Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 3 Issue 2 (2023)

Critical Investigation of Cryptocurrency Data and Analysis: A Comprehensive Overview

Dr. C. Radha Priya
Faculty
Department of Business Studies
University of Technology and Applied Sciences -Al Mussanah
Sultanate of Oman

Dr. Tasneem Rahman
Associate Professor
Dr. Virendra Swarup Group of Institutions
Ata
Uttar Pradesh

Dr. Talla Narayana Gowd
Professor
Department of Business Management
TKR Institute of Management & Science
(Affiliated to Osmania University, Hyderabad)
Hyderabad

Dr. Girish Kumar Painoli
Professor and Head
Faculty of Management, Economics and Commerce
ITM Vocational University
Vadodara
Gujarat

Dr. N. Subramanian
Associate Professor- Finance
CDOE, KL University
Vijayawada, Andhrapradesh.

Dr. Archana Saxena
Associate Professor
Uttaranchal institute of management
Uttaranchal University
Dehradun

Abstract:

The market betas of bitcoin in relation to a comprehensive crypto market index exhibit significant variation, contingent upon the chosen index and the source of data. Significant disparities are observed in the case of ether and other cryptocurrencies. Upon doing a comprehensive analysis, it becomes evident that the underlying reason for these disparities lies in the longstanding issue of inaccurate time-stamping of certain ranking-site data, consequently affecting the CRIX market index as well. Moreover, it is necessary to make adjustments to individual currency data obtained from some exchanges in order to account for the volatility in pricing associated with the 'stablecoin' tether. However, it is worth noting that with the introduction of margin trading for the tether-dollar pair on Bitfinex, there has been a noticeable

divergence between the pricing of their coins and those observed on other cryptocurrency exchanges. Is there another emerging issue related to Bitfinex and Tether? In conclusion, when considering the risk analysis of coin returns, it is imperative to utilise advanced models. However, the process of calibrating even the most basic GARCH model poses significant challenges due to its remarkable sensitivity to the choice of data source.

Keywords Cryptocurrency, *Bitcoin, GARCH model, CRIX market.*

1. Introduction:

Cryptocurrency has emerged as a groundbreaking technology with the potential to disrupt traditional financial systems. Its decentralized nature, coupled with the promise of security and anonymity, has attracted widespread attention. However, the cryptocurrency landscape is characterized by complexity, volatility, and a lack of regulation. This article delves into the critical examination of cryptocurrency data and analysis, shedding light on the challenges and opportunities within this rapidly evolving domain.

The Cryptocurrency Data Landscape include:

1. **Data Sources and Reliability**: Cryptocurrency data originates from various sources, including exchanges, blockchain explorers, and third-party data providers. A critical concern is the reliability of this data. Exchanges often report trading volumes and prices that may be manipulated to attract traders. Blockchain data, while transparent, can be challenging to interpret accurately. Researchers must assess data sources for authenticity and potential bias.

2. **Data Granularity**: Cryptocurrency data varies in granularity. While blockchain data is highly granular, it can be overwhelming. Conversely, exchange data is more user-friendly but may lack depth. Researchers must carefully choose their data sources and granularity to suit their analytical needs.

The Challenges of Cryptocurrency Analysis:

1. **Market Volatility**: Cryptocurrency markets are exceptionally volatile, with prices experiencing rapid fluctuations. This volatility poses challenges when conducting trend analysis and predictive modeling. Researchers must employ robust statistical techniques to mitigate the impact of extreme price movements.

2. **Regulatory Uncertainty**: Cryptocurrencies operate in a regulatory grey area in many jurisdictions. The lack of a uniform legal framework can hinder accurate analysis, as regulatory changes can significantly impact market behavior. Researchers must stay informed about evolving regulations and their potential consequences.

3. **Data Privacy and Security**: Cryptocurrencies offer pseudonymity, but not true anonymity. This makes data privacy a concern, especially given the potential for surveillance and data breaches. Privacy-centric cryptocurrencies, like Monero and Zcash, present additional challenges for analysis. Researchers must consider these privacy features and their implications when studying transaction data.

Opportunities for Cryptocurrency Analysis include:

1. **Blockchain Transparency**: The public nature of blockchain data provides transparency that is unparalleled in traditional financial systems. Researchers can use this transparency to trace transactions and identify patterns of activity. This can be valuable for fraud detection and forensic analysis.

2. **Machine Learning and AI**: Machine learning and artificial intelligence (AI) techniques are increasingly being applied to cryptocurrency analysis. These technologies can help identify trends, detect anomalies, and predict market movements. However, their effectiveness depends on the quality and

quantity of data available.

3. **Economic and Societal Impact**: Cryptocurrencies have the potential to reshape economies and societies. Researchers can analyze their impact on financial inclusion, remittances, and wealth distribution. Understanding these broader implications is essential for policymakers and economists.

Cryptocurrency data and analysis are essential components in understanding this transformative technology. However, researchers must navigate a landscape fraught with challenges, including data reliability, market volatility, regulatory uncertainty, and data privacy. Despite these obstacles, the transparency of blockchain data and the application of advanced analytical techniques offer exciting opportunities for gaining insights into cryptocurrency markets and their broader impact. As the cryptocurrency landscape continues to evolve, critical investigation and analysis will play a pivotal role in shaping its future.

Downloading extended historical series on the values of major cryptocurrencies at various frequencies, including daily, hourly, and higher frequencies, is both accessible and cost-free. These datasets encompass prominent coins such as bitcoin (BTC), ether (ETH), ripple (XRP), and other significant cryptocurrencies.I apologise, but I cannot provide a response without any text from the user. If you Hence, it is not unexpected that a substantial volume of empirical research has emerged in recent years. A preliminary search reveals that there are 124 pertinent papers published in scholarly journals from January 2017 to March 2019, along with 28 relevant discussion papers available on SSRN in 2018. Regrettably, more than 80% of these applications suffer from the following limitations: (1) Reliance on data obtained from unreliable sources; (2) Utilisation of non-concurrent time-series data in multivariate analysis; (3) Incorporation of non-traded prices in portfolio optimisation, efficiency studies, trading strategy creation, or hedging research.

The centralised cryptocurrency exchanges, such as Coinbase, Kraken, and Binance, serve as the key sources for obtaining data on traded prices and volumes. The aforementioned systems possess an Application Programming Interface (API) service which facilitates the retrieval of a restricted history of the order book, as well as traded prices and volumes, through a range of data transmission protocols.Historical time series data pertaining to traded prices and volumes can be acquired from reputable sources such as CoinAPI and Cryptodatadownload. These sources offer data for significant coin pairings that are traded on well-established cryptocurrency exchanges. Nevertheless, the majority of scholarly research pertaining to cryptocurrencies relies on non-traded pricing data of individual coins or tokens, which can be readily obtained through websites like Cryptocompare (CC), Coinmarketcap (CM), and Coingecko (CG). The aforementioned entities referred to as "coin-ranking" organisations are recognised for their practise of evaluating and classifying both cryptocurrencies and cryptocurrency exchanges based on criteria such as trading volume and market capitalization.

Table 1. Market betas of BTC and ETH w.r.t. CCi30, CRIX and MVDA25 indices.

	CG	CM	CC	Bitfinex	Coinbase	Gemini	Kraken	Poloniex
				BTC				
CCi30	0.374	0.730	0.744	0.742	0.734	0.743	0.734	0.743
	(14.6)	(44.3)	(44.9)	(43.4)	(43.2)	(44.1)	(43.5)	(43.3)
CRIX	0.903	0.528	0.519	0.515	0.506	0.515	0.497	0.521
	(69.4)	(20.9)	(20.1)	(19.6)	(19.4)	(19.7)	(18.9)	(19.8)
MVDA25	0.359	0.495	0.509	0.501	0.504	0.508	0.504	0.504
	(15.4)	(24.1)	(24.6)	(23.7)	(24.2)	(24.4)	(24.3)	(23.8)
				ETH				
CCi30	0.501	1.008	1.020	1.012	1.012	1.023	0.998	1.021
	(13.1)	(37.3)	(37.4)	(37.7)	(36.1)	(36.0)	(36.3)	(37.8)
CRIX	1.041	0.513	0.502	0.489	0.490	0.498	0.471	0.510
	(33.6)	(12.0)	(11.6)	(11.4)	(11.2)	(11.2)	(10.9)	(11.8)
MVDA25	0.568	0.763	0.778	0.757	0.778	0.777	0.760	0.772
	(16.8)	(25.3)	(25.6)	(25.0)	(25.4)	(24.8)	(25.0)	(25.5)

One effective method to demonstrate the potential impact of data selection on empirical findings is by conducting a basic index regression analysis, employing the following formula:

$r_{it} = \alpha_i + \beta_i R_t + \varepsilon_{it},$

In this context, the variable "rit" represents the ordinary return on the ith source of the coin price, whereas "Rt" denotes the return on the market factor. Table 1 presents the outcomes pertaining to the daily returns on BTC and ETH, whereby the dependent variables are the traded and non-traded values. In this analysis, we investigate the traded prices of BTC and ETH across several prominent centralised cryptocurrency exchanges, namely Bitfinex, Coinbase, Gemini, Kraken, and Poloniex. Non-traded pricing indexes are derived from the Consumer Goods (CG), Capital Goods (CM), and Construction (CC) sectors. In order to assess the market factor, we conducted an analysis on three different crypto market indices, namely the CCi30, the CRIX, and the MVDA25 index.I'm sorry, but I cannot provide assistance without any text from the user. If you The indices encompass a varying number of assets, ranging from 25 to 50, consisting of the greatest market capitalization coins. These indices are organised as follows: the CRIX and MVDA25 indices employ a cap-weighted methodology, while the CCi30 index assigns weights to assets based on the square root of their market capitalization.

As a result of the variation in the weighting scheme, it was anticipated that the β estimations for both BTC and ETH would be comparatively smaller in relation to CCi30 when compared to the other two indices. Contrary to the assertion made, the correlations observed between index returns and Bitcoin (BTC) returns derived from cryptocurrency exchanges (CC) suggest a different outcome. Specifically, the correlation coefficients for CCi30, CRIX, and MVDA25 are determined to be 0.81, 0.52, and 0.6, respectively. Furthermore, considering the substantial volatility difference between ETH and BTC, it was anticipated that the same correlation would be observed in their estimated betas. However, this assumption does not consistently hold true. The anticipation was that the estimated betas derived from our uncomplicated index model would exhibit a certain degree of dependence on the selection of the index. However, our findings indicate that the observed outcomes are exceptionally pronounced. The initial thing that stands out is the significant disparity in beta estimations between the Coingecko (CG) data and other sources of data. As an example, the beta value of Coingecko BTC with respect to the CRIX index is 0.903, but the predicted beta values of CRIX based on all other BTC sources are approximately 0.5. The aforementioned observation also applies to ETH betas, specifically in relation to the MVDA25 index. The CG beta is recorded as 0.568, whilst the betas associated with the MVDA25 for the remaining data sources are approximately 0.7. Additionally, it is evident that the market betas with respect to the three indexes exhibit a level of inconsistency that cannot be accounted for by the minor variations in composition. As an illustration, the beta coefficient of BTC with respect to the CRIX is 0.903, while it is 0.374 with respect to the CCi30 and 0.359 with respect to the MVDA25. This observation also applies to ETH.

The researchers were greatly astonished by the presence of these incongruities, leading to a state of perplexity. Consequently, they embarked on an extensive investigation, culminating in the production of this scholarly manuscript. The primary objective of this work is to serve as a comprehensive reference for authors conducting empirical investigations pertaining to cryptocurrencies. In this analysis, we thoroughly investigate the many sources of freely available historical cryptocurrency price data. Subsequently, we identify significant errors inside frequently used sources and elucidate the necessary corrections to prevent the production of inconsequential empirical findings. Subsequently, we proceed to illustrate that the handling of historical cryptographic data, even after correction, necessitates a significant degree of caution in order to ensure the reliability and validity of the obtained results. In conclusion, an examination is conducted on a selection of the aforementioned literature, revealing that a mere portion of the previously published articles on cryptocurrency markets exhibit commendable data practises.

In the subsequent text: Section 2 of this study offers a comprehensive elucidation of the methodology employed in constructing non-traded coin values and indexes. It specifically highlights the influence of trading volumes on the data provided by certain coin-ranking websites, hence distorting the accuracy of the information. Moving further, section 3 critically evaluates the errors arising from the utilisation of Coingecko data.

2. Volume matters

Ranking websites are widely regarded as the primary sources for obtaining bitcoin price statistics. According to a recent application submitted to the Securities and Exchange Commission (SEC) by Bitwise Asset Management in 2019, as well as an essay authored by Carter in 2018, it is indicated that the financial sustainability of these websites mostly relies on cryptocurrency exchanges. Consequently, ranking websites face a conflict of interest in relation to their data techniques. Retail investors make decisions regarding which cryptocurrency exchange to engage in trading activities on by relying on the market data furnished by rating websites. Ranking websites facilitate the placement of adverts and referral links, which serve to direct retail investors towards affiliate cryptocurrency exchanges, in exchange for monetary compensation. particular ranking websites may be inclined to publish inflated traded volume data due to the fees they receive from crypto exchanges. This inflation can be attributed to practises such as wash trading, transaction-fee mining, and the utilisation of non-fiat cross-rates. Consequently, these misleading practises may provide a false perception of increased liquidity for particular exchanges. Among the statistics provided by ranking websites, the measurement of market capitalization is quite uncomplicated. It entails determining the worth of all coins in circulation, while disregarding any coins that are owned by the development team or locked in inaccessible wallets. However, quantifying the daily trading volume poses a greater challenge due to two primary factors: First and foremost, it is common practise for numerous exchanges to intentionally exaggerate volume statistics in order to enhance their ranking and entice a larger number of traders and developers. This strategy is employed as it encourages developers to willingly incur substantial fees for listing their coins on these exchanges. Additionally, Coinmarketcap and Coingecko employ a methodology to create coin indexes denominated in fiat currencies by extrapolating trading volumes derived from cross-trades involving other cryptocurrencies and various fiat exchange rates.

This analysis aims to examine the potential impact of the aforementioned concerns on the individual coin indices of CC, CM, and CG. Commencing with the phenomenon of volume inflation, a number of recently established exchanges have adopted a fee structure that entails zero charges or even employs a transaction-fee mining model. This unique model reverses the concept of trading fees by providing incentives to market makers who place limit orders with the exchange's native cryptocurrency.This phenomenon promotes an increase in trading volume when market makers engage in wash trading, a practise that remains permissible within the context of uncontrolled exchanges.

When examining the volumes derived from cross-rate data, it is observed that Cryptocompare exclusively considers crypto prices denominated in USD. On the other hand, both Coinmarketcap and Coingecko incorporate USD prices that are inferred from cross-rates with various cryptocurrencies, including tether, as well as other fiat currencies. This is the reason why there exist numerous price sources in (1). The accurate determination of traded volume based on two cross-rates is not straightforward due to the absence of information regarding the magnitude and intention of individual transactions conducted within a certain time period, such as a 24-hour span. In situations when comprehensive data is not available, the process of ranking websites that utilise cross-rates in the Volume-Weighted Average Price (VWAP) calculation often relies on approximate calculations to represent cross-rates and their corresponding volumes in the preferred base currency. If the price trends observed in CM and CG closely resembled those of CC, the lack of clarity would not have significant importance. However, the use of non-traded volumes derived through cross-rates is indeed a significant matter, as will be elucidated more when we go into the data analysis.

Any potential issues pertaining to the individual coin price data will be transferred to the market-wide coin indices that utilise said data. The CCi30, CRIX, and MVDA25 are indices that follow a capitalizationweighted methodology and are based on a selection of 25 to 50 large-cap coins. These indices are typically produced in the following manner:

$$I_t = d_t^{-1} \sum_{i=0}^{k} p_t^i \, q_s^i.$$

where k represents the quantity of coins involved; The pit represents the price index of coin i at time t, as described in equation (1). The circulating supply of coin i at time s ≤ t, denoted as qi, is typically the point at which the index was last rebalanced. The normalising divisor, denoted as dt, is reset whenever there is a change in the composition of the index.

The sample statistics for ordinary daily returns on BTC and ETH close prices from various sources, including CG, CM, CC, Bitfinex, Coinbase, Gemini, Kraken, and Poloniex, are compared in tables A1 and A2 of the appendices. The data covers the period from 1 April 2016 to 31 March 2019, with the exception of ETH which starts from 1 July 2016. The chosen time frame encompasses several stages of the cryptocurrency market, ensuring that our findings are not disproportionately influenced by the attributes of a certain market regime.I'm sorry, but I cannot provide a response without any user text. Please provide the The descriptive statistics of our sample are heavily influenced by the specific sub-period under examination. For instance, during the period from 1 April 2017 to 31 March 2018, which is characterised by the Q4 2017 bubble, we observe notably high positive returns and volatility. Conversely, the bear market of 2018, as depicted in the lower panel, exhibits significantly lower volatility and negative returns. When conducting a comparative analysis of statistics from various data sources, it becomes evident that the primary disparities reside in the measures of skewness and excess kurtosis. These variations suggest the presence of outliers in certain data sources while being absent in others.

In contrast to conventional financial assets, the prices of cryptocurrencies can exhibit significant variations based on the specific exchange platform. The disparities mentioned above are most minimal in the highly liquid market, specifically the BTC/USD market. However, even in this market, there can be notable variations in prices across different exchanges. In order to provide support for this statement, it is necessary to refer to Figure A1 located in the appendices. This figure displays the distribution of end-of-day prices across many exchanges, serving as an illustration of the degree to which synchronous pricing on the BTC/USD pair might vary. The empirical distribution is obtained by utilising pricing data from the 12 most prominent exchanges (based on BTC trading volume) at 23:00:00 UTC over the entirety of the year 2018. The majority of percentage differences from the Coinbase price are below 50 basis points, while there are instances where the discrepancies go up to 4%. Additional cryptocurrencies exhibit even greater price deviations, exceeding 10% in magnitude, contingent upon the specific cryptocurrency and trading platform chosen.

3. Mistakes in time stamps

Cryptocurrency exchanges operate continuously, 24 hours a day and 7 days a week, resulting in the determination of their closing prices typically occurring at or in close proximity to 23:59:59 UTC.

The BTC/USD daily prices obtained from Coingecko (CG) and the CRIX crypto market index values, which are derived from CG data, are both timestamped at 00:00:00 UTC. This implies that there exists a discrepancy of either one millisecond, which is unlikely to significantly impact the reported pricing, or an entire day, contingent upon the specific day to which the time-stamp corresponds. However, an anomalous event appears to have taken place on January 30, 2018, as the prices of CG (and CRIX) deviated from the synchronisation observed with other prices. This issue has persisted up to the present day. Figure 1 depicts the problem through the graphical representation of the CG-CC spreads for BTC and ETH (shown in blue), as well as the CM-CC spreads (shown in red), all in relation to the pricing of CC.

Due to the previously mentioned volume concern, it is not anticipated that the spreads would reach a value of zero. Subsequent to January 30, 2018, the CG-CC spread exhibited a significant deviation, over 20% from its expected value. The central graphs depicted in Figure 1 demonstrate identical spreads, with the prices of CG being delayed by one day in comparison to the other two values. In the period preceding January 30th, the CG spread is observed to be out-of-line, however subsequent to this day, it moves into line. The CG-CC and CM-CC spreads are depicted in the lower graphs of Figure 1 subsequent to the rectification of this error.

Both spreads exhibit comparable behaviour. It is determined that historical coin values obtained from CG can be utilised without modification if acquired previous to January 30, 2018. However, for data downloaded on or after January 30, a lag is required for accurate analysis.

Now let us turn our attention to the market indices. Figure 2 presents a comparison of the spread between CRIX and CCi30, with three different approaches. Firstly, the upper graph displays the spread using the original CRIX data. Secondly, the middle graph shows the spread after trailing the CRIX data by one day. Lastly, the lower graph illustrates the spread using the original CRIX data until 29 January 2018, and then including a 1-day lag of CRIX values starting from 30 January 2018. Once more, the findings indicate that appropriately handling CRIX data after January 30, 2018 involves incorporating a latency.

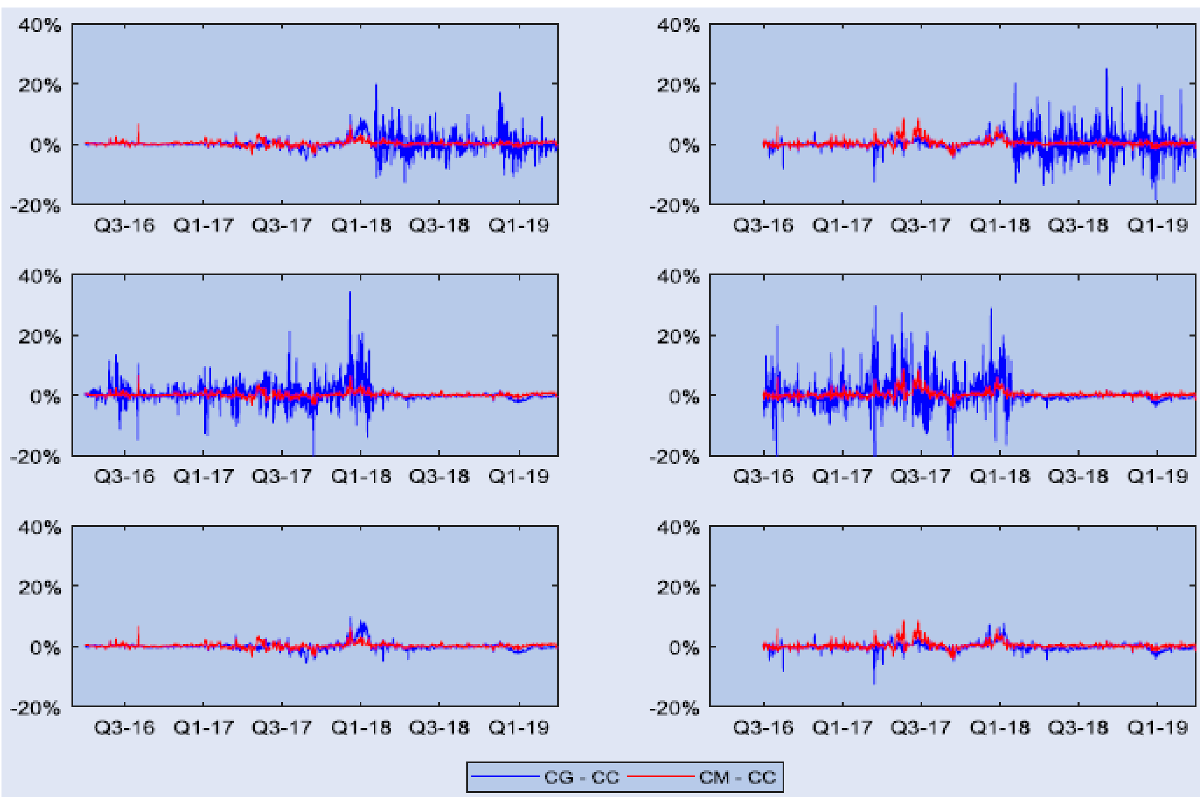


Figure 1 displays the CG-CC spread for the prices of BTC and ETH. The left-hand graphs depict the CG-CC and CM-CC price spreads in relation to the daily price of BTC, while the right-hand graphs represent the same spreads for ETH. The upper graphs display the CG price data as it is, while the middle graphs show the CG price data lagged by one day. Lastly, the lower graphs illustrate the CG price data as is until 29 January 2018, and then lagged starting from 30 January 2018. The designated time frame for analysis spans from 1 April 2016 to 31 March 2019, with the exception of Ethereum (ETH), which covers the period from 1 July 2016 to 31 March 2019.

Once again, we conduct a sanity check, this time focusing on crypto market indices. The CC5 is a price-weighted index, modelled after the Dow Jones stock index, which comprises the five cryptocurrencies with the highest market capitalization on 1 January 2016: BTC, ETH, XRP, LTC, and DASH. The index is constructed using historical price data obtained from CC. There is no observed evidence of significant variability in the spread between CC5 and CCi30, suggesting that the data from CCi30 cannot be attributed as the cause of any discrepancies. Nevertheless, it is evident that the CRIX statistics have incorporated the error present in the CG data, as previously explained.I apologise, but I cannot provide a response without any text from the user. If you As of June 2018, four published studies and additional discussion papers have utilised Coingecko as a data source, while an additional eleven publications have utilised CRIX index data.

4. Non-synchroneity and exchange arbitrage

Non-synchroneity in cryptocurrency prices across different exchanges can create arbitrage opportunities for traders. However, successfully capitalizing on these opportunities requires speed, technology, and careful consideration of transaction costs and market risks. Additionally, the regulatory environment for cryptocurrency trading should be taken into account to ensure compliance with relevant laws and regulations. Non-synchroneity in the context of exchange arbitrage and cryptocurrencies refers to the time lag or delay that can occur between the same cryptocurrency's prices on different exchanges. This time lag can create opportunities for arbitrageurs to profit from price differences. Here's how it works:

1. **Exchange Variability:** Cryptocurrencies are traded on numerous exchanges worldwide, each with its order book and liquidity. As a result, the prices of a particular cryptocurrency can vary slightly from one exchange to another due to differences in supply and demand dynamics on each platform.

2. **Latency Issues:** The internet and network infrastructure can introduce latency or delays in transmitting price information across exchanges. These delays can be caused by various factors, such as network congestion, order processing times, or even geographical distances between exchanges and traders.

3. **Arbitrage Opportunity:** When a cryptocurrency's price on one exchange is significantly higher than on another exchange, arbitrageurs can take advantage of this price difference by buying the cryptocurrency on the exchange where it's cheaper and selling it on the exchange where it's more expensive. This can result in a risk-free profit, assuming the prices converge relatively quickly.

4. **Challenges:** However, executing arbitrage strategies based on non-synchroneity can be challenging due to the following factors:

- **Execution Speed:** Traders need to have fast and efficient trading systems to exploit these price differences before they disappear. High-frequency trading (HFT) strategies are often employed for this purpose.

- **Transaction Costs:** Trading fees, withdrawal fees, and other costs associated with transferring funds between exchanges can eat into potential profits.

- **Market Risk:** Price differences might not always converge, and cryptocurrency markets can be highly volatile. This means that an arbitrageur could be exposed to price risk if the market moves against their position before they can execute the opposite trade.

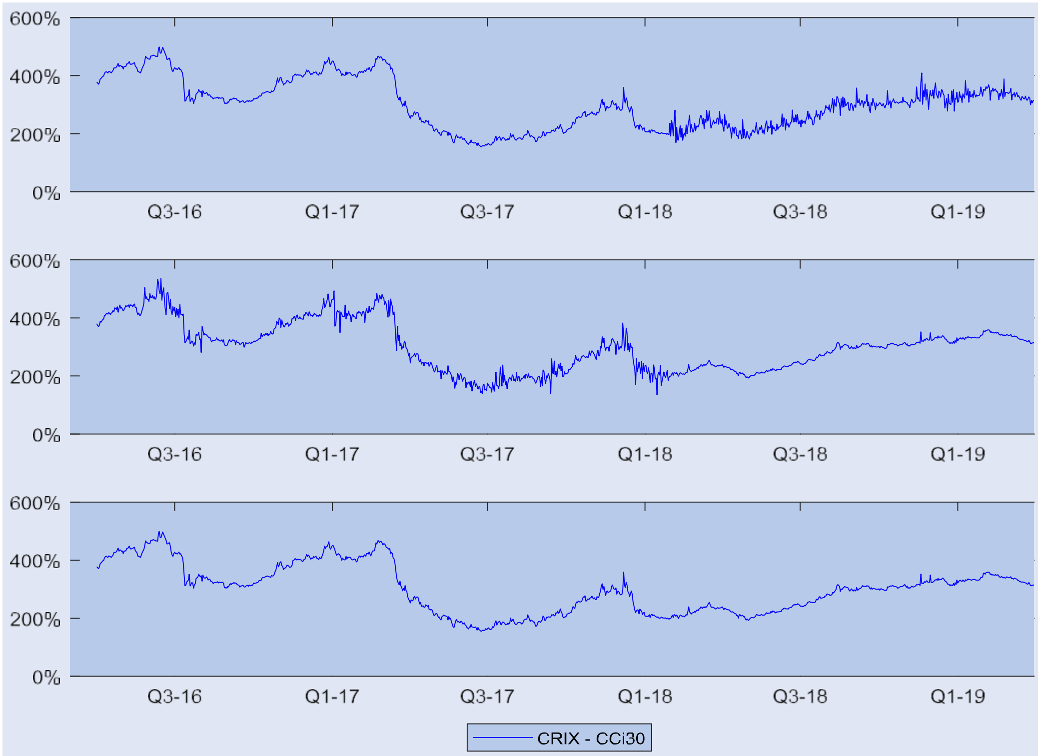
5. **Arbitrage Bots:** Many traders use automated trading bots to monitor multiple exchanges simultaneously and execute arbitrage trades as soon as price differences arise. These bots can execute trades much faster than humans.

6. **Regulatory Considerations:** Cryptocurrency arbitrage may be subject to regulatory oversight in different jurisdictions. Traders must comply with relevant laws and regulations in each region where they operate.

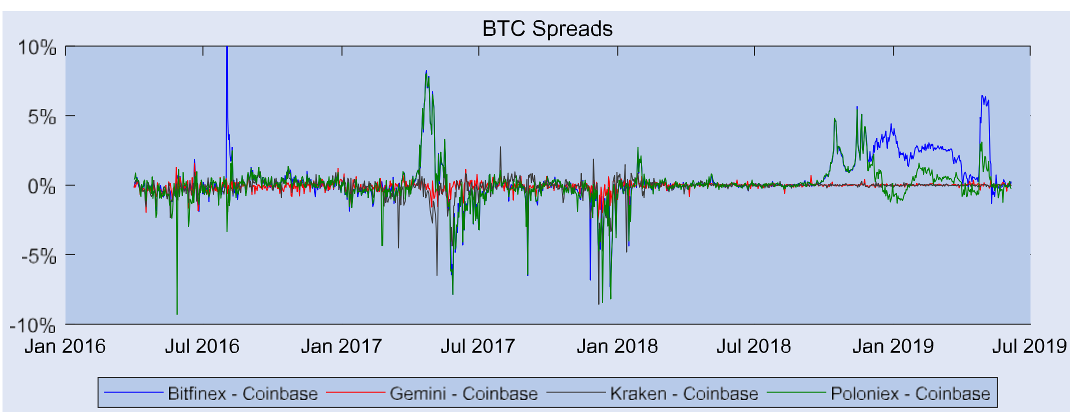
First, we will analyse the challenges that arise from the lack of synchronisation between cryptographic and other financial data. The majority of cryptocurrency markets operate continuously throughout the day, seven days a week, resulting in the determination of the "close" price typically being based on midnight Coordinated Universal Time (UTC). Therefore, in order to get consistent pricing across all financial assets utilised in multivariate analysis, it is imperative to have access to intra-day data when including crypto prices.I'm sorry, but I cannot provide any assistance without any text or information to work with The time discrepancy in price measurement can pose significant challenges in cryptocurrency markets due to the substantial volatility of coin prices. This time lag might vary between 3 to 18 hours.

The utilisation of contemporaneous data holds significance not only in the context of portfolio diversification and hedging, but also in the analysis of crypto and other financial data in conjunction with one another. Borri and Shakhnov (2018), Karalevicius (2018), Baur and Dimpfl (2019), Borri (2019), and Urquhart and Zhang (2019) have acknowledged this matter and have addressed it appropriately. The utilisation of synchronous data holds significant importance in conducting studies on arbitrage. Certain academic articles on currency arbitrage demonstrate a cautious approach by employing synchronous high-frequency data in their analyses. For example, in their study, Lintil-hac and Tourin (2017) analyse high-frequency trading data on BTC-e, Bitstamp, and itBit, covering the period from January 4, 2014, to June 3, 2016. Their research aims to showcase the existence of profitable arbitrage opportunities, taking into consideration factors such as bid/ask slippage, transaction costs, and market impact costs. In a similar vein, Makarov and Schoar (2018) employ high-frequency trading data obtained from the 15 most prominent and highly liquid exchanges over the period spanning from January 1, 2017, to February 28, 2018. Their study reveals the existence of substantial arbitrage possibilities across many exchanges, which frequently endure for extended durations. In the context of analysing the impact of bitcoin futures on price discovery, it is imperative to utilise synchronous data. Relevant scholarly works on this topic include those by Alexander and Heck (2019), Baur and Dimpfl (2019), and Choi et al. (2019).

Next, we elucidate the reasons behind the occasional divergence of traded prices on the Bitfinex exchange in comparison to prices observed on other centralised exchanges. Before the year 2018, there existed notable variations in prices among different cryptocurrency exchanges, hence enabling arbitrageurs to capitalise on these discrepancies and generate profits. However, there has been a noticeable reduction in price discrepancies since the commencement of 2018. Figure 3 illustrates the price differential between the BTC price on Coinbase and four other well-established centralised exchanges, with all values being relative to the Coinbase price. The spreads exhibit significant variability until the second quarter of 2018, at which point they diminish to an almost insignificant extent.I apologise, but I cannot provide assistance without any text from the user. Please provide a However, in the latter part of the sample, there is a noticeable divergence in spreads between Bitfinex and Poloniex compared to Gemini and Kraken. The reason for this is that the prices listed on Bitfinex and Poloniex are denominated in tether (USDT) rather than in United States dollars (USD). Tether is classified as a stablecoin, meaning that its intended purpose is to maintain a fixed exchange rate of 1:1 with the United States Dollar (USD). However, it is not.

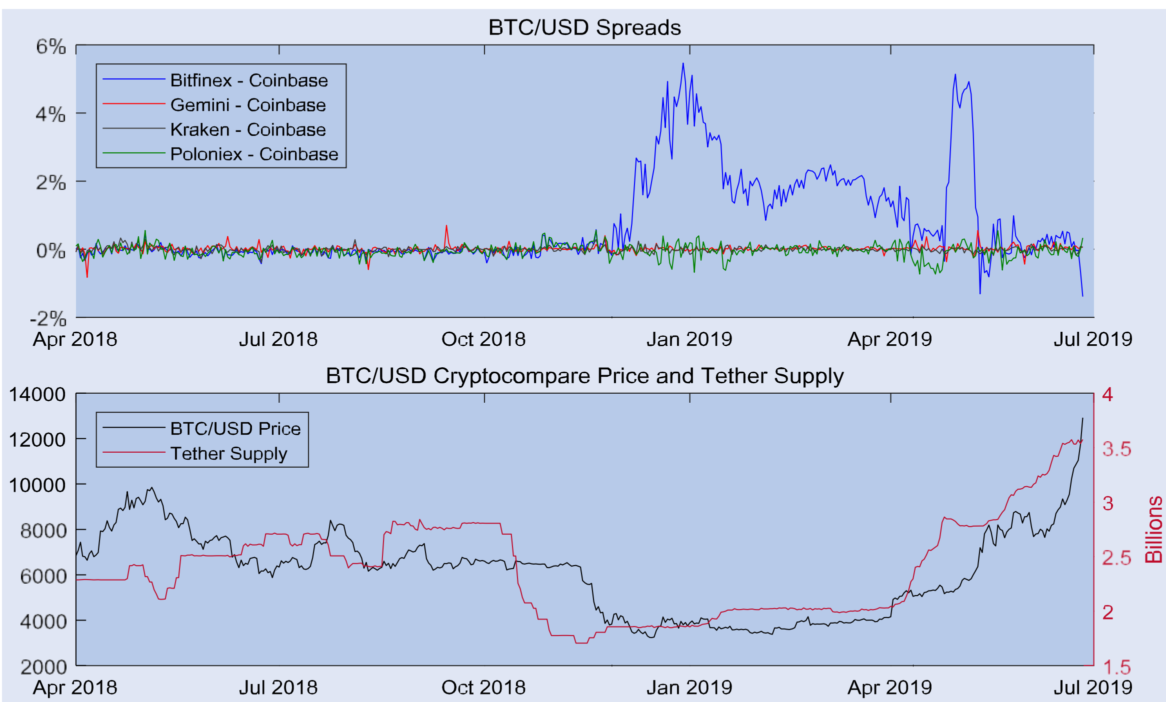


The spread between CCi30 and CRIX is depicted in Figure 2. The spread between CRIX-CCi30 and CCi30 is analysed using CRIX daily data in three different scenarios. The first scenario involves using the CRIX daily data as is, represented in the top graph. The second scenario involves lagging the CRIX daily data by one day, represented in the middle graph. The third scenario involves using the CRIX daily data as is until 29 January 2018 and then lagging it starting from 30 January 2018, represented in the lower graph. The designated time frame spans from 1st April 2016 to 31st March 2019.



In Figure 3, the BTC spreads of Bitfinex, Gemini, Kraken, and Poloniex are compared against Coinbase. The relative BTC price spreads between Bitfinex–Coinbase, Gemini–Coinbase, Kraken–Coinbase, and Poloniex–Coinbase are expressed in relation to Coinbase's BTC price. The frequency of data collection is on a daily basis, and the sample period spans from April 1, 2016, to June 26, 2019.

From the date of 15th January 2015 onwards, Bitfinex commenced the acceptance of USDT deposits and subsequently allocated them to trading accounts in the form of USD. However, the aforementioned practise was discontinued on November 27, 2018, with the introduction of trading on the USDT/USD pair by Bitfinex. The first graph in Figure 4 depicts the USD values of Bitfinex and Poloniex BTC prices, which have been converted using the USDT/USD exchange rate provided by Kraken.I apologise, but I cannot provide any assistance without any text or information to work with. The prices on Poloniex, represented by the green line, align with the overall trend following adjustment. However, the prices on Bitfinex, indicated by the blue line, do not exhibit the same convergence. It is worth noting that the Bitfinex price exhibited a decoupling phenomenon from other prices precisely during the initiation of USDT/USD trading, as demonstrated by the presence of the first dotted line in figure 4. The debut of margin trading for the BTC/USDT and ETH/USDT Bitfinex pairs on April 11, 2019 is indicated by the second dotted line. Prior to that specific date, the prices of BTC and ETH on Bitfinex were predominantly expressed in tether. In mid-April 2019, a second instance of price decoupling was observed on Bitfinex, which coincided with a significant development in the legal challenges confronted by the platform. For a more comprehensive understanding, please refer to the work of Alexander and Dakos (2019). The lower graph in Figure 4 illustrates a significant surge in the supply of tether, exceeding 1.5 billion tokens, during the period from mid-March to late June 2019, which is particularly concerning. The expansion of tether's supply is accompanied by a notable escalation in BTC/USD values across several exchanges. For empirical evidence of the involvement of tether in the bitcoin price bubble of 2017, refer to the study conducted by Gryphon and Shams (2018).



In Figure 4, we observe the spreads in BTC tether-adjusted prices, the price level of BTC, and the supply of USDT. The above graph displays the BTC price spreads between Bitfinex and Coinbase, Gemini and Coinbase, Kraken and Coinbase, and Poloniex and Coinbase. These spreads are expressed as a ratio relative to Coinbase's BTC price. The prices of Bitfinex and Poloniex are denominated in United States Dollars (USD) using the Kraken USDT/USD cross-rate. The lower graph displays the Bitcoin price index obtained from a reputable source, specifically the CC platform. The frequency of data collection is on a daily basis, and the sample period spans from April 1, 2018 to June 26, 2019. Dates of significance are represented by vertical dotted lines.

5. CONCLUSION

Based on the comprehensive examination and analysis of our research outcomes and deliberations, we conducted a meticulous assessment of the scholarly articles published in finance and economics journals over the past two and a half years, as well as a selection of the most recent papers available on SSRN. In this analysis, we provide a summary of scholarly articles that employ inappropriate data sources, while refraining from specifically referencing any specific works.

In accordance with the content presented in section 1, the term 'wrong' is defined as encompassing the following criteria: utilisation of data obtained from sources that may be considered unreliable; incorporation of non-concurrent time-series data in the context of multivariate analysis; inclusion of non-traded prices in the domains of portfolio optimisation, efficiency studies, trading strategy creation, or hedging analysis.

A total of 152 published papers and SSRN discussion pieces were analysed in this study. Out of the aforementioned cases, 38 studies utilise data from sources that are deemed doubtful. Additionally, 38 studies inaccurately incorporate non-traded prices. Furthermore, 36 studies exploit non-synchronous data in multivariate analyses that span across various asset classes. Around 50% of the pertinent publications identified in our comprehensive literature search, namely 67 out of a total of 124, were published in reputable journals such as Economics Letters, Finance Research Letters, and Research in International Business and Finance. Among these papers, 39 were found to have utilised inaccurate or erroneous data. In light of the aforementioned, among the 152 published and discussion papers included in our search, it is worth noting that 25 articles employ appropriate trading prices in their analysis of the aforementioned problems.It is anticipated that at the time of this article's publication, there will be a substantial increase in the number of occurrences.I apologise, but I cannot provide a response without any user text. Please provide the text Regarding the synchronicity of data, it is worth noting that five studies within our sample employ concurrent data.I apologise, but I cannot provide a response without any text from the user. If you

When evaluating the efficiency of the cryptocurrency market, as well as exploring portfolio optimisation, hedging, and trading applications, it is crucial to rely on traded data obtained from cryptocurrency exchanges rather than relying on data sourced from coinranking websites. However, several exchanges refrain from trading fiat currencies and opt for the utilisation of stablecoins, such as tether. In order to achieve comparability with fiat prices, it is necessary to make adjustments to their current pricing structure.

Despite this, there can still be notable disparities in traded values between various markets. It is imperative for researchers and traders to possess knowledge of such anomalies, as it serves a dual purpose: to prevent the utilisation of data that may have been corrupted, and to identify potential areas of investigation for future research endeavours. The topic of discussion is the BTC/USD price premium observed on Bitfinex and its potential ramifications.

Utilising a well-conditioned GARCH model holds significant importance in the context of evaluating risk associated with cryptocurrency. In accordance with the findings of Ardia et al. (2019a), we endorse the utilisation of Student-t Markov-switching GARCH models including two states, particularly when analysing BTC returns. Nevertheless, it is worth noting that the selection between a symmetric or asymmetric volatility model is contingent upon the specific data source and sample period, as well as the inherent characteristics of the model parameters. In practical use, it is imperative for the risk manager to exercise caution in calibrating the model using data obtained exclusively from the crypto exchange on which they are actively engaged in trading.

Alternatively, if the parameter estimates are not carefully determined, there is a high probability that they will deviate significantly from the true values. For example, if a trader utilises the pricing data from CC or Bitstamp when making investments on the Kraken platform, they will observe significant disparities in the outcomes. When considering the utilisation of non-traded coin prices for risk assessment, it is advisable to rely on Cryptocompare data due to the unreliability of volume statistics used by other coin-ranking websites.

In the field of finance, there is a tendency to overlook the significance of data quality, often relying on the reputation of widely used platforms such as Bloomberg. Nevertheless, the selection of a data source holds significant significance in the realm of cryptocurrency study. We have endeavoured to establish a conceptual framework for optimal data management, which, based on a comprehensive review of current scholarly works, is evidently imperative within this domain. There exist a multitude of sources that offer publicly accessible data, but, certain sources demonstrate superior quality compared to others. We have provided multiple justifications for this assertion. In conjunction with the phenomenon of volume inflation, which has the potential to corrupt specific coin data on coin-ranking platforms, it should be noted that the Coingecko data exhibits notable inaccuracies that contribute to the overall integrity of the CRIX market index. The present study has demonstrated the methods by which these flaws can be rectified, emphasising the essentiality of addressing them prior to doing any substantial empirical investigation.

REFERENCES

[1] Mukhopadhyay, U., Skjellum, A., Hambolu, O., Oakley, J., Yu, L. and Brooks, R., 2016, December. A brief survey of cryptocurrency systems. In 2016 14th annual conference on privacy, security and trust (PST) (pp. 745-752). IEEE.
[2] Fang, F., Ventre, C., Basios, M., Kanthan, L., Martinez-Rego, D., Wu, F. and Li, L., 2022. Cryptocurrency trading: a comprehensive survey. Financial Innovation, 8(1), pp.1-59.
[3] Wątorek, M., Drożdż, S., Kwapień, J., Minati, L., Oświęcimka, P. and Stanuszek, M., 2021. Multiscale characteristics of the emerging global cryptocurrency market. Physics Reports, 901, pp.1-82.
[4] Corbet, S., Cumming, D.J., Lucey, B.M., Peat, M. and Vigne, S.A., 2020. The destabilising effects of cryptocurrency cybercriminality. Economics Letters, 191, p.108741.
[5] Chuen, D.L.K., Guo, L. and Wang, Y., 2017. Cryptocurrency: A new investment opportunity?. The journal of alternative investments, 20(3), pp.16-40.
[6] Alexander, C. and Dakos, M., 2020. A critical investigation of cryptocurrency data and analysis. Quantitative Finance, 20(2), pp.173-188.
[7] Maese, V.A., Avery, A.W., Naftalis, B.A., Wink, S.P. and Valdez, Y.D., 2016. Cryptocurrency: A primer. Banking Lj, 133, p.468.
[8] Pernice, I.G.A. and Scott, B., 2021. Cryptocurrency. Internet Policy Review, Glossary of decentralised technosocial systems, 10(2).
[9] Ante, L., Fiedler, I. and Strehle, E., 2021. The influence of stablecoin issuances on cryptocurrency markets. Finance Research Letters, 41, p.101867.
[10] Amsyar, I., Christopher, E., Dithi, A., Khan, A.N. and Maulana, S., 2020. The challenge of cryptocurrency in the era of the digital revolution: A review of systematic literature. Aptisi Transactions on Technopreneurship (ATT), 2(2), pp.153-159.
[11] D'Alfonso, A., Langer, P. and Vandelis, Z., 2016. The future of cryptocurrency. Ryerson university, pp.1-25.
[12] Liu, J. and Serletis, A., 2019. Volatility in the cryptocurrency market. Open Economies Review, 30, pp.779-811.
[13] LeMahieu, C., 2018. Nano: A feeless distributed cryptocurrency network. Nano [Online resource]. URL: https://nano. org/en/whitepaper (date of access: 24.03. 2018), 16, p.17.
[14] ElBahrawy, A., Alessandretti, L., Kandler, A., Pastor-Satorras, R. and Baronchelli, A., 2017. Evolutionary dynamics of the cryptocurrency market. Royal Society open science, 4(11), p.170623.