Applied Economics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/raec20

Crypto market dynamics in stressful conditions

Christophe Desagre, Paolo Mazza & Mikael Petitjean

To cite this article: Christophe Desagre, Paolo Mazza & Mikael Petitjean (2023) Crypto market dynamics in stressful conditions, Applied Economics, 55:27, 3121-3153, DOI: 10.1080/00036846.2022.2108754

To link to this article: https://doi.org/10.1080/00036846.2022.2108754

APPLIED ECONOMICS
2023, VOL. 55, NO. 27, 3121–3153
https://doi.org/10.1080/00036846.2022.2108754

Routledge
Taylor & Francis Group

Crypto market dynamics in stressful conditions

Christophe Desagre[a,b], Paolo Mazza[c] and Mikael Petitjean[a,c]

[a]Louvain School of Management (UCLouvain, LFIN/LIDAM), Mons, Belgium; [b]ICHEC Brussels Management School, Brussels, Belgium; [c]IESEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie Management, Lille, France

ABSTRACT
Understanding market liquidity and trading dynamics in one of the most innovative and volatile markets in the world, is crucial from the standpoint of both regulators and investors. In contrast to stocks, very little is known about the functioning of cryptos around extreme returns (ERs). Using high-frequency order-book and trade data for the 8 most widespread cryptos on 16 trading platforms over three years, we examine the contemporaneous and lagged influence of trading activity and liquidity on the occurrence of extreme returns (ERs) in a logistic regression framework adapted to rare events. Despite its huge volatility, we show that the trading and liquidity dynamics on the crypto market around ERs is not orthogonal to what traditional markets experience in stressful conditions. The number of trades is a particularly robust driver to explain the occurrence of ERs, followed by the relative spread. The same drivers are identified for traditional markets.

JEL CODES
G15

JEL CLASSIFICATION
Liquidity; trading; shocks; bubble; bitcoin; cryptos

I. Introduction

Over the last two decades, technological change in financial markets has been profound and the rise of crypto assets has contributed to it by allowing for a new digital form of payment to be globally traded. Crypto assets share the following stylized features: (1) their supply is deterministically fixed; (2) there is no central counterparty, as it is the case when a central bank chooses the quantity of money in circulation; (3) transactions are recorded in a register called the *blockchain*; (4) miners ensure the stability of the network by solving cryptographic problems; and (5) these miners are rewarded with cryptos in exchange of their service. A consensus seems to emerge from the academic literature on their nature: they are non-cash flow generating assets and do not currently perform at least one of the three traditional functions of money: medium of exchange, store of value and unit of account.[1]

In about 10 years, the crypto market has become one of the most innovative and volatile markets in the world, but both regulators and investors still know very little about how trading and liquidity on this market evolve in stressful conditions, in contrast to what has been already discussed in the existing literature for stocks and bonds (Brogaard et al. 2018; Broto and Lamas 2020).

Cryptos are traded in decentralized markets 24/7 against other cryptos and traditional currencies, such as USD, EUR, JPY, or CNY. Quickly after they were used as investment vehicles, a number of derivatives were created. A tracker was introduced by the NASDAQ OMX in May 2015 followed by two futures contracts by the CBOE and CME in December 2017, while TeraExchange launched a USD/Bitcoin swap in late 2014. The development of these derivatives was an important milestone on Bitcoin's way to legitimacy as a financial asset. However, there is a still vivid debate on whether cryptos can reshape the financial system since shutdowns of platforms such as MtGox and websites such as Silk Road have questioned the reliability of the whole system. The

CONTACT Christophe Desagre christophe.desagre@ichec.be ICHEC Brussels Management School, Bd Brand Whitlock 4, 1150 Woluwe-Saint-Pierre, Brussels, Belgium

[1]There is debate on the best way to value, price, or even classify cryptos. Still, cryptos do not behave like commodities, whether scarce or not. Even bitcoin, which was initially dubbed `digital gold' because of its fixed-supply scarcity, is not a safe haven. On the contrary, cryptos are very sensitive to the level of risk aversion and seem to increasingly behave like small-cap, tech-oriented, publicly-quoted companies whose returns depend on the value of their intangible assets. In the case of cryptos, the intangible assets are their brand and in some cases their underlying technology. Interestingly, we find similarities in liquidity and trading dynamics between cryptos and stocks.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/00036846.2022.2108754.

renewed cryptocurrency rollercoaster ride in 2022 has also frayed investor nerves again. At the time of writing in July 2022, the current maximum drawdown on the bitcoin is 72% (in USD), a dramatic fall in just 222 days.

As the potential of cryptos is growing, market stability is at stake and central banks are fully aware of that threat. Their best response until now has been to investigate the issuance of their own digital currencies. Cryptos have also gone through several successive phases of sharp rises and drops in prices. This is in this context that Nouriel Roubini told the U.S. Senate Committee on Banking, Housing and Community Affairs at a hearing on 11 October 2018 that 'crypto is the mother or father of all scams and bubbles'. Nobody can deny that there is indeed no bright future for cryptos without further market stability. In particular, investors willing to include cryptos in their portfolio should care about the extreme returns (ERs) that cryptos can experience in a very short time.

Although the growing popularity of cryptos has attracted the attention of academics, regulators, and central banks (Makarov and Schoar 2020; Easley, O'Hara, and Basu 2019; Ali et al. 2014; McLeay, Radia, and Thomas 2014), the market functioning of cryptos in stressful conditions has been barely touched upon. This paper fills that gap by zooming in on the liquidity and trading dynamics of cryptos when ERs occur.

The literature on extreme returns in traditional securities markets is abundant and it is beyond the scope of this paper to review it extensively. Among the factors identified to explain returns, liquidity, trading volume, and order flow of informed trading are most commonly cited. There are several methodologies that identify important price changes. For example, Weber and Rosenow (2006) use price changes exceeding 5 standard deviations; Lee and Mykland (2012) propose a non-parametric test to distinguish between noise and jumps in the asset value; Brogaard et al. (2018, 253) define stressful periods as '*unexpected and rapidly developing extreme returns that belong to the* 99.9th*percentile of the return distribution*'.

Among these studies, we choose to follow Brogaard et al. (2018) who use 10-second intervals for stocks traded on NASDAQ. We take one-hour intervals as cryptos are not traded at the same regularity and frequency, being much more illiquid than traditional large cap companies. Another difference between crypto markets and more conventional stock markets is that cryptos do not generate any cash flow and do not have any underlying fundamental value. As such, the definition of informed traders in crypto markets needs to be altered. Contrary to Hasbrouck (1995) who defines informed trading relative to the fundamental value, we define it as the ability to monitor the order flow dynamics in the order book to make relevant trading decisions.

Some papers have documented sharp variations in crypto returns. For example, Chevapatrakul and Mascia (2019, 373) mention that '*a daily loss of around 26% was observed between* 17th*and* 18th*December 2013 and a monthly gain of 171% was realized between October and November 2013*'. Bitcoin also plummeted by 18% on 10 March 2017 following the SEC denial to launch an ETF. According to Thies and Molnár (2018), daily returns can vary from −48.52% to +40.14%, while Donier and Bouchaud (2015) report that Bitcoin lost half of its value on April 2013 in a few hours. They also provide a standard liquidity analysis of the platform MtGox between December 2011 and January 2014.

While some papers look at the relationship between returns and their drivers (Balcilar et al. 2017; Fousekis and Tzaferi 2021), there are very few papers which focus on the extreme tails of the return distribution to explain these large price variations. Chaim and Laurini (2018) analyse daily volatility and return jumps in Bitcoin between April 2013 and May 2018, and use a methodology that detects jumps in line with Scaillet, Treccani, and Trevisan (2020) who focus exclusively on the now defunct Mt. Gox exchange from June 2011 to November 2013. Chevapatrakul and Mascia (2019) find evidence of investor overreaction when returns are in the extreme tails of their distribution. When extreme price variations are identified by using percentiles of the return distribution, the 5th and 95th percentiles are often used due to data constraints. For example, Vidal-Tomás, Ibáñez, and Farinós (2019, 182) define '*extreme down (up) market as 5% of the lower (upper) tail* *of the market return distribution'*. Blau (2017) uses the same threshold. We instead use the thresholds of 1% and 99% to identify the extreme returns. On the one hand, the use of a larger tail size would allow for the inclusion of more returns which, in the presence of extreme volatility, would results in noisy price shocks. On the other hand, the use of a smaller tail size would significantly reduce the amount of extreme returns. All in all, we believe that the threshold used our study is the best compromise when it comes to studying the more illiquid crypto market.[2]

It is also noteworthy that most of the empirical studies on cryptos published in the literature use daily data based on prices and volume information only. As such, they do not bring any information neither on the intraday price dynamics nor on the order book dynamics. Often, these past studies investigate the Bitcoin only, over a short time period, before the 2017 price burst, and on a single platform. We extend previous studies by providing an in-depth intraday, cross-cryptos and cross-platforms analysis of trading and liquidity dynamics in extreme market conditions, before, during and after the bubble period.

Our fundamental goal is to look for evidence, or the absence of evidence, of market dysfunctions, examining in particular the intraday liquidity and trading dynamics around ERs. Not only we identify the drivers of ERs but we also analyse their degree of dependence. We start our analysis by studying the Bitcoin traded on the Bitfinex platform. We then extend the analysis by considering the 8 most widespread cryptos in the world across the 16 most active trading platforms. From a methodological point of view, we perform a multivariate logistic regression analysis, in both lagged and contemporaneous bases, a multivariate VAR, and Granger causality tests, allowing for structural breaks as well, with the objective of better understanding how crypto markets operate and identify dysfunctions. Finally, we test for the presence of herding in crypto markets and whether ERs are related to herding.

Our main findings can be summarized as follows. We show that ERs in cryptos are accompanied by a sharp increase in trading volume, spreads and depth. Our results suggest that liquidity takers seem to become more aggressive and consume liquidity from the order book, generating a larger trading volume and enlarging the relative spread. Using the logistic regression framework adapted to rare events, we show that trade-based and order-book-based variables help explain the occurrence of ERs. The number of trades is a particularly robust driver to explain the occurrence of ERs, followed by the relative spread. When we look at the contemporaneous relationship, we identify the same usual suspects than for traditional markets. This holds true whether we extend the analysis to a multi-platform and a multi-cryptocurrency analysis, condition it on the identification of the bubble period or on volatility regimes, and distinguish permanent from transitory ERs. Most drivers are positively interconnected, whether on a contemporaneous or a lagged basis, and display short-term reversion dynamics. However, the number of trades is the only driver which significantly explains the future variations of returns in Granger causality tests allowing for structural breaks, confirming its pivotal role in explaining price movements in cryptos. Our results also point to the presence of herding when returns are positive and large in magnitude, in particular during the late 2017 sharp rise in prices. Overall, our analysis provides robust evidence that cryptocurrency markets are less dysfunctional in stressful conditions than anticipated given their sharp ups and downs.

The remainder of this paper is as follows. Section II contains a description of our data and variables. In Section III, we explain our methodology and report our empirical findings. In Section IV, we carry some robustness checks. The last section concludes.

II. Data and variables

We obtain data from Kaiko, an independent data provider that collects data directly from the exchanges. The database is made of two distinct datasets. The first dataset records all the trades that

[2]The choice of this threshold is also in line with the literature on extreme risk. For example, standard measures of risk, such as the VaR and CVaR, are typically estimated at the 99% threshold. This is the default threshold recommended by the Basel Accord which requires large international banks to hold regulatory capital for the trading book based on a 99% VaR over a 10-day holding period. This threshold is also used in recent papers such as Ji et al. (2020).

occurred on each platform with date and time, price, number of cryptos exchanged, as well as a boolean variable indicating whether the trade is buyer- or seller-initiated. The second dataset contains order book snapshots with bid/ask prices and quantities up to the tenth limit. The analysis is conducted on a period ranging from May 2015 to July 2018. This is the maximum sample size for which we have access to both trade and order book information. Although the trade dataset starts in 2010, the earliest order book data provided dates back to May 2015.

Table 1 presents key statistics for all the 16 platforms on which BTCUSD is traded, i.e. Bitfinex, Bitflyer, Bitstamp, Bittrex, BTCC, BTCE, Cexio, Coinbase, Gatecoin, Gemini, Hitbtc, Huobi, Itbit, Kraken, OkCoin, and Quoine. Regarding the daily average number of observations, 1,440 observations are required to reach the 1-minute frequency of observation. 5 platforms have more than one observation per minute, with more than 2 observations per minute for the most recent Bittrex platform. Technical glitches, platform upgrades, or simply a lack of trading may affect this number as it is the case for Gatecoin which is the third most inactive platform in terms of the daily average number of trades.

Figure 1 represents the monthly market share of each platform for BTCUSD from July 2010 to September 2018.[3] We observe that in the beginning of the period, there was a monopolistic situation held by MtGox. As we mentioned previously, this platform shut down in February 2014, which resulted in a loss of more than 400 millions of dollars for its users according to Forbes.[4] At the end of our sample period, this is Bitfinex (in dark blue) which holds the largest market share in terms of trading activity, with more than 38,000 trades per day, followed by Coinbase, Hitbtc, Huobi, and Bitstamp.

Table 1 reports descriptive statistics on all platforms on which BTCUSD is traded. The table presents the start and the end of the period for which we have order book information, the number of days, the number of observations, the daily average number of observations, the total number of trades, and the daily average of trades.

Table 2 includes the same statistics for all the major cryptos traded against USD on the Bitfinex platform, i.e. Bitcoin Cash (BCH), Bitcoin (BTC), EOS (EOS), Ethereum (ETH), Litecoin (LTC), Stellar (×LM), Monero (×MR), and Ripple (×RP). Over the full sample across

Table 1. Descriptive statistics – BTCUSD on all exchanges.

Exchange	Start	End	Number of days	Number of observations	Daily average number of observations	Number of trades	Daily average Number of trades
Bitfinex	15 May 2015	21 July 2018	1,164	1,537,507	1,321	45,133,393	38,774
Bitflyer	18 April 2018	21 July 2018	95	239,226	2,518	109,869	1,157
Bitstamp	15 May 2015	20 July 2018	1,163	1,528,025	1,314	19,608,226	16,860
Bittrex	1 June 2018	20 July 2018	50	143,323	2,866	9,948	199
Btcc	13 February 2018	20 June 2018	128	276,122	2,157	15,235	119
Btce	15 May 2015	21 July 2018	1,164	1,458,635	1,253	16,288,069	13,993
Cexio	11 December 2017	20 July 2018	222	442,639	1,994	2,885,221	12,996
Coinbase	15 May 2015	20 July 2018	1,163	1,574,932	1,354	42,520,453	36,561
Gatecoin	18 February 2016	21 July 2018	885	731,403	826	322,237	364
Gemini	12 October 2015	21 July 2018	1,018	1,379,246	1,355	8,601,783	8,450
Hitbtc	26 August 2017	21 July 2018	330	605,017	1,833	9,075,776	27,502
Huobi	10 November 2015	13 September 2017	674	687,214	1,020	13,895,640	20,617
Itbit	7 October 2015	21 July 2018	1,019	1,336,893	1,312	2,684,271	2,634
Kraken	25 August 2015	21 July 2018	1,062	1,447,116	1,363	10,298,238	9,697
Okcoin	15 May 2015	21 July 2018	1,164	1,566,039	1,345	10,403,684	8,938
Quoine	22 September 2016	21 July 2018	667	948,843	1,420	2,127,941	3,186
TOTAL				**15,902,026**		**183,979,984**	

[3]Market share is determined by the relative proportion of trades on each platform.
[4]Source: Forbes (2014). Bitcoin's Mt. Gox Goes Offline, Loses $409 M – Recovery Steps and Taking Your Tax Losses. https://bit.ly/3dqVsgx.

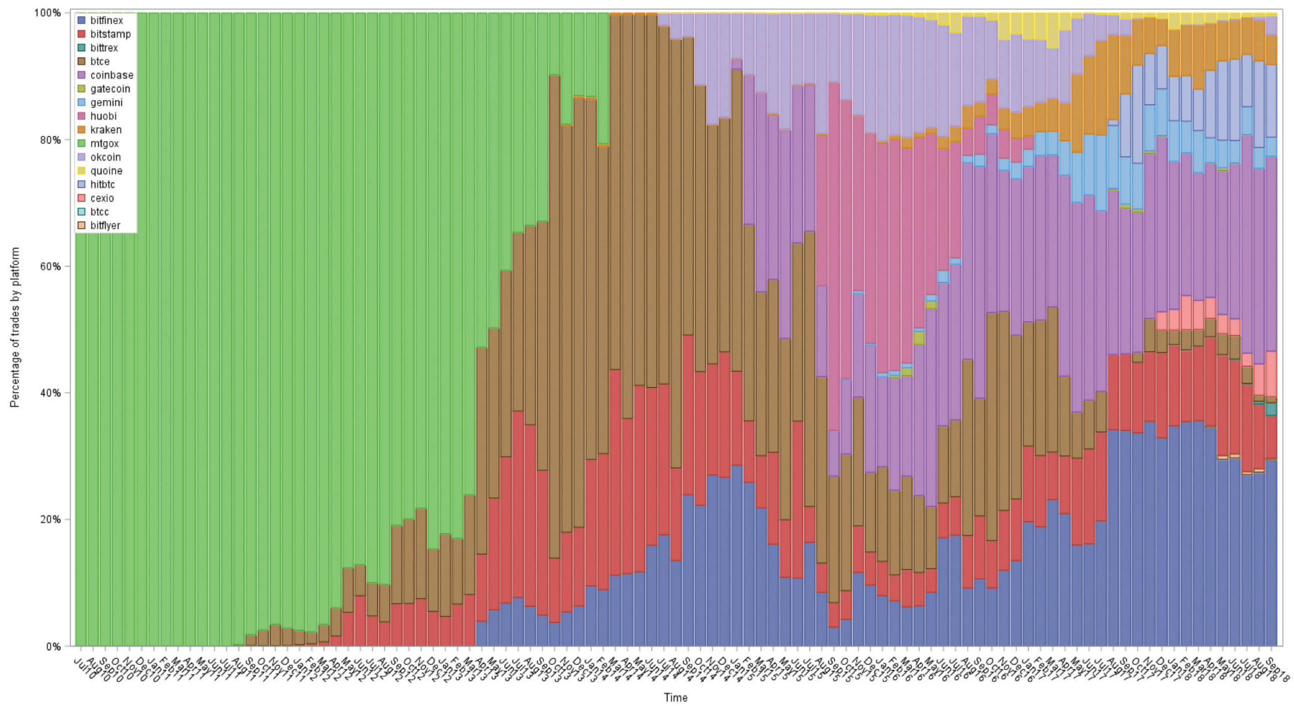

Figure 1. Market shares - BTCUSD on all exchanges. This figure represents the platforms' monthly market share for BTCUSD from July 2010 to September 2018. This figure was produced using the SAS software.

Table 2. Descriptive statistics - All cryptos on Bitfinex.

Currency	Start	End	Number of days	Number of observations	Daily average number of observations	Number of trades	Daily average Number of trades
BCHUSD	10 August 2017	21 July 2018	345	615,299	1,783	9,982,619	28,935
BTCUSD	15 May 2015	21 July 2018	1,163	1,537,507	1,322	45,133,393	38,808
EOSUSD	10 August 2017	21 July 2018	345	636,848	1,846	12,757,782	36,979
ETHUSD	28 April 2016	21 July 2018	814	1,048,549	1,288	24,009,056	29,495
LTCUSD	14 September 2016	21 July 2018	675	941,654	1,395	14,194,655	21,029
XLMUSD	2 May 2018	21 July 2018	80	206,667	2,583	46,179	577
XMRUSD	10 August 2017	21 July 2018	345	633,362	1,836	2,620,766	7,596
XRPUSD	10 August 2017	21 July 2018	345	616,421	1,787	13,484,949	39,087
TOTAL				**6,236,307**		**123,823,971**	

all platforms, we have more than 260 million trades and more than 20 million order book snapshots.

Table 2 reports descriptive statistics on all the cryptos traded in USD on Bitfinex. The table indicates the start and the end of the period for which we have order book information, the number of days, the number of observations, the daily average number of observations, the total number of trades, and the daily average of trades.

From the trade dataset, we compute the number of trades (NT), the quantities of cryptos traded (QT), the volume traded (VT, in thousands of dollars), the average trade size ($ATS = QT/NT$, in quantities), and the average trade volume ($ATV = VT/NT$, in dollars).

We compute the returns using trade prices. This approach presents the main drawback of including the bid-ask bounce, but it better reflects the economic reality in terms of returns for an illiquid market. Cryptos are fairly new assets and these markets are not as

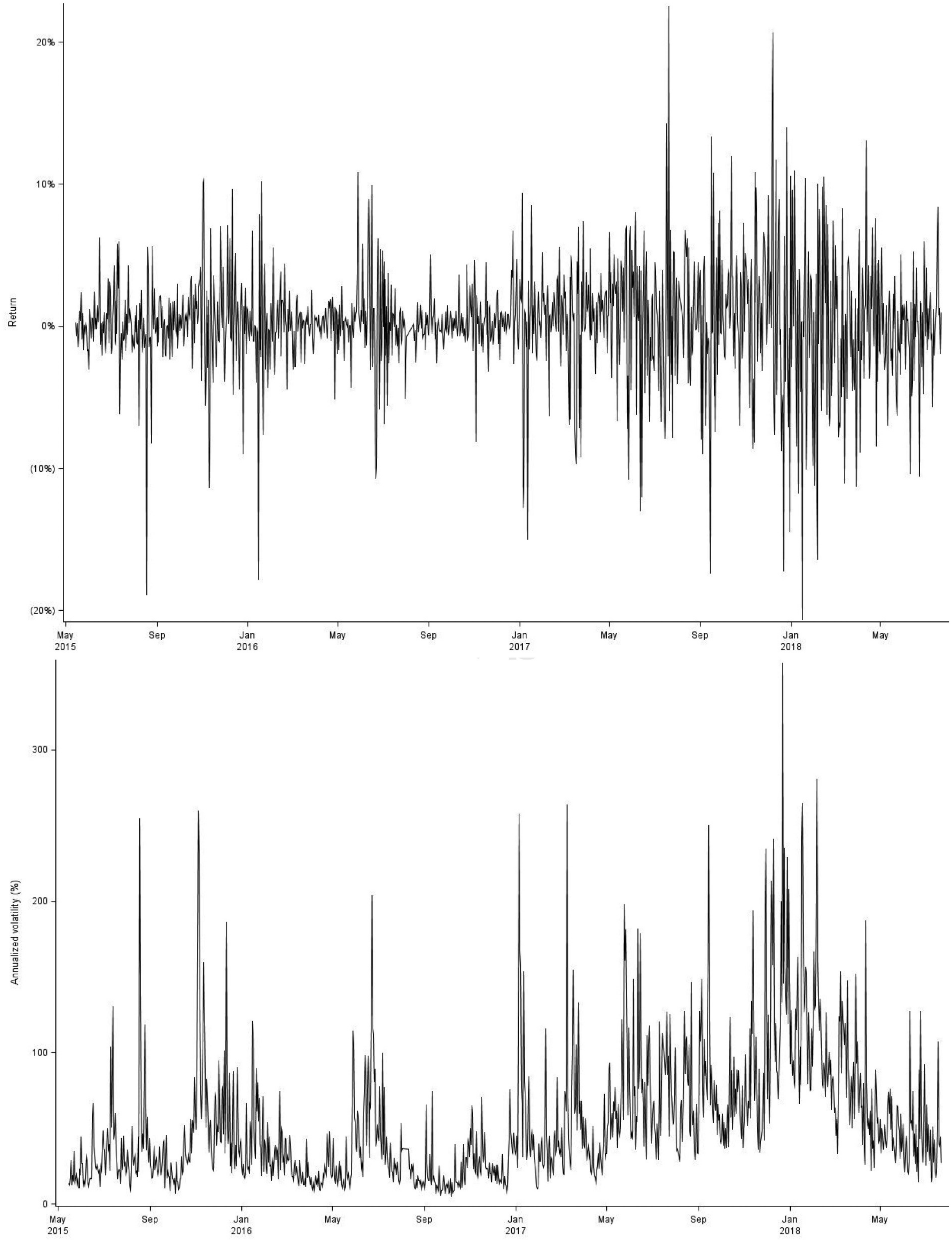

Figure 2. BTCUSD on Bitfinex - Returns (above) and volatility (below). These figures represent Bitcoin returns (above) and Bitcoin volatility (below) from May 2015 to July 2018, using hourly data. These figures were produced using the SAS software.

mature as equity, bond or forex markets. In other words, the midpoint is not a fair proxy of the prices at which investors actually trade. We plot these returns in Figure 2 for BTCUSD on Bitfinex. Bitcoin returns exhibit high variability over time.[5] To account for volatility, we compute the annualized daily volatility as[6]

$$Annualized\_Daily\_Vol_d = \sqrt{252 \times \sum_{t=1}^{24} R_t^2} \quad (1)$$

where t is the subscript for hourly frequency and d is the subscript for daily frequency.

Cryptos incur important periods of volatility as represented in Figure 2.[7] Over the sample period, the mean annualized daily volatility is 54% for BTCUSD, which is significantly higher than the typical volatility observed in equity markets.

As trade imbalance can trigger large price movements (Brogaard et al. 2018), we also estimate the trading imbalance (T_Imb) as:

$$T\_Imb_t = \frac{BUY_t - SELL_t}{BUY_t + SELL_t} \quad (2)$$

where BUY_t ($SELL_t$) is the number of buyer-initiated (seller-initiated) trades during interval t. The trades are already signed in the Kaiko database and as such, we do not use Lee and Ready (1991)'s algorithm. While this algorithm has proven to be relatively effective in equity markets, the use of the signed trades is anyway a better option.

We measure *ex ante* liquidity from the order book data set by including the quoted spread (QS, in dollars), i.e. the difference between the best bid and the best ask, the relative spread (RS), i.e. the quoted spread divided by the midpoint in %, the depth available at the best quote and at the 5 best quotes ($DEPTH$ and $DEPTH5$, in thousands of dollars), and the order book imbalance at the best quote and at the 5 best quotes (OB_Imb and OB_Imb5, in proportion of total available quantities). For each hourly interval t, we compute the average and median values of these variables by cryptocurrency and by platform.

III. Empirical analysis

We follow the methodology provided by Brogaard et al. (2018) and flag as ERs intervals during which the return exceeds the 99^{th} percentile. To ease notation, we refer to them as ER_{99}. We first consider the case of BTCUSD on the platform Bitfinex, before extending the analysis across platforms and cryptos.

The case of BTCUSD on Bitfinex

An ER_{99} occurs when the BTCUSD hourly absolute log-return exceeds 3.57% on Bitfinex. We identify 275 ER_{99} in BTCUSD on the platform Bitfinex, among which 156 are upward ERs while 119 are downward ERs.

Figure 3 shows the distribution of ERs across the day at the 99^{th} percentile. There is no clear intraday pattern in the occurrence of these ERs over the day. We conjecture that the absence of intraday pattern with respect to ERs in crypto markets comes from the fact that they are open 24/7. This is not the case for equities as documented by Brogaard et al. (2018) who find more ERs at the beginning and at the end of the NASDAQ trading session. This is one of the many stylized facts where cryptos depart from more traditional asset classes. We also find no statistically significant difference in the magnitude of downward and upward ERs.

Table 3 presents the average values for BTCUSD on Bitfinex between May 2015 and July 2018. Over all hourly intervals, the average trade size is 1.367BTC and the average trade volume is 1,681$. The number of trades sightly exceeds 1,600 trades and 1,139 units of bitcoins being traded each hour, resulting in more than 5 million dollar volume. Trade imbalance is equal to −2.3%, showing that there are more seller-initiated than buyer-initiated trades and that this difference amounts to 2.3% of

[5]We report the same figure for other cryptos, i.e. BCHUSD, EOSUSD, ETHUSD, LTCUSD, XLMUSD, XMRUSD, and XRPUSD, in the online Appendix (Figure 1).
[6]As a robustness check, we also compute the annualized daily volatility based on 288 5-minute intervals. The Pearson correlation coefficient between both measures is equal to 94%.:
[7]We report the same figure for other cryptos, i.e. BCHUSD, EOSUSD, ETHUSD, LTCUSD, XLMUSD, XMRUSD, and XRPUSD, in the online Appendix (Figure 2).

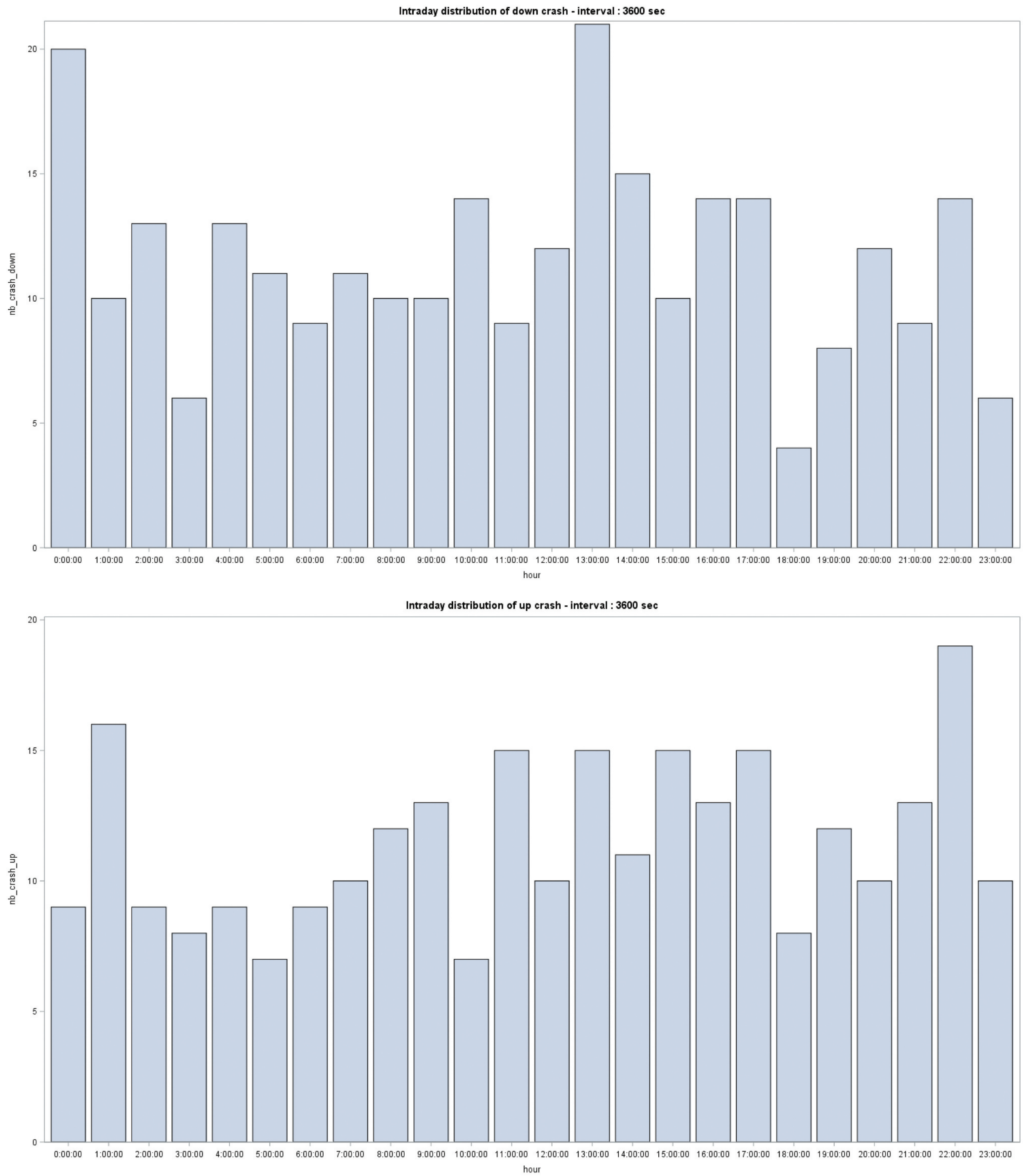

Figure 3. Intraday distribution of ERs in BTCUSD on the platform Bitfinex. These figures represent the intraday distribution of ERs in BTCUSD at the 99th percentile on Bitfinex. We separate the graphs between negative ERs (down crashes, top) and positive ERs (up crashes, bottom). These figures were produced using the SAS software.

all trades. Order book imbalance at the best quotes is 1%, meaning that there are more quantities displayed at the bid than at the ask and that this difference represents 1% of total quantities being displayed. Considering the 5 best quotes, order imbalance goes down to −0.3%. The depth

Table 3. Liquidity variables – BTCUSD on Bitfinex.

	Full	Full w/o ER	- ER	Variation	+ ER	Variation
Trade-based variables						
ATS	1.367	1.368	1.420	3.8%	1.043	(23.8%) ***
ATV	1,681	1,664	3,252	95.4% ***	3,479	109.0% ***
NT	1,616	1,539	8,636	461.1% ***	10,199	563% ***
QT	1,139	1,079	7,732	616.8% ***	6,467	499.5% ***
VT	5,186	4,854	35,300	627.2% ***	41,559	756.1% ***
T_IMB	(0.024)	(0.023)	(0.148)	(532.1%) ***	0.072	408.2% ***
Order-book-based variables						
OB_IMB	0.010	0.010	0.012	31.2%	0.056	491.7% ***
OB_IMB5	(0.003)	(0.003)	0.007	307.0%	0.061	2028.5% ***
Depth	34.83	34.57	55.24	59.8% ***	68.12	97.0% ***
Depth5	108.14	107.07	200.46	87.2% ***	234.10	118.6% ***
QS	0.637	0.611	3.233	429.5% ***	3.250	432.4% ***
RS	0.036	0.036	0.084	133.8% ***	0.057	59.7% ***

available at the best quotes amounts to 34.83k$ while it rises to 107.07k$ at the five best quotes. The quoted spread is equal to 0.64$ and the relative spread is equal to 0.036%. We also report average values across negative and positive ERs and compare them to the average values estimated over all hourly intervals *excluding* ERs to run appropriate two-sample *t*-test for difference between means.

The *t*-tests show that both the trade-based and order-book-based variables vary on average very significantly when ERs occur. Trading activity strongly increases, whether measured by the number of trades, quantities traded, or monetary volume, by 461.1%, 616.8%, and 627.2% for down ERs respectively, and by 562.7%, 499.5%, and 756.1% for up ERs respectively. In each case, the variation is statistically different from zero at the 1% level, showing strong evidence that there is a much higher demand for liquidity during ERs. The changes in trade imbalance are also in line with Brogaard et al. (2018) since it rises by 408.2% during positive ERs and decreases by 532.1% during negative ERs indicating that more transactions are initiated by buyers and sellers respectively.

Table 3 reports descriptive statistics about average trade size (*ATS*, in quantities), average trade volume (*ATV*, in dollars), depth at best quotes (*Depth*, in thousands of dollars), and at the 5 best quotes (*Depth*5, in thousands of dollars), number of trades (*NT*), order book imbalance at the best quotes (*OB_Imb*, in proportion of total available quantities), and at the 5 best quotes (*OB_Imb*5, in proportion of total available quantities), relative spread (*RS*, in %), quoted spread (*QS*, in dollars), quantities traded (*QT*), trading imbalance (*T_Imb*, in proportion of total number of trades), and volume traded (*VT*, in thousands of dollars). All variables are defined in section II. For each variable, we report the mean in the full sample, the full sample without ERs, in down (negative) and up (positive) ERs, and their respective variation with the full sample without ERs. Using a *t*-test, we test for the statistical significance of this variation. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Regarding *ex ante* (or order-book-based) liquidity, both the quoted and relative spreads increase significantly as well in the two ER samples, pointing to a higher cost of immediacy during these extreme events. Both depth proxies also significantly increase, showing that there are more quantities outstanding on both sides of the order book when an ER occurs. This may seem counter-intuitive as in this type of events, liquidity usually deteriorates in traditional markets. This can be explained by the fact that non competitive limits associated with larger quantities climb up the limit order book too quickly for being cancelled. Or simply that there are more patient traders willing to earn the spread by posting passive limit orders. While the cost of immediacy increases, larger quantities are present at both the demand and supply sides. Therefore, when there is a rally or a sell-off, there are still patient traders who do not agree to pay the increased spread but display more quantities, hoping to be lifted on the ask side or hit on the bid side. In markets where the order book is replenished less often, non competitive limits, which are associated with larger quantities as indicated above, can climb up the order book during price rallies or sell-offs. Trade prices vary, spreads widen, and the order book stabilizes after larger quantities were made available on the opposite side of the price move.

In three cases only, the null of the t-test is not rejected. First, the average trade size does not significantly increase when negative ERs occur, implying that participants do not buy more quantities of BTCUSD when its price falls down quickly. Second, negative ERs are not associated with lower order imbalance, be it at the best quote or the first five quotes. Negative shocks to the BTCUSD price do not discourage patient buyers to show up their interests: they increase the quantities displayed at the bid to a larger extent than patient sellers do at the ask side. This rise is nevertheless not strong enough to be statistically significant and trade imbalance still indicates that there are many more seller-initiated transactions during negative ERs, as expected.

We also zoom in on negative and positive ERs by aggregating our proxies at the 5-minute interval to better characterize their dynamics over time. The first 12 intervals correspond to the pre-event 1-hour window while the next 12 cover the ER itself, leading to 24 intervals. Figure 4 plots the average of the standardized number of trades (in green), trading volume (in red), and relative spread (in blue) by ER for both downward and upward ERs, together with the midpoint (in black). The number of trades seems to be a relevant proxy to identify ERs, as it strongly increases before the ER. This is confirmed in the next Section.

Addressing the drivers of ERs

To identify the key drivers of ERs in a multivariate framework, we first set the dependent variable, $ER_{99,i,j,t}$, as a dummy variable which equals 1 when an ER occurs at time t for the cryptocurrency j traded on platform i, and 0 otherwise. Since the dependent variable is a binary response variable, we implement a logistic regression framework (LOGIT) in order to address the determinants of the occurrence of ERs, while appropriately fitting the response in [0,1]. Our model is specified as follows:

$$Prob\left(ER_{99,i,j,t} = 1 | x'_{i,j,t-1}\beta, \alpha_i, \alpha_j\right) = \frac{exp(x'_{i,j,t-1}\beta + \alpha_i + \alpha_j)}{1 + exp(x'_{i,j,t-1}\beta + \alpha_i + \alpha_j)} \quad (3)$$

with

$$x'_{i,j,t-1}\beta = \alpha_0 + \beta_1 NT_{i,j,t-1} + \beta_2 T\_Imb_{i,j,t-1} + \beta_3 R_{i,j,t-1} + \beta_4 RS_{i,j,t-1} + \epsilon_{i,j,t-1} \quad (4)$$

Our selection of explanatory variables in the logistic regressions is based on Brogaard et al. (2018). We include the number of trades (NT), the trading imbalance (T_Imb), the absolute log-return (R), and the relative spread (RS), plus an intercept. All these non-dummy variables are standardized and lagged by one period. α_i and α_j denote the fixed effects for platform i and cryptocurrency j, respectively.

While Brogaard et al. (2018) use a measure of high frequency trading (HFT) activity, HFT^{NET}, this variable is irrelevant in our setting since there is no academic evidence of HFT within cryptocurrency markets. We replace this variable by a measure of trading imbalance. We also use the number of trades and not the share volume since the number of Bitcoin traded is closely related to its price level, as illustrated in Figure 5.

There are two main sample-related issues in our logistic specification. First, ERs are rare events and generate by definition a huge disequilibrium between events and non-events. For a LOGIT specification to be unbiased, a more balanced proportion of events and non-events is required. Second, the inclusion of fixed effects in the maximum likelihood function yields inconsistent estimates. This issue, called the incidental parameter problem, has been extensively discussed in the literature (Neyman and Scott et al. 1948; McFadden

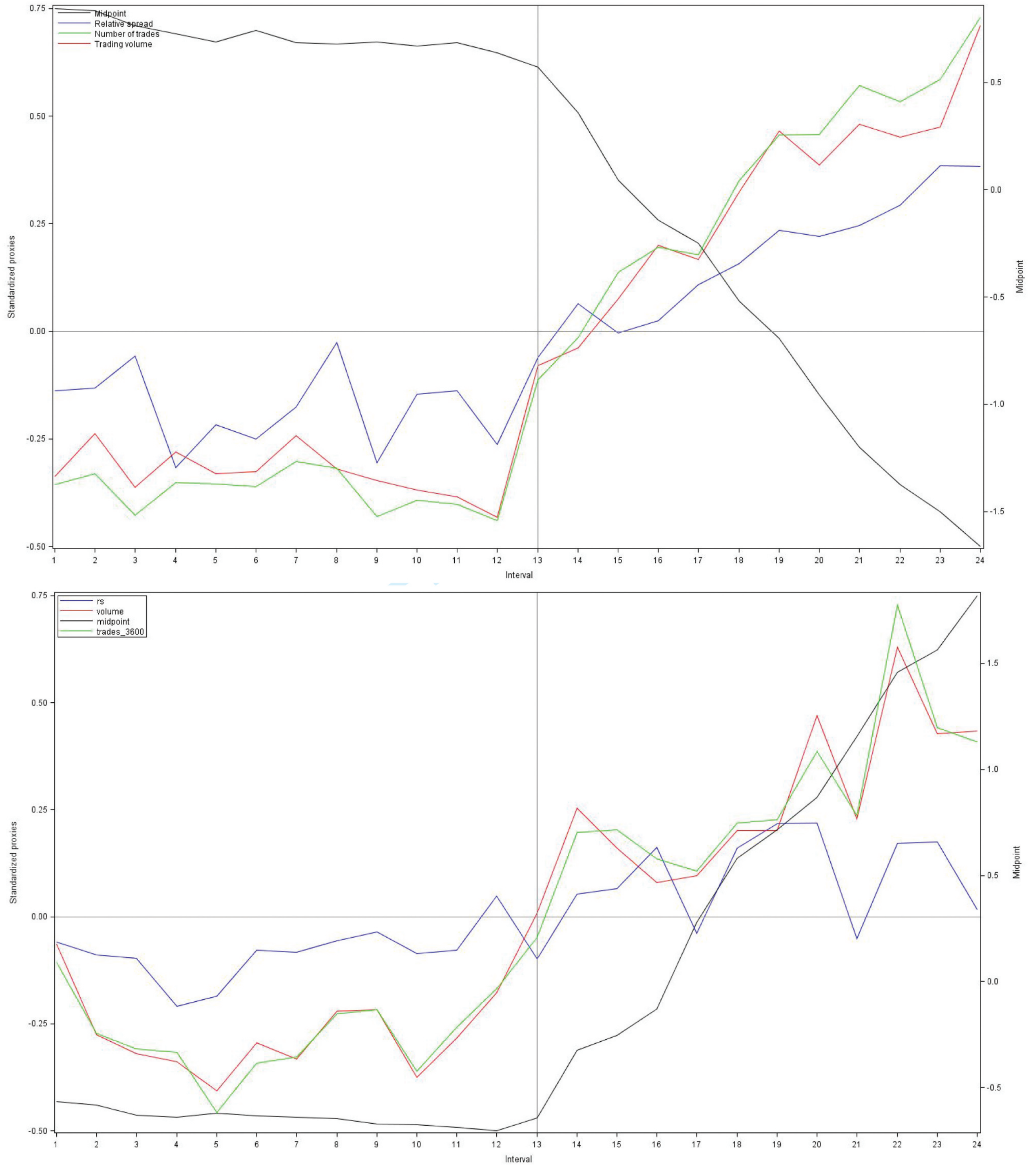

Figure 4. Zooming on the hour before the ER. These figures represent the number of trades (in green), the trading volume (in red), and the relative spread (in blue) during an ER (starting at time 13, after 60 minutes). All variables are standardized. We distinguish between downward ERs (above) and upward ERs (below). We also report the midpoint (in black) on the right-axis. These figures were produced using the SAS software.

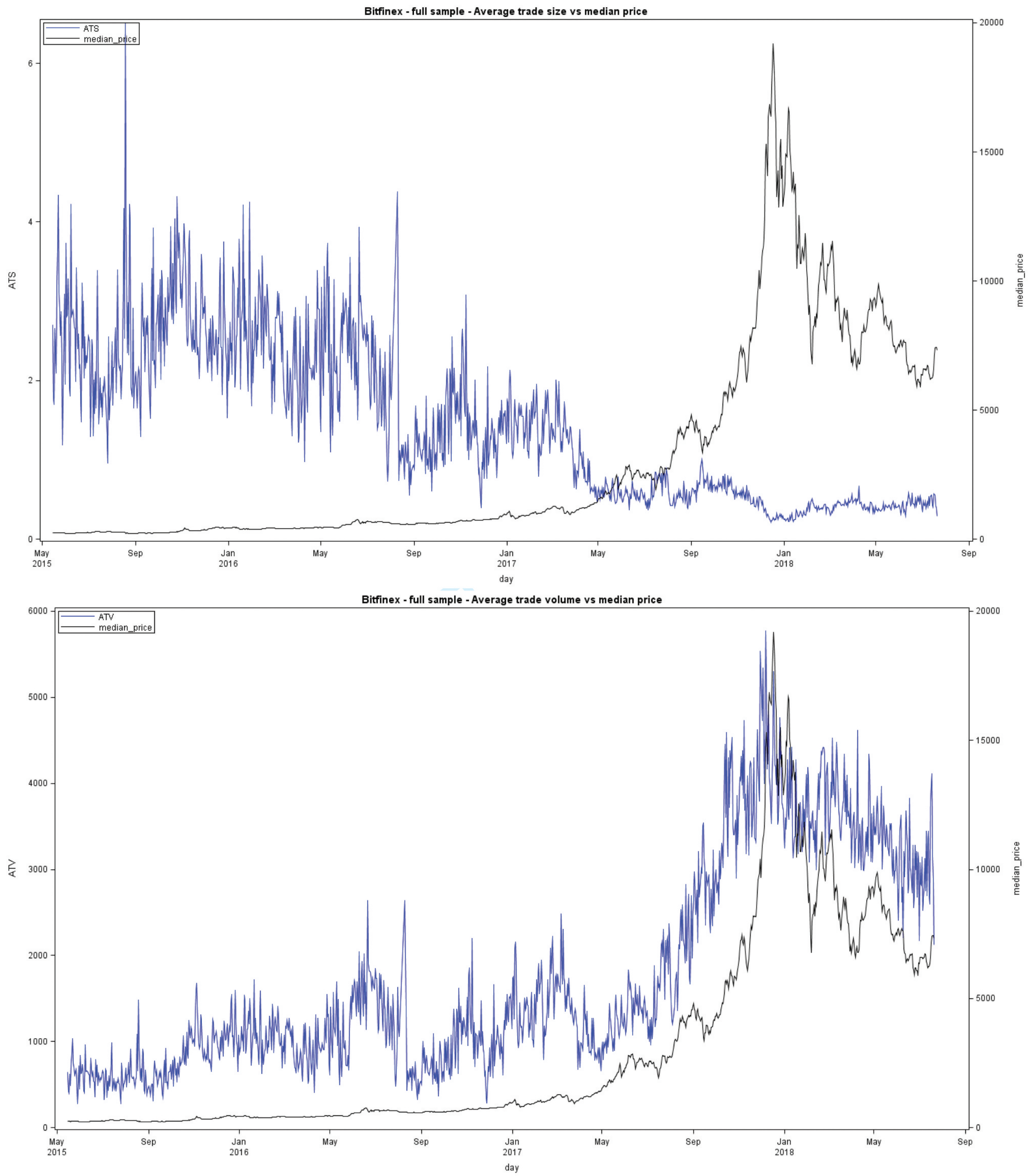

Figure 5. ATS and ATV. These figures represent the relationship between the daily median price (black line) and the average trade size (ATS, blue line) (above) and between the daily median price (black line) and the average trade volume (ATV, blue line) from May 2015 to July 2018. These figures were produced using the SAS software.

1973; Chamberlain 1980). It comes from the fact that fixed effects do not disappear from the differentiated likelihood function in non-linear frameworks.

To address the first bias, we rely on Firth (1993) and implement a penalized maximum likelihood estimation, which was initially designed to deal with cases of separation and quasi-complete separation.[8] In the case of the above-mentioned logistic model, the different parameter estimates β_k $(k = 1, \ldots, 4)$ are the solutions of the partial differential score equations $\partial logL/\partial\beta_k \equiv U(\beta_k) = 0$, with $log\ L$ being the log likelihood function. Firth (1993) proposes to correct the score equations for small sample bias, which generate quasi-complete separation, i.e. one regressor almost perfectly categorizes events and non-events.

In the case of a general logistic model, the score equation for the parameter estimate β_k is specified as:

$$U(\beta_k)* = \sum_{i=1}^{n}[y_i - \pi_i + h_i(1/2 - \pi_i)]x_{ik} = 0, \tag{5}$$

where h_i are the ith diagonal elements of the H matrix, with $H = W^{1/2}X(X^TWX)^{-1}X^TW^{1/2}$ and $W = diag[\pi_i(1 - \pi_i)]$ is the weighting matrix. We implement this estimation method to our logistic regression framework.

The second issue, i.e. the presence of fixed effects, is less important in our case. This bias is heavily influenced by the number of data points per individual. In our case, since the number of time intervals, t, is large and the number of individuals i (j), i.e. the number of platforms (cryptos) is small, the bias is negligible (Mazza 2020; Katz 2001; Greene 2004; Coupé 2005).

We first estimate Equation (3) for one single cryptocurrency, BTCUSD, on one single platform, Bitfinex. Table 4 presents the results in three different settings. Panel A shows the results of the standard LOGIT specification performed on lagged independent variables, as presented in Equation (3). Panel B displays the results for the same model based on contemporaneous regressors, excluding the contemporaneous absolute log-return as regressor since it defines the ER itself. Finally, Panel C presents the results of a GMM-LOGIT on contemporaneous regressors with the objective of addressing potential endogeneity issues. In all panels, we report the coefficient estimates, the odds ratio (OR), the pseudo R^2, the total number of observations, N, and the number of non-events (events), $N_{y=0}$ ($N_{y=1}$).

Table 4 reports the results of Equation (3). The dependent variable is the occurrence of an ER at time t and the independent variables include an intercept, the number of trades (NT), the absolute trade imbalance (Abs_T_Imb), the trade imbalance (T_Imb), the absolute log-return (R), and the relative spread (RS). All variables, excepting the intercept, are standardized. In Panel A, we estimate a LOGIT regression with Firth (1993)'s correction, with confidence intervals computed from the profile penalized log likelihood. The variables are lagged by one period. In Panel B, we estimate with contemporaneous variables. In Panel C, we use a GMM-LOGIT estimation. We also report the R-squared. The odds ratio (OR) is the exponential of the coefficient estimate. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. These estimates have been produced using the R software.

As indicated in Panel A, all explanatory variables are significant drivers of ERs when considering positive and negative shocks together. The number of trades, absolute log-return, and relative spread display positive coefficients. Among them, the number of trades has the strongest impact. Since the variables are standardized, the parameter estimates can be interpreted as follows: a one-standard deviation increase in the number of trades, absolute log-return, and relative spread increases the odds of observing an ER during the next hour by 53.0%, 21.5%, and 3.4%, respectively. The absolute trading imbalance displays a negative coefficient and an odds ratio of 0.600, thereby implying that a one-standard deviation increase in the absolute trading imbalance leads to a decrease in the odds of having an ER in the next hour by 40.0%. In other words, an

[8]Interested readers should refer to Heinze and Schemper (2002) and Heinze (2006).

Table 4. LOGIT - Bitfinex – BTCUSD.

	ERs		Down ERs		Up ERs	
Panel A: LOGIT(t-1)	Coeff.	OR	Coeff.	OR	Coeff.	OR
α_0	−5.1118***		−5.478***		−5.8593***	
NT_{t-1}	0.4255***	1.530	0.4034***	1.497	0.4909***	1.634
$Abs\_T\_Imb_{t-1}$	−0.5102***	0.600				
T_Imb_{t-1}			0.0342	1.035	−0.1616	0.851
R_{t-1}	0.1950***	1.215	0.1745***	1.191	0.1567***	1.170
RS_{t-1}	0.0339**	1.035	0.0376**	1.038	0.0415**	1.042
Pseudo R^2	0.0162		0.0068		0.0081	
$N_{y=0}$	24,924		25,034		25,064	
$N_{y=1}$	250		140		110	
N	25,174		25,174		25,174	
	ERs		Down ERs		Up ERs	
Panel B: LOGIT(t)	Coeff.	OR	Coeff.	OR	Coeff.	OR
α_0	−5.4569***		−6.0695***		−6.5356***	
NT_t	0.8316***	2.297	0.7387***	2.093	0.8208***	2.272
$Abs\_T\_Imb_t$	−0.0842	0.919				
T_Imb_t			−0.7424***	0.476	0.8518***	2.344
RS_t	0.0433***	1.044	0.0489***	1.050	0.0446***	1.046
Pseudo R^2	0.0316		0.0164		0.0162	
$N_{y=0}$	25,069		25,179		25,210	
$N_{y=1}$	251		141		110	
N	25,320		25,320		25,320	
	ER		ER DOWN		ER UP	
Panel C: GMM-LOGIT(t)	Coeff.	OR	Coeff.	OR	Coeff.	OR
α_0	−5.4671***		−6.0821***		−6.5533***	
NT_t	0.8327***	2.230	0.7403***	2.097	0.8232***	2.278
$Abs\_T\_Imb_t$	−0.0898	0.914				
T_Imb_t			−0.7415***	0.476	0.8500***	2.340
RS_t	0.0409***	1.042	0.0446***	1.046	0.0325***	1.033
$N_{y=0}$	25,069		25,179		25,210	
$N_{y=1}$	251		141		110	
N	25,320		25,320		25,320	

ER is less likely to occur when sellers or buyers have already put pressure in the market by trading more aggressively during the last hour.

When we distinguish between down and up ERs (in columns 4 to 7), we obtain very similar results with respect to the number of trades, the absolute log-return, and the relative spread. The signs and magnitudes of parameter estimates are consistent across the three models on ERs, down ERs and up ERs. Since we model directional ERs, we must use the signed trade imbalance (*T_Imb*) instead of the absolute trade imbalance (*Abs_T_Imb*). The signed trade imbalance is not significant. This suggests that the relationship between trade imbalance and ERs would be contemporaneous and not lagged, questioning the significance of the absolute trade imbalance in the model where all ERs are combined.

In Panel B, all determinants are contemporaneous, which allows us to further discuss the effect of trading imbalance. Consistent with the literature, a negative (positive) coefficient for down (up) ERs is observed for signed trade imbalance. The parameter estimates are significant at the 1% level and their magnitude is strong, i.e. −0.7424 and 0.8518 for down and up ERs respectively. The model in which all ERs are combined seems to hide the true information content of trade imbalance since *Abs_T_Imb* is not statistically significant. In both the lagged and contemporaneous models,

Table 5. LOGIT - All platforms - All cryptos.

	ER		ER DOWN		ER UP	
	Coeff.	OR	Coeff.	OR	Coeff.	OR
Panel A: LOGIT(*t*-1)						
α_0	−4.9754***		−5.4291***		−5.7531***	
NT_{t-1}	0.2592***	1.296	0.2162***	1.241	0.2551***	1.291
$Abs\_T\_Imb_{t-1}$	−0.2547***	0.775				
T_Imb_{t-1}			−0.0729***	0.930	0.0929***	0.911
R_{t-1}	0.3438***	1.410	0.3051***	1.357	0.3213***	1.379
RS_{t-1}	0.1726***	1.188	0.0872***	1.091	0.0959***	1.101
Pseudo R^2	0.0134		0.0056		0.0081	
$N_{y=0}$	317,050		318,548		318,785	
$N_{y=1}$	3,233		1,735		1,498	
N	320,283		320,283		320,283	
Panel B: LOGIT(t)						
α_0	−5.1472***		−5.7973***		−5.9099***	
NT_t	0.6706***	1.955	0.5812***	1.788	0.5738***	1.775
$Abs\_T\_Imb_t$	−0.0216	0.979				
T_Imb_t			−0.7234***	0.485	0.4321***	1.541
RS_t	0.3258***	1.385	0.2065***	1.229	0.1815***	1.199
Pseudo R^2	0.0216		0.0118		0.0097	
$N_{y=0}$	335,954		337,473		337,710	
$N_{y=1}$	3,275		1,756		1,519	
N	339,229		339,229		339,229	
Panel C: GMM-LOGIT(t)						
α_0	−5.1508***		−5.8695***		−5.914***	
NT_t	0.6562***	1.927	0.5657***	1.761	0.5658***	1.761
$Abs\_T\_Imb_t$	−0.0385*	0.962				
T_Imb_t			−0.6931***	0.500	0.4127***	1.511
RS_t	0.3175***	1.374	0.2023***	1.224	0.1831***	1.201
$N_{y=0}$	335,954		337,473		339,229	
$N_{y=1}$	3,275		1,756		1,519	
N	339,229		339,229		1,519	

we obtain positive and significant parameter estimates for the number of trades and the relative spread, with the number of trades playing an even greater role contemporaneously. As in Panel A, we find that negative and positive ERs are both driven by the same determinants.

In Panel C, we estimate the regression of Panel B using the generalized method of moments (GMM) to control for endogeneity. We obtain very similar results. The number of trades, return, and relative spread exhibit a positive and statistically significant coefficient; the trading imbalance is significant and again exhibits a negative (positive) sign for down (up) ERs.

While Bitfinex is the most important platform in terms of trading activity, it represents 25% of trading activity in BTCUSD in our sample. As indicated in Table 1, there are 15 other platforms in our sample where BTCUSD is exchanged, and many other cryptos. To take the cross-cryptos and cross-platforms dynamics into account, we estimate Equation (3) on the full sample. All non-dummy variables are standardized at the cryptocurrency platform level. The results are depicted in Table 5.

All signs and significance levels are consistent with those obtained in the case of BTCUSD on Bitfinex. Differences are mostly observed in the magnitude of the effects. Again, we find that the number of trades display a positive and statistically significant coefficient at the 1% level, as does the relative spread. Signed trade imbalance shows the same signs and level of significance than in Table 4, being also significant in the lagged model specification.

All in all, we identify early warning signs of an ER in both the order book and trading dynamics. One hour before the occurrence of an ER, liquidity tends to deteriorate with spreads widening; trading intensifies as indicated by the rise in the number of trades; and there is more volatility as proxied by the absolute log return. The estimates in the contemporaneous specifications presented in Panel B and Panel C are always very close to the theoretical explanations and the empirical results found in the literature on more traditional markets.[9]

Table 5 reports the results of Equation (3). The dependent variable is the occurrence of an ER at time t and the independent variables include an intercept, the number of trades (NT), the absolute trade imbalance (Abs_T_Imb), the trade imbalance (T_Imb), the absolute log-return (R), and the relative spread (RS). All variables, excepting the intercept, are standardized. In Panel A, we estimate a LOGIT regression with Firth (1993)'s correction, with confidence intervals computed from the profile penalized log likelihood. The variables are lagged by one period. In Panel B, we estimate with contemporaneous variables. In Panel C, we use a GMM-LOGIT estimation. We also report the R-squared. The odds ratio (OR) is the exponential of the coefficient estimate. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. These estimates have been produced using the R software.

Pre-bubble, bubble, and post-bubble

Bitcoin has gone through several market cycles since its launch. It first exceeded the $1,000 gap in January 2017 (Corbet and Katsiampa 2020). It then experienced a sharp price increase, reaching almost $20,000 in December 2017, before declining to approximately $3,000 a year later. Not surprisingly, there has been quite strong empirical evidence of bubbles in cryptocurrency markets since then (Bouri, Gupta, and Roubaud 2019; Chen and Hafner 2019; Chaim and Laurini 2019; Ji et al. 2019).

Liu et al. (2019) use Phillips, Wu, and Yu (2011)'s and Phillips and Yu (2011)'s methodologies to date the bubble of the Bitcoin. They distinguish three periods of time: pre-bubble (before 24 May 2017), bubble (between 25 May 2017 and 28 January 2018), and post-bubble (after 28 January 2018). Over the sample period, the volatility is 54% on average, with strong disparities across sub-periods. On average, the volatility is equal to 38%, 91%, and 68%, before, during, and after the bubble respectively. These differences in means are statistically significant at the 1% level.[10]

We replicate the previous analyses by estimating Equation (3) on three subsample periods, i.e. the pre-, bubble, and post-bubble subsample periods. Table 6 replicates the lagged Logit analysis, Table 7, the contemporaneous Logit, and Table 8, the contemporaneous GMM-Logit. We standardize the variables within each of these three periods of time. There are 741 days between 15 May 2015 and 24 May 2017; 249 days between 25 May 2017 and 28 January 2018; and 174 days between 29 January 2018 and 21 July 2018. As expected, the majority of ERs, i.e. 130, occurs during the bubble period even if it is relatively short. This is consistent with the fact that the absolute log-returns are higher in more volatile times.

Table 6 reports the results of Equation (3). The dependent variable is the occurrence of an ER at time t and the independent variables include an intercept, the number of trades (NT), the absolute trade imbalance (Abs_T_Imb), the trade imbalance (T_Imb), the absolute log-return (R), and the relative spread (RS). All variables, excepting the intercept, are standardized. We estimate a LOGIT regression with Firth (1993)'s correction, with confidence intervals computed from the profile penalized log likelihood. The variables are lagged by one period.The subsamples are before the bubble (before 24 May 2017), during the bubble (between 25 May 2017 and 28 January 2018), and after the bubble (after 28 January 2018). The odds ratio (OR) is the exponential of the coefficient estimate. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. These estimates have been produced using the R software.

[9]We conducted these analyses with and without fixed effects for the LOGIT specifications and the results are not sensitive to their inclusion or removal.
[10]Figure 3 in the online Appendix demonstrates these dynamics graphically by depicting the distribution of volatility estimates across the three time periods.

Table 6. LOGIT - Bitfinex – BTCUSD - Lagged model around the bubble.

	ER		ER DOWN		ER UP	
	Coeff.	OR	Coeff.	OR	Coeff.	OR
Panel A : Pre-bubble						
LOGIT(*t*-1)						
α_0	−5.7716***		−6.0458***		−6.7217***	
NT_{t-1}	0.2176***	1.243	0.3223***	1.380	0.1654	1.180
$Abs\_T\_Imb_{t-1}$	−0.4805***	0.619				
T_Imb_{t-1}			0.1865	1.205	−0.3354	0.715
R_{t-1}	0.2675***	1.307	0.1445***	1.156	0.2140**	1.239
RS_{t-1}	0.0422**	1.043	0.0475**	1.049	0.0562**	1.058
Pseudo R^2	0.0084		0.0049		0.0029	
$N_{y=0}$	16,239		12,263		16,288	
$N_{y=1}$	73		49		24	
N	16,312		16,312		16,312	
Panel B : Bubble						
LOGIT(*t*-1)						
α_0	−4.0247***		−4.5582***		−4.8238***	
NT_{t-1}	0.5880***	1.800	0.5974***	1.817	0.5327***	1.704
$Abs\_T\_Imb_{t-1}$	−0.1712	0.843				
T_Imb_{t-1}			−0.0206	0.980	−0.1904	0.827
R_{t-1}	0.0604	1.062	0.0366	1.037	0.0781	1.081
RS_{t-1}	0.1997***	1.221	0.0077	1.008	0.2606***	1.298
Pseudo R^2	0.0298		0.0136		0.0163	
$N_{y=0}$	4,583		4,642		4,654	
$N_{y=1}$	130		71		59	
N	4,713		4,713		4,713	
Panel C : Post-bubble						
LOGIT(*t*-1)						
α_0	−4.8835***		−5.9411***		−5.3690***	
NT_{t-1}	0.3732***	1.452	0.0992	1.104	0.5202***	1.682
$Abs\_T\_Imb_{t-1}$	0.1160	1.123				
T_Imb_{t-1}			−0.5144*	0.598	0.0792	1.082
R_{t-1}	−0.2566**	0.774	−0.5402**	0.583	−0.1342	0.874
RS_{t-1}	0.5768***	1.780	0.8738***	2.396	0.2976**	1.347
Pseudo R^2	0.0201		0.0124		0.0107	
$N_{y=0}$	4,102		4,129		4,122	
$N_{y=1}$	47		20		27	
N	4,149		4,149		4,149	

Table 7 reports the results of Equation (3). The dependent variable is the occurrence of an ER at time t and the independent variables include an intercept, the number of trades (NT), the absolute trade imbalance (Abs_T_Imb), the trade imbalance (T_Imb), the absolute log-return (R), and the relative spread (RS). All variables, excepting the intercept, are standardized. We estimate a LOGIT regression with Firth (1993)'s correction, with confidence intervals computed from the profile penalized log likelihood. The variables are lagged by one period.The subsamples are before the bubble (before 24 May 2017), during the bubble (between 25 May 2017 and 28 January 2018), and after the bubble (after 28 January 2018). The odds ratio (OR) is the exponential of the coefficient estimate. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. These estimates have been produced using the R software.

Table 8 reports the results of Equation (3). The dependent variable is the occurrence of an ER at time t and the independent variables include an intercept, the number of trades (NT), the absolute trade imbalance (Abs_T_Imb), the trade imbalance (T_Imb), the absolute log-return (R), and the relative spread (RS). All variables, excepting the intercept, are standardized. We estimate a LOGIT regression with Firth (1993)'s correction, with

Table 7. LOGIT - Bitfinex – BTCUSD - Contemporaneous model (OLS) around the bubble.

	ER		ER DOWN		ER UP	
	Coeff.	OR	Coeff.	OR	Coeff.	OR
Panel A : Pre-bubble						
LOGIT(t)						
α_0	−6.6873***		−7.6626***		−7.3885***	
NT_t	0.8927***	2.442	0.8869***	2.427	0.6096***	1.840
$Abs\_T\_Imb_t$	0.1017	1.107				
T_Imb_t			−1.1074***	0.330	0.8639***	2.372
RS_t	0.0565***	1.058	0.0685***	1.071	0.0580**	1.060
Pseudo R^2	0.0261		0.0201		0.0060	
$N_{y=0}$	16,301		16,325		16,249	
$N_{y=1}$	72		48		24	
N	16,373		16,373		16,373	
Panel B : Bubble						
LOGIT(t)						
α_0	−4.4820***		−5.3400***		−5.4296***	
NT_t	0.9805***	2.666	0.8440***	2.326	0.9050***	2.472
$Abs\_T\_Imb_t$	0.2492**	1.283				
T_Imb_t			−0.9837***	0.374	0.8884***	2.431
RS_t	0.3104***	1.364	0.2815***	1.325	0.2269***	1.255
Pseudo R^2	0.0668		0.0416		0.0318	
$N_{y=0}$	4,634		4,693		4,707	
$N_{y=1}$	132		73		59	
N	4,766		4,766		4,766	
Panel C : Post-bubble						
LOGIT(t)						
α_0	−6.0096***		−6.6904***		−6.9862***	
NT_t	1.2456***	3.475	0.8738***	2.396	1.1957***	3.306
$Abs\_T\_Imb_t$	−0.0170	0.983				
T_Imb_t			−0.8631***	0.422	1.2779***	3.589
RS_t	0.0514	1.053	0.1828	1.201	0.0440	1.045
Pseudo R^2	0.0566		0.0218		0.0368	
$N_{y=0}$	4,134		4,161		4,154	
$N_{y=1}$	47		20		27	
N	4,181		4,181		4,181	

confidence intervals computed from the profile penalized log likelihood. The variables are lagged by one period.The subsamples are before the bubble (before 24 May 2017), during the bubble (between 25 May 2017 and 28 January 2018), and after the bubble (after 28 January 2018). The odds ratio (OR) is the exponential of the coefficient estimate. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. These estimates have been produced using the R software.

In the lagged logistic model specification detailed in Table 6, we see that the number of trades and the relative spread display positive and statistically significant coefficients in almost every case. When combining all ERs, we see that a one-standard deviation increase in the number of trades (the relative spread) increases the odds of observing an ER during the next hour by 24.3% (4.3%) before the bubble, by 80% (22.1%) during the bubble, and by 45.2% (78%) after the bubble. While the *ceteris paribus* effect of the number of trades on the occurrence of ERs is the highest during the bubble, the coefficient estimate of the relative spread variable increases and reaches its maximum value after the bubble, becoming even more important than the number of trades. This is evidence that order-book-based liquidity conditions have become increasingly important over time in determining market stability. Comparing Panels A, B and C, trading activity, i.e. the number of trades, and the relative spread remains the most significant and consistent drivers over time with the relative spread best anticipating the occurrence of an ER after the bubble.

Table 8. LOGIT - Bitfinex – BTCUSD - Contemporaneous model (GMM) around the bubble.

	ER		ER DOWN		ER UP	
	Coeff.	OR	Coeff.	OR	Coeff.	OR
Panel A : Pre-bubble						
GMM-LOGIT(t)						
α_0	−6.7281***		−7.7313***		−7.4601***	
NT_t	0.8982***	2.455	0.8961***	2.450	0.6154***	1.850
$Abs\_T\_Imb_t$	0.0894	1.093				
T_Imb_t			−1.1159***	0.328	0.8702***	2.387
RS_t	0.0417***	1.043	0.0510***	1.052	0.0175	1.018
$N_{y=0}$	16,301		16,325		16,249	
$N_{y=1}$	72		48		24	
N	16,373		16,373		16,373	
Panel B : Bubble						
GMM-LOGIT(t)						
α_0	−4.5024***		−5.3715***		−5.4669***	
NT_t	0.9858***	2.680	0.8551***	2.352	0.9196***	2.508
$Abs\_T\_Imb_t$	0.2399**	1.271				
T_Imb_t			−0.9863***	0.373	0.8857***	2.425
RS_t	0.3089***	1.362	0.2655***	1.304	0.2028**	1.225
$N_{y=0}$	4,634		4,693		4,707	
$N_{y=1}$	132		73		59	
N	4,766		4,766		4,766	
Panel C : Post-bubble						
GMM-LOGIT(t)						
α_0	−6.0951***		−6.8136***		−7.1158***	
NT_t	1.2618***	3.532	0.8855***	2.424	1.2166***	3.376
$Abs\_T\_Imb_t$	−0.0460	0.955				
T_Imb_t			−0.8710***	0.419	1.2938***	3.647
RS_t	0.0523	1.054	0.1885	1.207	0.0476	1.049
$N_{y=0}$	4,134		4,161		4,154	
$N_{y=1}$	47		20		27	
N	4,181		4,181		4,181	

The informational content of the absolute trading imbalance remains unclear and appears to be much less decisive than the number of trades, as it is statistically significant during the pre-bubble period only. It also becomes insignificant when the directional trade imbalance is used to explain negative and positive ERs.

As to the absolute log-return variable, it is positive before the bubble, providing evidence of positive momentum, i.e. past absolute log-returns are positively correlated with subsequent extreme shocks in log-returns. As the market matures over time, its coefficient sign turns negative. This indicates that markets have become more contrarian after the bubble, since observing an important return in one period decreases the odds of having an ER in the following period.

Table 7 and Table 8, which refer to the analysis of the contemporaneous effects of the regressors, also highlight interesting facts. The number of trades is still positive and very significant. The behaviour of the signed trading imbalance is also significant and very consistent with the literature. The most striking finding concerns the relative spread which is not significantly correlated anymore with ERs in the post-bubble period, all else equal. Overall, it means that ERs in the post-bubble period are associated with trading dynamics and not with liquidity dynamics since variations in the relative spread do not drive them anymore. This is a common feature of more mature markets. The two analyses presented in Tables 7 and 8 lead to the same conclusions.

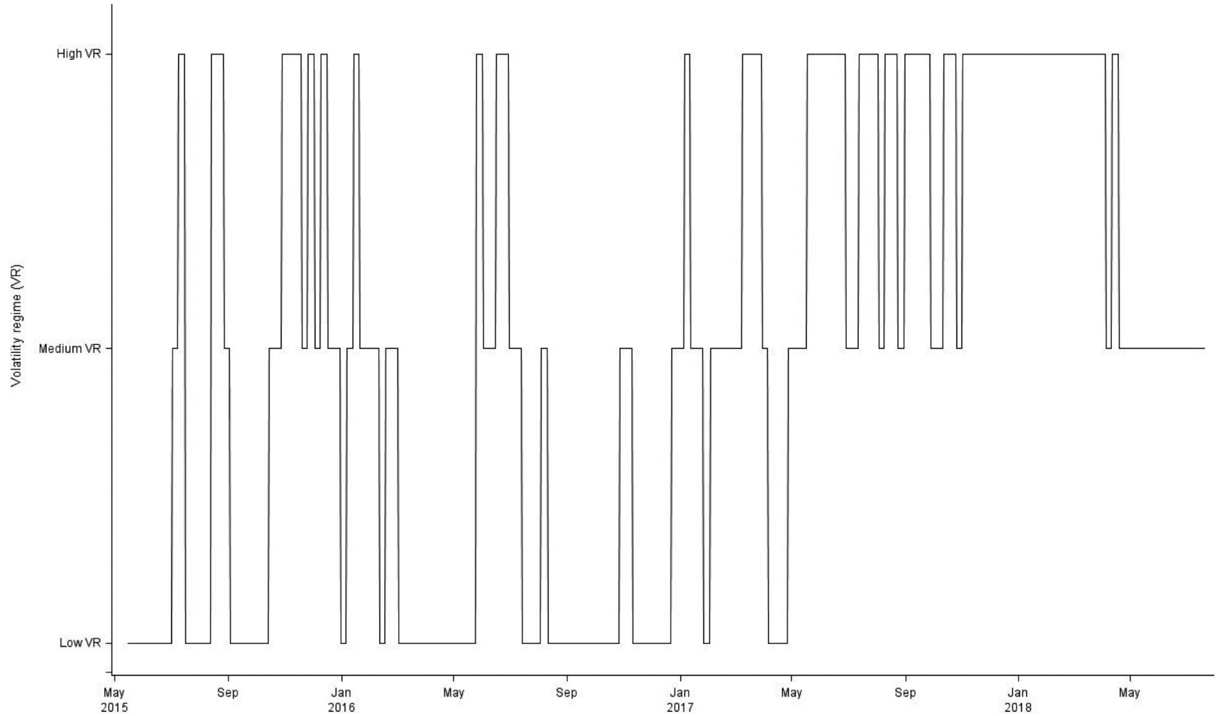

Figure 6. Volatility regimes on Bitfinex for BTCUSD. This figure represents the different volatility regimes (VR). The X-axis represents the time and the Y-axis represents the three VR. This figure was produced using the SAS software.

Volatility regimes

Instead of dividing the sample according to the occurrence of the bubble, we split our sample into three volatility regimes (VR), motivated by the differences in volatility observed over time, around the bubble period. Daily volatility is averaged by week and each week is assigned to one of the three VR, each regime including one third of the weeks. In the low volatility regime, the weekly volatility is always below 31%; in the medium regime, the weekly volatility ranges from 31% and 63%; and in the high regime, it is always higher than 63%. We also identify the ERs conditionally on the volatility regime. This analysis is complementary to the previous analyses because the three volatility regimes are not necessarily associated with successive time periods. As indicated in Figure 6, the market seems to switch frequently from one regime to the other.

We estimate Equation (3) using this new method and report the results in Tables 9, Tables 10 and 11. In Table 9, the number of trades remains a significant driver of ERs in both low and high volatility regimes, although the pseudo R^2 are lower. The absolute log-return variable often displays a positive and significant coefficient estimate in all three regimes, indicating that past absolute log-returns are positively correlated with subsequent extreme shocks in returns irrespective of the level of volatility. The relative spread is positive and significant in the three regimes but its significance and the magnitude of its coefficient estimate are higher in the low and high volatility regimes, implying that liquidity conditions drive market stability to a larger extent when volatility is 'abnormal'.

In the logit contemporaneous analysis detailed in Table 10, the number of trades and the relative spread are associated with positive and highly significant parameter estimates in all regimes. The signed trade imbalance results are also totally in line with the previous results obtained. All in all, this analysis confirms our findings from the previous analyses. From the results depicted in Table 11, which controls for endogeneity issues, no significant changes can be detected. Controlling for endogeneity via GMM does not change the conclusions.

Table 9 reports the results of Equation (3). The dependent variable is the occurrence of an ER at

Table 9. LOGIT - Bitfinex – BTCUSD - Lagged model with volatility regimes.

	ER		ER DOWN		ER UP	
	Coeff.	OR	Coeff.	OR	Coeff.	OR
Panel A : Low volatility regime						
LOGIT(*t*-1)						
α_0	−4.7259***		−5.2570***		−5.5993***	
NT_{t-1}	0.2417***	1.273	0.2614**	1.299	0.2354*	1.265
$Abs\_T\_Imb_{t-1}$	−0.1100	0.896				
T_Imb_{t-1}			−0.0229	0.977	0.2357	1.266
R_{t-1}	0.1593**	1.173	0.1376	1.148	0.1762*	1.193
RS_{t-1}	0.0695**	1.072	0.0805**	1.084	0.0721*	1.075
Pseudo R^2	0.0043		0.0027		0.0024	
$N_{y=0}$	8,442		8,476		8,489	
$N_{y=1}$	81		47		34	
N	8,523		8,523		8,523	
Panel B : Medium volatility regime						
LOGIT(*t*-1)						
α_0	−4.6556***		−5.3827***		−5.3200***	
NT_{t-1}	0.1162	1.123	0.0284	1.029	0.1808*	1.198
$Abs\_T\_Imb_{t-1}$	0.0989	1.104				
T_Imb_{t-1}			−0.2792*	0.756	0.0849	1.089
R_{t-1}	0.1780**	1.195	0.2438**	1.276	0.1366	1.146
RS_{t-1}	0.0477	1.049	0.0570*	1.059	0.0613*	1.063
Pseudo R^2	0.0018		0.0136		0.0013	
$N_{y=0}$	8,347		8,389		8,387	
$N_{y=1}$	82		40		42	
N	8,429		8,429		8,429	
Panel C : High volatility regime						
LOGIT(*t*-1)						
α_0	−4.8599***		−5.3002***		−5.8332***	
NT_{t-1}	0.4186***	1.520	0.2539**	1.289	0.5499***	1.733
$Abs\_T\_Imb_{t-1}$	−0.0985	0.906				
T_Imb_{t-1}			0.0866	1.090	−0.3002	0.741
R_{t-1}	0.2309***	1.260	0.2288***	1.257	0.2142***	1.239
RS_{t-1}	0.0511**	1.052	0.0533**	1.055	0.0645**	1.067
Pseudo R^2	0.0100		0.0124		0.0086	
$N_{y=0}$	8,096		8,135		8,142	
$N_{y=1}$	85		46		39	
N	8,181		8,181		8,181	

time t and the independent variables include an intercept, the number of trades (NT), the absolute trade imbalance (Abs_T_Imb), the trade imbalance (T_Imb), the absolute log-return (R), and the relative spread (RS). All variables, excepting the intercept, are standardized. We estimate a LOGIT regression with Firth (1993)'s correction, with confidence intervals computed from the profile penalized log likelihood. The variables are lagged by one period.The subsamples are before the bubble (before 24 May 2017), during the bubble (between 25 May 2017 and 28 January 2018), and after the bubble (after 28 January 2018). The odds ratio (OR) is the exponential of the coefficient estimate. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 10 reports the results of Equation (3). The dependent variable is the occurrence of an ER at time t and the independent variables include an intercept, the number of trades (NT), the absolute trade imbalance (Abs_T_Imb), the trade imbalance (T_Imb), the absolute log-return (R), and the relative spread (RS). All variables, excepting the intercept, are standardized. We estimate a LOGIT regression with Firth (1993)'s correction, with confidence intervals computed from the profile

Table 10. LOGIT - Bitfinex – BTCUSD - Contemporaneous model with volatility regimes (OLS).

	ER		ER DOWN		ER UP	
	Coeff.	OR	Coeff.	OR	Coeff.	OR
Panel A : Low volatility regime						
LOGIT(t)						
α_0	−5.5636***		−5.9995***		−6.6236***	
NT_t	0.9119***	2.489	0.7347***	2.085	0.8540***	2.349
$Abs\_T\_Imb_t$	−0.3488**	0.706				
T_Imb_t			−0.5720***	0.564	0.7852***	2.193
RS_t	0.0973***	1.102	0.1090***	1.115	0.0859**	1.090
Pseudo R^2	0.0371		0.0181		0.0192	
$N_{y=0}$	8,463		8,498		8,509	
$N_{y=1}$	81		46		35	
N	8,544		8,544		8,544	
Panel B : Medium volatility regime						
LOGIT(t)						
α_0	−5.2699***		−6.1134***		−6.5099***	
NT_t	0.7474***	2.112	0.5634***	1.757	0.7904***	2.204
$Abs\_T\_Imb_t$	−0.0205	0.980				
T_Imb_t			−0.8799***	0.415	1.0532***	2.867
RS_t	0.0726**	1.075	0.0737**	1.077	0.1018***	1.107
Pseudo R^2	0.0289		0.0142		0.0200	
$N_{y=0}$	8,392		8,434		8,434	
$N_{y=1}$	84		42		42	
N	8,476		8,476		8,476	
Panel C : High volatility regime						
LOGIT(t)						
α_0	−5.3773***		−6.1584***		−6.7951***	
NT_t	0.9444***	2.571	0.7874***	2.198	1.0700***	2.915
$Abs\_T\_Imb_t$	0.2477*	1.281				
T_Imb_t			−1.0219***	0.360	1.1117***	3.039
RS_t	0.0654***	1.068	0.0768***	1.080	0.0703***	1.073
Pseudo R^2	0.0279		0.0142		0.0195	
$N_{y=0}$	8,171		8,210		8,218	
$N_{y=1}$	86		47		39	
N	8,257		8,257		8,257	

penalized log likelihood. The variables are lagged by one period.The subsamples are before the bubble (before 24 May 2017), during the bubble (between 25 May 2017 and 28 January 2018), and after the bubble (after 28 January 2018). The odds ratio (OR) is the exponential of the coefficient estimate. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 11 reports the results of Equation (3). The dependent variable is the occurrence of an ER at time t and the independent variables include an intercept, the number of trades (NT), the absolute trade imbalance (Abs_T_Imb), the trade imbalance (T_Imb), the absolute log-return (R), and the relative spread (RS). All variables, excepting the intercept, are standardized. We estimate a LOGIT regression with Firth (1993)'s correction, with confidence intervals computed from the profile penalized log likelihood. The variables are lagged by one period.The subsamples are before the bubble (before 24 May 2017), during the bubble (between 25 May 2017 and 28 January 2018), and after the bubble (after 28 January 2018). The odds ratio (OR) is the exponential of the coefficient estimate. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Relationship between drivers

To complement the identification of the drivers of ERs, we further look at the relationships between the key four drivers used in the previous sections, i.e. absolute log-return, number of trades, trade imbalance, and relative spread. We use a Vector Autoregressive Approach (VAR). We deliberately do not include the ER_{99} variable in the VAR setting, since it is a binary response variable.[11] To avoid any non-stationarity issues, we take the first difference of these variables[12] and perform two unit root tests allowing for structural breaks, i.e. the Zivot and Andrews (2002) and Enders and Lee (2012) tests. They both point to stationarity for all the differentiated variables.[13] We estimate an interdependent system of four linear regression equations, one for each driver. The explanatory variables include the other three drivers on a contemporaneous basis as well as the lagged values of the four drivers, from 1 to N lags. We estimate the following system of equations with control variables up to 10 lags:

$$\begin{cases} \Delta R_t = \alpha_{10} + \alpha_{11}\Delta NT_t + \alpha_{12}\Delta Imb_t + \alpha_{13}\Delta RS_t \\ \quad + \sum_{i=1}^{N} (\beta_{1i}\Delta NT_{t-i} + \gamma_{1i}\Delta aImb_{t-i} \\ \quad + \omega_{1i}\Delta RS_{t-i} + \zeta_{1i}\Delta R_{t-i}) + \in_{1t} \\ \Delta N_t = \alpha_{20} + \alpha_{21}\Delta R_t + \alpha_{22}\Delta Imb_t + \alpha_{23}\Delta RS_t \\ \quad + \sum_{i=1}^{N} (\beta_{2i}\Delta NT_{t-i} + \gamma_{2i}\Delta Imb_{t-i} \\ \quad + \omega_{2i}\Delta RS_{t-i} + \zeta_{2i}\Delta R_{t-i}) + \in_{2t} \\ \Delta Imb_t = \alpha_{30} + \alpha_{31}\Delta NT_t + \alpha_{32}\Delta R_t + \alpha_{33}\Delta RS_t \\ \quad + \sum_{i=1}^{N} (\beta_{3i}\Delta NT_{t-i} + \gamma_{3i}\Delta Imb_{t-i} \\ \quad + \omega_{3i}\Delta RS_{t-i} + \zeta_{3i}\Delta R_{t-i}) + \in_{3t} \\ \Delta RS_t = \alpha_{40} + \alpha_{41}\Delta NT_t + \alpha_{42}\Delta R_t + \alpha_{43}\Delta Imb_t \\ \quad + \sum_{i=1}^{N} (\beta_{4i}\Delta NT_{t-i} + \gamma_{4i}\Delta Imb_{t-i} \\ \quad + \omega_{4i}\Delta RS_{t-i} + \zeta_{4i}\Delta R_{t-i} + \in_{4t} \end{cases} \tag{6}$$

In Figure 7, we report the results with one lag in the system of Equation (6).[14] We only report the contemporaneous (C) and/or the first lag (L) coefficient when it is statistically significant. The autocorrelation coefficient estimate of the first lag is negative for each variable, pointing to negative autocorrelation and short-term reversion dynamics in each driver. Second, there is a positive relationship between trades and returns and this relationship is bi-directional, both contemporaneously and with a lag: when the number of trades increases, both contemporaneous and subsequent absolute log-returns increase; the reverse is true as well. Third, there is a positive and bi-directional relationship between return and spread measures, with the exception of the lagged coefficient from spread to return which is not significant. Fourth, there is a positive and bi-directional relationship between trades and imbalance, with the exception of the lagged insignificant coefficient from trades to imbalance. Fifth, there is a positive and bi-directional relationship between return and imbalance, with the exception of the insignificant lagged coefficient from return to imbalance. Finally, there is only a significant negative lagged relationship from imbalance to spread.

To sum up, there are short-term reversion dynamics in each driver. All drivers are positively related to each other, from a contemporaneous perspective or not, with the exception of imbalance and spread which nearly exhibit zero interdependence. Overall, there seems to be more systematic interdependence on a contemporaneous basis than on a lag basis, given the magnitude of the coefficient estimates. The number of trades and the absolute log-returns exhibit the strongest links, in line with our previous findings.

Granger causality between drivers

To further investigate the dynamic relationship between the drivers of ERs, we run Granger causality tests which can be formalized as follows, using Equation (7) and (8) as the unconstrained and constrained models respectively:

[11] Dueker (2005) proposes a QUAL VAR methodology to forecast a binary variable with a VAR. However, El-Shagi and Von Schweinitz (2016, 293) argue that this methodology is *`inadvisable when the chain of causality matters'*.
[12] We use the Δ to indicate the first-difference of a variable.
[13] More detailed results are available upon request. The conclusions were identical with the standard Augmented Dicky-Fuller (ADF), Phillips-Perron (PP), and Kwiatkowski – Phillips–Schmidt – Shin (KPSS) tests.
[14] We have also replicated this analysis by sub-periods, i.e. before, during, and after the bubble and in low, medium, and high volatility regimes. The results are highly similar and the figures are available upon request.

Table 11. LOGIT - Bitfinex – BTCUSD - Contemporaneous model with volatility regimes (GMM).

	ER		ER DOWN		ER UP	
	Coeff.	OR	Coeff.	OR	Coeff.	OR
Panel A : Low volatility regime						
GMM-LOGIT(t)						
α_0	−5.5916***		−6.0400***		−6.6870***	
NT_t	0.9173***	2.502	0.7394***	2.095	0.8630***	2.370
$Abs\_T\_Imb_t$	−0.3593***	0.698				
T_Imb_t			−0.5812***	0.559	0.7958***	2.216
RS_t	0.0677***	1.070	0.0780***	1.081	0.0384	1.039
$N_{y=0}$	8,463		8,498		8,509	
$N_{y=1}$	81		46		35	
N	8,544		8,544		8,544	
Panel B : Medium volatility regime						
GMM-LOGIT(t)						
α_0	−5.2921***		−6.1527***		−6.5601***	
NT_t	0.7504***	2.118	0.5651***	1.760	0.7953***	2.215
$Abs\_T\_Imb_t$	−0.0270	0.973				
T_Imb_t			−0.8845***	0.413	1.0619***	2.892
RS_t	0.0543***	1.056	0.0501**	1.051	0.0652***	1.067
$N_{y=0}$	8,392		8,434		8,434	
$N_{y=1}$	84		42		42	
N	8,476		8,476		8,476	
Panel C : High volatility regime						
GMM-LOGIT(t)						
α_0	−5.4015***		−6.1893***		−6.8452***	
NT_t	0.9480***	2.581	0.7906***	2.205	1.0779***	2.939
$Abs\_T\_Imb_t$	0.2358**	1.266				
T_Imb_t			−1.0187***	0.361	1.1037***	3.015
RS_t	0.0570***	1.059	0.0647***	1.067	0.0462***	1.047
$N_{y=0}$	8,171		8,210		8,218	
$N_{y=1}$	86		47		39	
N	8,257		8,257		8,257	

$$Y_t = \alpha_0 + \sum_{i=1}^{N} \alpha_i Y_{t-i} + \sum_{i=1}^{N} \beta_i X_{t-i} + \in_t \quad (7)$$

$$Y_t = \alpha_0 + \sum_{i=1}^{N} \alpha_i Y_{t-i} + \mu_t \quad (8)$$

If the R^2 in the unconstrained equation is sufficiently higher than the corresponding R^2 in the constrained equation, then the null hypothesis of X *not Granger causing* Y is rejected with a p-value lower than 5%.[15] The optimal number of lags, N, is determined using the Akaike Information Criterion (AIC). The results are reported in Table 12.

Table 12 reports the results of the Granger causality tests between the four drivers, i.e. number of trades, absolute trading imbalance, absolute log return, and relative spread. All variables are differenced. We report the optimal number of lags (with a maximum of 10). This choice is based on the Akaike Information Criterion (AIC) and the p-value of the test. The reader should understand the table as follows: the variable in the first column Granger causes the variable mentioned in one of the other columns. These estimates have been produced using the R software.

We find that the number of trades Granger causes trading imbalance and absolute log returns, but the relationship is bivariate since the reverse is true as well. Interestingly,

[15]Causality in the sense of Granger does not necessarily mean that there is a strict chain of causality existing between two series. Granger causality is associated to the notion of time precedence, i.e. it tests whether one series precedes another series based on a number of lags and a time sequence. The nature of this precedence can be the direct consequence of a causal relationship, the result of pure luck, or a mixture of both effects. Granger causality tests cannot assess to which of these categories the highlighted relationship belongs.

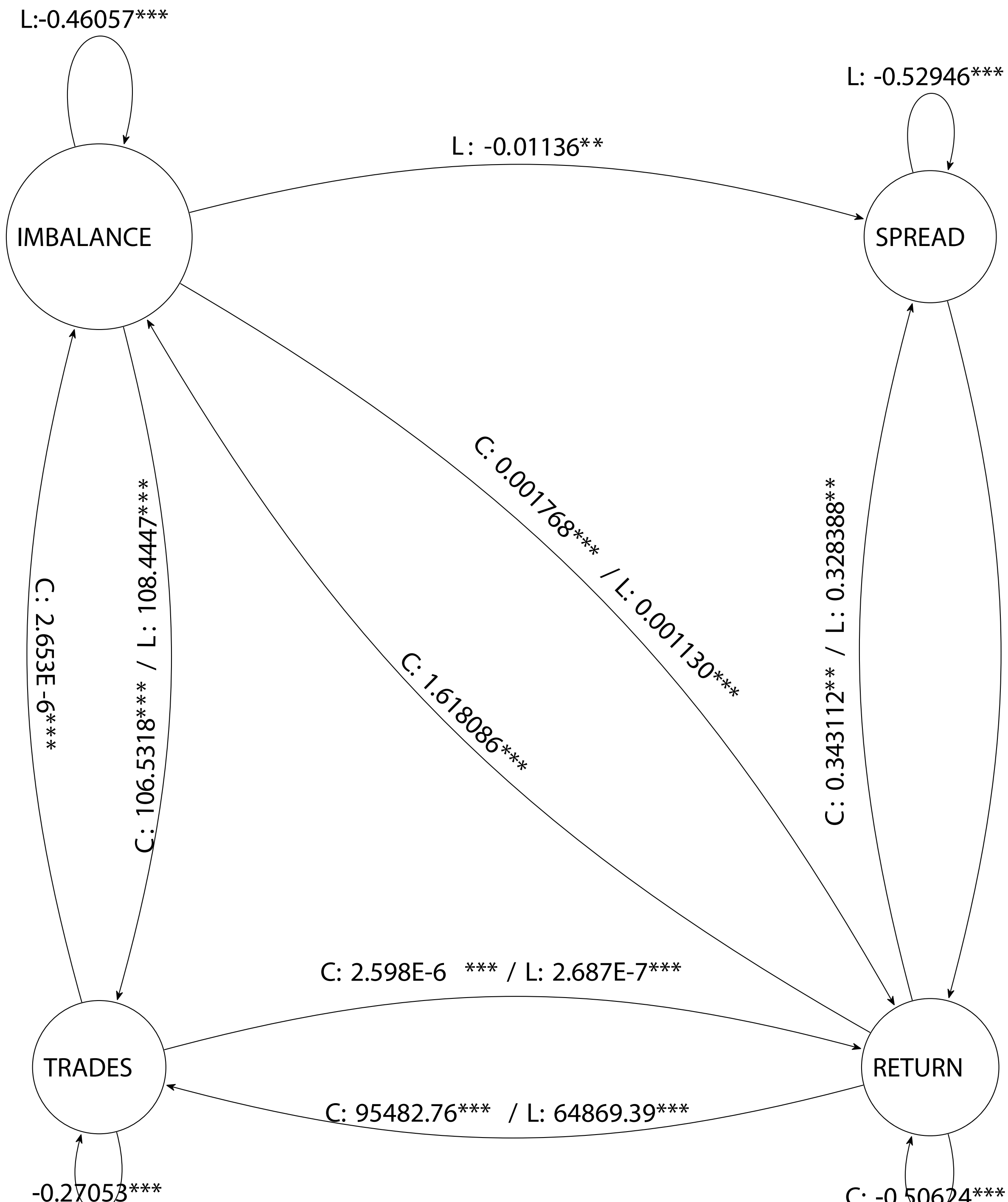

Figure 7. Relationship between drivers. This figure represents the relationships across the four drivers, i.E. absolute trading imbalance, spread, number of trades, and absolute log return. All variables are differencied. We report the contemporaneous effect (C) and the first-lag effect (L). *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 12. Granger causality.

Granger causes	ΔN_t	ΔImb_t	ΔR_t	ΔRS_t
ΔN_t		Optlag = 10 p-val < 0.001	Optlag = 4 p-val < 0.001	Optlag = 10 p-val = 0.76
ΔImb_t	Optlag = 10 p-val = 0.11		Optlag = 6 p-val = 0.42	Optlag = 6 p-val = 0.11
ΔR_t	Optlag = 10 p-val < 0.001	Optlag = 10 p-val < 0.001		Optlag = 10 p-val < 0.001
ΔRS_t	Optlag = 10 p-val < 0.001	Optlag = 10 p-val < 0.001	Optlag = 4 p-val < 0.001	

liquidity, as measured by the relative spread, is the only variable Granger causing the three other variables. Although the relative spread was ranked behind the number of trades in terms of its impact on ERs, this is the only ER driver which significantly explains the future variations in all the other ER drivers. The effect of liquidity is therefore both direct and indirect: liquidity directly affects the occurrence of ERs since the relative spread was a significant driver, but its indirect effect matters as well, since liquidity Granger causes the other factors. This is exactly the opposite conclusion for trade imbalance which never Granger causes the other drivers.

To alleviate concerns about the presence of different regimes, we run the Granger causality tests on the three sub-periods, i.e. before, during, and after the bubble. Most results remain consistent. There are a few noticeable differences. First, changes in the number of trades Granger causes changes in the relative spread in each sub-sample period, while it was not the case over the full sample period. This is additional evidence that the formation of the bubble may have significant influence on the interdependence dynamics between ER drivers. Second, imbalance also Granger causes the spread during the bubble period. Third, trading imbalance Granger causes absolute log return (only after the bubble) and the number of trades (both during and after the bubble). Finally, we also allow for the presence of structural breaks in the Granger causality test, and follow Gormus, Nazlioglu, and Soytas (2018) in augmenting the Toda – Yamamoto method with a Fourier approximation. The test clearly shows the dominance of the number of trades, which is the only driver which significantly explains the future variations of returns. This confirms the pivotal role that the number of trades plays in explaining price movements in cryptos, as first revealed in the previous Sections.[16]

Herding in volatile market conditions

Herding is the tendency of investors to blindly follow the market.[17] As such, ERs might be directly related to herding: When people herd in trading, price movements are exacerbated since the crowd pushes the price in the same direction. Even when traders in the crowd have no prior information about the market, they may still observe trading activity and volumes as they follow the herd. The herding issue in crypto markets has already received a lot of attention from the research community, e.g. Ballis and Drakos (2020), Blasco, Corredor, and Satrústegui (2022), Bouri, Gupta, and Roubaud (2019), Coskun, Lau, and Kahyaoglu (2020), da Gama Silva et al. (2019), King and Koutmos (2021), Mandaci and Cagli (2022), Raimundo Júnior et al. (2022), Ren and Lucey (2022), Rubbaniy et al. (2022), Stavroyiannis and Babalos (2019), and Vidal-Tomás, Ibáñez, and Farinós (2019). Seven of the above-listed papers use the Cross-Sectional Standard Deviation (CSSD) and/or the Cross Sectional Absolute Deviation (CSAD) of returns. These methodologies were respectively proposed

[16]Detailed results are available upon request. To complement this analysis, Figure 4 presents the results of the impulse response functions in the online Appendix. We use the same four variables, i.e. number of trades, relative spread, imbalance, and absolute log returns. A shock in the absolute log return tend to increase the spread, which is in accordance with our results. We also observe that a shock in the spread increases the absolute log return, suggesting a bivariate relationship between the variables. Finally, a shock in the number of trades also positively impacts the absolute log return. Again, the reaction of the dynamic system in response to external changes is sound as there is no evidence of market dysfunction.

[17]We thank an anonymous reviewer for suggesting to incorporate this analysis in the paper.

Table 13. CSSD and CSAD.

	CSSD		CSAD	CSAD - $R_{mt} > 0$	CSAD - $R_{mt} < 0$
α_0	0.009***	α_0	0.0048***	0.0049***	0.0047***
Down	0.0095***	$\lvert R_{mt} \rvert$	0.2156***	0.2700***	0.1660***
Up	0.0092***	R^2_{mt}	−0.6629**	−1.0645**	−0.2381
R^2	2.50%		11.70%	12.60%	11.90%
N	8,102		8,102	4,261	3,840

by Christie and Huang (1995) and Chang, Cheng, and Khorana (2000), and seem to be the most widespread techniques.

The CSSD is computed as follows:

$$CSSD_t = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(R_{it} - R_{mt})^2} \qquad (9)$$

where R_{it} and R_{mt} are respectively the return for crypto i and the market at time t. Given that there are no 'market' for cryptocurrencies, these authors either take an equally-weighted average of the crypto's returns or a market-cap-weighted average. Consistent with Vidal-Tomás, Ibáñez, and Farinós (2019), we use an equally-weighted average.

The CSSD is then regressed on two dummy variables capturing down and up market movements respectively, as follows.

$$CSSD_t = \alpha_0 + \beta_1 Down_t + \beta_2 Up_t + \in_t \qquad (10)$$

More specifically, $Down_t$ (Up_t) is equal to 1 if the market return is below (above) the 1st (99th) percentile. These thresholds correspond to −4.27% and 3.78% and are consistent with the thresholds we apply in our main analysis, although, they apply to the market return while it was applied to each crypto separately in our analysis.

The CSAD is computed as follows:

$$CSAD_t = \frac{1}{N}\sum_{i=1}^{N}|R_{it} - R_{mt}| \qquad (11)$$

This measure is then regressed on the absolute market return (

$$CSAD_t = \alpha_0 + \beta_1|R_{mt}| + \beta_2 R^2_{mt} + \in_t \qquad (12)$$

Following Ballis and Drakos (2020), we also estimate Equation (12) only when $R_{mt} > 0$ and $R_{mt} < 0$ to allow for an asymmetric effect.

As indicated in Table 13, the CSSD results do not support the hypothesis that crypto returns are

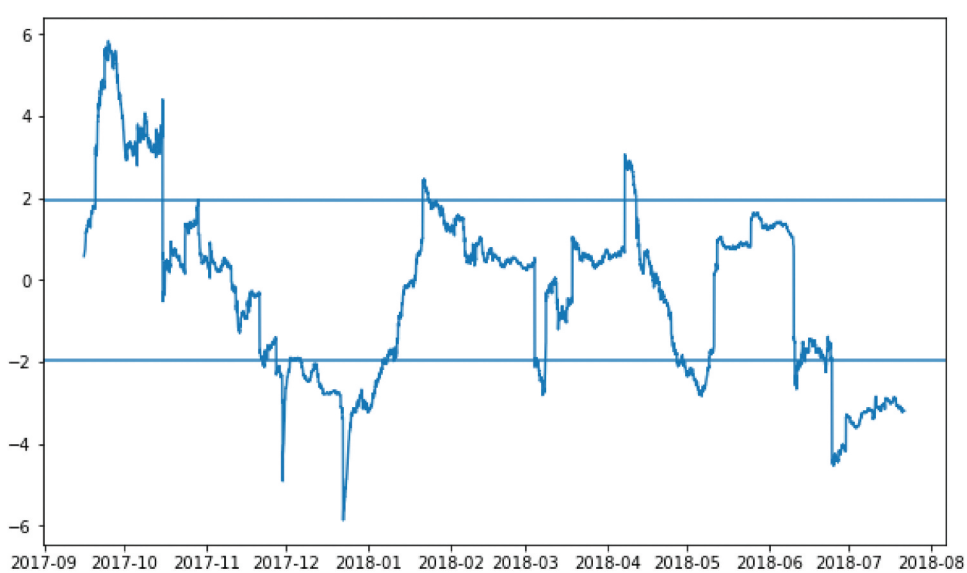

Figure 8. Time-Varying t-value. This figure represents the evolution of the t-stat from the β_2 coefficient of Equation (12). The two horizontal lines are respectively the 1.96 and −1.96 significance level thresholds.

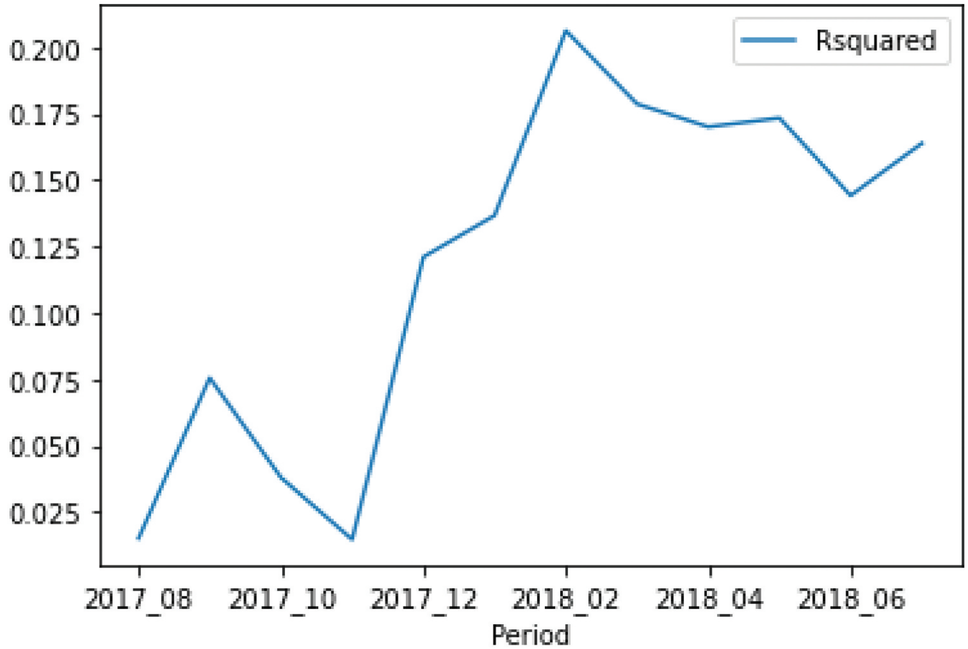

Figure 9. Average R^2 of Equation (13). This figure represents the evolution of the R^2 for Equation (13) which has been estimated on a monthly basis.

related to herding among market participants in extreme market circumstances. The beta coefficients in column 1 are both positive and significant, pointing to increased individuality in forming trading decisions when market returns are extreme. For evidence of herding, both coefficients should be *negative* and significant. There is nevertheless evidence of herding when we instead rely on the CSAD indicator (column 2). The parameter estimate of R^2_{mt} is negative and significant, indicating that as volatility increases the *CSAD* is disproportionately reduced compared to what is expected by rational asset pricing. When up and down market days are separated in two samples, herding is only detected in up markets (column 3).

Table 13 - column (1) reports the results of Equation (10). The dependent variable is the CSSD and the independent variables are two dummy variables, *Down* and *Up* which are equal to 1 if the market return is below (above) the 1st (99th) percentile. Column (2) reports the results of Equation (12). The dependent variable is the CSAD and the independent variables are the absolute market return,

To complement these findings, we conduct two additional tests. We first follow Bouri, Gupta, and Roubaud (2019) and estimate Equation (13) on a rolling-window basis, using 720 observations (which corresponds to approximately one month of data). We represent the evolution of the t-value with respect to the coefficient of interest, β_2 in Figure 8.

This figure confirms our previous findings. Herding increases during rallies, when market returns are positive. This is indicated by the β_2 parameter estimates becoming negative and reaching its lower level in the end of 2017, during the first Bitcoin bubble. Herding also increased in April and July 2018 when smaller bubbles occurred. Sell-offs, on the contrary, do not seem to be associated with herding.

Finally, we follow Chung and Kim (2017)'s methodology and estimate the following regression:

$$T\_Imb_t = \alpha_0 + \beta_1 \overline{T\_Imb}_{Mt} + \in_t \qquad (13)$$

where T_Imb_t ($\overline{T\_Imb}_{Mt}$) is the trading imbalance in the crypto (resp. the market) at time t. Contrary to Chung and Kim (2017)'s regression, we do not include an industry imbalance, as it is not relevant in our setting.

This equation is estimated separately for each crypto and for each month. We start our estimation in August 2017 to cover at least seven cryptos (See Table 2). When computing

Table 14. Pvalue counters.

	LOGIT - Lagged Model							LOGIT - Contemporaneous Model						
	Negative Est.				Positive Est.			Negative Est.				Positive Est.		
	1%	5%	10%	NS	10%	5%	1%	1%	5%	10%	NS	10%	5%	1%
ER														
NT		1		4	2	1	17				1			24
ABS_T_IMB	9	3	1	9	1		2	3	2	1	12		2	5
RET				4		2	19							
RS				4		3	18				1			24
ER down														
NT	1	1		3	1	4	15					1	1	23
T_IMB	1	2		20			2	17	1	1	4		1	1
RET				7		2	16							
RS				3	1	10	11				3			22
ER up														
NT	1			6	1	2	15				2			23
T_IMB	1	1	1	19		1	2				7		2	16
RET				5	1	1	18							
RS				5		3	17				2	2	4	17

$\overline{T\_Imb}_{Mt}$, we use equally-weighted averages while excluding trading imbalance in the crypto used as the dependent variable in the regression. Finally, we compute the mean R^2 of these regression by month. We plot the evolution of this average in Figure 9.

These results confirm our previous findings based on the CSAD indicator, highlighting a sharp increase in herding behaviour during the late 2017 bubble, when returns are positive and large in magnitude.

IV. Robustness tests

We conduct three robustness checks. First, we opt for a *p*-value counter methodology, i.e. we estimate Equation (3) on each platform/cryptocurrency combination. Second, we estimate Equation (3) separately for permanent and transitory ERs. Third, we estimate a RELOGIT specification, to take the scarcity of events into account.

P-value counters

We estimate Equation (3) on each 'platform/cryptocurrency' combination, leading to 25 estimations. We report in Table 14 the number of positive and negative coefficients, as well as their significance level for the contemporaneous and lagged LOGIT models.

When all ERs are included in the sample (irrespective of their direction) to estimate in the lagged model, the coefficients of the number of trades, absolute log returns, and relative spread are positive and significant in the lagged model in 20, 21, and 21 cases out of 25 respectively. As suggested before, results are more noisy for the absolute trade imbalance: its coefficient is negative and significant in 13 cases, 10 times non significant and positive and significant in 3 cases only.

In the contemporaneous LOGIT model specification, the results are even clearer for the number of trades and the relative spread, with 24 out of 25 positive and significant parameter estimates. The absolute log return is removed to avoid perfect collinearity. The absolute trade imbalance results are again mixed.

When we move to subsamples of upward and downward ERs, the *p*-value counters are highly similar, confirming that our results are not affected by the direction of ERs. Signed trade imbalance is also much less noisy than the absolute measure. We can therefore be confident about the stability of our estimates across cryptos and platforms.

Table 14 reports the number of positive or negative coefficient that we obtain when we estimate Equation (3) on each combination cryptocurrency / platform. The dependent variable is the occurrence of an ER at time *t* and the independent variables include an intercept, the number of trades (*NT*), the absolute trade imbalance (*Abs_T_Imb*), the absolute log-return (*R*), and the relative spread (*RS*). All variables, excepting the intercept, are lagged by one period and are standardized. N. S. stands for 'not significant'.

Table 15. RELOGIT - Lagged Model including all ERS.

	Coeff.	OR
α_0	−5.1266***	
NT_{t-1}	0.4249***	1.5295
$Abs\_T\_Imb_{t-1}$	−0.5164***	0.5967
R_{t-1}	0.1954***	1.2158
RS_{t-1}	0.0274***	1.0278
N	25,174	
$N_{y=0}$	24,924	99.01%
$N_{y=1}$	250	0.99%
Pseudo R^2	1.59%	

Transitory and permanent ERs

We also estimate Equation (3) while distinguishing transitory from permanent ERs to check whether the drivers differ in that respect. Following Brogaard et al. (2018), transitory (permanent) ERs are those who recover by more than 2/3 (less than 1/3) in 24 hours. We obtain 148 permanent and 80 transitory ERs. All previously highlighted results are robust to the transitory-permanent nature of ERs.[18]

Relogit

ERs are by definition rare events and the use of a LOGIT regression may introduce an estimation bias due to the disequilibrium between the number of events and non-events. All the LOGIT models presented in the previous Sections were therefore estimated by using Firth (1993)'s penalized maximum likelihood estimation. As a robustness check, we extend our baseline LOGIT model to a Rare Event LOGIT (RELOGIT). This method is discussed by King and Zeng (2001a, 2001b) and by Cook, Hays, and Franzese (2020). We estimate Equation (3) by using the RELOGIT specification and report the results in Table 15.

Table 15 reports results of Equation (3) when we apply the RELOGIT correction. The dependent variable is the occurrence of an ER at time t and the independent variables include an intercept, the number of trades (NT), the absolute trade imbalance (Abs_T_Imb), the absolute log-return (R), and the relative spread (RS). All variables, excepting the intercept, are lagged by one period and are standardized. N is the number of observations, $N_{y=0}$ ($N_{y=1}$) is the number of non-events (events). We also report the R-squared. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. These estimates have been produced using the R software.

It is worth noting that the scarcity of the ERs was already controlled for in some way in our previous estimations, in the sense that the return threshold depends on how rare the events are. The proportion of rare events in the data, $\bar{y}$, almost matches the theoretical proportion of these events, τ. The correction in the RELOGIT should therefore bring little changes in comparison with the results of Table 4. We indeed find very consistent results in terms of both the magnitude of the coefficient estimates and the level of significance.[19]

V. Conclusion

The recent advent of cryptos has drawn the attention of regulators, investors, and central banks for good and bad reasons. While the crypto market is undoubtedly one of the most innovative markets in the world, it exhibits an extremely high level of volatility as well. Unsurprisingly, the unique features of cryptos have raise eyebrows, in particular among central banks which have responded by investigating the introduction of their own central bank digital currencies (CBDCs). These new currencies will not share the same properties as traditional cryptos, namely decentralization and the use of a distributed ledger. CBDCs will be much closer to the concept of currency than cryptos, but it remains to be tested how different the price and volatility dynamics of these CBDCs will be.

In this paper, we investigate how trading and liquidity dynamics evolve in stressful conditions in the crypto market. Our study uses very detailed trade and order book data on the 8 most widespread cryptos traded in the 16 most active trading platforms over the period ranging from May 2015 to July 2018. We show that ERs in cryptos are accompanied by a sharp increase in trading activity, whether measured by the number of trades,

[18]The empirical results are available in the online Appendix in Tables A1 to A3.

[19]The results of the contemporaneous model as well as the separate outcomes for downward and upward ERs are not presented here but available upon request. These results are also highly consistent with the ones presented in the previous Sections.

quantities traded, or monetary volume. Regarding ex ante (or order-book-based) liquidity, both the quoted and relative spreads increase significantly, pointing to a higher cost of immediacy during these extreme events. Liquidity takers become aggressive and sell at the bid price by submitting market or marketable limit orders in order to ensure that their position is closed, or that liquidity providers vanish from the market at that moment because they fear that they will trade with a better informed trader. Depth also significantly increases, showing that there are more quantities outstanding on both sides of the order book when an ER occurs. When we zoom in on ERs at the 5-minute interval to better characterize the dynamics of liquidity and trading over time, the rise in the number of trades seems to be a leading indicator.

Using the logistic regression framework adapted to rare events, we show that there exist early warning signals of ERs, both in the order book and in trading activity. The number of trades is a particularly robust driver to explain the occurrence of ERs, followed by the relative spread. When we look at the contemporaneous relationship, we identify the same usual suspects than for traditional markets, including the signed trading imbalance. This holds true whether we focus on the Bitcoin on the most active Bitfinex platform, extend the analysis to a multi-platform and a multi-cryptocurrency analysis, condition it on the identification of the bubble period or on volatility regimes, and distinguish permanent from transitory ERs. The use of the RELOGIT framework adapted for rare events does not modify the conclusions.

We also analyse the relationship between the key ER drivers by running both VAR models and Granger causality tests. Most variables are positively interconnected, whether on a contemporaneous or a lagged basis, albeit stronger in the first case. Each driver displays short-term reversion dynamics. The number of trades and the absolute log-returns exhibit the strongest links, in line with our regression analysis. This is confirmed when we run a Granger causality test allowing for structural breaks: the number of trades is the only driver which significantly Granger causes the returns, confirming the pivotal role that the number of trades plays in explaining price movements in cryptos. Our results also seem to indicate that herding occurs when returns are positive and large in magnitude, in particular during the late 2017 sharp rise in prices.

There is overall no evidence that liquidity and trading dynamics around ERs on cryptos markets are orthogonal to what more traditional markets experience in stressful market conditions.

Acknowledgement

We thank the Editor, two anonymous reviewers, Thiago Winkler Alves, Neharika Sobti, Kotaro Miwa, Catherine D'Hondt, Yue Zhang, Ian W. Marsh, Rudy De Winne, Gunther Wuyts, as well as participants at the Australasian Finance and Banking Conference (Sydney, AUS), New Zealand Finance Meeting (Auckland, NZ), and ISCEF (Paris, FR). Christophe Desagre gratefully acknowledges financial support from the LFIN research center and the FNRS. Any errors are the full responsibility of the authors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Ali, R., J. Barrdear, R. Clews, and J. Southgate. 2014. "The Economics of Digital Currencies." *Quarterly Bulletin* Q3: 276–286.

Balcilar, M., E. Bouri, R. Gupta, and D. Roubaud. 2017. "Can Volume Predict Bitcoin Returns and Volatility? a Quantiles-Based Approach." *Economic Modelling* 64: 74–81. doi:10.1016/j.econmod.2017.03.019.

Ballis, A., and K. Drakos. 2020. "Testing for Herding in the Cryptocurrency Market." *Finance Research Letters* 33: 101210. doi:10.1016/j.frl.2019.06.008.

Blasco, N., P. Corredor, and N. Satrústegui. 2022. "The Witching Week of Herding on Bitcoin Exchanges." *Financial Innovation* 8 (1): 1–18. doi:10.1186/s40854-021-00323-4.

Blau, B. M. 2017. "Price Dynamics and Speculative Trading in Bitcoin." *Research in International Business and Finance* 41: 493–499. doi:10.1016/j.ribaf.2017.05.010.

Bouri, E., R. Gupta, and D. Roubaud. 2019. "Herding Behaviour in Cryptocurrencies." *Finance Research Letters* 29: 216–221. doi:10.1016/j.frl.2018.07.008.

Brogaard, J., A. Carrion, T. Moyaert, R. Riordan, A. Shkilko, and K. Sokolov. 2018. "High Frequency Trading and Extreme Price Movements." *Journal of Financial Economics* 128 (2): 253–265. doi:10.1016/j.jfineco.2018.02.002.

Broto, C., and M. Lamas. 2020. “Is Market Liquidity Less Resilient After the Financial Crisis? Evidence for US Treasuries.” *Economic Modelling* 93: 217–229. doi:10.1016/j.econmod.2020.08.001.

Chaim, P., and M. P. Laurini. 2018. “Volatility and Return Jumps in Bitcoin.” *Economics Letters* 173: 158–163. doi:10.1016/j.econlet.2018.10.011.

Chaim, P., and M. P. Laurini. 2019. “Is Bitcoin a Bubble?” *Physica A* 517: 222–232. doi:10.1016/j.physa.2018.11.031.

Chamberlain, G. 1980. “Analysis of Covariance with Qualitative Data.” *The Review of Economic Studies* 47 (1): 225–238. doi:10.2307/2297110.

Chang, E. C., J. W. Cheng, and A. Khorana. 2000. “An Examination of Herd Behavior in Equity Markets: An International Perspective.” *Journal of Banking & Finance* 24 (10): 1651–1679. doi:10.1016/S0378-4266(99)00096-5.

Chen, C.-Y.-H., and C. M. Hafner. 2019. “Sentiment-Induced Bubbles in the Cryptocurrency Market.” *Journal of Risk and Financial Management* 12 (2): 53. doi:10.3390/jrfm12020053.

Chevapatrakul, T., and D. V. Mascia. 2019. “Detecting Overreaction in the Bitcoin Market: A Quantile Autoregression Approach.” *Finance Research Letters* 30: 371–377. doi:10.1016/j.frl.2018.11.004.

Christie, W. G., and R. D. Huang. 1995. “Following the Pied Piper: Do Individual Returns Herd Around the Market?” *Financial Analysts Journal* 51 (4): 31–37. doi:10.2469/faj.v51.n4.1918.

Chung, P. Y., and T. S. Kim. 2017. “Extreme Returns and Herding of Trade Imbalances.” *Review of Finance* 21 (6): 2379–2399. doi:10.1093/rof/rfx004.

Cook, S. J., J. C. Hays, and R. J. Franzese. 2020. “Fixed Effects in Rare Events Data: A Penalized Maximum Likelihood Solution.” *Political Science Research and Methods* 8 (1): 92–105. doi:10.1017/psrm.2018.40.

Corbet, S., and P. Katsiampa. 2020. “Asymmetric Mean Reversion of Bitcoin Price Returns.” *International Review of Financial Analysis* 71: 101267. doi:10.1016/j.irfa.2018.10.004.

Coskun, E. A., C. K. M. Lau, and H. Kahyaoglu. 2020. “Uncertainty and Herding Behavior: Evidence from Cryptocurrencies.” *Research in International Business and Finance* 54: 101284. doi:10.1016/j.ribaf.2020.101284.

Coupé, T. 2005. “Bias in Conditional and Unconditional Fixed Effects Logit Estimation: A Correction.” *Political Analysis* 13 (3): 292–295. doi:10.1093/pan/mpi019.

da Gama Silva, P. V. J., M. C. Klotzle, A. C. F. Pinto, and L. L. Gomes. 2019. “Herding Behavior and Contagion in the Cryptocurrency Market.” *Journal of Behavioral andExperimental Finance* 22: 41–50.

Donier, J., and J.-P. Bouchaud. 2015. “Why Do Markets Crash? Bitcoin Data Offers Unprecedented Insights.” *PloS One* 10 (10): e0139356. doi:10.1371/journal.pone.0139356.

Dueker, M. 2005. “Dynamic Forecasts of Qualitative Variables: A Qual VAR Model of US Recessions.” *Journal of Business & Economic Statistics* 23 (1): 96–104. doi:10.1198/073500104000000613.

Easley, D., M. O’Hara, and S. Basu. 2019. “From Mining to Markets: The Evolution of Bitcoin Transaction Fees.” *Journal of Financial Economics* 134: 91–109. doi:10.1016/j.jfineco.2019.03.004.

El-Shagi, M., and G. Von Schweinitz. 2016. “Qual VAR Revisited: Good Forecast, Bad Story.” *Journal of Applied Economics* 19 (2): 293–321. doi:10.1016/S1514-0326(16)30012-5.

Enders, W., and J. Lee. 2012. “A Unit Root Test Using a Fourier Series to Approximate Smooth Breaks.” *Oxford Bulletin of Economics and Statistics* 74 (4): 574–599. doi:10.1111/j.1468-0084.2011.00662.x.

Firth, D. 1993. “Bias Reduction of Maximum Likelihood Estimates.” *Biometrika* 80 (1): 27–38. doi:10.1093/biomet/80.1.27.

Fousekis, P., and D. Tzaferi. 2021. “Returns and Volume: Frequency Connectedness in Cryptocurrency Markets.” *Economic Modelling* 95: 13–20. doi:10.1016/j.econmod.2020.11.013.

Gormus, A., S. Nazlioglu, and U. Soytas. 2018. “High-Yield Bond and Energy Markets.” *Energy Economics* 69 (C): 101–110. doi:10.1016/j.eneco.2017.10.037.

Greene, W. 2004. “The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects.” *The Econometrics Journal* 7 (1): 98–119. doi:10.1111/j.1368-423X.2004.00123.x.

Hasbrouck, J. 1995. “One Security, Many Markets: Determining the Contributions to Price Discovery.” *The Journal of Finance* 50 (4): 1175–1199. doi:10.1111/j.1540-6261.1995.tb04054.x.

Heinze, G. 2006. “A Comparative Investigation of Methods for Logistic Regression with Separated or Nearly Separated Data.” *Statistics in Medicine* 25 (24): 4216–4226. doi:10.1002/sim.2687.

Heinze, G., and M. Schemper. 2002. “A Solution to the Problem of Separation in Logistic Regression.” *Statistics in Medicine* 21 (16): 2409–2419. doi:10.1002/sim.1047.

Ji, Q., E. Bouri, C. K. M. Lau, and D. Roubaud. 2019. “Dynamic Connectedness and Integration in Cryptocurrency Markets.” *International Review of Financial Analysis* 63: 257–272. doi:10.1016/j.irfa.2018.12.002.

Ji, J., D. Wang, D. Xu, and C. Xu. 2020. “Combining a Self-Exciting Point Process with the Truncated Generalized Pareto Distribution: An Extreme Risk Analysis Under Price Limits.” *Journal of Empirical Finance* 57: 52–70. doi:10.1016/j.jempfin.2020.03.003.

Katz, E. 2001. “Bias in Conditional and Unconditional Fixed Effects Logit Estimation.” *Political Analysis* 9: 379–384. doi:10.1093/oxfordjournals.pan.a004876.

King, T., and D. Koutmos. 2021. “Herding and Feedback Trading in Cryptocurrency Markets.” *Annals of Operations Research* 300 (1): 79–96.

King, G., and L. Zeng. 2001a. “Explaining Rare Events in International Relations.” *International Organization* 55 (3): 693–715. doi:10.1162/00208180152507597.

King, G., and L. Zeng. 2001b. “Logistic Regression in Rare Events Data.” *Political Analysis* 9 (2): 137–163. doi:10.1093/oxfordjournals.pan.a004868.

Lee, S. S., and P. A. Mykland. 2012. “Jumps in Equilibrium Prices and Market Microstructure Noise.” *Journal of Econometrics* 168 (2): 396–406. doi:10.1016/j.jeconom.2012.03.001.

Lee, C., and M. J. Ready. 1991. “Inferring Trade Direction from Intraday Data.” *The Journal of Finance* 46 (2): 733–746. doi:10.1111/j.1540-6261.1991.tb02683.x.

Liu, J., I. W. Marsh, P. Mazza, and M. Petitjean. 2019. ”Factor Structure in Cryptocurrency Returns and Volatility.” *Available at SSRN 3389152*. doi:10.2139/ssrn.3389152

Makarov, I., and A. Schoar. 2020. “Trading and Arbitrage in Cryptocurrency Markets.” *Journal of Financial Economics* 135: 293–319. doi:10.1016/j.jfineco.2019.07.001.

Mandaci, P. E., and E. C. Cagli. 2022. “Herding Intensity and Volatility in Cryptocurrency Markets During the Covid-19.” *Finance Research Letters* 46: 102382. doi:10.1016/j.frl.2021.102382.

Mazza, P. (2020). Controlling for Event Disequilibrium and Fixed Effects in a Logistic Regression Framework: A Review of the Alternatives. *working paper*.

McFadden, D. 1973. “Conditional Logit Analysis of Qualitative Choice Behavior.” In *Frontiers in Economics*, edited by P. Z, 105–142. New York: Academic Press.

McLeay, M., A. Radia, and R. Thomas. 2014. “Money Creation in the Modern Economy.” *Bank of England Quarterly Bulletin* Q1: 14–27.

Neyman, J., E. L. Scott, et al. 1948. ”Consistent Estimates Based on Partially Consistent Observations.” *Econometrica* 16 (1): 1–32. doi:10.2307/1914288.

Phillips, P. C., Y. Wu, and J. Yu. 2011. “Explosive Behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values?” *International Economic Review* 52 (1): 201–226. doi:10.1111/j.1468-2354.2010.00625.x.

Phillips, P. C., and J. Yu. 2011. “Dating the Timeline of Financial Bubbles During the Subprime Crisis.” *Quantitative Economics* 2 (3): 455–491. doi:10.3982/QE82.

Raimundo Júnior, G. D. S., R. B. Palazzi, R. D. S. Tavares, and M. C. Klotzle. 2022. “Market Stress and Herding: A New Approach to the Cryptocurrency Market.” *Journal of Behavioral Finance* 23 (1): 43–57. doi:10.1080/15427560.2020.1821688.

Ren, B., and B. Lucey. 2022. “Do Clean and Dirty Cryptocurrency Markets Herd Differently?” *Finance Research Letters* 47: 102795. doi:10.1016/j.frl.2022.102795.

Rubbaniy, G., K. Tee, P. Iren, and S. Abdennadher. 2022. “Investors’ Mood and Herd Investing: A Quantile-On-Quantile Regression Explanation from Crypto Market.” *Finance Research Letters* 47: 102585. doi:10.1016/j.frl.2021.102585.

Scaillet, O., A. Treccani, and C. Trevisan. 2020. “High-Frequency Jump Analysis of the Bitcoin Market.” *Journal of Financial Econometrics* 18 (2): 209–232.

Stavroyiannis, S., and V. Babalos. 2019. “Herding Behavior in Cryptocurrencies Revisited: Novel Evidence from a Tvp Model.” *Journal of Behavioral and Experimental Finance* 22: 57–63. doi:10.1016/j.jbef.2019.02.007.

Thies, S., and P. Molnár. 2018. “Bayesian Change Point Analysis of Bitcoin Returns.” *Finance Research Letters* 27: 223–227. doi:10.1016/j.frl.2018.03.018.

Vidal-Tomás, D., A. M. Ibáñez, and J. E. Farinós. 2019. “Herding in the Cryptocurrency Market: CSSD and CSAD Approaches.” *Finance Research Letters* 30: 181–186. doi:10.1016/j.frl.2018.09.008.

Weber, P., and B. Rosenow. 2006. “Large Stock Price Changes: Volume or Liquidity?” *Quantitative Finance* 6 (1): 7–14. doi:10.1080/14697680500168008.

Zivot, E., and D. W. K. Andrews. 2002. “Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis.” *Journal of Business & Economic Statistics* 20 (1): 25–44. doi:10.1198/073500102753410372.