TOO GOOD TO BE TRUE: ARBITRAGE IN BITCOIN MARKETS

EVIDENCE FROM THE U.S., EUROPE, SOUTH KOREA, AND JAPAN

HERMAN ROGEFORS

ELLINOR MYRMAN

Bachelor Thesis

Stockholm School of Economics

2021

Too Good to be True: Arbitrage in Bitcoin Markets

Abstract:
This paper examines arbitrage opportunities between cross-regional bitcoin exchanges from tick-level data from September 1, 2020, to January 31, 2021. We continue with a regression analysis to provide explanations for the cross-regional price differences. Our findings support that arbitrage opportunities exist, although to a lesser extent than three years ago. The reasoning behind this follows increased integration of markets, adoption of the instrument, and a more trustworthy sample. The largest price difference persists between the U.S. and South Korea, with a maximum of approximately 6%. In contrast, the largest average price difference prevails between the U.S. and Japan. Regressing cross-regional price differences on investor attention and stock market development show no unambiguous significance. Not unexpectedly, the random walk theory seems invincible in predicting the price differences due to the non-integrated nature of bitcoin markets worldwide.

Keywords:

Bitcoin, arbitrage, capital control, investor attention, Newey-West estimator, random walk theory

Authors:

Herman Rogefors (24440)

Ellinor Myrman (24497)

Tutors:

Dong Yan, Assistant Professor, Department of Finance

Karl Wärneryd, Assistant Professor, Department of Economics

Examiner:

Adrien d’Avernas, Assistant Professor, Department of Finance

Bachelor Thesis

Bachelor Program in Business and Economics
Stockholm School of Economics

1. Introduction

Bitcoin, the most prominent cryptocurrency, has experienced substantial price movements and recognition from both individual and institutional investors during the last years. Although a relatively new phenomenon, with its inception in 2009, it is of the most debated financial instruments today, with extensive media coverage each week. In 2017-2018, the bitcoin market experienced a boom, and both the number of active traders as well as prices increased rapidly, although followed by a subsequent fall. The bitcoin media coverage and price have not reached the levels of 2017 until now. During the fall of 2020, prices rose rapidly, along with increased interest from investors. It is an ongoing debate whether the bitcoin market in 2020-2021 is experiencing a bubble once more, and if so, what might differ this time around. fields

Prior research has studied the pricing mechanisms to partially explain bitcoin price booms (Cheah and Fry, 2015; Corbet et al., 2018). Additionally, there is an ongoing debate on the market characteristics of bitcoin and whether it can be said to violate the efficient market hypothesis or not. Following this debate, several papers' findings are divided regarding the inefficiency of bitcoin (Urquhart, 2016; Bariviera, 2017; Nadarajah and Chu, 2017). Building on the price booms and the proposed market inefficiency, Makarov and Schoar (2019) research whether arbitrage opportunities are present between geographical regions during the 2017-2018 price boom. No one has yet examined whether the market inefficiencies and regional price differences persist during today's price boom of 2020-2021. Furthermore, no research has yet examined what drivers are behind these price differences in the first place, in addition to capital controls. This is the gap that we attempt to fill with this paper.

We show that price differences across regional exchanges persist, even though to a lesser degree than in previous years. We form an arbitrage index by analyzing datasets of tick-level bitcoin transactions over four months, using the method outlined by Makarov and Schoar (2019). The transactions are used to form a minute-level arbitrage index, which is then aggregated to a daily level. The study is limited to one exchange from each of the four regions to ensure that only trustworthy exchanges are included. The four regions form a cross-regional arbitrage index. The exchanges and regions in our paper are Bitstamp in the U.S., Kraken in Europe, Korbit in South Korea, and bitFlyer in Japan. The most significant price divergence for all regions is 9.2% between Japan and South Korea, while the largest dispersion with the U.S. as a basis is 6% between South Korea and the U.S.

Moreover, there is consistently more than 3% price difference across all regions during the 2.5 months from mid-November 2020 to the end of January 2021. Comparatively, the arbitrage opportunities in 2017-2018 reached a maximum of approximately 60%. The results show that when bitcoin experiences a rapid appreciation, arbitrage opportunities open up across all regions. Thus, our findings further confirm that the proposed arbitrage opportunities persist over long periods, which is unique for bitcoin compared to other, more integrated instruments. Although profits are possible by trading between the U.S. and Asian regions, no unison direction of trade can be established, since ultimately prices are lower in Japan and higher in South Korea, compared to the U.S.

Deviations from the law of one price have been proved to exist due to limitations of arbitrage (De longe et al., 1990; Gromb & Vayanos, 2002; Gromb & Vayanos, 2018). In order to shed light upon this occurrence, we study potential drivers of the price difference between the regions by regressing attention through the proxy of Google searches and stock market development in each of the countries on the price differences.

We construct a model regressing cross-regional price differences pairwise on investor attention through the proxy of Google searches and stock market development. These independents are chosen given that they previously have been able to predict the pricing of instruments, where attention has been shown to predict stock prices (Da, Engelberg & Gao, 2011). For the stock market development, the correlation between bitcoin and the S&P 500 spiked in 2020, reaching an all-time high of 0.229, on average (Morningstar 2021, cited in Vaneck, 2021). Our findings show that solely regressing price differences on attention and stock market development does not have significant explanatory value. This result is not entirely unexpected since we go beyond domestic pricing and look at price differences. Thus, the non-integrated nature of bitcoin markets will largely affect the results, ultimately making it difficult to find unison domestic price drivers. At the cross-regional level, primarily affected by segmented markets, price differences tend to take on a more random pattern. We confirm that the random walk theory seems invincible in this case, and not unexpectedly, is the main explanation for price differences between regions.

The rest of the paper is organized as follows. Section two displays background on the bitcoin ecosystem and current literature on closely related topics. This section also covers how our paper is positioned with regards to current findings, as well as our hypothesis development. Section three covers our data, statistical analysis, and variable construction. In section four, we display the results for our hypotheses. In section five, we discuss the findings for arbitrage opportunities and the drivers of price differences. In section six, we evaluate limitations, propose coming research orientation, and provide a conclusion of our study.

2. Background & Literature review

2.1 Background

The first application of the blockchain technique was introduced in a paper by the pseudonym Nakamoto (2008), while bitcoin was released in 2009 as open-source software. Since then, it has evolved from a niche software with limited use to a recognized currency with over 300 000 transactions per day and a market cap of over 1.2 trillion dollars (Blockchain, 2021). What differentiates cryptocurrencies from conventional fiat currencies is the peer-to-peer network that verifies transactions, making third-party authorization bodies, such as central banks, obsolete (Nakamoto, 2008). The system is fully decentralized, and users of the network sign each transaction digitally in a public record called *blockchain*. Blockchain technology guarantees that each transaction is authentic and prohibits the fraudulent risk of double-spending that digital currencies face. Anyone can contribute to the process of signing transactions by allocating computer power to the network, and those who do are compensated with new bitcoins and called "miners". Therefore, the supply of bitcoin will continue to increase as the miners are rewarded for their effort. Moreover, there is a finite supply of bitcoin with a total of 21 million coins expected to exist in the year 2140, compared to the current circulation of 18.7 million (Blockchain, 2021).

A report by Bitwise Asset Management (2019) proved that 95% of reported volume among the 81 largest exchanges was fake. The fraudulent exchanges proved to contain certain elements, such as significantly larger spreads which allowed for outlier transactions to be reported. Misreporting volume and fraudulent behavior on bitcoin exchanges have also been examined and confirmed in the academic world by Gandal et al. (2018).

Unlike the stock market, the bitcoin universe does not have a unifying mechanism to ensure investors will get the best price at each transaction. As a result, there are grounds for arbitrage opportunities to arise for investors who trade on multiple exchanges. Arbitrage is defined as "the simultaneous purchase and sale of the same, or essentially similar, security in two different markets for advantageously different prices" (Sharpe & Alexander, 1990). As Shleifer and Vishny (1997) describe, the theoretical definition of arbitrage necessitates no capital and entails no risk, while it often requires capital and entails risk in practice, which is the case for arbitrage across bitcoin exchanges.

There are two ways to act on bitcoin arbitrage opportunities: either through simultaneous trades or through the transfer of bitcoin between exchanges (Makarov & Schoar, 2019). In the first case, a trader would need capital on two exchanges and simultaneously buy bitcoin on the exchange with a lower price while short selling on the other exchange, if short selling is allowed on the exchange in question. The capital needed to execute this strategy will be challenging to borrow due to banks' strict money laundering policies, with an especially hesitant view on crypto trading, and additionally, there is convergence risk involved. In the second case, a trader would buy bitcoin on the exchange with a lower price, transfer the coins to the other exchange, and sell them for a higher price. This strategy also has complicating issues, such as the time required to transfer coins from one change to another.

Makarov and Schoar (2019) finds capital controls to be the primary reason for the hefty price differences shown in 2017-2018 across regions. Capital controls entail limiting the in- and outflow of capital in a region and usually take shape as taxes, tariffs, legislation, volume restrictions, or market-based forces. These constraints are enforced by the government or other regional regulatory bodies. Fernández et al. (2016) has presented a dataset covering capital controls over the world, and during the period from 1990 to 2017, it is clear that capital controls tend to be quite stable over time, and hence the data presented for 2017 can be assumed to prevail today, three years later.

In this paper, most of the researched currencies, JPY, USD, and EUR, do not have substantial restrictions on their flow on the foreign exchange. They are considered to be traded in an unlimited sense on the foreign exchange, and although the need for declaration exists, they do not take on the same complex nature as the South Korean won, KRW. The KRW is a complex currency, and there is a debate whether it is a restricted currency. In the 1990s, the South Korean government eased its regulations and opted to allow the currency to float freely on foreign exchanges. However, further restrictions were imposed in 2010 due to South Korea having foreign debt rising to 60% of their foreign reserves. These regulations targeted the equity and derivative market, and domestic banks' derivatives were capped at 50% of their equity capital.

What investors are actively paying attention to is difficult to measure. Choi and Varian (2009) show that a Search Volume Index (SVI) built on Google searches can effectively predict house and automotive sales as well as tourism patterns. Da, Engelberg, and Gao (2011) further build on this, arguing that Google searches are suitable for capturing what the vast majority are paying attention to and constitutes a revealed, direct attention measure. To quantify such a qualitative measure as attention can be difficult since attention is scarce, and there is no direct way to measure it. Prior to Da, Engelberg, and Gao's introduction of Google searches as an SVI, more indirect, traditional attention measures were commonly used, such as extreme returns, trading volume, and news. A search in Google is a revealed attention measure since the act of searching for something in Google reveals that the person is actively paying attention to it. Da, Engelberg, and Gao further state that Google searches tend to capture the attention of uninformed investors on the lookout for more information. Google continues to be the market leading search engine, holding an 86.6% market share worldwide in February 2021. Google also provides the tool *Google Trends* to analyze what people are searching for, which reports regional daily and weekly search data.

We conduct a joint test for heteroscedasticity, following a Breusch-Pagan (1979) methodology and with Cook-Weisberg (1983) suggestions. The test examines whether the variance of error terms is dependent on the independent variables. If so, heteroscedasticity is present in the model, which violates the OLS assumption of homoscedasticity.

Secondly, the Breusch-Godfrey test can test the hypothesis of small autocorrelation in the error terms and determine whether the OLS assumption of no autocorrelation holds. The Breusch-Godfrey test is built upon a suggestion to the first presented test for autocorrelation, the Durbin-Watson (1950) technique. The problem with the Durbin-Watson is that it only can test for the first-order autocorrelation, which is inefficient and non-valid for dynamic models (Maddala, 1992).

Should the Breusch-Pagan/Cook-Weisberg and Breusch-Godfrey tests show results indicating a dataset with heteroscedasticity or autocorrelation, the standard errors can be made robust for both heteroscedasticity and autocorrelation using a Newey-West estimator (Newey & West, 1987). This estimator will adjust the long-run variance for the model to be applicable even when standard OLS assumptions are violated. More specifically, it will create a covariance matrix estimate.

Morningstar (2021, cited in Vaneck, 2021) presents correlation statistics for 2020, suggesting that the correlation of bitcoin and S&P500 averaged at 0.229 throughout the year. The correlation indicates a vast increase compared to the eight-year average of 0.01 from 2012 to 2019. The reason behind such a difference in correlation might be the macroeconomic year of 2020 in general. On the other hand, Vaneck presents that increased adoption of bitcoin may cause such a correlation increase and the record-high trading volumes in the instrument. Furthermore, the OTC-traded bitcoin funds also saw a vast increase during the year, supporting the increased adoption of the instrument. Nevertheless, Vaneck expects this increased adoption to continue as an upgoing trend in subsequent years.

2.2 Literature review

Relating to our paper, Hayes (2018) finds support for a fundamental value of a bitcoin derived from the production price of mining, even though the actual price differs from the implied fundamental value for months at a time. In addition, several papers examine the volatility of bitcoin in relation to other assets and whether it is exposed to significant tail-risk. Gkillas and Longin (2018) and Bouri et al. (2017) find that bitcoin can act as a safe haven and diversifier in an equity portfolio, and Borri (2019) finds that cryptocurrencies are prone to tail-risk within cryptocurrency markets, but not in other asset classes.

In an efficient market, new information about said security is incorporated instantly in the price of a security, as defined in the efficient market hypothesis (EMH) by Fama (1970). Therefore, in an efficient market, investors will not achieve an above-average risk-adjusted return. Fama (1970) identifies three degrees of efficiency: weak, semi-strong, and strong. The weak form constitutes that abnormal returns cannot be attained from only past price information, as past prices and returns follow random walks. The semi-strong degree constitutes that prices only reflect public information, and hence investors with privately held information can achieve an excess return. The strong form constitutes that prices reflect both public and private information, and hence no excess return whatsoever can be attained.

The EMH is based on a solid theoretical framework and is useful both for assumptions of rational and irrational investors. The rational investor infers the value of a security by the net present value of future cash flows, which is discounted with regards to attributed risk. New information will rapidly be integrated into the prices, and hence the prices will swiftly move to the new fundamental value. The EMH can also hold in two cases of irrational investors, depending on whether they trade randomly. When irrational investors trade randomly using uncorrelated strategies and in sufficiently large numbers, their trades will cancel each other out. In the case of correlated trading strategies among irrational investors, the security will differ from the fundamental value. As a result, savvy investors or arbitrageurs will utilize the price difference, and the price will move towards the fundamental value (Shleifer, 2000). A significantly high number of small arbitrageurs will drive prices towards fundamental values (Shleifer and Vishny, 1997).

There are contradictory beliefs whether bitcoin experiences weak form efficiency, i.e., that past information cannot determine future returns. Urquhart (2016) concludes that bitcoin is inefficient using multiple robustness tests, while Nadarajah and Chu (2017) find returns to be weakly efficient when using power transformation in a replication of said article.

Building on this, multiple factors can act as drivers of bitcoin price. Research indicates a relatively high explanatory value of fundamental factors of currencies such as supply and demand, implementation as a means of trade, and price level itself as to affect bitcoin prices in the long run (Kristoufek, 2015; Athey et al., 2016; Ciaian et al., 2015). In addition, they find investor attention to have a significant impact on price development in the medium-long run but with varying results with different proxies. T The number of bitcoin-related tweets can predict next-day bitcoin trading volume and realized volatility but not returns (Shen et al., 2019), while Google searches do not affect the next-day realized volatility, volume, or returns (Urquhart, 2018).

Our work also relates to the law of one price, which declares that an asset traded on multiple platforms will have the same price. This reasoning relies on the major assumptions of no transaction- or transportation costs, no legal restrictions, and no market manipulation that will allow arbitrageurs to exercise arbitrage opportunities (Isard, 1977). Additionally, arbitrage constraints can explain that a breach of the law of one price exists in multiple markets (De longe et al., 1990; Gromb and Vayanos, 2002; Gromb and Vayanos, 2018). More narrowly, research on dual listing stocks reveals that significant price disparity can persist between markets(Froot & Dabora, 1999; Rosenthal & Young, 1990). Similarly, Makarov and Schoar (2019) find significant price differences in bitcoin exchanges across regions during 2017-2018, most eminently between the U.S. and South Korea.

Several academic papers have studied whether investor attention can predict prices, using Google Trends, which provides a Search Volume Index (SVI) that allows for comparison between countries. Da, Engelberg, and Gao (2011) concluded that investor attention through the proxy of this SVI could predict stock prices. They found that an increase in SVI could predict higher prices on Russell 3000 stocks within the two upcoming weeks and contribute to the abnormally high first-day returns and the long-run underperformance of IPOs. Furthermore, Han, Lv, and Yin (2017) use Google searches to determine whether this direct investor attention measure can predict oil prices and found that the SVI had explanatory value and significant predictive power for oil price movements. In addition, they found that the SVI did well on predictions for the short-term perspective, meaning the daily and weekly frequency. Likewise, they found it to be a highly relevant measure because the attention of retail investors tends to affect short-term decisions.

Our paper examines the law of one price, following a similar methodology as Makarov and Schoar (2019) but with more reliable data. We follow a similar method creating a Volume Weighted Average Price (VWAP) and a corresponding arbitrage index. Furthermore, we test similar hypotheses in a new setting, three years later and with a limit to only one exchange for each region, to test whether the cross-regional arbitrage opportunities persist today. Additionally, we aim to contribute to the debate around whether bitcoin markets experience weak form efficiency (Nadarajah & Schu, 2017) or are entirely inefficient (Urquhart, 2016) by building on existent literature by going broader in examining potential causes of price disparity that have not yet been used in previous research. For this section of the paper, we follow the approach of capturing non-institutional investors' active attention through the Search Volume Index (SVI) Google Trends, as proposed by Da, Engelberg, and Gao (2011) and confirmed by Han Lv and Yin (2017). We also examine whether regional stock market development can act as a price driver. By doing so, we aim to shed light on the conflict of increased correlation with the stock market and the proposed safe haven characteristics (Gkillas & Longin, 2018).

2.3 Hypotheses development

Our paper characterizes as a quantitative empirical study, ultimately observing and analyzing quantitative data to draw conclusions for the research at hand. The quantitative data includes tick-level bitcoin trades from four exchanges on different geographical regions, corresponding stock market indices, and a Search Volume Index (SVI) built on Google Trends. Distinguishing between two different research questions, where the second one acts as an extension to the first, the hypotheses will follow the same division. We use deductive methods to examine whether arbitrage opportunities remain today and to test how stock market development and investor attention might affect the regional difference in bitcoin pricing. En masse, the methodology is deductive. Nevertheless, we use some inductive approaches to look for relationships of predictiveness for the second part of the paper.

Existing literature has established that the price of bitcoin is not consistent with the law of one price during lasting periods and that arbitrage opportunities arise and prevail during long periods (Makarov & Schoar, 2019). Furthermore, there is a debate on whether the bitcoin market is weakly efficient (Nadarajah & Chu, 2017) or inefficient (Urquhart, 2017). It is difficult for investors to assess a fundamental value of a bitcoin, as it does not have any future cash flows to discount, and research has shown that the price differs from the implied fundamental value for extended periods (Hayes 2018). These factors make the assumption of irrational investors plausible, and as crypto tends to be bullish crypto (Alloway & Weisenthal, 2021), we can assume them to have correlated trading strategies. As a result, the EMH can still hold in theory if arbitrageurs can exercise the price differences. However, we see constraints to exercising arbitrage opportunities in bitcoin markets, and as a result, we do not assume the EMH to hold. Moreover, bitcoin markets have several constraints to exercising arbitrage that violates the assumptions of the law of one price, such as capital control and market manipulation.

Over the three years since bitcoin arbitrage opportunities last were researched, the number of active investors has increased together with an increased traded volume. This change suggests that prices should have reached a more fundamental level (Shleifer & Vishny 1997), indicating that potential arbitrage opportunities should have decreased. Likewise, our refined data set only includes four exchanges compared to 17 in the previous study, which would, all else being equal, decrease arbitrage opportunities. To build on this, Bitwise Asset Management (2019) reports that a large amount of volume on bitcoin exchanges is fraudulent, and exchanges included in our sample have a high proven amount of transparency. These factors indicate a lower arbitrage index of 2020-2021 compared to that of 2017-2018, and we form our first hypothesis:

H1: The arbitrage opportunities in bitcoin across four geographical regions between September 1, 2020, to January 31, 2021, are lower than those of January 1, 2017, to March 31, 2018.

Furthermore, previous research on investor attention has concluded that Google search data is a good measure of what retail investors actively pay attention to. Da, Engelberg, and Gao (2011) conclude that Google searches act as an active investor attention measure and that an increase in attention predicts higher prices in the coming weeks. Likewise, the increased correlation between bitcoin and the stock market in 2020 suggests that the general market view may influence bitcoin more today than it did in 2017-2018 when Makarov and Schoar concluded capital control to be the primary influencer. Hence, we form our second hypothesis:

H2: At least one of the independent variables investor attention and stock market development have significant explanatory value to the price differences between geographical regions.

3. Data, variable construction and methodology

3.1 Raw data

3.1.1 Bitcoin trade data & Volume Weighted Average Price (VWAP)

Our initial dataset consists of approximately 22.7 million bitcoin transactions at the tick-level obtained from Bitcoincharts.com, a public website that provides information on the bitcoin network since 2011. Bitcoincharts.com collects the data through querying each cryptocurrency exchange's Application Programming Interface (API) and then compiles them into datasets. The period of interest is September 1, 2020, to January 31, 2021, motivated by having the same price development characteristics as in Makarov and Schoar (2019). Both periods capture a quick rise followed by a subsequent fall. Please see Appendix A for the figures.

Our data covers four of the most liquid and trustworthy exchanges: Bitstamp, Kraken, bitFlyer, and Korbit. There are severe problems of misreporting trading volumes from cryptocurrency exchanges. The problem is based on the notion that there are incentives for the exchanges to report higher trading volume, as this will attract more customers via, for instance, broader media coverage. These dynamics incentivize exchanges to report substantial amounts of counterfeit volume, as reported by Bitwise Asset Management (2019). The report concludes that approximately 95% of the reported volume on the 81 largest exchanges were counterfeit during a sample week in 2017. Multiple factors were analyzed, including order book evaluation, historical trade patterns, and volume spikes.

Additionally, the fraudulent exchanges had significantly larger spreads, which allowed for outlier transactions to be reported. Gandal et al. (2018) has also examined and confirmed counterfeit volume on bitcoin exchanges to be significant. Thus, we include the degree of transparency in the selection of appropriate exchanges to examine, in addition to the self-reported volume. Subsequently, the trustworthiness is expected to decrease the level of price difference in our replication.

The degree of trustworthiness is based on Coinmarketcap's exchange score, which is based on web traffic, liquidity, and reported volume (Coinmarketcap, 2021). These parameters are in line with the report by Bitwise Asset Management (2019) regarding indicators of a trustworthy exchange. Each of the four exchanges in our dataset has been given the highest band *good*, above 6.0 on a 1-10 scale, with an average of 7.7. The exchanges are also cross-examined by Nomics' transparency rating, an index based on each exchange's readiness to hand over historical trade data (Nomics, 2021). We clean for outliers at a 10 percent level compared to the adjacent transaction. Price differences in tick-to-tick data averages around 0.2%, with a maximum of 2% throughout all the used data sets.

Bitstamp, Kraken, and bitFlyer operate in multiple countries and therefore host trading between more than one pair of bitcoin to fiat currency. bitFlyer only allows its customers to trade in the local currency. Contrarily, Kraken and Bitstamp permit customers to change base currency regardless of location and trade bitcoin for USD, GPB, or EUR. However, each currency pair within each exchange is separated in terms of order books. The typical investor only trades in his or her local currency, as this is the default when setting up an account. Therefore, most trades in Euro are executed by European residents, and hence we categorize these trades to reside in Europe. The same reasoning applies to the other currencies.

We collect the exchange rates from the European Central Bank through the proxy of excelrates.com, which compiles a range of currency pairs' rates. The reference rates are set at the daily level by the ECB in conjunction with other European central banks and are cross-examined to ensure reliability. The rates include four decimal points, making the exchange rate pair of Korean Won and USD inaccurate. As a result, we gather this exchange rate from Yahoo Finance, where the data set included six decimal points.

3.1.2 Attention

We create a Search Volume Index (SVI) with data gathered from Google Trends. Google dominates the search engine market, holding an 86.6% market share worldwide in January 2021. Looking at the markets analyzed in this paper, Google holds 92% of the US market, 90% of the European market, 69% of the Japanese market, and 81% of the South Korean market (Statcounter, 2021). Google Trends reports weekly and daily search data, and since it continues to be the dominant actor across all geographical areas, it captures attention in a good way. A search in Google becomes a revealed attention measure since the search for bitcoin in Google reveals that the person is actively paying attention.

The Google Trends data is adjusted in order to facilitate easier comparisons between regions. Each data point is divided by the geographic areas' total searches for the given period to provide a relative popularity measure. For our data set, a comparison is laid out across the four regions, using Germany, France, and the Netherlands as representatives for the European region, given their dominance in bitcoin trading in this region. The results for Germany, France, and the Netherlands are averaged out using their relative weight of European bitcoin trading as a basis, giving Germany a stake of 41%, France 35%, and the Netherlands 24% (Chainalysis, 2020).

We use the search frequency of the term "bitcoin" in the U.S. and Europe, the term "비트 코인" in South Korea (Korean for bitcoin) and "ビットコイン" in Japan (Japanese for bitcoin).

We study the general term "bitcoin" instead of specifying it further because the purpose is to capture what retail investors are paying attention to. Bitcoin is the most widely used term for investors who are interested in bitcoin. Until December 2019, the tool *Google Correlate* existed, where one could see which terms correlated the most to the term one was looking at. In 2017, Urquhart concluded that the terms with the highest correlation to bitcoin had a low relative volume, meaning that they were covered by the use of the general "bitcoin" as well. We assume this pattern to persist today.

The use of a general term will capture the attention of those who are solely interested in learning more about bitcoin and are not exclusive to those with an intent to invest. Therefore, we also look at the combined term "buy bitcoin" to see if it follows a similar pattern across the four regions. See Appendix B. The main reason not to include these data points in the regression is that the general term "bitcoin" has a higher volume, and hence people with an intent to trade are included in the general term. Another reason not to include combined terms is a question of language discrepancies. Relying on *Google Translate* to truly capture how domestic inhabitants would phrase their Google searches is quite naive. Since both Korean and Japanese are foreign languages, it is difficult to fact-check the exact wording and phrase. Therefore, the sole use of the term "bitcoin", but translated for these two countries, is more reliable.

3.1.3 Stock market development

The stock market data originates from Yahoo Finance, covering the period September 1, 2020, to January 31, 2021. When deciding the stock market development proxy, we aim to find a proxy representing the overall stock market. We follow the established conventional market proxy is the S&P 500 for the U.S. stock market. This index is created to represent the broad economy and includes both a variety of sectors and tends to be stable over time, with fewer than ten companies changing annually. For the remaining three regions, we choose indices based on having similar characteristics to the S&P 500 regarding the weighing technique, diversity characteristics, profitability, and trading liquidity.

We use the Nikkei 225 index for Japan and the KOSPI 200 index for South Korea. For Europe, we include one index from each of the three representative regions Germany, France, and the Netherlands. The three European stock indices receive the same weights per country as the attention indices, based on their share of total bitcoin trading in Europe. We include the HDAX index for Germany, CAC Large 60 for France, and the AEX index for the Netherlands.

3.1.4 Number of traders

It is difficult to quantify the number of active traders due to multiple reasons. Wallets can be either active or inactive, rendering the total number of wallets not valid as a measure. Likewise, the total number of transactions on the public blockchain is not appropriate because investors that trade on exchanges share wallets. Trades executed within an exchange will not result in additional transactions on the public blockchain, making the total number of transactions on the public blockchain inaccurate.

Consequently, we use the reported volume from the exchanges in our dataset to assess the number of active traders. Considering they are trustworthy, we extrapolate the change in the number of transactions to the market as a whole. The monthly number of trades in our four sample exchanges increased by an average of 10% compared to the fall of 2017 when bitcoin was subject to similar amounts of buying pressure. Hence, the calculated increase of 10% acts as a good proxy for the increase in the number of traders.

3.2 Summary statistics

3.2.1 Arbitrage index

Table 1. Trading volume

This table details the average daily traded volume in million USD, the average daily number of trades in thousands, and the average size of trades in USD. The exchanges and base currencies are bitFlyer (JPY), Bitstamp (USD), Kraken (EUR), and Korbit (KRW). Each exchange's base currency is transformed to USD. The studied period is September 1, 2020, to January 31, 2021.

Exchange	Average daily volume (million USD)	Average daily number of trades (thousands)	Average size of Trades (USD)
bitFlyer JPY	169.86 (161.08)	56.93 (28.32)	2983.36 (4949.32)
Bitstamp USD	200.75 (215.13)	37.73 (16.23)	5321.02 (6049.37)
Kraken EUR	135.04 (125.04)	46.06 (13.04)	2931.69 (3471.98)
Korbit KRW	6.39 (6.12)	7.44 (5.32)	858.08 (187.51)

Table 2. Summary statistics of returns

This table presents summary statistics of the returns used in the arbitrage index from September 1, 2020, to January 31, 2021. The table includes daily skewness, kurtosis and cross correlation, and the daily standard deviation is annualized. The statistics include the four exchanges in our sample.

Frequency	Standard Deviation	Skewness	Kurtosis	Cross Correlation
Daily	0.6328	-0.29	3.14	0.9464

The statistics are brought forward by using the daily returns over the period, which are averaged across the exchanges. The returns are calculated using the following formula:

$$r_t = \ln\,(p_t/p_{t-1})$$

Where r represents returns, t the day, (t-1) the previous day, and p price in terms of the volume-weighted average price.

The first surprising notation is the kurtosis of 3.14, meaning that it is relatively close to that of the normal distribution value of 3. However, cryptocurrencies are volatile and seldom assumed to follow a normal distribution in their returns. Furthermore, the return data is left-skewed, and it has a skewness of -0.29. Bitcoin is an instrument that, from time to other, is subject to significant tail risk (Borri, 2019). The fact that the dataset is left-skewed during this period is not surprising since significant increases in both traded volumes and prices occurred during the last month of the examined dataset. An annualized standard deviation on the daily frequency of 0.6328 supports the general view that bitcoin is a volatile instrument and returns fluctuate substantially. For comparative purposes, the annualized daily standard deviation of the S&P 500 index is around 10.5%, significantly lower than the calculated 63.28% for the bitcoin market proxied by the four exchanges. The calculated cross-correlation of 0.9464 suggests that the researched markets are not entirely integrated and efficient, and some price differences exist across regions.

3.2.2 Attention

The following results are found by comparing attention through the Google search “bitcoin”:

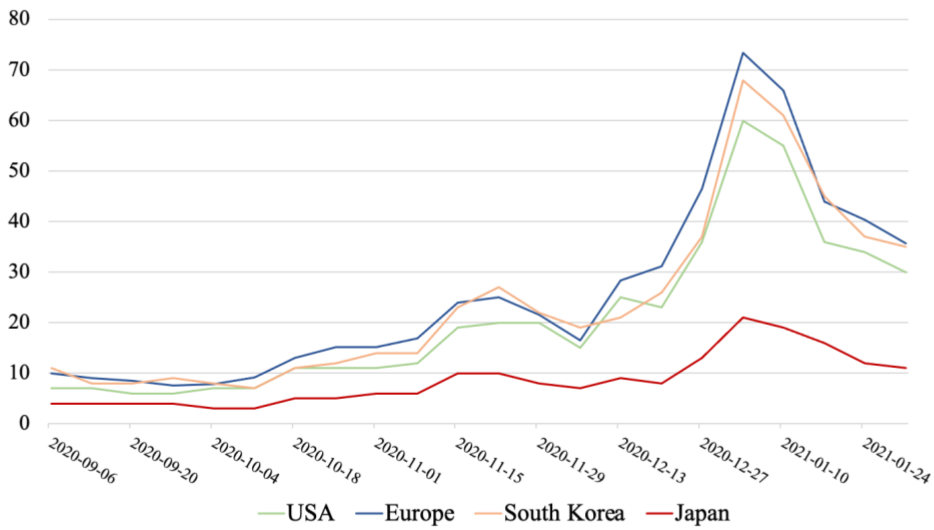

Figure 1. Attention comparison with weekly data.

This figure shows aggregated weekly Google searches for the term “bitcoin” in the U.S., Europe, South Korea, and Japan. Language adjustments have been made for Japan (비트 코인) and South Korea (ビットコイン). The y-axis plots the attention index from Google Trends, and the x-axis plots the date, covering September 1, 2020, to January 31, 2021.

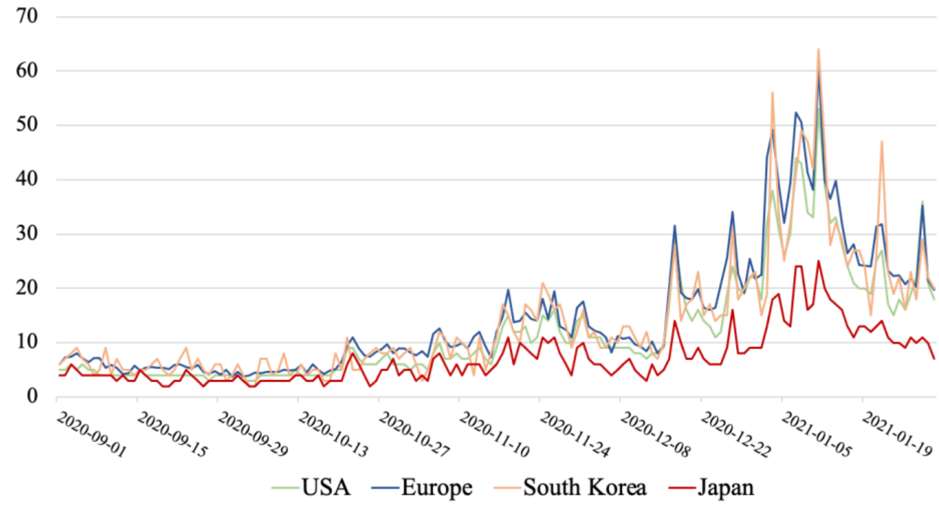

Figure 2. Attention comparison with daily data.

This figure shows the compared daily Google searches for the term "bitcoin" in the U.S., Europe, South Korea, and Japan. Language adjustments have been made for (비트 코인) and South Korea (ビットコイン). The y-axis plots the attention index from Google Trends, and the x-axis plots the date, covering September 1, 2020, to January 31, 2021.

The graphs suggest that Europe and South Korea are the two regions with the highest search frequency, and hence investor attention. The U.S. has slightly lower numbers, while Japan has significantly lower search volumes than the rest of the countries. If attention would be the only driver of bitcoin price, these results suggest that prices are higher in Europe and South Korea compared to the U.S., while prices in Japan are lower than in the U.S.

We also examine searches for "buy bitcoin" on the daily frequency, and the graph can be seen in Appendix B. The combined search term has similar trends, with the highest number of searches in the European region, although followed by the U.S. instead of South Korea. Looking at the results for South Korea and Japan, these cannot be fully relied upon. As previously discussed, searching for multiple combined words or whole sentences in Google Trends relies heavily on finding the correct phrasing. Given that both Korean and Japanese are foreign languages for us, the exact phrasing becomes difficult to fact-check. Thus, this can explain why the Japanese results for "buy bitcoin" have an SVI of zero during some days. Furthermore, the volumes for the term "bitcoin" solely are much larger for each region, meaning that the searches for "buy bitcoin" will be captured by the more general term as well. Thus, to avoid further noise in the model, we use the general term "bitcoin" in the analysis.

3.2.3 Stock market development

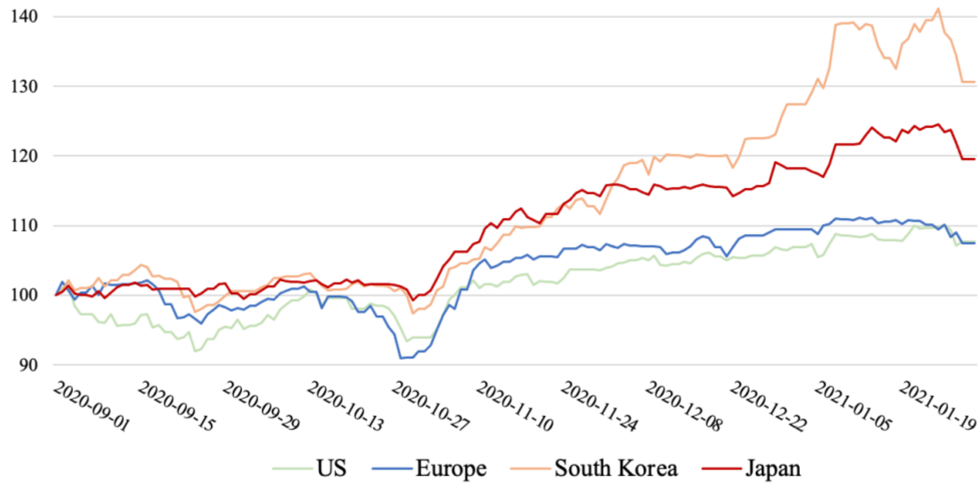

Figure 3. Stock market development

This graph illustrates the stock market development in the U.S., Europe, South Korea, and Japan. The indices are S&P 500 for the U.S, KOSPI 200 for South Korea, and Nikkei 225 for Japan. The European index is based on three regions, with HDAX for Germany, CAC Large 60 for France, and AEX for the Netherlands. The y-axis plots the stock index level, and the x-axis plots the date, covering September 1, 2020, to January 31, 2021.

Because the indices have varying levels, we start by decomposing each index. The new indices are computed with the 1 of September 2020 as a basis with the value of 100. Thus, each index takes on the value of 100 on September 1, and each development after this time is compared using this date as a base. The graph above depicts that the European and U.S. stock markets are reasonably well-integrated and co-move clearly. The integration holds for the complete comparison, with a cross-correlation across all four exchanges of 0.93. The differences between the European and U.S. stock indices are minor, with an average difference of 1.99. In contrast, the average difference between the U.S. and Japan is 8.16, and the U.S. and South Korea 11.53. We see a positive trend in all indices towards the end of the period, most evident in South Korea, with a bullish market by late 2020 and early 2021.

Interestingly, all indices fall in late October, suggesting a decline right before the presidential election in the U.S. This finding enhances the notion that 2020 was a volatile macroeconomic year, which affected the stock market accordingly.

3.3 Statistical analysis

To investigate whether and which macroeconomic factors drive the hypothesized price differences across regions, we conduct a regression analysis including the independent variables investor attention and stock market development. The dependent variable is the price difference between a region and the U.S. The bitcoin price in the U.S. is often established as a proxy for the world market price, and it allows us to compare price difference with findings in previous literature. Therefore, we use the U.S. as a base-case home country throughout the whole paper. We perform two regressions, one for the price difference between Japan and the U.S. and one for the price difference between Europe and the U.S. South Korea is excluded from the analysis. We exclude it due to its capital controls, making this market stand out as an outlier since it becomes more of an isolated market, differentiating it substantially from both the European and Japanese markets in terms of supply and demand.

For Japan, we code the dependent variable as the daily volume-weighted average price in Japan divided by the daily volume-weighted average price in the U.S. Thus, the dependent variable would have a value of 1 if there were no price differences, while the value of 1.05 would imply a 5% higher price in Japan than in the U.S. The dependent variable for Japan has a downward trend, and we make corresponding adjustments in our empirical analysis.

We follow the same methodology when computing the independent variables. To capture the difference in daily stock market fluctuations between the countries, we use the decomposed index for each region. The independent variable for Japan’s stock market difference compared to the U.S. is coded as the daily decomposed index value for Japan divided by the decomposed index value in the U.S. If there were no differences between the stock market developments from the beginning of the period to a specific day, the variable would have a value of 1 for that day. A value of 1.05 would imply that the stock market in Japan has increased 5% more from September 1, 2020, to that day compared to the stock market in the U.S.

To capture the difference in attention, we divide the countries’ Google trends search value for each day. For example, we code the independent variable for Japan as the daily Google trends value for Japan divided by the U.S. value. The variable would have a value of 1 if there were no differences in attention, while the value of 1.05 would imply 5% higher attention in Japan than in the U.S. The variables for Europe are computed following the same procedure as for Japan.

Table 3. Summary statistics for the drivers of price difference

This table presents the number of observations, mean, median, standard deviation, minimum and maximum for the variables used in the regression. Except for the number of observations and standard deviation, the value 1 indicates there is no difference between the countries for that variable. See Appendix D for variable definitions.

Variable	N	Mean	Median	Std. Dev.	min	max
price Europe US	153	1	1.001	.004	.974	1.013
price Japan US	153	.976	.972	.017	.951	1.015
stocks Europe US	103	1.019	1.023	.02	.954	1.065
stocks Japan US	97	1.078	1.084	.036	.994	1.148
attention Europe US	153	1.244	1.233	.157	.907	1.778
attention Japan US	153	.665	.667	.181	.306	1.25

As presented in table 3, the number of stock market observations differ due to a lack of data on weekends and other regional holidays. We eliminate these observations to ensure that the trend is regularly spaced for the time series analysis.

The dependent variable is continuous since all variables are built on time series data. Thus, an Ordinary Least-Squares (OLS) model seems most suitable. The regression model is built as follows, and full derivations can be found in Appendix C.

$$Y_n = \sum_{i=0}^{k} \beta_i X_{ni} + \mathcal{E}_n \quad (1)$$

This general formula is further developed to:

$$Y_n = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mathcal{E}_i \quad (2)$$

Which brings our fitted models:

$$Price\_diff_{Japan/US} = b_0 + b_1 \times Stocks_{Japan/US} + b_2 \times Attention_{Japan/US} + \mathcal{E}_i \quad (3)$$

$$Price\_diff_{Europe/US} = b_0 + b_1 \times Stocks_{Europe/US} + b_2 \times Attention_{Europe/US} + \mathcal{E}_i \quad (4)$$

There are a few assumptions that should be satisfied in order to be fully able to run an OLS model.

We test for heteroscedasticity and autocorrelation to see if the OLS model is suitable for our dataset. Autocorrelation and heteroskedasticity violate the assumptions that the OLS model is built upon and make the model inefficient. However, it does not make the model faulty since it will remain unbiased.

Firstly, we test for heteroscedasticity using both a Breusch-Pagan and a Cook-Weisberg test in three different versions in order for us to test if the null hypothesis of the error term variances are all equal should be rejected or not. In an OLS model, homoscedasticity is necessary. To not reject the null hypothesis, we would need a result with a high chi-square statistic and a correspondingly low p-value.

Secondly, we test for autocorrelation using a Breusch-Godfrey test. Since our data set includes time-series data, it is not unusual to find autocorrelation since values in the previous periods most likely will be relatively close to values in the current and following periods. This serial correlation is present because information not accounted for in the model will be left, causing the model to have a larger prediction interval than needed. The Breusch-Godfrey tests the null hypothesis that there is no autocorrelation in the error terms, and a small p-value will indicate that significant autocorrelation exists.

We use a Newey-West estimator if heteroscedasticity and autocorrelation are detected in the dataset. This model is used to produce Heteroscedasticity- and Autocorrelation-Consistent (HAC) standard errors. Here, the long-run variance of the OLS model (denoted as Ω) usually is derived as:

$$\Omega_t = \sum_{j=-(T-1)}^{T-1} \left(1 - \left|\frac{j}{T}\right|\right) \Gamma_j \rightarrow \sum_{j=-\infty}^{\infty} \Gamma_j \quad (5)$$

Using the Newey-West estimator, this will be replaced by:

$$\widehat{\Omega}^{NW} = \sum_{j=-m}^{m} \left(1 - \left|\frac{j}{m}\right|\right) \hat{\Gamma}_j \quad (6)$$

Where:

$$\hat{\Gamma}_j = \frac{1}{T} \sum_{t=1}^{T} \hat{Z}_t \hat{Z}_{t-j}{}' \quad (7)$$

Where:

$$\hat{Z}_t = X_t \hat{u}_t \quad (8)$$

Following a general rule-of-thumb dating back to the 1990s, the number of lags, captured by the denotation *m*, should be calculated as follows for the Newey-West estimator with no trend (Stock, 2015):

$$m = m_t = 0.75T^{\frac{1}{3}} (9)$$

Where *T* is the number of observations.

When a trend is present, we use the following calculation number of lags:

$$m = m_t = 1.4T^{\frac{1}{3}} \quad (10)$$

4. Results

4.1 Arbitrage index

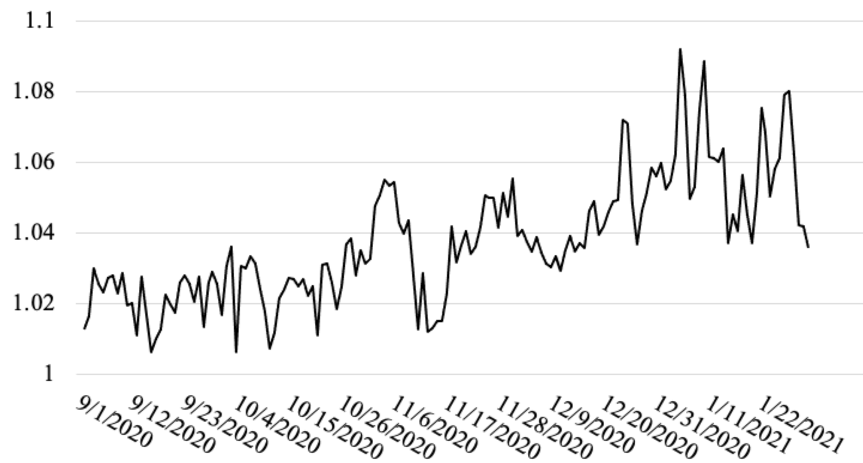

Figure 4. Arbitrage index.

This figure presents the level of arbitrage opportunities across the exchanges bitFlyer, Bitstamp, Kraken and Korbit. The arbitrage index is constructed using tick-level data, which is aggregated to minute-level volume-weighted average price and averaged at daily level. The y-axis plots the arbitrage index level, and the x-axis plots the date for the time period of September 1, 2020, to January 31, 2021.

An arbitrage index is constructed to measure the price differences on the daily level in the price data between markets. In an efficient market free from arbitrage, the index should be equal or very close to 1 at all times. The constructed arbitrage index has an average value of 1.038 over the period, suggesting systematic arbitrage opportunities of approximately 3.8%. The tick-level data is aggregated to minute-level and used to form a volume-weighted average price (VWAP). Using the VWAP as a basis for comparison, the minimum and maximum VWAP across all four exchanges should be similar in an efficient market. To examine whether this is the case, the maximum VWAP across the exchanges is divided by the minimum VWAP for the same minute. The results are averaged at the daily level and presented in the figure above. We see that although index levels are close to 1, differences do exist on the daily level. The maximum difference is approximately 9.2% across all exchanges. The cross-correlation for the arbitrage index of 0.95 also suggests that the bitcoin market is not entirely efficient.

Furthermore, we compare two exchanges at a time, using the U.S. as the basis for comparison. We present the results in the figures below.

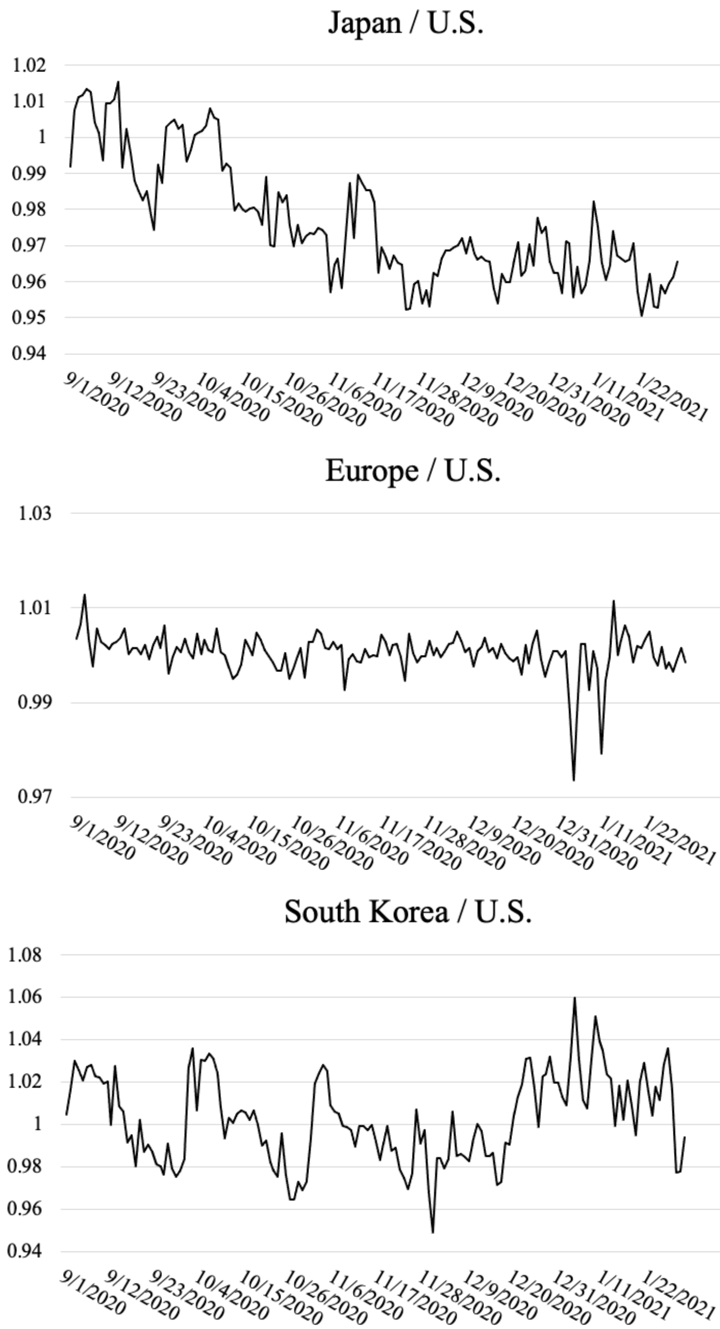

Figure 5-7. Arbitrage index across regions.

These figures present the level of arbitrage opportunities across pairs of exchanges for the time period of September 1, 2020, to January 31, 2021. The first pair is bitFlyer in Japan and Bitstamp in the U.S., the second pair is Kraken in Europe and Bitstamp in the U.S., and the third pair is Korbit in South Korea and Bitstamp in the U.S. The arbitrage index is constructed using tick-level data, which is aggregated to minute-level volume-weighted average price and averaged at daily level. The y-axis plots the arbitrage index level, and the x-axis plots the date.

The largest difference is between South Korea and the U.S. The lowest point in this comparison is 0.949, and the highest point 1.059, suggesting a spread of approximately 11 pps. The Kimchi premium states that crypto prices in South Korea consistently tend to be higher than those in the U.S., which is unambiguous in our results. We also see significant volatility in the comparison between the U.S. and Japan, although with a much lower spread, about 6 pps. Similar market conditions apply in the U.S. and Europe in terms of, for example, capital controls. Thus, the price differences in this comparison are, not surprisingly, minimal and stable. This finding indicates that the non-integration of bitcoin markets should account for a large part of the prevailing price differences between the remaining regions

Looking at the average differences, the largest difference is found between the U.S. and Japan, with an average difference of –2.366%, meaning that prices, in general, are lower in Japan. For South Korea, the average difference is only +0.262%. However, this small average difference may be due to the high volatility and large spread, and because the index is almost as much below 1.0 as above 1.0. This suggests that although significant price differences prevail, and the index seldom stays at 1.0 over periods, no clear tendency as to whether one should buy or sell bitcoin in South Korea can be seen. The Kimchi premium found in 2017 does, however, prevail for long periods, and between December 18, 2020 – 27 January 2021, the premium is clear and averages at +2%. The average between Europe and the U.S. amounts to only +0.038% and suggests that the price differences between Europe and the U.S. are, in general, minimal.

4.2 Drivers of price differences

Table 4. Regression output

This table presents the results of regressing price difference between two regions on differences in attention and stock market development between the regions in an ordinary least squares regression with heteroskedasticity and autocorrelation consistent errors. The table includes two regressions: one for price difference between Japan and the U.S., and one for the price difference between Europe and the U.S. The full sample consists of 153 observations from September 1, 2020, to January 31, 2021, and it is cleaned to ensure the trend is regularly spaced. See Appendix D for variable definitions. Standard errors are presented in in parentheses, where *** = p < 0.01, ** = p < 0.05, * = p < 0.1.

Variables	Japan(/US)	Europe(/US)
Stocks Japan_US	0.475	
	(0.0730)	
Attention Japan US	-0.00791	
	(0.00712)	
Stocks EU US		0.00224
		(0.0147)
Attention EU US		0.00304**
		(0.00146)
Trend	-0.000590	
	(0.0000842)	
Constant	0.959***	0.995***
	(0.0768)	(0.0149)
Observations	97	103

We perform a multiple linear regression using the Newey-West estimator to predict the price difference between a country and the U.S. based on their stock market and attention. The reason for using the Newey-West estimator in contrast to an OLS is that both heteroscedasticity and autocorrelation are found in the samples, ultimately needing HAC standard errors. The predicted price difference for Japan is equal to 0.959+0.475(stocks)-0.00791(attention)-0.00059, where *stocks* is coded as the difference in stock market development, and *attention* is coded as the difference in Google Trend data. The results imply that an increase in the difference in stock market by one unit, which is 100 per cent, would increase the price difference by 47.5%, all else being equal. Contrarily, a one unit increase in attention difference, which is one unit in Google Trends, would decrease the price difference by 0.79%. The coefficient *Trend* implies the negative trend of -0.0059%. Neither stocks nor attention is significant predictors of the price difference between Japan and the U.S.

The findings for price difference between Europe and the U.S. are equal to 0.995+0.00224 (stocks)+0.00304 (attention), with the same coding as for Japan. Increasing the difference in stock market by one unit would increase the price difference by 0.22%, while increasing the difference in attention by one unit would increase the price difference by 0.3%. Attention is a significant predictor of the price difference between Europe and the U.S., while the stock market is not.

In an OLS regression analysis, exogeneity is assumed to prevail, ultimately meaning that the independent variables do not depend on the dependent variable or the error term, but rather the other way around. Inconsistency of the model would prevail if endogenous variables were included in the model. The simultaneity bias, in which the dependent variable predicts the independent variable, could be present in our model. Although, endogeneity is not as critical in our case since we use time series data, and the desired outcome relates more to predictability than causality. Therefore, we need not have estimated parameters that precisely estimate specific causal effects.

We find no significant results in regressing attention and the stock market as price drivers between regions. There is reason to believe that, in accordance with prior research, we have not been able to conclude what drives cross-regional price differences in bitcoin. To find out whether this is the case, we test for the random walk theory. This test first normalizes the price differences and later benchmarks the correlation between the first and second periods in the sample against a positive and negative critical value. This critical value is based on the normal distribution's standard deviation of 1.96 and is calculated using the following formula:

$$\pm \frac{1.96}{\sqrt{n}}$$

Where *n* is the number of observations in the sample, which is 153 for our period. This results in a critical value of $\pm$ 0.158457, and any value *x* within the span of -0.158457 < *x* < 0.158457 would confirm the null hypothesis of a random walk. The following results are calculated for Japan and Europe, respectively:

Table 5. Summary statistics of returns

This table presents the results from the random walk test for the time period September 1, 2020, to January 31, 2021. Correlations are based on normalized values for both Japan and Europe, and the samples are separated into two parts, where the first one constitutes of observation 1-78 and the second part observation 78-153.

	Japan	Europe
Correlation	-0.09801	0.12722

The computed test statistics for both Japan and Europe lie within the spectrum for confirming the null hypothesis. Thus, both countries' results confirm the random walk to prevail. Ultimately, the results indicate that the model including only attention and stock market development does not have significant predictive value. The findings are expected given the insignificant coefficients found in the regression analysis, in addition to prior research failing to prove predictiveness for potential bitcoin price drivers.

5. Discussion of arbitrage, constraints & price drivers

5.1 Findings replication

The constructed arbitrage index shows some differences in the bitcoin price in the different regions. Following Makarov and Schoar (2019), indications of the same arbitrage patterns persist in today's trading environment, but to a lesser extent. The Kimchi premium is present during the last month of the period, with a high point of +6 %.

This result suggests that a trading strategy where one purchases bitcoin in the U.S. for USD and then sells for KRW in Korea should yield a risk-free profit. This strategy would have been particularly beneficial during the first half of January 2021, when one should have been able to yield a profit of approximately 6%. However, the strategy rests on the assumption that we live in a world without frictions which evidently cannot be true in practice. Although constituting as a textbook example of arbitrage, further analysis is needed to conclude whether this opportunity can be realized or not.

Furthermore, there are price differences between the U.S. and the other regions. On average, the largest price differences can be found between Japan and the U.S., averaging at –2.366%. During the period being, the index for Japan and the U.S. is mostly below one, suggesting lower prices in Japan compared to the U.S. This finding corresponds to a trading strategy where one would buy bitcoin in Japan and sell on the U.S. market. During the most lucrative period, this strategy should yield an arbitrary profit of approximately 5%. Since the prices in Japan are systematically lower than in the U.S., and the highest point across all indices originates in South Korea, the most profitable trading strategy would be to buy in Japan and sell in South Korea. However, the advantage of trading between Japan and the U.S. is that capital controls should not be as big of a constraint as it is between South Korea and the remaining regions. Nevertheless, all the proposed trading strategies are exposed to frictions in the real world.

The price differences between Europe and the U.S. are minor, averaging at +0.038%, although the price in Europe was approximately 2% lower than in the U.S. at two times during the period. As previously discussed, the European and U.S. markets are quite integrated. As seen with the stock market indices, these markets tend to have minor absolute differences. Nevertheless, it is possible to yield small arbitrary profits between these two markets. The fact that the two markets are so integrated could be beneficial in exercising arbitrage and avoiding major market frictions and constraints.

5.1.2 Trustworthiness

One should acknowledge that there is much fraudulent volume being reported by bitcoin exchanges. In 2017-2018 arbitrage opportunities of up to 60% were found (Makarov & Schoar, 2019). However, this dataset included exchanges that have later been found to not be fully transparent in their volume reporting, meaning that this result could in part be based on fake volume. Our dataset, with four exchanges proven to be transparent by different grading systems, does not include fake volume to the same extent could also partially explain the vast difference between our result of 9% and that of Makarov & Schoar (2019) of 60%. When accurate volume is analyzed, the actual market behaviors can be captured to a greater extent. Moreover, it behaves more efficiently in reality than what one might expect by just looking at reported volume across different exchanges.

5.1.3 Constraints

The strategy based on buying in a low-priced region, transferring the coins, and selling in a region with higher prices entails a few constraints. Firstly, there is the problem of transaction registration. Miners sign each transaction between bitcoin wallets digitally in the blockchain, a process that generally takes around ten minutes. Although constituting as a complicating issue, the arbitrage index showed that price differences remained for far more extended periods than ten minutes, meaning that this should not fully mitigate the arbitrage opportunities. Secondly, once a transaction has been registered, the translation to fiat currencies can take everything from a few hours up to days to complete. This factor might mitigate the arbitrage opportunities, and it is something that traders need to find their way around. One way to deal with this timing issue is to lock in the arbitrage through simultaneous trades in bitcoin on the two exchanges with different prices.

The alternative strategy of simultaneous trades with short selling could overcome the time constraints. However, out of the four exchanges examined in this paper, only Kraken allows for short selling. Bitstamp and bitFlyer allow for margin trading, although this strategy can expose the trader to significant convergence risk. To complicate things further, Korbit allows for neither short selling nor margin trading, ultimately complicating locking in the wished-for strategy, in which traders buy in the U.S. and sell in South Korea. Hence, in theory, an arbitrage opportunity could be mitigated by the transformation time to fiat currency.

Since the suggested problem relates more to fiat currency translation than bitcoin trading, one strategy is to hedge the currency using other derivatives instruments. For example, one could lock in the KRW/USD exchange rate prior to the trade, estimating when the trade will go through and how long the translation disturbance will last. A trader could enter a futures contract specifying the price of the U.S. dollar against the Korean won, given that they are reasonably sure of when the trade will go through and how much the exchange rate will swing during this period. Since the translation could take only a few hours, using a hedging strategy might not be necessary at all, but given that it is an error term, it should be taken into account before the trade. Different strategies to minimize this error term's effect would need further research to fully establish and hence go beyond the scope of this paper.

5.1.4 Capital control

The fact that bitcoin price differences tend to be volatile between South Korea and the U.S. most likely depends on some outside factors, which complicate the cross-regional transactions and hence lowers the willingness of traders to realize the proposed arbitrage opportunities. Following Makarov and Schoar (2019), cross-border capital control is another complicating factor with transaction costs and governance risk of cryptocurrencies. In South Korea, there is a constraint that any retail investor moving an amount of $50 000 or more out of the country annually needs to provide sufficient reasons and papers for the transfers to authorities. This constraint might cause some hesitation in realizing the arbitrage opportunities for investors trading large volumes of bitcoin and trying to act on prevailing price differences.

The capital controls are even more present for cryptocurrency-related trades, as the South Korean government is reluctant to offer permits to trades with origin in cryptocurrencies due to a fear of money laundering. Retail investors need to prove their physical presence in the country by providing an address, bank name, phone number, and an ID number issued by the South Korean government. There are consultancy agencies that have business relations to help single retail investors receive institutional contacts and, in that way, have a much higher daily cap on transfers. Retail investors could also tie themselves to an institution or form their own business relations to get around the demands from the Korean government. However, this strategy might be a much tougher process, and this is only one instance that could come in handy to get through the capital controls.

South Korea stands out in terms of capital controls, and out of the four examined regions, this is the only country that can be said to have capital controls. Therefore, capital control is likely to be one driver for the Kimchi premium since South Korea’s bitcoin market will be mainly driven by domestic supply and demand. In theory, investors would want to sell their bitcoin in South Korea since prices there tend to be higher than in the U.S. Thus, in theory, supply in South Korea is not driven by only domestic investors, and hence supply will be rather large compared to domestic demand. Nevertheless, different researchers and investors, such as the crypto journalist Larry Cermak, argue that supply does not match demand on the South Korean market, leading to bitcoin mispricing, and thus a Kimchi premium can rise. Since foreign investors cannot freely trade on South Korean exchanges, the increasing domestic demand within South Korea cannot be met by the existing domestic supply. It follows from the baseline of economic theory that when demand overstates supply, prices will increase. Furthermore, due to the extent of control on the flow of capital across the South Korean borders, the marginal investor in South Korea might benefit more from adopting cryptocurrencies since these do not meet the same strict controls. This is another example of how South Korea acts as an outlier as an entire market and how its capital controls significantly differentiate it from the three other researched markets.

5.1.5 Fees

Each transaction between bitcoin wallets needs to be confirmed by miners in the blockchain, and there is a cost associated with the process. The transaction cost is a flat fee that depends on the number of transactions and miners at a given time, and it is therefore volatile. The daily transaction fee ranges from 1.4 to 31.1 USD during our period. We do not view the transaction fee to prohibit the possibility to exercise arbitrage opportunities, as they are flat and, as concluded by Makarov and Schoar (2019), small relative to possible arbitrage.

The exchanges typically charge fees for market-making, deposits, and withdrawals. Active investors will be charged based on 30 days volume, with trading fees ranging from 0-0.5%. Deposits are often free except for credit card deposits, while the withdrawal fees range from 0-0.5%. It is also noteworthy that the more active and valuable investors will attain more lucrative deals, lowering their expenses. A significant investor is likely to face total fees of 0.5-0.75%. In conclusion, the fees are relatively low compared to possible arbitrage profits.

5.2 Findings drivers of price difference

5.2.1 Stock market development

The hypothesis behind the stock market development affecting bitcoin prices is that a bullish view on the equity market would also show higher bitcoin prices and vice versa. Thus, the general market view should affect bitcoin prices in the same direction, not least due to the increased correlation between the stock market and bitcoin. The regression results show no significant predictive value of the stock market development, neither for Europe on the European bitcoin prices nor within Japan on the Japanese bitcoin prices. This result is not necessarily surprising, given that the average year-over-year correlation between bitcoin and the stock market over the eight years 2012-2020 is 0.01. In theory, it is difficult to distinguish whether the domestic stock market is entirely uncorrelated with the domestic bitcoin price since we only examine cross-regional differences. Thus, it could very well be that only one of the two stock market indices included in the comparison is uncorrelated with the price difference in total. Although, given the low correlation between bitcoin and the stock market as well as the insignificant results for both Japan and Europe, it is not likely to be the case.

The insignificance might also be due to the highly correlated indices. The correlations between the Nikkei 225 and the S&P 500, as well as the European indices and the S&P 500, are put in clear contrast to the highly uncorrelated price differences between the same currency pairs. High degrees of co-movement between the stock market indices might fail to explain the uncorrelated prices on the same regions, meaning that a high correlation is difficult to use as an explanatory variable for a dependent variable built on low correlation.

The increase in correlation between the stock market and bitcoin during 2020 might depend on other extreme macroeconomic factors that faced the stock market in 2020, such as the coronavirus crash in early 2020 and the American election in the fall. Thus, it remains to establish whether the detected correlation depends on adverse swings in the stock market or in bitcoin. These factors might also have hit differently depending on the geographical region. The correlation of 0.229 relates to the S&P 500 index, and we find no data on bitcoin's correlation with other regional indices. Hence no generalization can be made. The increased correlation might also be due to the increased adoption of the instrument in general, with record-high volumes being traded across major bitcoin exchanges and the rise in OTC-traded bitcoin funds.

5.2.2 Attention

We find significant results for Europe regarding the SVI and its correlation with bitcoin prices on the different regional exchanges. Although the coefficient is small, 0.003, it is significant at the 95% significance level.

Europe has the largest observed attention for bitcoin during the period, and Japan the lowest. It is noteworthy that the coefficient for attention difference between Europe and the U.S. is statistically significant, and the difference in attention is large. Nevertheless, the random walk theory is proved to hold, which is in line with that only such a general term as bitcoin is used in the SVI. With the general term, every search including the term will be accounted for, and only a minority are likely to invest. Thus, attention can be cyclical and highly affected because the development of bitcoin prices had high similarities with a boom in late 2020 and early 2021. It might be a question of reverse causality, where the bitcoin prices rise, in an experienced boom, and ultimately draws the attention of the public. Given this reasoning, it is possible that increased attention during the last month of the dataset is a cause of the boom in bitcoin prices, rather than a driver of the bitcoin prices rising. Thus, it is also a possibility that the last month of the SVI includes more people without investing intent since the increased media coverage and public attention increases Google searches.

Given that the attention measure is not explicit for investors, these can be a minority of the total, meaning that their behaviour still becomes random when looking at the whole population's search patterns. The possibility also remains that the language discrepancies make the attention measure for Japan insignificant since this wording cannot. Further, there is also a possibility of different behavioural patterns surrounding the use of Google as a search engine.

Attention could also be one cause for the Kimchi premium, which is present during early 2021. As previously mentioned, we exclude South Korea from the regression because it is an isolated market. Nevertheless, we observe their SVI and see that their attention for "bitcoin" translated to

Korean is right below Europe on the weekly level and at times above Europe on the daily level. During late 2020 and early 2021, South Korea does, however, systematically have a higher SVI than Europe does, making them have the highest observed SVI for this subperiod. Thus, it becomes a question as to whether increased attention in South Korea affects the rise in prices or if the increasing prices themselves, caused by some other factor, cause a higher SVI.

5.2.3 Random walk

We find the random walk theory to prevail, and it is the best explanation given the choice of independent variables included. Thus, one never knows where the next step will be, and although concluded to be random with a probability of 50% for both directions, a discussion could be held regarding the behavioural characteristics of non-institutional investors. One plausible assumption is that non-institutional investors are more prone to continue investing in an instrument if the instrument has evolved in a positive direction compared to the previous day. This rationale would confirm the theory of crypto being bullish crypto. Furthermore, although insufficient results are found based on our dataset, a similar assumption could hold for the stock market in general. An investor's general bullish stock market outlook could affect the investor's probability function of investing in other instruments as well. These behavioural characteristics would still prove the random walk theory true, given that behaviour is highly individual.

The non-integrated nature of bitcoin markets might also explain why the random walk theory seems invincible in predicting price differences. We seek to explain and predict the differences in bitcoin price across international markets, and a unison explanatory variable is difficult to find since these markets are non-integrated and segmented. Thus, the resulting random walk is not surprising and confirms findings of prior research around the topic of within-regional bitcoin prices. Since within-regional markets can be assumed to be more integrated, our result might very well be due to segmented markets across geographical regions.

One reason for the random walk theory to seem invincible might also be the domestic views of each region. In a region with controls on capital flows to a great extent, cryptocurrencies might be an attractive alternative since the same capital controls do not exist for these kinds of currencies. Thus, this can be one explanation behind the, in general, higher prices on the South Korean exchanges. Furthermore, it might be a question of future lookouts and beliefs for bitcoin as a currency compared to an instrument. Within the U.S. and Europe, more large corporations are starting to adopt cryptocurrencies and are rumoured to begin accepting bitcoin as currency in exchange for their goods. Thus, the general market view that bitcoin is more of a currency than an instrument might be on the rise. Early adopters of bitcoin viewed it as a solely speculative investment, and it is still today a very volatile instrument compared to the stock market, why this view might persist among regional investors today. Ultimately, the view on bitcoin is different yet to this day across regions, which is an indicator that random walk should be the best explanation for the pricing of bitcoin since there are unobservable characteristics that might affect the pricing the most.

6. Conclusion

We show how arbitrage opportunities continue to exist between regional bitcoin markets, although to a lesser extent than in 2017-2018. These results are likely due to increased worldwide adoption of the instrument, together with using a dataset constituting more volume transparency. Using the U.S. as a base case, the largest price difference of 9.2% is found between the U.S. and South Korea, partially confirming the Kimchi premium phenomena. Nevertheless, large price differences persist across regions, and although some constraints of the arbitrary profits exist, any investor can, in theory, work their way around these and profit on the arbitrage opportunities. Interestingly, price differences persist for long periods, clearly showing the extent of market segmentation in bitcoin markets and making bitcoin markets unique compared to other instrument markets.

We find no significant results when regressing cross-regional price difference on investor attention and stock market development. The random walk theory proves to remain the best estimator of price differences, leading to drivers behind cross-regional price differences being challenging to explain. This result might be due to the difference in convenience yield across different regions. Nevertheless, our findings prove the volatility and unpredictive nature of the bitcoin markets. The fact that bitcoin markets, to a great extent, are non-integrated also supports that predictive models are difficult to build to explain the cross-regional price differences.

Acknowledging the limitations to our study, we begin with the smaller sample size. In contrast to Makarov and Schoar (2019) using 17 exchanges in their arbitrage index, we only use one exchange for each region. The smaller sample size is validated because the four selected exchanges display a higher degree of transparency, and the limited number of exchanges included may in itself lead to smaller arbitrage opportunities. For the second research question, one obvious limitation is the behavioural schemes around how people use Google, as well as the language discrepancies. It is possible that only looking at Google searches leads to a biased sample since the people who turn to Google might very well have similar reasons for doing so. Meanwhile, it is difficult to assess whether Google is used for the same purposes across all four covered regions.

For the stock market development, the main limitation is the comparativeness of the included indices for each region and the fact that some indices only include 30 companies, while others include 500 companies. This difference might affect the diversity characteristics of the index but is at large overcome by the fact that the indices are all created to be market proxies with similar techniques.

For following research, the interesting question is ultimately the pricing of bitcoin. Even though our research provides some explanations for what might drive prices in bitcoin markets, it is far from providing the whole picture, and we need to acknowledge that quite a lot of information is yet to be found in the residuals. We show how increased adoption of the instrument, together with the news-value settling and the markets at least starting to become integrated, have, as hypothesized, lessened the arbitrage opportunities during the past three years. Nonetheless, the bitcoin markets are far from being efficient. Although even further research on whether these opportunities will continue to prevail in the future will be highly interesting, especially due to the increased adoption of bitcoin by large corporations, the weak efficiency is a fact that continues to be confirmed. Thus, future research examining why prevailing price differences are not being exercised to a greater extent would be a highly interesting extension of our findings. Many fields within bitcoin are yet to be explored, and we regard the mentioned approaches to be a good starting point.

7. References

Alloway, T., & Weisenthal, J. (2021). (audio podcast episode). The Ex-Jane Street Trader Who's Building a Multi-Billion Crypto Empire, *Bloomberg*.

Athey, S., Parashkevov, I., Sarukkai, V., & Xia, J. (2016). Bitcoin pricing, adoption, and usage: Theory and evidence, unpublished.

Bariviera, A. (2017). The inefficiency of Bitcoin revisited: A dynamic approach, Economics Letters, 161, pp. 1-4. doi: 10.1016/j.econlet.2017.09.013.

Blockchain (2021). Total Bitcoins. (online) Available at: https://www.blockchain.com/charts/ (Accessed: 15 February 2021).

Borri, N. (2019). Conditional tail-risk in cryptocurrency markets, Journal Of Empirical Finance, 50, 1-19. doi: 10.1016/j.jempfin.2018.11.002

Bouri, E., Molnár, P., Azzi, G., Roubaud, D., & Hagfors, L. (2017). On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?, Finance Research Letters, 20, 192-198. doi: 10.1016/j.frl.2016.09.025

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation, Econometrica, Sep. 1979, Vol. 47, No. 5 pp. 1287-1294.

Ciaian, P., Rajcaniova, M., & Kancs, D. A. (2016). The economics of BitCoin price formation, *Applied Economics*, *48*(19), 1799-1815.

Chainanalysis (2020). The 2020 Geography of Cryptocurrency Report. (online) Available at https://go.chainalysis.com/2020-geography-of-crypto-report.html (Accessed: 12 March 2021).

Cheah, E. & Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin, Economics Letters, 130, pp. 32-36. doi: 10.1016/j.econlet.2015.02.029.

Choi, H. & Varian, H. (2009). Predicting Initial Claims for Unemployment Benefits, unpublished.

Coinmarketcap (2021). Top Cryptocurrency Spot Exchanges. (online) Available at https://coinmarketcap.com/rankings/exchanges/ (Accessed: 15 February 2021).

Corbet, S., Lucey, B. & Yarovaya, L. (2018) Datestamping the Bitcoin and Ethereum bubbles, Finance Research Letters, 26, pp. 81-88. doi: 10.1016/j.frl.2017.12.006.

Nadarajah, S. & Chu, J. (2017). On the inefficiency of Bitcoin, Economics Letters, 150, pp. 6-9. doi: 10.1016/j.econlet.2016.10.033.

Da, Z., Engelberg, J. & Gao, P. (2011). In Search of Attention, *The Journal of Finance*, 66(5), pp. 1461-1499. doi: 10.1111/j.1540-6261.2011.01679.x

DeLong, J.B., Shleifer, A., Summers, H., & Waldmann, R.J. (1990). Noise trader risk in financial markets, *Journal of Political Economy, 98,* 703-738.

Durbin, J. & ,Watson G.S. (1950) Testing for Serial Correlation in Least Squares Regression: I, *Biometrika*, Vol. 37, No. 3/4 (Dec., 1950), pp. 409-428.

Fama, E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work, *The Journal of Finance. Vol. 25, No. 2.*

Fernández, A., Klein, M., & Rebucci, A. (2016). Capital Control Measures: A New Dataset, *IMF Econ Rev 64***,** 548–574. *https://doi.org/10.1057/imfer.2016.11*

Froot, K. A., & Dabora, E. M. (1999). How are stock prices affected by the location of trade?. *Journal of financial economics*, *53*(2), 189-216.

Gandal, N., Hamrick, T., Moore, & Oberman, T. (2018). Price manipulation in the Bitcoin ecosystem, Journal of Monetary Economics, 95, pp. 86-96. doi: 10.1016/j.jmoneco.2017.12.004.

Gkillas, K., & Longin, F. (2018). Is Bitcoin the New Digital Gold? Evidence From Extreme Price Movements in Financial Markets, SSRN Electronic Journal. doi: 10.2139/ssrn.3245571

Gromb, D. & Vayanos, D. (2002). Equilibrium and welfare in markets with financially constrained arbitrageurs, Journal of Financial Economics, 66(2-3), pp. 361-407. doi: 10.1016/s0304-405x(02)00228-3.

Gromb, D. & Vayanos, D. (2018). The Dynamics of Financially Constrained Arbitrage", The Journal of Finance, 73(4), pp. 1713-1750. doi: 10.1111/jofi.12689.

Han, L., Lv, Q., & Yin, L. (2017). Can investor attention predict oil prices?, Energy Economics, 66, 547-558.

Hayes, A. (2017). Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. Telematics And Informatics, 34(7), 1308-1321. doi: 10.1016/j.tele.2016.05.005

Isard, P. (1977). How Far Can We Push the "Law of One Price"?, The American Economic

Review, 67(5), 942-948.

Kristoufek, L. (2015). What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. Plos one, 10(4), e0123923. doi: 10.1371/journal.pone.0123923

Maddala, G. S. (1992). Introduction to econometrics (Vol. 2). New York: Macmillan.

Makarov, I. & Schoar, A. (2019). Trading and arbitrage in cryptocurrency markets, *Journal of Financial Economics*, 135(2), pp. 293-319. doi: 10.1016/j.jfineco.2019.07.001.

Nakamoto, S., 2008. Bitcoin: a peer-to-peer electronic cash system. (Working paper) Available at: https://bitcoin.org/bitcoin.pdf (Accessed 10 February 2021).

Newey, W. K., & West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix (No. t0055). National Bureau of Economic Research.

Nomics (2021). Top Global Cryptoexchanges. (online) Available at https://nomics.com/exchanges (Accessed: 15 February 2021).

Rosenthal, L., & Young, C. (1990). The seemingly anomalous price behavior of Royal Dutch/Shell and Unilever NV/PLC, *Journal of Financial Economics, 26*(1), 123-141.

Shen, D., Urquhart, A., & Wang, P. (2019). Does twitter predict Bitcoin?. *Economics Letters, 174*, 118-122. doi: 10.1016/j.econlet.2018.11.007

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. The Journal of finance, 52(1), 35-55.

Shleifer, A. (2000) Inefficient Markets: An Introduction to Behavioural Finance. OUP Oxford.

Statcounter (2021). Search Engine Marketshare Worldwide. (online) Available at https://gs.statcounter.com/search-engine-market-share (Accessed: 10 March 2021).

Stock, J.H. (2015) Heteroskedasticity- and Autocorrelation-Robust Inference, Harvard University. (online) Available at https://scholar.harvard.edu/files/stock/files/aea_2015_lecture4_har_rev.pdf (Accessed 25 March 2021)

Urquhart, A. (2016). The inefficiency of Bitcoin, *Economics Letters*, 148, pp. 80-82. doi: 10.1016/j.econlet.2016.09.019.

Urquhart, A. (2018). What causes the attention of Bitcoin?, *Economics Letters*, 166, pp. 40-44. doi: 10.1016/j.econlet.2018.02.017.

Vaneck (2021). Bitcoin's Correlation to Markets Hits a Record in 2020. (online) Available at https://www.vaneck.com/us/en/blogs/digital-assets/bitcoins-correlation-to-markets-hits-a-record-in-2020/ (Accessed: 15 March 2021).

Appendix

A – Bitcoin Prices

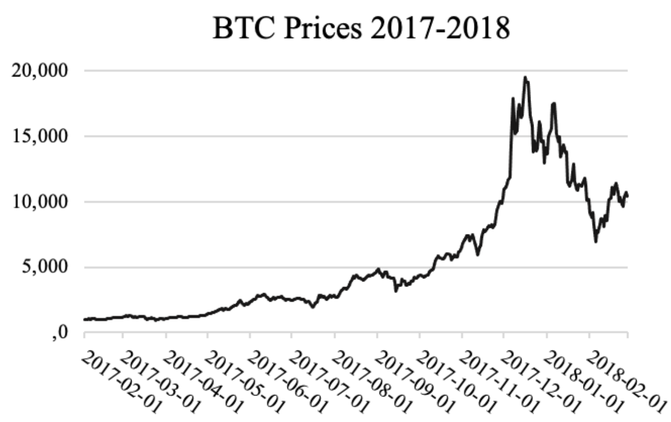

Figure A1: bitcoin price in the U.S.
The figure illustrates the price development for bitcoin prices in USD between February 2017 and February 18. The y-axis plots the price in USD and the x-axis plots the date.

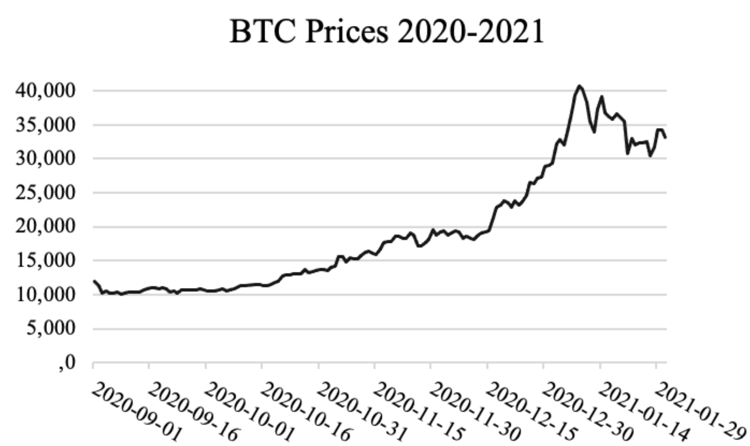

Figure A2: bitcoin price in the U.S.
The figure illustrates the price development for bitcoin prices in USD between September 2020 and January 2021. The y-axis plots the price in USD and the x-axis plots the date.

B - Attention for "Buy Bitcoin"

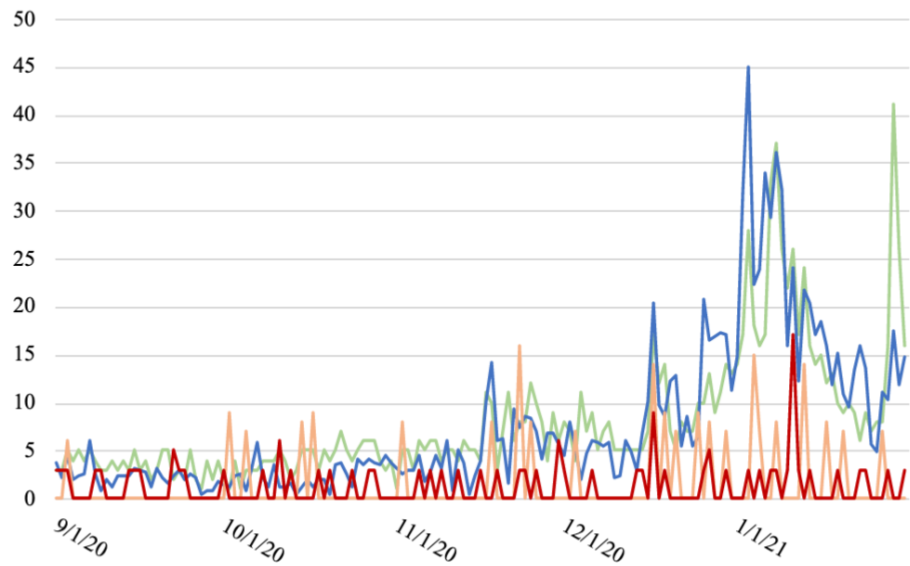

Figure B. Attention comparison "buy bitcoin" with daily data.

This figure shows aggregated, compared daily Google searches for the term "buy bitcoin" in the U.S., Europe, South Korea, and Japan. Language adjustments have been made for Japan and South Korea. The y-axis plots the attention index from Google Trends, and the x-axis plots the date, covering September 1, 2020, to January 31, 2021.

C - Complete Calculations

Complete derivations of models and equations used in the regression analysis, second research question.

The OLS Model:

$Y_n = \sum_{i=0}^{k} \beta_i X_{ni} + \mathcal{E}_n$, becomes:

$y_i = \alpha + \beta \times x_i + \mathcal{E}_i$, becomes:

$Y_n = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mathcal{E}_i$, and our fitted models:

$Price\_diff_{Japan/US} = b_0 + b_1 \times Stocks_{Japan/US} + b_2 \times Attention_{Japan/US} + \mathcal{E}_i$

$Price\_diff_{Europe/US} = b_0 + b_1 \times Stocks_{Europe/US} + b_2 \times Attention_{Europe/US} + \mathcal{E}_i$

Where Epsilon, $\mathcal{E}$, represent the error terms, described as the random component of the (linear) relationship between the independents and the dependent.

To find the intercept and slope coefficient, a minimization problem needs to be solved:

$\min\limits_{\beta_0, \beta_1} \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$, taking the first derivative equals:

$\frac{\partial W}{\partial \hat{\beta}_0} = \sum_{i=1}^{N} -2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$, and:

$\frac{\partial W}{\partial \hat{\beta}_1} = \sum_{i=1}^{N} -2x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$, solving so that:

$N\hat{\beta}_0 = N\bar{y} - N\hat{\beta}_1\bar{x}$, dividing by N we get:

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$

Solving for $\beta_1$by substituting in the equation above:

$\sum_{i=1}^{N} x_i y_i - \bar{y}\sum_{i=1}^{N} x_i + \hat{\beta}_1\bar{x}\sum_{i=1}^{N} x_i - \hat{\beta}_1\sum_{i=1}^{N} x_i^2$, and:

$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}}{\sum_{i=1}^{N} x_i^2 - N\bar{x}^2}$, ultimately leading to:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

The Newey-West Estimator:

The long-run variance of $X_t u_t$ in the OLS, when $Z_t = X_t u_t$, $EZ_t = 0$ and Z_t is second order stationary, is:

$\Omega_t = var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\right) = E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\right)^2$which yields:

$\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E(Z_t\, Z_s{}') = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\Gamma_{t-s}$, which yields:

$\frac{1}{T}\sum_{j=-(T-1)}^{T-1}(T-|j|)\Gamma_{t-s} = \sum_{j=-(T-1)}^{T-1}\left(1-\left|\frac{j}{T}\right|\right)\Gamma_j \rightarrow \sum_{j=-\infty}^{\infty}\Gamma_j$, so:

$$\Omega = \sum_{j=-\infty}^{\infty}\Gamma_j = 2\pi S_Z(0)$$

The Newey-West estimator is:

$\hat{\Omega}^{NW} = \sum_{j=-m}^{m}\left(1-\left|\frac{j}{m}\right|\right)\hat{\Gamma}_j$, where:

$\hat{\Gamma}_j = \frac{1}{T}\sum_{t=1}^{T}\hat{Z}_t\hat{Z}'_{t-j}$, where $\hat{Z}_t = X_t\hat{u}_t$

D – Variable Explanations

attention_Europe_US: A comparative variable examining the difference between the SVI for Europe with the SVI for the U.S.

attention_Japan_US: A comparative variable examining the difference between the SVI for Japan with the SVI for the U.S.

price_Europe_US: A comparative variable examining the difference between the VWAP for Europe with the VWAP for the U.S.

price_Japan_US: A comparative variable examining the difference between the VWAP for Japan with the VWAP for the U.S.

stocks_Europe_US: A comparative variable examining the difference between stocks_indexed_Europe and stocks_indexed_US

stocks_Japan_US: A comparative variable examining the difference between stocks_indexed_Japan and stocks_indexed_US