah (2016) with the Gaussian noise replaced by a Poisson difference or Skellam Distribution (Karlis and Ntzoufras, 2003). Replacing v by $-v$ in the above gives a modified version of the speculative bubble model in Cheah and Fry (2015) leading to a model for Flash Rallies (Cui and Gozluzku, 2016) where, rather than crashing, prices rise dramatically in the short-term before quickly returning to normal levels. According to this model Flash Crashes occur as high-frequency traders drive down the price before profiting as the price instantaneously recovers. Hence this model reconstructs qualitative aspects of predatory algorithmic trading practices (see e.g. Lewis, 2014; Serbera and Paumard, 2016; Aldridge and Krawciw, 2017) and mirrors the way that Flash Crashes have led to a collapse in the bid prices of certain stocks (see Section 5).

From Assumption 1 it follows that

$$b[\lambda_B(t) - \lambda_S(t)] + vh(t) = \mu;\; b[\lambda_B(t) - \lambda_S(t)] = \mu - vh(t). \tag{6}$$

Thus, equation (6) shows that the Flash Crash depresses the price. Moreover, the implication is that the effects will be sufficiently strong as to be essentially unpredictable in advance. From Assumption 2 it follows that

$$b^2[\lambda_B(t) + \lambda_S(t)] + v^2 h(t) = \sigma^2. \tag{7}$$

The simultaneous equations (6-7) can be solved to give

$$\lambda_B(t) = \frac{\sigma^2 + b\mu}{2b^2} - \frac{(vb + v^2)h(t)}{2b^2};\; \lambda_S(t) = \frac{\sigma^2 - b\mu}{2b^2} + \frac{(vb - v^2)h(t)}{2b^2}. \tag{8}$$

3.1 Market distortion

In the sequel we consider a model without Flash Crash risk ($v = 0$). In this case equation (8) reduces to

$$\lambda_B(t) = \frac{\sigma^2 + b\mu}{2b^2}; \ \lambda_S(t) = \frac{\sigma^2 - b\mu}{2b^2}. \tag{9}$$

The importance of equation (9) is threefold. Firstly, equation (9) emphasises that some level of price risk σ^2 is essential for the proper functioning of the market. Secondly, equation (9) generalises a model for high-frequency returns in Alzaid and Omair (2010). Thirdly, equation (9) is stylistically similar to practitioner models of latency-sensitive trading briefly discussed in Schlepper (2016). Following Fry and Cheah (2016) define the fundamental value to be the expected price when $v = 0$:

$$P_F(t) := E[P_t] = \frac{P(0)E[e^{bB_t}]}{E[e^{bS_t}]} = P(0)e^{(\lambda_B - \lambda_S)t(e^b-1)} = P_0 e^{\frac{\mu}{b}t(e^b-1)}, \tag{10}$$

since if $X{\sim}Po(\lambda)$ then

$$E[e^{tX}] = e^{\lambda(e^t-1)}$$

(Grimmett and Stirzaker 2003).

During the Flash Crash ($v{\neq}0$) we have that

$$\begin{aligned} P_{NB}(t) := E[P_t] &= P(0)\exp\left\{\left(\frac{\mu}{b}t - \frac{v}{b}H(t)\right)(e^b-1)\right\}. \\ &= P_F(t)\exp\left\{-\frac{v}{b}H(t)(e^b-1)\right\}, \end{aligned} \tag{11}$$

where $P_F(t)$ is given by equation (10) and $H(t) = \int_0^t h(u)du$. Following Fry and Cheah (2016) we can estimate the size of the effect in terms of the average distance between observed and fundamental prices:

$$M := \frac{1}{T}\int_0^T \left(1 - \frac{P_F(t)}{P_{NB}(t)}\right) dt = 1 - \frac{1}{T}\int_0^T \exp\left\{\frac{v}{b}H(t)(e^b-1)\right\} dt \tag{12}$$

4 Policy implications

During a Flash Crash the level of market distortion is given by equation (12). Differentiating (12) with respect to b it follows that

$$\frac{\partial M}{\partial b} = -\frac{1}{T}\int_0^T vH(t)\left(\frac{1 - e^b + be^b}{b^2}\right)\exp\left\{\frac{v}{b}H(t)(e^b-1)\right\} dt{\leq}0. \tag{13}$$

Thus, equation (13) suggests reducing Flash-Crash risk by limiting the market impact of individual trades (e.g. by spreading out high-frequency transaction orders over longer periods of time; Lewis, 2014). There is a suggestion that Flash Crashes are associated with extreme forms of market illiquidity (Easley et al., 2011) and market concentration (Bethel et al., 2011). Maintaining liquidity, especially in times of market stress, is hugely significant (Schlepper, 2016).

Similarly, differentiating (12) with respect to v it follows that

$$\frac{\partial M}{\partial v} = -\frac{(e^b - 1)}{bT} \int_0^T H(t) \exp\left\{\frac{v}{b} H(t)(e^b - 1)\right\} dt \leq 0. \tag{14}$$

Thus, equation (14) suggests that as v increases then the effects of the Flash Crash become more extreme. Similarly, M also decreases ("gets worse") as $H(t)$ increases. Both results suggest we may reduce the severity of Flash Crashes by limiting the profitability of predatory trading and AT more generally. From a theoretical perspective various authors have previously made a link between AT and Flash-Crash risk (see e.g. Brandt and Neumann, 2015; Hudson et al., 2015; Leal et al., 2016). Practical steps limiting the effects of flash trading include the development of time delays on the IEX exchange (Lewis, 2014). Schlepper (2016) suggests reducing the attractiveness of predatory trading strategies by incentivising ATs to generate more informational trades (e.g. by imposing Order to Trade Ratio (OTR) limits). Aït-Sahalia and Sağlam (2017) consider the effect of time delays and transaction taxes to limit the profitability of high-frequency trading.

5 Empirical application

Following Shaw and Schofield (2015) we consider an empirical application to the Flash Crash of May 6th 2010. A plot of the Accenture stock price is shown in Figure 2. The speed and scale of the crash is dramatic – the bid price of Accenture almost completely collapses within less than 1 minute. Following Cheah and Fry (2015) we use choose

$$h(t) = \frac{\beta t^{\beta-1}}{\alpha^\beta + t^\beta}. \tag{15}$$

Under this specification the estimated market distortion given by equation (12) reduces to

$$M = 1 - \frac{1}{T} \int_0^T \left(\alpha^\beta + t^\beta\right)^{\frac{v}{b}(e^b - 1)} dt. \tag{16}$$

In the sequel we fit the Flash Crash model shown in equation (8) to this data. A likelihood ratio test of the null hypothesis $v = 0$ shows that the derived model offers a statistically significant description of historical data ($\chi^2 = 296.0588$, $p = 0.000$). Further, numerical evaluation of the integral shown in equation (16) shows that the overall level of market distortion is substantial. The clear implication is that Flash Crashes may present hugely profitable opportunities for rogue traders (Aldridge and Krawciw, 2017).

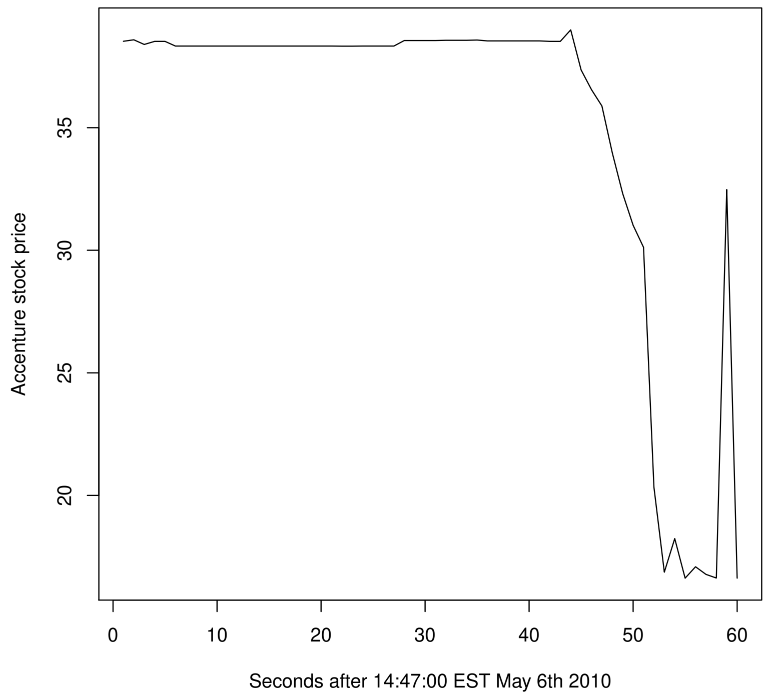

Figure 2: Stock price of Accenture 14:47-14.48 EST May 6th 2010. Source: `www.dxfeed.com`.

6 Conclusions and further work

Algorithmic trading represents a new financial reality and has had a dramatic impact upon global financial markets. Throughout financial history the transition to new technologies has rarely run smoothly (Reinhart and Rogoff, 2009). The flash crash of May 6th 2010 brought into sharp focus the threat that high-frequency traders may pose to the global financial system. Though associated with extreme forms of liquidity and market concentration flash crashes appear to be essentially unpredictable in advance. The sinister threat posed by flash crashes is exacerbated by uncomfortable questions about the sometimes perverse incentives facing Algorithmic Traders. Highly capitalised stocks appear particularly prone to flash crashes. Moreover, as our model shows, the potential clearly exists for Algorithmic Traders to profit handsomely from dramatic price oscillations.

In this paper we provide a new way of characterising both flash crashes and flash rallies. This tractability builds on a burgeoning literature that has sought to model flash crashes (Aït-Sahalia and Sağlam, 2017; Cliff and Northrup, 2012; Hardiman et al,. 2013; Shaw and Schofield, 2015). Further, we contribute to on-going debates regarding the regulation of Algorithmic Trading and flash crashes. Regulation is a vexed issue and the efficiency of various proposals is often hotly debated (Aït-Sahalia and Sağlam, 2017; Aldridge and Krawciw, 2017). However, the implication of our model is that some degree of flash crash mitigation is possible by placing limits on the profitability of predatory trading, and of Algorithmic Trading more generally, and by reducing the market impact of individual trades.

Algorithmic Trading, and its attendant risks, are undoubtedly here to stay. Future flash crashes are inevitable. Future work will examine additional applications to algorithmic trading and high-frequency finance (see e.g. Barndorff-Nielsen et al., 2012). Future work will also examine the application of related methods to risk and failure in other complex social systems (Chernov and Sornette, 2016).

Appendix: Widely documented predatory algorithmic trading strategies

Algorithm sniffers are designed to detect the Volume-Weighted Average Price (VWAP) execution algorithms typically used by large institutional investors to ensure that trades are conducted in line with volumes actively being traded on the market. So-called "algo sniffers" typically "ping" very small market orders to detect liquidity and hidden orders. Once the sniffer has detected, say, a large buying order they can profit by front running. Simply buy the shares faster than the VWAP and then sell the shares on to the VWAP at a profit.

Spoofing consists of a set of strategies specifically designed to fool other traders – especially ATs. For example, bogus limit buy orders may be placed at fractions below the current market price with the purpose of luring other participants, especially ATs specialised in tape reading, into anticipating an upward price due to the presence of a large number of bids. The aim is to incentivise other participants to post buying quotes above the large (spoof) buying order. In the meantime the spoofers can cancel their buy order and sell their existing share holdings at a higher price. However, as spoofing has become more high profile it has come under increased scrutiny from regulatory agencies with the Commodity Futures Trading Commission (CFTC) and the London Stock Exchange both issuing penalties for spoofing in recent years.

Quote stuffing consists of predatory arbitrage. It entails sending unusually and unexpected high volumes of traffic orders. For example, 5000 quotes were sent in one second on May 6th 2010 during the Flash Crash and were directed to public and private exchanges, with the express purpose of slowing down their data systems. Once systems are slowed by this manipulation ATs can realize a profit by arbitraging artificially increased spreads. The generation of a sizeable number of quotes, that rival ATs have to analyse and process, allows for substantial time gains - especially given the super-human speed of trading involved. In addition, this strategy also affects the National Best Bid and Offer (NBBO) prices, without any trades occurring, leading to potentially highly profitable arbitrage opportunities.

Liquidity anticipation in flash trading is used to front-run large orders. Faster ATs can detect, by posting small quantities of orders on all stocks, when bids are hit by large orders on a specific market (such as BATS, the nearest exchange from Wall-Street). However, the non-filled quantities still have to be executed on other markets. The faster AT front-runs the execution algorithm by buying the stocks which it can then sell on at a higher price when the rest of the large buy order finally arrives a few microseconds later.

Latency arbitrage is facilitated by a loophole in market regulation in the face of continued technological advancement. In terms of market microstructure the NBBO is determined by the Securities Information Processor (SIP) which collects data from the fourteen official US stock markets. The process takes time (milliseconds) to gather prices, devise a harmonised quote and then disseminate the NBBO prices. The SIP depends mainly upon the technology used by stock exchanges. These stock exchanges are not incentivised, e.g. by the Regulation National Market System (Reg NMS), to update their technology regularly. When ATs have faster private technologies, coupled with a direct feed to the exchange, they can effectively see the future prices in advance (the prospective NBBO) and profit accordingly. The potential profits available from latency arbitrage are substantial and have been estimated to be worth as much as up to $3 billion annually.

To ensure liquidity for large block trades several investment banks established private exchanges or dark pools - eschewing the clarity and transparency of public exchanges. Liquidity re-direction allows ATs to manipulate the "best available price" rule. By using repeated pinging of small orders (e.g. several 100-share orders), instead of executing the large orders all at once, ATs signal activity and exhaust the supply of counterparties willing to trade at current prices. The next best available price is then only available on the bank's own internal dark pool which channels investors' funds there.

Marking the close involves the artificial manipulation of prices at the time of the market close. A rush of orders can artificially inflate or depress the closing price of a security and may have a particularly strong impact upon Imbalance Only Orders (IO) that come into effect around the time of market close. Having been accused of such price manipulations an AT firm in New York, Athena Capital, reached a $1 million settlement with the SEC in October 2014.

References

[1] Aït-Sahalia, Y., Sağlam, M., 2017 High frequency market making: Implications for liquidity. Preprint.

[2] Aldridge, I., 2014. High-frequency runs and flash-crash predictability. J. of Portf. Manag. 40, 113-123.

[3] Aldridge, I., Krawciw, S., 2017. Real-time risk what investors should know about FinTech, high-frequency trading and flash crashes. Wiley.

[4] Alzaid, A. A., Omair, M. A., 2013. On the Poisson difference distribution inference and applications. Bull. of the Malays. Math. Sci. Soc. 33, 17-45.

[5] Barndorff-Nielsen, O., Pollard, D., Sheppard, N., 2012. Integer-valued Lévy processes and low-latency financial econometrics. Quant. Financ. 12, 587-605.

[6] Benos, E., Sagade, S., 2012. High-frequency trading behaviour and its impact on market quality: evidence from the UK equity market. Bank Of Engl. Q. Bull. 52, 370.

[7] Bershova, N., Rakhlin, D., 2013. High-frequency trading and long-term investors: a view from the buy-side. J. of Invest. Strateg. 2, 25-69.

[8] Bethel, E. W., Leinweber, D., Rübel, O., Wu, K., 2011. Federal market information technology in the post flash crash era: roles for supercomputing, in Proceedings of the fourth workshop on High performance computational finance (WHPCF '11), ACM, New York, NY, USA, pp. 23-30.

[9] Bingham, N. H., Fry, J. M., 2010. Regression: Linear models in statistics. Springer, London Dordtrecht Heidelberg New York.

[10] Boultona, T. J., Braga-Alvesb, M. V., Kulchania, M., 2014. The flash crash: An examination of shareholder wealth and market quality. J. of Financ. Intermed. 23, 140-156.

[11] Brandt, T., Neumann, D., 2015. Chasing lemmings: Modeling IT-induced misperceptions about the strategic situation as a reason for Flash Crashes. J. Of Manag. Inf. Syst. 31, 88-108.

[12] Brogaard, J., Hendershott, J., Hunt, S., Latza, T., Pedace, L., Ysusi, C., 2012. High-frequency trading and the execution costs of institutional investors. Driver Review 21, Foresight Project on the Future of Computer Trading in Financial Markets.

[13] Brogaard, J., Hendershott, T., Riordan, R., 2013. High frequency trading and price discovery. Rev. of Financ. Stud. 27, 2267-2306.

[14] CFTC and SEC, 2010. Findings regarding the market events of 6 May 2010. Report by the US Commodity Futures Trading Commission and the US Securities and Exchanges Commission.

[15] Chaboud, A. P., Chiqouine, B., Hjalmarsson, E., Vega, C., 2014. Rise of the machines: algorithmic trading in the foreign exchange market. J. of Financ. 49, 2045-2084.

[16] Cheah, E-T., Fry, J. M., 2015. Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. Econ. Lett. 130, 32-36.

[17] Chernov, D. Sornette, D.. 2016. Man-made catastrophes and risk information concealment. Springer, Heidelberg New York Dordtrecht London.

[18] Cliff, D., Northrop, L., 2012. The global financial markets: an ultra-large-scale systems perspective, in Calinescu, R. and Garlan, D. (Eds.), Monterey Workshop 2012, LNCS 7539, pp. 29-70.

[19] Cui, B., Gozluzku, A. E., 2016. Intraday rallies and crashes: Spillovers of trading halts. Int. J. of Financ. and Econ. 21, 472-501.

[20] Easley, D., Lópezde Prado, M. M., O'Hara, M., 2011. The microstructure of the "Flash Crash": Flow toxicity, liquidity crashes and the probability of informed trading. J. of Portf. Manag. 37, 118-128.

[21] Filimonov, V., Sornette, D., 2012. Quantifying reflexivity in financial markets: Toward a prediction of Flash Crashes. Phys. Rev. E 85, 056108

[22] Fry, J., 2012. Exogenous and endogenous crashes as phase transitions in complex financial systems. Eur. Phys. J. B 85, 405.

[23] Fry, J., Cheah, E-T., 2016. Negative bubbles and shocks in cryptocurrency markets. Int. Rev. of Financ. Anal. 47, 343-352.

[24] Grimmett, G. R., Stirzaker, D., 2003. Probability and random processes, third ed. Oxford University Press, Oxford.

[25] Hagstromer, B., Norden, L., 2013. The diversity of high-frequency traders. J. of Financ. Mark. 16, 741-770.

[26] Hardiman, S., Berot, N., Bouchaud, J-P., 2013. Critical reflexivity in financial markets: a Hawkes process analysis. Eur. Phys. J. B 86, 1-9.

[27] Hudson, R. S., Cai, C. X., Keasey, K., Harris, J. H., 2015. Informed trading and market structure. Eur. Financ. Manag. 21, 148-177.

[28] Jain, C., Jain, P., McInish, T. H., 2012. Short Selling: The Impact of SEC Rule 201 of 2010. Financ. Rev 47, 37-64.

[29] Johansen, A., Ledoit, O., Sornette, D., 2000. Crashes as critical points. Int. J. of Theor. and Appl. Financ. 3, 219-255.

[30] Karlis, D., Ntzoufras, I., 2003. Analysis of sports data using bivariate Poisson models. J. of the R. Stat. Soc. D 52, 381-393.

[31] Leal, S. J., Napoletano, M., Roventini, A., Fagiolo, G., 2016. Rock around the clock: An agent-based model of low – and high – frequency trading. J. of Evol. Econ. 26, 49-76.

[32] Lewis, M., 2014. Flashboys: A Wall Street Revolt. W. W. Norton & Company, New York.

[33] Luckock, H., 2003. A steady-state model of the continuous double auction. Quant. Financ. 3, 385-404.

[34] Madhavan, A., 2012. Exchange-traded funds, market structure and the Flash Crash. Financ. Anal. J. 68, 20-35.

[35] Malinova, K. Park, A., Riordan, R., 2013. Do retail traders suffer from high frequency traders? Preprint.

[36] Manahov, V., 2016. Front-running scalping strategies and market manipulation; Why does high-frequency trading need stricter regulation? The Financ. Rev. 51, 363-402.

[37] McInish, T., Upson, J., Wood, R. A., 2014. The Flash Crash: Trading aggressiveness, liquidity supply, and the impact of Intermarket Sweep Orders. Financ. Rev. 49, 481-509.

[38] Menkveld, A. J., 2014. High-Frequency Traders and Market Structure. Financ. Rev. 49, 333-344.

[39] Miller, R. S., Shorter, G. 2016. High frequency trading: Overview of recent developments. *Congressional Research Service Report R44443.*

[40] Nagata, S., Inui, K., 2014. Does High-Speed Trading Enhance Market Efficiency? Empirical Analysis on “Arrowhead” of the Tokyo Stock Exchange. The J. of Trading 9, 37-47.

[41] Reinhart, C. M., Rogoff, K., 2009 This time its different: Eight centuries of financial folly. Princeton Univesity Press, Princeton.

[42] Schlepper, K., 2016. High-frequency trading in the bund futures market. Bundesbank Discussion Paper No. 15/2016.

[43] Serbera, J.-P., Paumard, P., 2016. The fall of high-frequency trading: a survey of competition and profits. Res. in Int. Bus. and Financ. 36, 271-287.

[44] Serritella, D. M., 2010. High speed trading begets high speed regulation: SEC response to flash crash. U. Ill. J. Tech. and Pol., 2010, 433-444.

[45] Shaw, W. T., Schofield, M., 2015. A model of returns for the post-credit-crunch reality: hybrid Brownian motion with price feedback. Quant. Financ. 15, 975-998.

[46] Tokic, D., 2015. ‘Cancel if close to market’: The curious case of market manipulation. J. of Asset Manag. 16, 149-155.