Financial Conduct Authority

Occasional Paper 37

June 2018

Flash Crash in an OTC Market

Florian Schroeder, Matthew Allan, Andrew Lepone, Henry Leung and Stephen Satchell

FCA occasional papers in financial regulation

The FCA occasional papers

The FCA is committed to encouraging debate on all aspects of financial regulation and to creating rigorous evidence to support its decision-making. To facilitate this, we publish a series of Occasional Papers, extending across economics and other disciplines.

The main factor in accepting papers is that they should make substantial contributions to knowledge and understanding of financial regulation. If you want to contribute to this series or comment on these papers, please contact Karen Croxson or Kevin James at karen.croxson@fca.org.uk and kevin.james@fca.org.uk.

Disclaimer

Occasional Papers contribute to the work of the FCA by providing rigorous research results and stimulating debate. While they may not necessarily represent the position of the FCA, they are one source of evidence that the FCA may use while discharging its functions and to inform its views. The FCA endeavours to ensure that research outputs are correct, through checks including independent referee reports, but the nature of such research and choice of research methods is a matter for the authors using their expert judgement. To the extent that Occasional Papers contain any errors or omissions, they should be attributed to the individual authors, rather than to the FCA.

Authors

Florian Schroeder, Matthew Allan, Andrew Lepone, Henry Leung and Stephen Satchell.

Florian Schroeder and Matthew Allan were working in the Markets Policy Department of the Financial Conduct Authority while conducting the research. Andrew Lepone is a Professor at Macquarie University in Sydney, Henry Leung is a lecturer at the University of Sydney Business School and Stephen Satchell is a Professor at the University of Sydney Business School and a Fellow of Trinity College Cambridge.

Acknowledgements

We would like to thank Matteo Aquilina, Karen Croxson, Edwin Schooling Latter, Tom Steffen, Toby Wallis and Terry Walter for their helpful comments and discussions over the course of the research project. Florian Schroeder thanks the Capital Markets Cooperative Research Centre (CMCRC) for funding a period as a visiting researcher at the FCA during this study. The views expressed in the paper are those of the authors. All remaining errors and omissions are our own.

All our publications are available to download from www.fca.org.uk. If you would like to receive this paper in an alternative format, please call 020 7066 9644 or email publications_graphics @fca.org.uk or write to Editorial and Digital Department, Financial Conduct Authority, 25 The North Colonnade, London E14 5HS or Financial Conduct Authority, 12 Endeavour Square, London E20 1JN (new address from 1 July 2018).

Contents

1 Overview 3

Summary 3

Key findings 4

2 Introduction 6

The GBP/USD flash crash 6

Theoretical frameworks 6

Empirical studies 7

Contribution of this paper 9

3 Data and method 10

Data sample 10

Methodology 11

4 Results 17

Descriptive statistics during the flash crash 17

Contributions of different market participants to the fall in liquidity during the flash crash period 22

Explanations for dealers’ liquidity withdrawal 27

5 Conclusion 33

1 Overview

Summary

The foreign exchange ('FX') spot market is the biggest and most liquid in financial services. Over 5 trillion US Dollars' worth of currencies are traded in this market every day. Corporates, financial institutions and private individuals can exchange currencies immediately in the spot market, enabling them to buy products in foreign countries, or speculate on currency price movements. Alongside the spot market, there is an over-the-counter ('OTC') FX derivatives market. This large market mainly consists of forward contracts[1], which are non-standardised contracts between 2 parties to buy or sell a specified amount of currency at a pre-determined price in the future. This market is used by large corporates or financial institutions to hedge foreign exchange risk, for arbitrage, or for speculation.

In this paper, using proprietary data reported to the FCA under EMIR (the 'European Market Infrastructure Regulation'), we examine the underlying drivers of the flash crash[2] in the spot rate for Pound Sterling vs US Dollar (GBP/USD) in October 2016. To our knowledge, this is the first paper to take use these trade reports to analyse how different market participants react in times of market stress and their impact on the liquidity dry-up in a flash crash. Our research is the first study to examine the underlying drivers of a flash crash in an opaque OTC market; previous analyses have focused exclusively on flash crashes in exchange-traded markets. Finally, our paper is also the first to investigate the impact of derivatives on the underlying spot market during a flash crash. This allows us to test 3 competing theories of flash crashes in an OTC market: (i) order flow toxicity[3]; (ii) limited risk-bearing capacity of market makers[4]; (iii) developments in a related derivative market.

We do not investigate the initial trigger of the GBP/USD flash crash in this paper as it has been discussed comprehensively in previous literature (BIS, 2017). Furthermore, a recently published report by the Bank of England provides an in-depth analysis of the liquidity deterioration in GBP/USD using a number of different liquidity metrics (Noss et al., 2017).

However, there is no research on the behaviour of OTC market participants during flash crash periods. This is largely due to the lack of publicly available transaction data. The behaviour of market participants in OTC markets is of special interest given the significant market volumes executed for particular asset classes (eg foreign exchange) and the different trading structure involved. In contrast to an asset traded on a centralised multilateral trading platform, an asset traded in an OTC market may trade simultaneously

[1]According to our GBP/USD OTC derivatives data set, forward contracts account for 95% of the foreign exchange derivatives market.

[2]A flash crash is defined as a rapid drop in the price of a financial asset occurring within an extremely short time period followed by a quick recovery.

[3]See for example Easley, López de Prado, and O'Hara (2011), Andersen and Bondarenko (2014).

[4]See Kirilenko et al. (2017).

at different prices. This is because trades are agreed bilaterally between 2 counterparties such as an investor and a dealer.[5] Also, dealers have no obligation to provide liquidity on OTC markets. These unique features of the trading architecture mean it is particularly interesting to study how a sudden price movement propagates through the market.

Key findings

Three main behaviours are found to have contributed to the drying-up of liquidity in the GBP/USD OTC during the October 2016 flash crash:

1. Dealers (including investment banks and other banks) withdrew liquidity by widening their spreads and in some cases withdrawing from the market altogether.
    - We show that dealers' trading activity led to higher transaction costs (the round-trip costs were about 60 times higher during the flash crash compared to normal times) as they charged higher bid-ask spreads when providing liquidity to their clients.
    - Dealers also impacted trading volume negatively. Their reduced trading activity led to the drying up of liquidity during the flash crash (their trading volume was less than 1% of its average level during normal times).
    - Other financial firms (eg hedge funds and asset managers) stepped in during the flash crash period and provided liquidity by taking long positions. However, they offered to buy at less competitive prices. Other financial firms were involved in 98% of the traded volume during the flash crash period compared to 35% in normal times. So who did they trade with if dealers had withdrawn from the market? About 55% of these trades were with non-financial firms who mainly took the short position during the flash crash, while 38% of the trades were with each other.
2. The inter-dealer part of this market, which is exclusively used by dealers to hedge their client trades with each other, collapsed almost completely during the flash crash period. During normal trading this part of the market accounts for 61% of all transactions, but this share fell to just 2% during the flash crash.
    - The absence of this key market during the flash crash meant that dealers could effectively hedge only 31% of their client trades during this episode. This is turn may explain why they withdrew liquidity in the non-interdealer part of the market.
    - Without the inter-dealer market, dealers have to face the inventory holding risk for every transaction undertaken. Dealers in OTC markets are only willing to accumulate additional inventory during times of stress if there are large price concessions. This induces a downward pressure on prices and potentially also withdrawal of liquidity.
3. The existence of the FX OTC derivatives market in the spot rate for GBP/USD amplified the initial effects of the flash crash in the underlying spot market.
    - We show cross-market effects and bidirectional causalities between liquidity in the OTC derivatives market and its underlying spot market.

[5] See Deuskar, Gupta, and Subrahmanyam (2011).

- The channel for these bidirectional causalities is that dealers in the derivatives market learn from the underlying spot market (and vice versa) and this can cause a feedback loop in illiquidity between the 2 markets. We can confirm this amplification channel via transaction costs, price dispersion and trading volume.

2 Introduction

The GBP/USD flash crash

While not commonplace, flash crashes have been an increasing phenomenon in financial markets in recent years.[6] These often high-profile episodes can be thought of as short-lived malfunctions of capital markets typically involving a substantial price change and a drying up of liquidity followed by a price reversal.[7]

Our study focuses on the GBP/USD flash crash in October 2016. The GBP/USD rate is one of the major currency pairs used by a large number of corporates and investors every day to conduct cross-border transactions between 2 of the biggest industrialised nations in the world: the United Kingdom and the United States of America. During the 21 minutes of this particular flash crash, the FX rate between GBP/USD dropped by approximately 9% and volatility increased to 17 times its typical level. This was accompanied by a pronounced short-term drying up of liquidity. Institutional and retail investors were affected by the volatile trading during the flash crash when, for example, stop-losses were triggered and positions sold at a significantly reduced value.[8]

Theoretical frameworks

A large body of theoretical literature has addressed the behaviour of market participants during periods of financial distress.

As market making is increasingly provided by participants without formal obligations, a number of papers have focused on the behaviour of non-designated intraday market makers during flash crashes. We see 3 main approaches in this theoretical literature, focusing on: (i) order flow toxicity[9]; (ii) inventory holding costs of dealers[10]; and (iii) spillover effects between connected markets (eg ETFs and the underlying assets of ETFs)[11]. We give brief summaries of these 3 classes of theoretical framework.

(i) Order flow toxicity:

Easley et al. (2012) explain how market makers may induce market crashes in high-frequency trading environments. They focus on the idea of `order flow toxicity'. From the perspective of a market maker, an order flow is toxic where it contains primarily orders of traders who have better information about the fundamental asset price. Where order flow is toxic in this sense, liquidity is likely to be provided at a loss and is consequently

[6]The term 'flash crash' was used the first time in May 2010 when the US equity market crashed at high speed but recovered within 36 minutes. Since then there have been several flash crashes – as well as flash rallies – in other asset classes including for example the flash event in US treasuries in October 2014 or the flash crash in GBP/USD.

[7]See Kirilenko, Kyle, Samadi, and Tuzun (2017).

[8]See 'Understand the Flash Crash' by BlackRock in 2010 which commissioned a survey of 380 retail financial advisors about how their clients were affected by the flash crash in May 2010.

[9]See Easley, López de Prado, and O'Hara (2012).

[10]See Kirilenko et al. (2017).

[11]See Cespa and Foucault (2014).

withdrawn in short order. So, order flow toxicity can cause market makers to suddenly leave the market, setting the stage for episodic illiquidity.

(ii) Inventory holding costs of dealers:

Kirilenko et al. (2017) show that in the absence of a fundamental shock, a flash crash can be triggered by a large sell order from any single trader, leading to a large net change in her daily position.

Kirilenko et al. (2017) ascribe the liquidity withdrawal of non-designated intraday market makers to their limited ability to bear risk in the absence of large price concessions. They build on the equilibrium model of Huang and Wang (2008) which shows that simply the cost of maintaining continuous market presence can be a factor behind or the basis for market crashes, even in the absence of fundamental shocks. The respective price drops need to be large enough to compensate the increasingly reluctant market makers for taking on additional risky inventory and the recovery of demand-side liquidity.

(iii) Spillover effects between connected markets

Using a 2 asset framework, Cespa and Foucault (2014) link liquidity crashes to spillovers effects that can arise when dealers who are specialised in different assets learn from each other's prices. When 1 asset becomes less liquid due to an initial shock, its price becomes less informative for dealers who are specialised in the other asset. These dealers face more uncertainty and require larger price concessions for providing liquidity in their chosen asset market. The fall in liquidity propagates from the first asset to the second. By the same mechanism, the drop in liquidity in the second asset market feeds back to the first asset market, further eroding liquidity and amplifying the initial shock. The ultimate impact on the liquidity of each asset is bigger than the immediate impact of the shock.

Empirical studies

A number of academic studies examine the underlying drivers of the flash crash of 6 May 2010 empirically.

The theoretical framework of order flow toxicity is studied by Easley et al. (2011). They examine the flash crash in US equities and present evidence that during this period order flow was becoming increasingly toxic for market makers by using a new measure called VPIN ('Volume Synchronised Probability of Informed Trading').

Dealers' inventory holding costs as an underlying driver for flash crashes is studied by Kirilenko et al. (2017). They use audit trail transaction data for the E-Mini S&P 500 futures market and provide evidence that the behaviour of non-designated market makers is consistent with the theory of limited risk-bearing capacity. Nevertheless, the most active intraday intermediaries (high-frequency traders) do not significantly alter their inventory dynamics when faced with large liquidity imbalances.

In addition, Menkveld and Yueshen (2016) provide evidence that cross-market arbitrage first weakened and then broke down completely in the minutes leading up to the flash crash. Madhavan (2012) finds that the impact of the flash crash across stocks is systematically related to prior market fragmentation which is measured by quote competition between exchanges. He also shows divergent intraday behaviour of trade and quote fragmentation on the day of the flash crash itself. Finally, McInish et al. (2014)

examine the trading aggressiveness proxied by intermarket sweeping orders (ISO) which sweep down the price levels of order books of several market places. They find that it was significantly higher during the flash crash.

These studies focus solely on the flash crash in US equities on 6 May 2010 and contribute to the overall understanding of flash crashes in exchange traded markets. However, so far there are no empirical studies about the flash crash in OTC US treasuries on 15 October 2015 or the flash crash in GBP/USD on 28 October 2016. The problem researchers usually face is the lack of and access to data about OTC transactions.

There are a few empirical studies examining OTC markets in financial crises but not particularly on flash crashes. For example, Dick-Nielsen, Feldhütter, and Lando (2012) analyse the liquidity components of corporate bond spreads during 2005-2009 using TRACE (Trade Reporting and Compliance Engine) transaction data. They provide evidence that the component of the bond spread originating from illiquidity increases dramatically with the onset of the subprime crisis and this increase is stronger for short-lived speculative grade bonds than for investment grade bonds. Moreover, they show that bonds become less liquid when financial distress hits a lead underwriter resulting in the drying-up of the liquidity of bonds issued by financial firms under crises. Benos and Zikes (2014) use proprietary transaction data of the UK government bond (gilt) secondary market between January 2008 and June 2011 to study the determinants of gilt liquidity. They conclude that liquidity deteriorated significantly during the global financial crisis due to increased funding costs and aggregate market uncertainty. In addition, they document that the reduction in market liquidity was associated with frictions in the inter-dealer market above and beyond the effect of funding costs and aggregate uncertainty. Gissler (2017) studies corporate bond liquidity using regulatory data including transactions from the U.S. corporate bond market from 2005 to 2013. He finds that the liquidity of corporate bonds co-moves with other bond liquidity if they are traded by the same dealer. Furthermore, he employs a case study of bonds that are mainly traded by a major dealer that went bankrupt in 2008. He shows that these bonds were still more illiquid than comparable bonds after 1 month of bankruptcy.

A recent series of FCA publications has studied flash crashes in equity markets. In the FCA publication 'Catching a falling knife: an analysis of circuit breakers in UK equity markets', the authors found that high-frequency trading (HFT) firms in aggregate act as a partial stabilizing force during these periods, by buying on order-book markets during the price fall. In contrast, large investment banks appear to contribute the most to the price fall by selling heavily during the same period.[12]

A second FCA publication named 'How do participants behave during Flash events? Evidence from the UK equity market' examines 'mini flash-crashes/rallies' in FTSE350 stocks. These are smaller abnormal price movements that reverse within a short time window. They are associated with high levels of traded volume. The authors similarly find that large investment banks appear to drive the extreme price movements by trading aggressively in the direction of the price change. High frequency traders (HFTs) initially lean against the wind, by trading in the opposite direction, but then follow and exacerbate the initial price change. Both types of firm however continue to provide some liquidity and do not withdraw completely from the order book during the events.[13]

[12]See Allan and Bercich (2017).
[13]See Aquilina et al. (Forthcoming).

Contribution of this paper

The focus of the current paper is the October 2016 flash crash in the GBP/USD exchange rate.

Using proprietary FCA derivatives transaction data, we investigate the dynamics of the behaviour of market participants in the opaque over-the-counter (OTC) market during this period of extreme volatility. Our research is innovative in studying the opaque OTC market, where previous research focuses solely on exchange-traded markets.

We find that, dealers (including investment banks) contribute the most to the liquidity dry-up during this time. Dealers reduce their market making activity significantly and widen spreads. We find that other financial firms (such as HFTs and hedge funds) step in to provide some liquidity in their place.

To understand factors driving this dealer behaviour, we test the 3 theoretical frameworks discussed above using our novel dataset.

Firstly, we cannot provide evidence for the theoretical framework of order flow toxicity by Easley et al. (2012). In their model, adverse selection risk for dealers increases during the flash crash period; that is, the extent to which dealers are being adversely selected by informed traders increases. In contrast, we find that dealers are better informed than financial and non-financial firms during the flash crash. One possible explanation could be the unique trading architecture of OTC markets in which dealers set the price in their bilateral negotiations with clients and are not constrained to prices which are determined by the automatic price discovery process of exchange-traded markets.

Secondly, we test the theory of dealers' inventory holding costs developed by Kirilenko et al. (2017) in the context of the OTC markets. We find evidence to support their idea that inventory holding costs of dealers may cause flash crashes in OTC markets. The increased inventory holding risk of dealers can be explained by the collapse of the inter-dealer market during the flash crash, leaving dealers sufficiently less able to effectively hedge their client trades.

Lastly, our study is the first to provide evidence that spillover effects between connected markets play a key role in contributing to illiquidity during a flash crash beyond and above the other channels. We use the theoretical framework of Cespa and Foucault (2014) and show cross-market effects and bidirectional causalities between returns, volatility and liquidity in the OTC derivatives market and its underlying spot market. Dealers in the OTC derivatives market learn from the underlying spot market (and vice versa) and this can cause a feedback loop in illiquidity between the 2 markets.

Overall, our paper's findings are consistent with the conclusions of other FCA studies. We, too, find that dealers (mainly investment banks) contribute the most to the liquidity dry-up during the flash crash. In this case by dramatically reducing their activity and widening spreads. We find that other financial firms (such as HFTs and hedge funds) step in to provide some liquidity in their place.

3 Data and method

Data sample

To conduct our empirical analysis of the GBP/USD flash crash, we use 2 different data sets, which contain GBP/USD currency pair transactions from the OTC foreign exchange (FX) derivatives market and the spot (cash) FX market.

Firstly, we use the FCA's proprietary EMIR data which covers all OTC derivatives transactions executed by UK counterparties or foreign counterparties in the UK, or on UK products (for example, securities denominated in GBP). We collect the EMIR reporting data from 3 trade repositories, namely The Depository Trust & Clearing Corporation ('DTCC'), UnaVista and the Chicago Mercantile Exchange ('CME'). This is estimated to cover about 90% of the GBP/USD derivatives market according to analysis by the European Systemic Risk Board [14]. EMIR reporting data include double-sided reporting [15] and occasional misreporting, which we account for and delete. We created a software algorithm to clean the data which identifies identical derivative transactions based on several product and transaction details. In addition, we exclude transactions with an abnormally high notional amount (due to misreporting) at the top 1% level using a winsorisation method to ensure high data quality. To provide the highest timestamp accuracy, we compare the timestamps of double reports included in our EMIR reporting dataset. As derivative transactions must be reported by both counterparties according to EMIR, we check if both counterparties report the same transaction execution time. If the counterparties report a different timestamp, we create an indicator which measures the data quality of the report provided by a counterparty[16]. We assume that the counterparty which submitted the report with the higher data quality also submitted the more accurate timestamp for the derivative transaction. Our cleaning algorithm deletes the double report with the lower data quality. If the data quality indicator of both reports is equal we choose the report with the earliest timestamp.

Secondly, we use spot FX rates obtained by Thomson Reuters Tick History. According to Thomson Reuters, these are 'rates which have been contributed to the Thomson Reuters network, as indicative deal-able rates, by Thomson Reuters FX customers. These are customers who are usually active in the interbank or wholesale/institutional FX markets'[17].

Both data sets cover the period from 30 September 2016 to 13 October 2016. As there are potential day-of-week effects in the FX market, our dataset is chosen to cover 5 weekdays before and after the flash crash on 6 October 2016. We typically refer to the 5 days before and 5 days after the flash (excluding the 21 minutes during which the flash crash occurred) in our report as 'normal times'. In using a relatively brief period before and after the flash crash, the intention is to reduce the likelihood that other events significantly impacted the

[14]See https://www.esrb.europa.eu/pub/pdf/occasional/20160922_occasional_paper_11.en.pdf.

[15]Double-sided reporting means that both sides of a derivatives trade have to report their trades to be compliant with EMIR.

[16]Our data quality indicator sums up how many fields are missing in the report. Important variables like notional amount and price of a derivative are weighted higher. The higher the indicator the worse is the quality of the report.

[17]See https://www.sirca.org.au/2010/11/tick-history-foreign-exchange-instrument-codes.

FX GBP/USD rate during our event window. The roughly 21-minute period where the sterling price falls approximately 9% and then mostly recovers is what we describe as the 'flash crash period'.

Methodology

Classification of market participants:

To analyse the contributions of different market participants to the flash crash, we classify each trader into 1 of 3 groups:

Dealers: This group consists of large investment banks and medium-size banks which act as dealers in the OTC derivatives market by executing OTC derivatives contracts with their clients (financial firms or non-financial firms). Dealers provide liquidity and charge a bid-ask spread to their clients. The bulk of the transactions take place in the interdealer part of the OTC derivatives market where dealers hedge the risk of their derivatives positions with other dealer firms.

Financial Firms ('FF'): Financial firms in our dataset consist of asset managers, small brokers, exchange-traded funds ('ETF'), hedge funds, high frequency traders ('HFT'), insurance companies, pension funds, unclassified funds and others (eg financial advisers). These firms are expected to have financial knowledge and superior market information compared to non-financial firms.

Non-Financial firms ('NF'): This group includes corporates and private individuals. They are expected to be uninformed investors who have a less sophisticated understanding of asset valuations and do not have superior market information.

Key variables:

Our main variables are returns, volatility and liquidity (where we deploy 4 separate measures).

Returns: We calculate the return as the logarithmic difference of the current price and the previous price. Prices are the exchanges rates between GBP/USD available in the OTC derivatives market ('Reto')[18] and spot market ('Rets') at time t.

Volatility: We apply a GARCH(1, 1) model to estimate the variance ϑ_t^2 of the exchange rates in the OTC derivatives market and the spot market. It is based on the following equation where r_t + is the return at time t and ω,ɑ,β are positive parameters:

$$1)\ \boldsymbol{\vartheta_t^2 = \omega + \alpha r_{t-1}^2 + \beta \vartheta_{t-1}^2}$$

A maximum likelihood method is used to estimate ω, ɑ and β. The estimated volatility ϑ_t is from now on expressed as 'Volo' and 'Vols' for the OTC derivatives market and the spot market, respectively.

Liquidity spot: We measure liquidity in the spot FX market by calculating the spread between bid and ask rates. The transaction costs are from now on expressed as 'Liqs'.

Liquidity OTC – round trip costs: To calculate spreads of transactions in the OTC derivatives market, we use an approach described by Feldhütter (2011). In dealer markets, Feldhütter (2011) suggests an estimate of round-trip costs which is the difference between the price

[18]EMIR reporting data include the price of the derivative's underlying (e.g. exchange rate for currencies).

at which a dealer sells an asset to a customer and the price at which a dealer buys an asset from a customer. Regarding derivative markets, the buy-side belongs to the payer of leg 1 and the sell side will be the payer of leg 2. EMIR reporting data includes a flag stating the counterparty side for each derivative transaction. To estimate the round-trip costs, we subtract the average buy prices from average sell prices[19]. The estimated liquidity in both markets is from now on expressed as 'Liqrt'.

Liquidity OTC – price dispersion: We calculate price dispersion in the OTC market as an additional measure of liquidity, following Jankowitsch, Nashikkar, and Subrahmanyam (2011). Traded prices may deviate from the expected market valuation of an asset due to the presence of inventory risk for dealers and search costs for investors. Suppose we have $\mathrm{NoT_t}$ transactions during period t.

$$2)\ \mathbf{Liqdisp_t} = \sqrt{\sum_{k=1}^{\mathrm{NoT_t}} \frac{Vol_{k,t}}{\mathrm{Liqvol_t}} \left(\frac{P_{k,t}-m_t}{m_t}\right)^2}$$

The measure can be described by $\mathrm{Vol_{k,t}}$ and $\mathrm{P_{k,t}}$ with k=1,…, $\mathrm{NoT_t}$ where $\mathrm{Vol_{k,t}}$ is the volume of the k th transaction and $\mathrm{P_{k,t}}$ is the corresponding price. $\mathrm{m_t}$ is the average price during this period and $\mathrm{Liqvol_t}$ is the aggregated volume at time t.

Liquidity OTC – trading volume: To measure the liquidity we use the aggregated trading volume $Liqvol_t$and number of transactions NoT_t.

All variables are timestamped to the second.

Empirical models:

Our empirical analysis is structured as follows. Firstly, we conduct a panel regression to analyse the impact of different market participants on the drying up of liquidity in the OTC derivatives market. Since we find that dealers contribute most to the dry up in liquidity, we examine the underlying drivers of behaviour of dealers using further techniques. We test the 3 theoretical explanations for dealers withdrawing liquidity discussed in our introduction: (1) toxic order flow (2) dealers' inventory holding costs (3) spillover effects between connected markets. Below we describe our regression models in more detail:

Panel regression to determine the contributions to illiquidity of different trader types

We study the contributions of different types of market participants to the drying up of liquidity in the OTC derivatives market. Our empirical approach involves: i) calculating 3 different liquidity metrics including 1 metric proxying the transaction costs (round-trip costs), 1 metric proxying the price dispersion and 1 metric proxying the trading volume (aggregated notional amount) in the OTC derivatives market; and ii) relating the liquidity metrics to the trading behaviour of different market participants via panel regressions[20] on a per second basis as this is the most granular interval possible in our dataset due to the high liquidity in the OTC GBP/USD market. To check the robustness of our results we also re-run our panel regressions on a 10 second interval as the OTC market is still dominated by humans and their reaction time is longer than 1 second. For the panel regressions, we use 3 different statistical methods: i) the standard OLS regression, ii) the OLS regression

[19]See Hong and Warga (2000), Chakravarty and Sarkar (2003).

[20]The cross-sectional dimension in our panel regressions is the type of market participant (dealers, financial firms and non-financial firms) observed at time t,

with fixed effects (fixed market participants to account for issues related to clustered data) and iii) the 2-stage least squares (2SLS) regression.

By using 2-stage least squares (2SLS) regression, we address the potential endogeneity between the trading behaviour of market participants (their trading volume) and round-trip costs or price dispersion in the OTC derivatives market. As described by Feldhütter (2011), sophisticated investors (financial firms) which usually have higher trading volume than unsophisticated investors (non-financial firms) bargain lower transaction costs or less dispersed prices in OTC markets due to their higher search intensity. In the contrary direction, low transaction costs attract investors to enter the market and the trading volume increases. To resolve this potential endogeneity problem, we use 3 different instruments. Quantity is 1 of the 3 components employed in the calculation of the notional amount of derivatives (our proxy for logarithmic trading volume). This is found to be highly correlated to the trading volume but not to the transaction costs. We also use the maturity of a derivative and the lagged logarithmic trading volume (notional amount) as instruments for the trading volume as we can prove that these are highly correlated to the notional amount of a derivative but less to the transaction costs. In line with Bound, Jaeger, and Baker (1995), we test the validity of our instruments by examining the partial R^2 and F statistic on the excluded instruments in the first-stage regression. Based on this validity test, we find that the 3 instruments used are valid and can properly be excluded from the outcome equation.

We will primarily focus our discussions around the 2SLS regression results and employ the OLS and fixed effect regression results as a robustness check.

We estimate the following equation for each second:

3) $$\boldsymbol{Liqo_t = \alpha + \sum_{i=1}^{3} \beta_i Vol_{it} * FC_t + \theta_{FC} FC_t + \sum_{j=1}^{3} \gamma_j C_{Spot_{jt}} + \sum_{j=1}^{2} \delta_j C_{OTC_{jt}} + \varepsilon_t}$$

In our regression model, we use $Liqo_t$ which is 1 of the 3 liquidity measures in the OTC market (Liqrt, Liqdisp or Liqvol) as the dependent variable. The coefficient β_i is used to capture the incremental effects on liquidity that are particular to the trading behaviour of different market participants Vol_{it} (measured as the trading volume in GBP) in normal times and during the flash crash period (measured by the dummy variable FC_t). In addition, the coefficient θ_{FC} captures the general effect of the flash crash on the liquidity in the OTC derivatives market. Finally, we use 3 different control variables $C_{Spot_{jt}}$ (vols, rets and liqs) for the spot market whose incremental effect is captured by the coefficient γ_j. We also use 2 control variables $C_{OTC_{jt}}$ (reto and volo) for the OTC derivatives market whose incremental effect is captured by the coefficient δ_{jt}. Furthermore, α is the intercept and ε_t are the innovations in our regression model.

Testing theoretical explanations for why dealers withdraw liquidity during a flash crash

Hypothesis 1: Order flow toxicity

To test the theoretical framework of order flow toxicity provided by Easley et al. (2012), we examine how adverse selection risk changes during the flash crash period, that is, the extent to which dealers are being adversely selected by informed traders. Specifically, we test the hypothesis that dealers withdraw liquidity during flash crashes because they would mainly trade with market participants who have better information about the fundamental asset price. In this sense, liquidity is likely to be provided at a loss.

So, we measure the information content of trades executed by financial firms, dealers and non-financial firms by adapting the VAR regression framework of Hasbrouck (1991). We use the following system:

$$
4)\quad
\begin{aligned}
r_t &= \gamma^r + \sum_i^n \gamma_i^r\, r_{t-i} + \sum_i^n \gamma_i^{FF} x_{t-1}^{FF} + \sum_i^n \gamma_i^{Dealer} x_{t-1}^{Dealer} + \sum_i^n \gamma_i^{NF} x_{t-1}^{NF} + \varepsilon_t^r \\
x_t^{FF} &= \gamma^r + \sum_i^n \delta_i^r\, r_{t-i} + \sum_i^n \delta_i^{FF} x_{t-1}^{FF} + \sum_i^n \delta_i^{Dealer} x_{t-1}^{Dealer} + \sum_i^n \delta_i^{NF} x_{t-1}^{NF} + \varepsilon_t^{FF} \\
x_t^{Dealer} &= \gamma^r + \sum_i^n \theta_i^r\, r_{t-i} + \sum_i^n \theta_i^{FF} x_{t-1}^{FF} + \sum_i^n \theta_i^{Dealer} x_{t-1}^{Dealer} + \sum_i^n \theta_i^{NF} x_{t-1}^{NF} + \varepsilon_t^{Dealer} \\
x_t^{NF} &= \gamma^r + \sum_i^n \vartheta_i^r\, r_{t-i} + \sum_i^n \vartheta_i^{FF} x_{t-1}^{FF} + \sum_i^n \vartheta_i^{Dealer} x_{t-1}^{Dealer} + \sum_i^n \vartheta_i^{NF} x_{t-1}^{NF} + \varepsilon_t^{NF}
\end{aligned}
$$

r_t is the absolute change in the GBP/USD OTC derivatives prices and x_t^{FF}, x_t^{Dealer}, x_t^{NF} are the signed transaction volumes (in GBP) executed by financial firms (FF), dealers or non-financial firms (NF). In our VAR regression system, we use the coefficient $\gamma_i^{r,FF,Dealer,NF}$ to capture the incremental effects on price changes, $\delta_i^{r,FF,Dealer,NF}$ on the signed transaction volume of financial firms, $\theta_i^{r,FF,Dealer,NF}$ on the signed transaction volume of dealers and $\vartheta_i^{r,FF,Dealer,NF}$ on the signed transaction volume of non-financial firms. ε_t^r are innovations based on public information and $\varepsilon_t^{FF,Dealer,NF}$ innovations based on private information owned by the different types of market participants. γ^r is the intercept in our regression model.

Hasbrouck (1991) defines the information impact of a trade as the ultimate impact on the price resulting from the unexpected component of the trade, ie, the persistent price impact of the trade innovation. For this purpose, we estimate the VAR regression framework above. Based on our estimated VAR regression model, we examine the cumulative impulse response and calculate the persistent price impact after 10 seconds in time of a transaction executed by each type of market participant.

Hypothesis 2: Inventory holding costs of dealers

Here we test the hypothesis that the withdrawal of liquidity during the flash crash is due to limitations on the risk-bearing capacity of dealers during that period. We apply the approach of Kirilenko et al. (2017) and empirically study the second-by-second co-movement of dealer inventory changes and price changes but we adjust their framework to the special characteristics of OTC derivatives markets. The special characteristic of the OTC derivatives market is that it more easily allows traders to hedge or bet on falling prices by selling an OTC derivatives contract (short-selling). In contrast, selling an asset in an exchange-traded market usually leads to the closing of the trader's position.

The change in inventory ΔInv_t of a OTC derivative dealer (also known as net exposure) is determined by:

5) $\Delta \boldsymbol{Inv_t} = \boldsymbol{Vol_t^{Buy}} - \boldsymbol{Vol_t^{Sell}}$

with $\mathrm{Vol}_t^{\mathrm{Buy}}$ being the trading volume on the buy-side and $\mathrm{Vol}_t^{\mathrm{Sell}}$ being the trading volume on the sell side. To test whether there is a statistical relationship between dealers' inventory changes and price changes in the OTC derivatives market and whether it significantly changed during the flash crash, we estimate the following regression:

6) $\Delta \boldsymbol{Inv_t} = \boldsymbol{\alpha} + \boldsymbol{FC_t}(\boldsymbol{\delta} \Delta \boldsymbol{Inv_{t-1}} + \sum_{i=0}^{20} \boldsymbol{\beta_i} \Delta \boldsymbol{p_{t-i}}) + \boldsymbol{\varepsilon_t}$

with $\Delta\mathrm{Inv}_t$ being the inventory change as well as $\Delta\mathrm{p}_t$ being the lagged price changes. Δ and β_i, are the incremental effects of these changes and FC_t is a dummy variable being 1 during the flash crash period and nil in normal times. Furthermore, α is the intercept and ε_t the innovations in our regressions model.

Hypothesis 3: Spillover effects between connected markets

This empirical analysis tests the hypothesis that the withdrawal of dealer liquidity is due to spillover effects between the spot GBP/USD market and the OTC derivatives market. We provide evidence on the interaction between the spot GBP/USD market and its OTC derivatives market. We also examine how flash crashes modelled as unexpected shocks can spill over from one market to the other.

We test the theory of Cespa and Foucault (2014), which shows cross-market effects and bidirectional causalities, and adopt a 6-equation vector autoregression model VAR(K) to produce the following system of equations:

7)
$$X_t = \sum_{j=1}^{K} a_{1j} * X_{t-j} + \sum_{j=1}^{K} b_{1j} * Y_{t-j} + u_t$$
$$Y_t = \sum_{j=1}^{K} a_{2j} * X_{t-j} + \sum_{j=1}^{K} b_{2j} * Y_{t-j} + v_t$$

X_t is a vector including 3 variables (Reto, Volo and Liqrt) of the GBP/USD OTC derivatives market and Y_t is a vector including three variables (Rets, Vols and Liqs) of the GBP/USD spot market.

a_{1j} describes the incremental intermarket effects and a_{2j} the incremental cross-market effects of the lagged variables in the OTC derivatives market. b_{1j} are the incremental intermarket effects and b_{2j} the incremental cross-market effects of the lagged cross-market variables in the GBP/USD OTC derivatives market. K, which is the number of lags in the following system, is based on the Akaike information criterion. Methods and notations used to estimate the VAR regression model are based on Lütkepohl (2005).

Based on the VAR regression model, we employ Granger causality tests and cumulated impulse response functions to provide evidence for the interaction of the spot FX and OTC derivatives market.

Finally, we analyse the underlying channel for the liquidity spillover between the spot market and the OTC derivatives market. The underlying channel posited by Cespa and Foucault (2014) involves dealers in 1 asset class (OTC derivatives GBP/USD) learning from other asset prices (from the underlying spot FX rates). So, we implement the following regression model:

8) $$Liqo_t = \beta_0 + \beta_1 * Ps_t + \beta_2 * Reto_t + \beta_3 * Volo_t + \varepsilon_t$$

We apply the model on all transactions in our data set in which a dealer is involved. As a robustness check, we use 3 different liquidity metrics $\mathrm{Liqo_t}$ namely round-trip costs (Liqrt), price dispersion (Liqdisp) and the logarithmic transaction volume (Liqvol) as the dependent variable. We control for returns $\mathrm{Reto_t}$ and volatility $\mathrm{Volo_t}$ in the OTC derivatives market as we have already shown that these variables have a significant impact on liquidity in the OTC derivatives market. $\mathrm{Ps_t}$ is the spot price at time t. β_0, β_1 and β_3 are the coefficients we are estimating with our regression model and ε_t are the innovations.

As we have information about the trades of market participants in the OTC derivatives market only, we can test for whether dealers learn from prices in the underlying spot market only and not whether dealers learn from prices in the OTC derivatives market.

4 Results

Descriptive statistics during the flash crash

Descriptive statistics for the OTC derivatives and spot market:

Table 1: EMIR Transaction Reporting Dataset

	Pre-Crash	Flash Crash	Post-Crash	Total
Number of OTC derivatives transactions	615,317	30,442	706,359	1,351,938
Number of unique market participants	31,408	4,406	29,018	43,579
Number of unique interlinkages	33,758	4,391	30,698	47,204

Notes: The Pre-Crash period includes transactions from 30/09/16 00:00am to 06/10/16 11:07pm. The Flash Crash period includes transactions on 06/10/16 from 11:07pm to 11:28pm. The Post-Crash period includes transactions from 06/10/16 11:28pm to 13/10/16 11:59pm.

Our proprietary FCA EMIR reporting dataset includes in total 1,351,938 GBP/USD OTC derivatives transactions executed between 30/09/2016 and 13/10/2016. During this timeframe, 43,579 unique market participants were involved in these transactions. There were 47,204 unique pairs of counterparties observed to trade with each other in our complete dataset. The low ratio of unique pairs of trading partners to overall market participants reflects the reality that OTC derivatives are traded in a 'hub-and-spoke' market which is dominated by a few large dealer banks.

Table 2: Descriptive statistics for key variables

		Normal Period				Flash Crash Period				
		Count	Mean	Std. Dev.	Median	Count	Mean	Std. Dev.	Median	Δ Mean
OTC Derivatives	Reto ($*10^{-6}$)	410,015	-0.11	5427.10	0.00	1.214	-16.59	12447.47	-23.84	-16.48
	Volo ($*10^{-6}$)	410,015	3,894	3,794	3,046	1,214	8,964	6,255	6,347	5,071
	Volume (in GBPm)	410,015	5.28	99.44	0.04	1.214	0.81	2.96	0.13	-4.48
	Liqrt (in BP)	214,685	0.54	33.83	0.20	1.094	32.11	104.91	6.61	31.58
	Liqdisp (in BP)	263,384	3.69	29.23	0.38	1.161	52.61	88.74	10.75	48.92
Spot	Rets ($*10^{-6}$)	350,481	0.04	54.93	0.00	810	-17.26	1523.57	0.00	-17.22
	Vols ($*10^{-6}$)	350,481	52	22	47	810	863	1.091	413	811
	Liqs (in BP)	350,481	3.64	1.47	3.60	810	5.94	2.91	5.33	2.30

Notes: This table includes the descriptive statistics of returns (Reto), volatility (Volo), trading volume (Volume), round-trip costs (Liqrt) and price dispersion (Liqdisp) in the GBP/USD OTC derivatives market and returns (Rets), volatility (Vols) and bid-ask spread (Liqs) in the GBP/USD spot market. The statistics are distinguished between a normal period (27/09/16 00:00am to 06/10/16 11:07pm and 06/10/16 11:28pm to 18/10/16 11:59pm) and the flash crash period (06/10/16 11:07pm to 11:28pm). The last right column shows the difference between the mean in the flash crash period and the normal period.

The descriptive statistics confirm high liquidity in both OTC FX derivatives and spot FX markets, as evidenced by the tight spreads in both markets as well as the high average volume (notional amount) transacted in the OTC derivatives market (5.3m GBP per second).

Average spreads are slightly lower in the OTC market compared with the spot FX rates in our data sample (0.54 in OTC derivatives market vs. 3.64 basis points in the spot market). This likely reflects the large amount of inter-dealer activity. However, volatility in OTC derivatives prices is higher compared to spot FX rates, possibly due to the non-transparent nature of the OTC derivatives market.

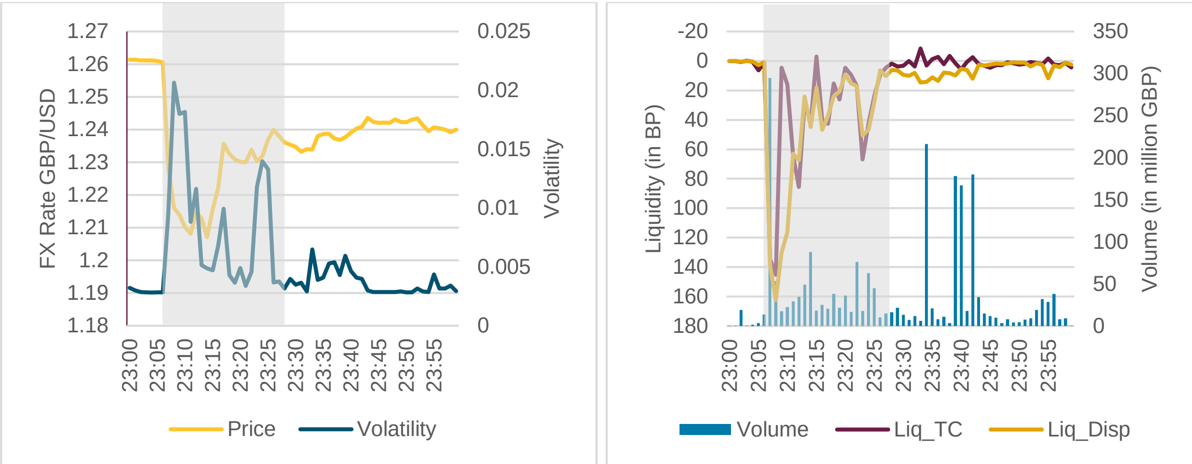
In both markets, volatility is higher during the flash crash period than in normal times. The mean average volume transacted per second in the OTC derivatives market is 0.8m during the flash crash, compared to 5.3m GBP. However, the median volume is higher, which is an indication that fewer large transactions are executed during the flash crash period (see also figure 2). Spreads are wider in both markets during the flash crash periods, but especially in the OTC derivatives market. Round-trip costs go up from 0.54 basis points in normal times to 32.11 basis points during the flash crash period, while price dispersion increases from 3.69 basis points to 52.61 basis points. In the spot market, the increase in spreads is smaller (these widen from 3.64 basis during normal times to 5.94 basis points during the crash) which could be an indication that this market is more resilient under stress.

Figure 1 shows the time-path of key variables in the OTC derivatives market during the flash crash. At 11:07pm on 6 October 2016, the average OTC derivatives price for GBP/USD decreases from 1.27 to 1.21. After about 10 minutes it recovers to 1.24, and then continues to rise slightly for the rest of the hour. The volatility variable shows some very large movements during the flash crash period from 11:07pm and 11:27pm, and afterwards it drops back to the levels prevalent before the flash crash.

Illiquidity, as proxied by round-trip costs and price dispersion, increases considerably in the OTC derivatives market during the flash crash period. In the beginning of the flash crash period, illiquidity increases to 160 basis points, confirming the liquidity dry-up of the market during a flash crash as explained by the theoretical model of Kirilenko et al. (2017). However, it recovers quickly, within 15 minutes, to the old level.

As investors are uncertain about the direction of the market, the average volume of OTC derivatives transactions decreases during the flash crash period. Smaller transactions continue to be executed in the OTC derivatives market during the flash crash period but there are very few large transactions (ie those with a relatively large notional amount).

Figure 1: OTC GBP/USD derivatives during flash crash



Notes: The figure above shows the changes in exchange rates, volatility, liquidity (round-trip costs in basis points) and trading volume (notional amount in million GBP) of OTC GBP/USD derivatives during the flash crash period at October, 6 2016. It starts at 11:07pm and ends at 11:27pm.

Most noticeable in Figure 1 is the high peak in transacted volume (295m GBP) in the OTC derivatives market at 11:07pm. It shows the quick reaction of several market participants in the OTC derivatives market to the flash crash in the underlying spot market. There are no observable large transactions by single market participants.

Drivers of illiquidity during the flash crash:

Table 3 compares the behaviour of the different types of market participants in the normal period to the flash crash period by analysing the number of trades and the transacted volume per second in both periods.

Table 3: Transaction statistics by type of market participant

Panel A: Number of transactions per second by market participants

NoT	Buy (Long Position)			Sell (Short Position)			Net Buy-Sell	
	Normal	Flash Crash	Δ	Normal	Flash Crash	Δ	Normal	Flash Crash
NoT_FF	1.80	11.54	539%	1.80	7.08	294%	0.00	4.46
NoT_NF	1.79	5.94	231%	1.79	15.63	773%	0.00	-9.68
NoT_Dealer	1.55	1.42	-9%	1.68	1.31	-22%	-0.13	0.11

Panel B: Trading volume in GBPm per second by market participants

Volume	Buy (Long Position)			Sell (Short Position)			Net Buy-Sell	
	Normal	Flash Crash	Δ	Normal	Flash Crash	Δ	Normal	Flash Crash
Vol_FF	1.03	0.35	-67%	0.72	0.29	-60%	0.32	0.06
Vol_NF	0.55	0.22	-61%	1.03	0.33	-68%	-0.48	-0.11
Vol_Dealer	37.31	0.24	-99%	37.00	0.35	-99%	0.31	-0.11

Notes: Panel A of the table shows the average number of transactions per second (NoT) and panel B of the table shows the average volume of transactions per second (in GBP). The numbers are broken down into 3 types of market participants: Financial Firms (FF), Non-Financial Firms (NF) and Dealer. There is also a differentiation as to whether the market participant has the long or short position of the GBP/USD OTC derivative. Finally, the numbers are split into transactions executed in normal times (before and after the flash crash in our data sample) and transactions executed during the flash crash. The delta (Δ) shows the difference in % between these 2 periods.

Table 3 shows that dealers reduce the number of GBP/USD OTC derivative transactions per second by about 9% (long positions) and by about 22% (short positions) during the flash crash period. Financial firms in aggregate start to buy GBP/USD OTC derivatives (ie increase their long positions by 539%) during the flash crash period while non-financial firms mostly sell the same derivatives in this period (increase their short positions by 773%).

The findings indicate that non-financial firms try to hedge against falling prices of GBP or speculate on even greater losses of the currency which could indicate following the crowd. In contrast, financial firms have already identified the short-term phenomenon of the falling prices during the flash crash and speculate for a recovery of the currency in the longer term. Dealers in the OTC markets leave in turbulent times and stop providing liquidity.

Table 3 Panel B indicates that the average volume of GBP/USD OTC derivatives transactions per second declined considerably during the flash crash period. Dealers decreased the volume of their transactions the most among all 3 types of market participants (-99%). The decrease of average volume during the flash crash period is independent of the side of the transaction.

Underlying reasons for liquidity withdrawal:

Panel A of Table 4 shows the combinations of market participants in GBP/USD OTC derivatives transactions in normal times compared with the flash crash period. In normal times, dealers are involved in 85% of all transactions in the GBP/USD OTC derivatives market. The interdealer market (dealer-dealer) represents the largest proportion of the GBP/USD OTC derivatives transactions (61%) as the core function of this market is to quickly lay off risk to other dealers incurred in trading with customers. However, the proportion completely changes during the flash crash period and the largest proportion of transactions are executed between financial firms and non-financial firms (55%) and between financial firms with other financial firms (38%). Dealers are involved in just 7% of the total transaction volume during the flash crash period.

Table 4: Combinations of market participants in GBP/USD OTC derivatives transactions

Panel A: Combinations of counterparties

Normal Period			Flash Crash Period		
Combination	Trading Volume (in GBPm)	in %	Combination	Trading Volume (in GBPm)	in %
Dealer-Dealer	1,318,334	61%	NF-FF	539	55%
Dealer-FF	427,949	20%	FF-FF	377	38%
NF-FF	274,931	13%	Dealer-FF	48	5%
Dealer-NF	92,843	4%	Dealer-Dealer	15	2%
FF-FF	48,722	2%	Dealer-NF	0	0%
NF-NF	3,437	0%	NF-NF	0	0%
Total	2,166,215	100%	Total	979	100%

Panel B: Dealer market

Normal Period		Flash Crash Period	
Combination	Trading Volume (in GBPm)	Combination	Trading Volume (in GBPm)
Dealer-Dealer	1,318,334	Dealer-Dealer	15
Dealer-Client	520,791	Dealer-Client	48
Ratio	2.53	Ratio	0.31

Notes: Panel A of the table below shows the combinations of both counterparties in a GBP/USD OTC derivatives transaction and panel B focus on the dealer market. The statistics are distinguished between a normal period (27/09/16 00:00am to 06/10/16 11:07pm and 06/10/16 11:28pm to 13/10/16 11:59pm) and the flash crash period (06/10/16 11:07pm to- 11:28pm). 'FF' stands for Financial Firms (excluding dealers) and 'NF' stands for Non-Financial Firms.

In panel B of Table 4, we focus on GBP/USD OTC derivatives dealer transactions. We distinguish between transactions involving a dealer and a client, as well as transactions between 2 dealers (the interdealer market). The ratio between dealer-dealer transactions and dealer-client transactions shows that dealers unwind all their customer trades immediately in the interdealer market in normal times. By pairing all of it in the interdealer market, dealers can eliminate their inventory holding costs, shift the risk to other counterparties and earn a market-riskless spread for providing liquidity to clients. However, this ratio reduces during the flash crash period as the interdealer market dries up during this period. So, dealers can pair only about 31% of their client trades in the interdealer market and would have to bear the risk of the remaining trades. This could be a potential reason why dealers stop providing liquidity during the flash crash period.

The numbers in Table 4 show that dealers stop providing liquidity in the OTC derivatives market as they are uncertain about the direction of the market. This appears to support the theoretical framework of Kirilenko et al. (2017) showing that dealers leave the market in times of market stress due to their limited risk-bearing capacity without large price concessions. The respective price drop during the flash crash needs to be large enough to compensate increasingly reluctant dealers for taking on additional risky inventory and recovering the liquidity.

Contributions of different market participants to the fall in liquidity during the flash crash period

In Table 5, we report regression results based on the round-trip costs in OTC markets as the dependent variable. As described by Feldhütter (2011), round-trip cost is a metric which measures the spread between the price at which a dealer is on the sell and the price the dealer is on the buy-side. So, it measures how much a dealer charges for intermediation or a client must pay for the dealer's service. Having this understanding of round-trip costs in OTC markets, our results provide evidence that the flash crash has a significant positive impact on the round-trip costs (Dummy_Flash in OLS, OLS(Fixed Effect), 2SLS are significant at 1% level) which proves that round-trip costs increase during the flash crash period and liquidity decreases.

This is particularly ascribed to dealer's activity during the flash crash period. Dealer's trading activity is positively significant with round-trip transaction costs at the 1% level. One possible explanation is that they charge a higher intermediation fee due to the higher uncertainty / risk during this period. This result is robust at the 1% level using 10 second intervals considering a potential longer reaction time of dealers. Moreover, we provide evidence that the trading activity of financial firms during the flash crash period leads to a significant increase in round-trip costs at the 1% level for OLS and OLS (Fixed Effects). However, the economic magnitude of market participant's effect on the round-trip costs is quite small compared to the effect of returns or volatility in the spot and OTC derivatives market.

Table 6 shows the regression results using the price dispersion as the dependent variable. Price dispersion occurs when traded prices deviate from the expected market valuation of an asset. The reasons are trading frictions like the presence of inventory risk for dealers and search costs for investors. Our results provide evidence that the flash crash period has a significant positive impact on price dispersion which means that liquidity decreases during this time. The results are statistically significant at 1% for our 2SLS, OLS (Fixed Effect) and OLS models (neither on a 1 second basis and/or 10 second basis). However, our 2SLS regressions show a significant negative impact on price dispersion of the trading activity of dealer at 1% level during the flash crash.

This shows that dealer's trading activity in times of market stress leads to less dispersed prices as they take on only additional inventory with a price discount and, so, compensate the inventory risk. In addition, trading activity of non-financial firms has a significant negative impact on price dispersion at a 1% level (2SLS regression) during the flash crash period. Again, market participant's trading activity has just a small effect on price dispersion compared to returns and volatility in the spot and OTC derivatives market. Understanding price dispersion as their search costs, this is consistent with Feldhütter (2011) who argues that times of selling pressure are less forceful for unsophisticated investors. In a later section, we will examine the underlying reasons and look at inventory holding costs in more detail.

In Table 7 we examine which type of market participant has an effect on the trading volume (notional amount) in the GBP/USD OTC derivatives market. As the trading activity (volume) of the market participants (independent variables) is a component of the total trading volume in the OTC market (dependent variable), we attempt to overcome endogeneity issues using the 2SLS approach. We find evidence that all types of market participants have a significant positive impact on liquidity (measured by trading volume) during normal times at the 1% level. The flash crash has a significant negative effect on the liquidity in the OTC market at the 1% level according our 2SLS regression (also shown in the descriptive statistics in Table 2).

The bottom of it is that dealers reduce their trading activity in times of stress and are just involved in small transactions (see Table 3). This result is statistically significant at 1% level for our OLS and OLS (Fixed Effects) regressions. Surprisingly, financial firms have a significant positive impact on the trading volume (liquidity) during the flash crash at a 1% level according our 2SLS regression. However, the same effect is significant negative according our OLS and OLS (fixed effect) regressions but our inclination is to rely on the econometrically correct 2SLS approach. Comparing the economic magnitude of the variables, we find that returns and volatility in the spot and OTC market are the main drivers of the trading volume and market participants' trading activity accounts for a small part. Nevertheless, we provide evidence that financial firms provide some liquidity during the flash crash period whereas dealers stop providing liquidity during this time.

Table 5: Effects of market participants on liquidity (round-trip costs)

Metric A: Round trip cost												
Method	OLS		OLS		OLS (Fixed Effects)		OLS (Fixed Effects)		2SLS		2SLS	
Time Interval	1s		10s		1s		10s		1s		10s	
Variables	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Intercept	3.37***	12.20	3.45***	7.68	-0.60	-1.25	-3.71***	-4.62	4.17***	7.31	3.82***	5.53
Reto	766.35***	81.65	931.72***	48.77	763.99***	81.40	923.44***	48.34	604.34***	36.38	943.26***	39.15
Rets	-2352.53***	-4.94	5,367.70***	5.28	-2,342.81***	-4.92	5,254.16***	5.17	-2,144.90***	-3.09	8,468.50***	6.67
Liqs	377.45	0.95	570.46***	0.71	366.78	0.92	667.81	0.84	-169.98	-0.20	375.30	0.30
Vols	-4995.80***	-6.65	7,451.08***	5.05	-5,156.59***	-6.86	7,013.80***	4.75	-2,383.30***	-2.00	14,670.00***	7.74
Volo	-417.70***	-29.34	-451.59***	-19.27	-418.36***	-29.40	-452.31***	-19.31	-414.68***	-14.07	-543.84***	-16.31
Vol_FF	-0.10***	-4.87	-0.15***	-5.84	0.29***	6.47	0.46***	7.24	-0.15***	-4.07	-0.16***	-4.52
Vol_NF	0.02*	1.70	0.02	1.50	-0.45***	-7.42	-0.88***	-9.83	0.01	0.55	0.02	0.78
Vol_Dealer	0.03**	2.13	0.01	0.01	-0.43***	-7.73	-0.47***	-7.46	0.01	-0.05	-0.01	-0.45
Dummy_Flash	13.48***	3.61	29.16***	4.29	13.50***	3.62	27.75***	4.08	24.16***	3.88	9.01	1.04
Vol_FF_Flash	1.47***	4.29	-0.24	-0.51	1.46***	4.26	-0.20	-0.42	0.87	1.57	0.69	1.17
Vol_NF_Flash	0.10	0.61	0.32	1.06	0.13	0.81	0.45	1.48	-0.36	-1.41	0.34	0.95
Vol_Dealer _Flash	1.80***	5.48	0.38***	0.80	1.74***	5.31	0.43	0.91	1.43***	2.96	1.83***	3.07
NOBS	359,676		140,380		359,676		140,380		129,322		91,824	
F-Statistic	710		273		517		210		2,142		2,325	
R^2	0.023		0.023		0.024		0.025		0.016		0.025	

Notes: The table shows the results from our panel regression models. The dependent variable is transaction costs (round-trip costs). The key independent variables Vol_FF, Vol_NF and Vol_Dealer, are the log trading volume in GBP. The same variables ending with 'Flash' are the trading volume in GBP multiplied by a flash crash dummy variable (equals one during flash crash period). We control for further variables (rets, reto, vols, volo, liqs) which should theoretically have an impact on the liquidity in the OTC market. The coefficients of all independet variables are multiplied by 10,000 for better comparability. We apply 3 different models: i) OLS, ii) OLS with market participant fixed effects and iii) 2SLS using instrument variables (quantity, maturity and lagged log volume). The t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively. To check the robustness of our results we apply these models on our dataset with a 1 second and a 10 second interval.

Table 6: Effects of market participants on liquidity (price dispersion)

Metric B: Price dispersion												
Method	OLS		OLS		OLS (Fixed Effects)		OLS (Fixed Effects)		2SLS		2SLS	
Time Interval	1s		10s		1s		10s		1s		10s	
Variables	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Intercept	-6.99***	-35.35	-5.38***	-17.69	-9.51***	-27.34	-13.40***	-24.56	-5.43***	-13.14	-5.58***	-12.22
Reto	-3,324.18***	-494.62	-3,940.86***	-311.50	-3,326.72***	-495.61	-3,951.37***	-313.16	-3,915.90***	-327.35	-4,176.10***	-269.91
Rets	2,253.53***	6.23	5,850.88***	8.16	2,268.69***	6.28	5,584.81***	7.82	2,364.70***	4.44	6,356.20***	7.27
Liqs	410.63	1.44	1,053.79*	1.95	392.91	1.38	1,329.50**	2.46	1,152.70*	1.81	2,224.00***	2.71
Vols	25,170.00***	44.17	24,900.00***	23.85	24,573.54***	43.17	23,714.52***	22.76	22,950.00***	25.16	25,520.00***	19.53
Volo	2,063.52***	200.35	2,165.01***	133.58	2,060.11***	200.33	2,158.50***	133.63	2,572.80***	117.89	2,550.10***	112.94
Vol_FF	0.03*	1.70	-0.19***	-11.04	0.32***	10.02	0.54***	12.49	-0.26***	-9.73	-0.28***	-12.23
Vol_NF	0.10***	13.29	0.12***	12.00	0.45***	10.46	-0.16***	-2.61	0.13***	7.10	0.13***	8.38
Vol_Dealer	0.85***	72.22	0.69***	51.31	0.57***	14.03	0.24***	5.73	0.67***	34.04	0.59***	30.81
Dummy_Flash	0.47	0.17	21.95***	4.58	0.96	0.34	20.28***	4.24	17.20***	3.63	22.19***	3.73
Vol_FF_Flash	1.04***	3.96	-0.66**	-2.00	1.00***	3.82	-0.60*	-1.82	-0.24	-0.56	-0.88**	-2.18
Vol_NF_Flash	-0.06	-0.50	-0.13	-0.60	-0.04	-0.29	-0.01	-0.04	-0.57***	-2.92	-0.15	-0.59
Vol_Dealer _Flash	-0.12	-0.47	-0.85**	-2.53	-0.10	-0.39	-0.68**	-2.04	-0.79**	-2.14	-0.83**	-2.01
NOBS	420,053		149,777		420,053		149,777		141,597		96,283	
F-Statistic	25,070		10,890		17,830		7,806		137,900		100,800	
R^2	0.417		0.466		0.419		0.469		0.493		0.511	

Notes: The table below shows the results from our panel regression models. The dependent variable is price dispersion. The key independent variables Vol_FF, Vol_NF and Vol_Dealer, are the log trading volume in GBP. The same variables ending with 'Flash' are the trading volume in GBP multiplied by a flash crash dummy variable (equals 1 during flash crash period). We control for further variables (rets, reto, vols, volo, liqs) which should theoretically have an impact on the liquidity in the OTC market. The coefficients of all independet variables are multiplied by 10,000 for reasons of better comparability. We apply 3 different models: i) OLS, ii) OLS with market participant fixed effects and iii) 2SLS using instrument variables (quantity, maturity and lagged log volume). The t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively. To check the robustness of our results we apply these models on our dataset with a 1 second and a 10 second interval.

Table 7: Effects of market participants on liquidity (trading volume)

Metric C: Trading volume												
Method	OLS		OLS		OLS (Fixed Effects)		OLS (Fixed Effects)		2SLS		2SLS	
Time Interval	1s		10s		1s		10s		1s		10s	
Variables	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Intercept	6.44***	531.86	11.36***	431.26	4.89***	279.19	8.76***	215.38	11.24***	433.77	14.09***	386.86
Reto	-8.02***	-18.45	0.69	0.62	-6.61***	-19.59	1.93**	2.07	-8.33***	-11.76	-1.86	-1.64
Rets	-14.11	-0.57	358.00***	5.35	1.18	0.06	182.54***	3.22	-16.11	-0.48	351.10***	5.10
Liqs	-60.87***	-3.52	-1,063.00***	-23.01	-74.17***	-5.53	-848.71***	-21.68	-33.47	-0.87	-943.60***	-15.40
Vols	1,293.00***	32.68	1,767.00***	17.98	677.04***	22.06	809.44***	9.71	1,153.00***	19.95	1,521.00***	14.72
Volo	14.89***	22.99	33.61***	22.95	10.79***	21.49	26.53***	21.38	15.47***	11.68	30.81***	17.67
Vol_FF	0.43***	444.39	0.19***	127.65	0.63***	383.66	0.46***	142.27	0.01***	4.46	0.02***	9.58
Vol_NF	0.01***	28.48	0.02***	18.98	0.30***	144.83	0.41***	93.98	0.02***	15.48	0.01***	5.67
Vol_Dealer	0.39***	507.69	0.25***	208.99	0.52***	255.68	0.26***	81.53	0.17***	129.26	0.11***	65.11
Dummy_Flash	0.85***	4.40	2.22***	4.98	0.46***	3.10	0.40	1.06	-1.57***	-4.95	-0.42	-0.86
Vol_FF_Flash	-0.11***	-6.17	-0.20***	-6.50	-0.07***	-4.74	-0.07**	-2.79	0.10***	3.47	-0.02	-0.48
Vol_NF_Flash	-0.01	-1.48	-0.01	-0.64	-0.01	-1.14	-0.02	-1.06	-0.01	-0.91	-0.01	-0.26
Vol_Dealer_Flash	-0.15***	-8.80	-0.22***	-6.83	-0.05***	-3.68	-0.04	-1.58	-0.01	-0.31	-0.05	-1.32
NOBS	539,559		167,331		539,559		167,331		155,372		102,253	
F-Statistic	28,370		3,947		54,330		7,763		28,550		8,459	
R^2	0.387		0.221		0.631		0.441		0.268		0.143	

Notes: The table shows the results from our panel regression models. The dependent variable is trading volume (notional amount). The key independent variables Vol_FF, Vol_NF and Vol_Dealer, are the log trading volume in GBP. The same variables ending with 'Flash' are the trading volume in GBP multiplied by a flash crash dummy variable (equals one during flash crash period). We control for further variables (rets, reto, vols, volo, liqs) which should theoretically have an impact on the liquidity in the OTC market. We apply 3 different models: i) OLS, ii) OLS with market participant fixed effects and iii) 2SLS using instrument variables (quantity, maturity and lagged log volume). The t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively. To check the robustness of our results we apply these models on our dataset with a 1 second and a 10 second interval.

Explanations for dealers' liquidity withdrawal

Order flow toxicity:

In Table 8, we show the persistent impact on the price of a GBP/USD OTC derivatives (in bps) of a trade (in 1,000,000 GBP) executed by a Financial Firm (FF), Dealer and a Non-Financial Firm (NF). All calculated persistent price impacts are located between the upper and lower IRF Monte Carlo integrated error bands.

During the flash period, the persistent price impacts of a trade executed by a financial firm and dealer are negative (-5.3 bps and -10.9 bps) whereas a trade of a non-financial firm has a positive persistent price impact (0.5 bps). The absolute persistent price impact of all 3 types of market participants are higher during the flash crash period compared to the non-flash crash period which is significant at the 1% level using a Welsh test. Our analysis provides evidence that the price sensitivity increases and the market becomes more vulnerable in times of market stress. The increased price impact during the flash crash period could be due to the illiquid market conditions (low trading volume) during the flash crash and, so, a trade (in 1,000,000 GBP) of a market participant has a much bigger price impact than in more liquid times.

Ranking the absolute persistent price impact among the different market participants illustrated in Table 8, shows that financial firms have the largest persistent price impact (0.007 bps) during the non-flash crash period followed by non-financial firms (0.006 bps) and dealer (0.001 bps). Using a paired t-test, we provide evidence that the persistent price impacts are significantly different between financial firms and dealers, between dealer and non-financial firms as well as financial firms and non-financial firms at the 1% level. According to Hasbrouck (1991), these results provide evidence that financial firms are the most informed market participants during the non-flash crash period, followed by non-financial firms and dealers. This could lead dealers to charge higher bid-ask spreads to their clients to compensate for the adverse selection risk during normal times.

This absolute ranking changes in the flash crash period and dealers firms possess the highest amount of private information having a negative persistent price impact of -10.9 bps followed by financial firms (-5.3 bps) and non-financial firms (0.5 bps). The negative impact of dealers on prices during the flash crash could be due to inventory holding risk as they just trade during times of market stress if they get a large price discount (described in more detail in the next section). Using a paired t-test, we provide evidence that the persistent price impacts are significantly different between financial firms and dealers, between dealer and non-financial firms as well as financial firms and non-financial firms at the 1% level.

In contrast to the theoretical framework of Easley et al. (2012), the order flow of dealers does not become more toxic during the flash crash as indicated by the higher persistent price impacts of dealers to financial firms and non-financial firms. As shown in Hasbrouck (1991), the persistent price impact is an indicator for the information content of trades and, according to our analysis, market participants in the OTC GBP/USD derivatives market are unequally informed during a flash crash with dealers having the most private information. In conclusion, order flow toxicity does not appear to be a reason why dealers leave the OTC GBP/USD derivatives market during the flash crash period and cause the liquidity dry-up and price-drop during this time. A possible explanation could be the unique trading architecture of OTC markets in which dealers set the price in their bilateral negotiations with clients and are not constrained to prices which are determined by the automatic price discovery process of exchange-traded markets.

Table 8: Information content of trades by different market participants

Cumulative impulse response (persistent price impacts)

	Non-Flash Crash			Flash Crash Period			
	PPI	Lower Band	Upper Band	PPI	Lower Band	Upper Band	Non-Flash vs Flash
FF	0.007	-0.001	0.023	-5.336	-10.377	-0.353	107.4***
Dealer	0.001	-0.002	0.004	-10.942	-40.872	22.087	30.6***
NF	0.006	-0.06	0.025	0.465	-6.705	5.412	-6.9***

Notes: The table shows the persistent price impact (in BPS) on the GBP/USD OTC derivatives of a trade (in 1,000,000 GBP) executed by a Financial Firm (FF), Dealer and a Non-Financial Firm (NF). We report the persistent price impact (PPI) measured in a cumulative impulse response function after 10 seconds in time and also provide the lower and upper IRF Monte Carlo integrated error bands of the cumulative effect (99% confidence interval). In the last right column, we report results of a Welsh test indicating the statistical difference of the persistent price impact in the non-flash crash against the flash crash period. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively.

Inventory holding costs of dealers

Table 9 reports the empirical results of our OLS regression which is solely based on the transactions in our data set in which a dealer is involved. During the non-flash crash period, we cannot find a significant negative relationship between the contemporaneous price change and inventory change in normal times as provided by Kirilenko et al. (2017). The reason for our deviating results could be the special microstructure of OTC derivatives markets compared to order-book markets. As formalised by Ho and Stoll (1983), the OTC derivatives market contains an inter-dealer market which plays a key role in facilitating risk sharing in OTC markets. In Table 3, we show that the inter-dealer market is 2.5 times bigger than dealer-client market. So, dealers are able to hedge all their client trades in the inter-dealer market during normal times and don’t face inventory holding risk during this time.

During the flash crash period, we find some significant lagged negative price changes: ΔP^{F}_{t-6} , ΔP^{F}_{t-13}, ΔP^{F}_{t-14}, which is largely consistent (except the specific number of lags) with the results of Kirilenko et al. (2017). As market conditions become uncertain during the flash crash, dealers are unwilling to take the opposing side of a hedge of other dealers and the inter-dealer market collapses temporarily (see Table 4). This could be due to binding value-at-risk (VaR) constraints as these are often cited as a prime suspect for the ‘liquidity crunch’ of financial crises[21]. So, dealers in OTC markets who face the inventory holding risk for every trade are only willing to do so with a large price concession.

In summary, we employ the framework of Kirilenko et al. (2017) in the context of the OTC derivatives markets and find evidence that inventory holding costs of dealers may cause flash crashes in OTC derivatives markets. Due to their limited risk-bearing capacity, dealers in OTC markets are only willing to accumulate additional inventory with a large

[21]See Brunnermeier and Pedersen (2008).

price concession during times of stress. This induces a selling pressure which results in large temporary price drops and the drying up of liquidity.

Table 9: Inventory holding costs of dealers

Dependent variable: ΔInv_t					
Variable	Coefficient	t	Variable (cont)	Coefficient	t
α	-1,210.52	0.00			
D_t^F	0.11	0.35			
ΔInv_{t-1}	-0.16	-1.25	ΔInv_{t-1}^F	0.39**	2.59
ΔP_t	6.92	0.39	ΔP_t^F	-8.49	-0.47
ΔP_{t-1}	-14.40	-0.93	ΔP_{t-1}^F	14.21	0.89
ΔP_{t-2}	-22.53**	-2.00	ΔP_{t-2}^F	23.75**	2.04
ΔP_{t-3}	-15.96	-1.10	ΔP_{t-3}^F	18.01	1.20
ΔP_{t-4}	13.68	1.37	ΔP_{t-4}^F	-17.43	-1.62
ΔP_{t-5}	19.32	1.33	ΔP_{t-5}^F	-18.93	-1.29
ΔP_{t-6}	26.64*	1.84	ΔP_{t-6}^F	-33.30**	-2.25
ΔP_{t-7}	5.24	0.45	ΔP_{t-7}^F	-9.66	-0.81
ΔP_{t-8}	-20.99*	-1.83	ΔP_{t-8}^F	10.66	0.77
ΔP_{t-9}	5.28	0.73	ΔP_{t-9}^F	-3.29	-0.34
ΔP_{t-10}	11.95	1.33	ΔP_{t-10}^F	-8.81	-0.87
ΔP_{t-11}	-15.48	-1.45	ΔP_{t-11}^F	12.26	1.12
ΔP_{t-12}	3.64	0.48	ΔP_{t-12}^F	-8.76	-1.06
ΔP_{t-13}	17.92***	2.74	ΔP_{t-13}^F	-19.55***	-2.73
ΔP_{t-14}	12.57	1.46	ΔP_{t-14}^F	-15.45*	-1.73
ΔP_{t-15}	-10.20	-1.16	ΔP_{t-15}^F	10.07	1.09
ΔP_{t-16}	-33.44*	-1.93	ΔP_{t-16}^F	28.80	1.64
ΔP_{t-17}	-9.93	-1.09	ΔP_{t-17}^F	8.10	0.86
ΔP_{t-18}	1.70	0.21	ΔP_{t-18}^F	-3.71	-0.44
ΔP_{t-19}	-15.72	-1.10	ΔP_{t-19}^F	15.40	1.07
ΔP_{t-20}	-3.63	-0.49	ΔP_{t-20}^F	3.34	0.43
NOBS			71,274		
F-Statistic			2.74		
R^2			0.026		

Notes: The table shows our estimated coefficients of the regression shown in (6) based on the transactions of our data set in which a dealer is involved. Similar to Kirilenko et al. (2017), we use the change in inventory as the dependent variable and lagged inventory and price changes as the independent variables. The dummy variable D_t^F indicates the time period from 11:07pm o 11:28pm on October, 6 2016. All coefficients shown are multiplied by 10^6 except for the lagged changes in inventory ΔInv_{t-1}. The t-statistics are corrected for heteroscedasticity in the standard errors using a Newey-West estimator. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively.

Spillover effects between connected markets:

In the following tables and figures, we provide results from our VAR model. First, our approach uses a Granger causality test to examine the interactions between the 2 markets. Secondly, we apply impulse response functions to test how unexpected shocks in 1 of the variables influence the other variables.

Table 10: Granger causality tests

		OTC			**Spot**		
		Reto	Volo	Liqrt	Rets	Vols	Liqs
OTC	Reto	3717.29***	1115.67***	137.30***	5.87***	7.95***	1.12
	Volo	1077.09***	361,559.11***	14.88***	1.09	7.67***	2.59**
	Liqrt	22.13***	11.58***	70.01***	25.28***	17.77***	1.72
Spot	Rets	8.39***	3.90***	21.28***	88.74***	280.49***	1.81
	Vols	3.98***	4.41***	23.31***	243.61***	643,976.51***	3.45***
	Liqs	0.38	4.92***	3.67***	7.68***	112.75***	1317.57***

Notes: The table shows the results of pairwise Granger causality tests between the endogenous variables. The null hypothesis is that variable i does not Granger-cause variable j. We test whether the lag coefficients of i are jointly nil when j is the dependent variable in the VAR. It is estimated with 5 lags. The cell associated with the *i*th row variable and the *j*th column variable shows the F statistic associated with this test. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively.

Table 10 (top left and bottom right corner) provides evidence that returns, volatility and liquidity in one market are Granger-caused by the lagged variables from the same market. Most importantly, returns and volatilities impact liquidity in the same market and vice versa. This has already been examined in several studies[22]. Chordia, Sarkar, and Subrahmanyam (2004) explain it by factors which include the premium for greater trading costs, the psychological bias of loss aversion and inventory risks.

However, our main interest in this analysis is the interaction between the GBP/USD spot market and the respective OTC derivatives market. The results are shown in the top right and bottom left corner of Table 10. We provide evidence that there are significant return and volatility spillovers from the spot to the derivative market even after controlling for the remaining variables. Volatility in the spot market also Granger-causes returns and liquidity in the OTC derivative market. Lastly, returns in the spot market have a significant influence on the liquidity in the OTC derivative market.

The Granger causality test also shows evidence of bidirectional spot and OTC derivatives market causations. There are significant return, volatility and liquidity spillovers from the OTC derivatives market to the spot market after controlling for the remaining variables. Volatility in the OTC derivatives market also Granger-causes returns and liquidity in spot market. Finally, liquidity in the OTC derivatives market Granger-causes returns and volatility in the spot market. Returns in the OTC derivatives market also Granger causes volatility in the spot market.

The empirical evidence confirms the bidirectional causal relationship between the derivatives market and its underlying primary market. In accordance with Cespa and

[22]See for example Benston and Hagerman (1974), Amihud and Mendelson (1986), Subrahmanyam (1994) and Odean (1998).

Foucault (2014), we provide evidence that the prices of the 2 markets are interconnected which is shown in a significant return spillover between the 2 markets. The underlying channel for this phenomenon, as described by Cespa and Foucault (2014), is that liquidity providers / dealers in 1 asset class (GBP/USD OTC derivatives) often learn from other asset prices (from the underlying spot FX rates). If the liquidity of the spot GBP/USD drops, its price becomes less informative for dealers in OTC derivatives. This can lead to a feedback loop and returns, volatility and liquidity spills back from the OTC derivatives market to the underlying spot FX market as dealers in the spot market also learn from the prices in the derivatives market.

Table 11: Cumulated impulse response function

		Unit Impulse					
		Vols	Volo	Rets	Reto	Liqs	Liqrt
Cumulated Response	Vols	40.5778	0.0249	-1.0641	-0.0009	0.1625	0.0048
	Volo	132.5522	17.3908	-2.8030	-0.6492	1.7855	-0.0811
	Rets	1.4910	0.0016	1.0046	0.0003	0.0005	0.0002
	Reto	7.3404	0.7196	-0.0655	0.6942	0.3901	-0.0683
	Liqs	5.6919	0.0133	-0.0951	-0.0006	1.5536	0.0019
	Liqrt	31.2165	-0.0945	-0.4192	0.0226	0.5422	1.0004

Notes: The table illustrates the cumulated response of a variable to a unit standard deviation shock in the endogenous variables. All impulse responses are within Monte Carlo 2 standard error bands which provides evidence for the statistical significance of the response. As the results of the impulse response functions are very sensitive to the specific ordering of the endogenous variables, we fix our ordering of the endogenous variables as following: 'Vols', 'Volo', 'Rets', 'Reto', 'Liqs', 'Liqrt'.

Table 11 provides evidence surrounding the cross-market dynamics between GBP/USD spot and the OTC derivatives market. This analysis reveals how flash crashes (modelled as unexpected shocks) transmit between spot and derivative markets. We focus on the dynamics of volatility, return and liquidity spillover between the 2 markets (highlighted grey in table). In line with the theoretical model of Cespa and Foucault (2014), we assume that the ordering of endogenous variables is the sequence in which the variables are affected[23].

A positive shock in volatility in the spot market leads to a positive persistent volatility effect in the OTC derivatives market. A positive persistent effect is also observed in the other direction – from the OTC derivatives market to the spot market. This is in line with the theoretical framework of Cespa and Foucault (2014) in which derivatives and primary markets are connected through cross-asset learning of dealers. Uncertainty of dealers in asset prices in 1 asset class leads to uncertainty of dealers in the other market.

We also find evidence that a positive shock in returns in the spot market leads to a negative persistent return effect in the OTC derivatives market. On the other side, a positive shock in returns in the OTC derivatives market leads to a positive persistent return effect in the spot market 'Rets'.

Lastly, an unexpected positive liquidity shock in the spot market leads to a persistent liquidity effect in the OTC market. A positive persistent effect is also observed in the other direction – from the OTC market to the spot market. This result provides evidence that a

[23]See Figure 1 on page 1617 in Cespa and Foucault (2014) which demonstrates the feedback loop and the sequence of the affected variables.

flash crash in the spot FX market that eventually causes a liquidity dry-up also leads to a liquidity dry-up in the OTC derivatives market.

Based on a paired t-test, the absolute persistent effect is for all 3 variables stronger in the OTC derivatives market with an unexpected shock from the spot market. However, there is also a persistent but smaller effect from the spot market to the OTC derivatives market which provide evidence for a feedback loop ('multiplier effect') as described by Cespa and Foucault (2014). This shows that a shock in the spot market leads to a persistent effect in the OTC derivatives market, then feeds back to the spot market leading that the ultimate impact is bigger than its immediate effect.

Table 12: Cross-asset learning of dealers

Dependent Variable	Liqrt		Liqdisp		Liqvol	
Variables	coef	t	coef	t	coef	t
Intercept	0.005***	2.70	0.006***	5.24	5.84***	8.69
Ps	-0.003***	-2.47	-0.005***	-5.33	5.19***	9.73
Reto	0.067*	1.79	-0.469***	-20.96	-12.04***	-8.59
Volo	-0.072***	-2.62	0.308***	10.58	7.79**	2.55
NOBS	72,351		80,023		87,054	
F-Statistic	3.8		316.0		64.4	
R^2	0.014		0.539		0.003	

Notes: The table below shows our estimated coefficients of the regression shown in (8) based on the transactions of our data set in which a dealer is involved. The t-statistics are corrected for heteroscedasticity in the standard errors using a Newey-West estimator. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively.

Table 12 provides evidence that dealers in the OTC derivatives market learn from the underlying spot prices. Spot prices have a significant negative impact on round trip costs and price dispersion in the OTC derivatives market. This means that if prices in the spot market go down (for example during an illiquid period in the spot market) dealers in the OTC derivatives market charge a higher spread and the OTC derivatives market also becomes illiquid. This result is confirmed by the positive impact of spot prices on the logarithm of transaction volume. If spot prices go down dealers reduce their transaction volume in the OTC derivatives market and it becomes illiquid. Our results are significant at a 1% level for all 3 metrics using Newey-West estimator to account for heteroscedasticity in the standard errors.

In summary, we shed light on 3 important questions. Firstly, we provide evidence that there are bidirectional and significant volatility and return spillover as well as one-directional liquidity spillover between spot FX and derivatives markets using Granger-causality tests. Secondly, we show the bidirectional dynamics of unexpected shocks between spot FX and derivatives markets, where the interaction effect is even greater than in normal times. Lastly, we provide evidence that these interaction between spot and derivatives market is due to cross-asset learning of dealers as described by Cespa and Foucault (2014)

5 Conclusion

We analysed the underlying drivers of the flash crash in GBP/USD in October 2016 using proprietary FCA derivatives transaction data reported under EMIR. To our knowledge, this study is the first of its kind to examine how the flash crash unfolded in an Over-The-Counter (OTC) market. Other studies have focused on the more transparent exchange-traded markets. The richness of our dataset allows us to study the contribution of different market participants to the drying up of liquidity during the flash crash.

We provide evidence that dealers (mainly investment banks) contributed most to the fall in liquidity during this particular flash crash. By using a 2-stage least squares (2SLS) regression, we address the potential bidirectional relationship between the trading behaviour of market participants (their trading volume) and liquidity. We show that dealers' trading activity led to higher round-trip costs during the flash crash as they charged a higher intermediation fee due to high uncertainty and risk during this time. The significant negative impact of dealers on market liquidity can be confirmed by looking at the impact of dealers on trading volume using OLS and fixed effect regressions. We find that other financial firms (eg hedge funds, asset managers, HFT firms et cetera) step in during the flash crash period and provide some liquidity by taking the long position (according to our 2SLS regression). However, their trading activity still leads to higher round-trip costs as they buy at lower prices to compensate for the increased risk.

To explain dealers' behaviour in the OTC derivatives market during flash crash periods, we empirically study the 3 theoretical frameworks offering potential explanations: order flow toxicity by Easley et al. (2012), inventory holding costs of dealers by Kirilenko et al. (2017) and spillover effects between connected markets by Cespa and Foucault (2014).

Firstly, we cannot provide evidence for the theoretical framework of order flow toxicity by Easley et al. (2012) in the OTC market and cannot provide evidence that dealers are less informed than financial and non-financial firms during the flash crash. By using an approach defined by Hasbrouck (1991) measuring the information impact of trades, we find that adverse selection risks significantly decrease during the flash crash period. Our empirical analysis finds that dealers have the largest persistent absolute price impact followed by financial firms and non-financial firms. Dealers are better informed than financial and non-financial firms during flash crashes and order flow toxicity does not appear to be a reason for dealers leaving the OTC derivatives market during this time. One possible explanation could be the unique trading architecture of OTC markets in which dealers set the price in their bilateral negotiations with clients and are not constrained to prices which are determined by the automatic price discovery process of exchange-traded markets.

Secondly, we employ the framework of inventory holding costs of dealers by Kirilenko et al. (2017) in the context of the OTC markets and find evidence that inventory holding costs of dealers may cause flash crashes in OTC markets. During the flash crash, we show that the inter-dealer market crashes and dealers are not able to effectively hedge their client trades anymore. They would have to face the inventory holding risk for every transaction and, due to limits to their willingness to bear inventory risk, dealers in OTC markets are only willing to accumulate additional inventory with a large price concession during times of stress. This induces a selling pressure which results in large temporary price drops and the drying up of liquidity.

Lastly, our study is the first of its kind which provides evidence that spillover effects between connected markets can be a key factor behind observed illiquidity during a flash crash, above and beyond other channels. We use the theoretical framework of Cespa and Foucault (2014) and show cross-market effects and bidirectional causalities between the OTC derivatives market and its underlying spot market. Dealers in the OTC derivatives market learn from the underlying spot market (and vice versa) and this can cause a feedback loop in illiquidity between the 2 markets. We can confirm this channel for round-trip costs (spreads), price dispersion and trading volume of OTC derivatives transactions in which dealers are involved.

While we have not carried out a comparative study of how OTC and exchange-traded markets react to flash crashes, we can show 1 clear similarity. In both trading architectures, dealers (OTC markets) or market makers (exchange-traded markets) leave the market and cause a liquidity dry-up during the flash crash period due to inventory costs (in OTC and exchange traded markets) and/or order flow toxicity (in exchange-traded markets). Other financial firms step in and provide some liquidity by taking the long position. However, there exist big differences in implementing countermeasures for times of market stress in both market types. In Allan and Bercich (2017), the authors showed that for example, circuit breakers can be a suitable measure to reduce the impact of flash crashes in exchange-traded markets, whereas circuit breakers are not applicable in OTC markets.

Overall, our results deepen the understanding of flash events and the potential risks they pose. A number of steps are already being taken by authorities to limit the impact of flash crashes. These include mandating that certain derivatives trade on more transparent exchange-traded venues rather than primarily on OTC markets, and requiring those venues to have appropriate systems and controls in place to manage excessive volatility, such as circuit breakers where applicable.

References

Allan, M., & Bercich, J. (2017). Catching a falling knife: analysis of circuit breakers in UK equity markets. ‘FCA Insight Article’.

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. 'Journal of Financial Economics, 17(2)', 223-249.

Andersen, T. G., & Bondarenko, O. (2014). VPIN and the flash crash. 'Journal of Financial Markets, 17', 1-46.

Aquilina, M., Eyles, B., Shao, J., & Ysusi, C. (Forthcoming). How do participants behave during Flash events? Evidence from the UK equity market. 'FCA Research Note'.

Benos, E., & Zikes, F. (2014). Liquidity and Dealer Activity in the UK Gilt Market During the Financial Crisis.

Benston, G. J., & Hagerman, R. L. (1974). Determinants of bid-asked spreads in the over-the-counter market. 'Journal of Financial Economics, 1(4)', 353-364.

BIS. (2017). 'The sterling 'flash event' of 7 October 2016'.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. 'Journal of the American Statistical Association, 90'(430), 443-450.

Brunnermeier, M. K., & Pedersen, L. H. (2008). Market liquidity and funding liquidity. 'The Review of Financial Studies, 22'(6), 2201-2238.

Cespa, G., & Foucault, T. (2014). Illiquidity Contagion and Liquidity Crashes. 'Review of Financial Studies, 27'(6), 1615-1660.

Chakravarty, S., & Sarkar, A. (2003). Trading costs in three US bond markets. 'The Journal of Fixed Income, 13'(1), 39-48.

Chordia, T., Sarkar, A., & Subrahmanyam, A. (2004). An empirical analysis of stock and bond market liquidity. 'The Review of Financial Studies, *18*'(1), 85-129.

Deuskar, P., Gupta, A., & Subrahmanyam, M. G. (2011). Liquidity effect in OTC options markets: Premium or discount? 'Journal of Financial Markets, 14'(1), 127-160.

Dick-Nielsen, J., Feldhütter, P., & Lando, D. (2012). Corporate bond liquidity before and after the onset of the subprime crisis. 'Journal of Financial Economics, 103'(3), 471-492.

Easley, D., López de Prado, M. M., & O'Hara, M. (2012). Flow toxicity and liquidity in a high-frequency world. 'The Review of Financial Studies, 25'(5), 1457-1493.

Easley, D., López de Prado, M. M., & O’Hara, M. (2011). The microstructure of the “flash crash”: flow toxicity, liquidity crashes, and the probability of informed trading. 'The Journal of Portfolio Management, 37'(2), 118-128.

Feldhütter, P. (2011). The same bond at different prices: identifying search frictions and selling pressures. 'The Review of Financial Studies, 25'(4), 1155-1206.

Gissler, S. (2017). Lockstep in liquidity: Common dealers and co-movement in bond liquidity. 'Journal of Financial Markets, 33', 1-21.

Hasbrouck, J. (1991). Measuring the information content of stock trades. 'The Journal of Finance, 46'(1), 179-207.

Ho, T. S., & Stoll, H. R. (1983). The dynamics of dealer markets under competition. 'The Journal of Finance, 38'(4), 1053-1074.

Hong, G., & Warga, A. (2000). An empirical study of bond market transactions. 'Financial Analysts Journal, 56'(2), 32-46.

Huang, J., & Wang, J. (2008). Liquidity and market crashes. 'The Review of Financial Studies, 22'(7), 2607-2643.

Jankowitsch, R., Nashikkar, A., & Subrahmanyam, M. G. (2011). Price dispersion in OTC markets: A new measure of liquidity. 'Journal of Banking & Finance, 35'(2), 343-357.

Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The Flash Crash: High-Frequency Trading in an Electronic Market. 'The Journal of Finance, 72', 967-998.

Lütkepohl, H. (2005). Finite Order Vector Autoregressive Processes. 'New Introduction to Multiple Time Series Analysis, Springer-Verlag, Berlin, Heidelberg, New York'.

Madhavan, A. (2012). Exchange-traded funds, market structure, and the flash crash. 'Financial Analysts Journal, 68'(4), 20-35.

McInish, T., Upson, J., & Wood, R. A. (2014). The flash crash: trading aggressiveness, liquidity supply, and the impact of intermarket sweep orders. 'Financial Review, 49'(3), 481-509.

Menkveld, A. J., & Yueshen, B. Z. (2016). The flash crash: A cautionary tale about highly fragmented markets. 'Working paper, Vrije Universiteit Amsterdam'.

Noss, J., Pedace, L., Tobek, O., Linton, O., & Crowley-Ready, L. (2017). The October 2016 sterling flash episode: when liquidity disappeared from one of the world's most liquid markets. 'Bank of England, Staff Working Paper No. 687'.

Odean, T. (1998). Are investors reluctant to realize their losses? 'The Journal of Finance, 53'(5), 1775-1798.

Subrahmanyam, A. (1994). Circuit breakers and market volatility: A theoretical perspective. 'The Journal of Finance, 49'(1), 237-254.

25 The North Colonnade, Canary Wharf, London E14 5HS
Telephone: +44 (0)20 7066 1000
Website: www.fca.org.uk