POLITECNICO MILANO 1863

Thesis Project

Anno accademico: 2020/2021

Stablecoins: Stability Analysis and Price Manipulation

Relatore:
Emilio Barucci

Studente:
Piervito Coletta

Abstract

Following the same rise in popularity of Bitcoin and Ethereum, Stablecoins are set to become the future of digital transactions. Indeed, given a completely different structure, these new Coin aims to maintain a stable value compared to the already proven high volatility of Cryptocurrencies. The increasing interest, along with the general lack of solid and well-oiled regulation of the Exchange Markets, has led, however, to the diffusion of illicit behavior directly connected to these Cryptocurrencies. This project will focus on one particular scheme, the Pump and Dump Scheme, which has generated over 300M$ of profit in a single year. After a stability analysis of the price, the main goal is to determine whether the Stablecoins considered, DAI and USD Coin, have been targeted by this kind of activity and can be susceptible in case of future attacks.

Contents

Chapter 1

Introduction

Since the creation of the first Bitcoin, digital currencies have always been associated with various illegal activities, undermining the credibility of this new possible meter of wealth. From Dark Web transactions to price manipulations, the life of Cryptocurrencies has never been relatively calm. In particular, due to the nature of these coins, based on decentralization, and the underestimation of the Crypto Phenomenon, which now values almost 2 Trillion, governing financial institutions left Exchange Markets unregulated for almost a decade.

In this unsupervised environment, numerous price manipulation has taken place. The first massive one occurred on the Mt. Gox Exchange in 2013. Gandal et al.[12] shows how the event developed and its direct consequences. Hackers stole a substantial amount of coins (approximately 650.000 units) from the exchange servers. In order to avoid the collapse of the Exchange due to the lack of capital, Mt. Gox created a scheme to attract fiat capital. To begin with, they began to place fake massive purchase orders. The first bot covered a longer time interval, buying a smaller amount of coins daily, up to 335K coins. In contrast, the activity of the second bot was more intense, leading to a purchase of 250K coins in a fourth of the time interval covered by the first bot. Therefore the Mt.Gox exchange bought almost 600K of coins, yet no actual transaction took place. Indeed, the cryptocurrency exchange market functions as banks where customers buy and sell coins but typically maintain balances of fiat currencies and coins on the Exchange without retaining direct access to the currency. In

this way, Mt. Gox created interest, increasing trading volume, fraudulent activity set apart, of approximately 2% during the active bots period.
At the end of this operation, the Exchange gained money from collecting transaction fees and, more importantly, converted consumer bitcoin balances into fiat money, hiding the missing coins. As long as the user would remain confident in the stability of the Exchange, the default could be avoided. However, as the story tells, the consumers tried to withdraw money from the ecosystem of the Exchange, unveiling the fraud that the Mt.Gox owner put in place. The Mt.Gox owner didn't expect the massive consequences that this price manipulation would create. Indeed after a descriptive analysis Gandal et al.[12], discovered that all the other exchange that were operating the same pair BTC-USD, were influenced by the bots. The magnitude of the increase can be considered similar to the one occurred in the Mt.Gox, up to 5% in percentage rate change of the price.

What happened in 2013 was unprecedented in crypto framework and almost surely never to occur again due to the regulations that governing institution are imposing. Nowadays what is concerning most is the impact of a large community that as an organized entity matching the fire power of an investment bank can do in this online trading system. It already happened with Gamestop and the sudden upward shift of 300% of the stock value of firm. The goal of this project is to understand how much stable coins can handle this kind of activity.

1.1 Cryptocurrency as Currency

Since the first coin was created, Cryptocurrencies have drawn enormous attention, introducing an actual application of the concept of decentralization, thanks to the distributed ledger (DLT). A DLT is essentially a record of information or database shared across a network without the need for a central validation process. It is considered by most of the important central banks in the world a powerful tool which potential has yet to be exploited. The infrastructure offered through DLT, be it in public or private form, could serve as a record of holdings and be used to transfer various kinds of assets. So when Bitcoin White Paper was released in 2008, the discussion about the future of money was triggered, leading to the creation of more than 2000 cryptocurrencies. However, it was clear from the beginning that none of these crypto-assets could be considered the future of digital transactions. Indeed the European Central Bank [8] defines three different functions that have to perform a tool in order to be considered as "money":

- **Medium of exchange for buying things**: it is a means of payment with a value that everyone trusts.
- **Unit of account for pricing**: it is a unit of account allowing goods and services to be priced.
- **Store of value for savings**: it represents a meter of wealth and indeed only a portion is actually circulating.

Cryptocurrencies cannot meet the requirements to be considered a form of money given the definition that the ECB [8] is using:

- **Crypto-Asset**: *"a new type of asset recorded in digital form and enabled by the use of cryptography that does not represent a financial claim on, or a liability of, any identifiable entity"*.

The lack of an underlying makes the asset highly unstable, leading to a risky application as a store of savings.

1.2 Stablecoin Design

In search of stability in cryptocurrencies, developers around the globe created Stablecoins. Again this brand new tool, as for the cryptocurrencies in 2008, lacks an agreed definition, so the one used by the ECB [8] will be considered:

- **Stablecoin**: *"digital units of value that are not a form of any specific currency (or basket thereof) but rely on a set of stabilisation tools which are supposed to minimize fluctuations of their price in such currency(ies)"*.

Each type of Stablecoin differs in three criteria: i) the existence/absence of an issuer that is responsible for satisfying any attached claim; ii) the decentralization/centralization of responsibilities over the Stablecoin initiative; iii) what underpins the value of a Stablecoin and its stability in the currency of reference. By these criteria, four classes of Stablecoins can be identified:

- **Tokenised Funds**: supported by funds, which implies the issuer's commitment to their redeemability and the need for someone (possibly a custodian) to take responsibility for their safekeeping.
- **Off-chain collateralised stablecoins**: supported by other traditional asset classes, which require a custodian for their safekeeping and are in the possession of the issuer only as long as the user does not claim them back.
- **On-chain collateralised stablecoins**: supported by assets, typically crypto-assets, which can be held for safekeeping in a decentralised manner and do not need an issuer to be identified.
- **Algorithmic stablecoins**: supported solely by users' expectations about the future purchasing poweer of their holdings, which does not require the accountability of any part, nor the custody of any underlying asset.

To better understand the differences of these stablecoins design a *crypto-cube* can be produced where each criteria occupy an axis.
The analysis are going to be performed on two Stablecoins, **USD Coin** (Tokenised Funds) and **DAI** (On-chain collateralised Stablecoins).

1.2.1 Tokenized Funds: USD Coin

Units of monetary value that are stored electronically in a distributed ledger to represent a claim on the issuer and are issued, for the purpose of making payment transactions to persons other than the issuer, are often labelled "fiat-backed stablecoins". These coin does not represent actually a new form of asset but rather represent axisting currency units in a distributed ledger. So the fund undergo a tokenization process making possible transaction also on the DLT. To own a tokenised funds corresponds to a claim on the issuer over the funds it received from users. In the case a custodian is needed to channel the funds, the issuer has to be identifiable and accountable in order to enter into an agreement with the custodian of the funds. Moreover the funds must be redeemable

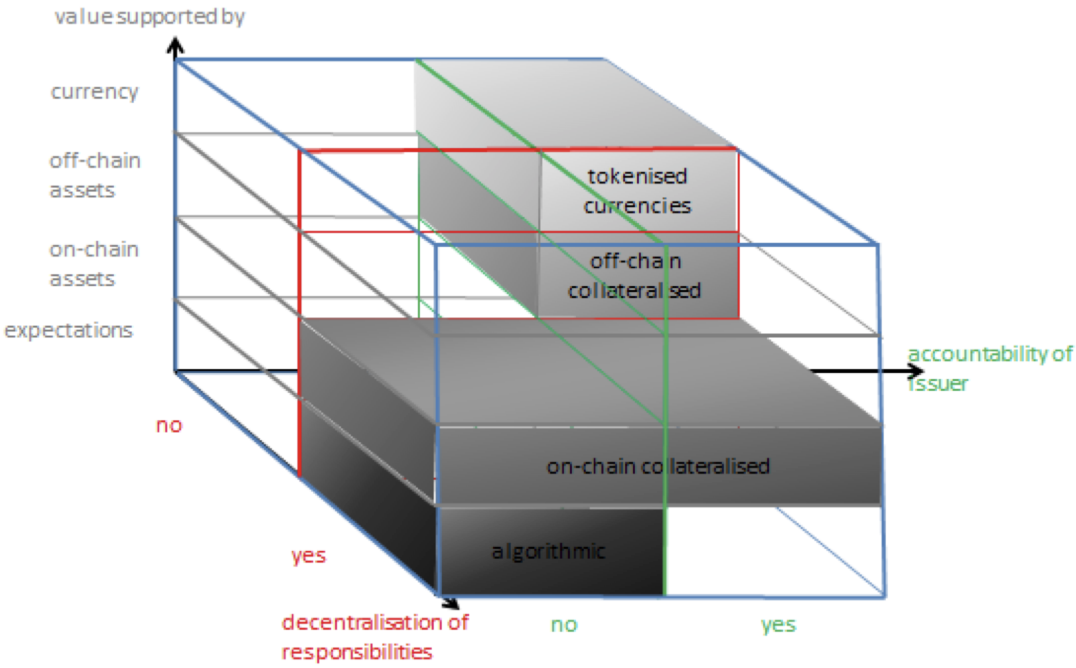


Figure 1.1: Taxonomy of stablecoins within the “crypto-cube" [8]

according to the terms of service communicated to users. Three operations characterize a tokenized funds:

- **Issuance**: User transfer funds to the issuer’s account, opened with the custodian who keep them safe. Upon confermation of the transfer, the issuer creates and allocates an equivalent amount of Coins through the smart contract it mantains.
- **Transfer**: Sender initiates a transfer to a receiving user by instructing the smart contract. DLT participants verify and validate the transfer.
- **Redeeming**: User sends units of tokenized funds to the address specified by the user who will withdraw them from circulating, in order to mantain redeemability of circulating units. Then custodian transfer corresponding amount of funds back to the user.

The USD Coin (USDC), in particular, was the product of the direct tokenization of the US Dollar(USD) until June 2021, when Circle, the issuer of the coin, changed the fund backing the stablecoin, in "fully reserved assets" redeemable on a 1:1 basis for US Dollars. The USDC nowadays occupies the second position, between all stablecoin, according to market capitalization which amount to a total of 27 Billion USD.

1.2.2 On-chain Collateralized Stablecoins: DAI

Collateralized stablecoins are backed by units of an asset (or assets) that the coin owner can redeem. However, the collateral price can fluctuate over time in the currency of reference, in contrast to tokenized funds whose smart contract guarantees the ratio at which it is possible to redeem the value of the stablecoins. Therefore this type of coin is characterized by over-collateralization to ensure that every stablecoin is backed by collateral valued at par in the currency of reference. In this way, users can store the proceeds from crypto-assets without the need to go through the service of the trading platform and can avoid the penalty fee associated with the default of collateral position. The On-Chain collateralized stablecoin relates to assets in digital form and, for this reason, can be completely decentralized, delegating the control process to the DLT participants. Thus the operations of issuance, transfer, and redeeming are completely different from tokenized funds:

- **Issuance**: The user sends directly to the address of the smart contract that governs the scheme the on-chain collateral. So the smart contract then creates the coins and sends them to the user

- **Transfer**: Unlike the tokenized funds, no central party is needed so that DLT participants maintain the smart contract governing stablecoin transfer.

- **Redeeming**: It can be voluntary. Thus user sends stablecoins to the smart contract, which burns them and returns to the user address the equivalent on-chain collateral. Alternatively, it can be compulsory. It can happen when the circulating stablecoins are under-collateralized due to a possible default of the collateral. In this case, smart contracts search for new funds to retrieve the under-collateralized coins and issue new secondary backed coins.

The DAI, in particular, is a stablecoin backed by on-chain collateral (Ethereum more precisely) with a floating peg to 1USD, associated with the secondary token MKR, issued both by MakerDAO. It is a "decentral autonomous organization represented by rules encoded as a computer program transparent and controlled by MKR holders", which have governance duties along with failure responsibilities, thus also incentivized by potential seigniorage revenue.

1.3 Conditional Volatility

The analysis of the volatility is probably the most common process that has been performed talking about Bitcoin. So it is commonly accepted the highly volatile characteristics associated with Bitcoin in contrast to the traditional currencies. Indeed, numerous model has been implemented to understand and predict the volatility of cryptocurrencies. The most used and analyzed model is the GARCH used to compute conditional volatility, which is considered the best quantity to describe the stability.

Introduced by Bollerslev(1982) [6], the GARCH model has been the key to compute the variance of a time series conditional on the past values, leaving the unconditional variance constant. Starting from the ARCH model introduced by Engle(1982) [11], the Generalized Autoregressive Conditional Heteroscedasticity model showed a longer memory and a more flexible lag structure. Nevertheless, GARCH models showed some limitations, according to Nelson(1991). The starting point of the critics relied on the symmetry property, which could not distinguish the negative jumps correctly from the positive ones. Then he proceeded to list other weaknesses: the adverse correlation between assets' returns and changes in return volatility, the nonnegativity constraints of the parameters complications, and the persistence of shocks.
Thus four features were introduced in order to classify a model as a good choice. First, the model should show volatility clustering (a period of high volatility must follow another with a similar volatility level). Second, the model must present the mean reversion component, meaning that after some shocks in the time series and thus in volatility, the latter's level must return to some average level. Third, the model must implement an asymmetric structure. Otherwise,

the impact of positive and negative shocks will be considered the same. Finally, exogenous variables have to be added in case the phenomenon considered requires it.
As a consequence, a considerable amount of versions of the GARCH model were introduced. In Chu et al. [9], a large-scale study has been conducted to determine the best fitting GARCH model in the cryptocurrencies framework. The research interested twelve different GARCH models fitted on the seven most popular cryptocurrencies traded at that time. The selection process was based on a comparison of five quantities, AIC, BIC, CAIC, AICc, and HQC. The results show that IGARCH and GJR-GARCH are the models that provide the best fits. However, the choice of the IGARCH can lead to some significant alteration of the predicted volatility due to its infinite memory feature.

1.3.1 Stablecoin Stability

Regarding Stablecoins, state of the art in volatility analysis is not as consistent as the bitcoin one. In Hoang et al. [14], a stability analysis focused on studying the volatility of the six largest stablecoins by market capitalization. The project manages to fit an AR(1)-TGARCH(1,1) (asymmetric model) to the time series of the exchange rates against the US Dollar in order to compute the conditional volatility. Moreover, they fitted the same model also to the prices of Bitcoin, two fiat currencies (EUR and USD) and gold. Then to determine whether the stablecoins could maintain their promised stability, confronted the conditional volatility through a correlation computation. Results were not promising, showing high volatility at intraday level, precisely 5-min and hourly frequencies.
Moreover, Hoang et al.[14] warned against the use of the Realized Volatility Estimation in this kind of stability analysis. Indeed, BTC and USDT showed different distributions at different frequencies. Thus when computing the RV, which values considerably the frequency distributions summing up squared returns during the day, a poor and biased measure of relative volatility was obtained.

1.4 Pump and Dump Scheme

Pump and Dump scheme (P&D) is the form of price manipulation considered in this project, that involves artificially inflating an asset price before selling the cheaply purchased assets at higher price.

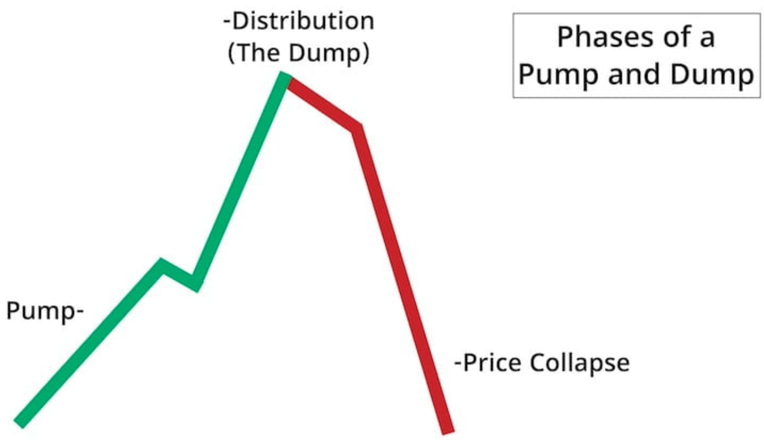


Figure 1.2: Pump and Dump Pattern

This practice is deemed illegal in the stock market by most countries worldwide. Nevertheless, due to the weaker regulations that the Crypto exchanges have, opposed to the stock one, it has taken hold in the cryptocurrencies framework, essentially focusing on microcap coins.
A P&D is a trade-based manipulation of the price, meaning it involves manipulating the price of a security by buying and then selling or vice versa. Thus to operate a successful P&D is necessary that a large amount of user participate. Thanks to the use of the Telegram channel, the pump organizer invited other investors to join the group advertising the operation on social media as an easy and highly profitable method. Then, using the Telegram channels or Discord groups, the operator announces the target date, time, and exchange usually a day in advance, not disclosing, however, the cryptocurrency to purchase until the scheduled time. When the pump begins, the operator, and probably another tiny group of people affiliated to, sells the coins they previously bought

in large quantities, triggering the dumping effect that intensifies when all the other participants, fearing losses, sell, in turn, their coins.

1.4.1 Players

To understand how this manipulation takes place, Dhawan et al.[10] classify the participants into two classes, non-manipulators (the ones that don't know the coin to target until the Pump begin) and maniputaltors (the actual organizers of the scheme). Then they separates the non-manipulators into two category:

- Overconfident participants: people who actually believe at the false advertisment the manipulators posts on telegram channel, overestimating the ability to sell the assets at its peak.
- Gambler: people who involve manipulating the price of a security by buying and then selling or vice versa.

Independently of the type of non-manipulator, the P&D is a negative-sum game. Indeed, according to Dhawan et al.[10], the Pump & Dump can be modeled as a four period game.

In the Period 0, the manipulators select the coin to target, which price is P_0. Then, in Period 1, manipulators take long positions of M total units in the coin selected and inform the non-manipulators of the coming P&D. Assuming market orders have linear price impacts, the prices are determined by the function $P_t = P_0 + \beta X_t$ where β is the price impact parameters between 0 and 1, P_t is the price at time t and X_t is the cumulative volume of buys received by the market at time t. So, taking a long position means pumping the price up to $P_1 = P_0 + M\beta$. In Period 2 the Pump signal is launched to the N non-manipulators. Given Exchange markets place order sequentially in a queue, it is possible to define the price at which each non-manipulator buys as $\{(P_1 + 1\beta), (P_1 + 2\beta), ...\}$, so that the combined price impact corresponds to $N\beta$ with price $P_2 = P_0 + M\beta + N\beta$ at the end of Period 2.
Finally in Period 3 the players exit the pump selling the coins bought. As before, sell orders are placed sequentially, executed at prices $\{(P_2 - 1\beta), (P_2 + 2\beta), ...\}$

making the price drop to the initial level $P_3 = P_0$.

Assuming entry prices P_{entry} and exit prices P_{exit} are uniformly distributed respectively $\{(P_0 + \beta(M+1)), (P_0 + \beta(M+2)), ..., (P_0 + \beta(M+N))$ and $\{(P_0 + \beta(M+N-1)), ..., (P_0)$ the expected profit of a non-manipulator individual is:

$$E[\pi_{nm}] = E[P_{exit} - P_{entry}] = -\frac{\beta(2+M)}{2}$$

The player expects a loss composed of half the initial price impact and the round-trip trade cost, not considering any transaction fees. These results imply that no rational player with a risk-averse or risk-neutral utility function would ever enter a Pump and Dump game.
Dhawan et al.[10] also computed the expected profit of a manipulator player:

$$E[\pi_m] = E[P_{exit} - P_{entry}] = \frac{\beta M}{2}(N - 2M)$$

Thus also the manipulator faces some risks which depends on the level of involvement in the event. The value computed, of course, is the sum of all the losses of the non-manipulators participants.

1.4.2 Detection

Exchange regulators never offered an optimal tool to detect an ongoing pump and dump to block the illicit activity from completing. The only detection that has been conducted is post-event, mainly due to the reduced duration of a crypto P&D. In Victor and Hagemann[23], they identify three classes of P&D events that can occur:

- Sustained Pump: This type of pump leads to elevated price levels that are sustained hours after the pump. Although there typically is an immediate price peak, from which there will be a drop, the price does not return to the initial levels.
- Short-term Pump: When the price briefly increases, but returns to or below initial levels, then the pump was only short term, immediately followed by a selloff (dump).

- Failed Pump: When the price briefly increases, but returns to or below initial levels, then the pump was only short term, immediately followed by a selloff (dump).

Thanks to this classification and the telegram chat history, they train a classifier conducting a supervised detection. The precision obtained from the XGBoost in the Cross-Validation test was acceptable, yet in an unsupervised framework, even with a trained classifier, numerous P&D events were detected, not guaranteeing, of course, that an actual P&D scheme has been successfully conducted. Unfortunately, in this project, a telegram history chat could not be retrieved. Thus it has been implemented only an unsupervised analysis following the instructions of the Kamps et al.[17] paper.

1.4.3 Indicators

Generally, a crypto P&D lasts for only several minutes, in contrast to the stock market, which can reach months. Moreover, the pumping effect is due to the combined purchase of thousand of people and is not based on the release of false information. The only case of the Pumping effect in the stock market that had similarity to the one on crypto exchanges is the Gamestop case, where almost 2 million users bought an incredible amount of stocks pumping up to 300% the initial value.
The Kamps et al. [17] provides the peculiarities that distinguish a crypto P&D from a penny stock one in Table 1.1:

	Stock	**Crypto**
Target	Low market cap Low volume Low price Lack of reliable information	Low market cap Low volume Low price Lack of reliable information
Tactic	Misinformation Privately organised (smaller scale)	Real-Time Misinformation Public or private group scams (larger scale)
Timescale	Medium (days to weeks)	Short (minutes to hours)

Table 1.1: Comparison of traditional and crypto pump-and- dump schemes

Moreover Kamps et al. [17] provides a list of indicators, using the identified characteristics of a crypto P&D dividing them in two groups, to determine if a price manipulation occurred:

- **Breakout Indicators**: Signals that will always be present during a Pump and Dump Scheme
- **Reinforcers**: Indicators which increase confidence that observed data are the result of a price manipulation

Indicators of pump-and-dumps per temporal dimension and indicator type

	Temporal Dimension	
	Real Time Indicators	**Post-pump Indicators**
Breakout Indicators		
Volume	Has the volume at the current data point been significantly higher than in the estimation window?	Was there a decline in volume after the event window where a pump was detected?
Price	Has the price at the current data point been significantly higher than in the estimation window?	Was there a decline in price after the event window where a pump was detected?
Reinforcers		
Market Cap	Is the market cap of the coin relatively low? (+)	
Number of Exchanges	Whether the coin is listed on multiple exchanges and the indicators only spike on one (+)	
	Whether the coin is not listed on multiple exchanges (+)	
Symbol Pair	Whether the coin is trading for BTC or some other cryptocurrency (+)	
	Whether the coin is trading for USD or some other fiat currency (-)	

Table 1.2: (+) corresponds to an increase in confidence that data shows a P&D, while (-) corresponds to a decrease in confidence

1.4.4 Susceptibility of Price to Volume Shocks

Due to the difficulties that the detection of the P&D poses, an analysis of the presence of Granger Causality of the Volume on the price can be useful. Indeed, in a trade-based manipulation, it is crucial that the price of the asset targeted inflates as the trading volume increases. Thus, what a P&D causes is a shock in the price pattern, meaning a non parametric Causality-in-Quantiles test can implemented following the work of Balcilar et al[2].
As first step, they checked if a linear Granger cause test could be applied, looking for non-linearity or structural break in the time series of the bitcoin returns. The BDS test results showed signs of non-linearity, meaning a linear approach would lead to misspecification and thus unreliable results. Then they implemented a non-parametric Causality-in-Quantiles test to understand if the volume could predict the returns.
Results showed a significant level of causality over the quantiles range of $0,25$ to 0.75 of the conditional distribution of returns. However, When the market is in bearish or bullish phases all information about volume is irrelevant, only past values of the returns can be useful.

Chapter 2

Data

The main data for this project are tick-level trading data of the pair USDC/USD and DAI/USD. The variables contained in the datasets are timestamp of the transaction at the millisecond level, the price at which the trade occurred, the volume of coins traded and a logic array corresponding to the label *sold* or *bought*. For each coin there is a substantial difference between the number of transaction on each exchange (10 times bigger), leading to issue to simple computation such correlation between returns.
The time interval covered starts on the 1^{st} June 2020 and ends on the 30^{th} September 2020 for both exchanges for a total of 112 days. Given the highly irregular structure of the datasets, following the work of Brownlees and Gallo[7], data were downsampled using different approach for different frequencies:

5min and Hour	**Daily**
First Element	*Linear Poin Interpolation*
$y_j^* = y_f$ where $t_f = mint_i \vert t_i \in (t_{j-1}^*, t_j^*]$;	$y_j^* = (1 - \frac{t_j^* - t_p}{t_n - t_p})y_p + \frac{t_j^* - t_p}{t_n - t_p}y_n$

Table 2.1: y: real value, y^*: downsampled value, f:first, n: next, p: previous

The problem in using these methods, is that they might employ information not available at t_j. However the data are dense enough to apply both methods without any danger. In the case of DAI, data regarding transactions on the 15^{th} August 2020 on Coinbase exchange are missing, thus the date will not be considered in the following computations for both exchanges.

2.1 USDC

Data regarding the USD Coin comes from two market exchanges, Bittrex and Kraken. Unfortunately, due to the lower interest in the Bittrex platform, traded volumes belongs to two different orders of magnitude. Indeed before the downsampling, in the period considered, the former registered about $80K$ transactions while the latter more than $1M$. This difference is evident also in the daily traded volume which is represented below:

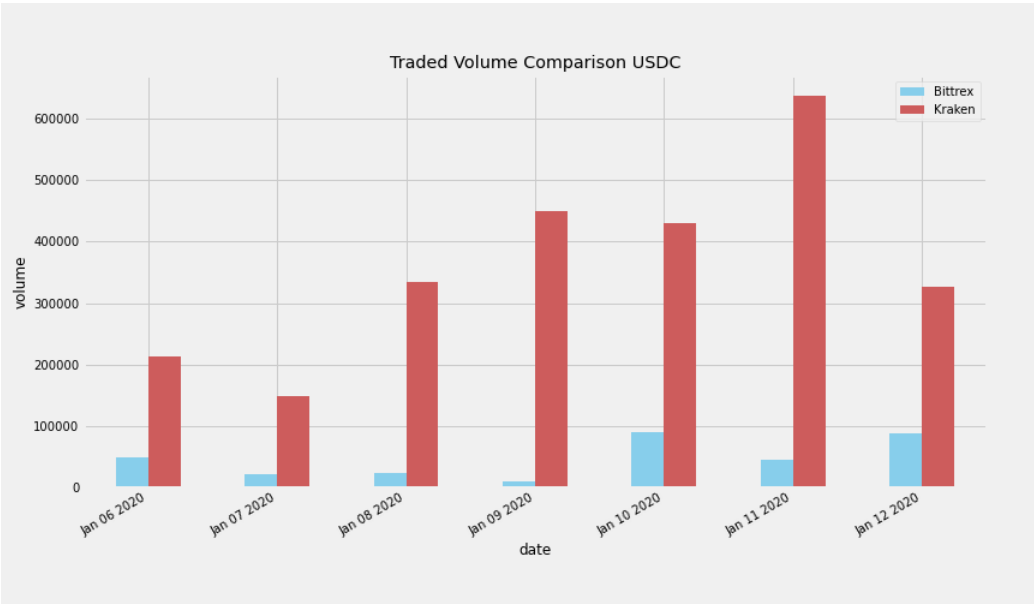


Figure 2.1: Traded volume of USD Coin in the range between 06 June and 12 June

Price Analysis

A preliminary analysis has been conducted to visually understand the pattern behavior of the time series, test the stationarity, and finally check if the data had to be cleaned. The following figure represents the price series sampled at 5 minutes and the moving mean sampled at the hour and day level. A possible pump and dump event signal in the price analysis is the lasting shock present in higher frequencies and remains in lower ones.

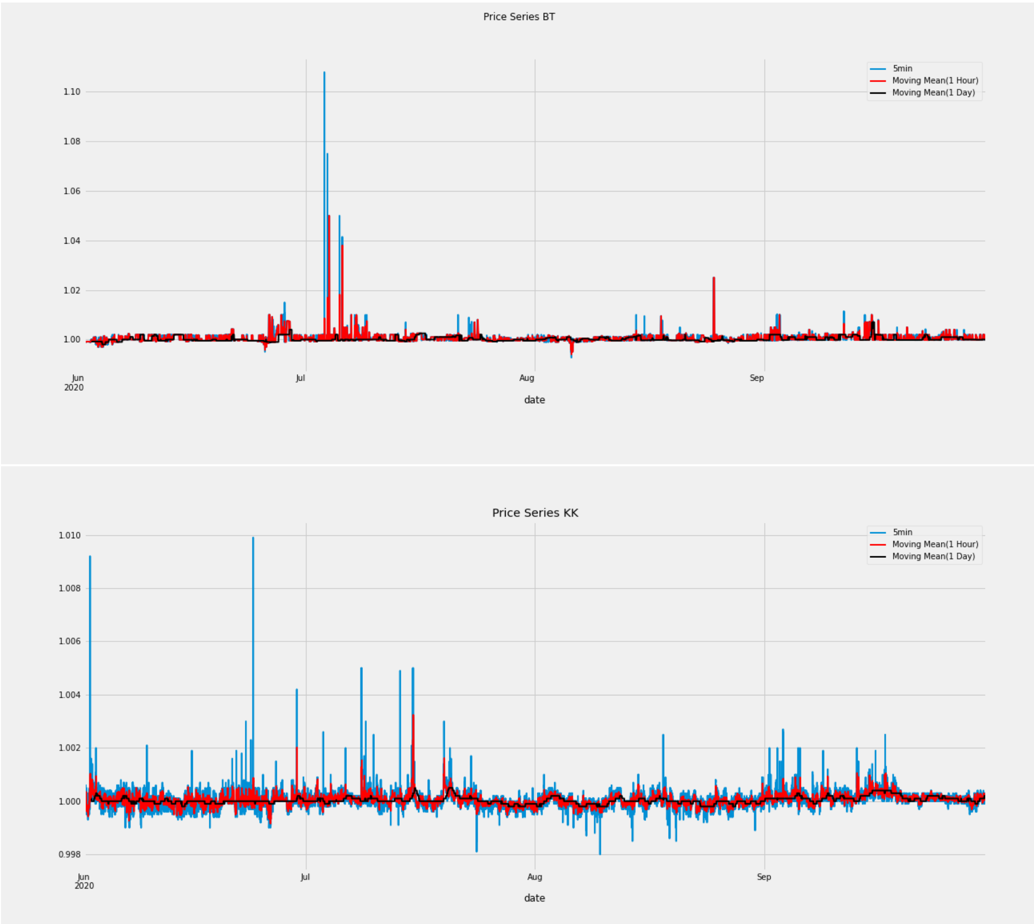


Figure 2.2: 5-min Price dynamics and hourly and daily rolling window on Bittrex(BT) and Kraken(KK)

The two exchanges exhibit different price dynamics, which are persistent at the hourly level; a correlation analysis is necessary to understand whether a possible arbitrage opportunity can be exploited in this framework. It has to be taken into consideration. However, the differences in volume registered to make the price much stable in the case of Bittrex exchange. Apart from these differences, the price at the daily level (in black) has promising behavior, always remaining close to the 1 US Dollar reference value.

Price Analysis: Additive Decomposition

Subsequently a naive analysis of the seasonality can be exploited, eliminating a possible disturbance in the stability and detection computations. Starting from the assumption that the time series is composed by a trend component T_t, a seasonal component S_t and a residual term e_t so that the time series is:

$$X_t = T_t + S_t + e_t$$

The results are obtained by first estimating the trend by applying a convolution filter to the data. The trend is then removed from the series and the average of this de-trended series for each period is the returned seasonal component. The method used is Seasonal and Trend decomposition using Loess, thus a decomposition based on a non linear least square regression.

Figure 2.3 shows that the trend generally confirms the stability highlighted by the boxplots analysis, while the seasonality decomposition can be avoided. Indeed, the quantity ranges between $-0,0004\$$ and $+0,0004\$$ for Bittrex and between $-0,0001\$$ and $+0,0001\$$ for Kraken, thus a negligible quantity. Moreover, there is no real explanation for this kind of seasonality. Thus results can be discarded.

Figure 2.3: Additive Decomposition at hour level of the price on Bittrex(BT) and Kraken(KK)

Returns

All the computation are based on the returns, to avoid the possibility of non-stationarity of the price time series. They are computed as: Where the choice of Log-returns is consistent with the hourly and daily frequencies.
The following figures represent the return dynamics alongside the volume and number of transactions, which occurred in 5 minutes intervals.

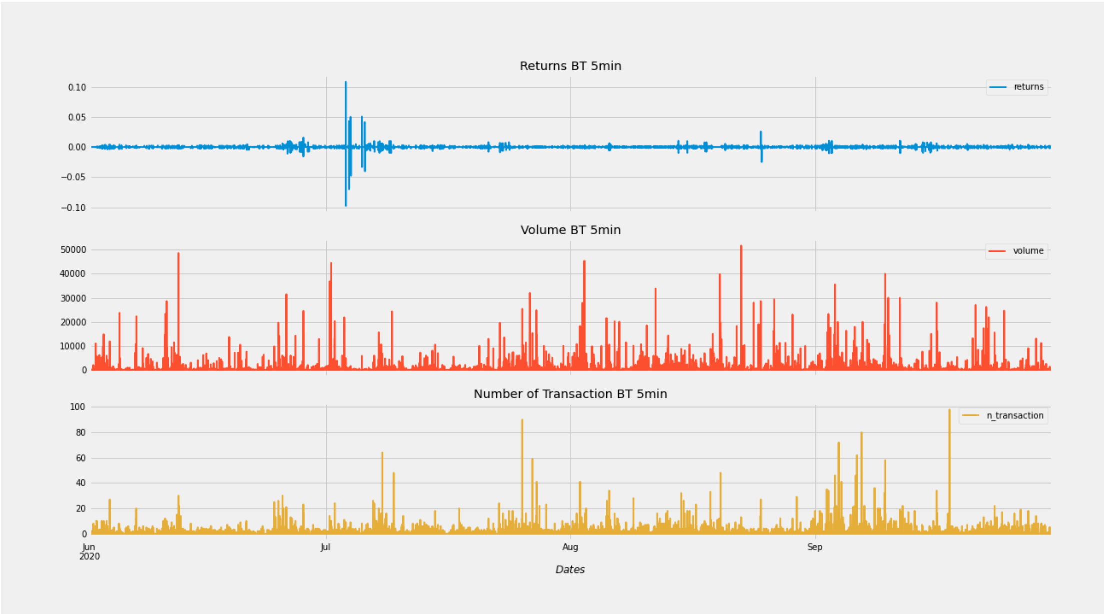


Figure 2.4: Returns, Volume and Number of Transaction dynamics on Bittrex(BT)

Regarding Bittrex returns, the values remain close to zero apart from rare spikes, validating possible stability. Moreover, it can be interesting to check if the volume and number of transaction spikes also correspond to price sudden increase. A promising absence of this correspondence makes the pump and dump event highly improbable.

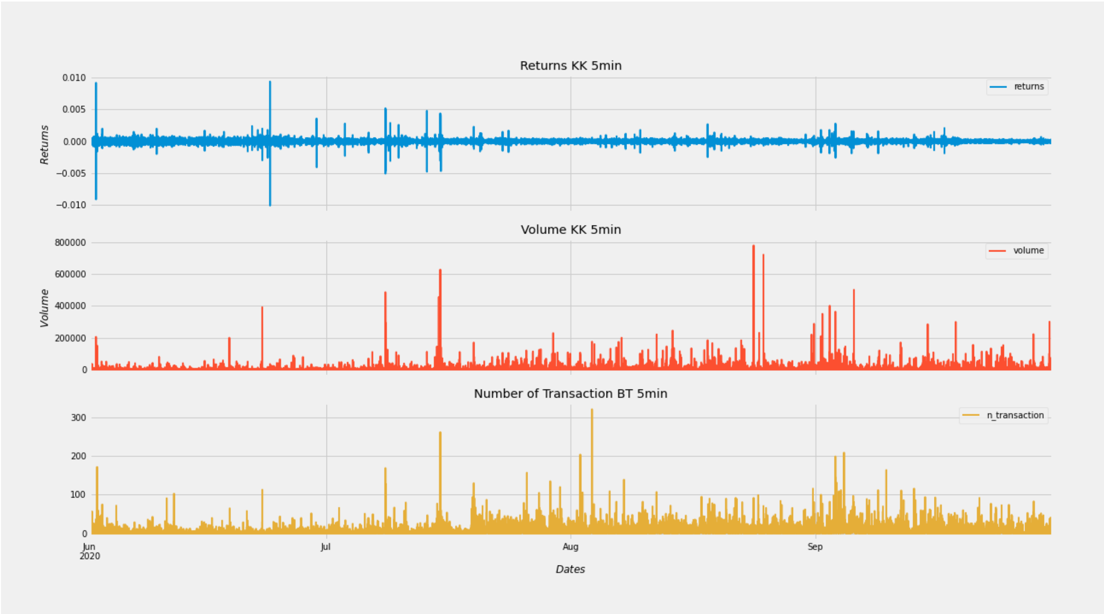


Figure 2.5: Returns, Volume and Number of Transaction on Kraken(KK)

In the Kraken exchange, we have general stability around the zero values as in the previous exchange, but there are much more price spikes that belong to the same time interval of a volume and several transactions spikes. If a pump and dump event occurs, the spikes must happen only on the exchange considered and is not present on the Bittrex exchange.

Returns: Boxplot

A useful view of the volatility of the trading pair is offered also by the boxplot analysis at different level. The colored bar represent the Interquantile Range $IQR = Q_{0,75} - Q_{0,25}$, where Q_α is the quantile of order α of the distribution considered, while rhombuses represents the outliers.

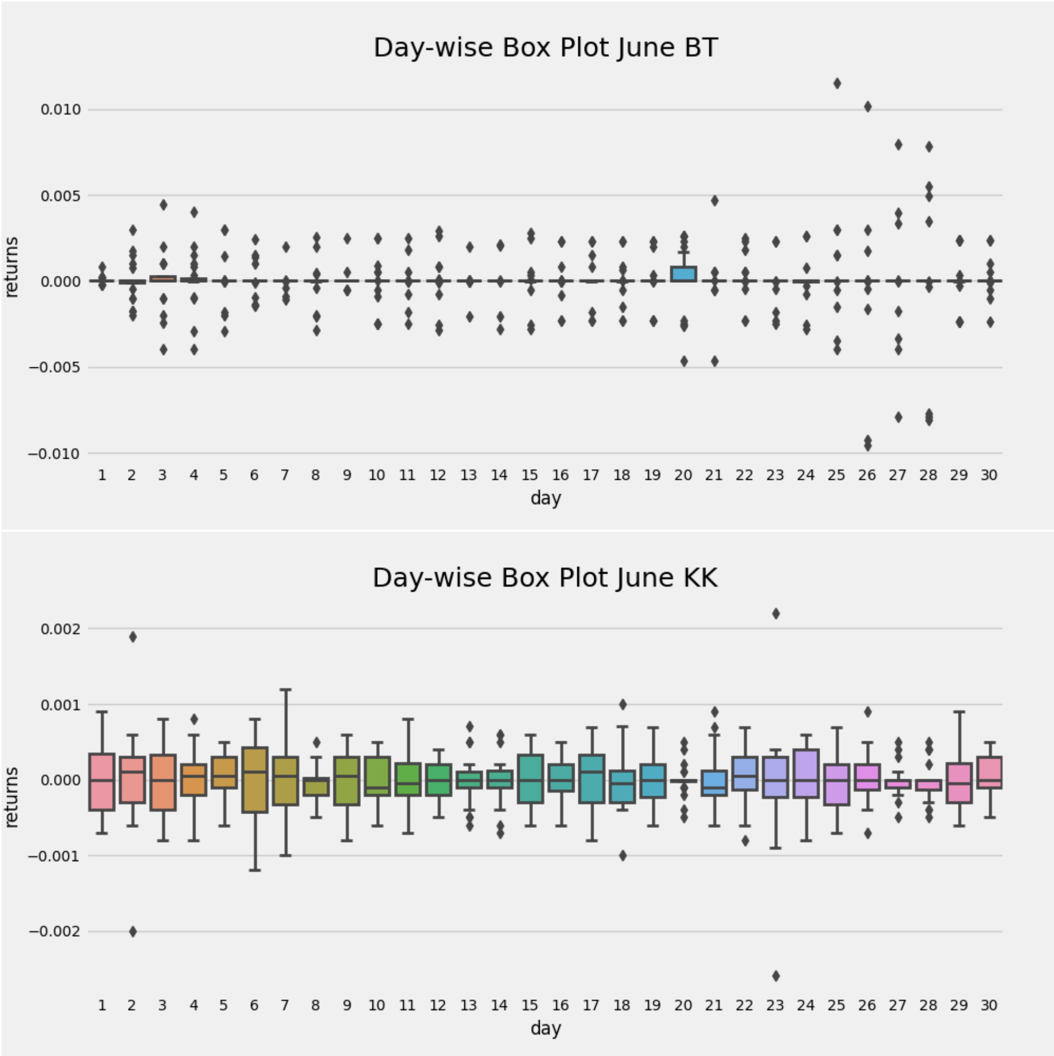


Figure 2.6: Boxplots of the Daily Price on Bittrex(BT) and Kraken(KK)

The boxplots are computed on the 5min distributions of the returns for each day. The Figure 2.6 shows low volatility with interquantile range IQR always under the $0,1\%$ on both exchanges. The higher amount of outliers for the USDC derives from a more anomalous pattern of the price, with multiple spikes, an indicator of a possible presence of Pumping effect.

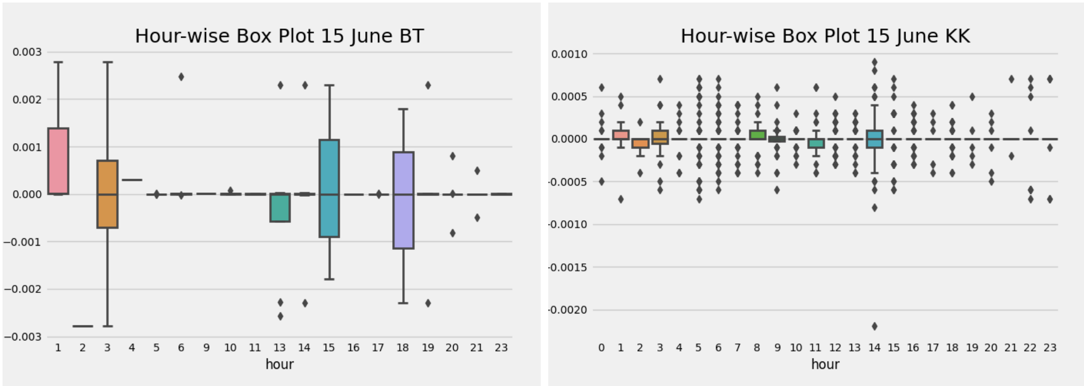


Figure 2.7: Boxplots of the Hourly Price on Bittrex(BT) and Kraken(KK)

Instead in Figure 2.7 boxplots are computed on the 5min distribution of returns for each hour. As expected the intra-day returns are more noisy, with an IQR around the zero values and an abundant number of outliers in the most volatile day identified.

Ruturns Analysis: Summary Statistics

Moreover a general summary statistics has been computed to better have a general idea of the returns pattern that characterize these time series, along with the results of the ADF test, necessary in order to use the GARCH modeling

	BT 5min	KK 5min	BT hour	KK hour	BT day	KK day
Mean	8,1e-07	3,1e-08	3,1e-06	4,9e-08	2,6e-06	-1,6e-06
St.Dev.	1,2e-03	2,5e-04	2,3e-03	3,1e-04	2,3e-03	3,0e-04
Min	-0,1	-0,1	-0,0047	-0,0042	-0,0098	-0,0042
Max	0,108	0,009	0,05	0,0042	0,011	0,0008

Table 2.2: Statistics of different frequencies of the two exchanges

Results show two different pictures of the returns of the two exchanges. Indeed the standard deviation is one order of magnitude smaller in the case of the Kraken exchange in contrast to the Bittrex one. Also, looking at the maximum values, there are some differences: in the case of the Kraken is one-tenth the Bittrex value.

In order to proceed with the fitting of GARCH model and the Causality in Quantiles test, returns must be stationary, thus the classical Augmented Dickey–Fuller test has been implemented.
Starting from the regression model to predict r_t based on

$$r_t = c + \beta t + \alpha r_{t-1} \phi_1 \Delta r_{t-1} + ... + \phi_p \Delta r_{t-p} + \epsilon_t$$

where r_{t-1} is the lag 1 of the return time series, while Δr_{t-1} is the first difference of the series at time $t-1$. If the slope coefficient β is not significantly different from one, than we cannot reject the null hypothesis that the series is non stationary. However, if β is significantly less than one, then we can reject the null hypothesis:

$$H_0 : \beta = 0$$
$$H_1 : \beta < 0$$

ADF Test Results						
	BT 5min	KK 5min	BT hour	KK hour	BT day	KK day
Test Statistics	-42	-37	-16	-18	-7,2	-6,6
P-Value	0,0	0,0	0,0	0,0	0,0	0,0

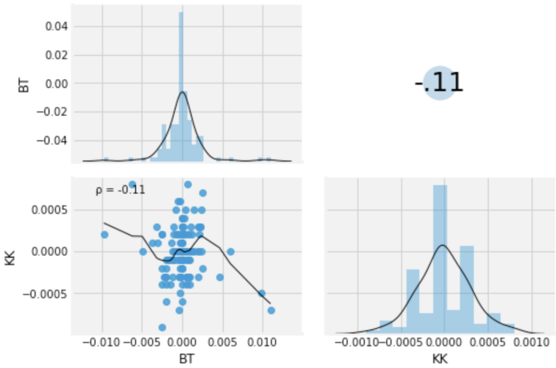
Table 2.3: If p-val less than 0.05, null hypothesis of non-sationarity rejected

Results in Table 4.1 show high confidence in stationarity of the returns.

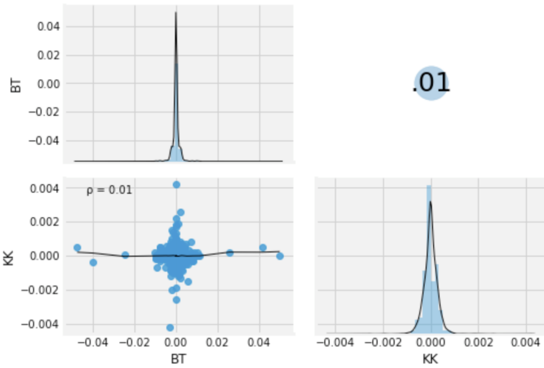
Returns analysis: Correlation analysis

Finally, regarding the USD Coin, a correlation analysis is performed to understand the price's behavior in the exchanges to check if the coin can suffer from an arbitrage opportunity. The correlation quantity computed is the Pearson's Correlation Value for time series.

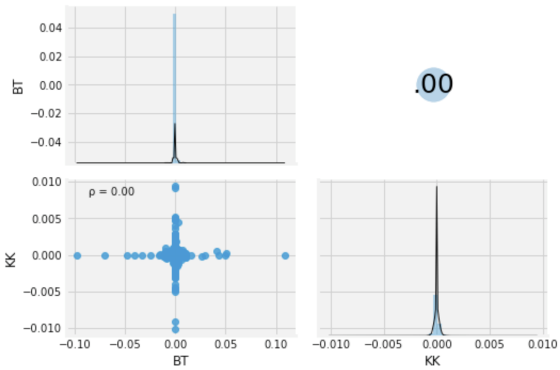
In Figure 2.8, on the diagonal are represented the distributions of the returns, while on the antidiagonal, respectively the Pearson's Correlation value and the returns of both exchanges. Results shows a strange non-correlation between exchanges in the returns. This represents a worrying situation in which an arbitrage opportunity can be exploited, making suffer the credibility of the exchanges and coins.



(a) 5min Level



(b) Hour Level



(c) Day Level

Figure 2.8: Correlation at three different levels

2.2 DAI

Data regarding the DAI coin comes from two market exchanges, Coinbase and Kraken. Unfortunately, due to the outstanding popularity of the Coinbase platform, traded volumes belong to two different orders of magnitude. Indeed before the downsampling, in the period considered, the former registered about $4,5M$ transactions while the latter less than $500K$. This difference is evident also in the daily traded volume, which is represented below:

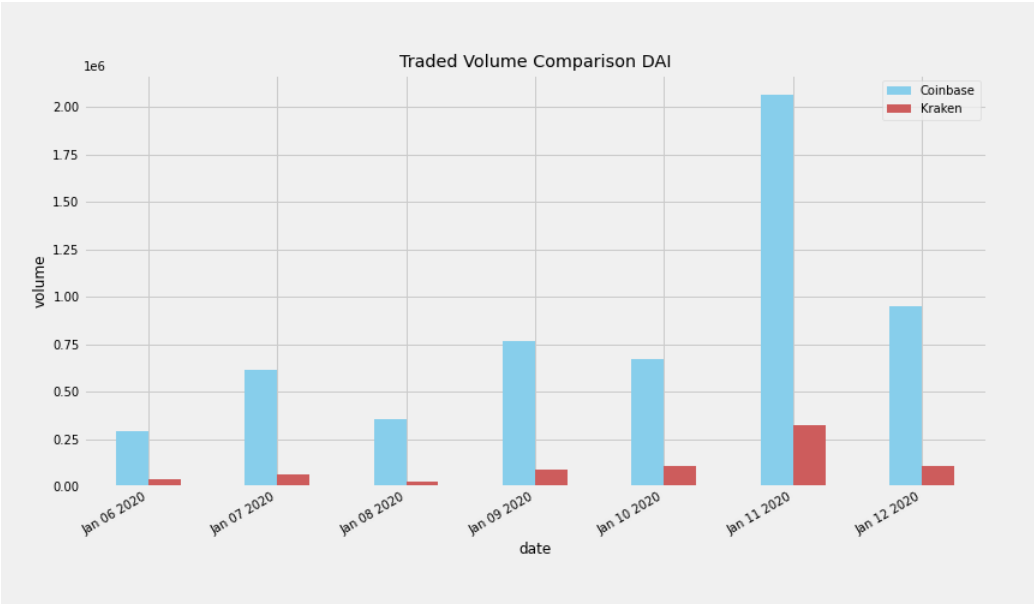


Figure 2.9: Traded volume of DAI coin in the range between 06 June and 12 June

Price Analysis

As for the USD Coin, the same type of preliminary analysis has been conducted on the DAI price and returns time series. The following figure represents the price series sampled at 5 Minutes along with the moving mean sampled at hour and day level.

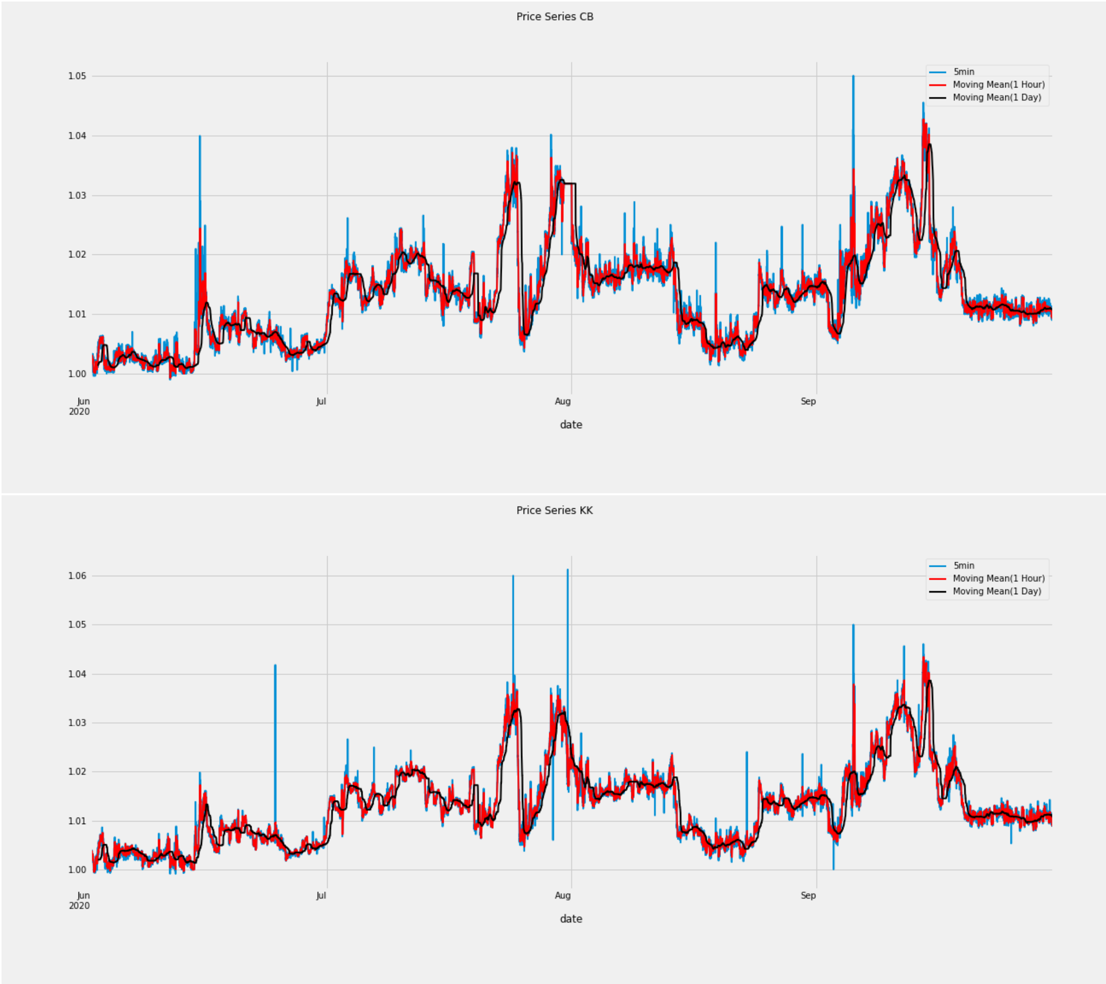


Figure 2.10: 5-min Price dynamics and hourly and daily rolling window on Coinbase(CB) and Kraken(KK)

The price dynamics of the DAI differ entirely from the USD Coin. There is a high level of instability of the price with multiple spikes. Identifying a possible Pump and Dump event on these kinds of Stablecoin can be extremely difficult using the simple moving average methods. Indeed multiple positive signals will come out from the detection analysis, yet no info can be retrieved to understand if the signal corresponds, actually, to a P&D. However, in contrast with the USD Coin, the two exchanges shows a high level of correlation given the price follows the same dynamics on the two exchanges.

Price Analysis: Additive Decomposition

Subsequently, the addtive decomposition has been computed, eliminating a possible disturbance in the stability and detection computations as for USDC.

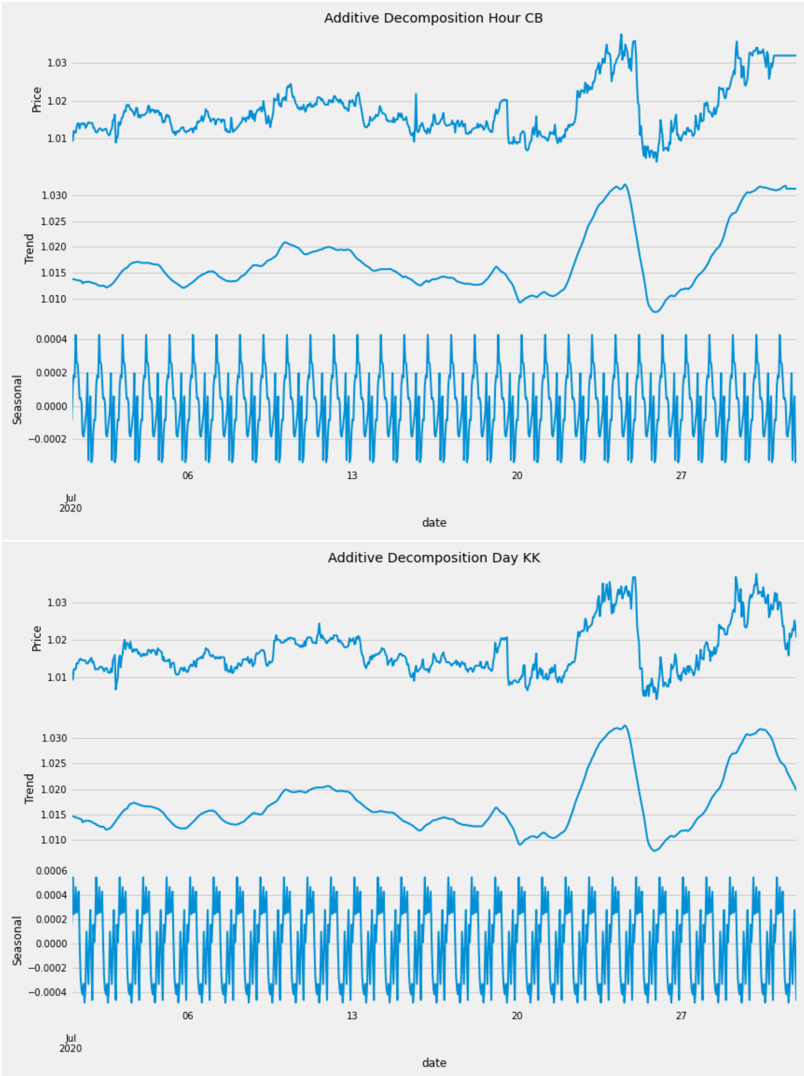


Figure 2.11: Additive Decomposition at hour level of the price on Coinbase(CB) and Kraken(KK)

Figure 2.11 shows that the trend has not the stability desired, while the seasonality decomposition can be avoided. Indeed, the quantity ranges between

−0,0002$ and +0,0004$ for Coinbase and between −0,0004$ and +0,0006$ for Kraken, thus a negligible quantity. Also, the frequency poses some difficulties in understanding the meaning.

Returns

Computations on returns are the same as for the USD Coin, including the choices of the log returns. The following figures represent the return dynamics along-side the volume and number of transactions, which occurred in 5 minutes intervals.

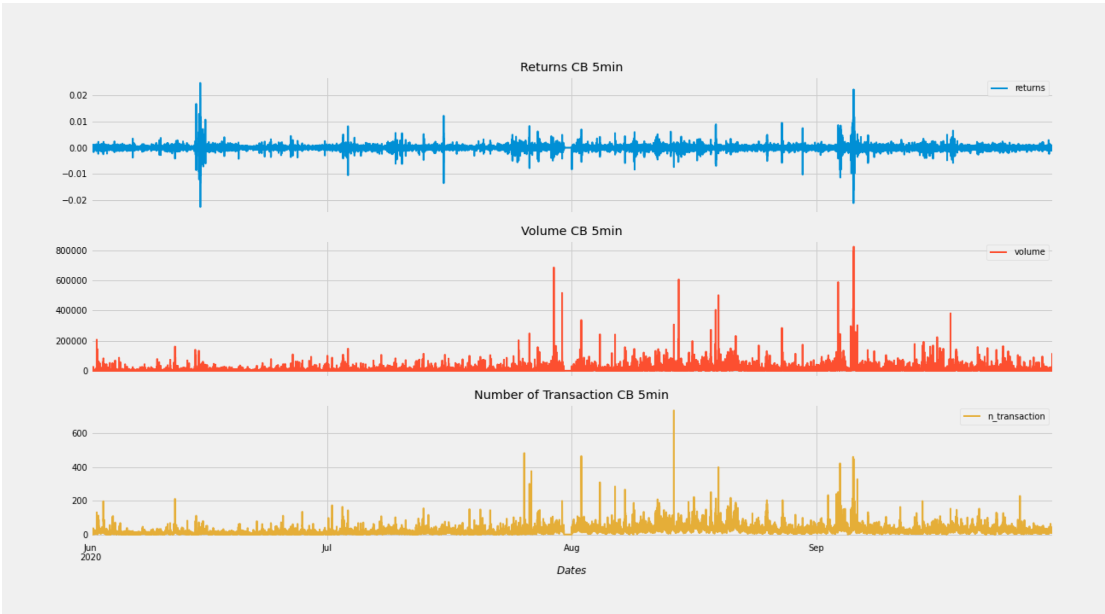


Figure 2.12: Returns, Volume and Number of Transaction dynamics on Coinbase(CB)

Regarding Coinbase returns, the values present some spikes, which occurred in the exact moment of volume and number of transactions. The probability of a P&D event with similar dynamics is high.

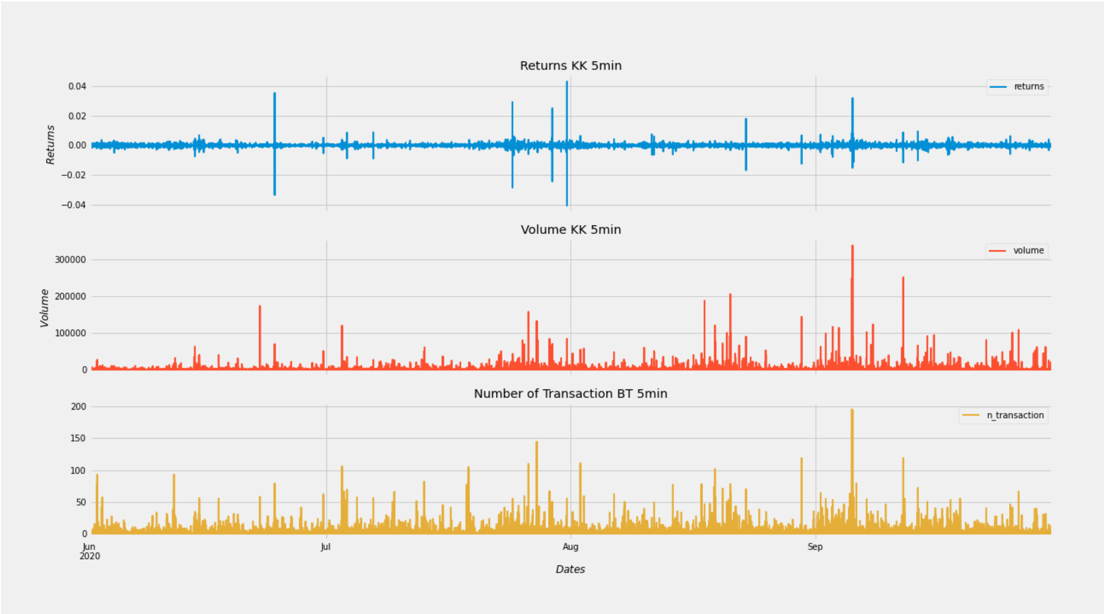


Figure 2.13: Returns, Volume and Number of Transaction on Kraken(KK)

In the Kraken exchange, we have higher instability with returns that hit 4%. Moreover, two important spikes belong to the same time interval of a volume and several transactions spikes. Also, in this case, the probability of similar dynamics to the P&D is high, but it must be considered that one spike occurred at the exact moment of the Coinbase exchange. Thus it is more probable that it was a moment of high profitability of the DAI rather than a Pumping event.

Returns: Boxplot

As for the USDC, also for DAI, according to the methodology introduced, the boxplots have been computed with colored bar representing the Interquantile Range IQR while rhombuses representing the outliers:
In Figure 2.14 the boxplots are computed on the 5min distributions of the returns for each day. The IQR for DAI is a little bit higher than the USD Coin, reaching $0,5\%$, with more consistency between the two exchanges.

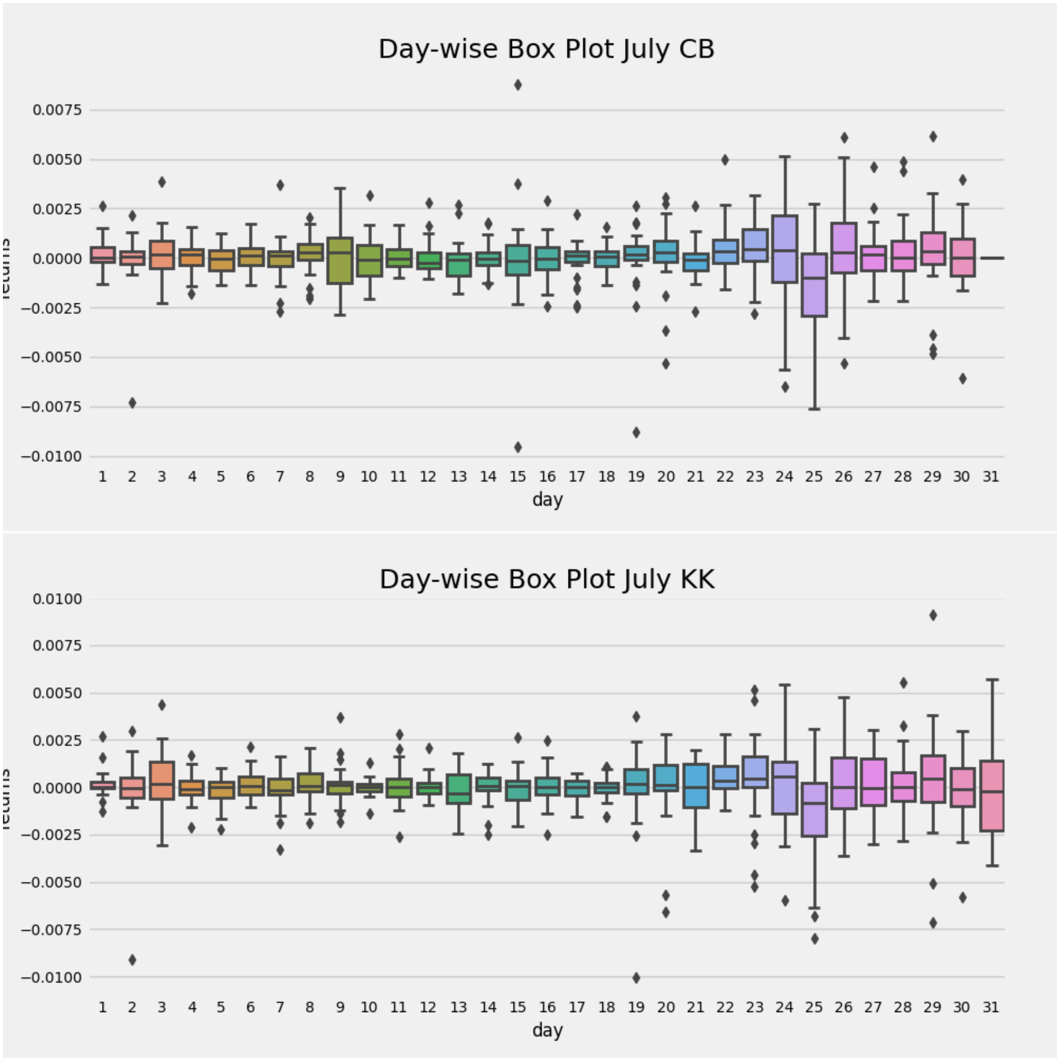


Figure 2.14: Boxplots of the Daily Price on Coinbase(CB) and Kraken(KK)

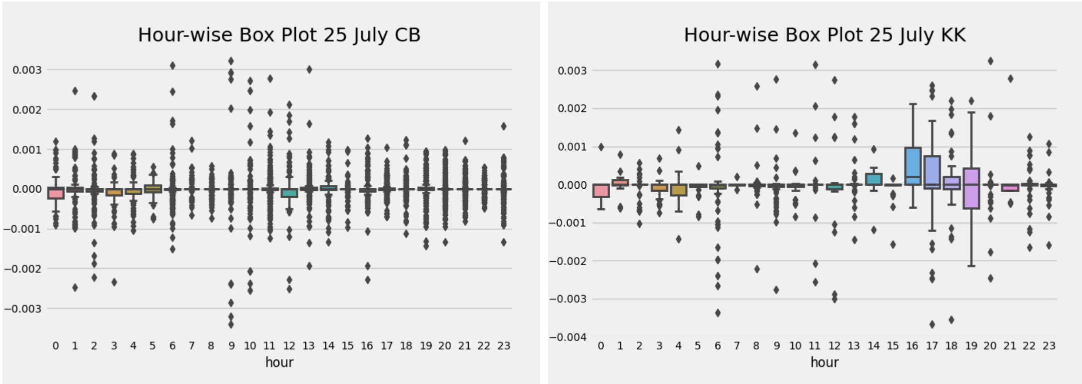


Figure 2.15: Boxplots of the Hourly Price on Coinbase(CB) and Kraken(KK)

In Figure 2.15 boxplots are computed on the 5min distribution of returns for each hour. The intra-day boxplots highlights a less stable price, with large amount of outliers.

Ruturns Analysis: Summary Statistics

Moreover, a general summary statistics has been computed to have a better general idea of the returns pattern that characterizes these time series, along with the results of the ADF test, necessary to use the GARCH modeling.

	CB 5min	KK 5min	CB hour	KK hour	CB day	KK day
Mean	6,2e-07	5,8e-07	4,0e-06	4,0e-06	6,2e-05	6,1e-05
St.Dev.	8,4e-04	7,8e-04	1,6e-03	1,5e-03	5,2e-03	5,3e-04
Min	-0,022	-0,041	-0,026	-0,022	-0,031	-0,031
Max	0,024	0,043	0,028	0,012	0,014	0,015

Table 2.4: Statistics of different frequencies of the two exchanges

In contrast with the USD Coin results, the DAI returns show a higher correlation between the two exchanges, with standard deviations and the same order as for the mean. While talking about the maximum value, the value is considerably higher for the Kraken exchange. The same can also be said for the minimum value.

ADF Test Results						
	BT 5min	KK 5min	BT hour	KK hour	BT day	KK day
Test Statistics	-29	-51	-28	-25	-8,3	-10
P-Value	0,0	0,0	0,0	0,0	0,0	0,0

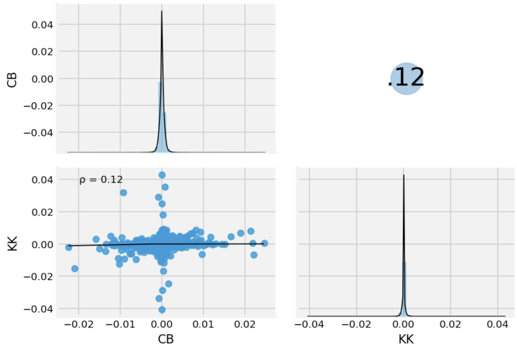
Table 2.5: If p-val less than 0.05, null hypothesis of non-sationarity rejected

Table 4.3 shows high confidence in stationarity of the returns, thus no more manipulation of the data will be needed.

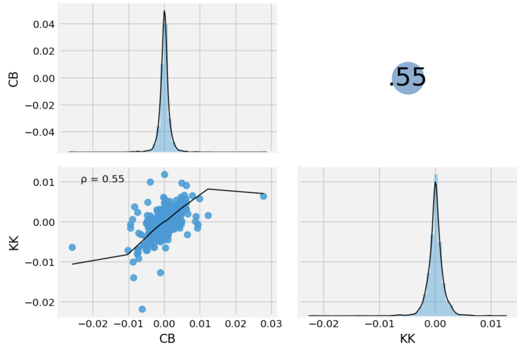
Returns analysis: Correlation analysis

Finally, the correlation analysis is performed to understand the price behavior in the exchanges and check if the coin can suffer from an arbitrage opportunity. The correlation quantity computed is the Pearsons Correlation Value for time series as for the USD Coin.

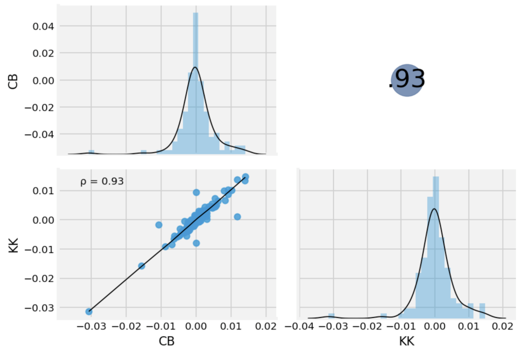
As for the USDC results, in Figure 2.16, on the diagonal are represented the distributions of the returns, while on the antidiagonal, respectively the Pearson's Correlation value and the returns of both exchanges. As expected, Figure 2.16 highlights a more realistic scenario than the one depicted for USD Coin, where the intra-hour correlation is low but increases with the frequency, reaching the 0.93 level. Thus an arbitrage opportunity is highly improbable.



(a) 5min Level



(b) Hour Level



(c) Day Level

Figure 2.16: Correlation at three different levels

2.3 Bitcoin

Part of the stability analysis is based on comparing the bitcoin returns dynamics, which is considered highly volatile, with the stablecoins ones. Thus a time series of the price has been retrieved, unfortunately only at an hour and day level, from the CoinMarketCap.com platform covering the same period of the DAI and USD Coin. The retrieved price is the BTC/USD trading pair. However, the platform from which time-series are retrieved is not an exchange platform. Thus the price of any crypto asset is a volume-weighted average of market pair prices for the crypto asset. The higher percentage of volume contributed from the pair, the more influence it has on the average price. So this situation will not provide the best consistency in results computation. In particular, this kind of computation should lower the volatility of the BTC price, not enough, however, to be defined as stable.

Price Analysis

In the case of BTC, it has been not possible to retrieve more granular data, so the figure 2.17 represents the price dynamics at hour level and the daily rolling window:

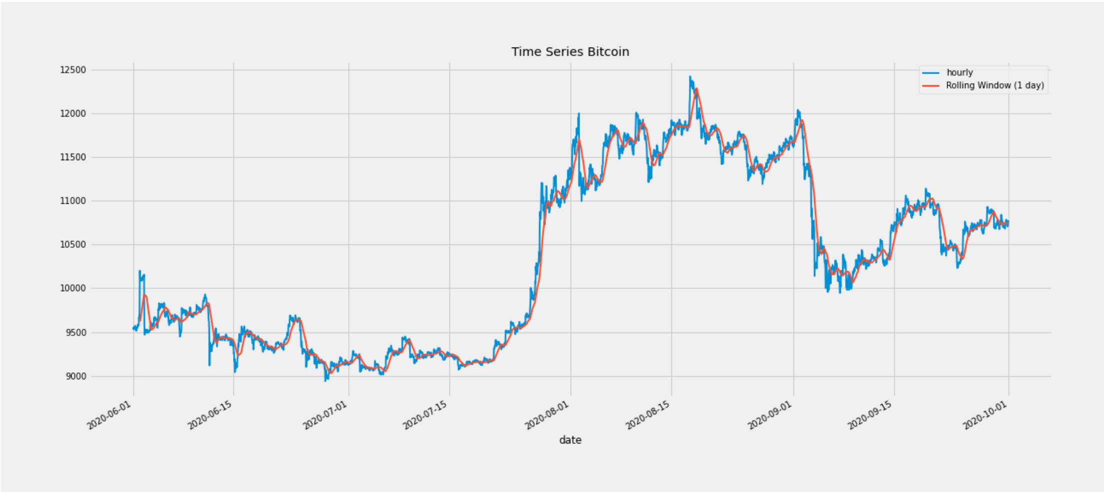


Figure 2.17: Hour Price dynamics and daily rolling window

There are some consistent spikes, which will be higlighted by the returns computations.

Returns

The computations are the same as the previous coins. Following figures represents the returns dynamics along the day-wise bloxplot in the month of higher instability, August:

Bitcoin returns offer a completely different picture, with values consistently higher and an IQR generally equal to 3%, one order of magnitude higher than stablecoins. The source of the BTC data probably has made the dynamics less noisy but not less volatile.

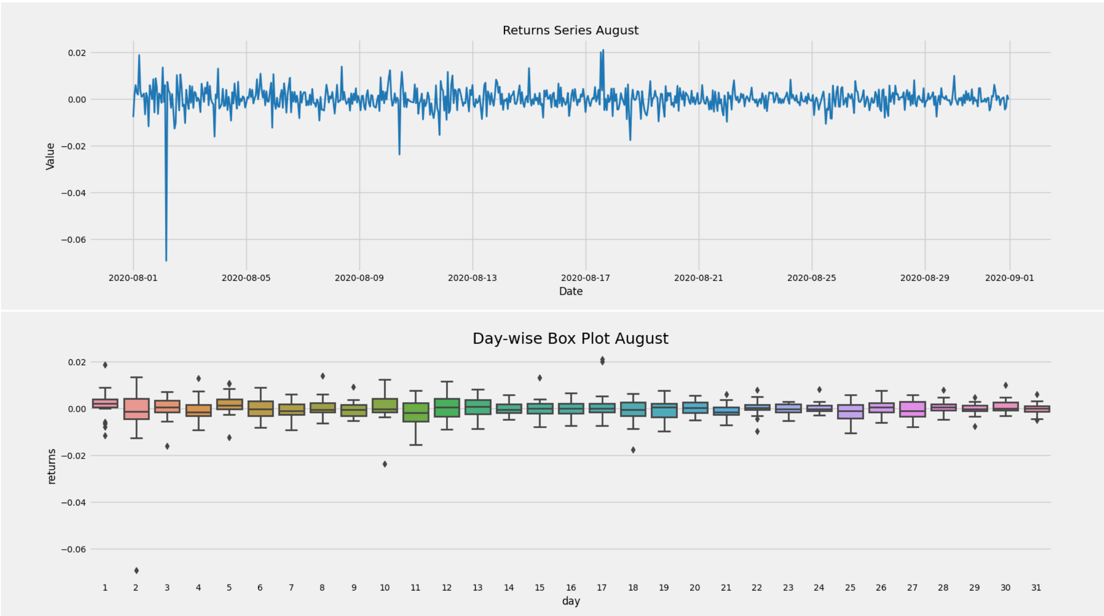


Figure 2.18: Returns dynamics and Day-wise Boxplot

Chapter 3

Stability Analysis

3.1 Methodology

In this section, according to Hoang and Baur's [14] work, a stability analysis will be performed to understand if the stablecoin can compete as a possible alternative as digital currency or if the anomaly detection process that follows will suffer from unexpected high volatility. To be consistent, the stability analysis is divided into Absolute, based on computation on unconditional volatility, and in Relative, based on computation on conditional volatility.

3.1.1 Absolute Stability

The Absolute stability will be determined using a simple χ^2-test on standard deviation. A perfect stable coin should show no variation, which means zero standard deviation. However, a zero variance does not exist in reality so that the following hypothesis will characterize the test:

$$H_0: \quad \sigma_{sc} < \sigma_0 \tag{3.1}$$

$$H_1: \quad \sigma_{sc} \geq \sigma_0 \tag{3.2}$$

where σ_{sc} is the standard deviation of the stablcoin considered and the $\sigma_0 =$ 0.1%, so that σ_{sc} can vary a little bit. The Test Statistics on which the critical region is computed is:

$$T = \frac{(N-1) * \sigma_{sc}}{\sigma_0} \overset{H_0}{\sim} \chi^2(N-1)$$

3.1.2 Relative Stability

The relative stability is more complex to be determined. The analysis starts with a statistical test on the variance ratio between the stablecoin selected and the bitcoin, which is highly volatile. The hypothesis of the test are:

$$H_0: \quad \frac{\sigma_{sc}}{\sigma_{btc}} \geq 1 \tag{3.3}$$

$$H_1: \quad \frac{\sigma_{sc}}{\sigma_{btc}} < 1 \tag{3.4}$$

where the test statistics is $T = \frac{\sigma_{sc}}{\sigma_b tc} \overset{H_0}{\sim} F(N-1, N-1)$. A non powerful test in this case, due to the non-gaussanity of the returns of stablecoin. Thus the relative stability analysis proceeds to compare the conditional volatility of the Bitcoin with the stablecoin ones, computing the Pearson's Correlation Coefficient. To compute the conditional volatility is necessary to check whether the time series shows signs of heteroscedasticity and then apply the version of the GARCH model selected.

GJR-GARCH(1,1)

The generalized autoregressive conditionally heteroscedastic or GARCH model main goal is to model the strong dependance of sudden burst of variability in series own past. The model is as follow:

$$r_t = \delta r_{t-1} + \epsilon_t \qquad with \quad \epsilon_t | \phi_t \sim N(0, \sigma_t^2) \tag{3.5}$$

$$\sigma_t^2 = \omega_0 + \alpha \epsilon_{t-1} + \beta \sigma_{t-1}^2 \tag{3.6}$$

where ϕ_t is the information set on which the residual of the AR(1) process is conditioned. The stationarity conditions tested before are necessary in order to ensure that the moments of the normal distribution are finite.
So the GARCH model captures volatility clustering. The volatility is more likely to be high at time t if it was also high at time $t1$. Another way of seeing this is noting that a shock at time $t1$ also impacts the variance at time t. However, as proved in Chu et al.[9], a better version of the GARCH, in modeling financial returns, is the GJR-GARCH version. The GJR-GARCH model was introduced for capturing asymmetries in volatility, distinguishing positive and negative

parts of the innovation process. Starting from the usual GARCH model, the GJR-GARCH includes an ulterior parameter γ:

$$r_t = \delta r_{t-1} + \epsilon_t \qquad with \quad \epsilon_t|\phi_t \sim N(0, \sigma_t^2) \tag{3.7}$$

$$\sigma_t^2 = \omega_0 + \alpha\epsilon_{t-1}^2 + \gamma\epsilon_{t-1}^2 1\{\epsilon_{t-1} < 0\} + \beta\sigma_{t-1}^2 \tag{3.8}$$

Thus the GARCH model is a restricted version of the GJR-GARCH, with $\gamma = 0$

3.2 Results

All the computation performed are based on the time series sampled at hour level, as no 5Min time series could be retrieved for the Bitcoin price.
Absolute Stability The results of the test are, as supposed in the **Data** section, not promising.

Chi-Square Test Results

	DAI CB	DAI KK	USDC BT	USDC KK
P-Value	0,0	0,0	0,0	0,0

Table 3.1: If p-value less than 0.05, null hypothesis of $\sigma_{sc} < \sigma_0$ rejected

The standard deviation is much higher than the $0,1\%$.

Relative Stability

The results of the Fisher test, instead, looks better, given the standard deviation is compared to the one of Bitcoin which is consistently higher.

Fisher Test Results ($\sigma_{sc} < \sigma_0$)

	DAI CB	DAI KK	USDC BT	USDC KK
P-Value	0,0	0,0	0,0	0,0

Table 3.2: If p-value less than 0.05, null hypothesis of $\sigma_{sc} < \sigma_0$) rejected

However, as introduced before, unconditional variance cannot be used to determine the stability of a coin, in particular, if it seems not to be present gaussianity in the returns.

So to implement the GJR-GARCH model, a preliminary test on the heteroscedasticity of the residuals is necessary. The test selected is the Breusch-Pagan, which compares the residual variance computed from the AR(1) fitting to the return series and the variance of the 1-lagged returns that correspond to the exogenous variables in the fitted model.

Breusch-Pagan Test Results

	DAI CB	DAI KK	USDC BT	USDC KK	BTC
P-Value	0,0	0,0	0,0	0,0	0,0

Table 3.3: If p-value less than 0.05, null hypothesis of $\sigma_{sc} > \sigma_{btc}$ rejected

The results confirm the presence of heteroscedasticity, the GJR-GARCH model can be implemented.

AR(1) - GJR-GARCH(1,1) Model Results DAI CB

	Coefficients	Std. Err.	P-value(t-stat)
Constant	6,47e-03	2,46e-03	8,4e-03
r_{t-1}	-0,161	2,69e-02	2,4e-09
ω_0	1,70e-03	6,93e-04	1,42e-02
α	0,373	0,105	3,98e-04
γ	-0,323	0,101	1,35e-03
β	0,749	5,88e-02	3,69e-37

Table 3.4: Model Results for DAI Coinbase

AR(1) - GJR-GARCH(1,1) Model Results DAI KK

	Coefficients	Std. Err.	P-value(t-stat)
Constant	4,43e-03	2,18e-03	4,29e-02
r_{t-1}	-0,221	2,43e-02	1,14e-19
ω_0	8,20e-04	3,19e-04	1,03e-02
α	0,226	4,82e-02	2,82e-06
γ	-0,168	4,05e-02	3,44e-05
β	0,833	4,35e-02	1,81e-81

Table 3.5: Model Results for DAI Kraken

For the DAI, the GJR-GARCH seems to be a good choice due to the low P-Value of all the coefficients. The main goal of this section is to understand the correlation of the bitcoin volatility with one of the other stable coins.
Regarding the USD Coin, the GJR-GARCH could not produce acceptable results, given that the γ coefficient was not significant. A basic GARCH model has been implemented instead. The same happened for Bitcoin, in contrast with the results of Peng et al[20]. In figure 4.2 the results of the GARCH model. Coefficients seem significant enough except for the ω one.

AR(1) - GARCH(1,1) Model Results USDC BT

	Coefficients	Std. Err.	P-value(t-stat)
Constant	5,14e-03	2,61e-03	4,90e-02
r_{t-1}	-0,309	3,36e-02	3,75e-20
ω_0	1,27e-03	7,94e-04	0,110
α	0,047	1,95e-02	1,74e-02
β	0,932	1,66e-02	0,0

Table 3.6: Model Results for USDC Bittrex

AR(1) - GARCH(1,1) Model Results USDC KK

	Coefficients	Std. Err.	P-value(t-stat)
Constant	1,93e-04	4,69e-04	0,681
r_{t-1}	-0,442	1,88e-02	1,76e-12
ω_0	8,02e-05	5,0e-05	0,109
α	0,1	2,0e-02	6,08e-07
β	0,8	8,57e-02	1,05e-20

Table 3.7: Model Results for USDC Kraken

AR(1) - GARCH(1,1) Model Results BTC

	Coefficients	Std. Err.	P-value(t-stat)
Constant	5,71e-03	7,48e-03	0,445
r_{t-1}	-0,055	2,66e-02	4,04e-02
ω_0	6,13e-03	4,37e-03	0,161
α	0,135	7,84e-02	8,61e-02
β	-0,859	7,17e-02	3,54e-33

Table 3.8: Model Results for BTC

The next step is to compute the conditional volatility of the model and compare values with bitcoin volatility pattern.

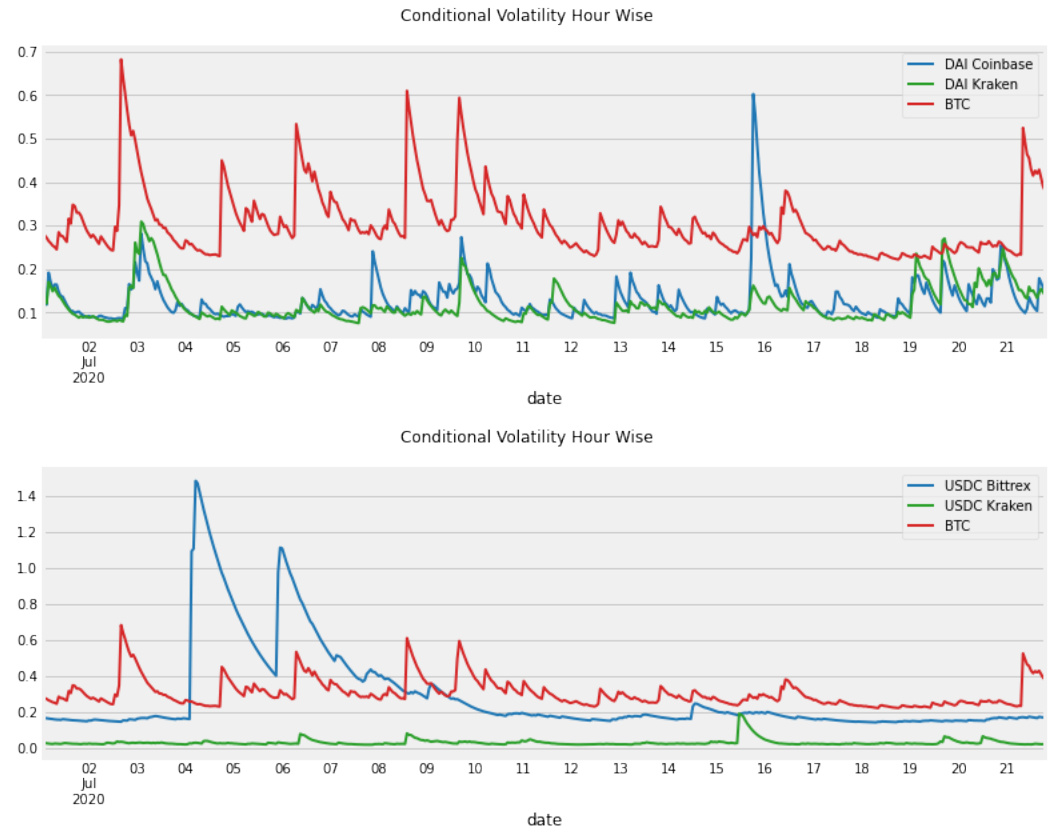


Figure 3.1: Conditional Volatility in period 01/07-01/08

In general, July has been a volatile month for all coins; thus, in the case of USDC on Bittrex, the conditional volatility has exceeded bitcoin values. In particular, this spike in volatility corresponds to the one seen in price in the **Data** section and has been carried for almost two days. The period is probably shorter due to low results produced by the GARCH model, yet it reached a compromising level.

For the DAI case instead, the volatility remains under the Bitcoin values. However, it follows a similar pattern with a notable amount of co-occurrencies of the spikes. Thus the correlation analysis should provide a high level of serial correlation.

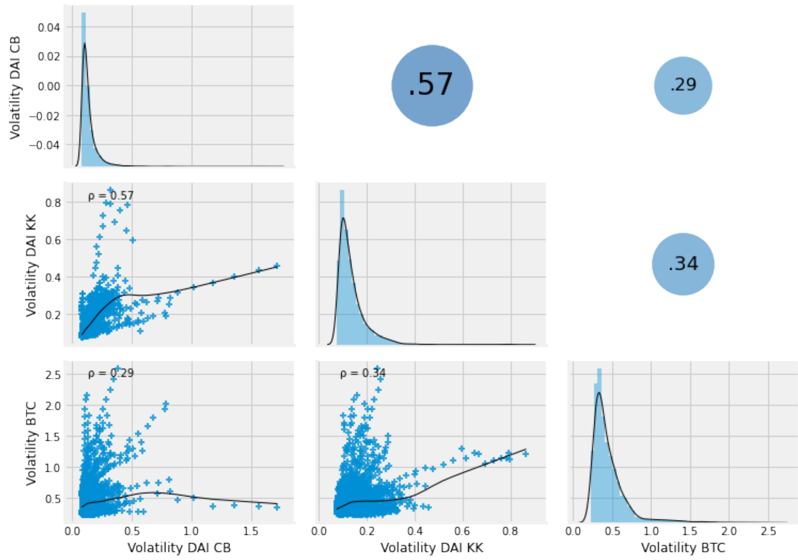


Figure 3.2: Correlation between DAI and BTC

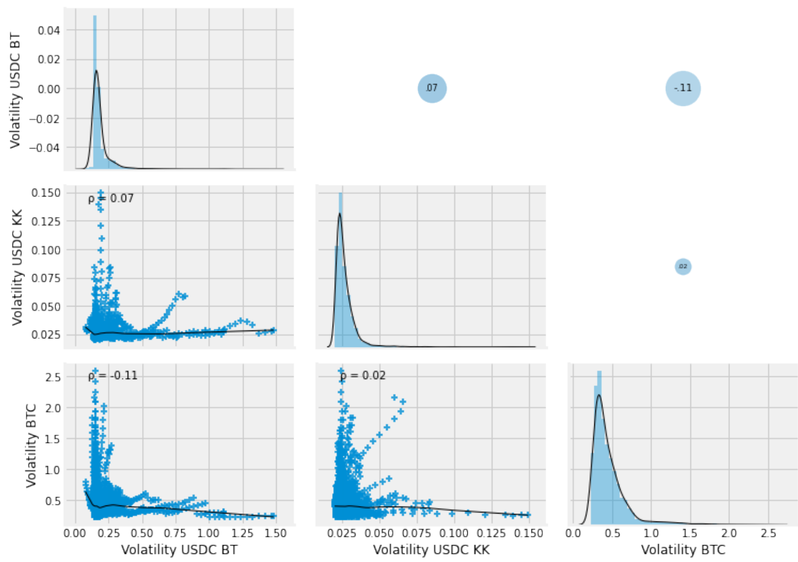


Figure 3.3: Correlation between USDC and BTC

The correlations plot shows the high level of correlation between DAI and BTC conditional volatility. Thus the stablecoin is much more stable than Bitcoin. At the same time, the USDC is much less correlated and more stable than the DAI coin. This implies that tokenized funds result in being the best tool in stabilizing the volatility of the crypto asset. Indeed the first coin by market capitalization is the USD Tether, the first tokenized funds stablecoin. In conclusion, the two stablecoin are more stable than the bitcoin but remain however, in general volatile. This can lead to detection of anomaly deriving from a volatile period rather than an actual Pump and Dump scheme taking place. So the Causality in Quantiles test comes in handy to better understand the dynamics in the most extreme values, searching for the Granger Causality in Volume and Number of Transaction.

Chapter 4

Pump and Dump

This section will apply the methods introduced by Kamps et al.[17] and by Jeong et al.[16] to understand if and how these stablecoins be affected by the P&D event. The first phase is dedicated to the detection of any anomaly in the price, volume and number of transactions time series using the moving average method. The goal is to identify the major spikes in price, comparing them to the average of the previous values, and to search for a co-occurance of the spike also in the volume and the number of transaction. If it is present, then, according to the breakout indicators introduced by Kamps et al.[17], they will be classified as a possible P&D or not. Moreover it will be implemented the Causality in Quantiles test, to search for Granger causality in price by volume and number of transaction, in the region of higher quantile. The results will explain if the volume or the number of transaction can affect the price, as would happen in a trade-based manipulation event.

4.1 Methodology: Anomaly Detection

What characterizes a P&D event is the pumping and dumping phase. In this case, the manipulation is trade-based. Thus the pumping can only occur if numerous purchase transactions happen in a relatively small amount of time. So a P&D event will exhibit, on a given time interval, volume, number of transactions, and price anomaly. An anomaly can be identified as a non-conforming value to the rest of the set, thus an outlier. The identification of the outlier

can be supervised, relying on a training dataset or unsupervised, which relies on the assumption that anomalies are a rare occurrence in the data to avoid false-positive signals. This project will focus on unsupervised detection, as no training dataset could be retrieved, most notably on detecting anomalies concerning recent history.
The anomaly detection technique is a thresholding technique. Thus for each 5-minutes value of the price, a simple moving average is computed, taking the average of previous values in a time window preselected, defined as the lag factor. The idea is to compare the past trend with the value selected to check if it conforms to that trend or can be classified as an outlier.

Price anomaly

Following the work of Kamps et al.[17] the anomaly will be detected if:

$$Price\_Anomaly_t = \begin{cases} True, & p_t^{high} > \epsilon * \mu_\gamma(p_t) \\ False, & p_t^{high} \leq \epsilon * \mu_\gamma(p_t) \end{cases}$$

where p_t^{high} is the highest price registered in the time interval t selected, ϵ is the percentage increase factor, γ is the lag factor and $\mu_\gamma(p_t)$ is the moving average computed as $\mu_\gamma(p_t) = \frac{\sum_{i=t-\gamma}^{t} p_t^{close}}{\gamma}$.
So if the price observed p_{high} is bigger than the ϵ percentage of the $\mu_\gamma(p_t)$ value computed over the γ past values of p_t^{high}, a price anomaly is detected.

Volume and Number of Transaction Anomaly

In the same way as for the price anomaly, the volume anomaly is detected if:

$$Volume\_Anomaly_t = \begin{cases} True, & v_t > \epsilon * \mu_\gamma(v_t) \\ False, & v_t \leq \epsilon * \mu_\gamma(v_t) \end{cases}$$

$$Num\_of\_Trans\_Anomaly_t = \begin{cases} True, & q_t > \epsilon * \mu_\gamma(q_t) \\ False, & q_t \leq \epsilon * \mu_\gamma(q_t) \end{cases}$$

where v_t and q_t is the total traded volume and total of transactions registered in the time interval t selected, ϵ is the percentage increase factor, γ is the lag

factor and $\mu_\gamma(v_t)$ is the moving average computed as $\mu_\gamma(v_t) = \frac{\sum_{i=t-\gamma}^{t} v_t}{\gamma}$. Of course for the q_t variable all the computations are the same.
So if the volume traded v_t is bigger than the ϵ percentage of the $\mu_\gamma(x_t)$ value computed over the γ past values of v_t, a volume anomaly is detected.

Pump-Connected Anomaly

The idea is to detect not only an anomaly on these three-time series but, most notably, a co-occurrence of the anomaly and an absence of the triple anomaly on other exchanges considered. Only in this situation, a P&D can be detected. Unfortunately, this method suffers incredibly from the parameters choices. Thus a selection phase is necessary to understand the best values to implement.

4.2 Methodology: Causality-in-Quantiles Test

In order to better classify the detected events in the previous phase, a causality in quantiles test is necessary. The Causality in Quantiles test is a test introduced by Jeong et al.[16] capable of capturing the Granger causality in a non-linear time series. Moreover, the causality can be exploited at different quantiles so that a more precise picture of the events taking place is obtained. The test is a non-parametric test based on the kernel method. This project will be used to test the significance of the regressors Volume and Number of Transactions. Given two time series y_t and w_t, the Granger Causality in quantile is defined as follow:

- w_t does not cause y_t in the θ-quantile with respect to $\{y_{t-1}, ..., y_{t-p}, w_{t-1}, ..., w_{t-p}\}$ if :

$$Q_\theta(y_t | y_{t-1}, ..., y_{t-p}, w_{t-1}, ..., w_{t-p}) = Q_\theta(y_t | y_{t-1}, ..., y_{t-p})$$

- w_t does cause y_t in the θ-quantile with respect to $\{y_{t-1}, ..., y_{t-p}, w_{t-1}, ..., w_{t-p}$ if :

$$Q_\theta(y_t | y_{t-1}, ..., y_{t-p}, w_{t-1}, ..., w_{t-p}) \neq Q_\theta(y_t | y_{t-1}, ..., y_{t-p})$$

where $Q_\theta(y_t|\cdot)$ is the θ^{th} conditional quantile of y_t given $\cdot$ which depends on t.
So the reporting the Jeong et al.[16] procedure, denote $x_t \equiv (y_{t-1},...,y_{t-p})$,
$z_t = (y_{t-1},...,y_{t-p},w_{t-1},...,w_{t-p})$ and the conditional distribution function y_t given $z_t(x_t)$ by $F_{y_t|z_t}(y_t|z_t)(F_{y_t|z_t}(y_t|x_t))$. According to Jeong et al.[16] $F_{y_t|z_t}(y_t|z_t)$ is absolutely continous in y for almost all (z_t, x_t).
So the hypothesis of the test can be defined as:

$$H_0 : P\{F_{y_t|z_t}(Q_\theta(x_t)|z_t) = \theta\} = 1 \quad a.s.$$
$$H_1 : P\{F_{y_t|z_t}(Q_\theta(x_t)|z_t) = \theta\} < 1 \quad a.s.$$

The null hypothesis can reformulated as $E[1\{y_t \leq Q_\theta(x_t)|z_t\}] = \theta$ or $1\{y_t \leq Q_\theta(x_t)|z_t\} = \theta+\epsilon_t$ where $E[\epsilon_t|z_t] = 0$ and $1(\cdot)$ is the indicator function. To implement the test, Jeong et al.[16] introduced the following distance measure which will be considered the test statistics:

$$J = E[\{F_{y_t|z_t}(Q_\theta(x_t)|z_t) - \theta\}^2 f_{z_t}(z_t)]$$

where $f_{z_t}(z_t)$ is the marginal density function of z_t. The formula is however impossible to implement, so starting from the assumption that the expected value $E[\epsilon_t|z_t] = F_{y_t|z_t}\{Q_\theta(x_t)|z_t\} - \theta$ the distance measure is:

$$J = E\{\epsilon_t E[\epsilon_t|z_t] f_{z_t}(z_t)\}$$

This implies that the estimation of the expected value of residuals of the indicator function can be computed using kernel methods. The estimator of the residuals are:

$$\hat{E}[\epsilon_t|z_t]\hat{f}_{z_t}(z_t) = \frac{1}{(T-1)h^m}\sum_{s\neq t}^{T} K_{ts}\epsilon_s$$

where $K_{ts} = K\{(z_t - z_s)/h\}$ is the Kernel function and h is the bandwidth.
For the θ-conditional quantile, as before, the estimator is computed using the non-parametric kernel method.

$$\hat{Q}_\theta(x_t) = \hat{F}^{-1}_{y_t|x_t}(\theta|x_t) \qquad where \quad \hat{F}_{y_t|x_t}(y_t|x_t) = \frac{\sum_{s\neq t} L_{ts} 1(y_s \leq y_t)}{\sum_{s\neq t} L_{ts}}$$

$\hat{F}_{y_t|x_t}(y_t|x_t)$ is the Nadaraya-Watson kernel estimator of $F_{y_t|x_t}(y_t|x_t)$ with $L_{ts} = L\{(x_t - x_s)/a\}$ the kernel function and a the bandwidth associated.
Finally Jeong et al.[16] presents the estimator to compute:

$$\hat{J}_t = \frac{1}{T(T-1)h^m} \sum_{t=1}^{T} \sum_{s \neq t}^{T} K_{ts} \epsilon_t \epsilon_s \tag{4.1}$$

$$= \frac{1}{T(T-1)h^m} \sum_{t=1}^{T} \sum_{s \neq t}^{T} K_{ts}[1\{y_t \leq \hat{Q}_\theta(x_t)\} - \theta][1\{y_s \leq \hat{Q}_\theta(x_s)\} - \theta] \tag{4.2}$$

To define the critical region, Jeong et al.[16] introduced the asymptotic distribution of the test statistics under the null hypothesis:

$$Th^{m/2}\hat{J}_t \xrightarrow{L} N(0, \sigma_0^2)$$

where the variance can be estimated as $\hat{\sigma}_0^2 = 2\theta^2(1-\theta)^2/(T(T-1)h^m)\sum_{s \neq t}^{T} K_{ts}^2$ finally obtaining the estimator:

$$Th^{m/2}\hat{J}_t/\hat{\sigma}_0^2 = \tag{4.3}$$

$$= \sqrt{\frac{T}{T-1}} \frac{\sum_{t=1}^{T} \sum_{s \neq t}^{T} K_{ts}[1\{y_t \leq \hat{Q}_\theta(x_t)\} - \theta][1\{y_s \leq \hat{Q}_\theta(x_s)\} - \theta]}{\sqrt{2}\theta(1-\theta)\sqrt{\sum_{s \neq t}^{T} K_{ts}^2}} \tag{4.4}$$

BDS Test

In order to apply the test, the assumption of non-linearity of the time series has to satisfy. According to Bisaglia e Gerolimetto [5] research, a robust test to verify the non-linearity is the BDS Test(Brock et al., 1987). BDS is a nonparametric test, originally designed to test for independence and identical distribution (iid), but shown to have also power against a large gamma of linear and nonlinear alternatives. The BDS statistics is based on the correlation integral, a measure of the number of times temporal patterns are repeated in the data. Given a time series X_t, $t = 1, 2, ..., n$ and define its m-history as $X_t^m = (x_t, x_{t1}, ..., x_{tm+1})$, the correlation integral at the embedding dimension m is:

$$C_{m,T}(\epsilon) = \sum_{t<s} I_\epsilon(X_t^m, X_s^m)\{\frac{2}{T_m(T_m-1)}\}$$

where $T_m = T - (m-1)$ and $I_\epsilon(X_t^m, X_s^m)$ is an indicator function is the $\|X_t^m - X_s^m\| < \epsilon$. Basically, $C_{m,T}(\epsilon)$ counts up the number of m-histories that lie within a hypercube of size ϵ of each other. Bisaglia e Gerolimetto [5] (1996) define the BDS statistics as follows:

$$V_{m\epsilon} = \sqrt{T}\frac{C_{m,T}(\epsilon) - C_{1,T}(\epsilon)^m}{s_{m,t}}$$

which converges to a $N(0,1)$.

Bandwidth Selection: Cross-Validation for Time Series

In the implementation of the Causality in Quantiles test, bandwidth selection plays a crucial role. The Cross-Validation method is the most used to compute the bandwidth. However, in the case of a time series, the time dependency has to be maintained. Following the work of Peter et al.[19], the selection procedure minimizes the quantity:

$$CV(\mathbf{h}^{(i)}) = \frac{1}{n - n(w)}\sum_{t=1}^{n} M_\theta(Y_t, \hat{\mu}_\theta^{(-t)}(\mathbf{X}_i))w(\mathbf{X}_t)$$

where $w : \mathbf{R}^d \rightarrow \mathbf{R}^d$ is a nonnegative weight function used to omit observation at boundaries, $n(w)$ is the number of observations that take zero values in $w(\mathbf{X}_t)$ and $\hat{\mu}_\theta^{(-t)}(\mathbf{X}_i)$ is the leave-one(block)-out estimate obtained as:

$$\hat{\mu}_\theta^{(-t)}(\mathbf{X}_i) = inf\{y \in \mathbf{R} | \hat{F}_{\mathbf{X}_i}^{(-t)}(y) > \theta\}$$

$\hat{F}_{\mathbf{X}_i}^{(-t)}(y)$ is the inverse of the Nadaraya-Watson kernel estimator, instead $M_\theta(y,\mu)$ is the loss function defined as:

$$M_\theta(y,\mu) = \theta|y-\mu|^+ + (1-\theta)|y-\mu|^-$$

with $|y-\mu|^+$ and $y-\mu|^-$ the absolute of negative and positive values respectively.

4.3 Results

4.3.1 Anomaly Detection

As mentioned in the methodology, the detection method suffers from parameter selection. The initial analysis has been conducted with a 2% increase in the moving mean of the price, 25% increase in the volume and number of transactions, while the lag factor is 12 hours.

DAI Anomalies

Number of Anomaly: Initial Parameters

Type of Anomaly	DAI CB	DAI KK
Price	2	4
Volume	776	763
N. Transactions	689	816

Table 4.1: Anomaly detected in time series separately analyzed

These initial parameters are already effective in price anomaly, while the volume and number of transactions time series show higher volatility. The dates identified for the Coinbase exchange in the DAI case are:

Coinbase

- 14/06/2020 21:00:00
- 05/09/2020 16:00:00

Kraken

- 24/06/2020 11:00:00
- 24/07/2020 14:00:00
- 31/07/2020 12:00:00
- 05/09/2020 16:00:00

As Kamps et al.[17] reminded, no P&D event can co-occur on two different exchanges. Thus the 05/09/2020 date is undoubtedly not a P&D event. For the other dates, a plot of the anomaly and a candlestick plot has been computed. In figure 5.1 the anomaly is evident. Indeed there is a positive spike, but soon after, a negative spike. It can be a P&D, but the volume under the price time series does not show any particularity of the specific property of a P&D event. There is an increase in the trading volume, but it seems that is caused by the price higher volatility, with an increasing trend.

Figure 4.1: Candlestick 14/06 DAI Coinbase

In the case of Kraken's anomalies, there is a positive spike on the 24th June with a difference between the Closing price and the Highest price of $0,025\$$ and no difference between Opening and Closing price. Thus, inside that hour, a considerable price increase has happened. Nevertheless, the coin quickly lost all the value acquired. This means that a Short-Term Pump maybe has happened. With a closer look at 5Min level analysis and an Hour-wise plot in figure 5.3, it is clear that the price manipulation did not occur. The duration is short enough to exclude the possibility. Moreover, that spike does not correspond to a volume spike.

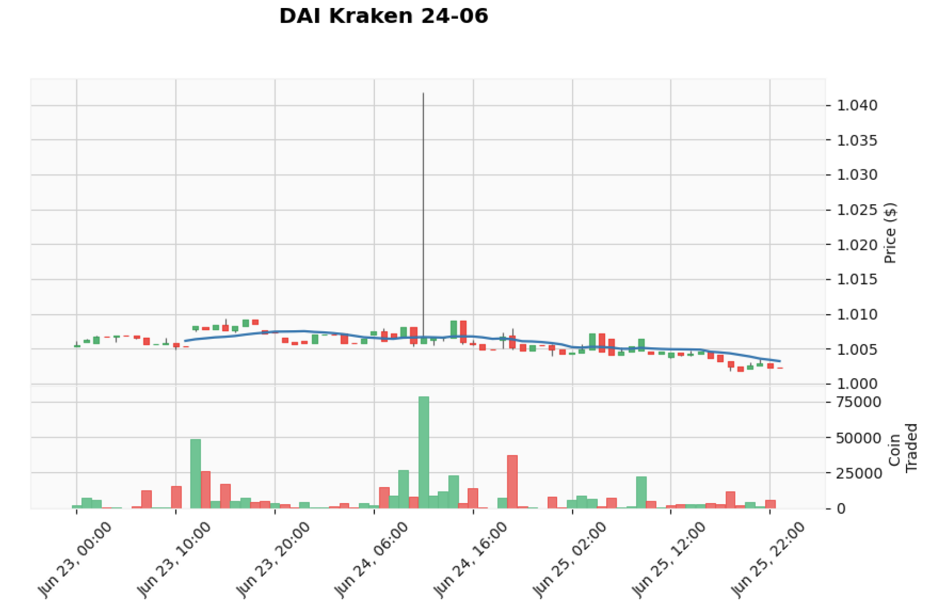


Figure 4.2: Candlestick 24/06 DAI Kraken

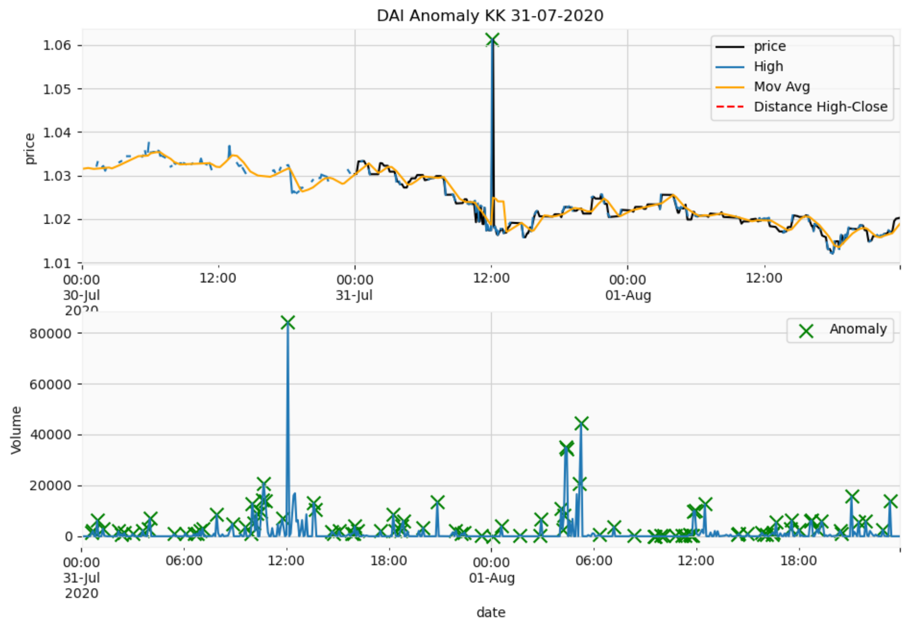


Figure 4.3: Anomaly 24/06 11:00 DAI Kraken

The rest of the other anomalies exhibit the same type of spike of the one occurred on the 24^{th} June.

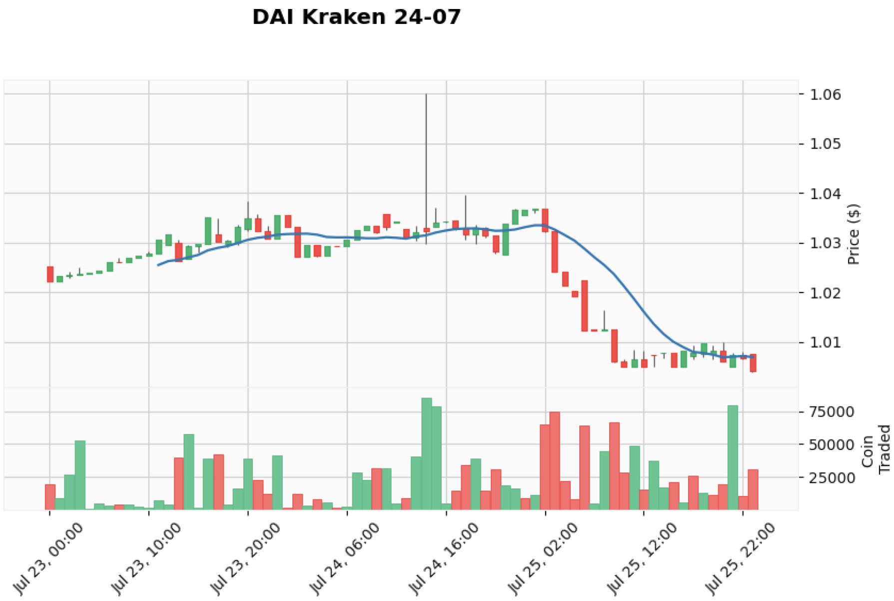
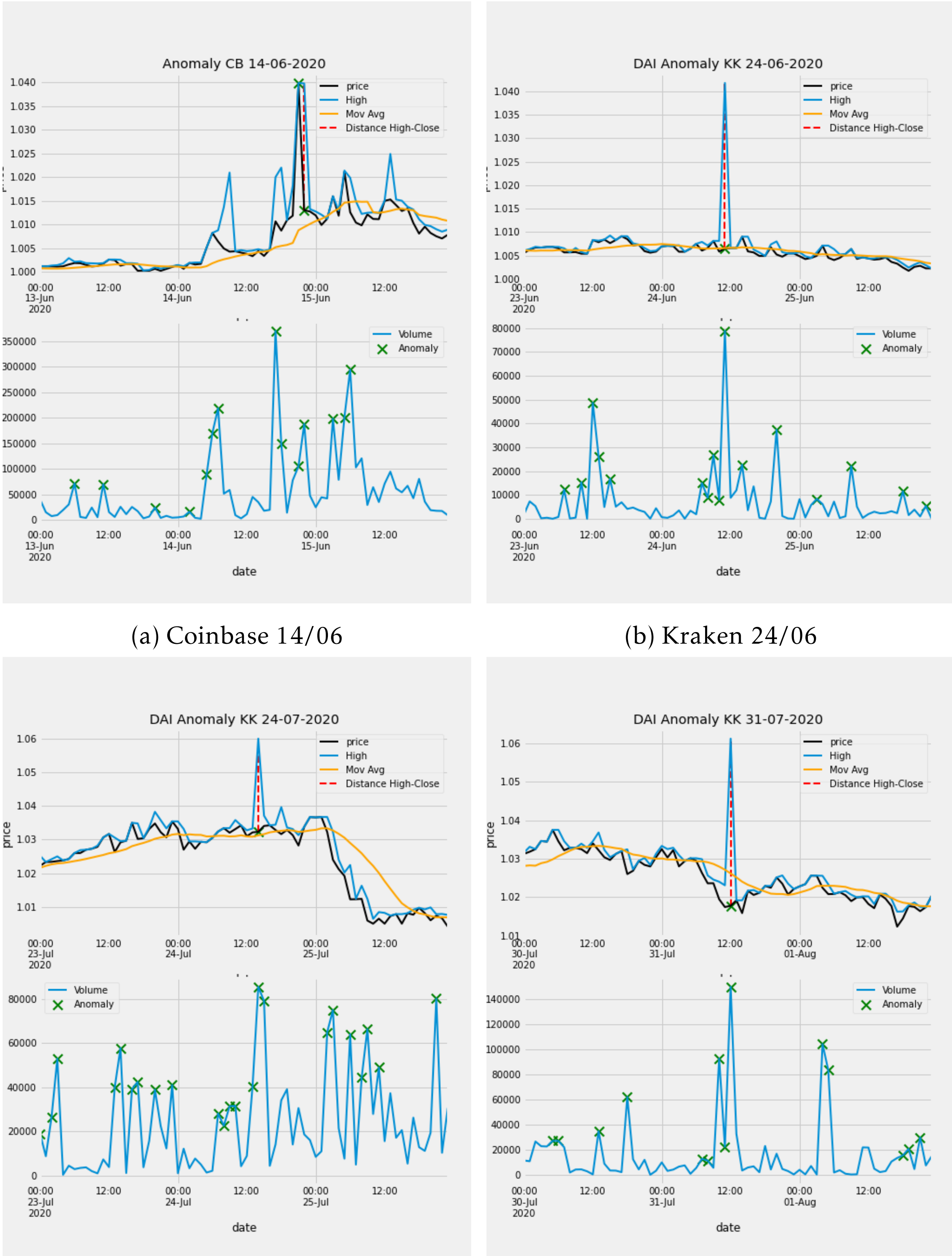


Figure 4.4: Candlestick 24/07 DAI Kraken

Figure 4.5: Candlestick 31/07 DAI Kraken



(a) Coinbase 14/06

(b) Kraken 24/06

(c) Kraken 24/07

(d) Kraken 31/07

Figure 4.6: Anomalies On DAI Coin

In conclusion, the initial parameters were just right to detect the price anomalies. More strict parameters provided no anomalies. Instead, with a 400% increase from the moving average, the volume and number of transactions reduced the number of detected spikes to less than 100 elements. No actual P&D events have been detected.

USDC Anomalies

Number of Anomaly: Initial Parameters

Type of Anomaly	USDC BT	USDC KK
Price	0	0
Volume	679	821
N. Transactions	812	794

Table 4.2: Anomaly detected in time series separately analyzed

In the USD coin case, no price anomalies have been detected with the same initial parameters chosen for the DAI coin. A second relaxation of the parameters would be non-sense, as a 2% increase is already low. Thus also the USD Coin seems has not been targeted by this kind of fraudulent activity.
These results confirm the initial hypothesis: these two coins have a market capitalization too big to be considered a possible target. What can be determined, also, the susceptibility of the price to volume and the number of transactions spikes? Given the absence of price anomaly, in contrast to numerous volume and number of transactions anomalies, the Causality in Quantiles test results should reject the null hypothesis.

4.3.2 Causality in Quantiles Test

Unfortunately, due to the lack of data from the Bittrex exchange for the USD Coin, the time series cannot be analyzed. Kernel method suffers the considerable number of zero elements.

BDS Test Results (m = 2)

Type of Anomaly	DAI CB	DAI KK	USDC KK
Test Statistics	13,94	3,80	5,25
P Value	0,0	2,9e-04	4,13e-07

Table 4.3: Results of the BDS Test with embedding dimension m = 2

The results highlight a non-linearity property for all the time series considered.

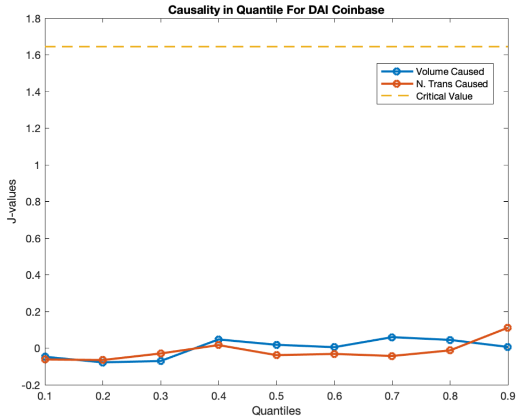
According to Jeong et al.[16], for the Kernel Estimation method, a gaussian kernel function has been chosen in both kernel computations, while the lag factor at which the test statistic has been computed is m = 2. Indeed the test has to check if the volume or the number of transaction influence the price in a short period, which corresponds to 10 minutes. The bandwidth computation instead was performed for ten quantiles, from 0,1 to 0,9, in the case of the Nadaraya-Watson Kernel estimator.

Bandwidth Nadaraya-Watson Kernel Estimator

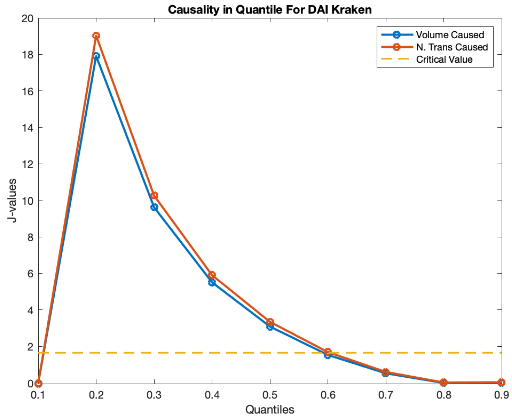
Quantiles	DAI CB	DAI KK	USDC KK
0,1	2,939e-04	8,033e-05	1,108e-04
0,2	1,291e-04	2,939e-04	1,108e-04
0,3	1,291e-04	2,573e-04	5,592e-05
0,4	1,230e-04	1,963e-04	3,761e-05
0,5	1,169e-04	1e-06	1,597e-04
0,6	1,902e-04	1e-06	1e-06
0,7	2,267e-04	1,291e-04	2,329e-04
0,8	2,878e-04	1,352e-04	1,780e-04
0,9	2,756e-04	2,329e-04	1,475e-04

Table 4.4: Bandwidth results from Cross Validation Procedure

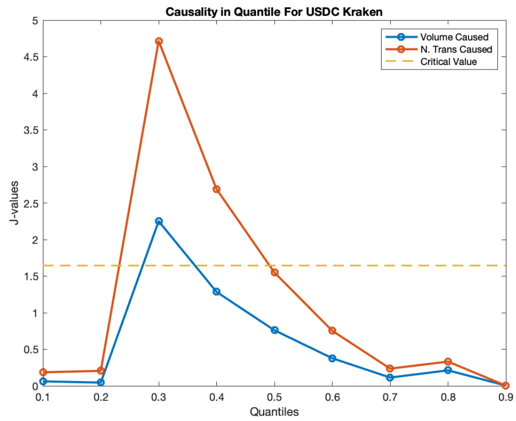
To avoid heavy computation, (the bandwidth selection with a dataset cutted to 2’000 elements from the originals 35’000 element, took almost 3 hours for each time series) the datasets has divided into 17 blocks of 2000 elements and for each block the J-distance estimator has been computed, and averaged for all the blocks result.
Figure 4.7 shows that the initial hypothesis made after the anomaly detection in the DAI case is correct. Indeed the null hypothesis of no causality cannot be rejected, as the value of , never reaches a critical value of 1.65, for both volume and number of transactions.



(a) DAI Coinbase



(b) DAI Kraken



(c) USDC Kraken

Figure 4.7: $\hat{J}$ distance estimator results

Instead, in the Kraken case, for both DAI and USD Coin, there is an initial causality in the first quantiles (from $0,1$ to 0.6 for DAI and $0,2$ to $0,5$ for USDC) but then the collapses under the critical value in the quantiles of interest for susceptibility to P&D event.
In conclusion, these coins seem to be strong enough not to suffer from these illicit activities. This can be said by looking at how they have never been targeted in anomaly detection and how the price does not suffer from volume spikes, even if they are frequent as in these exchanges.

Chapter 5

Conclusion

This project aimed to perform a general analysis on the stablecoins considered to understand if there could be a future regarding these coins. As introduced before, cryptocurrencies suffer from the illicit activity and high volatility, undermining their credibility. This is also the reason why investments banks prefer more classic financial tools to use.
In basic investment portfolio theory, to reduce the risk exposure, the simplest option is to buy Bonds, which guarantee low volatility at the expense of lower returns. Moreover, when risk is high, investments banks can also shift the risk to another type, liquidating all the assets and thus exposing to currency risk, in the case the latter is significantly low.
Unfortunately, these simple procedures of any Stock Exchange Market are not operable in the Cryptocurrency Exchange Markets. The liquidation fee of the coins in any Exchange Market is consistently higher than exchanging one coin for another, making the practice avoidable in most cases. Thus, a new kind of coin was necessary for this framework to limit exposure to risk without paying enormous fees. With these objectives, stablecoin was born. Thus, the main goal was to lower the volatility that characterizes cryptocurrencies. In this sense, both coins, DAI and USDC, achieved their ambitions. Indeed in the conditional volatility analysis, the DAI and USDC one are consistently lower than Bitcoin one. However, to be considered a safe haven (Baur [4]) against bitcoin's high volatility, their price volatility must also be uncorrelated to the other cryptocurrencies. Unfortunately, only the USDC manage to keep their price stable

against a high period of the volatility of other coins.
Moreover, thanks to lower fees in exchanging coins on the same platform, the Pump and Dump scheme has become one of the most popular fraudulent activities to perform. Thanks to Telegram, Discord, and Reddit, it has been possible to coordinate thousands of people to purchase the same coins. As said in the beginning, chances to detect, with unsupervised techniques, this kind of activity was not an easy task. The probability that these coins were targeted was low, as their value is backed by something not connected to the cryptocurrencies framework, the time series on which computations were made regarded the pairs USDC/USD and DAI/USD, and the market capitalization was much higher than usual targeted coins. Indeed results show that no P&D event took place in the period considered. Moreover, the Causality in Quantiles test showed that, at higher quantiles, the volume and the number of transactions registered on the exchange markets analyzed did not cause any spikes, so that if future attacks may happen, no success is guaranteed.
In conclusion, the best overall coin was the USDC, given the low correlation with Bitcoin and no non-linear Granger Causality was found above the $0,5$ quantiles. The DAI performed well but showed consistently higher volatility than its competitor, as expected, given its design (On-Chain Collaterilized Stablecoin) suffer the volatility of the units of assets that backs its value. In the DAI case, these assets are Cryptocurrencies (Ethereum), thus the high correlation of the conditional volatility with Bitcoin and less stable value. The Causality In Quantiles also showed no Granger Causality in volume and number of transactions.
In the end, stablecoins fulfill the space of the exchange market, which is occupied by Bonds in the Stock market. Maybe the design has to be perfected, but the initial results are promising. Indeed, not surprisingly, J.P. Morgan decided to invest in cryptocurrencies in 2020 publicly.

List of Figures

List of Tables

Bibliography

[1] Torben G. Andersen et al. "The Distribution of Realized Exchange Rate Volatility". In: *Journal of the American Statistical Association* (2001).

[2] Mehmet Balcilar et al. "Can Volume Predict Bitcoin Returns and Volatility? A Quantiles-Based Approach". In: *Economic Modelling* (2017).

[3] Mehmet Balcilar et al. "Does Economic Policy Uncertainty Predict Exchange Rate Returns and Volatility? Evidence from a Nonparametric Causality-in-Quantiles Test". In: *Open Econ Rev* (2016).

[4] Dirk G. Baur and Lai T. Hoang. "A crypto safe haven against Bitcoin". In: *Finance Research Letters* (2021).

[5] Luisa Bisaglia and Margherita Gerolimetto. "Testing for (non)linearity in economic time series: a Monte Carlo comparison". In: *Working Paper Series* 3 (2014).

[6] Tim BOLLERSLEV. "GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY". In: *Journal of Econometrics* (1982).

[7] C.T. Brownlees and G.M. Gallo. "Financial econometric analysis at ultra-high frequency: Data handling concerns". In: *Computational Statistics Data Analysis* 51.2232-2245 (2006).

[8] Dirk Bullmann, Jonas Klemm, and Andrea Pinna. "In search for stability in crypto-assets: are stablecoins the solution?" In: *Occasional Paper Series* (Aug. 2019).

[9] Jeffrey Chu et al. "GARCH Modelling of Cryptocurrencies". In: *Journal of Risk and Financial Management* (2017).

[10] Anirudh Dhawan and Tālis J. Putniņš. "A new wolf in town? Pump-and-dump manipulation in cryptocurrency markets". In: *SSRN Working Paper* (Jan. 2020).

[11] Robert F. Engle. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* (1982).

[12] Neil Gandal et al. "Price manipulation in the Bitcoin ecosystem". In: *Journal of Monetary Economics* 95.86-96 (2018).

[13] Zaid Harchaoui et al. "Kernel-Based Methods for Hypothesis Testing: A Unified View". In: *IEEE Signal Processing Magazine* Special Issue on Advances in Kernel-Based Learning for Signal Processing.30 (2013), pp. 87–97.

[14] Lai T. Hoang and Dirk G. Baur. "How stable are stablecoins?" In: *The European Journal of Finance* (June 2020).

[15] G. Reza Jafari, A. Bahraminasab, and P. Norouzzadeh. "Why does the Standard GARCH(1,1) model work well?" In: *International Journal of Modern Physics C* (July 2007).

[16] Kiho Jeong, Wolfgang K. Härdle, and Song Song. "A CONSISTENT NONPARAMETRIC TEST FOR CAUSALITY IN QUANTILE". In: *Econometric Theory* 28.4 (Aug. 2012), pp. 861–887.

[17] Josh Kamps and Bennett Kleinberg. "To the moon: defining and detecting cryptocurrency pump-and-dumps". In: *Crime Science* (2018).

[18] Tao Li, Donghwa Shin, and Baolian Wang. "Cryptocurrency Pump-and-Dump Schemes". In: *SSRN Working Paper* (July 2020).

[19] Peter Nyamuhanga Mwita. "Nonparametric Estimates for Conditional Quantiles of Time Series". In: *ResearchGate* (2003).

[20] Yaohao Peng et al. "The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with Support Vector Regression". In: *Expert Systems With Applications* 97.177-192 (2018).

[21] Robert H. Shumway and David S. Stoffer. *Time Series Analysis and Its Applications with R Examples*. Second. Springer Science+Business Media, LLC, 2006.

[22] Xiaojun Song and Abderrahim Taamouti. "Measuring Nonlinear Granger Causality in Quantiles". In: *Journal of Business Economic Statistics* (July 2017).

[23] Friedhelm Victor and Tanja Hagemann. "Cryptocurrency Pump-and-Dump Schemes: Quantification and Detection". In: *Conference Paper* (2019).