Trading and Arbitrage in Cryptocurrency Markets

Igor Makarov[1] and Antoinette Schoar*[2]

[1]London School of Economics
[2]MIT Sloan, NBER, CEPR, Ideas42

April 30, 2018

ABSTRACT

This paper studies the efficiency and price formation of bitcoin and other cryptocurrency markets. First, there are large recurrent arbitrage opportunities in cryptocurrency prices relative to fiat currencies across exchanges that often persist for several days or weeks. These price dispersions exist even in the face of significant trading volumes on many of the exchanges. The total size of arbitrage profits just from December 2017 to February 2018 is above of $1 billion. Second, arbitrage opportunities are much larger *across* than within the same region; they are particularly large between the US, Japan and Korea, but smaller between the US and Europe. But spreads are much smaller when trading one cryptocurrency against another, suggesting that cross-border controls on fiat currencies play an important role. Finally, we decompose signed volume on each exchange into a common component and an idiosyncratic, exchange-specific one. We show that the common component explains up to 85% of the variation in bitcoin returns and that the idiosyncratic components of order flow play an important role in explaining the size of the arbitrage spreads between exchanges.

*Igor Makarov: Houghton Street, London WC2A 2AE, UK. Email: i.makarov@lse.ac.uk. Antoinette Schoar: 62-638, 100 Main Street, Cambridge MA 02138, USA. Email: aschoar@mit.edu. We thank Yupeng Wang for outstanding research assistance. We thank seminar participants at the Brevan Howard Center at Imperial College and LSE, EPFL Lausanne, and Nova, as well as Simon Gervais, Dong Lou, Peter Kondor, Norman Schürhoff, and Adrien Verdelhan for helpful comments. Andreas Caravella, Robert Edström and Ambre Soubiran provided us with very useful information about the data. The data for this study was obtained from Kaiko Digital Assets.

Cryptocurrencies have had a meteoric rise over the last few years. These digital currencies are built on the blockchain technology which allows verification of payments and other transactions in the absence of a centralized custodian. Bitcoin, the most famous and earliest cryptocurrency, was originally introduced in a paper by Nakamto (2008) and came into existence in 2009. Since then the market for cryptocurrencies has evolved dramatically. A large number of new coins have been introduced, most notably ethereum, ripple, bitcoin cash, litecoin and others. The total number of these "altcoins" now counts in the 1000s. Today bitcoins and other cryptocurrencies are traded on more than 100 exchanges worldwide. The average daily volume traded in bitcoin reached $5 billion at the end of February 2018. For the entire cryptocurrency market, the estimates are as high as $40 billion. The current number of active traders is estimated to be above 15 million and include both retail and sophisticated institutional investors (such as DRW, Jump Trading, or Hehmeyer Trading). While significant attention has been paid to the dramatic increase in the volume and price of cryptocurrencies, and many commentators have highlighted their price volatility, there has not been a systematic analysis of the trading and efficiency of cryptocurrencies markets. In this paper we attempt to fill this gap.

A number of features make the cryptocurrency market a unique laboratory for studying arbitrage, price formation and price impact across exchanges. First, there are several non-integrated exchanges that exist in parallel across many countries and jurisdictions. Most of these cryptocurrency exchanges are lightly regulated and are privately owned and operated. On an individual basis the majority of these exchanges function like traditional equity markets where traders submit buy and sell orders and the exchange clears trades based on a centralized order book. However, in contrast to traditional regulated equity markets, the cryptocurrency market lacks any provisions to ensure that investors receive the best price when executing trades, such as the SEC's National Best Bid and Offer (NBBO) rule. The NBBO helps level the playing field for retail investors who might not have the resources to compare prices across multiple exchanges. The absence of such mechanisms means it is up to market participants to compare prices across different exchanges. Second, some of the largest and most liquid bitcoin exchanges as of 2018 are in Europe, Hong Kong, Japan, Korea, and the US. But foreign nationals are often prevented from opening up accounts or transferring capital to local exchanges. This segmentation of markets increases the potential role that large and sophisticated arbitrageurs can play across these exchanges.

In the following we document a number of stylized facts about cryptocurrency markets. First, we show that there are large arbitrage opportunities in bitcoin prices across exchanges that open up recurrently across different exchanges and often persist for several hours, and in some instances even days and weeks. These price dispersions exist even in the face of significant trading volumes on the exchanges. We construct an arbitrage index to show the maximum deviations in the price of bitcoin versus USD at different intervals. The index is calculated each second (each minute in the beginning of the sample), as the ratio of the maximum price of bitcoin on any exchange to the minimum price of bitcoin on any other exchange where bitcoin is traded. On an average day the arbitrage index is 1.05 in 2017. But there are several months where the arbitrage index shoots up to 1.5. For example, in December 2017 and January 2018 there were more than 15 days, where the maximum difference in bitcoin prices across exchanges was more than $3000.

To provide a sense of the magnitude of the money left on the table, we calculate the daily profits that could have been achieved in this market. We find that the daily amount of arbitrage profits between these markets was often more than $5 million a day, and for several days in December 2017 and January 2018, the daily profits reached $30 million. The total size of arbitrage profits in the period December 2017 – February 2018 is estimated to be at least $1 billion under the most conservative assumptions.[1]

Second, we show that arbitrage opportunities are much larger *across* countries (or regions) than *within* the same country. We recalculate our arbitrage index separately for each country where there are several significant cryptocurrency exchanges. We find that the average size of the arbitrage index within US, Europe, Japan and Korea has an average value of around 1.01 to 1.03, compared to a value of 1.15 to 1.6 for the total arbitrage index. Similarly, the daily average price ratio between the US and Korea from December 2017 until the beginning of February 2018 was more than 15 percent; and reached 40 percent for several days. This has been noted in the popular press as the "Kimchi premium". Similarly, the average price difference between Japan and the US was around 10 percent, and between US and Europe about 3 percent. The results suggest that regions which are more closely integrated show smaller cross-region arbitrage spreads.[2] Note, however, that even the *within* country arbitrage spreads are still large by comparison with more traditional asset markets. For example, Du, Tepper, and Verdelhan (2016) show that deviations from the covered interest rate parity in the currency market after the 2008 financial crisis range between 9 and 23 basis points on average on annualized values. These are an order of magnitude smaller than the ones observed in cryptocurrency markets.

[1]Our approach allows us to abstract from any assumptions about price impact of additional arbitrage trades or the speed of convergence. Since we rely on trades that were executed on the exchanges, it also eliminates concerns about stale prices or illiquid exchanges. The results are unchanged, when we repeat the calculations only for the most liquid exchanges.

[2]In fact, several of the exchange in Europe and the US have the same parent company, for example Bitstamp and Kraken operate exchanges both in USD and Euro.

Our findings suggest that there are significant barriers to arbitrage between regions and to a lesser extent even between exchanges in the same country. We show that mere transaction costs cannot explain these results since their magnitudes are small in comparison to the arbitrage spreads we document. The governance risk of cryptocurrency exchanges is also unlikely to explain these arbitrage spreads. First, the exchange risk would have to be correlated within a region to explain the large cross border arbitrages we observe. Second, one would expect that the exchange risk would be correlated with trading volume and bid-ask spreads. This is not supported by the data, since we find large heterogeneity in the liquidity of exchanges within a region but nevertheless arbitrage spreads are small between them.

Our analysis suggests that the most important factors that impede arbitrage are cross-border capital controls on fiat currencies. In further support of this interpretation we find that arbitrage spreads are an order of magnitude smaller in two way cryptocurrency trades (say bitcoin to ethereum) on the exact same exchanges where we see big (and persistent) arbitrage spreads relative to fiat currencies. We show that the arbitrage spread between bitcoin and ethereum in Korea versus the US is low, around 1.03 on average. But over the same time period the spread of bitcoin to Korean Won is more than 20 percent. Similar low arbitrage spreads between bitcoin and ethereum exist between the US and Japan or Europe. At the same time the price of ethereum (or ripple) to fiat currencies, shows similarly large arbitrage spreads as the bitcoin market. Since the main difference between fiat and cryptocurrencies is that capital controls cannot be enforced on cryptocurrency transactions our finding suggest that controls on fiat currency contribute to the large arbitrage spreads we find across regions.

Nevertheless, industry reports suggest that while capital controls are binding for retail investors large institutions are able to avoid these constraints, see for example a recent IMF working paper by Baba and Kokenyne (2011). Thus, capital controls should not impose insurmountable constraints to arbitrage across regions, but they add to the cost of arbitrage. They may be the reason why arbitrageurs are unable to scale up their trading strategies with the intensity of noise trader activity in a timely fashion. We observe a recurring pattern of arbitrage spreads opening up across different exchanges and times which might be the result of a delayed equilibration between noise traders and arbitrage capital.

In the second part of the paper, we ask how arbitrage opportunities arise in the first place. Previous research in other asset classes attributes the price pressure of net order flow to price discovery, but in the cryptocurrency market it is less obvious whether there are any traders who are more informed than others and what the nature of the information is. Nevertheless, we show that a strong positive relationship also exists between net order flows and prices in the cryptocurrency market. A common way to estimate the impact of net order flow is to regress returns over a particular time period on the signed volume of trades during the same period. The complication in the bitcoin market is that the same asset is traded simultaneously on multiple exchanges. When forming their demand investors might not only look at prices on their own exchange but also take into account prices on the other exchanges where bitcoin is traded. Therefore, we build on the approach used in traditional financial markets and decompose signed volume and returns on each exchange into a common component and an idiosyncratic, exchange-specific component; see Hasbrouck and Seppi (2001). We use factor analysis to extract the common factors from data at 5-minute, hourly, and daily frequencies. The common component of signed volume explains about 50% of the variation in returns at 5-minute and hour level, and up 85% at daily level. The price pressure at the daily level is mostly permanent. Buying 10,000 bitcoins raises returns by about 4%.

To investigate the role of signed volume in explaining price deviations across exchanges, we show that exchange-specific residuals of signed volume are significant at explaining variation in exchange-specific residuals of returns at 5-minute and hour level. We also show when the price on any exchange deviates above(below) from the average price on other exchanges, subsequent returns on this exchange are predicted to be lower(higher) than the returns on other exchanges. Furthermore, the predictive power of the local average price index is higher for exchanges in this local market. These results show that arbitrage spreads open up in periods when there are differential price pressures through idiosyncratic signed volume on one exchange relative to another. The arbitrage spreads are not arbitraged away immediately but they do predict subsequent relative returns on exchanges. This lends further support to our interpretation that cryptocurrency prices are the result of a balance between the idiosyncratic sentiments of noise traders and the efforts of arbitrageurs to equilibrate prices across exchanges.

Our paper is related to several streams of the literature. Research on cryptocurrencies in finance and economics is still in its beginning. The majority of papers in this literature focuses on the potential real effects of cryptocurrencies as a payment and transaction mechanism. Ciaian, Rajcaniova, and Kancs (2014), Harvey (2016), Bohme, Christin, Edelman, and Moore (2015), and Raskin and Yermack (2017), provide a broad perspective on the economics of cryptocurrencies and the blockchain technology they are built upon. Athey et al. (2016), and Pagnotta and Buraschi (2018) propose models of valuation of digital currencies. Cong, He, and Li (2018), Easley, O'Hara, and Basu (2017) and Huberman, Leshno and Moallemi (2017) study bitcoin mining fees and the incentives of miners in equilibrium. We view our paper as complementary to this literature. To our knowledge, we are the first to provide a systematic empirical study of trading in cryptocurrency markets using transaction level data.

Our paper is closely linked to the limits of arbitrage literature, especially the idea that noise traders can affect prices even in the presence of arbitrageurs, see e.g., DeLong, Shleifer, Summers, and Waldmann (1990), Gromb and Vayanos (2002), and Gromb and Vayanos (2017). On the empirical side, our paper is closest to the studies that analyze deviations from one price in different markets. In particular, Rosenthal and Young (1990) and Froot and Dabora (1999) study 'Siamese twin' companies. They show that prices of two types of shares, which are traded in different markets but have identical claims on the cash flows and assets of the same company, can nevertheless substantially deviate from each other. Similar to Rosenthal and Young (1990) and Froot and Dabora (1999) we show that bitcoin and other cryptocurrencies can be traded at vastly different prices on different exchanges. The deviation from the law of one price is even more striking in the case of cryptocurrencies, since unlike shares which are traded within specific countries, bitcoins can be transferred to any market. As a result, typical explanations such as tax-induced investor heterogenity or index membership do not apply in this case.

Finally, our paper is also related to research that documents a strong positive relation between asset prices and net order flow in "traditional" financial markets. For example, Evans and Lyons (2002), Berger et al. (2008), and Fourel et al. (2015) look at foreign exchange markets, Brandt and Kavajecz (2004) at U.S. Treasury markets, Deuskar and Johnson (2011) at the S&P 500 futures market, and Chordia et al. (2002) and Hendershott and Menkveld (2014) for NYSE stocks. These papers suggest that order flow imbalances typically explain about 30% percent of the day to day variation of market returns or treasury yields, and 50% of foreign exchange returns. We show that a very strong positive relationship exists in cryptocurrency markets as well. But the R-squared that we document for cryptocurrency markets are significantly higher, up to 80 percent.

To estimate price impact we follow most closely the econometric approach of Hasbrouck (1995) and Hasbrouck and Seppi (2001) who study common factors in stock returns and order flows in a cross-section of US stocks. Following Hasbrouck (1995) we rely on the idea that prices across different markets are cointegratated and thus cannot diverge too far from each other. This allows us to decompose the price on each exchange into the common component (what Hasbrouck calls the implicit efficient price) and an exchange-specific deviation from the common component. One notable difference of our approach from these earlier papers is that we are interested in estimating the magnitude of the price impact of signed volume on the common-component and an exchange-specific component. For this purpose we combine factor analysis with the price decomposition in Hasbrouck (1995) and impose appropriate constraints on factor loadings and weights.

The rest of the paper is structured as follows. Sections 1 and 2 explain the data and provide summary statistics of volume and returns. Section 3 presents the results of the arbitrage index and arbitrage profits within and across regions. In Section 4 we discuss where arbitrage spreads might arise and estimate the model of price pressure. Finally, Section 5 discusses the implication of our findings for aribitrage dynamics in the cryptocurrency market and Section 6 concludes.

1. Data Description

The main data for this project are tick level trading data obtained from Kaiko, a private firm that has been collecting trading information about cryptocurrencies since 2014. The Kaiko data cover the 15 largest and most liquid exchanges: Bitstamp, Kraken, BTCC, Bittrex, Coinbase, OkCoin, Bitfinex, Poloniex, Bithumb, Gemini, Quoine, bitFlyer, Huobi, Binance, and Zaif. Besides bitcoin, which is the first and most famous cryptocurrency, Kaiko also provides trading information on other coins such as ethereum, ripple, tether, and other cryptocurrencies. We also use aggregate data from coinmarketcap.com, an independent website that provides aggregate information on trading volumes by exchange, to verify that the exchanges covered by Kaiko are 15 out of the largest 20 exchanges.

Kaiko obtains the data by querying APIs provided by the exchanges. Where available, we confirm the quality of the data by comparing them with the data reported by the exchanges to bitcoincharts.com.[3] Naturally we rely on the exchanges for the quality of their data.

The variables contained in the data are the time stamp of the transaction at the second level, the price at which the trade happened, the amount of the trade, and an indicator whether the trade was buy or sell initiated. To get information about the bid-ask spread we also use Kaiko's order book data, which are obtained by querying the APIs of the exchanges and taking snapshots of their order books at the minute frequency.

We restrict our attention to the three most liquid and largest cryptocurrency markets: Bitcoin (BTC), Ethereum (ETH), and Ripple (XRP). We focus our analysis on the period from to January 1st, 2017 to February 28th 2018. This choice is moti-

[3] The list of exchanges available both in Kaiko and bitcoincharts.com include Coinbase, Bitstamp, bitFlyer, Bitfinex, Kraken, OkCoin, and Zaif.

vated by the market liquidity. Prior to these dates the liquidity in crypto markets was significantly lower than in later periods.

The 15 exchanges covered by Kaiko span the geographic distribution of bitcoin trading worldwide. The geographic focus of an exchange usually coincides with the main fiat currency used as a base currency. For example, the main trading platform of bitflyer, a Japanese exchange, offers trading of BTC to Japanese Yen. Our classification is as follows:

a. China: OkCoin, Huobi, and BTCC (base currency: Chinese Yuan)
b. Japan: bitFlyer, Zaif, Quoine (base currency: Japanese Yen)
c. Korean: Bithumb and Korbit[4] (base currency: Korean Won)
d. HK: Bitfinex (base currency: US dollar)

There are four major exchanges available in the US and Europe: Coinbase, Kraken, Bitstamp, and Gemini. The three largest of them, Bitstamp, Coinbase, and Kraken allow trading of BTC both in US dollars and Euro. These three exchanges explicitly allow US and European citizens to open an account with the exchange. But access to US citizens typically is in US dollar while European citizens would open an account in Euros. We therefore, classify the euro trades on Bitstamp, Coinbase, and Kraken as belonging to Europe and the US dollar trades on Coinbase, Kraken, Bitstamp, and Gemini as belonging to the US.

e. US: Coinbase, Kraken, Bitstamp, Gemini (base currency: US dollar)
f. Europe: Kraken, Coinbase, Bitstamp (base currency: Euro)

The remaining three exchanges Poloniex, Binance, and Bittrex only allow trading between different cryptocurrencies and not fiat currencies. The main base currency used on these exchanges is tether. Tether (USDT) is a cryptocurrency that was created in 2014; each token is supposed to be backed by one US dollar. The aim was to create a cryptocurrency that facilitates the digital transfer of fiat currencies with the stability of the US dollar. Tether has been highly traded since being used as a base currency for the above exchanges; with a value very close to the dollar.[5]

Daily exchange rate data are obtained from Bloomberg for the exchange rate pairs between Japanese Yen and USD, Korean Won and USD and Euro versus USD. We also obtained hourly exchange rates for Euro: USD; these are reported as the first minute of a given hour.[6]

---

[4] The data for Korbit exchange are obtained from bitcoincharts.com

[5] However, since December 2017 there have been concerns about the transparency of Tether Limited, the company that is backing the 1:1 dollar claim of tether. The U.S. Commodity Futures Trading Commission sent subpoenas to Tether on December 6, 2017.

[6] We thank Adrien Verdelhan for sharing this data with us.

We found a number of data coding errors while cleaning up the data. These adjustments will be helpful for any researcher using the data in the future. First, we found that the time stamp on Korean exchanges, Bithumb, is reported in local Korean time and not in Universal Coordinated Time (UTC). The rest of the exchanges follow the UTC time convention. Second, we found that Bithumb and Quoine also seem to have an error in the convention of signing volume: sell-originated volume seems to be reported as buy-originated volume and vice versa. We confirm that this is the case by looking at the price impact (lambda). For all exchanges except for Bithumb and Quoine lambda is positive while for Bithumb and Quoine the estimated lambda is negative irrespective of the time period of estimation. When we flip the sign, the lambda of course become positive and looks similar to the coefficients estimated for the other bitcoin exchanges (see Section 5).

We also extensively cleaned the data for outliers and stale prices on days where an exchange was closed or experienced major system problems. The code can be obtained from the authors.

2. Summary Statistics

2.1. Volume

We first document the total volume of bitcoin trading to their base currencies across the 15 exchanges. The choice of the base currency depends on the geographical focus of an exchange and can be found in Section 1. For example, for a Japanese exchange we would use the trading volume of bitcoin to Japanese Yen. Several of these exchanges also allow trading of bitcoin to other cryptocurrencies such as ethereum. We do not include this volume here, since the liquidity of these exchange rate pairs is often very limited. We will analyze the trading in bitcoin to other coins in more detail below. Figure 1 Panels A shows the average daily trading volume, averaged over the week, across all exchanges from January 2017 to the end of February 2018. We see that the daily trading volume was about 200,000 bitcoins across all exchanges at the beginning of 2017, which increase to almost 400,000 bitcoins in December and January 2018. We also see that a significant amount of volume is in HK, Japan and the US. In Panel B of Figure 1 we see, however, that the volume was an order of magnitude higher prior during 2016 and only dropped off at the beginning of 2017.

The reason for the significant drop off in volume at the end of 2016 is a change in the distribution of trading across different exchanges. In Figure 2 we now report the fraction of average daily volume split by the different regions. We see that in

January 2017 most of the volume in bitcoin trading (almost 95 percent) was driven by Chinese exchanges. Trading on bitcoin exchanges in the rest of the world combined constituted only a small amount in comparison. In Figure 1 Panel B, we confirm that this distribution was true for most of 2016, when the daily average volume in bitcoin trading was dominated by Chinese exchanges. The high volume on Chinese exchanges in 2016 was reportedly due to the fact that these exchanges had no trading fees and allowed margin trading with very high leverage. However, starting from January 2017 The People's Bank of China began exerting significant pressure on local exchanges to curb speculation and price volatility. In response, the exchanges implemented fixed trading fees of 0.2% per trade during that period and started migrating operations overseas or using peer to peer platforms.

Once the volume in China dropped off after January 2017, Figure 2 shows that the volume in other regions is reasonably evenly distributed. The US and Japanese exchanges each have about 20 percent by the end of February 2018, Europe and Korean around 10 percent. The graph also shows that the two fastest growing bitcoin exchanges are HK based exchange, Bitfinex, and the tether based exchanges: Poloniex, Binance, and Bittrex. Many industry observers believe that some of the volume from China moved to trading on these exchanges.

In Table 1 we provide summary statistics on trading volume, number of trades, average trade size and bid ask spreads for each of the 15 exchanges in our data. We break out the statistics for the two sub-periods, January 1st to July 31st 2017, and then August 1st 2017 to February 28th, 2018, to reflect the significant changes in the bitcoin market. In the first half of the year, average daily trading volume of BTC in US dollars ranges from $2.5M to about $20M. But starting from August 2017 to the end of February 2018, the average daily trading volume increases significantly. On the largest exchanges, such as Bitfinex, Binance or Coinbase the average daily trading volume is around $445M, $224M and $180M, respectively.[7] In comparison the smaller exchanges typically have a daily volume of around $45M. This is more than a tenfold increase in trading volume for most exchanges. The only exception are the three Chinese exchanges, Okcoin, BTCC and especially Huobi, which saw a dramatic drop in trading volume after January 2017.

The data for Binance and Bithumb are only available for the second half of 2017. While the daily trading volume increased significantly over this time period, it is still small relative to major forex markets. For example, the average daily volume of USD to Euro is $575 billion. But the magnitude is comparable to trading in a large firm

[7]In the appendix Table 1, we also report the average daily trading volume for all exchanges in bicoins. The increase in trading volume in bitcoins for most exchanges is more moderate, from 200 to 300 percent.

like Netflix, which has an average daily volume of about $3 billion.

A similar picture emerges when we look at the number of trades across exchanges, which also increases significantly. On most exchanges the number of trades almost triples between the first half of 2017 and the period from August 2017 to February 2018. Some of the biggest increases are on US and Japanese exchanges. Coinbase, the largest US exchange, went from 28,000 trades to about 84,000 trades a day. Similarly, bitFlyer the largest Japanese exchange increases from 32,000 trades per day to about 84,000. In comparison some of the smaller exchanges such as Korbit or Gemini on average have only 16,000 or 23,000 daily trades, respectively. Finally, the bid-ask spread on most of the exchanges is remarkably tight. The bid-ask spread on average is about 10 basis points, and on the most liquid exchanges it is about 2-3 basis points.

2.2. Prices and Returns

We now document the price dynamics of bitcoin. In Figure 3 we plot the volume-weighted daily price of bitcoin from coinbase and bitstamp, the two largest US based exchanges, from January 1st, 2016 to February 28, 2018. Our data confirms the steep increase in the bitcoin price from January 2017 to January 2018, which has drawn a lot of attention in the popular press. The price rose from less than $1000 to almost $20,000 at the end of 2017, with an especially rapid acceleration in the price of bitcoin after November 2017. The price fell back to just below $10,000 by the end of February 2018. Thus, from January 1st, 2016 to February 28, 2018 the return to bitcoin is about 900 percent.

Table 2 shows the higher moments of bitcoin returns at the daily, hourly and 5-minute level from January 1st 2017 to February 28th, 2018. These statistics are calculated by averaging the corresponding moments across all available exchanges. For each frequency we report the annualized standard deviation, skewness and kurtosis of returns; as well as the autocorrelation and cross-correlation across exchanges.

Column (1) of Table 2 reports the standard deviation. We see that the volatility of returns is very high. Even at the daily frequency the annualized standard deviation is 107%. In comparison, the annualized standard deviation of Nasdaq from 1985 to 2017 is 0.18. However, the kurtosis at the daily frequency is 3.86, which is not too far from that of the Normal distribution. The daily returns are positively skewed, which is perhaps not very surprising given the steep increase in the price of bitcoin over the considered time period. Columns (4) through (6) show the autocorrelation in returns for 1, 2 and 3 lags. We can see that even at the 5-minute frequency the autocorrelations are small, which shows that there is little predictability in the market.

Finally, in column (7) we report the average cross correlation of returns. We average across all the pairwise correlations but take out the diagonal, i.e. the autocorrelation of an exchange with itself. We see that at the 5-minute level the correlation between exchanges is quite low, only 57 percent, while at higher frequencies the correlation increases: It is 83 percent at the hourly level and 95 for daily returns. These results are similar to what is observed in other well-established markets, see for example, Budish, Cramton, and Shim (2015). However, if in equity markets the break of correlations happens at millisecond frequencies then here it is already present at the minute levels.

The lower correlations at higher frequencies point out to the existence of arbitrage opportunities between exchanges. In the following section will investigate the existence of arbitrage opportunities in more details.

3. Arbitrage

3.1. Arbitrage Index

The decentralized nature of trading in the cryptocurrency market, makes it an ideal setting to study arbitrage dynamics across exchanges. The autocorrelation in returns we document in the previous section, already suggested that this market might not be perfectly efficient. To document the amount of price dispersion between exchanges at a given point in time, we form an arbitrage index that compares the maximum difference in prices between exchanges. We start by calculating this arbitrage index at the minute level. For this purpose, we compute the value-weighted price per minute for each exchange. For a given minute we take the maximum price across all exchanges and divide it by the minimum price in that minute. We remove outliers by replacing any price movement of more than 10% between two adjacent transactions. Finally, we average the arbitrage index at the daily level to reduce the impact of intra-day volatility.

If the markets were completely integrated and arbitrage free this arbitrage index should be constant and equal to 1 at all times. We first report the arbitrage index across all exchanges in US, Japan, Korea, HK, and Europe. We exclude Chinese exchanges, since the significant government interventions and trading restrictions on these exchanges can potentially make their prices less liquid and integrated with those on other exchanges.[8]

Figure 4 shows that during the period from January 1st 2017 to February 28th, 2018, the arbitrage index is consistently above 1. On average over the year the arbitrage index

[8] Figure 1 in the appendix shows the arbitrage index with the Chinese exchanges included.

is about 1.2. This means in an average day the difference between the highest and the lowest prices across different exchanges was 20 percent. We also see from Figure 4 that there is significant variability during the year. Remarkably there are several months during the year where the index stays at about 1.5, for example May and June 2017, as well as December 2017 to mid-February 2018.

We also construct the arbitrage index for the time period of 2016. It is interesting that during the year 2016 the arbitrage spread is much lower, even with the Chinese exchanges included; on average over the year the index is about 1.05 (compared to 1.2 in 2017) and the maximum spikes occur in June and December. These results suggest that the outsized arbitrage opportunities seem to have opened up during a time when the volume and price volatility are going up significantly.

3.2. Arbitrage Index within Geographical Regions

To understand what might be driving the above results, we decompose the arbitrage index into price differences within regions versus across regions. We start by looking separately at the arbitrage index within each of the major regions, where bitcoin trading is prevalent and we have more than one exchange, i.e. Europe, Korea, Japan and US. For each region we have at least three exchanges, the only exception is Korea where we only have two exchanges Bithumb and Korbit.

In Figure 5 Panel A we report the arbitrage index for the four major operated in the US, Bitstamp, Coinbase, Gemini and Kraken from January 2017 to February 2018.[9] The calculations follow the exact same approach as above. We see that for most of the year the arbitrage index is very small; the average price dispersion is below 2 percent. But there are a few weeks in March, June and December where the arbitrage index is around 1.06. While the arbitrage spreads within the region are small compared to the total arbitrage index reported above, they are still large in comparison to more traditional financial markets.

We then repeat the same analysis for the arbitrage index among European exchanges for the period from January 2017 to February 2018. Again, we see in Figure 5 Panel B that the price discrepancies within Europe are much smaller than the overall arbitrage index. The index within Europe on average is 1.02. And there are only a few dates, in May 2017 and then December 2017 to January 2018 when the price dispersion is around 6 percent. And again, a similar picture emerges in Japan, see Figure 5 Panel C, where the within-region arbitrage index is on average less than 1.02 over the same time period. There are two short periods in January and December 2017, where the

[9]For Bitstamp, Coinbas, and Kraken we use the BTC to USD exchange rate, as discussed above, since it is more likely to be traded by the US investors.

arbitrage index goes to 1.05. The December increase is parallel to the patterns we observed in the other regions. But the increase January might have been idiosyncratic to Japan. And finally, in Figure 5 Panel D we look at Korea. Unfortunately, we only have data for Bithumb starting September of 2017. Therefore, we can only calculate an arbitrage index from that date on. We see that for most of September till December the arbitrage index within Korea was less than 1.03, which is similar to the other regions. But starting in the end of December till the end of January of 2018 the arbitrage index jumped to 1.05, which is still significantly lower than the full index.

Overall the results show that the arbitrage opportunities are much smaller within regions than across regions. This means that the most pronounced price deviations must be driven by across country price dispersion. Thus, our results suggest that cryptocurrency exchanges within a given country or region seem to be much better integrated than across regions.

3.3. Arbitrage Index between Geographical Regions

To confirm the conjecture that a significant part of the arbitrage spread is driven by cross price deviations across geographic regions, we plot for 2017 the price ratio at the minute level between the US and each of the other regions. Remarkably we see that for large parts of 2017 to February 2018, prices on Korean exchanges (indicated in blue here) were more than 20% above the US and had the highest prices overall. There were two distinct periods in June and December to January when the price ratio went as high as 1.6 for a sustained time period. The fact that Korean exchanges have a premium over most of 2017 has even been termed the "Kimchi premium". But we also see that Japan had significant price dispersions from the US during the same time periods as in Korea. But the price ratio between the Japan and the US had a max of 1.2. In contrast, HK saw the biggest difference in price to the US in May 2017 of around 10 percent above the US prices. This deviation seems to have been idiosyncratic to Bitfinex, the largest HK based exchange. It was caused by investigations by the SEC in the US that resulted in US banks stopping banking relationships with the exchange. The price differences between the US and Europe are small compared to other regions, which is perhaps not surprising given that the same set of exchanges operate in the US and Europe. The results suggest that a big fraction of the large arbitrage spreads we documented for the overall market are driven by price differences across regions. And these differences in many instances are persistent over long time periods.

3.4. Arbitrage Profits

So far, we have only looked at the maximum and average price dispersion between regions or exchanges. But this analysis does not take into account the volume traded at different prices. To capture the full magnitude of arbitrage opportunities between regions, we now calculate how much profit could have been made with cross exchange arbitrage. We only look at the period between November 2017 and February 2018, since these were the months that saw the maximum trading liquidity and also spikes in the arbitrage index.

We calculate the arbitrage profits at the second-level and aggregate at the daily level. To make sure that our results are not driven by price volatility we only look at seconds where the price difference from between the exchanges is larger than 2%.[10] For each second, we find the aggregate amount of low priced volume that could have been sold in a high price region. For that purpose, we calculate the sell-initiated volume in the region that has the lowest price in a given second. By using the sell initiated volume we know that we would have been able to buy at this price. Then we calculate the buy originated volume in the region with the highest price. Again, the same logic holds that this ensures that we could have sold in that region. We then compare the two numbers to see if all the sell-initiated volume could have been sold on the high price exchange. If the sell-initiated volume could have been sold, we calculate the value of this volume by multiplying the value times the price at which it was sold, and then we subtract the price at which that volume would have been bought on the low-price exchange. But if the seller-initiated volume is larger than the demand, we calculate which fraction of the low-price volume could have been sold on the high price exchange and calculate the value by multiplying with the higher price.

In Figure 7 Panel A we graph this arbitrage profits between the US and Japan. All profits are reported in USD in order to facilitate comparison between regions. In the beginning of November arbitrage profits on an average day are around $100,000. However, starting on November 28th, the daily profits jump to more than $2 million a day. These outsized profits persist in December, when the average profit per day is around $2 million but there is significant heterogeneity, with several days where the daily arbitrage profit reach $8 million and $16 million a day. Similarly, in January 2018, the average daily arbitrage profit over the month are about $5 million, but again there are several days in the middle of the month when the daily arbitrage profits spike to more than $7 million a day. The timeline of these profits of course closely follows the arbitrage index that we plotted before. In line with those findings we also see that

[10]In case when there are several trades per second, we construct the volume-weighted price per second.

the arbitrage profits drop off significantly at the end of January 2018. The total profits over the examined four months is $116 million.

We repeat the same exercise for the arbitrage profits between Korea versus the US in Figure 7 Panel B. As before we see that the level of the arbitrage profits is higher at the beginning of November 2017, with an average daily profit over the month of $500,000. But the profits jump to $15 million per day at the end of November. In December the average daily profits are about $10 million with several days in the middle of the month where the profits go up to $60 million or even $75 million a day. And as suggested by the plot of the arbitrage index we see that the daily average profits in January 2018 are more than $20 million, with many days higher than $30 million. However, by the end of January 2018 the arbitrage profits start dropping to about $7 to $10 million a day. As suggested by the results on the arbitrage index, the arbitrage profits become quite small starting in February. The total profits over the examined four months is $747 million.

Finally, in Figure 7 Panel C, we also document the arbitrage profits between Europe and the US. As expected the profits are much smaller between these two regions, since the prices seem more integrated. We document that the maximum daily arbitrage profits are around $2 million to $3 million, but for the modal day in December and January are around $500,000. The total profits over the examined four months is $23 million.

3.5. Arbitrage in Other Cryptocurrency Markets

In the previous section we showed that large arbitrage opportunities exist in the price of bitcoin to fiat currencies across regions, in particular the US, Japan and Korea. Now we show that arbitrage opportunities are not unique to the bitcoin market but also exist in other crypto markets.

Figure 8 plots the arbitrage index for ethereum and ripple. We can see that similar to the bitcoin arbitrage index there is significant variability during the year. Periods of relatively low levels of the arbitrage index alternate with prolonged spikes. Similar to bitcoin, at the height of its peak December and January the ethereum arbitrage index stays at about 1.5. The ripple index displays similar behavior, but the series only starts from August 2017 due to the data availability.

As in the case of bitcoin, a significant part of the arbitrage spread in ethereum and ripple prices is driven by price deviations across geographic regions. We do not report this analysis in the paper since they are very similar to the reported arbitrage indices for bitcoin, but can be obtained from the authors on request. Furthermore, by comparing the three arbitrage indices one can notice the high degree of correlation between them. All three arbitrage indices usually spike at about the same time and take similar levels. As we show in the next section, this is not a mere coincidence but the result of arbitrage across cryptocurrencies.

3.6. Arbitrage between Cryptocurrencies

To analyze if the same arbitrage opportunities persist between cryptocurrencies that we saw in the exchange rate of cryptocurrencies to fiat currencies, we now look at the exchange rate between bitcoin and ethereum. We focus on ethereum as the second most traded cryptocurrency after bitcoin.[11] We only look at the months of November 2017 to February 2018, since these are the time periods when the BTC price has the strongest price dispersion relative to fiat currencies. It is also the time period when trading in ethereum and other coins become more liquid. If constraints in the movement of capital contributes to the arbitrage profits between BTC and the local fiat currencies, then these price deviations should be much smaller across cryptocurrencies which by design do not obey the same restrictions.

In Figure 9 we plot the ratios of the value weighted price of bitcoin to ethereum at the minute level across different regions. This is very similar to the calculations in Section 3.3. However, here we make two modifications to the process. First, we calculate the volume-weighted price at the 5-second level since trading in ethereum has less volume than bitcoin. Second, since not all exchanges directly provide a platform to trade bitcoin to ethereum but often trade bitcoin and ethereum only to the local fiat currency we calculate the local exchange rate of bitcoin to ethereum as the cross rate. For example, to calculate the exchange rate ethereum to bitcoin on Bithumb we take the ratio of the exchange rate of ethereum to Korean Won to the exchange rate of bitcoin to Korean Won.

Figure 9 Panel A shows the ratio of the two exchange rates of bitcoin to ethereum between the US and Japanese exchanges from November 1st 2017 to February 28th 2018. As mentioned before, if there were no friction in the currency markets this ratio should be constant and equal to one all the time. We see that the price ratio in November is indeed close to one but during December and January it increases to an average of around 1.03 per day, with a few days jumping to 1.06 and 1.08. In February it falls back to around one.

We repeat the same calculations in Figure 9 Panel B for the price of bitcoin to ethereum but for the US and Korea. Again, we see that the deviations from the ratio

[11] We run the same analysis for ripple and obtain qualitatively similar results, which are available upon request from the authors.

of one are relatively small, and even in December and January it hovers around 1.03 to 0.97. In comparison these were the months when the Kimchi premium (price of bitcoin in fiat currency in Korea versus US) was highest, almost 50% for several days. This again confirms that the arbitrage opportunities are much less pronounced and persistent between different cryptocurrency markets than between cryptocurrency and fiat currency markets. We finally repeat the same exercise in Figure 9 Panel C for the US and Europe, and find that the difference in the price of BTC to ethereum across all three months are small; less than one percent for the average day; and there are only several days in mid-December where the ratio is around 1.03. However, this lack of price dispersion should not be too surprising since we have previously shown that even in the bitcoin to fiat currency market, the difference between US and European exchanges is smaller than in other regions.

4. Order Flow and Prices

To understand how arbitrage opportunities arise in the first place, we develop and estimate a model of order flows and prices. The existing literature documents the importance of net order flows for price formation in "traditional" financial markets (see Evans and Lyons (2002) for the study of foreign exchange markets, Brandt and Kavajecz (2004) for the U.S. government bond markets, and Deuskar and Johnson (2011) for the S&P 500 futures market). While previous research attributes the price pressure of net order flow to price discovery, it is less clear what the fundamentals are in the case of cryptocurrency markets and whether there are any traders who have more information than others. Nevertheless, in this section we show that a strong positive relationship between net order flow and prices also exists in cryptocurrency markets.

A common way to estimate the impact of net order flow is to regress price differences or returns over a particular time period on the signed volume of trades during the same period. The complication in the bitcoin market is that the same asset is traded simultaneously on multiple exchanges and as we showed earlier, often at different prices. Therefore, when forming their demand investors might not only look at prices on their own exchange but also take into account prices on the other exchanges where bitcoin is traded. As a result, a regression of returns on signed volume in each market separately may give a biased picture of the true impact of net order flow.

To accommodate the case of multiple exchanges we decompose signed volume on each exchange into a common component and an idiosyncratic, exchange-specific component:

$$s_{it} = \bar{s}_i + \beta_i^s s_t^* + \hat{s}_{it}, \tag{1}$$

$$E[s_t^*] = 0, \quad E[\hat{s}_{it}] = 0, \quad E[s_t^* \hat{s}_{it}] = 0.$$

Here s_{it} is signed volume on exchange i, s_t^* is the common component for all exchanges, s_{it} is an exchange specific component, and $\bar{s}_i$ is the exchange specific mean. Similarly, we decompose the log return on each exchange, $r_{it} = \ln(p_{it}/p_{it-1})$, into a common component and an idiosyncratic, exchange-specific component:

$$r_{it} = \bar{r}_i + \beta_i^r r_t^* + \hat{r}_{it}, \tag{2}$$

$$E[r_t^*] = 0, \quad E[r_{it}] = 0, \quad E[r_t^* \hat{r}_{it}] = 0.$$

The models (1) and (2) can be estimated either separately by factor analysis by assuming additionally that

$$E[\hat{s}_{it} \hat{s}_{jt}] = 0, \quad E[\hat{r}_{it} \hat{r}_{jt}] = 0, \quad \text{for} \quad i \neq j,$$

or jointly by the canonical correlation analysis (see Jollifee (2002) for a textbook treatment). The canonical correlation analysis estimates models (1) and (2) by maximizing the correlation between s_t^* and r_t^*. The factor analysis and the canonical correlation analysis are linear models. They both estimate the common factors as a linear combination of input data, that is

$$\begin{aligned} s_t^* &= \sum_i w_i^s \left(s_{it} - \bar{s}_i \right), & (3) \\ r_t^* &= \sum_i w_i^r \left(r_{it} - \bar{r}_i \right), & (4) \end{aligned}$$

where w_i^s and w_i^r are called the factor weights. Both models (1) and (2) imply that

$$\sum_i w_i^s \beta_i^s = 1, \quad \sum_i w_i^r \beta_i^r = 1.$$

We fix the scale of the common factor in signed volume by requiring that the sum of factor loadings, β_i^s, is equal to one. To fix the scale of the common factor in returns we require that the sum of factor weights, w_i^r is equal to one. Under this normalization the common factor in returns becomes a portfolio. Suppose one buys β_i^s bitcoins on each exchange i. Since both the sum of β_i^s and $w_i^s \beta_i^s$ is equal to one the total amount of bitcoins bought is equal to one and the common component in signed volume is increased by one. Hence, in the regression

$$r_t^* = \lambda s_t^* + \varepsilon_t, \tag{5}$$

the coefficient λ measures the price pressure of the aggregate order flow.

While the price of bitcoin across exchanges can be different for some period of time, as we show in the analysis above, one should expect the price of bitcoin across any two exchanges to be cointegrated. More generally, any linear combinations of prices where the sum of weights is equal to one should be cointegrated as well. Thus, the restriction that the sum of factor weights, w_i^r is equal to one allows us to decompose the price on each exchange into a common component and an exchange specific deviation from the common component:

$$p_{it} = p_t^* + \hat{p}_{it}, \quad p_t^* = \sum_i w_i^r p_{it}. \tag{6}$$

Unlike the common component of the price, p_t^*, each $\hat{p}_{it}$ must be a bounded process. Since we use the log-prices $\hat{p}_{it}$ measures the percentage deviation from the weighted average price across exchanges. If any of $\hat{p}_{it}$ were an unbounded process it would imply arbitrary large arbitrage opportunities.

4.1. Decomposition of Returns and Order Flow

We estimate the models (1) and (2) using data from 14 exchanges at 5-minute, hourly, and daily frequency. We exclude Binance and Chinese exchanges from this analysis because the Binance data are only available starting from November 2017, and the Chinese exchanges stopped trading in September 2017.

The results of this estimation are reported in Tables 3 and 4. The first panel of Table 3 reports the factor loadings, weights and R-squared of the factor analysis of signed volume using data at the 5-minute frequency. Not surprisingly, the factor loadings are highest for the exchanges with the largest volume. Bitfinex has the highest loading of 0.35 followed by Coinbase USD, Bitstamp USD, and bitFlyer. With the exception of Quoine, the common component explains from 30 to 60 percent of variation in exchange-specific signed volume. In the next two panels of Table 3, we repeat the same analysis but at hourly and daily frequency. The results show that at longer frequencies the common component of signed volume explains an even higher fraction of variation; R-squares go up to 70 percent. The coefficients on the factor loadings stay relatively stable, and the volume on the exchanges which are less liquid and less integrated with the rest of the markets, also load less heavily on the common component of signed volume at the longer frequencies.

In Table 4 we repeat a similar analysis for the common component of returns. As discussed above the one difference is that we now normalize the weights across the exchanges to sum up to one. We start with 5-minute frequency of the return data. For returns the common component is even more important than for signed volume. Even at the 5-minute frequency, the common component in returns explains about 80 percent of exchange returns on exchanges where the US dollar and tether are a base currency. The only exception is Kraken where the R-squared is only 40 percent. The R-squared is also low for Japanese and Korean exchanges. As shown in the next two panels of Table 4 the same pattern persists at longer frequencies. But at longer frequencies the common component becomes progressively more important. We find that the common component in returns explains around 90 percent in variation at hourly frequency and 96 percent at daily frequency. This is intuitive, since we have shown above that the average correlation in returns at the hourly and daily level are extremely high; upside of 95 percent.

We also estimate models (1) and (2) using the canonical correlation analysis. In all cases, the extracted factors are very similar to those estimated using the factor analysis. At 5-minute and hourly frequency the correlation between factors is above 98%. At the daily frequency the common factors in returns correlate at 98% with each other, and correlation between the common factors in signed volume is 90%. Therefore, in what follows we only report results based on the factors extracted using the factor analysis.

4.2. Common Components of Returns and Order Flow

In Table 5 we show that common component in signed volume explains a very large fraction of the common component in returns. The first three columns report the estimates at the 5-minute frequency, the next six columns show results at the hourly and then daily frequencies. In column (1) we regress the common component in returns on the contemporaneous common component in signed volume. The coefficient on the independent variable is 8.8×10^{-6} with a t-statistic of 80 and an R-squared of 54 percent, which shows that there is a very strong relationship between the common component in returns and signed volume. For any 10,000 bitcoin increase in buy volume across exchanges we see an 9 percent increase in the price on average. In columns (2) and (3) we look at the persistence of the price impact. In column (2) we add one lag of the common component in signed volume, and in column (3) we add five lags. We see that the coefficients at all five lags are negative and significant, which suggests that part of the price pressure in the common component is temporary. A bit less than half of the impact on returns reverses within the next five periods. The negative sign is also consistent with the negative sign of the first-order correlation of 5-minute returns.

In the next three columns we report the results of the price impact regressions at the hourly level. The price impact of the contemporaneous common factor in signed volume is still positive and highly significant but slightly smaller than that for the 5-minute frequency. Buying 10,000 bitcoin over an hour predicts a 6 percent increase in the price on average. Similar to the 5-minute frequency the coefficients on lagged signed volume are negative suggesting that even at the hourly frequency some of the price impact is temporary. Finally, at the daily frequency we see a smaller price impact on average (the coefficient on the contemporaneous common component in signed volume is only 3.6) and also much reduced mean reversion in lags two through five. Buying 10,000 bitcoin over a day predicts a 3.6 percent increase in the price on average. The past signed volume is only significant at the first lag bringing the total price impact to about 3% for every 10,000 bitcoins.

Overall, we show that the common component of signed volume explains a large fraction of the common component of returns at all frequencies and subsequent mean reversion is much smaller at the daily level than at higher frequencies.

4.3. Idiosyncratic Price Pressure

To find the exchange specific price pressure we follow Hasbrouck (1991) and estimate the VAR model of the idiosyncratic part of signed volume and the exchange-specific deviation from the common component:

$$\hat{s}_{it} = \sum_{s=1}^{\tau} b_{i,s}\hat{s}_{it-s} + \gamma_i \hat{p}_{it-1} + u_{it}, \tag{7}$$

$$\hat{p}_{it} = \sum_{s=1}^{\tau} a_{i,s}\hat{p}_{it-s} + \lambda_i \hat{s}_{it} + v_{it}, \tag{8}$$

where

$$E[v_{it}] = E[u_{it}] = 0, \quad E[v_{it}v_{is}] = E[u_{it}u_{is}] = 0, \text{ for } s \neq t, \quad E[v_{it}u_{is}] = 0.$$

We estimate these equations as a system of OLS equations and obtain the results for equation (7) in Table 6, and for equation (8) in Table 7. The idiosyncratic component of the price on each exchange is obtained as the residual value after taking out the common component from each price. The exchange specific deviations from the common component are then estimated as a function of contemporaneous signed volume as well as past deviations from the common price, we use three lags here. We do not include lagged values of idiosyncratic signed volume, since our estimates of equation (7) show that the lagged values have low correlation with the contemporaneous value of residual of signed volume. To conserve the space we report the results based on the 5-minute frequency.

Each column of Table 7 reports the result of equation (8) for each exchange as labeled on the top of the column. By looking at the coefficients on the lagged values of $\hat{p}_{it}$ it is apparent that $\hat{p}_{it}$ are highly persistent mean-reverting processes. All coefficients on the past three lags are positive. The sum of the three lags while below one is close to one for all exchanges. When the price on any exchange deviates above(below) from the average price on other exchanges then subsequent returns on this exchange are predicted to be lower(higher) than the returns on other exchanges. But the convergence to the common component is slow. The coefficient on the past lags of $\hat{p}_{it}$ are particularly high for Japanese and Korean exchanges, such as Bitflyer or Bithump. This confirms our prior results that arbitrage spreads persist for longer on these exchanges.

Compared to the price pressure we estimate for the common component, the idiosyncratic (exchange specific) price pressure is significantly higher on almost all exchanges. The price pressure is particularly high on smaller and less liquid exchanges, for example Zaif or Coinbase EUR. The two exchanges where the estimated coefficient on the idiosyncratic price impact are lower than the one on the common component are Bitfinex and Bitstamp. These are two of the largest and most liquid exchanges. Note that one should be careful with the interpretation of the exchange specific price pressure. When prices on one exchange are either very high or low traders might adjust which exchanges they trade and how they trade. Hence, the idiosyncratic part of signed volume might change endogenously and this can lead to non-linearity in the relationship between price and signed volume, which is not picked up in our model. Nevertheless, our results show that the exchange-specific part of signed volume play an important role at explaining the deviation of prices on an exchange from the common component.

5. Discussion of Arbitrage Stategiges

Our goal in this section is to outline the mechanics of the different arbitrage strategies that exist in the bitcoin market and the potential risks and costs that can impede the effectiveness of arbitrage. Consider the situation that the price in Korea is above the price in the US as we documented above. In the world without frictions this situation would constitute a riskless arbitrage. One could buy bitcoins in the US, transfer them to Korea, sell them for Korean Won, exchange Korean Won for dollars, and transfer dollars back to the US. In practice, this textbook arbitrage is not possible since the nature of bitcoin transactions implies that it takes about an hour for the transaction to be registered on the blockchain. Moreover, exchanges typically take from a few hours to several days to transfer fiat currency. In that time period the arbitrage opportunity might disappear. As a result, to lock in the arbitrage an arbitrageur has to simultaneously buy bitcoin on the exchange where the price is low and sell it on the exchange where the price is high.

Ideally, the arbitrageur would like to short sell bitcoin on the market where the price is high, say Korea, and buy bitcoin in the US. Then she would transfer bitcoin from the US to Korea, and so realize the riskfree profit. This trading strategy, however, is generally not feasible, because few if any exchanges allow short-sales. In the absence of short-sales the arbitrageur in cryptocurrency markets can, however, resort to two alternative arbitrage strategies.

First, she could establish a negative position in bitcoin by trading on margin, which is similar to short-sales, but does not allow for physical settlement. In this case, the arbitrageur can profit from the trade only if prices on the two exchanges converge in the future. Thus, the arbitrageur is subject to the convergence risk, which has been extensively studied in the limits of arbitrage literature, see for example, Shleifer and Vishny (1997), or Gromb and Vayanos (2002). While in theory prices across exchanges may not converge for a long time, Figure 6 shows that in practice, arbitrage opportunities in any given market open for less than two days on average and even in the extreme never existed for more than a month.

The second arbitrage strategy, is to hold a positive balance of bitcoins on both exchanges and simultaneously buy and sell bitcoins across the two exchanges whenever the price on one exchange deviates from that on the other. Naturally, the bitcoin balance of the arbitrageur will go down on the exchange where the price of bitcoin is high (since this is where she would sell bitcoin) and increase on the exchange where the price is low. To replenish it the arbitrageur needs to transfer bitcoins from the exchange with high bitcoin balance to the one with low balance. While this strategy does not expose the arbitrageur to convergence risk, a substantial drawback of this strategy is that the arbitrageur becomes exposed to bitcoin price fluctuations. To mitigate this risk she can borrow bitcoin from people who hold big amounts of bitcoin without an interest to sell, the so-called *hodlers*.[12] Of course, these hodlers themselves would be in a great position to do the arbitrage in cryptocurrency markets. Starting from the end of December 2017, the arbitrageur can also use CBOE and CME bitcoin futures contracts to hedge the price risk. The futures contracts track bitcoin price on major US dollar exchanges and have an average daily open interest of about 10,000 bitcoin.[13]

[12]The term hodler is a peculiarity of the bitcoin market since one investor in bitcoin wrote in a post on the Bitcoin talk forum in 2013 while prices were dropping *I AM HODLING*. This has become a meme for Hold On for Dear Life.

[13]The CBOE contract settlement price is determined by results of the auction on the Gemini exchange. The CME contract settlement price is based on the CME CF Bit-

In practice, the arbitrageur has to incur a number of transaction costs, but their magnitude is too small to prevent arbitrageurs from implementing the above trading strategies. To transfer bitcoins the transaction has to be recorded on the blockchain; this is the work of the so-called miners that provide certification of transactions. The fees are typically around \$10 per transaction, they peaked around \$40 in the end of December 2017 at the height of the bitcoin price, but since February have come down to below \$10. Since these are fixed cost, they are minuscule relative to the size of the potential arbitrage. In addition, exchanges have trading fees, which increase the cost of trading. We document in the appendix, the magnitude of the fees for the exchanges used in this paper. These fees range from 0.25% of the amount traded to 0.1%. Most exchanges do not charge fees on a trade by trade basis but assign them based on the trading volume in a given month or week. Furthermore, most exchanges charge zero fees for trades that add to the liquidity of the order book, e.g. limit orders. The exchange fees are comparable to the bid-ask spreads, which are on average between 1 and 10 bp. Finally, many exchanges charge withdrawal fees, these range from 10 to 50 bp per withdrawal for most of the exchanges. But all large exchanges state that for VIP, large traders they provide preferential customized fees that are much below the cost for retail investors. In sum, these fees are small for large transactions. Overall, we believe that for large players the round-up trading costs should be within 50 to 75 bp. These transaction costs are very low compared to the arbitrage spreads we documented, and therefore, cannot explain the arbitrage spreads we find in the paper.

Another factor that might limit the willingness of traders to engage in arbitrage is the governance risk of cryptocurrency exchanges. The governance risk of exchanges arises since in practice to trade on an exchange the arbitrageur has to transfer her bitcoins to the exchange and therefore to give up control of her coins to the exchange. Judging from many widely publicized hacks of exchanges these can lead to significant losses to investors who trade there.[14] However, it seems unlikely that this explains the arbitrage spreads we found. Concerns about the governance risk of an exchange should affect its volume and possibly bid-ask spreads. But we show that many of the exchanges with the largest arbitrage spreads, for example Bithump and Korbit, have very significant volumes and small bid ask spreads. Moreover, we show that arbitrage spreads are much larger *across* than within regions. For exchange risk to explain this pattern one would have to assume that it is correlated within a region. But this is not supported by our data since there is significant heterogeneity in the liquidity of

coin Reference Rate (BRR), which aggregates the price from major US dollar exchanges (see http://www.cmegroup.com/trading/equity-index/us-index/bitcoin.html for more details.)

[14]For example, in the notorious hack of Mt Gox in 2014 850,000 bitcoins were stolen from customers and the company.

exchanges within a region but nevertheless arbitrage spreads are small between them.

Finally, an important potential constraint to arbitrage are cross-border capital controls. As we described before unless the arbitrageur is willing to bet on price convergence between Korean and the US exchanges she would need to sell bitcoin in Korea and repatriate profits from Korea to the US. The regulation in some countries make cross-border transactions difficult for retail investors. For example, in Korea, local residents and companies moving more than $50,000 out of the country in a single year must submit documents to authorities proving their reasons for the transfers, which may not always be approved. Industry reports as well as descriptions from trading blogs suggest that these constraints are binding for retail investors. However, it is more difficult to quantify how binding these constraints are for large financial institutions that trade in multiple international financial markets.[15] There are a few reports which suggest that large institutions are able to avoid these constraints. In a recent IMF working paper Chikako and Kokenyne (2011) find that the effectiveness of capital controls in South Korea seems limited, since capital flows in and out of the country as well as the effectiveness of monetary policy does not seem to be significantly changed after the introduction of capital controls in the early 2000s. Similarly, industry reports suggest that there are networks of forex dealers which help institutions to transfer capital in and out of the country. Thus, capital controls should not impose insurmountable constraints to arbitrage across regions, especially for large traders, but they add to the cost of arbitrage. This interpretation is supported by our finding that arbitrage spreads are an order of magnitude smaller in two way cryptocurrency trades (say bitcoin to ethereum) on the exact same exchanges where we see big (and persistent) arbitrage spreads relative to fiat currencies. But even in case of the fiat currency the arbitrage spread does not stay open for more than a month and eventually close.

In summary, our analysis suggests that the history of bitcoin exchanges over the last years is marked by recurring episodes of arbitrage opportunities opening up and closing again and a few periods of extremely large arbitrage spreads that persists for several weeks. It appears that most of the time, arbitrageurs are able to equalize prices across markets.[16] But at times the arbitrage capital seems to get overwhelmed by the noise traders who are driving up the price in certain markets or lose heart when negative information about bitcoin comes out. One interpretation of our results is that this is a particular form of slow moving capital (see the 2010 AFA presidential address by

[15] A related constraint is that many retail investors face restrictions on which exchanges they can trade. For example, foreign nationals are typically prevented from opening up accounts and trading on local exchanges. But similar to capital controls large financial institutions should be able to bypass these restrictions and be able to operate across regions.

[16] Industry reports suggest that hedge funds and high frequency traders have been active across different cryptocurrency markets for several years.

Duffie (2010) for an extensive discussion), where arbitrageurs because of the described constraints and risks are unable to scale up their trading strategies with the intensity of noise trader activity in a timely fashion.

6. Conclusions

This paper studies price impact and arbitrage dynamics in the cryptocurrency market. We show that there are large arbitrage opportunities across different exchanges that open up recurrently and often persist for several days and weeks. These arbitrage opportunities are much larger *across* countries (or regions) than *within* the same country. The total arbitrage profits for the period from December 2017 to February 2018 are at least \$1 billion.

To analyze how these price deviations between exchanges emerge, we analyze the relationship between net order flows and prices in the cryptocurrency market. We decompose signed volume and returns on each exchange into a common component and an idiosyncratic, exchange-specific component. The common component of signed volume explains about 50% of the variation in returns at 5-minute and hour level, and up 85% at daily level. The exchange-specific residuals of signed volume are significant at explaining variation in exchange-specific residuals of returns at 5-minute and hour level. We also show that when the price on any exchange deviates above(below) from the average price on other exchanges then subsequent returns on this exchange are predicted to be lower(higher) than the returns on other exchanges.

These results suggest that trading in bitcoin markets seems to be a constant counterbalancing effort between noise traders whose changing sentiment about the cryptocurrency market creates price pressures on exchanges and arbitrageurs who are trying to equalize prices across exchanges. The findings also suggest that there are significant barriers to arbitrage between regions and to a lesser extent between exchanges in the same country. They appear to be related to capital controls and other operational constraints across countries.

References

Avdjiev, S., W. Du, C. Koch, and H. S. Shin ,(2016), "The Dollar, Bank Leverage and the Deviation from Covered Interest Parity" *Working Paper* 592, BIS.

Athey, S., Parashkevov, I., Sarukkai, V., and J. Xia, (2016), "Bitcoin pricing, adoption, and usage: Theory and evidence," *Working Paper.*

Berger, D., A. Chaboud, S. Chernenko, E. Howorka, and J. Wright, (2008), "Order Flow and Exchange Rate Dynamics in Electronic Brokerage System Data," *Journal of International Economics*, 75, 93–109.

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

Brandt, M. W., and K. A. Kavajecz, (2004), "Price Discovery in the U.S. Treasury Market: The impact of Order flow and Liquidity on the Yield Curve," *Journal of Finance*, 59, 2623–54.

Budish E., Cramton P., and J. Shim, (2015) "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," *The Quarterly Journal of Economics*, 130, 2015, Pages 1547–1621.

Chikako, B., and A. Kokenyne, (2011),"Effectiveness of Capital Controls in Selected Emerging Markets in the 2000," *IMF Working paper.*

Chordia, Tarun, Richard Roll, and Arvind Subrahmanyam, (2002), "Order imbalance, liquidity, and market returns," *Journal of Financial Economics* 65(1), 111–130.

Ciaian, P., Rajcaniova, M., and d'Artis Kancs, (2014), "The Economics of BitCoin Price Formation," *Working paper.*

Cong, L.W., He Z., and J. Li, (2018), "Decentralized Mining in Centralized Pools", *Working paper.*

DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, (1990). "Noise trader risk in financial markets", *Journal of Political Economy*, 98, 703–738.

Deuskar P. and T., C. Johnson, (2011)."Market Liquidity and Flow-driven Risk," *Review of Financial Studies*, 24(3), 721–753.

Du, W., A. Tepper, and A. Verdelhan (2016): "Deviations from Covered Interest Rate Parity," *Working Paper*, MIT.

Duffie, D. (2010), "Presidential Address: Asset Price Dynamics with SlowMoving Capital," *The Journal of Finance*, 65, 1237–1267.

Easley, D., M. O'Hara, and S. Basu, (2017), "From Mining to Markets: The Evolution of Bitcoin Transaction Fees," *Working Paper*.

Evans, M. D. D., and R. K. Lyons. 2002. "order flow and Exchange Rate Dynamics," *Journal of Political Economy*, 110, 170–180.

Fourel, Valere, Dagfinn Rime, Lucio Sarno, Maik Schmeling and Adrien Verdelhan, (2015), "Common Factors, Order Flows, and Exchange Rate Dynamics," *Working Paper*.

Froot K. A. and E. M. Dabora, (1999), "How are stock prices affected by the location of trade?," *Journal of Financial Economics*, 53, 189–216,

Gromb, D. and D. Vayanos, (2002). "Equilibrium and welfare in markets with financially constrained arbitrageurs," *Journal of Financial Economics* 66, 361–407.

Gromb, D. and D. Vayanos, (2017). "The Dynamics of Financially Constrained Arbitrage," *Working Paper*

Harvey, C. R, (2016). "Cryptofinance," *Working Paper*.

Hasbrouck, (1995), "Many Markets: Determining the Contributions to Price Discovery," *Journal of Finance* 50 (4), 1175–1199.

Hasbrouck, J. and J. D. Seppi, (2001). "Common factors in prices, order flows, and liquidity," *Journal of Financial* , 59(3), 383–411.

Hendershott, T. and A. J., Menkveld, (2014). "Price Pressures," *Journal of Financial Economics*, 114, 405–423.

Huberman, G., J. D. Leshno, and C. Moallemi, (2017), "Monopoly without a monopolist: An economic analysis of the bitcoin payment system," *Working Paper*.

Lamont, Owen and Richard Thaler, (2003), "The Law of One Price in Financial Markets," *Journal of Economic Perspectives* 17 (4), 191-202.

Pagnotta E.S. and A. Buraschi, (2018), "An Equilibrium Valuation of Bitcoin and Decentralized Network Assets," *Working Papers*.

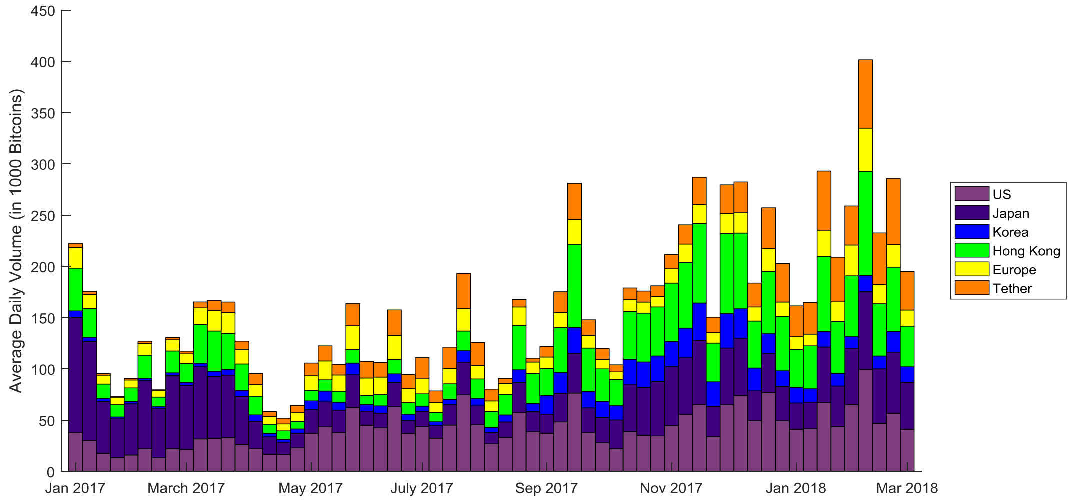
Rosenthal L. and C. Young, (1990), “The seemingly anomalous price behavior of Royal Dutch/Shell and Unilever N.V./PLC,” *Journal of Financial Economics*, 26, 123–141.

Rime, D., A. Schrimpf, and O. Syrstad, (2016), “Segmented money markets and covered interest parity arbitrage,” *BIS Working Papers*, 651.

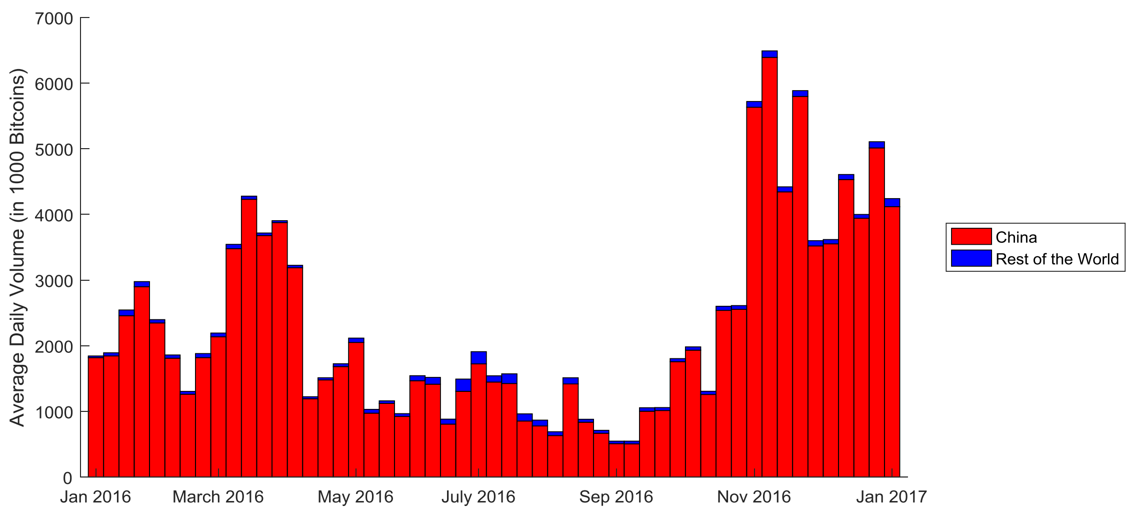
Shleifer, A., and R. Vishny, (1997), “The limits of arbitrage,” *Journal of Finance* 52, 35–55.

Raskin, M., and D. Yermack, (2017), “Digital currencie, decentralized ledgers, and the future of cental banking”, *NBER Working Paper 22238*.

Figures



Panel A



Panel B

Figure 1. Total trading volume. This figure shows the average daily volume of bitcoin to fiat currency trading per week (reported in 1000 BTC), from January 2017 until February 28^{th} 2018 (Panel A), and from January 2016 to December 2016 (Panel B). The volume is reported across all 15 exchanges in the Kaiko data. We exclude any volume that is coin to coin trading. The fraction of volume that is generated on exchanges in different regions is indicated with different colors. The regions are China, Europe, Hong Kong, Japan, Korea, US and Tether. In Panel A we report all regions apart from China. In Panel B we report all regions including China. The trading volume in China is indicated in red, while the volume in the rest of the world is in blue and was less than 5 percent.

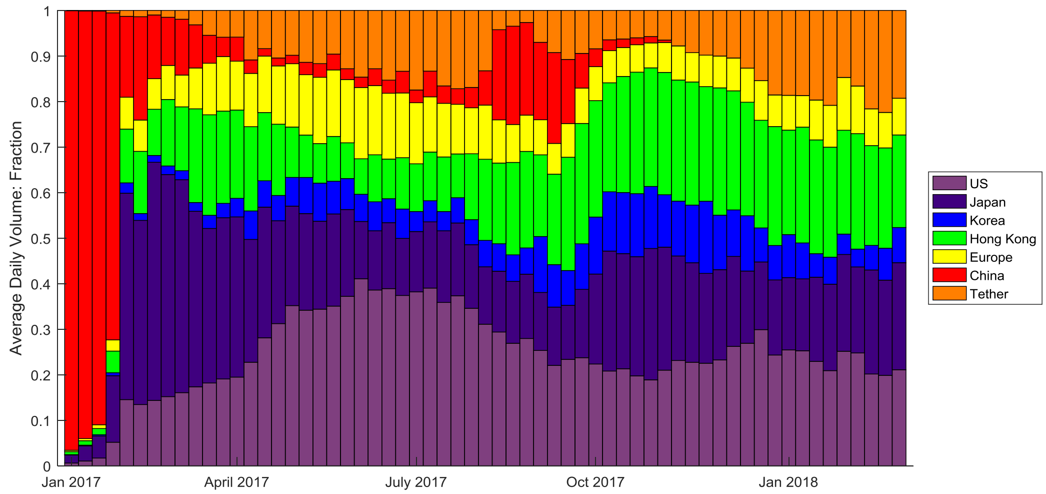


Figure 2. Volume shares. This figure shows the fraction of total bitcoin volume to fiat currency from January 2017 until February 28th 2018 across all 15 exchanges in the Kaiko data. The fraction of volume that is generated on exchanges in different regions is indicated with different colors. The regions are China, Europe, Hong Kong, Japan, Korea, US and Tether.

Figure 3. Bitcoin price: January 1st, 2016 – February 28, 2018. The price is calculated as the volume-weighted daily price of bitcoin based on the data from Coinbase and Bitstamp.

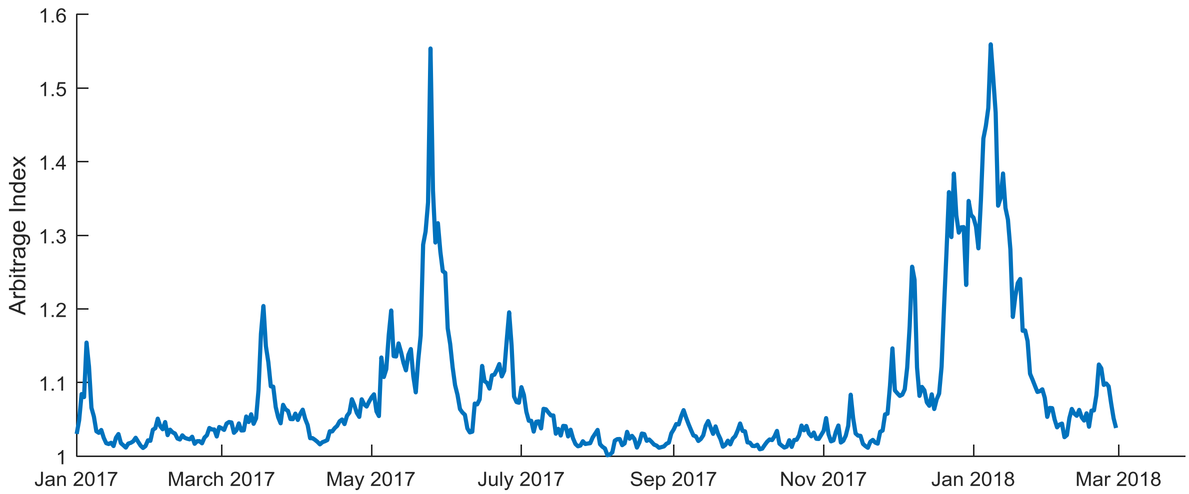


Figure 4. Arbitrage index. The arbitrage index is calculated based on the value-weighted price per minute for each exchange and averaged at the daily level. For a given minute the maximum price across all exchanges is divided by the minimum price in that minute. The regions included in the index are Europe, Hong Kong, Japan, Korea, US and Tether from January 2017 until February 28th 2018.

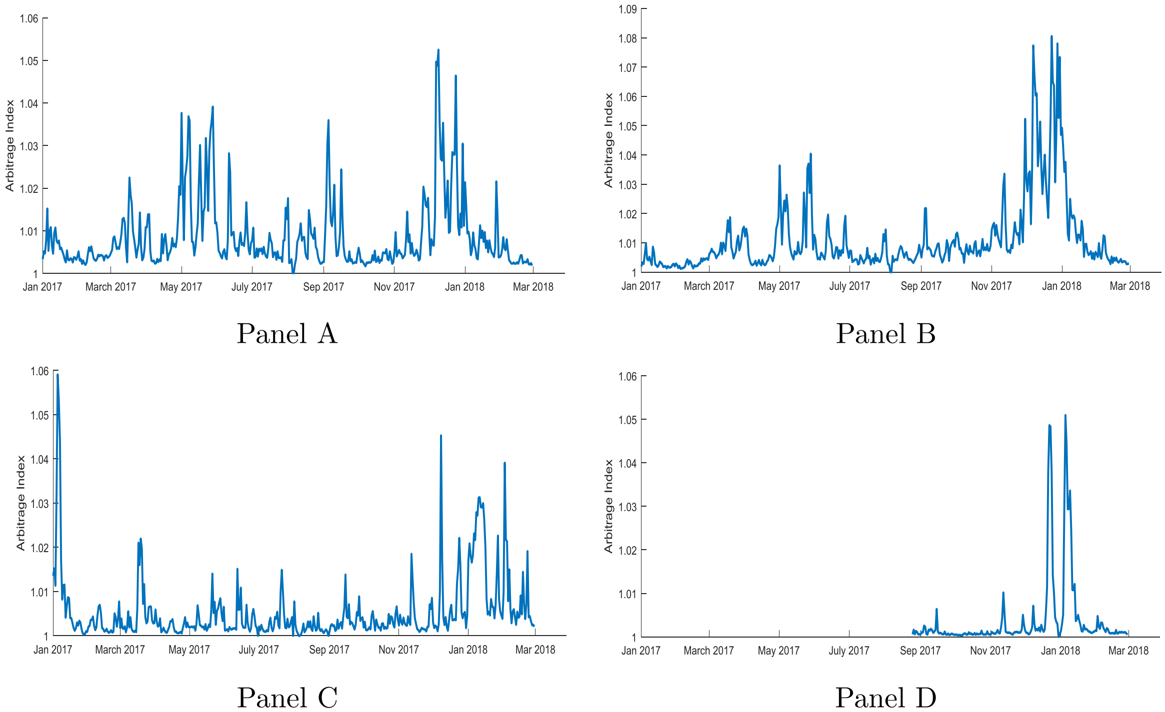
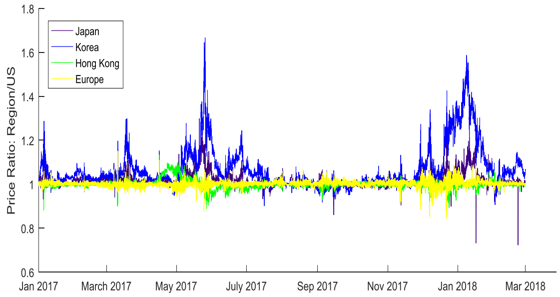
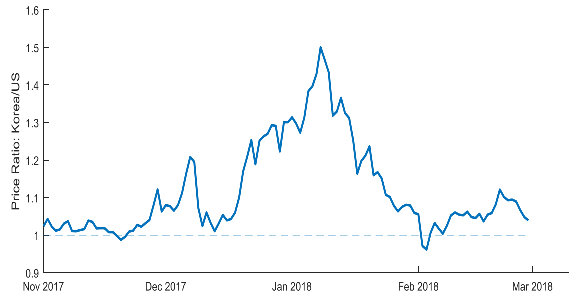


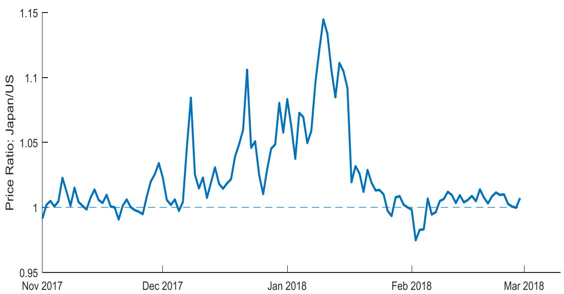
Figure 5. Arbitrage index within regions. This figure calculates the arbitrage index for bitcoin to fiat currency on all the exchanges within a region from January 2017 until February 28th 2018. The arbitrage index is calculated based on the volume-weighted price of bitcoin per minute for each exchange and averaged at the daily level. For a given minute the maximum price across all exchanges is divided by the minimum price in that minute. Outliers are removed by replacing any price movement of more than 10% between two adjacent transactions. Panel A uses data from US exchanges: Bitstamp, Coinbase, Gemini, Kraken: USD. Panel B uses data from Japanese exchanges: bitFlyer, Quonie and Zaif. Panel C uses data from European exchanges: Bitstamp, Kraken and Coinbase: EUR. Panel D uses data for Korean exchanges: Bithumb and Korbit.



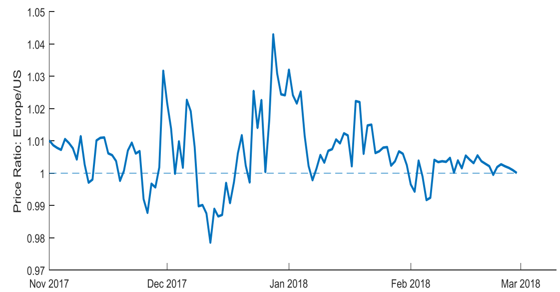
Panel A



Panel B

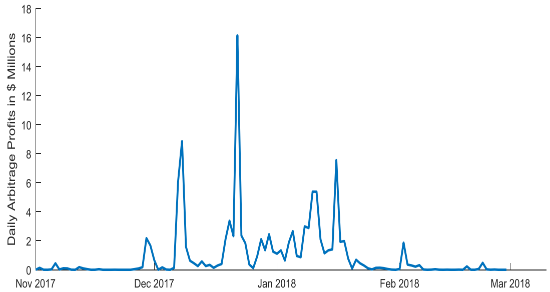


Panel C

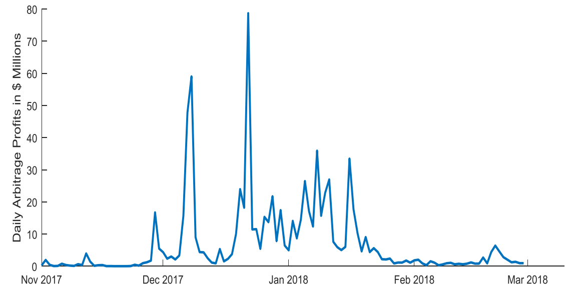


Panel D

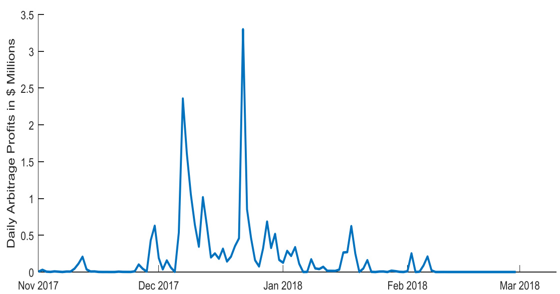
Figure 6. Price ratio across regions. This figure plots the average price ratio between the price of bitcoin to USD across pairs of regions, from January 2017 until February 28th 2018. Panel A reports the price ratio between the US and the other regions, Europe, Hong Kong, Japan and Korea. The ratio is calculated based on the volume-weighted price per minute across all the exchanges in a region and averaged at the hourly level. Panel B plots the price ratio averaged at the daily for the US versus Korea from November 1st, 2017 to February 28th, 2018. Panel C repeats the same calculation for the exchanges in Japan versus the US. Finally, Panel D reports these calculations for the US versus Europe.



Panel A

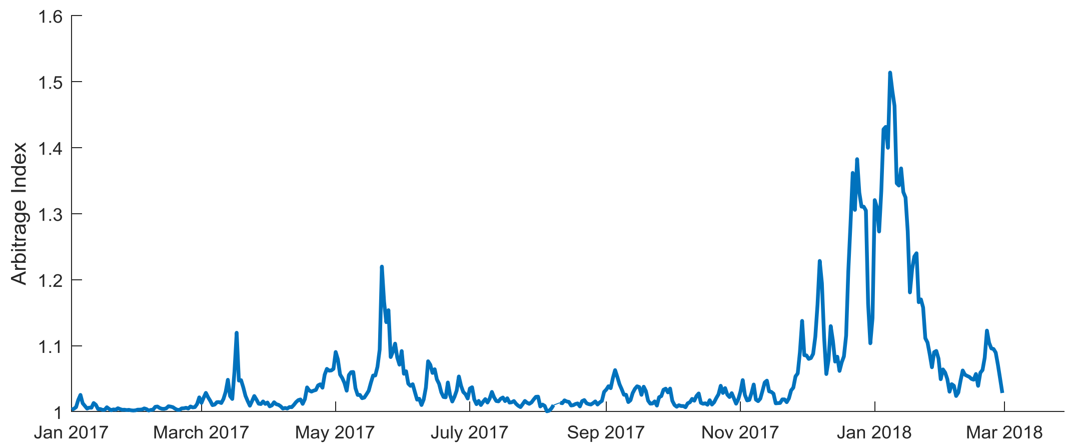


Panel B

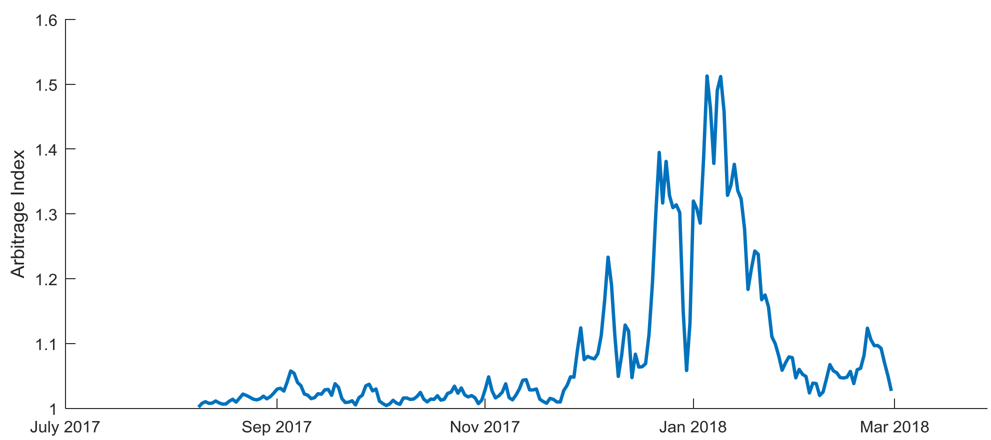


Panel C

Figure 7. Daily profits. This figure plots the arbitrage profits between two regions calculated at the second-level and then aggregated to the daily level. We include only price differences between exchanges if the price difference is larger than 2%. For each second, the aggregate amount of low priced volume that could have been sold in a high price region is subtracted from the sell-initiated volume in the region that has the highest price in a given second. Panel A shows the profit between the US and Korea, Panel B between the US an Japan and finally Panel C between the US and Europe.

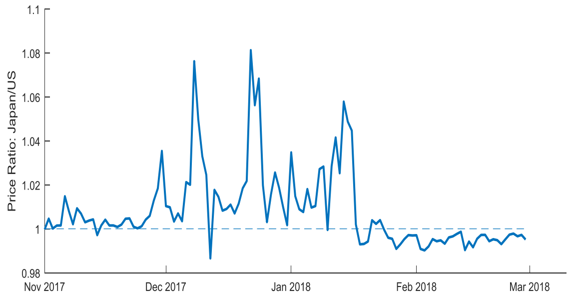


Panel A



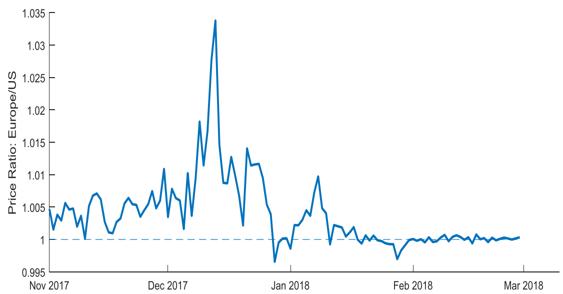
Panel B

Figure 8. Ethereum and ripple arbitrage index. This figure calculates the arbitrage index for ethereum and ripple to fiat currency on all the exchanges from January 2017 until February 28th 2018. Panel A shows the index for ethereum, Panel B for ripple. Each arbitrage index is calculated based on the volume-weighted minute price of the corresponding currency and is then averaged at the daily level. For a given minute the maximum price across all exchanges is divided by the minimum price in that minute.



Panel A

Panel B



Panel C

Figure 9. Ethereum-bitcoin exchange rate across regions. This figure plots the average price ratio between the price of ethereum to bitcoin across pairs of regions, from November 2017 until February 28th 2018. The ratio is calculated based on the volume-weighted price per minute across all the exchanges in a region and averaged at the daily level. Panel A plots the daily price ratio for the US versus Japan. Panel B shows the results for the US versus Korea and finally Panel C reports the results for US versus Europe.

Tables

Exchange	Jan 2017—Jul 2017: Average daily volume ($ millions)	Jan 2017—Jul 2017: Average daily number of trades (thousands)	Jan 2017—Jul 2017: Average size of trades ($ millions)	Jan 2017—Jul 2017: Average bid ask spread (basis points)	Aug 2017—Feb 2018: Average daily volume ($ millions)	Aug 2017—Feb 2018: Average daily number of trades (thousands)	Aug 2017—Feb 2018: Average size of trades ($ millions)	Aug 2017—Feb 2018: Average bid ask spread (basis points)
US								
Coinbase: USD	19.51	27.86	619.45	3.04	180.47	83.90	1,862.88	0.94
	(15.56)	*(13.17)*	*(272.11)*		*(172.03)*	*(56.32)*	*(654.92)*	
Bitstamp: USD	19.92	11.69	1,496.07	12.34	132.11	41.49	3,022.68	12.81
	(17.45)	*(7.18)*	*(485.48)*		*(104.15)*	*(28.24)*	*(594.30)*	
Gemini	13.57	6.90	1,698.51	4.28	77.66	23.41	3,078.75	2.55
	(15.77)	*(5.88)*	*(778.51)*		*(68.20)*	*(14.54)*	*(988.45)*	
Kraken: USD	8.59	9.20	862.10	21.54	45.44	24.45	1,792.53	9.37
	(7.69)	*(7.09)*	*(209.92)*		*(33.12)*	*(14.08)*	*(616.32)*	
Japan								
Bitflyer	23.49	31.91	783.35	6.50	173.98	88.42	1,922.94	6.20
	(11.71)	*(16.53)*	*(267.74)*		*(128.90)*	*(59.72)*	*(503.69)*	
Zaif	6.73	78.42	98.41	3.42	77.30	163.27	472.76	3.78
	(7.62)	*(86.69)*	*(31.98)*		*(50.58)*	*(53.31)*	*(305.25)*	
Quoine	22.18	14.06	1,495.37	18.58	104.52	41.01	2,552.31	11.03
	(19.57)	*(8.26)*	*(563.06)*		*(113.65)*	*(36.15)*	*(1583.85)*	
Korea								
Bithumb	-	-	-	-	142.82	50.66	2,621.91	7.10
	-	-	-		(81.69)	(16.76)	(814.16)	
Korbit	10.73	6.00	1,567.06	-	49.06	15.85	2,761.54	-
	(9.76)	*(4.24)*	*(485.97)*		*(45.37)*	*(9.30)*	*(815.11)*	
Hong Kong								
Bitfinex	25.76	18.14	1,369.84	3.96	445.36	117.88	3,460.44	2.39
	(16.70)	*(8.62)*	*(375.17)*		*(330.12)*	*(70.81)*	*(871.99)*	
Europe								
Kraken: Euro	19.14	20.71	856.02	8.88	78.20	38.71	1,788.59	9.23
	(13.28)	*(10.18)*	*(272.21)*		*(58.11)*	*(23.79)*	*(611.87)*	
Coinbase: Euro	2.05	8.65	213.64	10.48	33.99	42.95	675.67	3.06
	(1.80)	*(5.11)*	*(84.41)*		*(37.92)*	*(39.83)*	*(220.16)*	
Bitstamp: Euro	2.73	3.09	776.46	34.69	32.66	19.91	1,493.87	26.57
	(2.72)	*(2.34)*	*(290.93)*		*(31.98)*	*(17.04)*	*(490.92)*	
Tether								
Poloniex	-	25.48	-	15.34	68.33	61.70	1,134.81	10.40
	-	*(17.08)*	-		*(55.30)*	*(43.29)*	*(442.97)*	
Binance	-	-	-	-	224.41	156.21	1,580.26	-
	-	-	-		*(186.30)*	*(146.21)*	*(590.98)*	
Bitrrex	-	2.26	-	116.46	60.42	39.19	1,383.77	13.78
	-	*(4.67)*	-		*(47.34)*	*(23.97)*	*(368.23)*	
China								
Okcoin	19.48	145.38	875.61	15.25	44.57	30.24	1,529.70	23.18
	(53.30)	*(619.20)*	*(395.48)*		*(38.71)*	*(27.49)*	*(622.11)*	
BTCC	182.31	50.41	1,273.52	31.37	47.75	19.88	2,117.16	48.49
	(758.73)	*(159.83)*	*(1095.22)*		*(38.33)*	*(12.70)*	*(691.83)*	
Huobi	1,284.72	1,820.26	613.32	8.42	-	-	-	17.12
	(1,217.81)	*(784.94)*	*(440.19)*		-	-	-	

Table 1. We report average daily trading volume, number of trades and number of trades for each exchange from tick data. We also add a last column in each panel where we report the average bid ask spread based on data from the order book for each exchange. The first for columns cover the time period from 1/1/2017 to 7/31/2017 and the second from 8/1/2017 to 2/28/2018. Exchanges are organized by the seven regions: US, Japan, Korea, Hong Kong, Europe, tether and China. Tether is a cryptocurrency that serves as the base currency for several exchanges, which we group together, since they do not have a specific geographic region of trading.

Return frequency	Std. Dev	Skewness	Kurtosis	ρ_1	ρ_2	ρ_3	cross correlation
5-Minute	1.40	-0.30	259.76	0.07	-0.01	0.01	0.57
Hour	1.22	-0.06	13.86	-0.07	-0.05	-0.01	0.83
Daily	1.07	0.29	3.85	-0.01	0	0.02	0.95

Table 2. This table describes the higher moments of bitcoin returns at the daily, hourly and 5-minute level from January 1st 2017 to February 28th, 2018. For each frequency we report the annualized standard deviation, skewness and kurtosis of returns; as well as the autocorrelation at 1, 2 and 3 lags (1 to 3) and cross-correlation across exchanges. These statistics are calculated across all the exchanges in our data, but without the Chinese exchanges due to data availability.

	Bitfinex	Coinbase USD	Bitstamp USD	Gemini	Kraken USD	Kraken EUR	Coinbase EUR	Bitstamp EUR	bitFlyer	Quoine	Zaif	Bithumb	Poloniex	Bittrex
5-min frequency														
loadings	0.354	0.124	0.102	0.046	0.034	0.050	0.020	0.016	0.088	0.041	0.026	0.031	0.034	0.033
weights	0.442	1.173	1.195	0.702	1.548	1.148	4.718	1.900	0.951	0.278	1.962	1.284	1.712	1.931
$\mathbf{R^2}$	0.596	0.576	0.526	0.211	0.307	0.326	0.453	0.202	0.420	0.084	0.298	0.251	0.331	0.353
hourly frequency														
loadings	0.325	0.132	0.102	0.049	0.045	0.062	0.022	0.017	0.079	0.026	0.030	0.035	0.039	0.036
weights	0.420	0.809	1.210	0.825	2.537	1.580	3.978	1.681	0.679	0.102	1.471	0.862	1.802	1.733
$\mathbf{R^2}$	0.672	0.608	0.647	0.352	0.623	0.586	0.558	0.276	0.424	0.034	0.376	0.288	0.497	0.465
daily frequency														
loadings	0.312	0.125	0.106	0.052	0.052	0.071	0.015	0.019	0.073	0.016	0.032	0.042	0.042	0.042
weights	0.375	0.326	1.268	1.488	3.256	1.698	1.791	1.669	0.370	0.052	1.709	0.519	2.205	1.992
$\mathbf{R^2}$	0.669	0.393	0.704	0.559	0.758	0.677	0.294	0.330	0.297	0.013	0.471	0.256	0.608	0.580

Table 3. This table reports the results of a factor analysis applied to signed volume data on 14 exchanges as described by the model (1) in Section 4 . Each panel reports the factor loadings, weights and R-squared for the first common factor across our 14 main exchanges at 5-minute, hourly, and daily frequency, respectively.

	Bitfinex	**Coinbase USD**	**Bitstamp USD**	**Gemini**	**Kraken USD**	**Kraken EUR**	**Coinbase EUR**	**Bitstamp EUR**	**bitFlyer**	**Quoine**	**Zaif**	**Bithumb**	**Poloniex**	**Bittrex**
5-min frequency														
loadings	1.121	1.023	1.033	1.036	0.697	0.697	0.929	0.972	0.840	0.926	0.828	0.819	1.077	1.062
weights	0.162	0.112	0.121	0.164	0.032	0.030	0.048	0.050	0.051	0.024	0.021	0.032	0.099	0.054
$\mathbf{R^2}$	0.885	0.818	0.832	0.876	0.442	0.425	0.615	0.639	0.608	0.436	0.378	0.486	0.804	0.676
hourly frequency														
loadings	1.033	0.986	1.003	1.005	0.964	0.963	0.966	0.990	0.892	0.949	0.905	0.853	1.040	1.077
weights	0.137	0.117	0.142	0.153	0.062	0.041	0.031	0.077	0.022	0.023	0.017	0.015	0.098	0.064
$\mathbf{R^2}$	0.962	0.953	0.962	0.966	0.908	0.865	0.825	0.928	0.754	0.773	0.708	0.664	0.946	0.919
daily frequency														
loadings	1.032	0.984	0.998	1.001	0.965	0.982	0.951	0.977	1.097	1.115	1.123	0.986	1.018	1.021
weights	0.082	0.056	0.308	0.155	0.068	0.047	0.022	0.098	0.008	0.009	0.008	0.004	0.074	0.061
$\mathbf{R^2}$	0.988	0.982	0.998	0.994	0.985	0.978	0.951	0.990	0.888	0.903	0.887	0.796	0.987	0.984

Table 4. This table reports the result of factor analysis applied to log-return data from 14 exchanges as described by model (2) in Section 4 . Each panel reports the factor loadings, weights and R-squared for the first common factor across our 14 main exchanges at 5-minute, hourly, and daily frequency, respectively.

	5-min frequency			**hourly frequency**			**daily frequency**		
	$\lambda \times 10^4(\%)$			$\lambda \times 10^4(\%)$			$\lambda \times 10^4(\%)$		
s_t^*	8.8	9.9	10.1	6.0	6.6	6.6	3.6	3.9	4.0
	(80.06)	(86.19)	(88.05)	(35.12)	(39.7)	(40.41)	(16.92)	(19.93)	(18.96)
s_{t-1}^*		-3.1	-2.6		-2.1	-2.0		-1.1	-1.1
		(-36.54)	(-32.24)		(-16.53)	(-15.67)		(-4.05)	(-3.62)
s_{t-2}^*			-0.8			-0.4			-0.0
			(-11.68)			(-3.71)			(-0.2)
s_{t-3}^*			-0.5			-0.1			-0.1
			(-7.56)			(-1.22)			(-0.76)
s_{t-4}^*			-0.4			-0.3			-0.3
			(-6.88)			(-3.00)			(-1.71)
s_{t-5}^*			-0.3			-0.1			0.3
			(-5.24)			(-1.33)			(1.57)
R^2	0.54	0.60	0.61	0.6	0.66	0.67	0.69	0.75	0.76

Table 5. This table reports the results from time-series regressions of the common component of returns on the contemporaneous and lagged common component in signed volume extracted using data from our 14 main exchanges:

$$r_t^* = \lambda s_t^* + \sum_{\tau=1}^{T} \lambda_\tau s_{\tau-1}^* + \varepsilon_t.$$

The first three columns report the estimates at the 5-minute frequency, the next six columns show results at the hourly and then daily frequencies. T-statistics are computed using the MacKinnon and White's (1985) heteroskedasticity robust standard errors and are given in parentheses.

	Bitfinex	Coinbase USD	Bitstamp USD	Gemini	Kraken USD	Kraken EUR	Coinbase EUR	Bitstamp EUR	bitFlyer	Quoine	Zaif	Bithumb	Poloniex	Bittrex
5-min frequency														
$\gamma_i \times 10^{-2}$	-1.12	0.86	-1.29	-0.45	-1.61	-0.42	0.04	-0.22	-0.49	0.27	-0.15	0.05	0.14	-0.23
	(-2.4)	(4.12)	(-4.9)	(-2.66)	(-15.41)	(-4.54)	(1.27)	(-5.16)	(-7.78)	(2.92)	(-7.23)	(7.12)	(3.18)	(-5.35)
b_{1i}	0.09	0.17	0.08	0.06	0.07	0.09	0.13	0.15	0.15	0.22	0.23	0.21	0.13	0.15
	(5.57)	(17.72)	(6.1)	(6.18)	(6.58)	(8.39)	(11.6)	(11.73)	(17.55)	(16.83)	(21.31)	(32.29)	(9.9)	(13.42)
b_{2i}	0.06	0.08	0.07	0.02	0.03	0.05	0.05	0.04	0.05	0.08	0.07	0.09	0.06	0.04
	(6.52)	(9.24)	(6.15)	(2.85)	(2.74)	(4.64)	(5.35)	(3.91)	(6.76)	(7.52)	(8)	(14.23)	(4.67)	(4.71)
b_{3i}	0.05	0.08	0.04	0.02	0.02	0.03	0.08	0.04	0.06	0.09	0.07	0.08	0.03	0.06
	(5.19)	(9.82)	(5.13)	(3.04)	(2.32)	(3.7)	(8.8)	(5.34)	(7.9)	(8.18)	(7.87)	(12.34)	(2.52)	(7.01)
R^2	0.02	0.05	0.01	0	0.01	0.01	0.03	0.03	0.04	0.09	0.08	0.08	0.02	0.03

Table 6. This table reports the results from time-series regressions of the idiosyncratic component of the signed volume on each of the exchange list on the top of the column, regressed on the deviation of the price from the common price component and past three lags of the idiosyncratic component of the signed volume of the same exchange . The idiosyncratic components, $\hat{s}_{it}$ and $\hat{p}_{it}$ and are obtained as the residual values of signed volume and prices after taking out the common component from each.

$$\hat{s}_{it} = \gamma_i \hat{p}_{it-1} + b_{1i}\hat{s}_{it-1} + b_{2i}\hat{s}_{it-2} + b_{3i}\hat{s}_{it-3} + \varepsilon_{it}.$$

T-statistics are computed using the MacKinnon and White's (1985) heteroskedasticity robust standard errors and are given in parentheses.

	Bitfinex	Coinbase USD	Bitstamp USD	Gemini	Kraken USD	Kraken EUR	Coinbase EUR	Bitstamp EUR	bitFlyer	Quoine	Zaif	Bithumb	Poloniex	Bittrex
5-min frequency														
$\lambda_i \times 10^4(\%)$	2.86	17.35	5.76	8.37	40.95	41.66	172.03	15.8	17.13	4.35	59.61	32.1	20.1	22.66
	(16.49)	(22.83)	(9.18)	(14.35)	(21.14)	(27.66)	(25.64)	(7.43)	(22.26)	(6.58)	(13.34)	(25.13)	(12.28)	(14.00)
a_{1i}	0.6	0.63	0.55	0.59	0.56	0.63	0.73	0.5	0.83	0.79	0.84	0.83	0.61	0.6
	(48.44)	(16.28)	(56.57)	(34.58)	(43.48)	(40.07)	(29.02)	(25.25)	(40.69)	(26.36)	(14.73)	(50.95)	(54.99)	(61.34)
a_{2i}	0.23	0.18	0.23	0.24	0.2	0.19	0.16	0.26	0.12	0.15	0.01	0.12	0.21	0.21
	(17.07)	(5.58)	(21.47)	(13.5)	(14.75)	(11.51)	(4.18)	(18.78)	(4.8)	(5.55)	(0.08)	(6.45)	(19.32)	(21.32)
a_{3i}	0.16	0.18	0.2	0.16	0.21	0.16	0.1	0.23	0.04	0.05	0.15	0.05	0.17	0.18
	(12.84)	(5.51)	(21.89)	(11.18)	(19.68)	(9.62)	(4.1)	(13.54)	(1.79)	(2.59)	(3.3)	(3.64)	(16.56)	(18.78)
R-square	0.98	0.97	0.94	0.96	0.89	0.95	0.98	0.95	0.99	0.98	0.98	0.99	0.99	0.98

Table 7. This table reports the results from time-series regressions of the deviation of the price from the common price component on each of the exchange list on the top of the column, regressed on past three lags of the deviation of the price from the common price component and the idiosyncratic component of the signed volume of the same exchange . The idiosyncratic components, $\hat{s}_{it}$ and $\hat{p}_{it}$ and are obtained as the residual values of signed volume and prices after taking out the common component from each.

$$\hat{p}_{it} = \lambda_i \hat{s}_{it} + a_{1i}\hat{p}_{it-1} + a_{2i}\hat{p}_{it-2} + a_{3i}\hat{p}_{it-3} + \varepsilon_{it}.$$

T-statistics are computed using the MacKinnon and White's (1985) heteroskedasticity robust standard errors and are given in parentheses.