



Order flow analysis of cryptocurrency markets

Eduard Silantyev^{1,2,3}

Received: 4 October 2018 / Accepted: 18 March 2019 / Published online: 29 March 2019
© Springer Nature Switzerland AG 2019

Abstract

Order flow analysis studies the impact of individual order book events on resulting price change. Using data acquired from BitMex, the largest cryptocurrency exchange by traded volume, the study conducts an in-depth analysis on the trade and quote data of the XBTUSD perpetual contract. The study demonstrates that the trade flow imbalance is better at explaining contemporaneous price changes than the aggregate order flow imbalance. Overall, the contemporaneous price change exhibits a strong linear relationship with the order flow imbalance over large enough time intervals. Lack of depth and low update arrival rates in cryptocurrency markets are found to be the main differentiators between the nascent asset class market microstructure and that of the established markets.

Keywords Market microstructure · Cryptocurrency · Electronic trading · Market impact · Order flow

1 Introduction

Cryptocurrency is a nascent asset class. It was first conceptualised in the seminal Bitcoin whitepaper by anonymous programmer Satoshi Nakamoto (Antonopoulos 2017). Bitcoin was the first digital currency that achieved exchange of value without the need of a third party. Bitcoin network timestamps the transactions by hashing them, thus creating a hash-based proof of work chain of transactions that cannot be undone without redoing all the work done prior. It may help to think of Bitcoin and other cryptocurrency structures as immutable databases or ledgers of transactions, that cannot be changed because of so-called network agents—the miners. Miners solve computationally hard problems to arrive to common consensus about what

✉ Eduard Silantyev
ed.silantyev@thalesians.com

¹ Thalesians Ltd., London, UK

² BNP Paribas, Paris, France

³ London Institute of Banking and Finance, London, UK

chain of events (or transactions) took place over a defined number of events. Only the copy of the chain of events that is agreed upon by the majority of the network will be considered the true one. Bitcoin was the first proof of concept of the blockchain technology. Blockchain—the technology used for verifying and recording transactions and is the core of Bitcoin—is seen as having the potential to reshape the global financial system and possibly other industries. As such, blockchain removes the need for trust between the parties exchanging value or utility, protecting from common pitfalls such as double-spending, fraud and transaction reversibility. Nowadays, there are a few thousand different currencies/projects that employ blockchain technology. As public interest in cryptocurrencies grew, first exchanges between fiat currencies and cryptocurrencies started to take place.

Cryptocurrency first began trading hand-to-hand, usually via forum negotiations; the first ever documented transaction was a purchase of two Papa Johns pizzas for the region of 10,000 Bitcoins (Bitcointalk 2010). As growth progressed and the asset class became more popular, first exchanges started to facilitate fiat–Bitcoin trading. First such venues were Mt.Gox, BitStamp, BTC China and BTC-e (Halaburda and Gandal 2017). Technological inferiority of such “handmade” exchanges meant that not only were they susceptible to hacks and compromises, but also that trading of new digital assets was and still is fragmented and inefficient. As of mid-2013, 45% of exchanges have been shutdown because of hacks and owner-driven compromises (Steadman 2013). Bitcoin price appreciation reached its initial peak in late 2013/early 2014 when Mt.Gox, then the biggest exchange, shut down, reporting an attack that compromised roughly 850,000 Bitcoins, majority of which belonged to the customers of the exchange (Halaburda and Gandal 2017). Consequently, prices of cryptocurrency assets entered a period of downtrend, but the number of users of cryptocurrencies actually increased over the period. As of late 2017 and early 2018, CME and CBOE have listed cash-settled future contracts on Bitcoin/Dollar, which has made inflows of institutional capital into the asset class possible.

Market dynamics in cryptocurrency markets tend to be characterised by high volatility, thin liquidity and extreme sentiment swings (Chan 2017). Such market dynamics are produced by a number of drivers. First of all, there is no central limit order book or order routing between cryptocurrency venues, unlike for example in U.S. Equities. 90% of trading volume is fragmented across a dozen of exchanges, with notable price discrepancies (Verhage 2018). Another related driver of market dynamics is technological inferiority of cryptocurrency exchanges. Periods of high volume have seen many exchanges’ matching engines dysfunction at once, brightest example being December 2017 when value of Bitcoin crossed \$15,000 (Russo 2018). Another side effect of technological inferiority is the security breaches that happen in the form of hacking attacks and other compromises of private keys (Antonopoulos 2017). Such events usually result in market-wide panic and trigger high levels of volatility and thin liquidity.

In turn, volatility attracts a wide array of day-traders that feast off the price swings that are much harder to come by in mature markets. If we add the effects of leverage that some exchanges offer, the traders’ activity can be argued to amplify the price swings even further. In addition, there is no generally accepted way of valuing cryptocurrency. One could equate the intrinsic value of Bitcoin to the cost of

computational power that goes into solving the block's hash divided by the block reward. On the other hand, one can attribute value to adoption statistics such as network hash rate and number of unique wallets. However, as such, there is no agreement on fundamental value criteria. As a result, technology, regulation, sentiment, market participants and lack of fundamental value consensus drive the over-arching market macrostructure of cryptocurrency (Hileman and Rauchs 2017).

Shift in public opinion about cryptocurrency, changing market microstructure landscape, diversity of available data and gaps in current research merit a thorough analysis of order flow price impact in cryptocurrency markets. The idea of a decentralised economy is gaining traction and that demands better understanding of underlying dynamics of the currency (Zheng et al. 2016). To facilitate smooth adoption of cryptocurrency into everyday lives, market stability is essential. The study will examine the order flow impact on price, a fundamental characteristic of market microstructure of any asset. The research will likely benefit fields of optimal liquidity provision, optimal execution, and advancement of electronic trading in realm of cryptocurrency markets.

2 Research objectives and questions

The research objectives of this study are motivated by the gaps in current research, which will be evident in the next section. The study aims to be the primer on the cryptocurrency market microstructure, thus becoming a point of future reference for subsequent research in the area. The main question that this study addresses is as follows:

To what extent does order flow imbalance impact price change in cryptocurrency markets?

The specific definitions of order flow imbalance and price change follow in Sect. 4. The preliminary research yields the following hypotheses:

1. Order flow imbalance OFI_k has a positive linear relationship with contemporaneous mid-price change ΔMP_k , i.e. the price impact coefficient β_{OFI_k} is not equal to zero and is statistically significant at 1% significance level, i.e. p value ≤ 0.01 .

$$H_0 : \beta_{OFI} = 0;$$

$$H_1 : \beta_{OFI} \neq 0.$$

2. Trade flow imbalance TFI_k has a positive linear relationship with contemporaneous mid-price price change ΔMP_k , i.e. the price impact coefficient β_{TFI_k} is not equal to zero and is statistically significant at 1% significance level, i.e. p value ≤ 0.01

$$H_0 : \beta_{TFI} = 0;$$

$$H_1 : \beta_{TFI} \neq 0.$$

3. Order flow imbalance OFI_k has a stronger explanatory power than trade flow imbalance TFI_k on price ΔMP_k , as measured by the coefficient of determination— R^2 .

3 Literature review

This section provides an overview of current research in the field of cryptocurrency, market microstructure and the union of the two. Further, it identifies current gaps in the existing research and justifies the motivation to pursue the exploration of the chosen topic.

3.1 Market microstructure

Market microstructure concerns itself with study of the different agents within a confined market structure and events that occur between these agents, namely limit orders, market orders and cancellations. Order flow imbalance quantifies the difference between supply and demand within a limit order book (LOB) and has been formalised by Cont et al. (2014). Further studies have used order flow imbalance (OFI) and its variants to establish its predictive capacity over intra-day time frames, as opposed to measuring contemporaneous price change used in the initial study (Shen 2015; Jessen 2015). OFI examines how well a market can absorb the impact of order events. Understanding of this aspect of markets is essential for liquidity provision and market stability, both of which are of interest to financial institutions and government entities. The ability to forecast short-term price movements allows market makers to better position themselves in a stochastic environment to provide deeper levels of liquidity longer, thus dampening the effects of volatility (Bilokon 2018). By construction, market stability can be improved, which largely decreases the probability of evaporation-of-liquidity type of events, e.g. 2010 ES-mini Flash Crash (Kirelenko et al. 2017). In turn, market makers' ability to make both sides of the book based on a well-defined statistical edge that helps them avoid adverse selection. In parallel, understanding how a market functions on an intricate level also reduces second-order execution costs such as slippage and market impact.

3.2 Approaches to LOB analysis

Approaches to study of LOBs and their dynamics can be classified into two broad categories: theoretical (analytic) and empirical (data driven). Theoretical study of limit order books is motivated by replicating the processes observed in a LOB by means of a mathematical process. Many scientists converge that processes within LOB can be modelled by characteristics of a Markov process (Huang et al. 2015; Kelly and Yudovina 2017; Cont and De Larrard 2013), by those of a Hawkes process (Abergel and Jedidi 2015) and the hybrid of the two that evolves into a marked-point process (Morariu-Patrichi and Pakkanen 2017). Results are formulated by means of

mathematical analysis. Studies usually proceed to calibrate their models on empirical data to verify the validity of the models and find outstanding parameters.

Empirical approach to studying LOBs tackles the problem by first and foremost addressing the characteristics of real-world data. Such data-driven approaches were especially made attractive with development of (a) fast and efficient computation, (b) vast amounts of data and (c) modern Machine Learning algorithms that include classical statistical learning techniques (Hastie et al. 2001) and more recently, Deep Learning (Goodfellow et al. 2016). Goals of such studies vary from studying market impact and order book modelling (Donier and Bonart 2015; Cont et al. 2014), to extracting predictive capability from various market microstructure features (Dixon 2018; Sirignano and Cont 2018). Criticism around data-driven approach is centred around the fact that studies tackle the data head-on, often studying the statistical mechanics of the after-facts, rather than asking fundamental questions about possible origins of the underlying processes. Further styles of approach include econophysics approach, which attempts to model LOB dynamics by behaviour of sub-atomic particles (Chakraborti et al. 2011).

3.3 Birth of a new asset class

LOBs are a product of rapid technological development that took place over last 20 years. At its heart, LOBs attempt to solve the problem of supply and demand and target information asymmetry by indicating a state of the market at any given time (Black 1971). Electronification of exchanges, implementation of limit order books, emergence of high-frequency trading and automated execution have become essential features of the modern financial markets. However, financial technology has brought a lot of value to the financial industry outside of electronic trading enhancements—it has disrupted the way people exchange value and utility. The blockchain is in the core of Bitcoin function. A blockchain can be thought of as an immutable public ledger, in which all transactions are linked cryptographically. While the underlying technology is essential to understanding intrinsic value of an asset, this study will focus solely on order book dynamics of cryptocurrency markets.

3.4 Current research

The majority of current cryptocurrency research is based on blockchain technology (Zheng et al. 2016). A small fraction of research focuses on macroscopic price dynamics of cryptocurrencies (Osterrieder et al. 2017). To our current knowledge, there are only two academic studies that concern themselves with market microstructure of digital assets (Donier and Bonart 2015; Guo and Antulov-Fantulin 2018) and some blog posts that shine light on the subject (Heusser 2013). Technology research is concerned with various improvements of blockchain, such as scalability, throughput, applications to new industries and disruption of existing services by means of decentralisation. Current studies on price dynamics are mostly interested in predicting price series of cryptocurrency assets and try to understand how to value cryptocurrency objectively (Pagnottoni et al. 2018). Wheatley et al.

(2018) are able to model Bitcoin's market cap via application of Metcalfe's law to network size and show that Bitcoin price behaviour breaks its "fundamental" value on at least four occasions. Their analysis shows that such behaviour is well modelled by Log-Periodic Power Law Singularity (LPPLS) model, which parsimoniously captures diverse positive feedback phenomena, such as herding and imitation. Osterrieder et al. (2017) fits a large array of statistical models to cryptocurrency data to understand which distribution is best at modelling price dynamics. The study concludes that Bitcoin price is best explained by a hyperbolic distribution of returns.

Madan et al. (2015) explores avenues in automated trading of bitcoin, highlighting ease of market access and real-time feeds in data collection methodology. Researchers in Madan et al. (2015) propose Support Vector Machine (SVM), Binomial GLM and Random Forest algorithms that take various bitcoin network features, such as transaction count, mempool (number of unconfirmed transactions on Bitcoin blockchain network) and hash rate to classify 10-s and 10-min direction of return. They achieve maximum accuracy of 57% with deep Random Forest algorithm. Shah and Zhang (2014) successfully apply Bayesian regression to the problem of binary classification of bitcoin price direction. Jiang and Liang (2016) apply convolutional neural network to cryptocurrency portfolio management problem. The network takes price series of cryptocurrencies as input and outputs a weight vector constrained to being long-only portfolio. Study claims to achieve a tenfold return on investment within a space of few months.

Market dynamics studies that are enumerated above have clear limitations. First of all, they work with sampled trade data, omitting the order book dynamics. Such limited data can produce only limited backtests; it is not possible to simulate realistic execution and simulate slippage that would most likely occur in similar trading environments. Second, studying the strategies that attempt to maximise profit at such intervals does not merit much value to market stability and one may even argue them to be detrimental to stability of markets because the momentum price swings are likely to be amplified in such low-liquidity environments. Last but not least, validity of strategies that the studies come up with is questionable because data sample that they examine (mainly 2010–2017) is of trending nature, which may make these studies subject to overfitting.

3.5 Cryptocurrency market microstructure

Donier and Bonart (2015) examine market impact of meta-orders in Bitcoin/USD market. Researchers use a privileged dataset of trade data from Mt.Gox exchange that discloses trades of distinct traders by anonymous IDs. Among the subjects it studied are execution of large orders, market impact of meta-orders, intra-day volatility and market reaction to order flow. Perhaps the main result of the study is that square-root law of market impact holds for meta-orders in Bitcoin/USD market. That implies that the average relative price change between the first and the last trades of a meta-order is well approximated by the square root of the order volume, which is well documented for other more mature markets such as equities, futures and options (Bershova and Rakhlin 2013). The study proceeds to examine market impact

conditioned on various order flow predicates. Researchers highlight a very distinct feature of Bitcoin/USD market—during execution of meta-orders, impact is nicely approximated by square root of global market imbalance. They conclude that market impact is not a reaction to individual meta-orders, but to the whole order flow. This motivates further study of order flow imbalance in cryptocurrency markets. Guo and Antulov-Fantulin (2018) apply a range of Machine Learning techniques to predict Bitcoin prices using various LOB features, but are rather ambiguous about the features they use.

While data in Donier and Bonart (2015) are represented by a privileged dataset, its source is one of the first-organised exchanges and dataset ends in 2013 due to a hack that bankrupted the exchange. Technological inferiority of Mt.Gox exchange may have affected the robustness of its LOB and hence the data that were recorded from it. Last but not least, Mt.Gox offered a flat-fee trading schedule of 60 bps for both, makers and takers. This is no longer the case for many current cryptocurrency exchanges; fee schedules and especially, rebates to market makers are essential features of market microstructure landscape of any organised exchange.

4 Methodology

This section describes the data that are used throughout the study, defines the variables subject to analysis and specifies the models that are fit to the variables.

4.1 Data

4.1.1 Data collection

The data that are used in this study correspond to the time period beginning in September 2017 and ending in November 2017. The data were collected via application programming interface (API), publicly provided by BitMEX exchange. Due to computational constraints, such as random-access memory (RAM), the subset of the data which is used in the study starts from 1 October 2017 and ends on 23 October 2017. Quote and trade data contain 81.3 million and 38.9 million data points, respectively.

Data were collected for the XBTUSD pair, which is the most traded pair on Bitmex, see “[Appendix A](#)”. The data are initially stored as Comma Separated Values file (CSV) and are later partitioned by trading day and stored in kdb+, an in-memory high-frequency database. Each row of quote data corresponds to an event taking place at the top of the order book (best bid and best ask). In other words, if there is a limit, a cancellation or a market order that changes the state of the top of the book, a new row will reflect that change. This representation is also known as Level I order book.

Given the fragmentation of cryptocurrency markets and lack of interoperability between the venues, the sources of data were carefully considered. BitMEX was chosen due to its satisfaction of two main criteria: sufficient liquidity and

lack of continuous downtime. Being the biggest exchange by trading volume, BitMex ticks the box of sufficient liquidity, having average daily turnover of \$3 billion. Second criterion is also satisfied, as BitMex has had lowest downtime out of all existing exchanges as of this writing. See “[Appendix A](#)” for BitMex exchange specification.

4.1.2 Data format

Level I order book reflects any changes at the best bid and asks levels of the LOB. More formally, it is represented in the following format:

Timestamp	Bid price	Bid volume	Ask price	Ask volume
2017-10-11 03:10:34.852660	4753.6	6397	4753.7	59216
2017-10-11 03:10:35.095169	4753.6	6589	4753.7	59216
2017-10-11 03:10:35.168064	4753.6	6397	4753.7	59216
2017-10-11 03:10:35.354433	4753.6	6397	4753.7	54216
2017-10-11 03:10:35.393526	4753.6	6397	4753.7	56216

Columns represent:

- Timestamp: nanosecond timestamp.
- Bid price: highest price a market maker is willing to buy a cryptocurrency for.
- Ask price: lowest price a market maker is willing to sell a cryptocurrency for.
- Bid volume: current contract volume available at best bid price. Unitary.
- Ask volume: current contract volume available at best ask price. Unitary.

The collected data also include individual market orders, i.e. trades, which are represented by sequential time series, where each row corresponds to a market order:

Timestamp	Price	Volume	Side
2017-10-11 03:09:53.566447000	4754.0	66	Sell
2017-10-11 03:09:53.858378000	4754.0	24	Sell
2017-10-11 03:10:01.632378000	4754.1	10	Buy
2017-10-11 03:10:12.383103000	4754.0	4500	Sell

Interpretation of market orders:

- Timestamp: nanosecond timestamp.
- Price: trade price.
- Amount: trade volume.
- Side: buy/sell market order differentiator.

4.1.3 Benchmarking

In some instances of the study, there occurs a need to benchmark the findings to the facts about established asset classes. The research question is “To what extent does order flow impact prices in cryptocurrency markets?” Answering this question in absolute terms will not give an intuitive answer without a reference point, whereas benchmarking to more established asset class will make the results more relevant by comparison. For these purposes, Level I data for ES-mini contracts are obtained. Traded on CME, ES mini-contracts are cash settled based on S&P 500 index value and are the most liquid equity index futures in the world. ES-minis represent a very liquid, mature and hence, stable financial instrument that makes it a good reference point for a nascent asset class that cryptocurrencies represent. The dates of the dataset correspond to the period of May 2016, which is the only period for which such data are available. Format of the benchmark data is the same as that of cryptocurrency dataset—Level I TAQ (trades and quotes) data.

4.2 Definitions

4.2.1 Limit order book

A limit order book is a reflection of current supply and demand present respective of an asset at some time t . LOB is an implementation of an order-driven market. A state of a LOB can be characterised in terms of a collection of orders being that are present in a LOB. LOB consists of orders signifying interest to buy (bids) and orders signifying interest to sell (asks). Hence, limit bid orders can be thought of as indications of demand; inversely, limit ask orders can be thought of as indications of supply.

Participants in a LOB are predominantly separated into two groups: market makers and market takers. Market maker is an agent that posts liquidity onto a LOB by means of a limit order. Market taker is an agent that depletes the LOB liquidity by means of posting an order that matches an order of a market maker, usually known as a market order. The de facto LOB is a double auction model, whereby orders on either side are prioritised by price and at each price level distinct orders are prioritised on a first-in-first-out basis (Cartea et al. 2015).

State of a LOB changes with introduction of new order events. Recent electronic trading innovations have introduced a large number of order types, but predominantly, they consist of limit orders, market orders and cancellations (Johnson 2010). Limit order is a binding intention to either buy or sell a specified quantity of an asset for at least (for limit sell orders) or at most (for limit buy orders) some price p . Limit orders guarantee the price but not the execution. Market order is an intent of buying or selling at best available market price. Market orders guarantee the execution but not the price. Cancellation removes an unmatched or partially matched limit order from an order book. State of an order book evolves with arrival of these three base types of orders.

Nowadays, in face of fast pace of financial markets, majority of venues use LOBs to match buyers and sellers. The Hong Kong, Swiss, Tokyo, Moscow, Euronext and Australian Stock Exchanges now operate as pure LOBs. New York Stock Exchange (NYSE), NASDAQ and London Stock Exchange (LSE) operate a hybrid LOB system, which allows an execution through market specialist and floor brokers, as well as direct access to the exchange LOB (Gould et al. 2013). Majority of cryptocurrency trading venues utilise a vanilla all-to-all LOB model (Hileman and Rauchs 2017).

4.2.2 Order flow imbalance

Out of many interpretations of a LOB imbalance that have been developed over the last couple of decades, some researchers define the imbalance as an imbalance of trades, aggregated by their direction (Lee and Ready 1991; Chordia et al. 2002). Others consider the aggregate order flow imbalance, composed of all types of events taking place in a LOB (Cont et al. 2014). OFI is a quantification of supply and demand inequalities in a LOB during a given time frame. OFI rests on the fact that any event that changes the state of a LOB can be classified as either the event that changes the demand or the event that changes the supply currently present in a LOB. Namely

- Increase in demand in a LOB is signified by any of the following events:
 - Arrival of a limit bid order
- Decrease in demand in a LOB is signified by any of the following events:
 - Arrival of market sell order
 - Full or partial cancellation of a limit bid order
- Increase in supply in a LOB is signified by any of the following events:
 - Arrival of limit ask order
- Decrease in supply in a LOB is signified by any of the following events:
 - Arrival of a market buy order
 - Full or partial cancellation of a limit ask order

Cont et al. (2014) assumes a simplified model of a LOB, under which the price impact of any given order event is deterministic. The main assumption of the model is uniform distribution of liquidity across price levels; all price levels beyond best bid and ask are assumed to have a certain quantity of volume D present. By extension, cancellations and limit order arrivals occur only at best bid/ask. Under these assumptions, effects of individual order events are additive. Hence, over a specified time interval $[t, t + \Delta t]$, bid price change ΔP^b (in ticks) can be calculated by adding together impacts of three different event types:

$$\Delta P^b = [L^b - C^b - M^s]/D,$$

where L^b represents volume of limit bid orders, having a positive impact on bid price, C^b represents bid order cancellations (negative impact) and M^s represents market sell orders (negative impact). Ask price change ΔP^a is determined by the same method:

$$\Delta P^a = [-L^a + C^a + M^b]/D.$$

Given the order book model described above, calculating price changes over period $[t, t + \Delta t]$ become trivial. For demonstration purposes, let D be 10. Assume that in a given interval, the following events take place: limit bid order of size 3, bid order cancellation of size 12 and market sell order of size 5. Bid price change will be

$$\Delta P^b = [L^b - C^b - M^s]/D = [3 - 12 - 6]/5 = -3 \text{ ticks.}$$

Stylized parameter D makes price change calculations dependent only the net order flow. In reality, LOBs are not so ideal. LOBs are usually full of humps, gaps, sometimes thin and sometimes bearing hidden iceberg orders (Gould et al. 2013). Hence, the parameter D is far from constant in reality—it is constantly changing with very complex dynamics. Nevertheless, Cont et al. (2014) show strong linear relationship between net order flow and contemporaneous price change in US Equity markets.

The formal definition of OFI is derived from the above definitions. In general, the impact of a single event is quantified as follows:

$$e_n = I_{\{P_n^B \geq P_{n-1}^B\}} q_n^B - I_{\{P_n^B \leq P_{n-1}^B\}} q_{n-1}^B - I_{\{P_n^A \leq P_{n-1}^A\}} q_n^A + I_{\{P_n^A \geq P_{n-1}^A\}} q_{n-1}^A,$$

where P_n^A and P_n^B are best bid and best ask prices at index n respectively, q_n^B and q_n^A are bid and ask volumes, respectively, and I is the price-conditional identity function. To provide an intuition of mechanics of e_n , if q^B increases by some volume v , signifying an increase in demand via a limit bid order placement, e_n takes on the value of

$$q_n^B - q_{n-1}^B - q_n^A + q_{n-1}^A$$

since neither of the best bid and ask prices actually changed. By construction of Level I quote data,

$$-q_n^A + q_{n-1}^A = 0$$

since only one event can occur between the observations, which means that $q_n^A = q_{n-1}^A$. This implies that

$$e_n = q_n^B - q_{n-1}^B = v,$$

the size of the new limit order added to the bid queue v . In summary, e_n measures the supply/demand impact of n th order event.

Order flow imbalance is an aggregation of impacts e_n over a number of events that take place during time frame t :

$$\text{OFI}_k = \sum_{n=N(t_{k-1})+1}^{N(t_k)} e_n,$$

where $N(t)$ is the number of events occurring at Level I during time frame $[0, t]$. OFI can be seen as an accumulator of supply and demand changes over a given time frame.

The response variable is the contemporaneous mid-price change in number of ticks over the same time frame as OFI_k :

$$\Delta \text{MP}_k = (\text{MP}_k - \text{MP}_{k-1}) / \delta,$$

where MP_k is a simple mid-price defined as $\frac{P_t^B + P_t^A}{2}$ at time t and δ is the tick size, which is 10 cents in our data and is constant. Division by tick size is in line with assumptions made in Sect. 4.2.1.

The linear model that will be evaluated regresses contemporaneous price change on OFI:

$$\Delta \text{MP}_k = \hat{\alpha}_{\text{OFI}} + \hat{\beta}_{\text{OFI}} \text{OFI}_k + \epsilon_k,$$

where $\hat{\beta}_{\text{OFI}}$ is the price impact coefficient, ϵ_k is the error term and $\hat{\alpha}_{\text{OFI}}$ is the intercept. The model is to be fit by the method of Ordinary Least Squares (OLS). The chosen time intervals k are 1 s, 10 s, 1 min, 5 min, 10 min and 1 h.

4.2.3 Trade flow imbalance

Plerou et al. (2002), Karpoff (1987), Chordia et al. (2002) and Lee and Ready (1991) are only a few examples that study price changes as a function of trade flow. Some claim trade imbalance to work as a predictor in practical trading environments (Chan 2017). By definition, trade events are a subset of order book events considered in OFI. By intuition, trade flow imbalance may not have as big of an explanatory power as OFI because components of the latter are a superset of components of the former. However, when one places and cancels the order, he has virtually no cost of doing so, whereas to place a trade, one pays a commission and a bid/ask spread. Trade flow imbalance over time interval t is defined as

$$\text{TFI}_k = \sum_{n=N(t_{k-1})+1}^{N(t_k)} m_n,$$

where

$$m_n = -I_{M^s} + I_{M^b},$$

where M^s and M^b represent market sell and market buy orders, respectively, and $N(t)$ is the number of events occurring at Level I during $[0, t]$. I is the identity function that differentiates between market buy and sell events, marking them with according signs. This study will investigate to what extent trade flow events impact price in cryptocurrency markets by means of the following linear regression model:

$$\Delta \text{MP}_k = \hat{\alpha}_{\text{TFI}} + \hat{\beta}_{\text{TFI}} \text{TFI}_k + \epsilon_k,$$

where k is a time interval over which the magnitudes of signed market orders and mid-price change are calculated. $\hat{\beta}_{\text{TFI}}$ is the trade impact coefficient, $\hat{\alpha}_{\text{TFI}}$ is the intercept and ϵ_k is the error term. The model is to be fit via OLS.

5 Analysis and results

5.1 Statistical properties

5.1.1 Prices

Financial time series data are one of the most complex types of data sets that one can attempt to comprehend; datasets tend to have non-Gaussian and non-stationary properties (Bilokon 2018). The latter implies the dynamically changing statistical properties of financial data over time. Market microstructure introduces additional estimation difficulties, due to so-called “microstructure effects” (Aït-Sahalia and Jacod 2014). Cryptocurrency price series are especially subject to non-stationarity; sentiment swings and lack of fundamental pricing contribute to wild volatility of cryptocurrency, which can be orders of magnitudes larger than that of mature asset classes, e.g. U.S. Equities (Chu et al. 2017). To examine stationarity of the differenced price series, augmented Dickey–Fuller (ADF) and KPSS tests are conducted for all variants of k in ΔP_k variable. The tests confirm that price series are stationary for every sampling period k at 1% significance level.

Figure 1 illustrates a typical trading day featuring XBTUSD. Being a 24-h market, cryptocurrency is traded around the clock by traders globally. The price for the day starts at around \$5900, has a drawdown of roughly 10% to \$5400 and ends the day with a rally back to \$5700. The high volatility is precisely what attracts day-traders to cryptocurrency; such volatility is unheard of in established markets. Bottom panel illustrates a proxy for volatility—moving average of mid-price changes

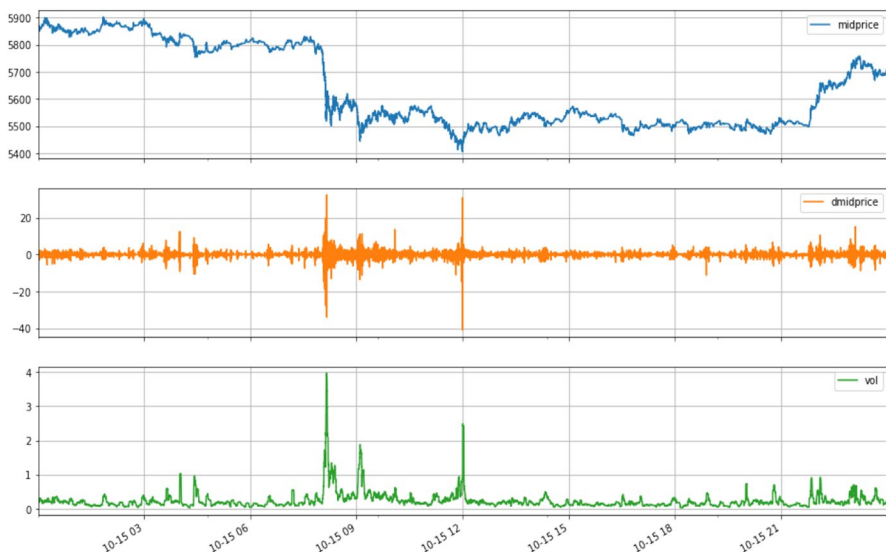


Fig. 1 XBTUSD midprice series P (top), differenced tick-to-tick mid-price series ΔMP (middle), rolling volatility (bottom), estimated by moving average of $SD \sigma_{\Delta MP}$ over 1000 tick window. Date: 15th October, 2017

standard deviation over 1000 ticks. Spikes in volatility are rather sudden, explained by low market depth at times of urgency, possibly explained by lack of liquidity providers.

5.1.2 Orders

The intensity of order arrivals is perhaps the biggest differentiator of cryptocurrency market microstructure dynamics in respect to that of other established asset classes. Empirical findings show that order arrival rates in cryptocurrency markets are orders of magnitude lower than those in mature markets. Level I updates are aggregated by count of updates per 1 s. The statistics are then benchmarked to ES-mini contracts.

Table 1 makes cryptocurrency low arrival rates apparent as compared to ES-mini contracts. It is especially staggering how different the maximum arrival rate is between the assets. Attributed mainly to number of market agents and technology of exchanges, ES-mini contracts Level I LOB experienced maximum arrival of 2387 orders in 1 s versus XBTUSD's 48. To examine this further, auto-correlation function of update arrival volumes is computed by summing the number of updates into 10-s buckets and performing auto-correlation. Update arrival volume ACF (Fig. 2) suggests that activity on best bid and ask levels is moderately time dependent. Thus, instances of high levels of activity are likely to be immediately succeeded by instances of high levels of activity and calm periods are usually succeeded by calm periods. ACF stays at roughly 30% after 8th lag due to the fact that update counts are always positive. Such nature of markets is well documented and consistent with others' findings (Cartea et al. 2015).

5.2 Order flow imbalance

Recall that order flow imbalance is a quantification of supply and demand activity over a given period of time k . Augmented Dickey–Fuller stationarity tests are conducted on non-differenced OFI_k variable, where k corresponds to 10-s sampling frequency. Test results, presented in Table 2, confirm that the order flow imbalance series are stationary by virtue of rejection of null hypothesis at 1%, 5% and 10% significance levels.

Table 1 XBTUSD versus ES-mini May 2016 contract: 1-s arrival rate descriptive statistics

	XBTUSD	ESM16
Mean	4.932	57.657
SD	5.427	96.234
Min	0.000	0.000
25%	1.000	4.000
50%	3.000	20.000
75%	7.000	64.000
Max	48.000	2387.000

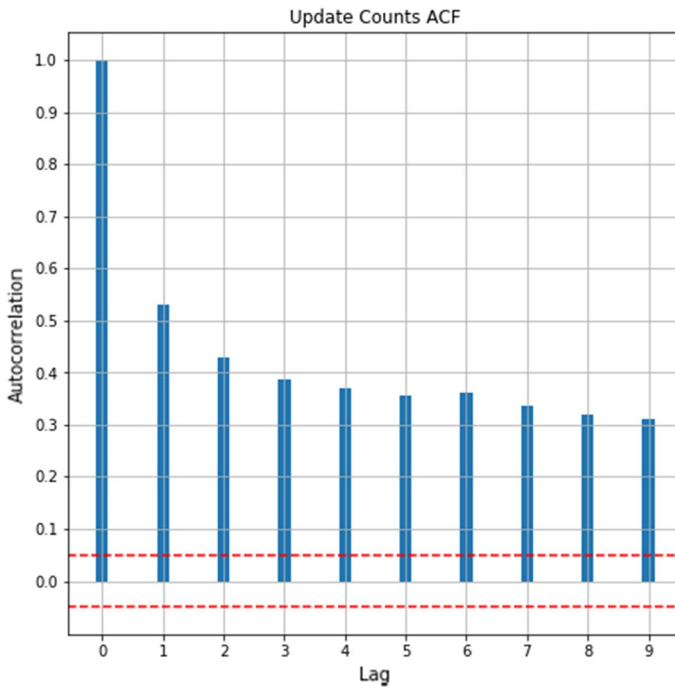


Fig. 2 XBTUSD 10-s update counts ACF

Table 2 OFI augmented Dickey–Fuller test results

ADF statistic	− 484.153707
<i>p</i> value	0.0000001
Critical values	
1%	− 3.430
5%	− 2.862
10%	− 2.567

Cont et al. (2014) start sampling from 1-s interval and then grow the sampling interval up to 10 min. We begin by sampling at 1-s intervals and extend the window up to 1 h to account for **lower arrival rates**. The reason for aggregation of events over a larger time grid is to accumulate a reasonable amount of order book events so that the eventual price change and imbalance are sufficiently observable. The first model is fit to data that are aggregated over 1-s intervals and demonstrates this subtle point and how it effects the model R^2 .

Self-dependence of OFI is evident from ACF plot (Fig. 3), which shows that 10-s OFI is positively auto-correlated up to third lag. Auto-correlation of order flow imbalance shows that direction and magnitude of order flow is self-dependent. In other words, periods of positive order flow (demand increasing or supply decreasing) are likely to be succeeded by positive order flow and vice versa for negative order flow.

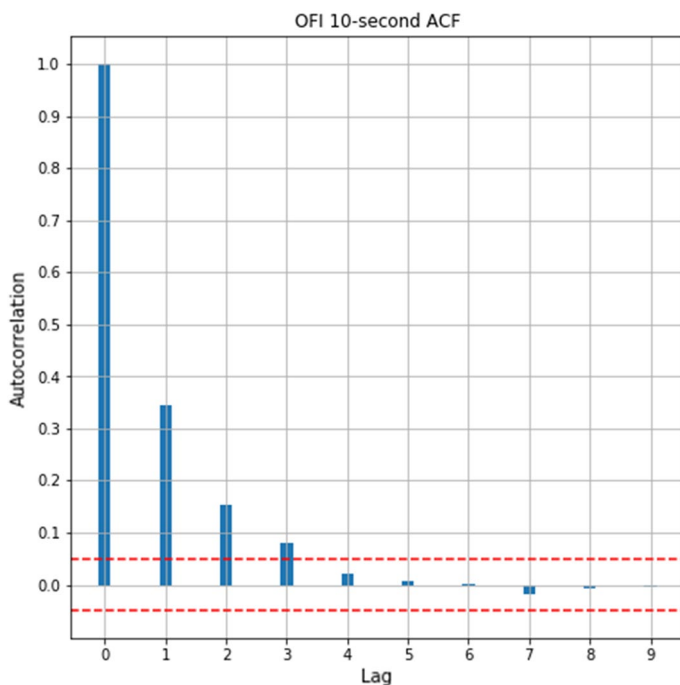


Fig. 3 XBTUSD 10-s order flow imbalance ACF

This largely extends our findings in Sect. 5.1.2; not only is there time dependence between absolute activity levels, but there is also dependence between the signs of the activity. Chordia et al. (2002) suggests that this nature is due to traders splitting their orders into multiple smaller orders, thus the succeeding order activity is likely to have the same sign.

The first relationship that is investigated is between OFI and the mid-price change (change in number of ticks) over a 1-s time frame by means of fitting an Ordinary Least Squares linear model defined in Sect. 4.2.2.

One-second OFI model demonstrates the point mentioned earlier—low update arrival rates require sampling over a bigger time frame to observe a substantial price change/order flow imbalance. It is visible that the scatter plot (Fig. 4) represents a “sliding cross” formation whereby not much activity is able to develop as most points lay close to the axes of the graph. Correspondingly, the linear relationship between OFI and price change is poor at this sampling window. R^2 of 1-s OFI linear model fit is 7.1%. All results are presented in Table 2 at the end of the section.

When k is set to 10 s, the linear model has a much better fit— $R^2 = 40.5\%$. The linear relationship starts to resemble the one that Cont et al. (2014) observe. One-minute time frame provides an even clearer demonstration of linear relationship between price change and order flow imbalance, see Fig. 5. The corresponding model with its estimated parameters is

$$\Delta MP = -0.19173 + 8.383e - 5 OFI.$$

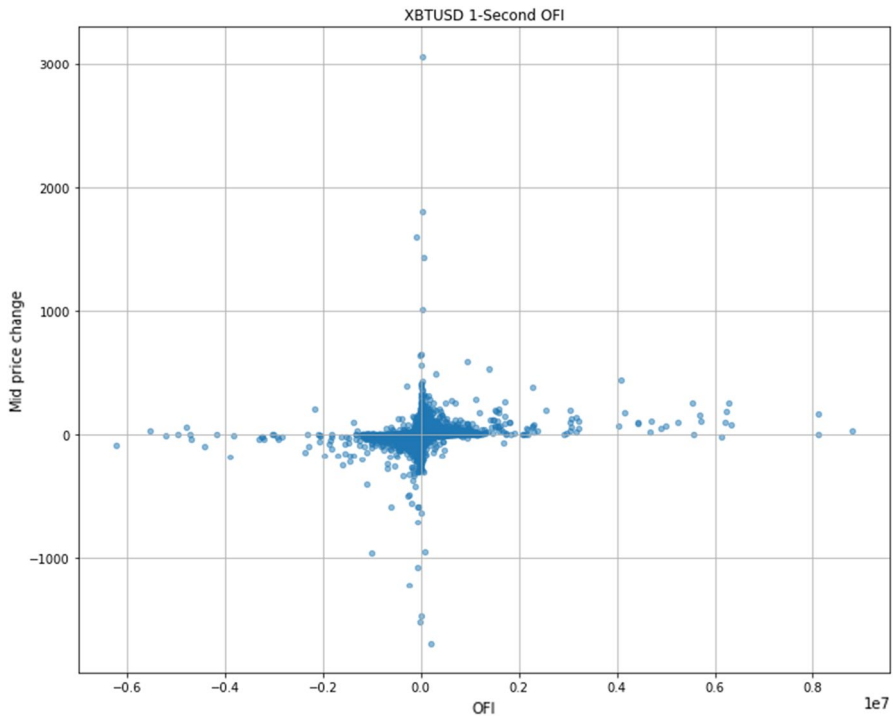


Fig. 4 XBTUSD 1-s order flow imbalance versus contemporaneous price change

The interpretation of the model is very intuitive: for 10,000 units of net order flow, the expected average mid-price change is 0.65 ticks. Note that the price impact coefficient $\hat{\beta}_{\text{OFI}}$, does not differentiate between types of order book events, hence generalising for cancellation, placement and trade order volume flows.

As the time frame is increased to 5-min, 10-min and 1-h intervals, there is an increase in the goodness of fit. R^2 eventually plateaus at 55%. At this point, we start seeing a linear relationship between OFI and the contemporaneous price change. The R^2 never gets as high as 65% as per results of Cont et al. (2014). The results merit the rejection of the first null hypothesis that OFI_k does not have a positive linear relationship with contemporaneous price change ΔMP_k at 1% significance level. The price impact coefficients $\hat{\beta}_{\text{OFI}}$ are statistically significant for all sampling windows as evidenced by the p values of the coefficients in Table 2.

5.3 Trade flow imbalance

This section examines how trade flow imbalance (TFI) affects the contemporaneous mid-price change. Cont et al. (2014) find that explanatory power of TFI in U.S. equities is weaker than that of OFI for all the stocks they examine.

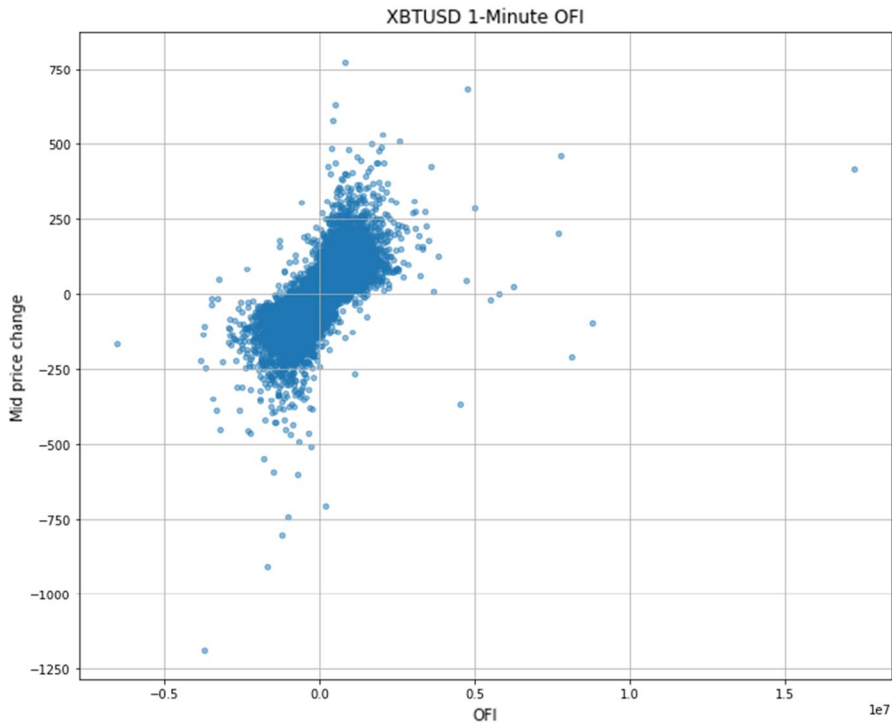


Fig. 5 XBTUSD 1-min order flow imbalance versus contemporaneous price change

The intervals that are used for calculation of TFI and contemporaneous price change are the same as the intervals used in OFI modelling: 1 s, 10 s, 5 min, 10 min and 1 h. Augmented Dickey–Fuller and KPSS stationarity tests are conducted on TFI_k for every sampling period k . The tests confirm stationarity of the variable for every sampling period at 1% significance level. As demonstrated by Fig. 6, trade flow demonstrates similar time dependence characteristics to aggregate order flow. It is evident that 10-s trade flow imbalance is significantly and positively auto-correlated with lags 1–5. This means that trades exhibit momentum towards the current direction of trade flow. Heusser (2013) finds the process of Bitcoin trades to be self-exciting, whereby time between trades is sparse and trades usually arrive in clusters. This study largely conforms to findings of Heusser (2013) and extends that the clusters tend to be uni-directional respective of the current trade flow.

The first model regresses TFI on contemporaneous mid-price change ΔMP sampled over 1-s intervals. The model produces a coefficient of determination of 12.8%, which is higher than the R^2 achieved for OFI over the same sampling period. At 10-s sampling interval R^2 of the TFI model is 37.3%, which is lower than its 10-s OFI counterpart, whose R^2 is 40.5%.

For sample periods higher than 10 s, however, TFI is a consistently better estimator of the price change than OFI. At 1-h sampling time grid R^2 is 75.2%, which is 20% higher than the order flow imbalance R^2 for the same time grid.

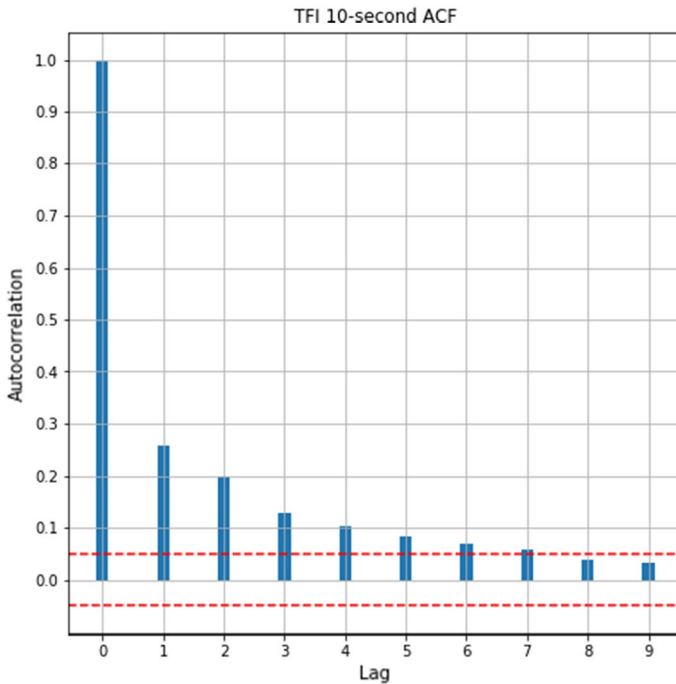


Fig. 6 XBTUSD 10-s trade flow imbalance ACF

Scatter plot exhibited in Fig. 7 visually demonstrates that the fit is more linear than that of OFI in Fig. 5. The estimated model and its parameters are

$$\Delta MP = 0.01297 + 1.1884e - 5TFI.$$

The interpretation of the model is as follows: for 10,000 units of net trade flow, the expected average mid-price change is 0.14 ticks.

These results are counter to the findings of Cont et al. (2014) that find that for all 50 U.S. stocks chosen for their analysis, order flow imbalance takes precedence of explaining contemporaneous price change for every single one. The initial hypothesis that aggregate order flow imbalance has stronger explanatory power than trade flow imbalance is rejected based on these results. The null hypothesis, which states that TFI does not have a positive linear relationship with contemporaneous price change ΔMP , is rejected at 1% significance level. Results in Table 3 confirm that $\hat{\beta}_{TFI}$ is statistically significant for all sampling periods k . Note that the p values are close to zero for all TFI and OFI coefficients, and instead of being reported, are instead subtracted from 100%, yielding the probabilities of the coefficients not being obtained by chance. t -Statistics are also included for every estimated coefficient.

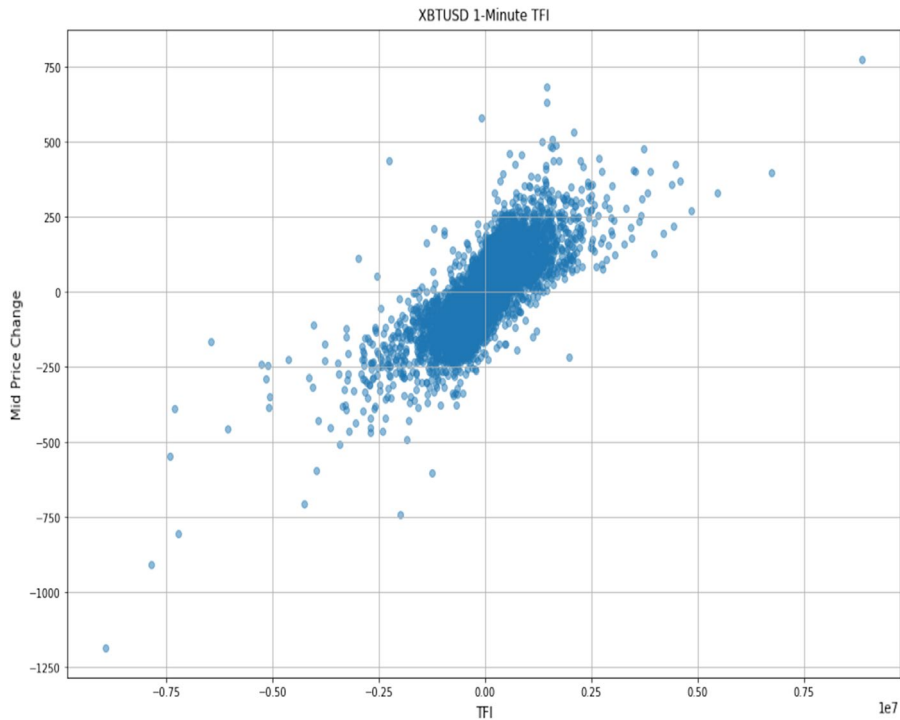


Fig. 7 XBTUSD 1-min trade flow imbalance versus contemporaneous price change. $R^2 = 58.1\%$

5.4 Discussion

Order flow imbalance provides a good approximation for realised mid-price change, and there are a few potential reasons why OFI does not provide a better fit. First of all, it helps to understand under which circumstances OFI provides an inferior estimate of contemporaneous price change. More crudely, under what predicates will the data points end up in second and fourth quadrants on scatter plot such as the one presented by Fig. 5. The stylised LOB model in Sect. 4.2.2 assumes that each and every level of the order book contains outstanding orders amounting to some constant quantity D and that activity takes place at best bid/ask levels only. Now, let us assume that at time t there is volume V_b present and best bid and V_a present at best ask, such that $V_a > V_b$. At time $t + 1$, a cancellation order arrives on ask side, cancelling amount $q_c < V_a$, thus, ceteris paribus, registering a positive effect on the current order flow calculation, and leaving the mid-price unchanged. At time $t + 2$, there is a sell market order of quantity q_m such that $q_m > V_b$ and $q_m < q_c$. This market order moves the mid-price down, but because $q_m < q_c$, current OFI value is still positive. The resulting data point will end up in the second quadrant of the scatter plot. Thus, it is unevenness of volume across the LOB price levels that exacerbates the estimation of price change by OFI.

Table 3 Results: order flow imbalance and trade flow imbalance model parameter estimation

k	Order flow imbalance				Trade flow imbalance					
	$\hat{\alpha}_{OFI}$	$\hat{\beta}_{OFI}$	$t(\hat{\beta}_{OFI})$	R^2	$\hat{\beta}_{OFI} \neq 0$ (%)	$\hat{\alpha}_{TFI}$	$\hat{\beta}_{TFI}$	$t(\hat{\beta}_{TFI})$	R^2	$\hat{\beta}_{TFI} \neq 0$ (%)
1 s	0.0029	4.442e-5	337.649	7.1	99.9	0.0044	7.712e-5	529.263	12.8	100
10 s	- 0.0206	7.568e-5	360.924	40.5	100	0.029	0.0001	337.434	37.3	100
1 min	- 0.1917	8.383e-5	197.709	55	99.9	0.130	0.0001	210.509	58.1	100
5 min	- 0.4276	7.117e-5	82.46	51.6	99.9	0.942	0.0001	117.727	68.5	100
10 min	- 0.5458	6.747e-5	55.893	49.5	99.9	2.147	9.238e-5	87.924	70.8	100
1 h	- 1.6048	6.416e-5	24.218	52.4	99.9	14.166	8.572e-5	40.065	75.2	100

Upon examining Level III data for a number of snapshots, numerous instances where best bid/ask and adjacent price levels are unevenly filled, and some not filled at all, exemplifying a thin order book, that would welcome such non-linear relationship between OFI and ΔMP . The original Cont et al. (2014) study investigated order flow in established markets (US Equities), which are more liquid and hence price impact can be modelled more accurately by order flow imbalance. Thus, the goodness of fit is a function of two main factors: (a) depth D at all price levels and (b) more realistically, dispersion of D , since all real-life markets will have non-constant D . If LOB price levels have a very “volatile” D , the effects of order flow will not even out as well as if D is not so dispersed. Concluding from statistics and empirical evidence, cryptocurrency prices are impacted by order flow in a much less deterministic fashion than established markets due to lower compliance with the stylised model of LOB that this study assumes.

The results also show that the impact of trade flow imbalance on prices is stronger than that of order flow imbalance. The explanatory power of TFI depends on the same depth parameter D and its dispersion across price levels. Circumstances under which trade flow will not be a good estimator of price change are, therefore, similar to circumstances under which order flow will not be a good estimator of price change.

The aggregate order flow already includes trades, so why does the trade flow on its own explain price movements better? The argument comes down to the fact that while aggregate order flow includes more information, in the realm of cryptocurrency market microstructure as well as macrostructure, such information may be of little value, due to noise. There are a few possible reasons that may help explain this phenomenon, both macrostructural as well as microstructural.

Unlike U.S. Equities, that are subject to multiple anti-spoofing policies including Dodd–Frank Wall Street Reform (Pasquale 2014) (spoofing constitutes an action of posting and cancelling limit orders in quick succession to disguise the intent of executing an order), there are no equal regulatory counterparts in cryptocurrency markets. This may have repercussions for why order flow may carry relatively lower information as opposed to trade flow in cryptocurrency markets. Traders who submit and quickly cancel orders to fake the intent of buying/selling are not legally constrained from doing so. Therefore, market agents are more inclined to post low-information orders of any magnitude into the LOB if that benefits their agenda. For example, a market maker that sits on a large inventory could choose to spoof in the direction that would benefit the value of his net inventory. This leads to ephemeral liquidity, i.e. orders that do not intend to be executed and, therefore, do not contribute to net price change. On the other hand, to execute a market order, a trader will pay a commission as well as a bid/ask spread, thus signifying a high-information intent that, as can be evidenced from the results, has a significant impact on price.

Stylised model of an order book described in 4.2.1 states that price change is inversely proportional to the depth parameter D , which, in the realm of theoretical model, is assumed fixed for all levels in a LOB. While it is only a theoretical relationship, we can clearly see how markets of lower liquidity abide to that relationship, exhibiting much higher average ΔP_k than U.S. Equities. In case of this study,

depth was not measured empirically, and it would make for a good basis for subsequent research, specifically in cryptocurrency markets.

D and its variance across price levels are the main factors that drive explanatory power of both OFI and TFI. The results also conclude that TFI has an overall better explanatory power than OFI, while the component events of the latter are a superset of component events of the former. This phenomenon is largely attributable to two things that are both, though indirectly, functions of parameter D . First of all, it is possible to consider the bid/ask spread having an effect on low explanatory power of OFI. The average spread of XBTUSD contract is 2.87 ticks, with standard deviation of 11 ticks, which is large and dispersed if compared to American equities, where large cap stocks rarely have average spreads larger than one tick (Upson and Van Ness 2017). When the spread is large, the mid-price can be manipulated at little or no cost by posting and cancelling limit orders at best bid and ask, whereas if the spread is almost always at one tick, there is no cost-less way of manipulating the price in the same way. In such circumstances, OFI is more likely to have a poor explanatory power. Cont et al. (2014) present that the CME Group stock that has an average spread of 103 ticks (the biggest of the group of selected stocks), also has the worst OFI R^2 of 35%, as compared to other stocks used in the study. Contrary to our results, however, CME's TFI has worse explanatory power than its OFI counterpart, which may be attributed to its below-average quote/trade ratio of 27.14. XBTUSD, on the other hand, has a quote/trade ratio of 2.08, which means that there is an average of only two quotes per trade. That suggests that there is very big propensity to trade (much higher than in U.S. Equities) in cryptocurrency markets. This propensity may imply a lack of market makers that are able to provide liquidity, and hence stabilise the depth across the order book. Such conditions may well justify the generous market maker rebates that BitMex pays to liquidity providing traders.

6 Conclusion

In conclusion, cryptocurrency market shares many features with conventional markets, specifically on microstructure levels. Main differences are attributed to lower average depths of the order book, which spawn other discrepancies related to how order books absorb order flow. One of the interesting findings that the study discovers is how well the price change can be explained by trade flow imbalance.

Further research may attempt to drill into this cause further. It would be of great use to analyse the linear model that combines both OFI and TFI as explanatory variables, whereby the noisiness of the former may become apparent. Bearing in mind that the study explored the biggest derivatives market for Bitcoin, which is also bigger than any other existing spot market by dollar turnover, it is highly advisable to replicate the research methodology on spot markets. Other exchanges have different characteristics, such as maker/taker fee schedules, volumes and participants. These factors are very likely to produce different landscape of market microstructure and hence, different results. Another direction that can be explored is the predictive capacity of the order flow in cryptocurrency markets. Existing studies that tackle this area, specifically in the realm of

cryptocurrencies, are rather ambiguous (Guo and Antulov-Fantulin 2018) with the input features used in generating predictive models. Without being able to have a stable forecasting apparatus, optimal liquidity provision is hardly attainable (Bilokon 2018). Yet further studies focusing on cryptocurrency market microstructure may also consider how underlying protocols of the currencies, such as mining algorithms and network statistics, manifest themselves in the microstructure.

The study began by saying that cryptocurrencies are a nascent asset class. As such, its value may continue being subject to sentiment shifts of different entities like governments and it might continue being an asset of high volatility that it is. One of the results of this study suggests that there is a clear lack of liquidity providers in this market. Brittle market depth and volatility create a “chicken and egg” problem, whereby cryptocurrency might continue lacking mass adoption and repel quality liquidity providers in face of its current volatility and thin markets.

Appendix A: Exchange specification

An essential feature of BitMex is that, above all, it is a marketplace for derivatives on cryptocurrency, as opposed to a spot market. All margin payments are carried out in Bitcoin, thus the only predicate for participating in the markets is a Bitcoin deposit. Another key feature of BitMex is leverage that it offers to traders. Currently, maximum leverage that one can take out on XBTUSD contract is $\times 100$.

XBTUSD is effectively a perpetual swap contract, where one contract is worth 1 USD of Bitcoin. XBTUSD never expires, but participants are may be subject to margin funding. The contract tracks the underlying price of Bitcoin, which is calculated as an index across various spot markets. The tracking mechanism is dependent upon funding ratio. In essence, to reduce tracking error, BitMex will calculate the deviation between current XBTUSD contract value and spot price index. If the value of the contract is above the reference index, than the implied interest rate of Bitcoin is higher than USD. Hence, to stabilise the price, the long contract holders will pay funding the short-sellers of the contract. This mechanism applies vice versa when contract value falls below the reference index and is what keeps the contract at fair price.

Trading fee structure on BitMex is very straightforward and highly shifted towards market makers when compared to other exchanges. Market makers get paid a constant 25 bps rebate, while takers pay 35 bps in commission.

Appendix B: Python code

Listing 1: Python Code for Data Import, Model Fitting and Stationarity Tests

```
import statsmodels.api as sm
from statsmodels.regression.linear_model import OLS
import statsmodels.tsa.stattools as ts
import pandas as pd
import numpy as np
%matplotlib inline
import matplotlib.pyplot as plt
plt.rcParams['figure.figsize'] = (12.0, 10.0)

def load_data_hdf(path, date_from, date_to = None):
    df = pd.read_hdf(path, index_col = 'timestamp')
    df.index = pd.to_datetime(df.index.values, format='%Y-%m-%dD%H:%M:%S.%f000')
    return df.loc[date_from:]

def ofi(quotes, dt):
    qdf = reset_index(quotes.copy())

    qdf['Mid-Price_Change'] = ((qdf['bidPrice'] + qdf['askPrice'])/2).diff().div(0.1)

    qdf['prev_bidprice'] = qdf['bidPrice'].shift()
    qdf['prev_bidsize'] = qdf['bidSize'].shift()
    qdf['prev_askprice'] = qdf['askPrice'].shift()
    qdf['prev_asksize'] = qdf['askSize'].shift()

    qdf.dropna(inplace=True)

    bid_geq = qdf['bidPrice'] >= qdf['prev_bidprice']
    bid_leq = qdf['bidPrice'] <= qdf['prev_bidprice']
    ask_geq = qdf['askPrice'] >= qdf['prev_askprice']
    ask_leq = qdf['askPrice'] <= qdf['prev_askprice']

    qdf['OFI'] = pd.Series(np.zeros(len(qdf)))
    qdf['OFI'].loc[bid_geq] += qdf['bidSize'].loc[bid_geq]
    qdf['OFI'].loc[bid_leq] -= qdf['prev_bidsize'].loc[bid_leq]
    qdf['OFI'].loc[ask_geq] += qdf['prev_asksize'][ask_geq]
    qdf['OFI'].loc[ask_leq] -= qdf['askSize'][ask_leq]

    qdf = set_nanosecond_index(qdf)

    qdf = qdf[['Mid-Price_Change', 'OFI']].resample(dt).sum().dropna()
    qdf.name = 'OFI_df.' + str(dt)

    return qdf

def tfi(trades, quotes, dt):
    tdf = trades.copy()
    qdf = quotes.copy()
    tdf['sgn_size'] = np.where(tdf['side'] == 'Buy', tdf['size'], -tdf['size'])
    qdf['midprice'] = (qdf['bidPrice'] + qdf['askPrice'])/2
    dmid = qdf['midprice'].diff().resample(dt).sum().fillna(0.0)
    tf = tdf['sgn_size'].resample(dt).sum().fillna(0.0)
    assert dmid.count() == tf.count()
    tf.name = 'TFI'
    dmid.name = 'Mid-Price_Change'
    return tf, dmid
```

```

def set_nanosecond_index(df):
    tmp = df.copy()
    tmp.set_index('index', inplace=True)
    tmp.index = pd.to_datetime(tmp.index.values, format='%Y-%m-%dH:%M%S.%f000')
    return tmp

def reset_index(df):
    tmp = df.copy()
    return tmp.reset_index()

def plot_ofi_and_ols_summary(df, dt):
    df.plot(kind='scatter', grid=True,
            x='OFI', y='Mid-Price_Change',
            title = 'XBTUSD_' + dt + '_OFI',
            alpha=0.5, figsize=(12,10))
    ofi_ = sm.add_constant(df['OFI'])
    ols = OLS(df['Mid-Price_Change'], ofi_).fit()
    print(ols.summary2())

def plot_tfi_and_ols_summary(tf, dmid, dt):
    pd.concat([tf, dmid], axis = 1).plot(kind='scatter', grid=True,
            x='TFI', y='Mid-Price_Change',
            title = 'XBTUSD_' + dt + '_TFI',
            alpha=0.5, figsize=(12,10))

    tf_ = sm.add_constant(tf)
    ols = OLS(dmid, tf_).fit()
    print(ols.summary2())

def adf_test(var):
    adf = ts.adfuller(var, 1)
    print('ADF_Statistic: %f' % adf[0])
    print('p-value: %f' % adf[1])
    print('Critical_Values:')
    for key, value in adf[4].items():
        print('\t%s: %3f' % (key, value))

def kpss_test(var):
    kpss = ts.kpss(var)
    print('KPSS_Statistic: %f' % kpss[0])
    print('p-value: %f' % kpss[1])
    print('Critical_Values:')
    for key, value in kpss[3].items():
        print('\t%s: %3f' % (key, value))

```

References

- Abergel, F., & Jedidi, A. (2015). Long-time behavior of a Hawkes process-based limit order book. *SIAM Journal on Financial Mathematics*, 6(1), 1026–1043.
- Aït-Sahalia, Y., & Jacod, J. (2014). *High-frequency financial econometrics*. Princeton: Princeton University Press.
- Antonopoulos, A. M. (2017). *Mastering Bitcoin: Unlocking digital cryptocurrencies*. Newton: O'Reilly Media, Inc.
- Bershova, N., & Rakhlin, D. (2013). The non-linear market impact of large trades: Evidence from buy-side order flow. *Quantitative Finance*, 13(11), 1759–1778.

- Bilokon, P. A. (2018). *Electronic market making as a paradigmatic machine learning and reactive computing challenge*. Working paper.
- Bitcoinalk. (2010). *Pizza for bitcoins?* <https://bitcointalk.org/index.php?topic=137.0>. Accessed 10 Mar 2018.
- Black, F. (1971). Toward a fully automated stock exchange, part i. *Financial Analysts Journal*, 27(4), 28–35.
- Cartea, A., Jaimungal, S., & Penalva, J. (2015). *Algorithmic and high-frequency trading*. Cambridge: Cambridge University Press.
- Chakraborti, A., Toke, I. M., Patriarca, M., & Abergel, F. (2011). Econophysics review: I. Empirical facts. *Quantitative Finance*, 11(7), 991–1012.
- Chan, E. P. (2017). *Machine trading: Deploying computer algorithms to conquer the markets*. Hoboken: Wiley.
- Chordia, T., Roll, R., & Subrahmanyam, A. (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics*, 65(1), 111–130.
- Chu, J., Chan, S., Nadarajah, S., & Osterrieder, J. (2017). Garch modelling of cryptocurrencies. *Journal of Risk and Financial Management*, 10(4), 17.
- Cont, R., & De Larrard, A. (2013). Price dynamics in a markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1), 1–25.
- Cont, R., Kukanov, A., & Stoikov, S. (2014). The price impact of order book events. *Journal of Financial Econometrics*, 12(1), 47–88.
- Dixon, M. (2018). Sequence classification of the limit order book using recurrent neural networks. *Journal of Computational Science*, 24, 277–286.
- Donier, J., & Bonart, J. (2015). A million metaorder analysis of market impact on the Bitcoin. *Market Microstructure and Liquidity*, 1(02), 1550008.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. <http://www.deeplearningbook.org>. Accessed 12 Mar 2018.
- Gould, M. D., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., & Howison, S. D. (2013). Limit order books. *Quantitative Finance*, 13(11), 1709–1742.
- Guo, T., & Antulov-Fantulin, N. (2018). Predicting short-term Bitcoin price fluctuations from buy and sell orders. arXiv preprint [arXiv:1802.04065](https://arxiv.org/abs/1802.04065).
- Halaburda, H., & Gandal, N. (2017). Competition in the cryptocurrency market. Available at SSRN 2506463.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. Springer series in statistics. New York, NY: Springer.
- Heusser, J. (2013). *Bitcoin trade arrival as self-exciting process*. <http://jheusser.github.io/2013/09/08/hawkes.htm>. Accessed 10 Feb 2018.
- Hileman, G., & Rauchs, M. (2017). Global cryptocurrency benchmarking study. *Cambridge Centre for Alternative Finance*. <https://doi.org/10.2139/ssrn.2965436>.
- Huang, W., Lehalle, C.-A., & Rosenbaum, M. (2015). Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509), 107–122.
- Jessen, C. R. (2015). *Implementation and evaluation of an order flow imbalance trading algorithm*. Unpublished MSc Thesis.
- Jiang, Z., & Liang, J. (2016). Cryptocurrency portfolio management with deep reinforcement learning. arXiv preprint [arXiv:1612.01277](https://arxiv.org/abs/1612.01277).
- Johnson, B. (2010). *Algorithmic trading and DMA*. London: 4Myeloma Press.
- Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, 22(1), 109–126.
- Kelly, F., & Yudovina, E. (2017). A Markov model of a limit order book: Thresholds, recurrence, and trading strategies. *Mathematics of Operations Research*, 43(1), 181–203.
- Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3), 967–998.
- Lee, C., & Ready, M. J. (1991). Inferring trade direction from intraday data. *The Journal of Finance*, 46(2), 733–746.
- Madan, I., Saluja, S., & Zhao, A. (2015). *Automated Bitcoin trading via machine learning algorithms* (Vol. 20). <http://cs229.stanford.edu/proj2014/Isaac%20Madan>.
- Morariu-Patrichi, M., & Pakkanen, M. S. (2017). *Hybrid marked point processes: Characterisation, existence and uniqueness*. arXiv preprint [arXiv:1707.06970](https://arxiv.org/abs/1707.06970).

- Osterrieder, J., Chan, S., Chu, J., & Nadarajah, S. (2017). A statistical analysis of cryptocurrencies. <https://doi.org/10.2139/ssrn.2948315>
- Pagnottoni, P., Dimpfl, T., & Baur, D. (2018). *Price discovery on Bitcoin markets*. Available at SSRN 3280261.
- Pasquale, F. (2014). Law's acceleration of finance: Redefining the problem of high-frequency trading. *Cardozo Law Review*, 36, 2085.
- Plerou, V., Gopikrishnan, P., Gabaix, X., & Stanley, H. E. (2002). Quantifying stock-price response to demand fluctuations. *Physical Review E*, 66(2), 027104.
- Russo, C. (2018). *One of the biggest crypto exchanges goes dark and users are getting nervous*. <https://www.bloomberg.com/news/articles/2018-01-12/crypto-exchange-kraken-goes-dark-and-user-anxiety-surges>. Accessed 14 Feb 2019.
- Shah, D., & Zhang, K. (2014). Bayesian regression and Bitcoin. In *2014 52nd annual Allerton conference on communication, control, and computing (Allerton)* (pp. 409–414). IEEE.
- Shen, D. (2015). *Order imbalance based strategy in high frequency trading*. Unpublished MSc Thesis.
- Sirignano, J., & Cont, R. (2018). *Universal features of price formation in financial markets: Perspectives from deep learning*. arXiv preprint [arXiv:1803.06917](https://arxiv.org/abs/1803.06917).
- Steadman, I. (2013). *Study: 45 percent of Bitcoin exchanges end up closing*. <http://www.wired.co.uk/article/large-bitcoin-exchanges-attacks>. Accessed 15 Feb 2018.
- Upson, J., & Van Ness, R. A. (2017). Multiple markets, algorithmic trading, and market liquidity. *Journal of Financial Markets*, 32, 49–68.
- Verhage, J. (2018). *Bitcoin's 43% arbitrage trade is a lot tougher than it looks*. <https://www.bloomberg.com/news/articles/2018-01-09/bitcoin-s-43-arbitrage-trade-is-a-lot-tougher-than-it-looks>. Accessed 14 Feb 2019.
- Wheatley, S., Sornette, D., Reppen, M., Huber, T., & Gantner, R. N. (2018). Are bitcoin bubbles predictable. Combining a Generalised Metcalfe's Law and the LPPLS Model, Swiss Finance Institute Research Paper (18–22).
- Zheng, Z., Xie, S., Dai, H.-N., & Wang, H. (2016). *Blockchain challenges and opportunities: A survey*. Work paper—2016.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.