BANK OF ENGLAND

Staff Working Paper No. 687

The October 2016 sterling flash episode: when liquidity disappeared from one of the world's most liquid markets

Joseph Noss, Lucas Pedace, Ondrej Tobek, Oliver Linton and Liam Crowley-Reidy

October 2017

BANK OF ENGLAND

Staff Working Paper No. 687

The October 2016 sterling flash episode: when liquidity disappeared from one of the world's most liquid markets

Joseph Noss,(1) Lucas Pedace,(2) Ondrej Tobek,(3) Oliver Linton(4) and Liam Crowley-Reidy(5)

Abstract

This paper provides an in-depth analysis of the evolution of liquidity during the flash episode in sterling during the early hours of 7 October 2016. It examines a number of estimates both of the cost of trading, and the price impact of executed transactions. These include a variant of the 'volatility over volume' measure of liquidity based on transaction data, which provides a better proxy of illiquidity — as given by measures based on high-frequency limit order book data — than other summary measures of price impact. The paper also shows that the fall in the value of sterling during the initial part of the flash episode was consistent with the estimated impact on prices of a large number of individually small — but in aggregate large — volume of orders to sell sterling during a normally quiet period of the trading day. However, the subsequent change in price was larger than that consistent with the estimated impact on prices of observed orders to sell sterling. This might support the suggestion, which was included in the report on the episode provided by the Bank for International Settlements, that the move in sterling may have been amplified by the pause in trading on the CME futures exchange.

Key words: Flash crash, foreign exchange market, liquidity, price impact.

JEL classification: F33, F37, G01, G12, G15.

(1) Bank of England. Email: joseph.noss@bankofengland.co.uk
(2) Bank of England. Email: lucas.pedace@bankofengland.co.uk
(3) University of Cambridge. Email: ot234@cam.ac.uk
(4) University of Cambridge. Email: obl20@cam.ac.uk
(5) Bank of England. Email: liam.crowley-reidy@bankofengland.co.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. We would like to thank Jon Relleen, Rhiannon Sowerbutts, Nick Vause and Jihyoung Yi for helpful comments, suggestions and advice. We are grateful to Thomson Reuters for providing data.

Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx

Publications and Design Team, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

A rapid price drop, or 'flash episode', occurred on 7 October 2016 in foreign exchange markets. The sterling-US dollar exchange rate (GBP/USD), the third most liquid currency pair in the world,[1] fell by 9.66%, from 1.2601 to 1.1491, within 40 seconds. Most of this movement was reversed within the ten minutes that followed.

The Bank for International Settlements (BIS) (2017) provided a report on the sterling flash episode. Rather than pointing to any single driver, it found the movement in the currency pair to have resulted from a confluence of factors. These included larger-than-normal trading (predominantly selling) volumes at a typically illiquid part of the trading day, as well as demand to sell sterling to hedge options positions and execute client orders in response to the initial fall in the exchange rate. The report also notes the potential amplifying role played by trading halts in GBP/USD futures contracts on the Chicago Mercantile Exchange (CME). This may have created larger price pressure on other trading platforms and increased the price impact of trades in the spot market.

The BIS (2017) report also provides a useful narrative of the events on 7 October, splitting them into three distinct phases:

- *Phase (1): Increasing order imbalance*

  At 00:07:00 British Summer Time (BST) 7 October 2016, a large number of trades were associated with a fall in the GBP/USD exchange rate. Sterling depreciated by 1.8% within 15 seconds. This period can be characterised by one-sided trading flow that was far larger than is customary for that time of the day. Changes in market prices during this period were, however, relatively orderly and continuous.

- *Phase (2): CME trading halt and a severe reduction in liquidity*

  The rapid drop in the exchange rate triggered a pause – or 'velocity logic event' – in trading of sterling FX futures on the CME at 00:07:15 BST. This meant that no trading could take place on the sterling futures exchange for the ten seconds that followed. At 00:07:29 BST a further 'downward price' limit was reached. This halted futures trading for two minutes and, during the two minutes that followed, no trades could take place at prices below this limit.

  During this period, liquidity in the cash spot FX market was highly impaired. And for short periods, resting orders to buy sterling on Thomson Reuters Matching – a key

[1] See BIS (2016).

interdealer platform – were completely exhausted. This meant that only a handful of low volume sell orders were matched during this period.

- *Phase (3): Recovery in price with larger than normal volatility and volumes*

  GBP/USD returned to a level of 1.24 at 00:17:00. But the remainder of the night and day that followed saw heightened volatility and trading volume. The volume traded over the entirety of 7 October was more than twice the average daily volume in the previous four days.

In this paper we deal predominantly with the first two phases of the flash episode. We provide a more in-depth description of the deterioration in liquidity that occurred during the episode than that in previous studies, using a number of estimates of market liquidity. These measure both the cost of trading and the impact on price of a transaction of a given size. The subsequent analysis also sheds further light on the causes of the flash episode in sterling, beyond that in previous studies, and the degree to which the movement in price was consistent with the estimated volume of orders to sell sterling.

Our contribution is broadly three fold:

First, we provide a detailed description of the price and order dynamics during the sterling flash episode, including the period when the market in the GBP/USD and EUR/GBP currency pairs experienced severe illiquidity with a near complete withdrawal of orders to buy sterling. In doing so, we utilise high frequency data on the price and quantity of every transaction and limit orders on the two currency pairs, taken from the Thomson Reuters Matching platform.

Second, we estimate a number of summary measures of liquidity during the flash episode. These include the cost of a round trip – that is, the cost of buying and immediately selling a given quantity of sterling – which is calculated using high frequency data from the limit order book of the Thomson Reuters platform. We also estimate two other summary measures of liquidity that are based on traded prices and volumes (but not limit orders). These include Effective Spreads (which measure the cost of trading); and a variant of the 'volatility over volume' (VoV) measure developed by Fong, Holden and Tobek (2017), which shows how price responds to a trade of a given volume (i.e. 'price impact'). According to all these measures, market liquidity deteriorated substantially during the flash episode.

The VoV measure is also found to give a better guide to liquidity – as given by a measure based on full limit order data – than does a simpler measure of price impact based on Amihud (2002). This is an interesting result, given that it is possible to calculate the VoV measure with less data, and at significantly lower computational cost, than measures of price impact based on limit order data.

Finally, we assess whether the change in price witnessed during the sterling flash episode was in line with the volume traded, given the empirical relationship between prices and volumes witnessed historically. Doing so requires us to develop a measure of price impact that is robust to the possibility that a large 'parent order' – or desire to trade by a single market participant – might result in a series of smaller orders. This is because participants in foreign exchange markets tend to split large transactions into a series of small orders in order to obtain a better price.

We find that the behaviour of the sterling exchange rate during the *initial* part of the flash episode is consistent with orders to sell sterling in a volume that vastly exceeded that which is usual at that time of day. The subsequent movement of the spot exchange rate is roughly consistent with our estimates of price impact given by the methodology of Kyle and Obizhaeva (2016a,b), which has been used previously to estimate price impact in equity markets. We also develop a new measure of price impact, based on that of Kyle (1985), and use this to cross-check the above result. This also provides some evidence that the estimates of price impact as a function of trade size found by Kyle and Obizhaeva (2016a,b) are also applicable to foreign exchange markets.

However, the *subsequent* fall in the value of sterling later in the episode – after the time of 00:07:15 BST – goes beyond that consistent with our estimates of the likely impact on prices given the quantity of orders to sell sterling. This may be in part due to the pause in trading at that time on CME – enforced by the triggering of its velocity logic mechanism. This may have led to the withdrawal of liquidity by market makers on other platforms, including Thomson Reuters Matching, because some market makers are thought to rely on the CME to provide a reference price for their liquidity provision in cash markets. This explanation is in line with the analysis of the flash episode in BIS (2017).

The paper is structured as follows. The next section describes the data provided by Thomson Reuters and describes in detail the dynamics of the limit order book during the flash episode. Section 3 develops and estimates a number of measures of liquidity during the episode. Section 4 examines the degree to which changes in price during the flash episode are consistent with the size and timing of observed transactions. Section 5 concludes.

2 Order book dynamics and liquidity around the sterling flash episode

The foreign exchange market trades 24 hours a day, five days a week. About one third of the market by volume is in spot contracts; the remainder is composed of derivative contracts such as forwards, futures, swaps and options.

Any attempt to examine dynamics in the sterling foreign exchange market is complicated by how trading is fragmented across trading venues. It is therefore hard both to evaluate what happened to the market as a whole, and to estimate the total volume of all trading activity during the flash crash.

This study, however, uses data from the Thomson Reuters Matching platform[2] for the GBP/USD and EUR/GBP currency pairs.[3] The data comprise of all trades (without counterparty information) in the spot market on this platform. It also shows updates to the limit order book – that is, the total quantity of orders to buy or sell at each price. This information is available at up to ten increments of price away from the best bid/ask prices. Data relate to the period between 19:00 BST on 2 October to 22:00 BST on 7 October.

The remainder of this section uses these data to describe the evolution of the limit order book around the flash episode. It also compares the evolution of liquidity during this episode to that over the prior four days.

2.1 Evolution of order book around the flash episode

Figures 1 and 2 give a visualisation of the order book around the event. Figure 2 is a magnified version of Figure 1 that covers the one minute of most severe illiquidity. The black line on the charts shows the mid-price; the coloured regions around this show the cumulative quantity of limit orders to buy/sell at prices between a given price and the mid-price. Coloured regions under the black line indicate limit orders to buy sterling, and those above the black line indicate limit orders to sell. For example, at 00:06:00, a trader wishing to sell £50 million could do so at a rate of 1.259 or lower. Our data

[2] Thomson Reuters Matching is a central limit order book that supports orders of type 'Good Until Cancelled' (which remain active until they are executed at the specified price or are cancelled) and 'Immediate Or Cancelled' (which are executed at a specified price immediately, or cancelled). Although orders are processed in continuous time, they are batched via a randomisation mechanism that queues similar orders for a short period and then randomly sends them for processing. Clients connected to Thomson Reuters via an Application Program Interface typically receive updated market data every 100ms.

[3] The Thomson Reuters Matching platform is thought to account for around 5-10% of trading in the sterling spot market, in normal times; see BIS (2016).

provide us only with ten levels of limit orders closest to the best bid and ask prices; we are unable to observe any limit orders beyond these.

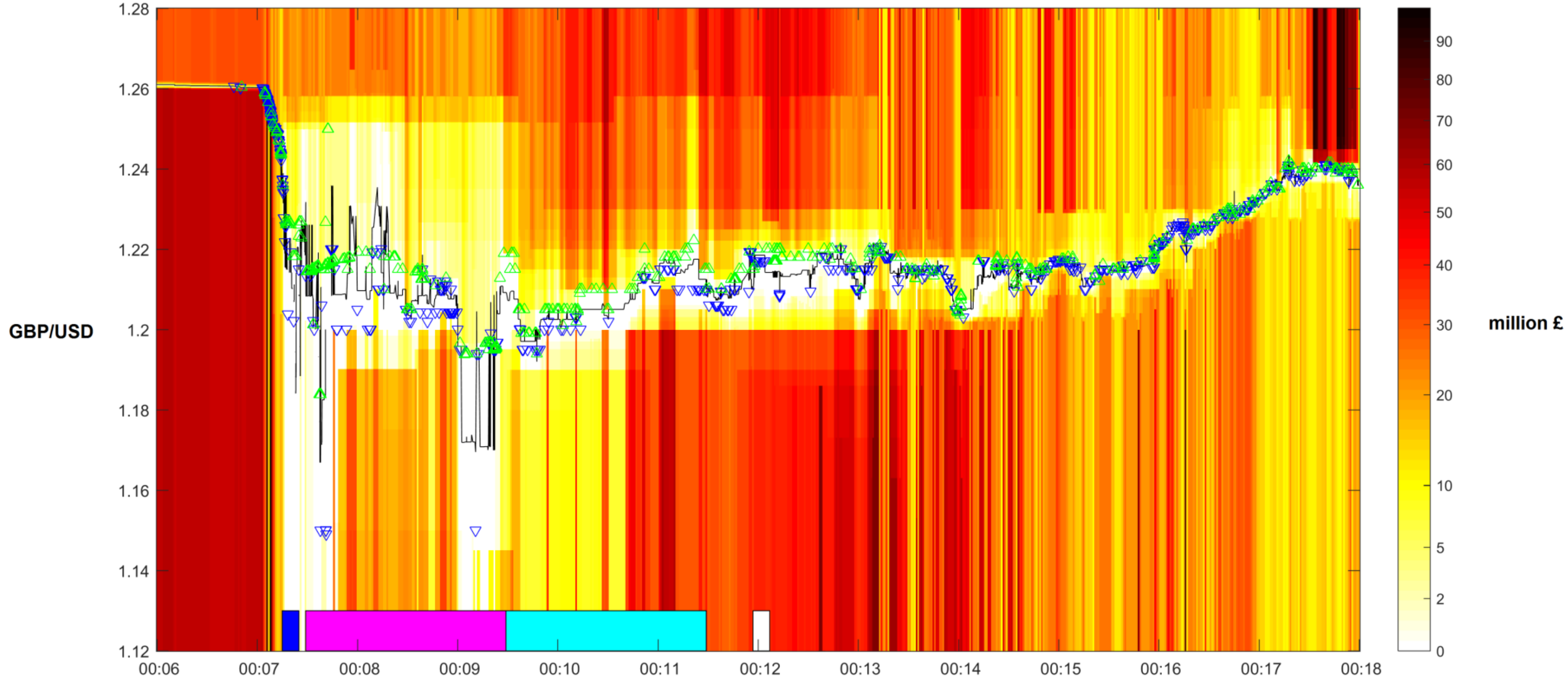
From these charts we observe:

- In the minute preceding the crash, between 00:06:00 and 00:07:00, the depth of limit orders was quite high just before the episode, with about £60 million of orders in the first ten levels of the book. The shape of the order book in the half hour immediately preceding the crash was similar to that between 00:06:00 and 00:07:00.
- At around 00:07:00 there develops an imbalance in the total quantity of limit orders to buy/sell sterling.
- This imbalance in limit orders became very severe around 00:07:17 BST. This can be seen from the large white areas in the graph, which indicate a total depth of limit orders to buy sterling of less than £2 million.
- The quantity of, and balance between, limit orders to buy/sell sterling recovered after about 30 seconds but deteriorated severely again after around one minute, shortly before 00:09 BST.
- The order book started to increase in depth around 00:09:30 BST, around 150 seconds after the initial sharp movement in price.

Throughout the episode there were several periods without any limit orders to buy, though these lasted less than about five seconds in total.

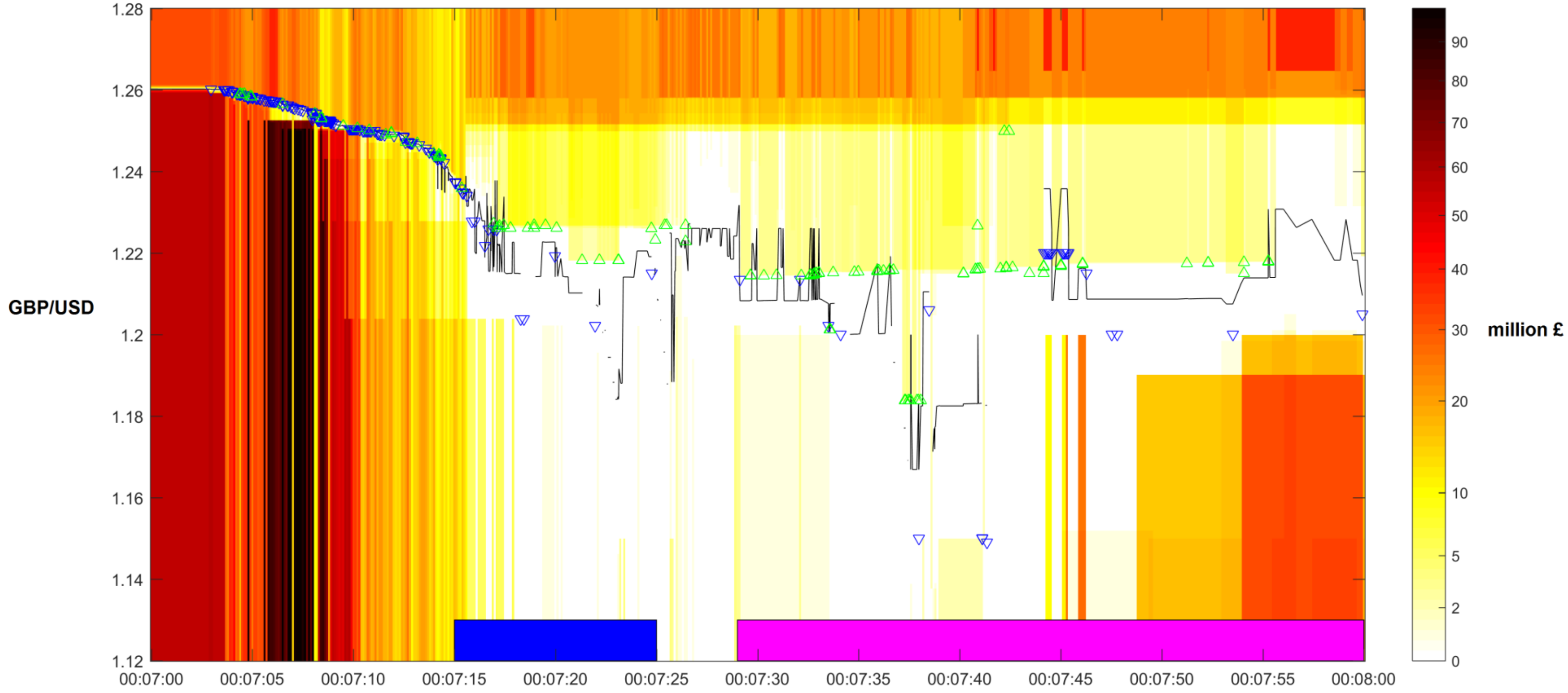
The remainder of the episode is characterized by rapid changes in market prices and a low quantity of limit orders to buy (of around £1 million to £2 million). Figure 11 in the Appendix provides the same graph for the EUR/GBP currency pair. Liquidity in this market was more adversely affected during the flash episode, with the order book depth on Thomson Reuters Matching completely exhausted on a number of occasions after trading on the CME was halted at 00:07:15. During this period, the cost of trading as given by the quoted spread – that is, the difference in best bid and ask prices, divided by their average – rose to as high as 20%. Trading volumes in other sterling currency pairs are comparatively low, so we focus exclusively on the GBP/USD and EUR/GBP pairs (see BIS (2016)).

Figure 1: The limit order book for the GBP/USD pair on the Thomson Reuters platform



- The black line shows midpoint between best bid and ask during the interval.
- Triangles show individual transactions: those in blue pointing down indicate transactions that involve the sale of sterling, and those in green pointing upwards indicate those to buy.
- Coloured regions in the graph show the cumulative quantity of limit orders to sell and buy between a given price and the best bid/ask price. Coloured regions under/above the black line indicate the quantity of limit orders to buy/sell sterling.
- The rectangles at the bottom of the chart show periods when trading was restricted on the CME.

Figure 2: The limit order book for the GBP/USD pair on the Thomson Reuters platform – the minute of most extreme illiquidity



- The black line shows midpoint between best bid and ask during the interval.
- Triangles show individual transactions: those in blue pointing down indicate transactions that involve the sale of sterling, and those in green pointing upwards indicate those to buy.
- Coloured regions in the graph show the cumulative quantity of limit orders to sell and buy between a given price and the best bid/ask price. Coloured regions under/above the black line indicate the quantity of limit orders to buy/sell sterling.
- The rectangles at the bottom of the chart show periods in which trading was restricted on the CME.

3 Measuring liquidity around the sterling flash episode

This section develops and examines a number of metrics of liquidity around the flash episode. It begins by directly measuring the cost of trading, as given by the 'effective spread' – that is a measure of the difference in price obtained on transactions, and those prevailing before them. It then goes on to estimate the price impact of trading – both via a static measure based on limit order data at any given point in time, and by the volatility over volume (VoV) measure, which is based on transactions over a given period.

3.1 The effective spread: a simple measure of the cost of trading

The effective spread is a simple measure of the cost of trading. It is defined as the volume weighted average of the difference in prices at which transactions take place, versus those prices prevailing one millisecond earlier; that is:

$$Effective\ Spread_k = \sum_{k \in t} Volume_k 2D_k \left(ln(P_k) - ln(M_k)\right) / \sum_{k \in t} Volume_k \quad (1)$$

where P_k is price of trade k, D_k is an indicator equal to 1 for buy and -1 for sell orders, and M_k is the prevailing midpoint price one millisecond before the transaction.

Figure 3: Effective Spread in GBP/USD pair

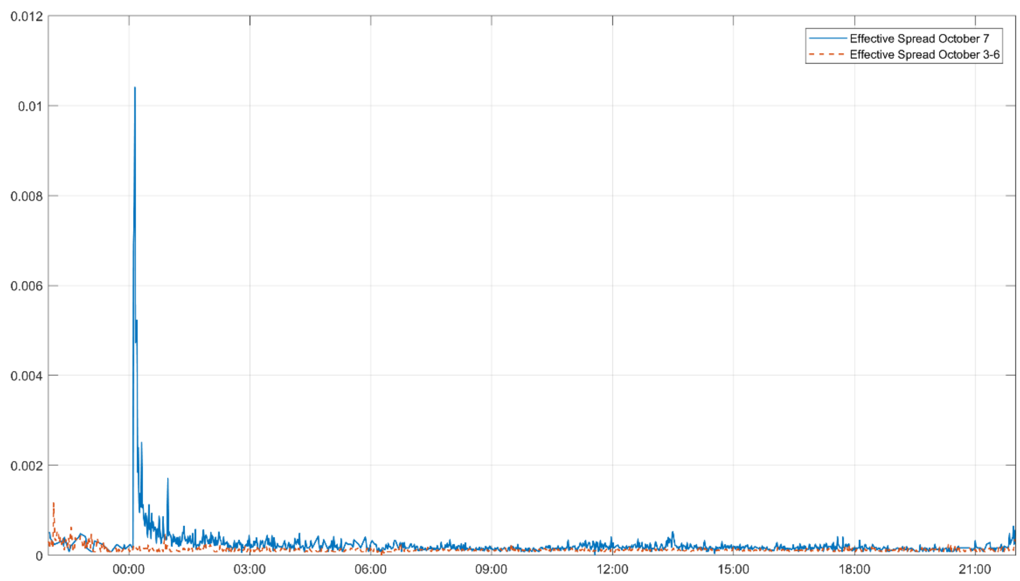


Figure 3 shows the average value of the effective spread four days before and on the day of the crash. It is evident that transactions costs were heavily elevated during the flash episode.

3.2 Measuring price impact in foreign exchange markets

Another important measure of market liquidity is price impact – that is, the degree to which an order of a given size results in a change in the prevailing price. Previous studies on foreign exchange markets – including BIS (2017) – have applied the methodology in Amihud (2002) to measure price impact around market events. This estimates price impact by comparing the absolute value of the percentage change in price that coincides with a trade, and that trade's volume.

We examine two alternative measures of price impact. The first – the 'cost of a round trip' (CRT) – is based on the totality of available limit order data. The second is 'Volatility of Volume' (VoV) – a summary measure based only on the price and volume of transactions. This is similar in intuition to that of Amihud (2002), but provides a closer approximation to estimates of price impact based on limit order data (see Section 3.3).

These measures are examined in turn below.

3.2.1 Cost of a round trip – an estimate of price impact based on limit order data

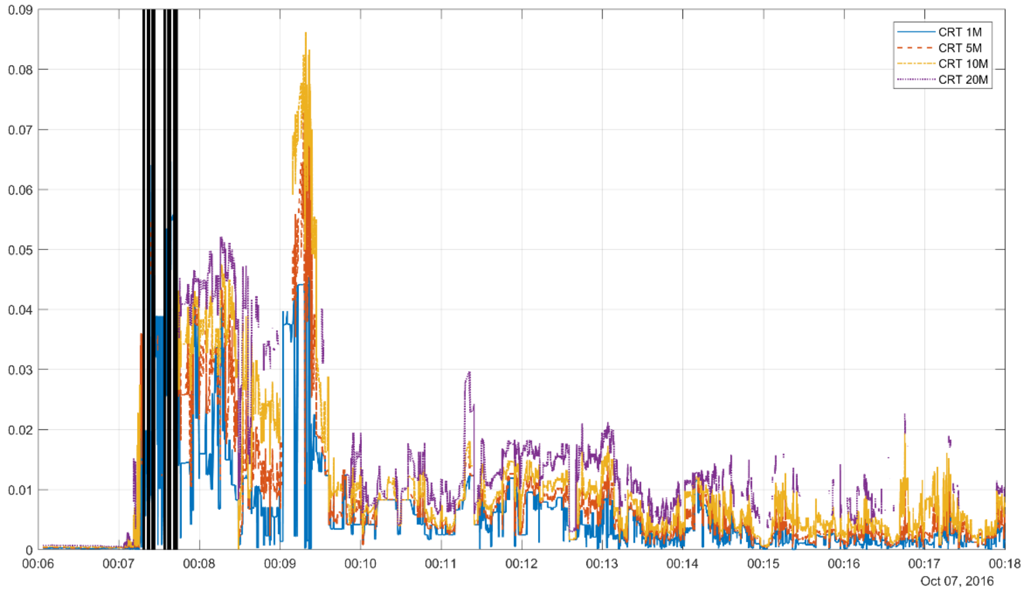
The cost of a round trip is defined as the cost of buying and immediately selling £*x* million of the currency divided by the midpoint price between the best bid and ask orders at that time. This we denote by the function CRT(x).This gives a direct indication of the degree to which the execution of orders to buy – and immediately sell – in a given quantity would impact market prices based on the available limit orders at a given point in time.

Figure 4 presents the cost of round trips of size £1, £5, £10, and £20 million for the GBP/USD currency pair. It is evident these measures were elevated during the episode. In addition, CRT(20M) – that is, the cost of a round trip of size £20 million – is undefined for much of this period, which indicates that transactions of this size were impossible to execute, given the available depth of limit orders. The black vertical lines on the graph indicate times where there were no quotes; i.e. it was impossible to transact a trade even of size £1 million.

This measure also allows for an estimate of the *marginal cost per additional unit of volume transacted*. This we define as the average of (CRT(20) – CRT(1))/19 and (CRT(10) – CRT(1))/9 – henceforth referred to as Price Impact 2-20M and Price Impact

2-10M, respectively (see Section (3.3). These measures exclude order book depth at the first £1 million from the best bid/ask (i.e. CRT(1)) because this is typically very large in size. Its inclusion would therefore cause estimates of transacting in large size to be heavily correlated with the effective spread, which – in both this and related studies – is analysed separately; see Karnaukh, Ranaldo, and Söderlind (2015).

Figure 4: Cost of round trip (CRT) in the GBP/USD currency pair



3.2.2 Volatility over volume – a measure of price impact based on transaction data

The second measure of price impact is the 'volatility over volume', or VoV, measure developed by Fong, Holden, and Tobek (2017). Unlike the cost of a round trip, this is based only on transaction data – specifically the ratio between the volatility of the price at which, and the volume in which, transactions take place. This measure of price impact increases with the level of market illiquidity. This is because, in a less liquid market, a transaction (or series of transactions) of a given size will have a larger impact on price, and be associated with greater volatility.

Volatility over volume is defined over a time interval *t* as the square root of the ratio of volatility of price to volume transacted, that is:

$$\mathrm{VoV}(\lambda)_t = \sqrt{\frac{MedRV_t}{volume_t}}, \tag{2}$$

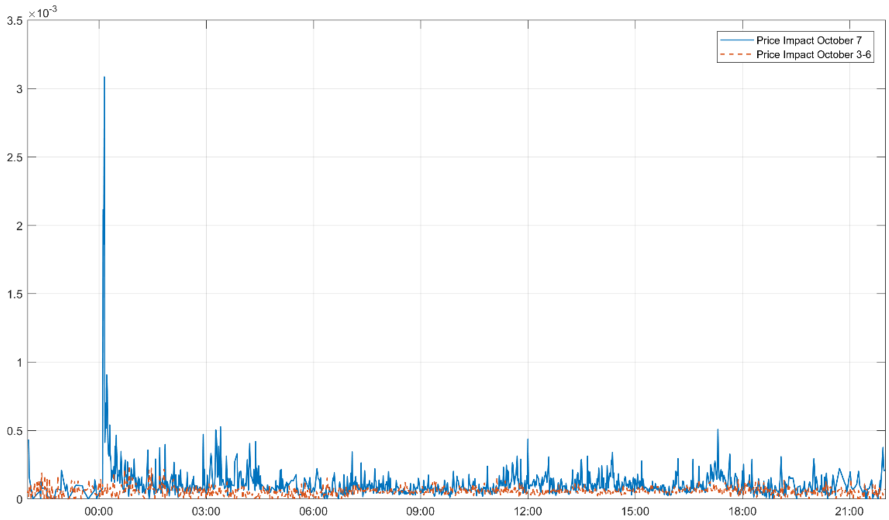
where MedRV is an estimator of realised variance (rescaled to a daily value) and *volume* is the traded volume, both over the time interval *t*.

MedRV is estimated using intraday data using the realised measure of Andersen, Dobrev, and Schaumburg (2012), which is robust to jumps. This is applied to the median of returns in three consecutive intervals, so that:

$$MedRV_t = \frac{\pi}{6 - 4\sqrt{3} + \pi}\left(\frac{N}{N-2}\right)\sum_{i=2}^{N-1} med(|r_{i-1}|, |r_i|, |r_{i+1}|)^2. \quad (3)$$

To enhance efficiency we apply the subsampling procedure discussed in Zhang, Mykland, and Aït-Sahalia (2005). To ensure robustness we also estimate MedRV as an average of its values computed with returns across one, five and fifteen second intervals.[4]

Figure 5: Price impact - VoV(λ) in GBP/USD pair



Note that this suggested measure is based on 'event time', whereby new information of relevance to the value of a security (or in this case, the exchange rate) is assumed to evolve as a function of the number of new transactions, rather than in proportion to the passage of time (see Kyle and Obizhaeva (2016) and Easley, de Prado, and O'Hara (2012)).

[4] We have tried applying filters to the data as advocated by Andersen, Dobrev, and Schaumburg (2012) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). Specifically, we tried deleting entries for which the mid-quote deviated by more than 10 mean absolute deviations from a rolling centred median (excluding the observation under consideration) of 50 observations (25 observations before and 25 after) but the procedure did not filter out any observations.

Figure 5 shows the VoV(λ) measure over the whole day and compares it to its mean value over the previous four days. It gives a very similar picture to other liquidity measures and further documents that markets remained relatively illiquid for the rest of the day after the crash.

Summary statistics for all the above measures are given in **Table 1**. Volatility is rescaled to have the same level as daily volatility (bottom row of Table 1).

Table 1: Summary Statistics for Measures in every 15 minute (basis points)

	GBP/USD			EUR/GBP		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
Effective Spread (bp)	1.610	1.237	2.821	1.963	1.150	4.574
Bid-Ask Spread (bp)	2.400	1.991	2.193	2.878	2.006	3.329
CRT10 million (bp)	5.225	3.824	4.829	6.488	4.044	8.664
CRT20 million (bp)	7.231	5.700	5.957	10.130	5.429	16.97
VoV(λ) (bp)	15.27	12.00	28.19	28.65	19.89	57.38
Amihud (bp)	0.888	0.675	0.903	1.116	0.778	1.327
Volatility (daily bp)	78.1	45.7	343.9	73.8	45.2	272.0

3.3 Benchmarking the Volatility over Volume (VoV) measure

This subsection assesses the performance of the VoV measure (based on transaction data) and compares it to a simpler measure of price impact given by Amihud (2002), which we define as the ratio of absolute returns over transaction volumes. Both measures are benchmarked with respect to Price Impact 2-10M and Price Impact 2-20M, as defined in Section 3.2.1: that is, the average cost of buying or selling an additional £ million of currency at any given point in time, based on limit order book data.

A similar exercise is performed by Fong, Holden, and Tobek (2017), who find the VoV measure to perform better than any other measure of price impact in equity markets, including that of Amihud (2002). This is, however – to the best of our knowledge – the first time such an exercise has been undertaken using FX market data.

Table 2 presents results from this benchmarking exercise. These are presented in the form of time series correlations between changes in benchmark measures of price impact and changes in the Amihud or VoV(λ) liquidity measures.[5] Bold numbers are significantly different from zero at a 5% confidence level. The VoV(λ) measure outperforms that based on Amihud (2002). The difference in performance is higher for EUR/GBP currency pair, and both liquidity measures outperform the benchmark in the case of the GBP/USD pair.

That the VoV measure gives a reasonable guide to liquidity is an interesting result. This is because this proxy for liquidity is based only on transactions – rather than limit order data – which makes it less computationally demanding to produce, and possible to calculate in circumstances where full limit order data are unavailable.

Table 2: Correlations between Amihud and VoV measures of price impact and benchmark measures (2-10M and 2-20M) based on limit order book data

	Price Impact 2-10M		Price Impact 2-20M	
	Amihud	VoV(λ)	Amihud	VoV(λ)
GBP/USD	**0.473**	**0.607**	**0.401**	**0.516**
EUR/GBP	0.040	**0.255**	0.044	**0.260**

4 Was the change in price during the flash episode consistent with order flow

This section examines the degree to which the change in market prices during the flash episode was consistent with the degree of imbalance between orders to buy and sell sterling just prior to the episode. There is good reason to expect that a significant imbalance in order flow should, all else being equal, increase the change in price that results from a trade (or trades) of a given size. One risk faced by those making markets is that of transacting with more informed traders. Market makers therefore charge a spread that compensates them for bearing the risk of transacting at prices that are 'stale' with respect to some new information (Copeland and Galai (1983)).[6] An imbalance in orders might be interpreted by market makers as a signal that another market participant is party to superior, or more up-to-date, information. This might incentivise market makers to adjust their prices, and widen their spreads, thereby increasing the cost of trading for other market participants (see O'Hara, Easley, and de Prado (2012)).

[5] This follows the methodology given in Karnaukh, Ranaldo, and Söderlind (2015).
[6] See O'Hara (1995) for a survey of relevant literature. Inventory explanations of market making behaviour also suggest that a build-up of inventory on one side can lead to a reluctance to quote on that side of the market; see Ait-Sahalia and Saglam (2017).

A natural way to estimate the effect of large trades on asset prices is through an estimate of the 'λ measure' suggested by Kyle (1985). This is defined as the coefficient on (the square root of) transaction volumes in a regression on returns for a given security.

But such traditional estimates of the responsiveness of prices to an order (or orders) of a given size – or 'price impact' – are not easily applicable to foreign exchange markets. This is firstly because, although foreign exchange markets typically have extremely small bid-ask spreads, these tend to give an indication only of the cost of trading in small quantity. Market participants thus typically break up larger trades into many smaller trades and execute them sequentially. Any estimate of price impact based on the average of that for individual (small) trades may therefore give an unreliable estimate of the price impact of a larger trade, whose impact is of interest here.

Previous attempts to circumvent this problem have sought to estimate the response of prices to trades as aggregated over a given interval. Doing so aims to reduce the bias in estimates of price impact that occurs when large orders are broken down into small trades. Goyenko, Holden and Trzcinka (2009) and Hasbrouck (2009), for example, estimate price impact as the coefficient on the size of trading volumes over a five minutes window, in a regression on the return over the same period.[7] This approach struggles, however, to capture the effect both (i) of *individual* large orders occurring *within* a given five minute interval, and (ii) of large orders executed via series of smaller trades over a horizon of *greater* than five minutes.

A second, more general, problem is that FX markets now contain a large degree of high-frequency algorithmic trading. This means that the reaction of prices to transactions initiated by algorithms is better understood by considering periods not of equal length in calendar time (e.g. five minute intervals), but of equal quantities of traded volume (O'Hara, Easley, and de Prado (2012)). Considering market activity in 'volume time' – rather than calendar time – also has the advantage of allowing for the estimation of a price impact function that applies regardless of the time of day, even though traded volumes tend to differ substantially during the day and night.

[7] Specifically, they estimate the price impact function as a regression coefficient λ in an equation $\mathrm{r_i} = \lambda S_{\mathrm{i}} + \mathrm{u_i}$ where $\mathrm{r_i}$ is the return, $\mathrm{S_i} = \sum_{\mathrm{ik}} \mathrm{sign}(\mathrm{vol_{ik}}) \sqrt{\mathrm{vol_{ik}}}$ is the signed square root of trading volume in the *i*th five-minute interval, and $\mathrm{u_i}$ is an error term.

4.1 Measuring the change in price consistent with observed transactions

The two subsections that follow propose measures of price impact that are robust to each of these concerns. These allow us to assess whether the change in price witnessed during the sterling flash episode was in line with the volume traded, given the empirical relationship between prices and volumes witnessed prior to the episode. The first subsection applies a measure proposed in Kyle and Obizhaeva (2016a) based on the principle of market microstructure invariance; the second subsection goes on to propose an extension of the λ measure of Kyle (1985) in volume time.

4.1.1 A measure of price impact based on the principle of market microstructure invariance Kyle and Obizhaeva (2016a)

Market microstructure invariance is the empirical hypothesis that the dynamics of prices – including the price impact of a trade of a given size – is invariant across different financial markets if examined at a scale that accounts for the differing rate at which risk is transferred. This is because the price impact of trades varies across markets due to differences in the rate at which risk changes hands. For actively traded assets, this transfer of risk – or 'bets' – arrives very rapidly; for less actively traded assets, this occurs more slowly. However, when price impact is measured against a measure of time that controls for the rate at which trading takes place, it is invariant across markets.

Drawing on this insight, Kyle and Obizhaeva (2016a) estimate the price impact of trades of a given size from a large asset manager in equity markets, as a function of market volume and volatility. They find this to take the functional form:

$$1 - \exp\left[-\frac{5}{10^4}\left(\frac{0.81 \cdot V}{1.26 \cdot 40}\right)^{\frac{1}{3}}\left(\frac{\sigma}{0.02}\right)^{\frac{4}{3}}\frac{X}{V}\right] \tag{5}$$

where V is the average daily volume, X is the size of the ('parent') order, and σ is expected volatility of market prices.[8]

Importantly, this function for price impact is found to hold even when such a parent order to buy or sell is decomposed into a number of smaller transactions with the same underlying motivation.[9] It is this that makes it a promising method with which to

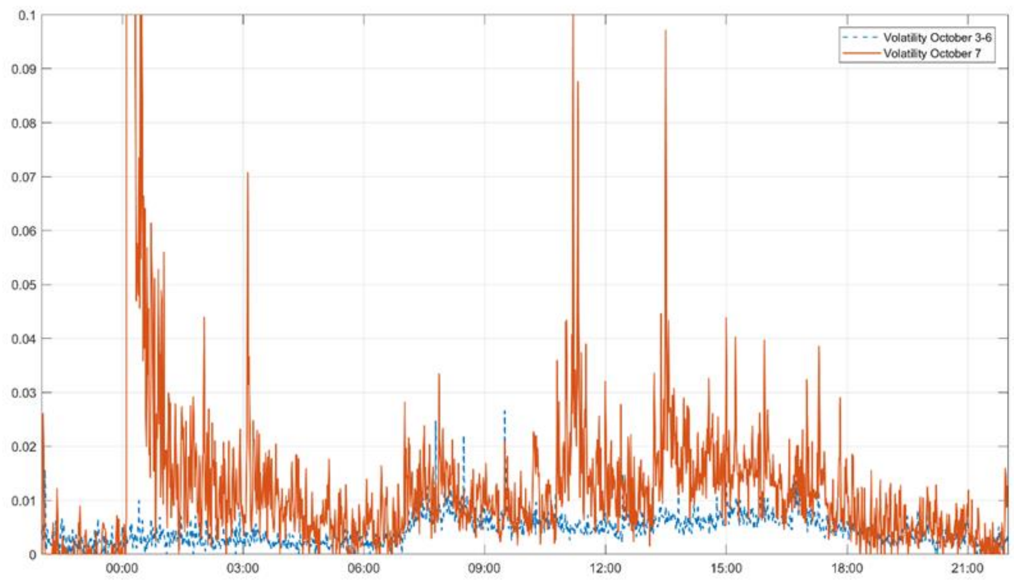
[8] These parameters are estimated using data from equity markets during the period 2002-2005.
[9] In fact, such transactions need not be placed by a single participant but can be composed of multiple orders that act on the same information; for example when a price decline forces a liquidation of multiple positions taken on margin in a given security.

estimate the price impact that might be expected to result from a large order – or orders – to sell, particularly when this may have been split up into smaller orders that were executed sequentially.

Applying this approach requires us to calculate the level of volatility and volume expected by market participants at the time of the trades. We do this in a manner similar to that in Kyle and Obizhaeva (2016b), as follows:

i. We estimate average <u>daily volatility</u> from FRED exchange rate data from 1 January to 6 October.[10] This is found to take the value 0.0094. It is slightly higher – at 0.0122 – in the period following the UK referendum on EU membership and lower – at 0.0054 – in the month preceding the sterling flash episode.

Figure 6: Volatility on the GBP/USD currency pair



ii. It is possible that market expectations of volatility ahead of the flash episode were elevated just prior to the flash episode given that the GBP/USD rate had fallen for four days in row. Realised volatility (rescaled to a daily value) in the four days preceding the crash was 0.36% for 00:00:00 to 01:00:00 period and 0.32% for 22:00:00 to 04:00:00 period (**Figure 6**). **We therefore assume that the expected daily volatility was somewhere between 0.0035 and 0.005**.[11] Daily <u>transaction volume</u> for the GBP/USD and EUR/GBP currency pairs average £370 billion and

[10] FRED provides data series for noon buying rates in New York City for cable transfers payable in foreign currencies. https://fred.stlouisfed.org/series/DEXUSUK

[11] The volatility presented there is estimated by the jump-robust estimator of Andersen et al. (2012). The volatility on the EUR/GBP pair has followed closely the volatility on the GBP/USD pair.

£80 billion, respectively, according to BIS (2016). Given around one third of trading activity occurs in the spot market, **we assume daily spot market volume across both currency pairs and across all platforms to average around £150 billion**.

The average daily trading volume we observe on Thomson Reuters Matching is around £10bn in total across the two currency pairs.

iii. Trading volume does, however, differ substantially over the course of the day. This is important, given that the flash episode occurred outside the main trading hours. Between 3 and 6 October the average GBP/USD trading volume on the Thomson Reuters platform between 21:00:00 and 06:00:00 was around ten times smaller compared to an average hourly volume during day time trading hours. **We therefore make the conservative assumption that the expected hourly volume during the night is 10 times smaller than during normal trading hours.**

iv. The volume traded on Thomson Reuters between 00:07:00 and 00:07:15 BST on 7 October was £301 million on GBP/USD pair and €62 million on EUR/GBP pair. Of these transactions, £273 million and €57 million were market orders to sell sterling, respectively. There was a further £17 million of sales of sterling in the two seconds that followed. **This leads us to estimate that there was, in aggregate about £290 million of parent orders to sell sterling via the GBP/USD currency pair alone, and £340 of parent orders to sell across the two currency pairs, on the Thomson Reuters platform.**

In what follows, we consider the two currency pairs together, given that a change in the price of one pair would likely have swiftly been reflected in the other, in order to remove the possibility of arbitrage.

v. **Together, these quantities lead to an estimate of X/V in equation (5)**; that is, the size of the parent order as a fraction of the daily volume. This we set to 0.035 in the case of orders occurring during the day (£340m total of one sided bets ((iv) above) divided by £10bn of total volume ((ii) above) across the two currency pairs.[12] We assume a corresponding value of X/V of 0.35 in the case of orders placed during the night: ten times (see (iii) above) that corresponding to the daytime.

[12] We make an assumption that orders observed on the Thomson Reuters matching platform were observed on other platforms as well. The BIS (2017) report shows that this was true for CME just before it shut down.

We further check how our results are altered if we change X/V to 0.2 and 0.02 for the night and day period, respectively; these more conservative values are selected arbitrarily.

Table 3 uses these estimates to presents an estimate of the change in the GBP/USD currency pair that could be attributed to a £290 million parent trade to sell sterling via the GBP/USD currency pair, or a £340 million trade on sterling as a whole, realised only on the Thomson Reuters Matching Platform.

Table 3: Predicted Movements in Price

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Panel A: Predicted price movement during night								
X/V	0.35	0.35	0.35	0.2	0.35	0.35	0.35	0.2
Expected volatility	0.35%	0.35%	0.35%	0.35%	0.5%	0.5%	0.5%	0.5%
Volume – Night (£bn)	0.81	37	45	45	0.81	37	45	45
Expected Return	0.50%	1.66%	1.79%	1.03%	0.80%	2.65%	2.87%	1.65%
Panel B: Predicted movement during the day								
X/V	0.035	0.035	0.035	0.02	0.035	0.035	0.035	0.02
Expected volatility	0.35%	0.35%	0.35%	0.35%	0.5%	0.5%	0.5%	0.5%
Volume – Day (£bn)	8.1	370	450	450	8.1	370	450	450
Expected Return	0.11%	0.36%	0.39%	0.22%	0.17%	0.58%	0.62%	0.36%

Panel A shows the expected change in price that would result from a parent order of such a size during the night. Columns (a) to (h) show this change in price based on different estimates of the size of the parent order, expected volatility and market volume, based on the figures above:

- Column (a) and (e) give the change in price assuming all the foreign exchange market activity occurred on the Thomson Reuters platform, for different values of expected volatility. Transaction volume is set equal to £0.81bn: that is the £8.14bn estimated average daily volume on Thomson Reuters Matching on the GBP/USD pair (see (ii) above), divided by a scale factor of 10 (see (iii) above).

- The estimated change in price – of 0.5%/0.8% – is smallest in this case – and smaller than that observed in reality. This is because the absolute size of the parent order is smaller when scaled to activity on the Thomson Reuters platform alone. Price changes in the other columns assume transactions to occur across the entirety of the FX market, including for transactions other than the spot market, and beyond those on the Thomson Reuters platform. Here transaction volume is set equal to £37bn (that is, £370bn estimated daily transaction volume in GBP/USD divided by ten), and £45bn in the case of both currency pairs (see (i) above).

  These assumptions correspond to an expected price movement between 1.03% and 2.87% depending on the choice of volatility and volume parameters.

  This is consistent with the observed price decline of 1.80% from 00:07:00 to 00:07:15, and suggests that this movement in price was consistent with the arrival of a large parent order to sell sterling, of the sort described above.

- But the change in price that occurred between 00:07:00 to 00:07:17 – when the CME futures exchange was closed but the cash market had not yet experienced the severest illiquidity – was even larger, at 2.73%. It follows that the overall drop in the exchange rate that followed cannot be explained by expected price impact of trades alone.[13]

Panel B provides corresponding figures for predicted movement in price based on the size of the observed trades during the daytime. This suggests that a parent order of the size arriving during the sterling flash episode would have had a smaller impact had it occurred during normal trading hours; the largest change in price consistent with the combinations of parameters above is 0.62%.

One drawback of these estimates is that they are predicated on the relationship between changes in price, and parent order size/volume, based on dynamics observed by Kyle and Obizhaeva (2016a) in equity markets. It is far from certain that this relationship also holds in foreign exchange markets. In the next subsection we therefore propose a new measure of price impact that addresses this concern.

[13] The assumption of rescaling the observed volume on Thomson Reuters Matching to the whole market likely provides an inflated estimate of the overall size of all the orders. The correct estimate is thus likely to lie somewhere in between the values for Thomson Reuters Matching alone and those for the whole market.

4.1.2 A measure of price impact based on volumes

This subsection proposes an alternative new means of estimating the degree to which the move in sterling was consistent with the impact of a large parent order. Unlike the measure in the previous subsection, it aims to estimate the price impact of selling a given volume of sterling, irrespective of the time at which this takes place. This helps circumvent the problem encountered in the previous section as to how the volume traded over a given time interval varies during the day and night.

The measure is inspired by the 'volume synchronised probability of informed trading' (VPIN) measure of order flow toxicity used in O'Hara et al. (2012). While its use in that paper is to predict the occurrence of a market crash, our aim here is to estimate the expected movement in price as a response to an order flow imbalance.

Our calculation proceeds as follows:

- Following O'Hara et al. (2012), we divide all trades within one day into consecutive buckets over time, each containing an equal volume of orders. We do so for four different sizes of buckets, £100 million, £200 million, £500 million, and £1 billion.
- Within each bucket we compute the order flow imbalance. This is defined as the quantity of observed transactions to buy minus those to sell, divided by the overall volume traded.
- We estimate a nonlinear relationship between returns and order flow imbalance using a smoothing spline method. This is shown in **Figure 7**.[14] The change in price increases with the size of volume buckets. This is because a given size of order imbalance is larger in nominal terms the larger the volume bucket. For example, an imbalance of 1 corresponds to £100 million of buys in the 100 million bucket but to £1 billion of buys in the 1 billion bucket.
  - For a given size of bucket, we then estimate an OLS regression of order imbalance on price returns:

$$r_i = \alpha + \bar{\lambda} I_i + u_i \tag{7}$$

  where r_i is returns, and I_i the order imbalance across orders in buckets of a given size.

[14] The smoothing spline s(*I*) minimizes the objective function:

$$0.95 \sum_i w_i \, (r_i - s(I_i))^2 + 0.05 \int \left(\frac{d^2 s}{dI^2}\right)^2 dx. \tag{6}$$

- Regression results are given in **Table 4** (t-statistics on the estimated coefficients are reported in parentheses).[15] Estimated coefficients suggest an order of size £1 billion (roughly corresponding to 1/8th of the daily volume) is associated with a 1.68% change in price. The R^2 in the regressions are rather low but this is unsurprising given that measuring bets with order flow imbalance is far noisier than that based on direct data on portfolio transitions.

Figure 7: Estimated Price Impact of Imbalance in GBP/USD pair

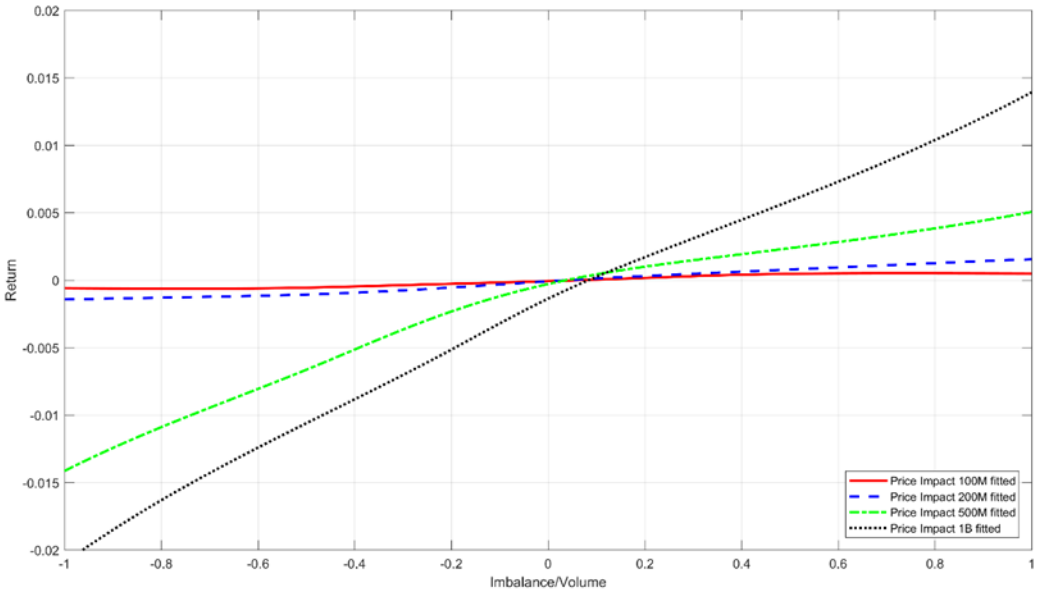


Table 4: OLS estimates of price impact given by the regression: $r_i = \alpha + \bar{\lambda} I_i + u_i$

	Bucket size (£)			
	100 million	200 million	500 million	1 billion
Intercept	-0.0000	-0.0002	-0.0004	-0.0008
	(-1.29)	(-1.22)	(-0.99)	(-0.85)
Imbalance	0.0012	0.0021	0.0083	0.0168
	(5.37)	(3.65)	(4.09)	(2.44)
Adjusted R^2	0.0499	0.0445	0.129	0.0852
Observations	531	266	107	54

[15] The above calculations includes all trades between 2 to 7 October. We exclude the period around the flash episode (00:00:00 to 01:00:00 on 7 October) from the estimation, as we seek to estimate the degree to which the price impact of trades during the episode was in line with that estimated based on trades outside of it. Including the flash crash in our sample strengthens the resulting estimate of price impact.

This result is roughly in line with the results of the previous subsection. This provides some evidence that the functional form of price impact given by Kyle and Obizhaeva (2016) for equity markets (see equation (5)) may also be valid for FX markets.

4.2 Consistency between change in price and order flow

The results suggest that the move in sterling during the first stage of the flash episode – up to 00:07:15 – was in line with the change in price that would have been expected of a large parent order to sell sterling. The estimates given above suggest that the observed traded quantities would normally, at this time of day, be associated with a change in price of between 0.5% and 2.87% – which encompasses the observed fall in sterling up to that point of 1.8%.

However, the overall drop in sterling of 9.66% during the entirety of the flash episode – *including the period following 00:07:15* – exceeds that consistent with these measure of price impact. To the extent that the measures and methodologies described here provide an accurate portrayal of the price movement that would result from orders during a period of the trading day where little trading takes place, this suggests a role for other factors in justifying the totality of the change in price observed during the episode. These could include the trading halts on the CME futures exchange which, according to BIS (2017), may have led to the withdrawal of liquidity by market makers on other trading venues.

5 Conclusion

This paper has provided an in-depth analysis of liquidity during the sterling flash episode of 7 October 2016. It documents the deterioration in illiquidity in the GBP/USD and EUR/GBP exchange rates against a number of metrics, and shows there was a short period in which there were no outstanding limit orders on the bid side of the order book on the Thomson Reuters Matching platform. It also shows that a variant of the 'volatility over volume' measure of liquidity, based on transaction data, provides a better proxy of illiquidity – when benchmarked against measures based on high-frequency limit order data – than other summary measures of liquidity.

It also analyses the extent to which the depreciation in sterling can be attributed to a large trade – or multiple consecutive smaller trades – to sell sterling, at a time of day when trading activity is normally thin. Findings suggest that the initial fall in sterling before 00:07:15 is consistent with the arrival of a large order to sell sterling. However, the entirety of the depreciation in sterling – including that later in the episode – goes beyond that consistent with estimates of the price impact of observed orders to sell sterling. This might support the suggestion in BIS (2017) that the move in sterling may have been amplified by the pause in trading on the CME futures exchange, which led to a withdrawal of liquidity on other platforms.

References

Ait-Sahalia, Yacine, Sağlam and Mehmet (2017), *High Frequency Market Making: Optimal Quoting*.

Amihud, Y (2002), *Illiquidity and stock returns: cross-section and time-series effects*, Journal of financial markets, Vol. *5*(1), p. 31-56.

Andersen, T. G., Dobrev, D and Schaumburg, E (2012), J*ump-robust volatility estimation using nearest neighbor truncation*, Journal of Econometrics, Vol*.*169*(*1), p. 75-93.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A and Shephard, N (2009), *Realised kernels in practice: trades and quotes*, The Econometrics Journal, Vol. 12(3).

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

BIS (2016), *Triennial Central Bank Survey of foreign exchange and OTC derivatives markets in 2016*.

Chordia, T., Roll, R and Subrahmanyam, A (2011), *Recent trends in trading activity and market quality. Journal of Financial Economics*, Vol 101(2), p. 243-263.

Copeland, T. E and Galai, D (1983), *Information effects on the bid-ask spread*, Journal of Finance, Vol. 38(5), p. 1457-1469.

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

Fong , Kingsley Y. L., Holden, Craig W and Tobek, O (2017), *Are Volatility Over Volume Liquidity Proxies Useful For Global Or US Research?*

Glosten, L. R., Jagannathan, R and Runkle, D. E (1993). *On the relation between the expected value and the volatility of the nominal excess return on stocks*, Journal of Finance, Vol. 48(5), p. 1779-1801.

Goyenko, R. Y., Holden, C. W and Trzcinka, C. A (2009), *Do liquidity measures measure liquidity?* Journal of Financial Economics, Vol. 92(2), p. 153-181.

Hasbrouck, J (2009), *Trading costs and returns for US equities: Estimating effective costs from daily data*, Journal of Finance, Vol. 64(3), p. 1445-1477.

Karnaukh, N., Ranaldo, A and Söderlind, P (2015), *Understanding FX liquidity, The Review of Financial Studies*, Vol. 28(11), p. 3073-3108.

Kyle, A. S (1985), *Continuous auctions and insider trading*, Econometrica: Journal of the Econometric Society, p. 1315-1335.

Kyle, A. S. and Obizhaeva, A. A (2016a), *Market microstructure invariance: Empirical hypotheses*, Econometrica, Vol. 84(4), p. 1345-1404.

Kyle, A. S. and Obizhaeva, A. A (2016b), *Large bets and stock market crashes*.

O'hara, M (1995), *Market microstructure theory*, Vol. 108, Cambridge.

Zhang, L., Mykland, P. A and Aït-Sahalia, Y (2005), *A tale of two time scales: Determining integrated volatility with noisy high-frequency data*, Journal of the American Statistical Association, Vol. 100(472), p. 1394-1411.

Appendix

Figure 8: Quoted Spread in Individual Minutes in GBP/USD pair

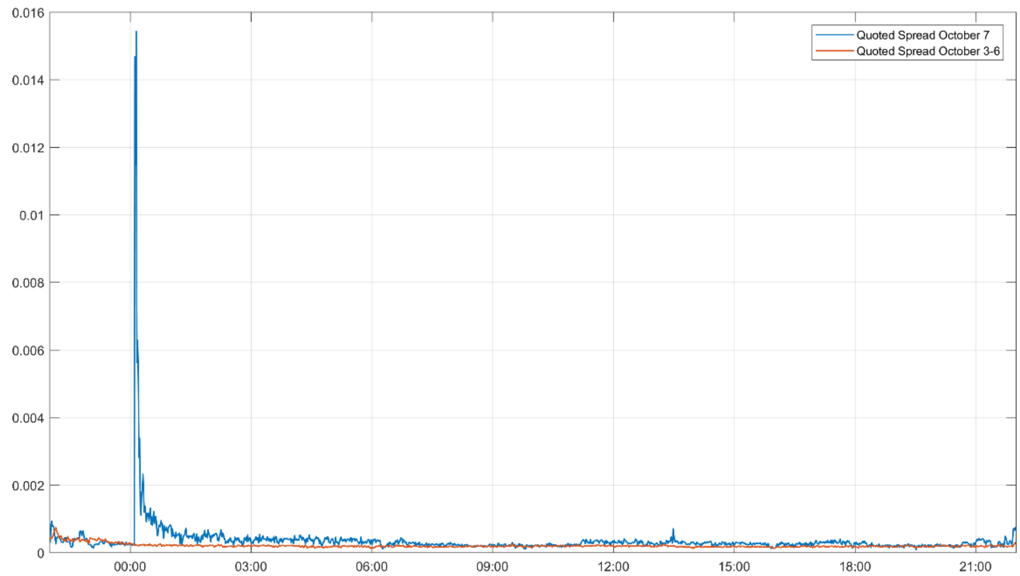


Figure 9: Cost of Round Trip for £10 million in Individual Minutes in GBP/USD pair

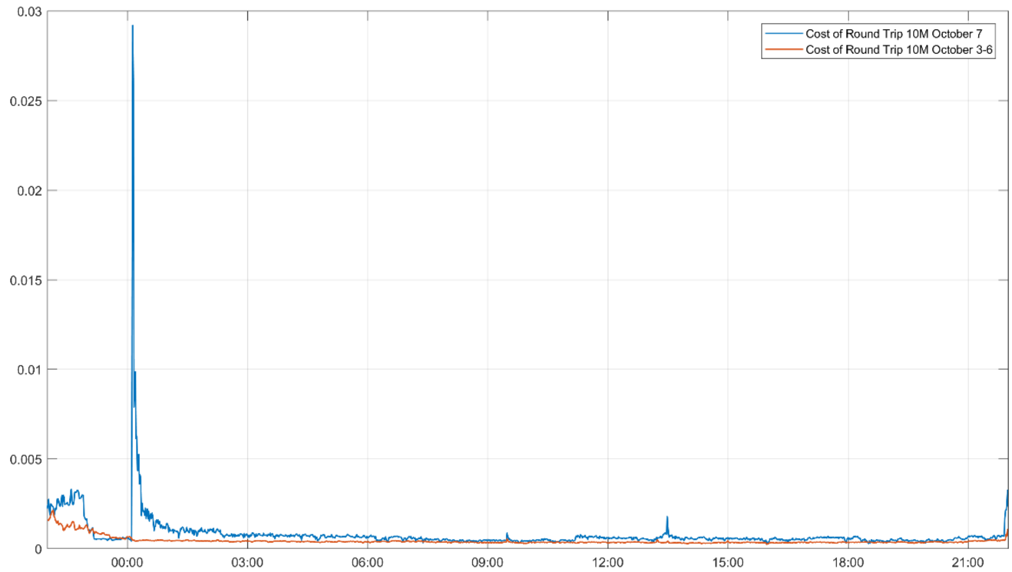


Figure 10: Cost of Round Trip for £10 million minus Cost of Round Trip for £10 in Individual Minutes in GBP/USD pair

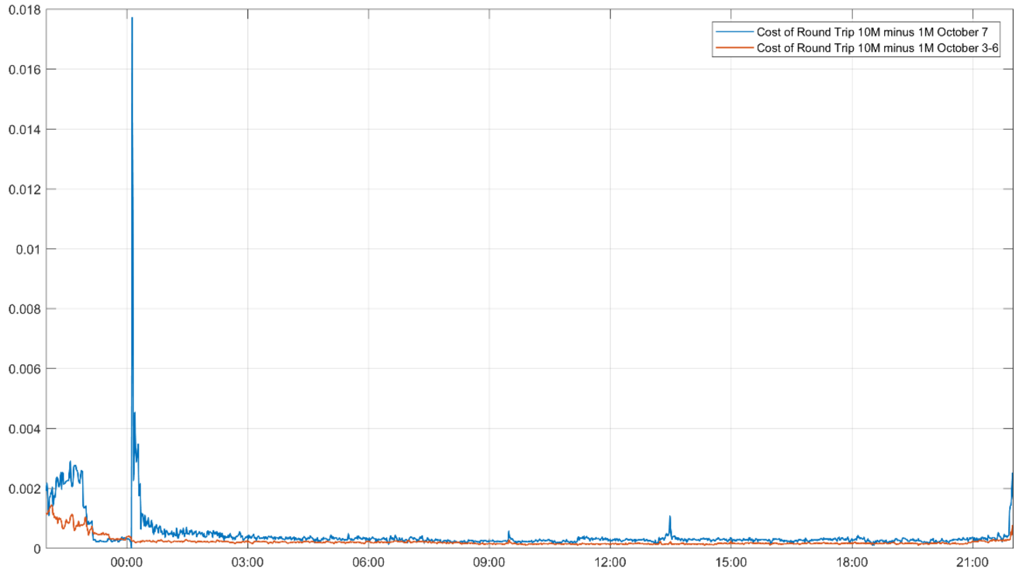
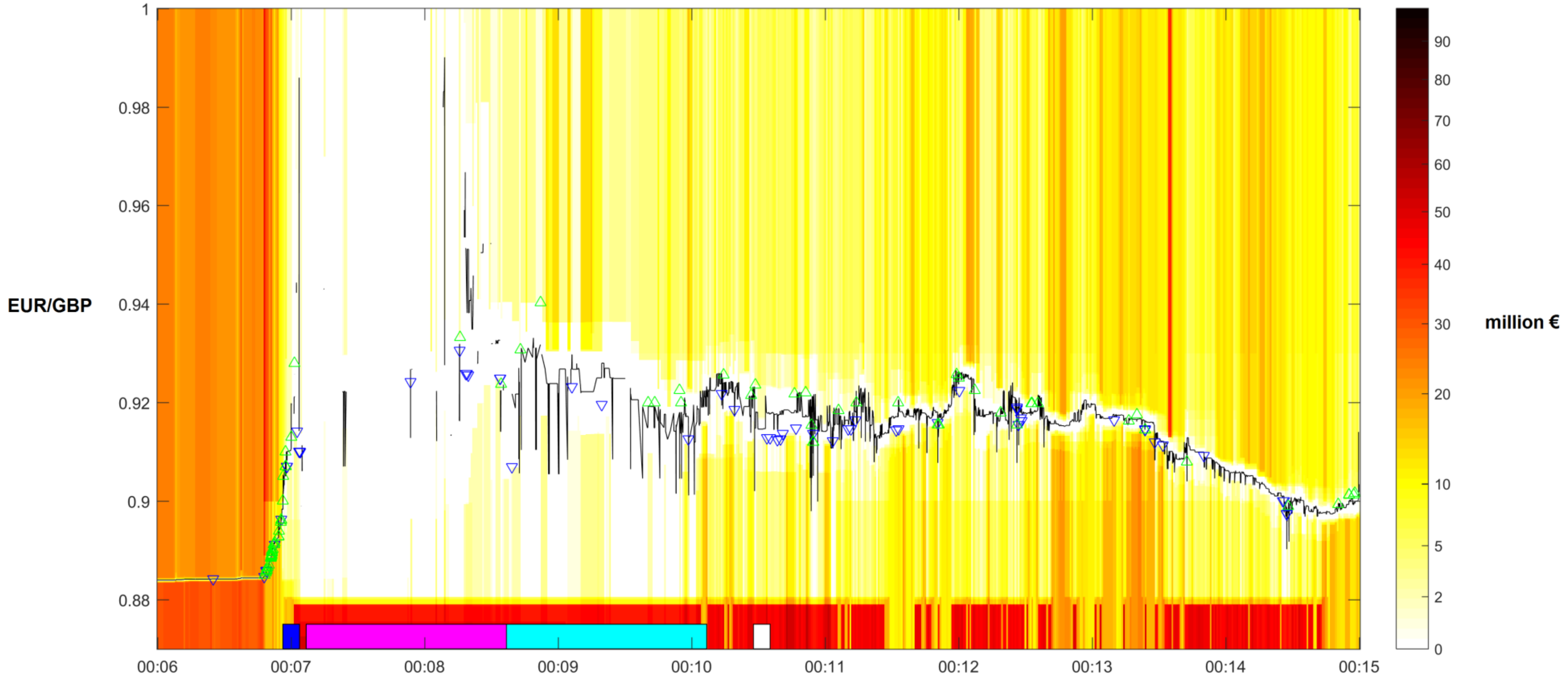


Figure 11: Visualisation of limit order book in EUR/GBP pair



- The black line shows midpoint between best bid and ask during the interval.
- Triangles show individual transactions: those in blue pointing down indicate transactions that involve the sale of sterling, and those in green pointing upwards indicate those to buy.
- Coloured regions in the graph show the cumulative quantity of limit orders to sell and buy between a given price and the best bid/ask price. Coloured regions under/above the black line indicate the quantity of limit orders to buy/sell sterling.
- The rectangles at the bottom of the chart show periods when trading was restricted on the CME.