UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Predictability of return and volatility in Bitcoin markets

Master Degree Project in Finance

Gothenburg University – School of Business, Economics, and Law

Institution: Centre for Finance

Graduate school

Supervisor: Marcin Zamojski

Martin Eimer & Lina Karlsson

Gothenburg, Sweden

Spring 2018

SCHOOL OF BUSINESS, ECONOMICS AND LAW AT THE UNIVERSITY OF GOTHENBURG

Abstract

Graduate School
Master of Science in Finance
Predictability of return and volatility in the Bitcoin market
By Lina Karlsson and Martin Eimer

We study how abnormal liquidity affects the predictive power of returns and volatility in Bitcoin markets. The presence of abnormal liquidity can be explained by price manipulation which is the result from previous studies that found manipulators present in the market. We find that abnormal liquidity has no predictive power for returns but that abnormal liquidity has some predictive power of volatility. We cannot conclude a presence of price manipulators but according to our results regarding volatility there are elements that show irregularities in the markets. Our results further highlight the mechanisms of the market and how traders deal with fees. The results for volatility indicates that the amount of abnormal liquidity have an effect on future volatility which tells us something about how the market is functioning. This study highlights the importance of further exploration of the effects abnormal liquidity has on the Bitcoin market.

Keywords: *Bitcoin, Price Manipulation, Abnormal Liquidity, Spoofing, Limit Order Book, High Frequency Trading*

Acknowledgements

We would like to first and foremost thank our thesis supervisor Marcin Zamojski for his never-ending help, constructive feedback, insightful comments and coding guidance during this process. We would also like to thank Axel Hellström for his technological insight when collecting our data and dealing with the API's.

List of Figures

List of Tables

Abbreviations

API – Application Programming Interface

HFT – High Frequency Trading

AT – Algorithmic Trading

RV – Realized Volatility

CME – Chicago Mercantile Exchange

1. Introduction

The purpose of this paper is to examine suspicious trading in Bitcoin markets and try to understand whether abnormal liquidity in the market predicts returns or volatility. Our approach is to look at sudden changes in liquidity, and try to establish if these are predictive of consequent price changes. We define liquidity as the dollar amount of orders outstanding in the limit order book of an exchange. In this manner, we seek to understand the impact of abnormal liquidity appearing in the Bitcoin market and provide further understanding of existing market mechanisms. Cryptocurrency markets have been in the spotlight for some time now and attracted a great deal of interest from the public, professional investors, and academics. However, due to its recent creation, there is not a great deal of research in the field of Bitcoin and cryptocurrencies more broadly.

When modeling returns and volatility we expect abnormal liquidity to affect predictability of returns and volatility. If the findings are consistent with previous research we can expect both returns and volatility to increase in an event of a spike in liquidity, on the ask side of the order book and decrease, on the bid side of the order book. This can be interpreted as a sign of price manipulation. There is also a possibility that the findings are inconsistent with previous research which means returns and volatility would decrease in an event of a spike in liquidity on the ask side of the order book and increase on the bid side of the order book.

The most recent article regarding price manipulation of Bitcoin was carried out by Gandal et al. (2017). They are able identify specific traders that manipulate the price of Bitcoin during a period of 30 days in 2014. Gandal et al. conclude that these traders use the method of spoofing to affect the price. Spoofing is the practice of posting limit orders that are not intended to be executed instead, these orders act as lure to increase the interest in the asset and in turn impact the price.

Spoofing is not the only way to manipulate the market and there is a wide area of research regarding price manipulation and its many approaches. Allen and Gale (1992) and Allen and Gorton (1991) both present a theoretical model where price manipulation is possible by uninformed traders through irrational trading. Additionally, Massoud et al. (2016) conclude that secret promotions by a hired 3rd party coincide with an increase in price and volume.

A phenomenon that has emerged in later years and affects market efficiency substantially is High Frequency Trading (HFT). Chaboud et al. (2014), Hendershott et al. (2011) and Brogaard et al. (2014) all find that HFT has a positive effect on market efficiency and liquidity. However, Jarrow and Protter (2012) find that HFT can act, unknowingly, as manipulators to the disadvantage of ordinary investors by creating mispricing in the market. Arguably, price manipulation and arbitrage opportunities affect market efficiency. However, markets can be efficient, but still be affected by manipulation. Nadarajah and Chu (2017) find that Bitcoin shows weak market efficiency and are supported by Baur et al. (2017). However, it is unclear if the market show evidence of semi-strong or strong efficiency and this could affect how Bitcoin is affected by price manipulation.

While previous research puts light on historical trading patterns when examining price manipulation, there is a lack of more present research examining the resilience of price manipulation in the Bitcoin market. Therefore, to use more recent data and to explore if abnormal liquidity predicts returns and volatility, which in turn could be interpreted as price manipulation, is of great interest. We focus on discovering spikes in liquidity in the limit order book that are predictive of subsequent returns. To the best of our knowledge this approach has not been considered in previous research. The subject is of high interest not only because of previous findings of price manipulation by Gandal et al. (2017), but also due to the massive general interest that cryptocurrencies have gathered over the last few years.
Our hypotheses to be tested are as follows:

H1: Abnormal changes in liquidity in the limit order book are predictive of subsequent returns.

H2: Abnormal changes in liquidity in the limit order book are predictive of subsequent volatility.

The fact that Bitcoin has created a general interest and is the first thing to come to mind when mentioning cryptocurrencies to a broad mass also makes it an interesting subject. However, the interest in all virtual currencies has steadily increased since the introduction of

Bitcoin in 2008. Among about 1300 cryptocurrencies, Bitcoin is the largest traded cryptocurrency and the market capitalization was roughly 141 billion U.S. dollars as of the May 22nd, 2018 (CoinMarketCap, 2018). The possibility to trade Bitcoin increased when Bitcoin futures were introduced on the world's largest futures exchange, the Chicago Mercantile Exchange (CME), in December 2017. It is now relatively easy to start trading in Bitcoins and other cryptocurrencies. At the same time, there have been several restrictions introduced both by financial institutions and tax authorities; for instance, China banned trading in Bitcoin in 2018 (South China Morning Post, 2018). Additionally, Foley et al. (2018) find that Bitcoin is used as a mean to avoid taxes, launder money, and to make illegal trades.

To examine whether abnormal liquidity has an effect on predictability we need a large amount of data over a selected sample period. We collect a dataset containing limit order book data for a two-week period in 2018 with the best 50 bids and the best 50 asks. This is done for two different trading platforms, Gdax and Bitfinex. The data is sampled every 5 seconds and we control for whether there are any periods of trading that look suspicious by modelling expected level of liquidity and look for relationships with return and volatility. We detect abnormal amounts of liquidities and test them against returns and volatility. We find it interesting to not only examine abnormal liquidity in the full order book but to see where in the order book the abnormal liquidity appears, which we will refer to as levels away from the mid-quote[1]. The different levels represent the accumulated liquidity, for example the liquidity between 1 USD and 2 USD is represented by one level.

We do not find evidence of price manipulation, and in particular of spoofing, in our results from return predictability. We find that the returns decrease when there are abnormal levels of liquidity for asks and increase when there are abnormal levels of liquidity for bids, which is the opposite of what we expect. However, what we find is still of interest as there is no previous research highlighting this which tells us something about how the market functions. The two scenarios, of abnormal liquidity for asks and bids can be explained by the fact that traders use a strategy where they limit their losses by avoiding paying fees. The findings

[1] Mid-quote is the price between the best price of the seller (ask) and the best price of the buyer (bid).

could be explained by the fact that there is no fee on limit orders which is used by the traders to limit their losses on trades that they already expect to make losses on. The results from predicting volatility are more consistent with the hypothesis. We find that abnormal liquidity predicts future volatility on all levels of abnormal liquidity for Gdax. Bitfinex is a more volatile market, and arguably, not as efficient as Gdax which gives us more obscure results. In conclusion, our findings highlight the mechanisms in Bitcoin markets and the results for Gdax imply that the abnormal liquidity in the market has an impact on returns during our sampling period. However, for Bitfinex we cannot draw any general conclusions regarding abnormal liquidity and the power of predictability it has. Our results show the importance of further research in the area.

2. Bitcoin and Blockchain Technology

In this section, we introduce Bitcoin and its development during the last year along with a brief explanation of the blockchain technology behind Bitcoin.

During 2017, the price of Bitcoin increased, from 1,020 USD per bitcoin on January 5, 2017 to 19 498 USD on December 18, 2017, when Bitcoin price has reached its peak. That is an increase of more than 1800% in less than one year. From the 18th of December 2017 until the 6th of February 2018 the market price had fallen by 69% to 7701 USD per unit (see Figure 1 in Appendix A.1). Despite this, the interest in Bitcoin is growing larger and more and more Bitcoins are being traded (CoinMarketCap, 2018). The increasing interest in cryptocurrencies along with a high volatility in the price makes it important to understand how the markets for Bitcoin functions.

The rise of Bitcoin began in 2008 when a single individual (or a group of people) under the pseudonym Satoshi Nakamoto released a paper titled “Bitcoin: A peer-to-peer electronic cash system” (Nakamoto, 2008). This paper laid the groundwork for the blockchain technology that enabled the creation of Bitcoin and other cryptocurrencies. The blockchain technology permits, allegedly, safe transactions between individuals without using financial institutions as intermediaries. This does not only apply to transaction of funds but to all information sent in a peer-to-peer environment. However, this attracts traders that wants to use the technology for ulterior motives due to the ability to hide the information from regulatory bodies. Tax evasion, money laundering, and purchases of illegal services or products are aided by the increased anonymity created by the technology, see Foley et al. (2018). Furthermore, views on Bitcoin and other cryptocurrencies differ greatly, some see them as the most important financial innovation since online transactions (The Guardian, 2016) while others see it as the biggest bubble since the Tulip mania in the 17th century (CNN, 2017). Among its opponents, we find business leaders such as Bill Gates and Warren Buffett as well as researchers, such as Nouriel Roubini who famously predicted the subprime crisis in 2006. However, the most salient aspect of cryptocurrencies in general and Bitcoin in particular, is the blockchain technology it is built on. Although Bitcoin has many uses, some of which are controversial or right out illegal, the blockchain technology is believed to have many potential uses that could be adopted in the future. Among these are digital contracts, voting systems, or supply chain records to name a few.

The blockchain technology can be likened to a ledger. It is essentially a database that contains information of any kind. The ledger is monitored, updated, and shared by all its users. None of its users own it or control it so there must be measures put into place so that its users can trust the information that it contains. The lack of control gives the users the freedom of choosing when or where to send information. In the case of many transfers, they can also do it quickly, without going through financial intermediaries that could take days to execute a transaction. Unlike fiat currencies, where a central body such as a central bank regulates the money market, cryptocurrencies rely on cryptographic technological solutions. This lack of regulatory oversight by a central institution is replaced by a consensus-based mechanism that relies on Proof of Work (POW). The POW can be described as a mathematical problem that has to be answered, for example by solving a random problem through a repetition process that requires computing power. However, when the problem has been solved the POW that follows with it is easy to verify and use as a proof that the block is valid. This mechanism is based on a set of cryptographic techniques that assure protection of the information, (Narayanan et al. 2016). Each new block that is created represents the creation of a new Bitcoin, and the process is called mining.

For a transaction to take place the submitter sends the information to the receiver represented in a block together with other transactions within it. The block is broadcasted to every party in the chain and approved that it is valid which leads to the block being added to the existing chain and the transfer is complete. The information required to verify that a block is legitimate is presented in the accompanying material of the block, as well as information regarding previous transactions with it and the POW. (Crosby et al. 2016)

3. Literature review

In this section, we give a summary of previous research regarding Bitcoin, high frequency trading, price manipulation and predictability of returns.

3.1 Research about Bitcoin and cryptocurrencies

There has not been a lot of research in the field of Bitcoin due to its short life as an asset, but there is a growing interest in cryptocurrencies due to their decentralized nature, and the exciting technology which enables its existence. Böhme et al. (2015) look closer at the blockchain technology and what disruptive powers it possesses towards both the financial and other industries as well as the governance of the different markets. The possible applications are many but the most interesting, and realistic uses, are that of transfers where both the sender and recipient cannot challenge the legitimacy of the information sent. Digital signatures, digital keys, digital contracts and digital stocks and bonds are just some of the examples.

Whether Bitcoin should be treated as a real currency or not is also of interest to researchers and end users. Bitcoin serves today both as a currency and as a financial asset. A currency should fulfil three main criteria (i) it should function as a medium of exchange, (ii) a store of value and (iii) a unit of account. Yermack (2015) argues Bitcoin is a poor performer in all three cases due to its high volatility, hacking and theft risks, and a lack of a regulatory body. He concludes that it behaves more like a highly speculative investment. Additionally, the volatility of the asset highly affects the utility of it as a currency. The relative value can change dramatically from one day to the other, which makes it difficult to use in a daily setting, as Yermack (2015) highlights. However, there exists currencies that have experienced similar inflation/deflation cycles, such as the pre-world war II German Reichsmark and the Zimbabwean Dollar in 2008 and there exist companies that accept it as legal tender which speaks for it being a currency. The Stock Exchange Commission (SEC), which regulates financial markets in USA, and Skatteverket, the tax agency in Sweden, both identify Bitcoin as a financial asset which further speaks against its identification as a currency. Speculation regarding how Bitcoin is used on black markets also creates attention and a need of further investigation (The Guardian 2017). Foley et al. (2018) claims that more than one quarter of all users and half of all transactions recorded on the blockchain are associated with illegal activity. However, they also conclude that the illegal share diminishes with rise in mainstream awareness and the development of other, more obscure, cryptocurrencies.

3.2 Research about market microstructure in general

Areas regarding the value drivers of cryptocurrencies is one of the more examined subjects by researchers. Several researchers have investigated this and they come to similar conclusions. Technological factors, such as the hash-rate, affect the price as well as social aspects, where public interest is shown to be significant by both Kristoufek (2015) and Li and Wang (2016). The hash-rate determines how quickly the blocks are produced at a given difficulty level, which is decided by how many people are trying to mine. This means that the higher the amount of computational power that is trying to mine Bitcoin the more difficult it is to create a new block of Bitcoin. Urquhart (2018) further focus on public interest and looks closer at what draws attention to the Bitcoin markets. He finds that volume and realized volatility are the main drivers of attention. Both Kristoufek (2015) and Li and Wang (2016) find that transaction volume and the trade exchange-ratio affect the price of cryptocurrencies.

O'Hara (2003) argues that the price of an asset is not only affected by its underlying fundamentals, but is also related to indirect costs of the marketplace it is present in. These indirect costs are associated with the characteristics of the marketplaces such as how liquid it is and what restrictions are present. According to this, a marketplace with higher liquidity will lead to a more efficient price discovery and therefore reduce the indirect costs of a slow price discovery. Symmetric information-based pricing models, such as CAPM, do not take liquidity into consideration and are therefore flawed according to O'Hara and she further advocates that the characteristics of market place will affect the asset price instead of only letting the price be a product of the underlying information, she states that "asset prices evolve in markets" (O'Hara, 2003, p.1). A solution to this problem could be to add a liquidity premium that accounts for the missing information. This type of argumentation could be an important factor in finding the price of Bitcoin, due to its lack of underlying fundamentals, and the liquidity, it could act as an important part in predicting future price.

Due to the increase of technology used in and around markets the new phenomenon, algorithmic trading has been introduced, enabling computers to execute trades after a set pattern, also referred as High Frequency Trading (HFT). However, it should be noted that not all algorithmic trading is HFT. One early article on this phenomenon is Jain (2005) who looks at exchanges in 120 countries and examines the impact of early stage automation, which is not equal to fully functioning HFT but a step in its direction. He finds that the equity premium in the CAPM decreases, especially in emerging markets. He also finds that there is a positive, short-term, effect on the price from the switch to algorithmic trading and that liquidity is enhanced by the automation. The increased liquidity in turn leads to lower cost of equity due to increase of information. High frequency trading has increased liquidity on markets and further increased the importance of investigating the effects liquidity has on price discovery.

In the track of algorithmic trading (AT), Chaboud et al. (2014) show that the increase of this type of trading has increased market efficiency through reducing arbitrage opportunities and increasing the speed of price discovery. Hendershott et al. (2011) concludes that AT narrows the bid-ask spreads, reduces adverse selection, and increases liquidity, especially for large stocks. Brogaard et al. (2014) follow up and look at high frequency traders and their role in creating more efficient markets. They conclude that the HFT create a higher level of efficiency in the market by trading in the direction of permanent price changes and off-setting the transitory pricing errors. These transitory pricing changes can be anomalies in the market that have no real support and therefore should not exist, and HFT helps battle these anomalies. Boehmer et al. (2015) come to a similar conclusion; to their mind, algorithmic trading increases liquidity and informational efficiency and in turn also trading volume on markets. However, this increase in trading volume is not observed in "good" markets but instead in markets that are of lesser quality, such as OTC. However, Jarrow and Protter (2012) find that this type of high frequency traders can create a mispricing of assets and exploit ordinary traders. Their paper highlights the fact that HFT may play a dysfunctional role in terms of mispricing generated via mutual and independent actions of high frequency traders.

3.3 Research about price manipulation

Previous research on price manipulation often looks at stocks with a low liquidity or stocks traded in Over the Counter (OTC) markets. Aggarwal and Wu (2006) examine the US SEC litigations against market manipulators and find that low liquidity and small companies that are subject to manipulation see increases in price, trading volume, and return volatility during the manipulation period and that this behaviour reverses quickly thereafter. Investigating the subject of price manipulation further, Massoud et al. (2016) look at OTC companies that hire promoters to engage in endorsing the stock, where they find that these promotions coincide with insider trading. Allen and Gale (1992) develop a theoretical framework in which uninformed speculators can affect the price of a stock. Their trade-based model is consistent with a rational, utility-maximizing individual. This means that large trades, seemingly without any fundamental change in the underlying value, can increase prices and therefore also be part of a manipulation process. Allen and Gorton (1991) look at the possibility of price manipulation where traders do not have any inside information or other type of edge over other traders. They conclude that it is possible for traders to manipulate the price by using the theories developed by Allen and Gale (1992) and further the research by testing them with the Glosten-Milgrom model[2], where they find that not only is the theory plausible but also show that there is an equilibrium level of manipulation. Just as Allen and Gale (1992), they conclude that a trade-based manipulation is possible under otherwise natural conditions. Notable is that, Allen and Gorton (1991) and Allen and Gale (1992) are purely theoretical studies and do not provide any empirical evidence.

Since the fall of the trading venue Mt. Gox, in 2014, there has been an increasing interest in if and how Bitcoin has been manipulated. Data from the exchange has since been leaked making analyses of trading and account activity possible. Gandal et al. (2017) argue that two bots on Mt. Gox have pushed the price up from 150 USD to 1000 USD, through the method

[2] Glosten-Milgrom develop a model where the bid ask spread arises from adverse selection from insider traders. They conclude that excess return in small firms is a result of insider trading in periods before positive news.

of spoofing. Spoofing is a systematic method where a spoofer places an abnormally large limit order on the market without the intention of filling it but with the intention of increasing or decreasing the price. This gives an illusion of a very liquid market. The spoofer uses the fact that traders on the market have different connection speeds and makes sure that they have better connection to the markets to be able to spoof. The higher connection speed enables them to quickly put orders and withdraw them from the markets without having them executed, creating a false picture for the average trader that are slow to react. The spoofer places an abnormally large order and often cancels it shortly thereafter in hope for a trader to take the bait and place a larger market order that reflects the spoofers offer. Right after the spoofer has withdrawn the offer he/she places a new, higher ask and waits for the order to get executed. If the spoofer succeeds, the spoofer has made a profit, the price increases and the trader had to pay additional transaction costs in the form of a larger bid-ask spread. Hence, spoofing is used to disorient the traders on the one hand and to control the market on the other (Lee et al. 2013).

Previous research concerning stock manipulation has also been heavily affected by the importance of liquidity for market efficiency. One approach is taken by Cumming et al. (2009) who look at trading rules and market liquidity. They find that stricter regulation of market manipulation and insider trading significantly affects the liquidity of stocks. The importance of liquidity in a market is a topic that has been widely discussed and its importance in the price discovery is broadly recognized.

3.4 Research about predictability of returns in high frequency

Yuferova (2017) looks at intraday return predictability in the stock market. She finds that limit orders are a key source for intraday return predictability by informed traders, and hence informed traders more often act as liquidity providers rather than liquidity demanders. She also finds that increased algorithmic trading and high frequency trading is associated with an increase in market orders rather than limit orders.

Instead of predicting returns, which is often difficult, it is possible to predict volatility. French et al. (1987) show that volatility is a predictor, in the way that there is a positive correlation between the risk premium of common stocks and volatility. Qiu et al. (2015) further examines this correlation and concludes that there exist a negative correlation between returns and volatility for the German stock exchange DAX and Chinese Indices. However, if the correlation between returns and liquidity is applicable for Bitcoin is not explored, to our knowledge.

Whether the market is efficient is of interest to researchers when exploring return predictability. If market efficiency cannot be established predictability from tests will be influenced by the inefficiency and it would therefore be difficult to draw any conclusions from the output. Urquhart (2016) runs several different statistical tests, both different models and over different time frames. He concludes that the Bitcoin market shows no sign of efficiency. In contrast, Baur et al. (2017) are looking at time-of-day, day-of-week, and month-of-year effects for bitcoin returns and conclude that the markets are efficient. By looking for time specific anomalies, they find no evidence of steady or persistent patterns across the period they look at. They mean that this is evidence for weak efficiency in the market. Market efficiency in this case is referring to predictability of the price instead of the asset being correctly valued based on fundamentals. However, this is not enough evidence to establish market efficiency. Nadarajah and Chu (2017) backed up Baur et al. and find evidence of weak market efficiency using different statistical tests and surprisingly include the ones used by Urquhart (2016) when he concluded that the markets were inefficient. Instead of using returns squared, as Urquhart (2016), Nadarajah and Chu (2017) use odd integer power of the Bitcoin returns when looking for efficiency, as they claim that using an odd power integer leads to no loss of information since positive and negative returns keep their original sign in front of the number when put in power of an odd number.

4. Data

In this section, we start by discussing the data gathering process of the two order books and continue with presenting the summary statistics for the data and variables.

4.1 Data gathering process

Bitcoin trading history to the extent that is needed for this study is not readily available and therefore has to be hand-collected over a given period. We collect time-series data on the 50 best bids and asks over a two-week period from two trading platforms: Gdax and Bitfinex. The data is collected every five seconds, twenty-four hours a day. The sample period is Feb

5, 2018 to Feb 19, 2018 and the period from which the data is collected is randomly chosen. The choice of a 5 second sampling frequency is due to technical restrictions as the two markets limit how often data can be collected with their APIs[3]. Limits like these, are common, and abuse is often penalized with denial of access. Choosing a longer interval would not be as precise as we do not expect to be able to detect spoofing in lower frequency data. In the collected order books, there are some missing observations. These missing observations are few and far apart and should not affect the results in our study.

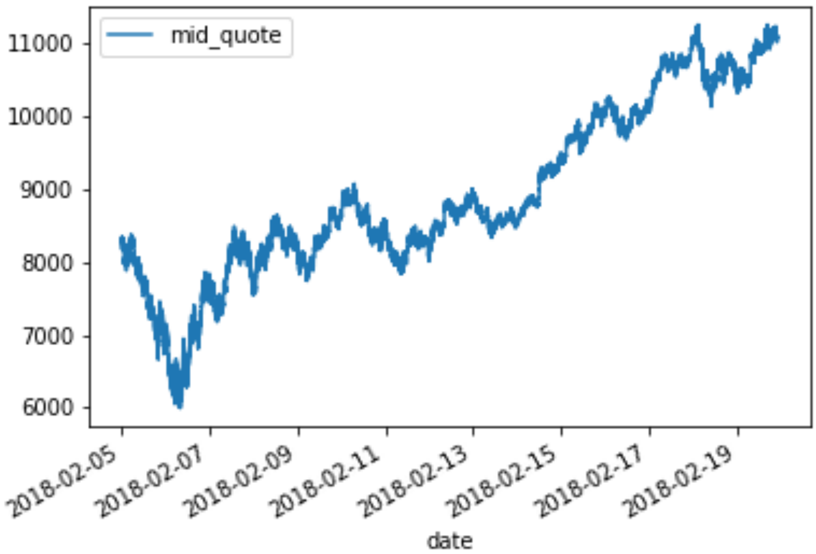
To obtain the time-series data we write a script in Node.js that collects the data continuously over two weeks from Gdax' and Bitfinex's APIs (Appendix A.2). The raw data is saved into a MongoDB database. The data includes a timestamp, price and quantity for every bid and ask. There are in total approximately 250,000 observations, each containing 50 best bids and 50 best asks. After the limit order book is collected the data is loaded in Python where we create several dummy variables to be used in our regressions (Appendix A.2).

As seen in Figure 4.1 there is an upward trend in the mid-quote (the price at where the bid and ask meet) during the time frame. To the best of our knowledge, there were no significant developments or announcements during the sample period. The increase in the price during this period could partly have been affected by the fact that a single trader raised his or her stake from 55,000 to 96,000 Bitcoins, worth 400 million USD, from the 9th of February to the 12th of February (Fortune, 2018). Additionally, prior to the sampling period there was a sharp decline in the price. This might be because many regulatory authorities and banks are attempting to ban trading with Bitcoin. For example, several major banks in America decided to prohibit their customers to buy Bitcoin with their credit cards during this period and China decided to ban foreign Bitcoin trading websites (Veckans Affärer, 2018).

[3] An application programming interface (API) is a set of functions or procedures to that allow interaction between a service and applications created by users.

Figure 4.1: Price trend during the sampling period

In this figure, the increase in Bitcoin price over the sampling period is illustrated. The price is expressed in USD, defined as mid-quote. The price moves from approximately 8000 USD at the first day of data sampling to approximately 11,000 USD on the last day.



The liquidity varies a lot over the day and below in Figure 4.2 and Figure 4.3 we illustrate an example of how the limit order book looks like. There is one example for Gdax and one for Bitfinex for how our time-series data is collected. The liquidity is represented on the x-axis and the different levels away from the mid-quote on the y-axis. On each level, there is the accumulated liquidity up to that level away from the mid-quote. For example, if the mid-quote is 8222 USD for one Bitcoin, then the first level of 0.01 USD away from the mid-quote would be 8222.01 for the ask and 8221.99 for the bid. The liquidity is therefore the accumulated liquidity up to 8222.01 and respectively up to 8221.99. We describe more in detail how we use these different levels to find abnormal liquidity in section 5.1.

Figure 4.2: Limit order book for Gdax

This chart illustrates how the limit order book for Gdax looks at 2018-02-05 00:07:30. On the x-axis we have the liquidity in USD and on the y-axis we have the different levels away from the mid-quote where 0 represent the mid-quote. The mid-quote at this time was 8222.9 USD. We have bids to the right and asks to the left in the graph.

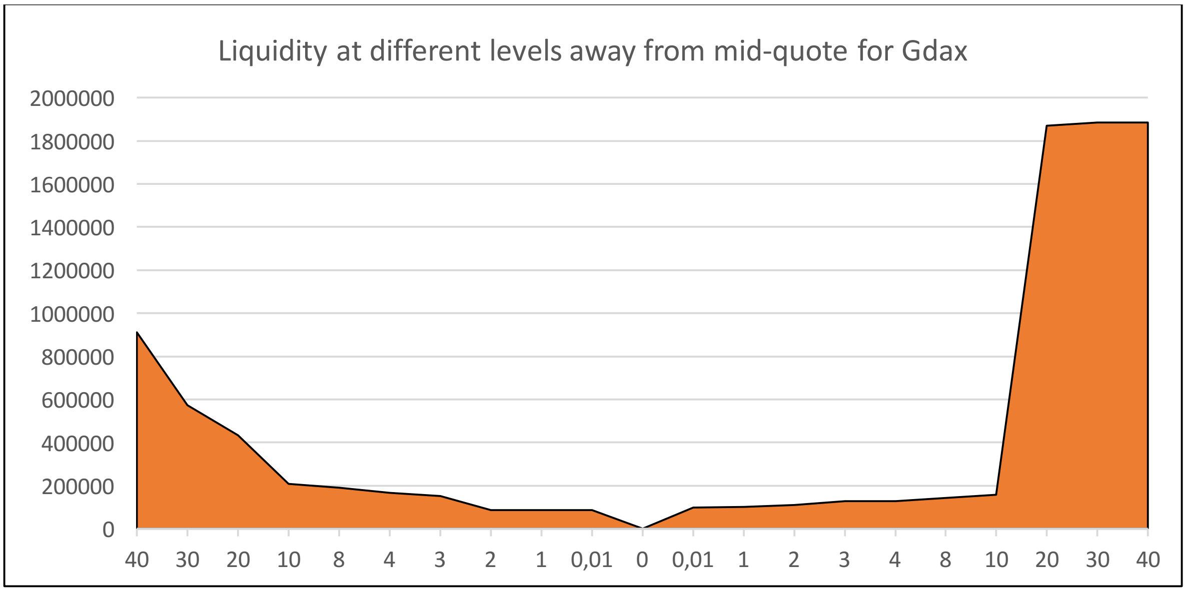
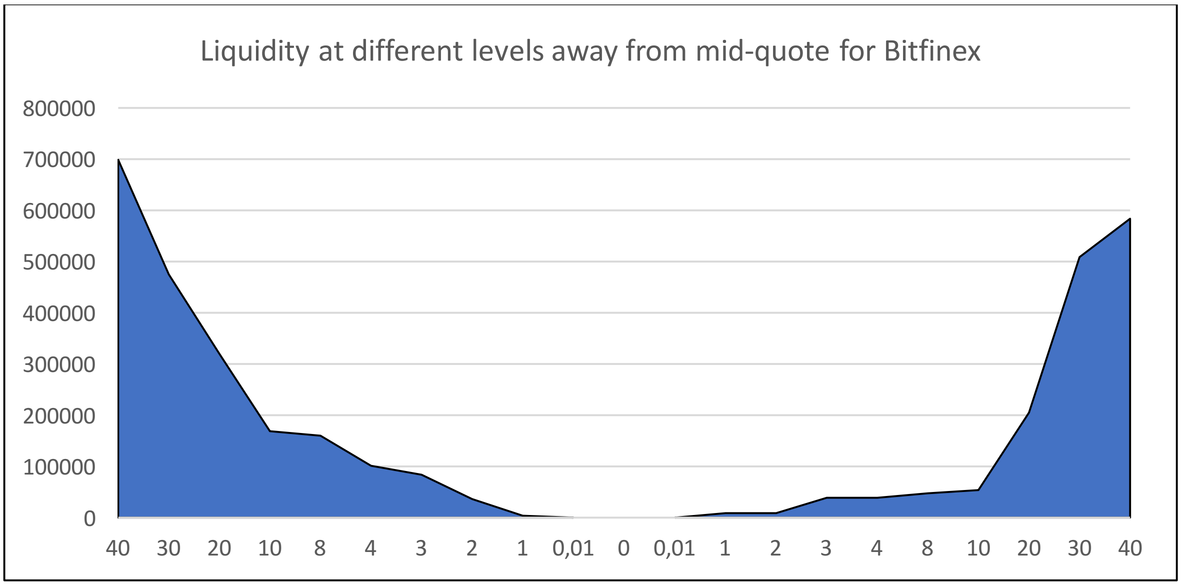


Figure 4.3: Limit order book for Bitfinex

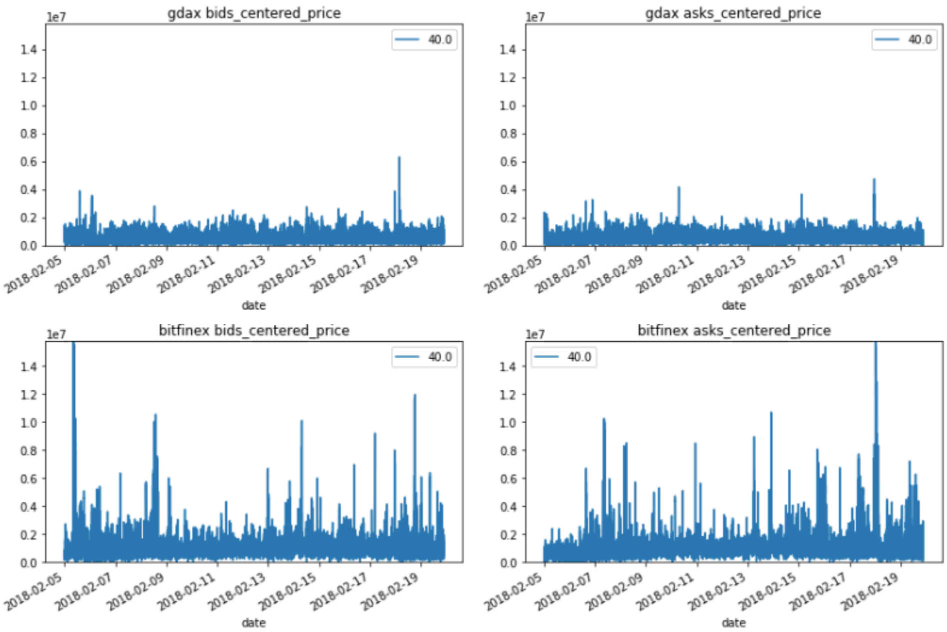
This chart illustrates how the limit order book for Bitfinex looks at 2018-02-05 00:07:30. On the x-axis we have the liquidity in USD and on the y-axis we have the different levels away from the mid-quote where 0 represent the mid-quote. The mid-quote at this time was 8280.9 USD. We have bids to the right and asks to the left in the graph.



The collected data from the given sampling period shows that liquidity differ a lot between Gdax and Bitfinex. Of the two venues, Bitfinex is the more volatile market. In Figure 4.4 the liquidity is illustrated for the two trading platforms for both bids and asks from the sampling period.

Figure 4.4: Liquidity level during the sampling period

These graphs illustrate the accumulated liquidity up to 40 USD from the mid-quote for Gdax and Bitfinex for both bids and asks, where centred price represent the price at that time.



4.1.1 Gdax and Bitfinex

This study is limited to data from only two trading platforms due to time constraint in processing and obtaining the data. The choice of trading platforms was decided by the availability, user friendliness, and access of their public API. Other platforms either do not have the ability of collecting data or they require investments in their products to gain access. Another factor that also impacted our choice of platforms was that they differ in how they are perceived by the public. Bitfinex has come under scrutiny due to allegations of breaches. It has, for example, been accused of laundering money (Bloomberg, 2018). In contrast, Gdax is considered as one of the most stable platforms for cryptocurrencies as they are embracing discussions with regulators and try meet regulatory requirements. They also aim to attract sophisticated and professional traders and focus on having advanced trading features (The balance, 2018). Bitfinex is the more volatile market as seen in Figure 4.4 while Gdax is a more stable market.

Note that there is a considerable difference in the setup of the two venues. In particular, the trading fees differ between these two exchanges which could have an impact on the trading activity. Both venues use a taker-maker fee model were a maker provides liquidity and a taker consumes liquidity from the market. Takers trade directly at the sell/buy price while makers put orders with a set limit price, so called limit order, and wait until their offer is accepted or withdraw it when they have lost patience. Gdax applies a 0% maker fee and between 0.1% to 0.3% taker fees based on the customers 30 day USD-equivalent trading volume. Bitcoin deposits and withdrawals are free and USD wire deposits are 20 USD and withdrawals are 25 USD (Gdax, 2018). In contrast, Bitfinex applies between 0% and 0.1% maker fee and between 0.1% and 0.2% taker fee based on the customers 30 day USD-equivalent trading volume. Bitcoin deposits are free if they are larger than 1000 USD equivalent otherwise 0.04% and Bitcoin withdrawals are 0.04% and USD Wire deposits are 0.1% (min 20 USD) and withdrawals are 0.1% (min 25 USD) (Bitfinex, 2018). It is free to have an account on both venues. Gdax and Bitfinex give us two sources that we believe may deliver different outcomes.

4.2 Summary statistics

Below in Table 4.1 and Table 4.2 we present the descriptive statistics for the most important variables in our models. For Bitfinex we exclude liquidity for the level 0.01 away from the mid-quote as they are zero in all cases. The level 0.01 is one of the chosen levels away from the mid-quote we look at to see where in the order book the abnormal liquidity appears. One noticeable feature regarding Bitfinex and Gdax' differences is the mean of the liquidity over the sample. Up until a certain point, around 8-10 USD from the centred price Gdax has a higher mean for liquidity. This means that around the mid quote Gdax is a market with higher liquidity which could be a product of it being a market with more market makers that trade closer or directly towards the most current offers and the mid quote. Bitfinex also has a substantially higher maximum of liquidity which could be a product of higher volatility. The standard deviation also differs for the two venues. Bitfinex has a substantially larger standard deviation. The data for Bitfinex is also more skewed which would strengthen the hypothesis of it being a more volatile market. It should also be noticed that for the levels of liquidity, both asks and bids, for both markets show evidence of extreme lows in liquidity.

For certain periods the Bid-Ask spread is 80 dollars which signifies a highly illiquid market. This is most evident for Bitfinex. This also applies to the mid-quote for the two markets.

Table 4.1: Summary statistics for Gdax

Table 4.1 presents the characteristics of the chosen variables for Gdax. It presents the number of observations, mean, max and min, standard deviation, skewness and kurtosis for all variables.

Gdax	N	mean	max	min	sd	skewness	kurtosis
mid_quote	257 360	8 917.07	11 299.10	5 873.01	1 210.39	0.13	2.34
Asks 0,01	257 360	74 006.62	3 265 079.55	0.00	94 329.09	3.11	27.39
Ask 1	257 360	93 100.90	3 467 782.90	0.00	115 424.73	3.06	24.53
Ask 2	257 360	107 035.31	3 787 347.35	0.00	133 376.78	3.39	28.32
Ask 3	257 360	118 760.27	3 800 352.82	0.00	145 338.55	3.31	25.24
Ask 4	257 360	130 128.71	4 089 984.26	0.00	154 430.66	3.22	25.19
Ask 8	257 360	183 836.63	4 737 012.89	0.00	195 503.66	3.21	28.69
Ask 10	257 360	213 886.02	4 737 012.89	0.00	210 875.28	2.83	22.43
Ask 20	257 360	364 206.81	4 737 012.89	0.00	257 872.35	1.92	11.86
Ask 30	257 360	468 292.77	4 737 012.89	290.13	279 152.82	1.57	9.08
Ask 40	257 360	515 950.35	4 737 012.89	388.26	290 160.98	1.39	7.88
Bids 0,01	257 360	74 465.78	2 417 214.91	0.00	93 416.91	2.63	17.26
Bid 1	257 360	92 974.21	2 419 324.10	0.00	115 329.72	2.79	18.61
Bid 2	257 360	104 779.29	2 491 125.74	0.00	130 428.61	3.15	23.23
Bid 3	257 360	114 781.43	2 491 349.59	0.00	140 022.67	3.04	21.26
Bid 4	257 360	124 568.65	2 492 541.62	0.00	148 504.72	2.98	20.27
Bid 8	257 360	169 663.72	3 289 008.35	0.00	180 289.20	2.56	15.12
Bid 10	257 360	195 814.90	3 289 008.35	0.00	194 819.96	2.30	12.38
Bid 20	257 360	334 376.10	3 880 563.21	0.00	247 428.48	1.76	9.31
Bid 30	257 360	444 198.02	5 509 265.97	0.22	276 812.31	1.66	9.69
Bid 40	257 360	497 174.17	6 292 196.12	135.39	290 955.29	1.58	10.14
Return_5sec_gdax	257 359	0.000	0.016	-0.014	0.001	-0.182	45.467
Return_10sec_gdax	128 679	0.000	0.016	-0.014	0.001	-0.036	28.984
Return_10min_gdax	2 126	0.000	0.049	-0.037	0.008	0.407	8.027
60secRet_5minRV	21 436	0.004	0.050	0.000	0.003	2.584	16.445
60secRet_10minRV	21 435	0.006	0.054	0.000	0.005	2.531	14.596

Table 4.2: Summary statistics for Bitfinex

Table 4.2 presents the characteristics of the chosen variables for Bitfinex. It presents the number of observations, mean, max and min, standard deviation, skewness and kurtosis for all variables.

Bitfinex	N	mean	max	min	sd	skewness	kurtosis
Mid-quote	257 339	8 912.88	11 249.50	6 000.05	1 192.66	0.16	2.30
Ask 1	257 339	42 990.20	7 539 309.80	0.00	174 505.82	19.91	590.80
Ask 2	257 339	61 054.86	7 561 031.32	0.00	203 758.66	16.31	394.94
Ask 3	257 339	75 467.44	8 006 607.85	0.00	218 739.20	14.81	333.93
Ask 4	257 339	90 011.04	8 006 607.85	0.00	234 073.69	13.40	274.87
Ask 8	257 339	156 074.15	9 257 656.63	0.00	305 199.01	11.30	206.55
Ask 10	257 339	196 194.30	10 046 222.27	0.00	341 906.45	10.26	174.48
Ask 20	257 339	474 076.93	16 959 313.97	0.00	549 450.44	6.22	69.87
Ask 30	257 339	828 761.65	17 368 810.93	0.00	734 583.03	5.87	66.14
Ask 40	257 339	1 108 813.88	20 688 207.96	0.00	890 946.05	6.09	65.80
Bid 1	257 339	46 358.09	16 110 114.84	0.00	213 725.38	34.59	1 772.75
Bid 2	257 339	65 950.52	16 142 644.51	0.00	239 857.69	29.26	1 327.72
Bid 3	257 339	82 385.53	16 145 827.69	0.00	256 801.12	26.14	1 091.49
Bid 4	257 339	98 826.59	16 146 601.23	0.00	272 548.52	23.32	891.38
Bid 8	257 339	169 720.03	16 225 107.89	0.00	334 926.28	16.59	484.94
Bid 10	257 339	210 836.49	16 303 253.37	0.00	365 340.71	14.41	375.62
Bid 20	257 339	478 432.07	17 942 757.84	0.00	558 305.24	10.83	214.92
Bid 30	257 339	818 124.50	18 141 922.82	0.00	662 169.07	8.75	148.22
Bid 40	257 339	1 069 135.00	18 141 922.82	0.00	703 188.89	8.19	126.82
Return_5sec_bitfinex	257 338	0.000	0.024	-0.014	0.001	0.355	62.591
Return_10sec_bitfinex	128 669	0.000	0.033	-0.014	0.001	0.435	46.375
Return_10min_bitfinex	2 144	0.000	0.053	-0.032	0.007	0.365	7.585
60secRet_5minRV	21 434	0.004	0.036	0.000	0.003	2.236	12.161
60secRet_10minRV	21 434	0.006	0.040	0.000	0.004	2.159	10.559

5. Methodology

In this section, we present the methodology we use to see if there is evidence that abnormal changes in liquidity in the limit order book are predictive of subsequent returns. In particular, the chosen models and variables are presented in detail. Overall, our methodology can be divided into two parts. The first part consists of finding abnormal liquidity in the sampling period and from this create dummy-variables. The second part consists of running several regressions to investigate if abnormal liquidity can predict future returns and volatility. We look at both asks and bids to see if prices move in any direction.

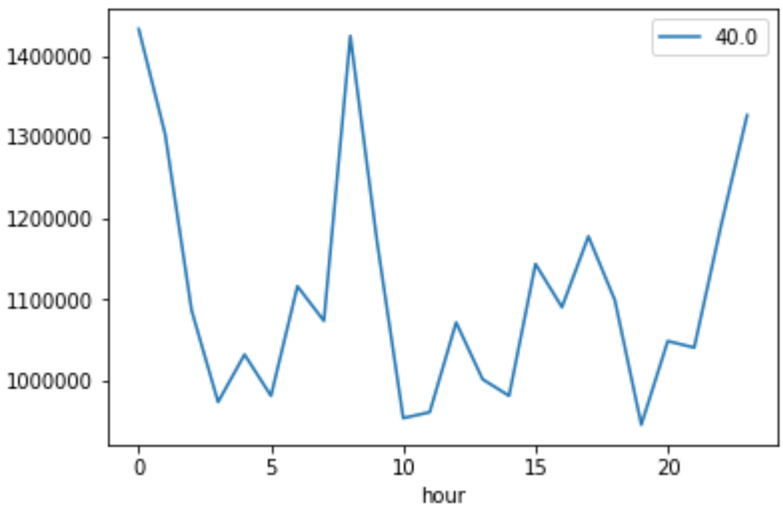
5.1 Abnormal liquidity and outlier detection

When modeling returns and volatility we expect abnormal liquidity to affect predictability of returns and volatility. We find it interesting to not only examine abnormal liquidity in the full order book but to see where in the order book the abnormal liquidity appears, which we will refer to as levels (distances) away from the mid-quote. The different levels represent the accumulated liquidity up to the each chosen level away from the mid-quote. We want to identify at which levels we have abnormal liquidity that can predict returns and volatility.

The first step is therefore to choose ten USD levels away from the mid-quote from where we model liquidity. The ten arbitrarily chosen levels are 0.01, 1, 2, 3, 4, 8, 10, 20, 30 and 40 USD from the mid-quote and we will test for abnormal liquidity up to each of these levels.

We start by looking for patterns, trends and other persistence over time in the data to find out what is normal liquidity. In Figure 5.1, we show average liquidity throughout a day for asks 40 USD above the mid-quote. We see that the liquidity differs greatly corresponding to which hour of the day it is. The average liquidity differs from that in the evening. We conclude from this that we have diurnal effects in our data that need to be controlled for when deciding what an abnormally high liquidity on the bids and asks is. What qualifies as an outlier depends on at what USD level from the mid-quote we are looking at, and we will most likely get different number of outliers at the ten chosen levels of liquidity.

Figure 5.1: Average liquidity over one day during the sampling period
This figure is an example of how the liquidity differ over the day. It shows the liquidity for asks at Bitfinex for the level 40 USD from the mid-quote expressed in thousand dollars.



Consequently, to estimate the normal level of liquidity for every hour of the day we need to control for the diurnal effects. If we would not control for these effects we would falsely classify observations as outliers and not take into consideration the normal changes in liquidity that occur over the day. Therefore, we include dummies for every hour of the day in our model which will correct for this. Additionally, we model the deseasonalized residual as an AR(1) process. We run regressions for each trading platform, for both bids and asks and for the ten different levels that we have selected. The time-series regression model (1) we will use for predicting the normal liquidity is as follows:

$$Liq_{t,k,l,x} = \beta_1 Liq_{t-1,k,l,x} + \beta_i Hour\_dummies_{t-1,k,l,x} + \varepsilon_t \ (1)$$

$$t = time, k = trading\ platform, l = level\ away\ from\ mid - quote, x = bids/asks, i = hour\ of\ the\ day$$

The dependent variable Liq_t represents the liquidity. The independent variables are the lagged liquidity Liq_{t-1} and 24 dummy variables for each hour of the day. We will predict the outliers from the residuals in this model. In the next part, we will tackle how to predict future return and volatility.

Model (1) gives us the normal level of liquidity and we will run this model for both trading platforms, for both bids and asks and at all ten levels. Resulting in 40 regressions in total. We assume normally distributed errors. For outliers, we look at the residuals and apply a one-sided 95% confidence interval test. The residuals that lie outside the interval are considered outliers. We expect to see a decreasing pattern of the distribution of the residuals, i.e. we expect many outliers close to the mid-quote and fewer outliers further away from the mid-quote. This is due to the nature of spoofing where the spoofer would want to lie close to the mid-quote but not too close due to the risk of the order being executed. Our residual dummy will take on the value 1 if there is shown to be an outlier and 0 if there is not an outlier. The outliers will convert into the dummy-variables for asks and bids we will use in model (2) and (3). Additionally, we will choose to only keep the first outlier at every observation. Meaning that if we have an outlier on several levels at the same observation, then we only keep the one that is on the lowest level as an outlier on the first level would affect the accumulated liquidity at all other levels. In summary, the outliers are what we identify as abnormal levels of liquidity, given the hour of the day in which they have been identified.

5.2 Return and volatility predictability

The next step in our model is to see if abnormal levels of liquidity predict return and/or volatility. We do this by using an AR(1) model where we additionally include our dummies for outliers. Predicting future returns is difficult, the amount of data required is large and the model which is chosen is also of high importance. One way to battle this is to instead try and predict future volatility, due to it being easier to predict. As shown by French et al. (1987) volatility can act, on itself, as an efficient predictor for returns. If this is applicable for Bitcoins has however not been tested to our knowledge, and since there exist market efficiency, at least according to Baur et al. (2017), and Nadarajah and Chu (2017), it would not hurt to further look at this aspect. We will use an AR(1) process to capture the essentials of past values impact and estimate the returns and volatility with an OLS regression. Consequently, we need to decide at what frequency the dependent variables return and volatility should be measured at. We will look at 5 second returns, 10 second returns, 10 minute returns, 5 minutes realized volatility based on 1 minute returns and 10 minutes realized volatility based on 1 minute return. The frequencies are chosen from the fact that we assume that abnormal liquidity will influence both returns and volatility within 10 minutes. We will run both regressions for the different frequencies to see if the results differ depending on which horizon we use when measuring return and volatility.

The time series regression model (2) for return predictability is defined by:

$$Ret_{t,k,l} = \beta_1 Ret_{t-1,k,l} + \beta_2 Bid_{t-1,k,l} + \beta_3 Ask_{t-1,k,l} + \varepsilon \ (2)$$
$$t = time, k = trading\ platform, l = level\ away\ from\ mid-quote$$

The dependent variable $Ret_{t,k.l}$ is Return. Return is calculated based on the mid-quotes with log-returns which gives a better estimate when dealing with time-series. We assume that the price is normally distributed. The independent variables are $Ret_{t-1,k.l}$, $Bid_{t-1,k.l}$ and $Ask_{t-1,k.l}$. $Ret_{t-1,k.l}$ is the lagged return by one period. $Bid_{t-1,k.l}$ and $Ask_{t-1,k.l}$ are the dummy-variables indicating if we have abnormal liquidity at the given period.

With this model (2), we test our first hypothesis that abnormal changes in liquidity in the limit order book are predictive of subsequent returns. Hence, the model gives us the impact abnormally large bids and asks have on returns.

We use the following time series regression model (3) to predict volatility:

$$Vol_{t,k,l} = \beta_1 Vol_{t-1,k,l} + \beta_4 Bid_{t,k,l} + \beta_5 Ask_{t,k,l} + \varepsilon \ (3)$$

$$t = time, k = trading\ platform, l = level\ away\ from\ mid - quote$$

The dependent variable $Vol_{t,k,l}$is volatility estimated using Realized Volatility (RV) which is one method among many competing measures of volatility. We chose to use RV as it is a widely-used method to estimate volatility and is a good measurement for high frequency data (Andersen and Benzoni, 2008). However, when using RV it is important to not have too high frequency in your data. A too high frequency would result in data contaminated by market microstructure noise and will not capture true variation in the price but capture other effects that are due to market mechanisms (Bandi & Russell, 2004). We will therefore use the returns over 60 seconds since we need a frequency that fits and is of a reasonable period. RV is calculated from the following formula:

$$\sqrt{RV_t} = \sqrt{\sum_{k=1}^{n} r_{t,k}^2}$$

$$t = time, k = trading\ platform$$

RV is calculated by summing the squared returns and taking the square root. The independent variables are $Vol_{t-1,k,l}$ which is the lagged return by one period, $Bid_{t,k,l}$and $Ask_{t,k,l}$ are the dummy-variables indicating if we have abnormal liquidity at the given period.

This model (3) tests our second hypothesis that abnormal changes in liquidity in the limit order book are predictive of subsequent volatility. Hence, the regression gives an estimate on volatility where we estimate the impact of bids and asks in the previous period.

Running the above regressions for both Gdax and Bitfinex, for all chosen levels, and for all chosen frequencies, results in 60 regressions for returns and 40 regressions for volatility. We will analyse the results from all regressions systematically and look for patterns and significant results that will tell us something about how abnormal liquidity affects the markets and that could possibly be interpreted as spoofing or price manipulation.

6. Results

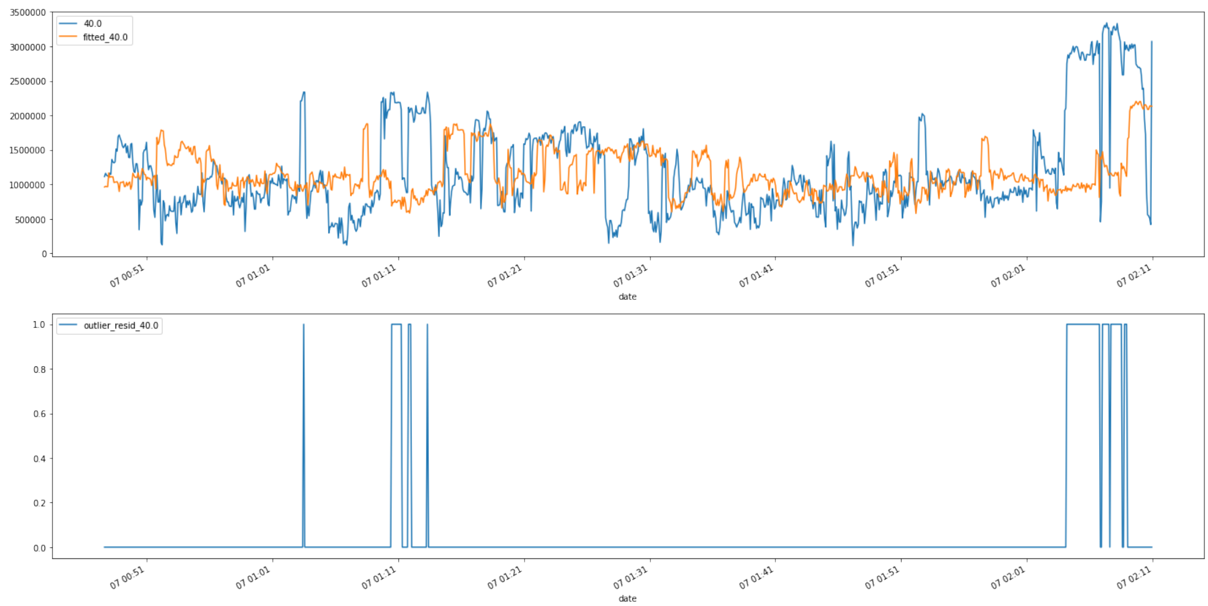
This section presents the results with which we explore the predictability of returns and volatility in the two Bitcoin markets, Gdax and Bitfinex. We employ OLS regressions on the two different markets and explore whether we can say something about predictability and spoofing being present in the two markets. Section 6.1 evaluates the outlier detection process and section 6.2 presents the results from the OLS regressions of return and volatility predictability.

6.1 Outlier detection

From the outlier detection process of localizing the abnormal levels of liquidity, we get the results from model (1) which gives us the outliers. We get 38 different regression tables over the two markets, bids and asks, and different levels from the mid-quote. The results are partly presented in Appendix A.3. The coefficients for the hour-dummies are the interesting parts and gives us which hour of the day has the largest liquidity for each level of liquidity from the mid-quote. A large coefficient indicates a high level of liquidity on that hour of the day, on average. In figure 6.1 we can see an example of how the liquidity behaves over a period of a day for one of the levels from the mid-quote and how our fitted model follows and detects outliers.

Figure 6.1: Modelled liquidity & detected outliers

The first figure display for a few minutes during Feb 7th , 2018, with the blue line, levels of liquidity for bids on Bitfinex at the level 40 USD from the mid-quote, and with the yellow line, our fitted AR(1) model. The second figure shows the detected outliers are during this period.



From the results from model (1) we find the normal level of liquidity for each hour of the day and by applying a one-sided 95% confidence interval we get the outliers. The number of outliers and at what level they occur is illustrated below in Figure 6.2, 6.3, 6.4 and 6.5.

Figure 6.2: Number of outliers for bids on Gdax on each level.

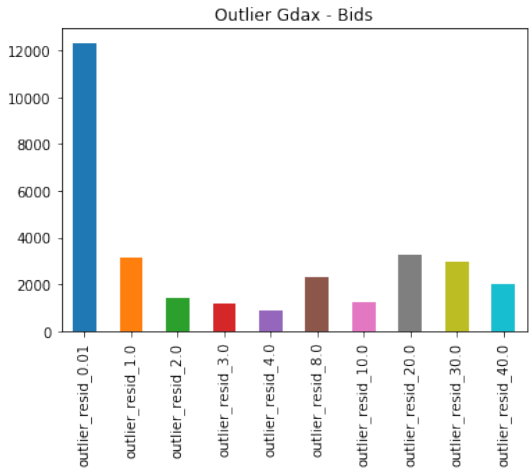


Figure 6.3: Number of outliers for asks on Gdax on each level.

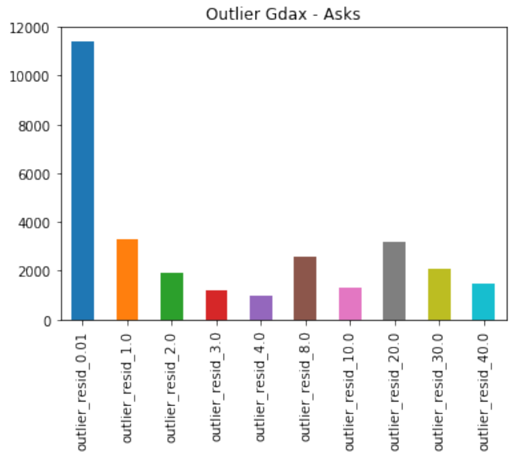


Figure 6.4: Number of outliers for bids on Bitfinex on each level.

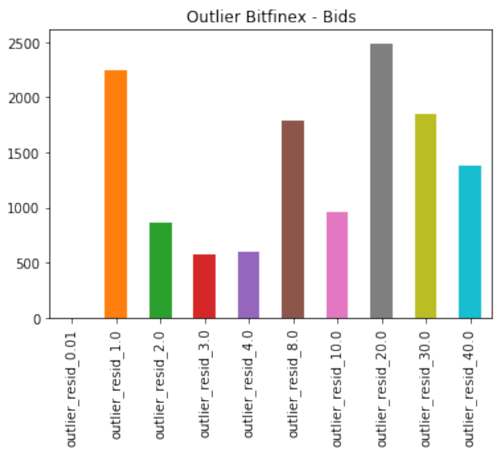
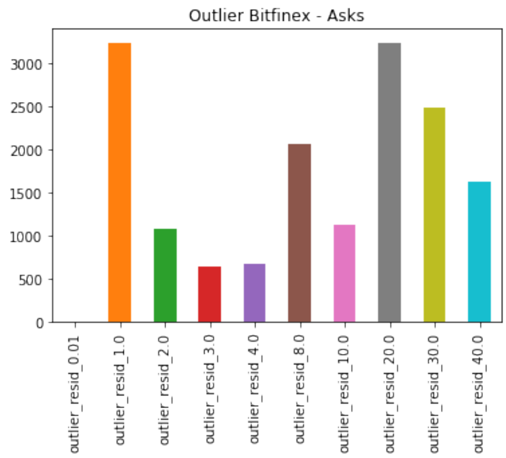


Figure 6.5: Number of outliers for asks on Bitfinex on each level.



We can see for Gdax in Figure 6.2 and Figure 6.3 that there are many outliers at the level of 0.01 while we don't have any outliers at all at the same level at Bitfinex in Figure 6.4 and Figure 6.5. One reason for this could be that Bitfinex is a more volatile market than Gdax. Except for the first level we have the same pattern for Bitfinex and Gdax with a lower number of outliers at the levels 2-4 and 10 USD from the mid-quote while we have a higher number of outliers at the levels 1, 8 and 20-30 USD from the mid-quote. We expected to see a decreasing number of outliers the further away from the mid-quote we examine. A reason why we see many outliers at the higher levels from the mid-quote might be because some traders expect the market to move in that direction or that people tend to put quotes on round numbers.

6.2 Return predictability for Gdax and Bitfinex

Below we present the results and analysis of the regressions for return predictability. We run three different regressions where return on a 5-second, 10-second and 10-minute basis have been set as the dependent variable.

The first model we run is where 5-second returns is set as the dependant variable. The outliers for each of the different levels of liquidity are almost all significant for Gdax and Bitfinex but there is a larger variation in the size of the coefficients in the model for Bitfinex. For both markets the asks are negative and the bids positive, this is the opposite of what we would have expected.

Table 6.1: 5-seconds return for Gdax

The dependent variable is Return_5sec_gdax. In the table, you find the different dollar levels, measured individually, away from the mid-quote on the horizontal axis and the independent variables on the vertical axis. Additionally, we have approximately 256 638 observations and a R-squared value of 0.056.

VARIABLES	level 0.01	level 1.0	level 2.0	level 3.0	level 4.0	level 8.0	level 10.0	level 20.0	level 30.0	level 40.0
Return_5sec_gdax_lagged	0.230***	0.230***	0.230***	0.230***	0.230***	0.231***	0.230***	0.230***	0.230***	0.230***
	(0.00192)	(0.00192)	(0.00192)	(0.00192)	(0.00192)	(0.00192)	(0.00192)	(0.00192)	(0.00192)	(0.00192)
outlier_resid_asks	-9.53e-05***	-8.33e-05***	-7.15e-05***	-6.93e-05***	-6.50e-05***	-4.48e-05***	-2.90e-05**	-6.25e-05***	-9.66e-05***	-0.000117***
	(4.68e-06)	(8.56e-06)	(1.12e-05)	(1.40e-05)	(1.56e-05)	(9.67e-06)	(1.35e-05)	(8.80e-06)	(1.08e-05)	(1.28e-05)
outlier_resid_bids	9.66e-05***	0.000103***	8.48e-05***	9.07e-05***	4.70e-05***	6.72e-05***	8.17e-05***	7.47e-05***	8.67e-05***	9.50e-05***
	(4.51e-06)	(8.80e-06)	(1.31e-05)	(1.44e-05)	(1.65e-05)	(1.02e-05)	(1.38e-05)	(8.59e-06)	(9.09e-06)	(1.10e-05)
Constant	5.31e-07	7.48e-07	1.00e-06	8.55e-07	1.02e-06	7.79e-07	6.83e-07	7.40e-07	7.21e-07	8.64e-07
	(1.01e-06)	(9.77e-07)	(9.72e-07)	(9.70e-07)	(9.69e-07)	(9.75e-07)	(9.70e-07)	(9.77e-07)	(9.75e-07)	(9.72e-07)
Observations	256,638	256,638	256,638	256,638	256,638	256,638	256,638	256,638	256,638	256,638
R-squared	0.056	0.054	0.053	0.053	0.053	0.053	0.053	0.054	0.054	0.054
Standard errors in parentheses										
*** p<0.01, ** p<0.05, * p<0.1										

Table 6.2: 5-seconds return for Bitfinex

The dependent variable is Return_5sec_bitfinex. In the table, you find the different dollar levels, measured individually, away from the mid-quote on the horizontal axis and the independent variables on the vertical axis. Additionally, we have approximately 256 617 observations and a R-squared value of 0.019.

VARIABLES	level 0.01	level 1.0	level 2.0	level 3.0	level 4.0	level 8.0	level 10.0	level 20.0	level 30.0	level 40.0
Return_5sec_bitfinex_lagged	0.139***	0.139***	0.139***	0.139***	0.139***	0.139***	0.139***	0.139***	0.139***	0.139***
	(0.00195)	(0.00195)	(0.00195)	(0.00195)	(0.00195)	(0.00195)	(0.00195)	(0.00195)	(0.00195)	(0.00195)
outlier_resid_asks	-	1.23e-07	-4.79e-05***	-4.22e-05**	-5.02e-05***	-3.35e-05***	-4.21e-05***	-3.05e-05***	-2.34e-05**	-6.91e-06
		(8.86e-06)	(1.53e-05)	(1.98e-05)	(1.93e-05)	(1.10e-05)	(1.50e-05)	(8.84e-06)	(1.01e-05)	(1.25e-05)
outlier_resid_bids	-	1.75e-05	7.51e-05***	6.84e-05***	6.53e-05***	5.12e-05***	3.15e-05*	4.26e-05***	3.47e-05***	4.10e-06
		(1.08e-05)	(1.71e-05)	(2.10e-05)	(2.07e-05)	(1.19e-05)	(1.62e-05)	(1.01e-05)	(1.17e-05)	(1.35e-05)
Constant	1.03e-06	8.82e-07	9.80e-07	9.86e-07	1.01e-06	9.49e-07	1.10e-06	1.01e-06	1.01e-06	1.05e-06
	(9.88e-07)	(9.98e-07)	(9.91e-07)	(9.90e-07)	(9.90e-07)	(9.95e-07)	(9.92e-07)	(9.98e-07)	(9.96e-07)	(9.93e-07)
Observations	256,617	256,617	256,617	256,617	256,617	256,617	256,617	256,617	256,617	256,617
R-squared	0.019	0.019	0.020	0.019	0.019	0.019	0.019	0.020	0.019	0.019
Standard errors in parentheses										
*** p<0.01, ** p<0.05, * p<0.1										

The next model we run has 10-second returns set as the dependent variable. For Gdax the results from the 10-second return model are similar with that for the 5-second return model, the coefficients are all highly significant and show the same signs in front of the coefficients as the model on 5-second returns, positive for bids and negative for asks.

Table 6.3: 10-seconds return for Gdax

The dependent variable is Return_10sec_gdax. In the table, you find the different dollar levels away from the mid-quote on the horizontal axis and the independent variables on the vertical axis. Additionally, we have approximately 128 318 observations and a R-squared value of 0.067.

VARIABLES	level 0.01	level 1.0	level 2.0	level 3.0	level 4.0	level 8.0	level 10.0	level 20.0	level 30.0	level 40.0
Return_10sec_gdax_lagged	0.244***	0.247***	0.247***	0.247***	0.247***	0.247***	0.247***	0.247***	0.246***	0.246***
	(0.00270)	(0.00270)	(0.00270)	(0.00270)	(0.00270)	(0.00270)	(0.00270)	(0.00270)	(0.00270)	(0.00271)
outlier_resid_asks	-0.000184***	-0.000158***	-0.000102***	-0.000101***	-0.000120***	-9.39e-05***	-8.91e-05***	-9.83e-05***	-0.000165***	-0.000196***
	(9.44e-06)	(1.63e-05)	(2.15e-05)	(2.68e-05)	(2.97e-05)	(1.90e-05)	(2.61e-05)	(1.72e-05)	(2.01e-05)	(2.33e-05)
outlier_resid_bids	0.000188***	0.000173***	0.000162***	0.000156***	0.000109***	0.000160***	0.000157***	0.000166***	0.000166***	0.000144***
	(9.07e-06)	(1.66e-05)	(2.41e-05)	(2.64e-05)	(3.03e-05)	(1.94e-05)	(2.60e-05)	(1.63e-05)	(1.67e-05)	(1.97e-05)
Constant	7.67e-07	1.69e-06	1.56e-06	1.44e-06	1.90e-06	1.07e-06	1.35e-06	4.88e-07	9.45e-07	1.78e-06
	(2.25e-06)	(2.17e-06)	(2.15e-06)	(2.15e-06)	(2.14e-06)	(2.16e-06)	(2.15e-06)	(2.17e-06)	(2.16e-06)	(2.15e-06)
Observations	128,318	128,318	128,318	128,318	128,318	128,318	128,318	128,318	128,318	128,318
R-squared	0.067	0.063	0.062	0.061	0.061	0.062	0.061	0.062	0.062	0.062
Standard errors in parentheses										
*** p<0.01, ** p<0.05, * p<0.1										

Bitfinex show significant results on the first levels of liquidity that are in line with spoofing, both asks, and bids are significant on the first level of liquidity, 1 USD away from the mid-quote, as well as having the expected coefficients, bids having a negative effect and asks a positive.

The outlier's significance can be the result of the previously mentioned need of the spoofers to catch the traders' attention. This should be close enough from the centre price to catch traders' attention, and at the same time far enough to not risk the offers to be traded upon. We can say that for this first level 0.01, abnormal changes in liquidity are predictive of future price movements. The rest of the results for Bitfinex are not as easily interpreted where asks are again negative and bids positive and they all show a high level of significance. However, if there were spoofers present on the market they would want to stay close to the mid-quote to create attention.

Table 6.4: 10-seconds return for Bitfinex

The dependent variable is Return_10sec_bitfinex. In the table, you find the different dollar levels away from the mid-quote on the horizontal axis and the independent variables on the vertical axis. Additionally, we have approximately 128 668 observations and a R-squared value of 0.048.

VARIABLES	level 0.01	level 1.0	level 2.0	level 3.0	level 4.0	level 8.0	level 10.0	level 20.0	level 30.0	level 40.0
Return_10sec_bitfinex_lagged	0.219***	0.219***	0.219***	0.219***	0.219***	0.219***	0.219***	0.219***	0.219***	0.219***
	(0.00272)	(0.00272)	(0.00272)	(0.00272)	(0.00272)	(0.00272)	(0.00272)	(0.00272)	(0.00272)	(0.00272)
outlier_resid_asks	-	5.24e-05***	-7.73e-05***	-6.40e-05*	-7.53e-05**	-6.34e-05***	-7.13e-05***	-8.32e-05***	-4.12e-05**	-2.71e-05
		(1.69e-05)	(2.71e-05)	(3.38e-05)	(3.34e-05)	(2.03e-05)	(2.64e-05)	(1.67e-05)	(1.89e-05)	(2.30e-05)
outlier_resid_bids	-	-5.04e-05**	9.65e-05***	7.13e-05**	0.000114***	6.80e-05***	7.53e-05***	7.53e-05***	5.71e-05***	2.36e-05
		(2.00e-05)	(2.97e-05)	(3.51e-05)	(3.47e-05)	(2.14e-05)	(2.84e-05)	(1.90e-05)	(2.14e-05)	(2.46e-05)
Constant	1.77e-06	1.52e-06	1.75e-06	1.76e-06	1.65e-06	1.79e-06	1.81e-06	2.15e-06	1.73e-06	1.82e-06
	(2.07e-06)	(2.10e-06)	(2.08e-06)	(2.08e-06)	(2.08e-06)	(2.09e-06)	(2.08e-06)	(2.10e-06)	(2.10e-06)	(2.09e-06)
Observations	128,668	128,668	128,668	128,668	128,668	128,668	128,668	128,668	128,668	128,668
R-squared	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048
Standard errors in parentheses										
*** p<0.01, ** p<0.05, * p<0.1										

These outcomes could arise through several ways. One explanation for the 5-seconds return model is that in the samples of roughly 250 000 observations for each market of returns, 180 000 are zero for Gdax and 81 000 are zeros for Bitfinex, and this could affect our models and their robustness. Having many observations in a sample, factors in the computation of confidence intervals and test statistics, regardless if the variables are connected or not, which could be the reason for the high significance in the 5 second models. The attempt of spoofing could also take more than 5-10 seconds to show in the returns. However, since most of the trading is managed by computers and the algorithms that are associated, this should be unlikely, as argued by a number of researchers, e.g., Jain (2005), Chaboud et al. (2014), and Boehmer et al. (2015). Nevertheless, high frequency trading and its impact on Bitcoin has not been fully explored and the effects of spoofing on the market could therefore take longer to impact the price than on a normal stock exchange. Another explanation for this behaviour could be explained by Urquhart (2018) who argues volume and volatility are the main drivers of attention to Bitcoin and we argue that this phenomenon is not as fast a process for the average trader and would be noticed more than 5 seconds after it appears. This could be described as a process where the order first are put in and withdrawn, raising the volume and price for a short period, followed by an increase in interest from the average traders, and this increased interest would lead to a higher demand for Bitcoin and an increase in price. This is called pump and dump and is another form of price manipulation. This could be one of the reasons for why the trading bots, mentioned by Gandal et al. (2017) were so successful in increasing the price from 150 USD to 1000 USD.

The repeated patterns Gdax show in its results could be the result of them not charging any fees on limit orders that are put on the book, meaning that every order that is not executed immediately is feeless. This could in turn be used by traders that want to sell, or buy, Bitcoin without paying the fees by putting the orders behind the paywall[4] and, for asks at a loss and bids at a profit, get rid of their inventory without fees. However, this is a high risk strategy. For asks, this is a desirable behaviour if the loss is less than the fee that would have been charged. These orders show up as outliers in our detection since they are so large, and because we assume that a lot of traders are doing it at the same time, the liquidity changes drastically and unnaturally. We do not consider this strategy to be applicable to Bitfinex since their trader pay fees on all trades smaller than 7 500 000 USD. Several reasons could explain why Bitfinex and Gdax show different results. The easier access traders have to Gdax could be a reason for the discrepancy, where no minimum deposit is required, opposite to Bitfinex where a minimum of 10 000 USD is required to start trading, as well as the possibility of trading without paying fees only applies to market makers where trades are

[4] The paywall is where offers to sell or buy assets meet market takers who are willing to trade at the prices that are currently offered in contrast to the limit order book where orders that currently don't meet the paywall are put and await to be traded on or withdrawn

larger than 7 500 000 USD. As of the 12th of January 2018, Bitfinex introduced a minimum account deposit of 10 000 USD, as they try to reduce the number of amateur traders on their venue. As the markets differ greatly in types of traders, liquidity, volume and barriers of entry, Gdax, arguably, should be a more efficient market due to its size and the fact that there is more trading closer to the mid-price indicating that it is a more liquid market. The relative easy access should therefore also be an indicator that the results from Gdax should be taken more seriously.

The last model we run is where returns on a 10-minute basis is set as the dependant variable. This model gives us results where the outliers have significance, albeit low, for the first level of liquidity in Gdax. For outliers 0.01 USD away from the mid-quote the asks are both significant, albeit low with a p-value<0.1, and positive. An explanation could be that the attention that is first created may lead to an effect on return, but not instantly. The coefficient is also larger and positive, meaning that the effect of the outlier will be higher for these time frames of returns, which is logical since the dependent variable, the average returns every 10-minutes will be higher, for both positive and negative returns, than the returns for every 5-seconds. The remaining levels of liquidity show the correct denominators in front of the coefficients, positive asks and negative bids but they are all insignificant. These results could be explained by the fact that spoofers act only on the first level of liquidity on Gdax, by using asks.

Table 6.5: 10-minutes return for Gdax

The dependent variable is Return_10min_gdax. In the table, you find the different dollar levels away from the mid-quote on the horizontal axis and the independent variables on the vertical axis. Additionally, we have approximately 2 119 observations and R-squared value of 0.004.

VARIABLES	level 0.01	level 1.0	level 2.0	level 3.0	level 4.0	level 8.0	level 10.0	level 20.0	level 30.0	level 40.0
Return_10min_gdax_lagged	-0.0465**	-0.0513**	-0.0530**	-0.0527**	-0.0537**	-0.0530**	-0.0536**	-0.0536**	-0.0523**	-0.0494**
	(0.0222)	(0.0220)	(0.0218)	(0.0218)	(0.0217)	(0.0218)	(0.0218)	(0.0217)	(0.0221)	(0.0226)
outlier_resid_asks	0.000598*	4.39e-05	0.000328	0.000179	3.51e-05	-0.000186	-4.49e-05	-4.88e-05	-4.60e-05	-7.33e-05
	(0.000335)	(0.000358)	(0.000419)	(0.000475)	(0.000498)	(0.000415)	(0.000489)	(0.000412)	(0.000420)	(0.000449)
outlier_resid_bids	-0.000287	-0.000242	-7.15e-05	-0.000118	0.000205	6.90e-05	9.95e-05	4.86e-05	-0.000149	-0.000429
	(0.000337)	(0.000359)	(0.000421)	(0.000450)	(0.000490)	(0.000392)	(0.000456)	(0.000376)	(0.000383)	(0.000420)
Constant	1.30e-06	0.000216	0.000105	0.000148	0.000124	0.000175	0.000145	0.000152	0.000200	0.000250
	(0.000262)	(0.000218)	(0.000194)	(0.000187)	(0.000184)	(0.000199)	(0.000186)	(0.000205)	(0.000205)	(0.000196)
Observations	2,119	2,119	2,119	2,119	2,119	2,119	2,119	2,119	2,119	2,119
R-squared	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

For Bitfinex there is almost no significance in the model. It is difficult to explain this behaviour, but one reason could be that the model is a bad fit for the data at this return basis. As in previous models, addressing the difference between the two markets, and at which level of measured return they show relevant results is hard and could perhaps best be explained by their differences as platform, as previously mentioned.

Table 6.6: 10-minutes return for Bitfinex

The dependent variable is Return_10min_bitfinex. In the table, you find the different dollar levels away from the mid-quote on the horizontal axis and the independent variables on the vertical axis. Additionally, we have approximately 2 137 observations and R-squared value of 0.001.

VARIABLES	level 0.01	level 1.0	level 2.0	level 3.0	level 4.0	level 8.0	level 10.0	level 20.0	level 30.0	level 40.0
Return_10min_bitfinex_lagged	-0.0320	-0.0321	-0.0305	-0.0316	-0.0327	-0.0286	-0.0300	-0.0281	-0.0304	-0.0295
	(0.0216)	(0.0216)	(0.0217)	(0.0217)	(0.0217)	(0.0217)	(0.0217)	(0.0218)	(0.0217)	(0.0216)
outlier_resid_asks	-	-0.000599	-0.000568	0.000422	0.000313	-0.000759**	-0.000611	-0.000507	-0.000275	-0.000658
		(0.000388)	(0.000457)	(0.000503)	(0.000521)	(0.000359)	(0.000465)	(0.000428)	(0.000474)	(0.000532)
outlier_resid_bids	-	9.00e-05	7.84e-05	0.000834	-0.000304	-2.84e-05	-2.68e-05	0.000607	0.000238	0.00120**
		(0.000415)	(0.000484)	(0.000507)	(0.000513)	(0.000428)	(0.000490)	(0.000450)	(0.000491)	(0.000529)
Constant	0.000148	0.000250	0.000214	1.74e-05	0.000148	0.000272	0.000228	0.000143	0.000154	9.66e-05
	(0.000154)	(0.000183)	(0.000173)	(0.000170)	(0.000169)	(0.000176)	(0.000172)	(0.000176)	(0.000173)	(0.000168)
Observations	2,137	2,137	2,137	2,137	2,137	2,137	2,137	2,137	2,137	2,137
R-squared	0.001	0.002	0.002	0.003	0.001	0.003	0.002	0.002	0.001	0.004
Standard errors in parentheses										
*** p<0.01, ** p<0.05, * p<0.1										

Price manipulation can take on many forms, from low liquidity markets being manipulated as Aggarwal and Wu (2006) examine to spoofing on high frequency markets which Lee et al. (2013) investigate. To say that one type is used over another would be misleading since we do not have enough data to conclude this. However, we can say with some confidence that the markets are affected by the abnormalities in liquidity which we have identified and we can take support from both Allen and Gales' (1992) and Allen and Gortons' (1992) theoretical frameworks that traders can affect the price of an asset by simply trading irrationally. Another aspect which is of importance is what affects return predictability. Returns are hard to predict but according to Yuferova (2017) limit orders are one efficient way of predicting future returns. However, we cannot support her results with our approach of looking at abnormal liquidity. Liquidity is of high importance for price discovery and how well a market is functioning in general. We have chosen to look closer at liquidity when trying to identify irregularities in the markets for that reason, O'Hara (2003), Johan and Li (2009).

6.3 Volatility predictability for Gdax and Bitfinex

In this section, we present our results of the measured impact of abnormal bids and asks on the volatility in the Bitcoin market. According to French et al. (1987), one way to predict returns is by predicting volatility. We use this approach and run two regressions with RV as the dependent variable for both Gdax and Bitfinex, all regressions has the same sampling frequency for the returns. The dependent variable in the first regression is RV measured every 5 minutes based on 1-minute returns and the dependent variable for the second regression is RV measured every 10 minutes based on 1-minute returns. In the results we expect both bids and asks to have positive coefficients since abnormal liquidity will affect volatility positively.

The results from the 5 minutes RV for both platforms show that the coefficient of the lagged RV is very close to one and has a high explanatory power. Results for both platforms are significant at all levels on a 95% level and the results for the lagged variable are what we expect, since volatility is easier to predict, French et al. (1987). Moreover, on the Gdax market we get significant results for the outliers for both bids and asks are. For Gdax, we find that abnormally large bids, and asks, have a positive effect on 5-minute RV, meaning that the volatility of the 1-minute returns increase within 5 minutes after the spike in liquidity. From the results we can interpret that the positive change in volatility on all levels, as when we have abnormal bids and asks in the market, has an impact on the returns. Hence for Gdax, it could be interpreted that the abnormal liquidity is an effect of spoofing, other types of price manipulation or a fee avoidance strategy.

Table 6.7: 5-minute RV based on 1 minute returns for Gdax

The dependent variable is Gdax_5minRV_60secRet. In the table, you find the different dollar levels away from the mid-quote on the horizontal axis and the independent variables on the vertical axis. Additionally, we have approximately 21,500 observations and R-squared value of 0.875.

VARIABLES	level 0.01	level 1.0	level 2.0	level 3.0	level 4.0	level 8.0	level 10.0	level 20.0	level 30.0	level 40.0
Gdax_5minRV_60secRet_lagged	0.937***	0.936***	0.936***	0.936***	0.935***	0.935***	0.935***	0.935***	0.935***	0.934***
	(0.00243)	(0.00242)	(0.00242)	(0.00242)	(0.00242)	(0.00242)	(0.00242)	(0.00242)	(0.00242)	(0.00243)
outlier_resid_asks	0.000125***	0.000219***	0.000143***	0.000183***	0.000126*	0.000132***	0.000182***	6.68e-05	8.60e-05*	0.000197***
	(2.60e-05)	(3.73e-05)	(5.00e-05)	(6.04e-05)	(6.54e-05)	(4.63e-05)	(6.00e-05)	(4.22e-05)	(4.53e-05)	(5.10e-05)
outlier_resid_bids	0.000142***	0.000104***	0.000168***	0.000121**	0.000176***	0.000196***	0.000214***	0.000143***	2.91e-05	9.49e-05**
	(2.54e-05)	(3.69e-05)	(5.00e-05)	(5.54e-05)	(6.29e-05)	(4.34e-05)	(5.60e-05)	(3.83e-05)	(3.82e-05)	(4.34e-05)
Constant	0.000218***	0.000237***	0.000248***	0.000250***	0.000252***	0.000245***	0.000249***	0.000248***	0.000254***	0.000253***
	(1.38e-05)	(1.32e-05)	(1.30e-05)	(1.29e-05)	(1.29e-05)	(1.30e-05)	(1.29e-05)	(1.30e-05)	(1.30e-05)	(1.28e-05)
Observations	21,435	21,435	21,435	21,435	21,435	21,435	21,435	21,435	21,435	21,435
R-squared	0.875	0.875	0.875	0.875	0.875	0.875	0.875	0.875	0.874	0.875
Root MSE 0.00123										
Standard errors in parentheses										
*** p<0.01, ** p<0.05, * p<0.1										

Looking at the results for the Bitfinex market with the 5 minutes RV we get less interesting results. Abnormal liquidity has a positive effect on RV on level one and level two, but the results are only significant on a 95% level at the first level for the asks. This could be a result of a higher variation in the volatility on the market. Several other factors could also be the reason behind Bitfinex results, such as traders that act irrationally or that the market is inefficient. One explanation to why we get mostly negative results might be because of trading circumstances or that people wait longer before they react on abnormal liquidities in the market. The results are difficult to explain and the market characteristics should be further investigated to be able to draw any general conclusions.

Table 6.8: 5-minute RV based on 1 minute returns for Bitfinex

The dependent variable is Bitfinex_5minRV_60secRet. In the table, you find the different levels away from the mid-quote on the horizontal axis and the independent variables on the vertical axis. Additionally, we have approximately 21,500 observations and R-squared value of 0.872.

VARIABLES	level 0.01	level 1.0	level 2.0	level 3.0	level 4.0	level 8.0	level 10.0	level 20.0	level 30.0	level 40.0
Bitfinex_5minRV_60secRet_lagged	0.934***	0.934***	0.934***	0.934***	0.934***	0.934***	0.934***	0.934***	0.934***	0.934***
	(0.00245)	(0.00245)	(0.00245)	(0.00245)	(0.00245)	(0.00245)	(0.00245)	(0.00245)	(0.00245)	(0.00245)
outlier_resid_asks	-	0.000188***	2.03e-05	-0.000183	-0.000150	-1.91e-05	-9.82e-05	-9.05e-06	-0.000203***	-0.000153*
		(6.43e-05)	(0.000114)	(0.000145)	(0.000164)	(8.10e-05)	(0.000111)	(6.35e-05)	(7.71e-05)	(9.19e-05)
outlier_resid_bids	-	7.03e-05	5.47e-05	0.000127	-5.55e-05	-0.000238***	-0.000158	-6.39e-05	-7.13e-05	0.000140
		(7.59e-05)	(0.000125)	(0.000164)	(0.000149)	(8.99e-05)	(0.000126)	(7.48e-05)	(8.66e-05)	(9.95e-05)
Constant	0.000277***	0.000273***	0.000277***	0.000277***	0.000278***	0.000279***	0.000278***	0.000278***	0.000280***	0.000277***
	(1.25e-05)	(1.26e-05)	(1.26e-05)	(1.26e-05)	(1.26e-05)	(1.26e-05)	(1.26e-05)	(1.26e-05)	(1.26e-05)	(1.26e-05)
Observations	21,434	21,434	21,434	21,434	21,434	21,434	21,434	21,434	21,434	21,434
R-squared	0.872	0.872	0.872	0.872	0.872	0.872	0.872	0.872	0.872	0.872
MSE Root 0.00106										
Standard errors in parentheses										
*** p<0.01, ** p<0.05, * p<0.1										

The results from the 10 minutes RV regressions for both platforms show that the lagged RV is very close to one and has a high explanatory power. It has a higher explanatory power for 10 minutes RV than for 5 minutes RV. The results for Gdax 10 minutes RV are significant for all levels up to three dollars and vary on the other levels which follows our hypothesis. This means that we have a higher impact of the outliers on the volatility within 5 minutes from when the order is placed than within 10 minutes from the placed order.

Table 6.9: 10-minute RV based on 1 minute returns for Gdax

The dependent variable is Gdax_10minRV_60secRet. In the table, you find the different levels away from the mid-quote on the horizontal axis and the independent variables on the vertical axis. Additionally, we have approximately 21,500 observations and R-squared value of 0.958.

VARIABLES	level 0.01	level 1.0	level 2.0	level 3.0	level 4.0	level 8.0	level 10.0	level 20.0	level 30.0	level 40.0
Gdax_10minRV_60secRet_lagged	0.980***	0.979***	0.979***	0.979***	0.979***	0.979***	0.979***	0.979***	0.979***	0.978***
	(0.00141)	(0.00140)	(0.00140)	(0.00140)	(0.00140)	(0.00140)	(0.00140)	(0.00140)	(0.00140)	(0.00140)
outlier_resid_asks	0.000119***	0.000165***	0.000111***	0.000161***	8.81e-05*	0.000116***	8.54e-05*	3.63e-05	7.21e-05**	0.000147***
	(1.98e-05)	(2.83e-05)	(3.79e-05)	(4.58e-05)	(4.96e-05)	(3.51e-05)	(4.56e-05)	(3.21e-05)	(3.44e-05)	(3.87e-05)
outlier_resid_bids	0.000108***	9.42e-05***	0.000159***	5.15e-05	0.000165***	0.000156***	5.41e-05	8.05e-05***	2.38e-06	6.78e-05**
	(1.93e-05)	(2.81e-05)	(3.80e-05)	(4.21e-05)	(4.77e-05)	(3.30e-05)	(4.25e-05)	(2.90e-05)	(2.90e-05)	(3.29e-05)
Constant	8.91e-05***	0.000108***	0.000116***	0.000120***	0.000120***	0.000114***	0.000122***	0.000120***	0.000123***	0.000122***
	(1.13e-05)	(1.08e-05)	(1.06e-05)	(1.06e-05)	(1.06e-05)	(1.06e-05)	(1.06e-05)	(1.06e-05)	(1.06e-05)	(1.05e-05)
Observations	21,435	21,435	21,435	21,435	21,435	21,435	21,435	21,435	21,435	21,435
R-squared	0.958	0.958	0.958	0.958	0.958	0.958	0.958	0.958	0.958	0.958
Root MSE 0.00093										
Standard errors in parentheses										
*** p<0.01, ** p<0.05, * p<0.1										

The results for Bitfinex with 10 minutes RV imply about the same results as for 5 minutes RV. We can see a significance on the one dollar level for both bids and asks. From this we can interpret that abnormally large bids and asks at one dollar from the mid-quote has a positive impact on the RV and hence returns, but we cannot say this for any other level tested.

Table 6.10: 10-minute RV based on 1 minute returns for Bitfinex

The dependent variable is Bitfinex_10minRV_60secRet. In the table, you find the different levels away from the mid-quote on the horizontal axis and the independent variables on the vertical axis. Additionally, we have approximately 21,500 observations and R-squared value of 0.958.

VARIABLES	level 0.01	level 1.0	level 2.0	level 3.0	level 4.0	level 8.0	level 10.0	level 20.0	level 30.0	level 40.0
Bitfinex_10minRV_60secRet_lagged	0.979***	0.979***	0.979***	0.979***	0.979***	0.979***	0.979***	0.979***	0.979***	0.979***
	(0.00140)	(0.00141)	(0.00140)	(0.00140)	(0.00140)	(0.00140)	(0.00140)	(0.00141)	(0.00141)	(0.00140)
outlier_resid_asks	-	0.000135***	-5.05e-06	-2.01e-05	2.34e-05	4.59e-05	-6.08e-05	2.27e-05	-0.000149**	2.46e-05
		(4.84e-05)	(8.60e-05)	(0.000109)	(0.000124)	(6.10e-05)	(8.32e-05)	(4.78e-05)	(5.81e-05)	(6.91e-05)
outlier_resid_bids	-	0.000142**	0.000135	-2.22e-05	8.26e-05	-0.000136**	-1.78e-05	-5.53e-05	-9.05e-05	4.90e-05
		(5.71e-05)	(9.39e-05)	(0.000124)	(0.000112)	(6.76e-05)	(9.45e-05)	(5.63e-05)	(6.52e-05)	(7.49e-05)
Constant	0.000131***	0.000127***	0.000130***	0.000131***	0.000131***	0.000131***	0.000131***	0.000131***	0.000133***	0.000130***
	(1.02e-05)	(1.02e-05)	(1.02e-05)	(1.02e-05)	(1.02e-05)	(1.02e-05)	(1.02e-05)	(1.02e-05)	(1.02e-05)	(1.02e-05)
Observations	21,434	21,434	21,434	21,434	21,434	21,434	21,434	21,434	21,434	21,434
R-squared	0.958	0.958	0.958	0.958	0.958	0.958	0.958	0.958	0.958	0.958
MSE Root 0.0008										
Standard errors in parentheses										
*** p<0.01, ** p<0.05, * p<0.1										

For Bitfinex we cannot say more about predictability in volatility than in returns but for Gdax we can. In contrast to other assets, Bitcoin is still in its infancy and models used for trading with algorithms might not be efficient enough and instead of making the markets more efficient, as proposed by a number of researchers such as Hendershott (2011), Boehmer et al. (2015) and Brogard (2014), they instead create chaos. Market efficiency is also an aspect which might lead to our results. Baur et al. (2017) and Nadarajah and Chu (2017) conclude that the market shows signs of weak form efficiency but semi-strong and strong efficiency tests are missing and to our knowledge there do not exist any research concerning this area. This lack of examination could mean that the market for Bitcoin is inefficient for the semi-strong and strong level which could be one of the reasons for our results. Nevertheless, for volatility and Gdax we get interesting results which supports our hypothesis of abnormal liquidity affecting subsequent return and volatility. As mentioned, French et al. (1987) concluded that volatility is correlated to returns and this can be used since volatility is easier to predict than returns.

In addition to looking at predicted returns and volatility with our original outliers we have also tried to distinguish what effect the nature of the outliers has on our models. First, by accumulating the outliers, i.e. adding a dummy of 1 for all levels after the first dummy appears on a particular level. If the dummy is a one on the second level, all following dummies will be converted to ones. This helps to address heteroskedasticity in periods where the price moves a lot and the bid ask spread becomes large and volatile. The results stay mostly the same as when looking at returns and keeping outliers the same as in the first results.

We have also tried to distinguish the duration of the abnormalities by separating short and long periods of abnormal liquidity. By short periods we define them as up to 10 seconds, meaning that if there are more than two outliers that follow each other they will be sorted into the long part. One explanation to shorter abnormalities could be price manipulation and longer abnormalities could be a result of fee avoidance strategies. The short outliers outweigh the long outliers, which could be a result of that price manipulation is present in our sample. However, the results do not change noticeably from what we have presented earlier.

As a final remark, one reason to why we do not get as remarkable results as we first expected might be due to manipulators explores more sophisticated methods once the simple ones stops working, moving away from liquidity. Meaning that the reveal of market manipulation by Gandal et al, 2017 has brought more attention to the problems the market suffers which defer further manipulation via known methods.

7. Conclusion

This paper examines the impact of abnormal liquidity in the Bitcoin market on returns. The hypotheses we test for are, *H1: Abnormal changes in liquidity in the limit order book are predictive of subsequent returns. H2: Abnormal changes in liquidity in the limit order book are predictive of subsequent volatility.* The results we get when testing the first hypothesis of return predictability show that they are predictive but not consistent with spoofing. The results do not show what we expect, but indicate that returns decrease when there is abnormal level of liquidity on the ask and the opposite for bids. When testing the second hypothesis we get results from Gdax that confirms our hypothesis. The volatility in the Bitcoin market increases when abnormal liquidity is present. The price move in the direction in line with spoofing when abnormal liquidity is added to the market.

Regarding the analysis of the two different markets, the results from Gdax are easier to analyse since the results show distinct pattern while the results from Bitfinex are more irregular. Therefore, we have chosen to put less emphasis on Bitfinex when stating our analysis. Aside from the differences the two markets show, our results for both return and volatility predictability are of utmost interest as this is one of the first studies paying attention to this subject, albeit what we find is only representative of a two-weeks period. Our study highlights the market mechanisms and give a perception of the amount of abnormal liquidity and its effect in the market. Furthermore, our study highlight the importance of future exploration of the effects abnormal liquidity has on the Bitcoin market.

7.1 Future research

Our study can be further extended to a longer sampling period than two weeks, more frequent data collection, more than two markets, and examine more than 10 different levels away from the mid-quote, and a larger part of the limit order book. All five areas are subject to improvement to be able to say something more general about return- and volatility predictability from abnormal liquidity. Methodologically, a more advanced model to estimate the outliers is of interest to improve the study. One way could be to include more lags and deciding the length from a formal model selection criteria as AIC or BIC. Another option could be to use an ARIMA process to model the liquidity. Moreover, the way of estimating volatility in the volatility predictability model could be estimated differently. Although RV is argued to be a well functioning model as mentioned by Andersen and Benzoni (2008) it might not always be the best option for predicting volatility. Ghysels et al. (2006) find that using realized power on intraday data is the best estimator of future volatility in competition with realized volatility. Another option could be to use a GARCH model. Additionally, it would be interesting to know the identity behind each offer in the market to look for patterns from the same trader, but which of course is very difficult as it would require a site to leak information about their traders. It would also be interesting to extend the model and study more closely at what orders are fleeting orders (with the intention of being filled) and exclude these orders as outliers in the data. This would be difficult to examine though, since we do not know the identity of each trader and we can therefore not distinguish between fleeting orders and spoofing attempts. Additionally, it would be interesting to examine other types of price manipulation, for example the interaction between liquidity and volume.

References

Articles

Andersen, T. G. and Benzoni, L. (2008). Realized Volatility. *Federal Reserve Bank of Chicago*

Aggarwal, R. K., & Wu, G. (2006). Stock market manipulations. *The Journal of Business*, *79*(4), 1915-1953.

Allen, F., & Gale, D. (1992). Stock-price manipulation. *The Review of Financial Studies*, *5*(3), 503-529.

Allen, F., & Gorton, G. (1991). *Stock price manipulation, market microstructure and asymmetric information* (No. w3862). National Bureau of Economic Research.

Bandi, F. M., and J.R. Russell (2004). Microstructure noise, realized variance, and optimal sampling. Working paper.

Baur, D., Cahill, D., Godfrey, K., & Liu, Z. F. (2017). Bitcoin Time-of-Day, Day-of-Week and Month-of-Year Effects in Returns and Trading Volume.

Boehmer, E., Fong, K., & Wu, J. (2015). International evidence on algorithmic trading.

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-frequency trading and price discovery. *The Review of Financial Studies*, *27*(8), 2267-2306.

Böhme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives*, *29*(2), 213-38.

Chaboud, A. P., Chiquoine, B., Hjalmarsson, E., & Vega, C. (2014). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, *69*(5), 2045-2084.

Crosby, M., Pattanayak, P., Verma, S., & Kalyanaraman, V. (2016). Blockchain technology: Beyond bitcoin. *Applied Innovation*, *2*, 6-10.

Cumming, D., Johan, S., & Li, D. (2011). Exchange trading rules and stock market liquidity. *Journal of Financial Economics*, *99*(3), 651-671.

Foley, S., Karlsen, J., & Putniņš, T. J. (2018). Sex, Drugs, and Bitcoin: How Much Illegal Activity is Financed Through Cryptocurrencies?.

French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of financial Economics*, *19*(1), 3-29.

Gandal, N., Hamrick, J. T., Moore, T., & Oberman, T. (2018). Price manipulation in the Bitcoin ecosystem. *Journal of Monetary Economics*.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, *131*(1-2), 59-95.

Hendershott, T., C. Jones, and A. Menkveld. 2011. Does algorithmic trading increase liquidity?

Journal of Finance 66:1–33.

Jain, P. K. (2005). Financial market design and the equity premium: Electronic versus floor trading. *The Journal of Finance*, *60*(6), 2955-2985.

Jarrow, R. A., & Protter, P. (2012). A dysfunctional role of high frequency trading in electronic markets. *International Journal of Theoretical and Applied Finance*, *15*(03), 1250022.

Kristoufek, L. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PloS one*, *10*(4), e0123923.

Lee, E. J., Eom, K. S., & Park, K. S. (2013). Microstructure-based manipulation: Strategic behavior and performance of spoofing traders. *Journal of Financial Markets*, *16*(2), 227-252.

Li, X., & Wang, C. A. (2017). The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin. Decision Support Systems, 95, 49-60.

Massoud, N., Ullah, S., & Scholnick, B. (2016). Does it help firms to secretly pay for stock promoters?. *Journal of Financial Stability*, *26*, 45-61.

Mills, Terence C. (1990). Time Series Techniques for Economists. *Cambridge University Press.*

Nadarajah, S., & Chu, J. (2017). On the inefficiency of Bitcoin. *Economics Letters*, *150*, 6-9.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*. Princeton University Press.

O'Hara, M. (2003). Presidential address: Liquidity and price discovery. *The Journal of Finance*, *58*(4), 1335-1354.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, *116*(2), 257-270.

Qiu, T., Zheng, B., Ren, F., & Trimper, S. (2006). Return-volatility correlation in financial dynamics. *Physical Review E*, *73*(6), 065103.

Swan, M. (2015). *Blockchain: Blueprint for a new economy*. " O'Reilly Media, Inc.".

Urquhart, A. (2018). What Causes the Attention of Bitcoin?.

Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters*, *148*, 80-82.

Yermack, D. (2015). Is Bitcoin a real currency? An economic appraisal. In *Handbook of digital currency* (pp. 31-43).

Yuferova, D. (2015). Intraday Return Predictability, Informed Limit Orders, and Algorithmic Trading.

Internet

Bitfinex, 2018. Fees schedule. https://www.bitfinex.com/fees . Last accessed April 30, 2018.

Bloomberg, 2018. U.S. Regulators Subpoena Crypto Exchange Bitfinex, Tether. https://www.bloomberg.com/news/articles/2018-01-30/crypto-exchange-bitfinex-tether-said-to-get-subpoenaed-by-cftc. Last accessed April 30, 2018.

CNN, 2017. Bitcoin vs history's biggest bubbles: They never end well. http://money.cnn.com/2017/12/08/investing/bitcoin-tulip-mania-bubbles-burst/index.html Last accessed May 1, 2018.

CoinMarketCap, 2018. Cryptocurrency market capitalizations. https://coinmarketcap.com/currencies/bitcoin/. Last accessed April 5, 2018.

Fortune, 2018. Someone Just Bought $400 Million Worth of Bitcoin. http://fortune.com/2018/02/19/400-million-bitcoin-anonymous-investor/. Last accessed April 30, 2018.

Gdax, 2018. Fee structure. https://www.gdax.com/fees. Last accessed April 30, 2018.

South China Morning Post, 2018. Beijing bans bitcoin, but when did it all go wrong for cryptocurrencies in China?. http://www.scmp.com/news/china/economy/article/2132119/beijing-bans-bitcoin-when-did-it-all-go-wrong-cryptocurrencies. Last accessed May 1, 2018.

The balance, 2018. What's the deal with Coinbase and GDAX? https://www.thebalance.com/whats-the-deal-with-coinbase-and-gdax-4063868 Last accessed April 30, 2018.

The Guardian, 2016. Is Blockchain the most important IT invention of our age?. https://www.theguardian.com/commentisfree/2016/jan/24/blockchain-bitcoin-technology-most-important-tech-invention-of-our-age-sir-mark-walport. Last accessed May 1, 2018.

Veckans affärer, 2018. Bitcoin tokrasar - nu på 6000 dollar.
https://www.va.se/nyheter/2018/02/06/bitcoin-tokrasar---nu-pa-6-000-dollar/.
Last accessed April 30, 2018.

Appendix

A.1 Bitcoin price development

Figure 1: This figure presents the development of the price on Bitcoin expressed in USD from April 5, 2017 to April 5, 2018.

A.2 Programming codes

Both codes for collecting and processing the data are presented below.

A.2.1 Javascript for collection of data from API:

```
const Model = require('./model.js')
const GdaxModel = Model.Gdax
const GdaxTickerModel = Model.GdaxTicker
const BitfinexModel = Model.Bitfinex
const BitfinexTickerModel = Model.BitfinexTicker

const request = require('request')
const schedule = require('node-schedule')

/*const Gdax = require('gdax');
const publicClient = new Gdax.PublicClient();*/

var mongoose = require('mongoose');
mongoose.connect('mongodb://localhost/bitcoin');

var db = mongoose.connection;
db.on('error', console.error.bind(console, 'Mongo connection error:'));
db.once('open', function() {
  console.log('connected to mongo')

  schedule.scheduleJob('*/5 * * * * *', () => {
    //publicClient.getProductOrderBook('BTC-USD', { level: 2 }, (error,
response, data) => {
    request({
      url: 'https://api.gdax.com/products/BTC-USD/book?level=2',
      method: 'GET',
      headers: {
          'Accept': 'application/json',
          'Accept-Charset': 'utf-8',
          'User-Agent': 'test'
      },
      agent: false,
      pool: {maxSockets: 100}
    },
    (error, response, gdaxData) => {
      if (error) {
        console.log('Error in gdax')
        console.log(new Date())
        console.log(error)
      } else if(gdaxData.substring(0, 15) != '<!DOCTYPE html>' && gdaxData)
{
        var data = JSON.parse(gdaxData)

        if(data.bids && data.asks && data.sequence) {

          var bidsArr = []

          for(var i = 0; i < data.bids.length; i++){
            var bidsObj = {
              price: data.bids[i][0],
              qt: data.bids[i][1],
              sp: data.bids[i][2]
            }
            bidsArr.push(bidsObj)
          }
```

```
        var asksArr = []

        for(var i = 0; i < data.asks.length; i++){
          var asksObj = {
            price: data.asks[i][0],
            qt: data.asks[i][1],
            sp: data.asks[i][2]
          }
          asksArr.push(asksObj)
        }

        var data = new GdaxModel({
          date: new Date(),
          sequence: data.sequence,
          bids: bidsArr,
          asks: asksArr
        });

        data.save(function (err, mongoData) {
          if (err) {
            console.log(err)
          } else {
            //console.log(mongoData)
            //console.log('gdax done')
          }
        });
      }
    }
  });
  //publicClient.getProductTicker('BTC-USD', (error, response, data) => {
  request({
    url: 'https://api.gdax.com/products/BTC-USD/ticker',
    method: 'GET',
    headers: {
        'Accept': 'application/json',
        'Accept-Charset': 'utf-8',
        'User-Agent': 'test'
    },
    agent: false,
    pool: {maxSockets: 100}
  },
  (error, response, gdaxData) => {
    if(error) {
      console.log('Error in gdax ticker')
      console.log(new Date())
      console.log(error)
    } else if(gdaxData.substring(0, 15) != '<!DOCTYPE html>' && gdaxData)
{

      var data = JSON.parse(gdaxData)

      if(data.price) {

        var mongoData = new GdaxTickerModel({
          date: new Date(),
          price: data.price
        });

        mongoData.save(function (err, mongoData) {
          if (err) {
```

```
                console.log(err)
              } else {
                //console.log(mongoData)
                //console.log('gdax ticker done')
              }
            });
          }
        }
      });
      request({
        url:
'https://api.bitfinex.com/v1/book/btcusd?limit_asks=50&limit_bids=50',
        method: 'GET',
        headers: {
            'Accept': 'application/json',
            'Accept-Charset': 'utf-8',
            'User-Agent': 'test'
        },
        agent: false,
        pool: {maxSockets: 100}
      },
      (error, response, bitfinexData) => {
        if(error) {
          console.log('Error in bitfinex')
          console.log(new Date())
          console.log(error)
        } else if(bitfinexData.substring(0, 15) != '<!DOCTYPE html>' &&
bitfinexData) {
          var data = JSON.parse(bitfinexData)

          if(data.bids && data.asks) {

            var bidsArr = []

            for(var i = 0; i < data.bids.length; i++){
              var bidsObj = {
                price: data.bids[i].price,
                qt: data.bids[i].amount,
                timestamp: data.bids[i].timestamp
              }
              bidsArr.push(bidsObj)
            }

            var asksArr = []

            for(var i = 0; i < data.asks.length; i++){
              var asksObj = {
                price: data.asks[i].price,
                qt: data.asks[i].amount,
                timestamp: data.asks[i].timestamp
              }
              asksArr.push(asksObj)
            }

            var mongoData = new BitfinexModel({
              date: new Date(),
              bids: bidsArr,
              asks: asksArr
            });

            mongoData.save(function (err, mongoRes) {
```

```
            if (err) {
              console.log(err)
            } else {
              //console.log('bitfinex done')
            }
          })
        }
      }
    })
    request({
      url: 'https://api.bitfinex.com/v2/ticker/tBTCUSD',
      method: 'GET',
      headers: {
          'Accept': 'application/json',
          'Accept-Charset': 'utf-8',
          'User-Agent': 'test'
      },
      agent: false,
      pool: {maxSockets: 100}
    },
    (error, response, bitfinexData) => {
      if(error) {
        console.log('Error in bitfinex ticker')
        console.log(new Date())
        console.log(error)
      } else if(bitfinexData.substring(0, 15) != '<!DOCTYPE html>' &&
bitfinexData) {
        var data = JSON.parse(bitfinexData)

        if(data[6]) {
          var mongoData = new BitfinexTickerModel({
            date: new Date(),
            price: data[6]
          });

          mongoData.save(function (err, mongoRes) {
            if (err) {
              console.log(err)
            } else {
              //console.log(mongoData)
              //console.log('bitfinex ticker done')
            }
          });
        }
      }
    })
  });
});


var mongoose = require('mongoose');
mongoose.connect('mongodb://localhost/bitcoin');

const Schema = mongoose.Schema

const GdaxSchema = new Schema({
  date: {
    type: Date
  },
  sequence: {
    type: Number
```

```
  },
  bids: [{
    price: Number,
    qt: Number,
    sp: Number
  }],
  asks: [{
    price: Number,
    qt: Number,
    sp: Number
  }]
})

const Gdax = mongoose.model('gdax', GdaxSchema)

const GdaxTickerSchema = new Schema({
  date: {
    type: Date
  },
  price: {
    type: Number
  }
})

const GdaxTicker = mongoose.model('gdax_ticker', GdaxTickerSchema)

const BitfinexSchema = new Schema({
  date: {
    type: Date
  },
  bids: [{
    price: Number,
    qt: Number,
    timestamp: Number
  }],
  asks: [{
    price: Number,
    qt: Number,
    timestamp: Number
  }]
})

const Bitfinex = mongoose.model('bitfinex', BitfinexSchema)
const BitfinexTickerSchema = new Schema({
  date: {
    type: Date
  },
  price: {
    type: Number
  }
})

const BitfinexTicker = mongoose.model('bitfinex_ticker',
BitfinexTickerSchema)

module.exports = {
  Gdax:Gdax,
  Bitfinex:Bitfinex,
  GdaxTicker:GdaxTicker,
  BitfinexTicker:BitfinexTicker
}
```

A.2.2 Python code for processing of data:

```
import json
import pandas as pd
import datetime
from datetime import datetime
import numpy as np
from collections import defaultdict
%matplotlib inline
import statsmodels.api as sm
from matplotlib import pyplot as plt

%%time
levels=np.array([0.01, 1.0, 2.0, 3.0, 4.0, 8.0, 10.0, 20.0, 30.0, 40.0])
data={'gdax':defaultdict(list), 'bitfinex':defaultdict(list)}
for file_name in data.keys():
    with open('{}.json'.format(file_name), 'r') as f:
        for i, line in enumerate(f):
              pass
        for i, rec in enumerate(f):
            rec_dict=json.loads(rec)
            df_bid=pd.DataFrame.from_dict(rec_dict['bids'])
            df_ask=pd.DataFrame.from_dict(rec_dict['asks'])
            mid_quote=(df_ask.loc[0, 'price']-df_bid.loc[0,
'price'])/2+df_bid.loc[0, 'price']
            rec_date=pd.to_datetime(rec_dict['date']['$date'])
            for df in (df_bid, df_ask):
                df['centered_price']=df['price']-mid_quote
            df_bid=df_bid[['centered_price', 'price']+[i for i in
df_bid.columns if i not in ['centered_price', 'price']]]
            df_ask=df_ask[['centered_price', 'price']+[i for i in
df_ask.columns if i not in ['centered_price', 'price']]]
            for l in levels:
                new_row=np.ones((1,df_bid.shape[1]))*np.nan
                if l not in df_ask['centered_price'].values:
                    new_row[0,0]=l
                    new_row[0,1]=l+mid_quote
                    df_ask=df_ask.append(pd.DataFrame(new_row,
columns=df_bid.columns))
                if -l not in df_bid['centered_price'].values:
                    new_row[0,0]=-l
                    new_row[0,1]=-l+mid_quote
                    df_bid=df_bid.append(pd.DataFrame(new_row,
columns=df_bid.columns))
            df_bid.sort_values(by='centered_price', ascending=False,
inplace=True)
            df_ask.sort_values(by='centered_price', ascending=True,
inplace=True)

            for df in (df_bid, df_ask):
                df['qt'].fillna(0.0, inplace=True)
                df['volume']=df['price']*df['qt']
                df['date']=rec_date
                df['liq']=df['volume'].cumsum()
                df['return_diff']=np.log(df['price'])-np.log(mid_quote)

            ix=df_bid['centered_price'].isin(-levels) # checks which rows
are levels that we want
            data[file_name]['bids_centered_price'].append(df_bid.loc[ix,
['liq','centered_price','return_diff','date']])
```

```
            ix=df_ask['centered_price'].isin(levels) # checks which rows
are levels that we want
            data[file_name]['asks_centered_price'].append(df_ask.loc[ix,
['liq','centered_price','return_diff','date']])
            data[file_name]['mid_quote'].append(mid_quote)

for exchange in data.keys():
    for f in ('bids_centered_price', 'asks_centered_price'):
        data[exchange][f]=pd.concat(data[exchange][f])
for exchange in data.keys():
    for f in ('bids_centered_price', 'asks_centered_price'):
        data[exchange][f].reset_index(drop=True, inplace=True)

#Create dummies for hour of the day
for exchange in data.keys():
    for f in ('bids_centered_price', 'asks_centered_price'):
        df=data[exchange][f]
        df['hour']=df['date'].apply(lambda x: x.hour)
        df_dummies=pd.get_dummies(df['hour'], prefix='hour')
        for c in df_dummies.columns:

#Change name of dummies for hours
hour_dummies_cols=[i for i in df.columns if 'hour_' in i]

for exchange in data.keys():
    for f in ('bids_centered_price', 'asks_centered_price'):
        display([exchange, f])
        #display(data[exchange][f].head())

        #Create pivot table with data, make bids into absolute numbers
        #add hour column, add column for mid_quote, start from date 2018-
02-05
        df=pd.pivot_table(data[exchange][f], values='liq',
columns='centered_price', index='date')
        if 'bids' in f:
            df.columns=map(np.abs, df.columns.values)
        df.reset_index(inplace=True)
        df['hour']=df['date'].apply(lambda x: x.hour)
        df['mid_quote']=data[exchange]['mid_quote']
        df.columns=map(str, df.columns.values.tolist())
        df.set_index('date', inplace=True)
        df=df['2018-02-05':]
        #df.head()

        #create new dataframe with the df and add dummies for hours
        df_regress=pd.get_dummies(df, columns=['hour'],
                                 # drop_first=True
                                 )
        #df_regress.head()


        #prints the levels applied
        levels_str=list(map(str, levels))
        #levels_str



        #Creates lagged variable
        for l_str in levels_str:

df_regress['Lagged_{}'.format(l_str)]=df_regress[l_str].shift(60)
```

```
df_regress['2Lagged_{}'.format(l_str)]=df_regress[l_str].shift(90)
        df_regress.dropna(inplace=True)
        #df_regress.head()

        dummy_columns=[i for i in df_regress.columns if 'hour_' in i]
        #dummy_columns

        for l_str in levels_str:
            model=sm.OLS(df_regress[l_str],
exog=df_regress[['Lagged_{}'.format(l_str)]+dummy_columns], hasconst=True)
            res=model.fit()
            print(res.summary())

            df_regress['fitted_{}'.format(l_str)]=res.fittedvalues
            df_regress['resid_{}'.format(l_str)]=res.resid

            std=np.std(df_regress['resid_{}'.format(l_str)])
            std196=std*1.96

df_regress['outlier_resid_{}'.format(l_str)]=df_regress['resid_{}'.format(l
_str)].apply(lambda x: 1.0 if x>std196 else 0.0)
        data[exchange][f]=df_regress
        display('Done')
print('Finished')

for exchange in data.keys():
    for f in ('bids_centered_price', 'asks_centered_price'):
        display([exchange, f])
        dfp=data[exchange][f]
        outlier_cols=[i for i in dfp.columns if 'outlier_' in i]
        dfp=dfp[outlier_cols].copy()

        display(set(data[exchange][f][outlier_cols].sum(1)))
        #ix_first_outlier_col=dfp.columns.map(lambda x: 'outlier_resid' in
str(x)).tolist().index(True)
        ix_first_outlier_col=0

        for it, t in enumerate(dfp.index):
            row=dfp.iloc[it,ix_first_outlier_col:]
            if 1.0 not in row.values:
                continue
            ix=row.values.tolist().index(1.0) #Index of the first level
with 1.0
            dfp.iloc[it,ix_first_outlier_col+ix+1:]=0

        for c in dfp.columns:
            data[exchange][f][c]=dfp[c]

        display(set(data[exchange][f][outlier_cols].sum(1)))
print('Finished')

for exchange in data.keys():
    for f in ('bids_centered_price', 'asks_centered_price'):
        display([exchange, f])
        data[exchange][f][outlier_cols].sum().plot('bar')
        plt.show()
            df[c]=df_dummies[c]
```

A.3 Output from model (1)

Below we give an example of the regression that is representative for the output from model (1). The regression output is for Bifinex, asks, at 1 USD level away from the mid-quote. Model (1):

$$Liq_{k,t,l,i} = \beta_{1,t}Liq_{t-1} + \beta_{2,t}First_{hour} + \beta_{3,t}Second_{hour} + \ldots + \beta_{25,t}Twentyfourth\_hour \quad (1)$$

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    1.0   R-squared:                       0.013
Model:                            OLS   Adj. R-squared:                  0.013
Method:                 Least Squares   F-statistic:                     138.2
Date:                Fri, 27 Apr 2018   Prob (F-statistic):               0.00
Time:                        19:58:11   Log-Likelihood:            -3.4695e+06
No. Observations:              257339   AIC:                         6.939e+06
Df Residuals:                  257314   BIC:                         6.939e+06
Df Model:                          24
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Lagged_1.0     0.0783      0.002     39.796      0.000       0.074       0.082
hour_0      7.282e+04   1679.065     43.367      0.000    6.95e+04    7.61e+04
hour_1      6.103e+04   1679.031     36.346      0.000    5.77e+04    6.43e+04
hour_2      3.193e+04   1670.161     19.120      0.000    2.87e+04    3.52e+04
hour_3      3.321e+04   1670.200     19.886      0.000    2.99e+04    3.65e+04
hour_4       3.59e+04   1670.215     21.492      0.000    3.26e+04    3.92e+04
hour_5      3.042e+04   1669.919     18.217      0.000    2.71e+04    3.37e+04
hour_6       3.45e+04   1670.314     20.653      0.000    3.12e+04    3.78e+04
hour_7       4.86e+04   1671.884     29.068      0.000    4.53e+04    5.19e+04
hour_8      7.273e+04   1676.538     43.381      0.000    6.94e+04     7.6e+04
hour_9      5.193e+04   1682.102     30.870      0.000    4.86e+04    5.52e+04
hour_10     3.012e+04   1669.834     18.040      0.000    2.69e+04    3.34e+04
hour_11       2.9e+04   1670.094     17.367      0.000    2.57e+04    3.23e+04
hour_12         4e+04   1671.961     23.925      0.000    3.67e+04    4.33e+04
hour_13     3.205e+04   1670.326     19.189      0.000    2.88e+04    3.53e+04
hour_14     3.016e+04   1670.049     18.058      0.000    2.69e+04    3.34e+04
hour_15     3.563e+04   1670.503     21.330      0.000    3.24e+04    3.89e+04
hour_16     2.633e+04   1669.845     15.770      0.000    2.31e+04    2.96e+04
hour_17     5.037e+04   1677.738     30.021      0.000    4.71e+04    5.37e+04
hour_18      3.52e+04   1670.350     21.072      0.000    3.19e+04    3.85e+04
hour_19     2.889e+04   1670.159     17.297      0.000    2.56e+04    3.22e+04
hour_20     3.186e+04   1670.343     19.075      0.000    2.86e+04    3.51e+04
hour_21     2.729e+04   1673.280     16.308      0.000     2.4e+04    3.06e+04
hour_22     3.707e+04   1729.053     21.442      0.000    3.37e+04    4.05e+04
hour_23      4.46e+04   1729.674     25.786      0.000    4.12e+04     4.8e+04
==============================================================================
Omnibus:                   558250.394   Durbin-Watson:                   0.299
Prob(Omnibus):                  0.000   Jarque-Bera (JB):       3579140875.893
Skew:                          19.627   Prob(JB):                         0.00
Kurtosis:                     579.418   Cond. No.                     9.16e+05
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
[2] The condition number is large, 9.16e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```