FIN-525: Financial Big Data

₿itcoin impact on several major cryptocurrencies

Léopold Bouraux - 257368
Kenyu Kobayashi - 314181

Directed by :
Prof. Damien Challet
T.A. Goutham Gopalakrishna

January 14, 2021

Abstract

Classical currencies are generally fairly correlated with the price of some commodities. For example, the Australian dollar is known to be a currency sensitive to basic products such as iron ore, or the Russian ruble is related to oil. In addition to being correlated to these commodities, currencies are sensitive to the US dollar rate. The same system we find with cryptomoney is similar. Each cryptomoney has its own utility and specificities. For example, Ethereum was one of the first cryptomoney systems to offer smart contracts, a way of programming the blockchain to be used in new contexts (securing a transparent agreement between two parties, automating payment and eliminating the risk of non-payment, reducing intermediate costs in the development, monitoring and conclusion of a contract...). Another example is Ripple, which aims to enable secure, instantaneous and almost free global financial transactions of any size without chargebacks. The value of these cryptocurrencies varies according to the utility they have at a given moment but also according to the price of Bitcoin, which is more speculative.

The current health crisis due to the Covid-19 epidemic raises doubts about a financial crisis in the coming months. This downturn could potentially be beneficial to cryptomarkets. In this context, our goal is to analyze the evolution of Bitcoin and its impact on several other cryptocurrencies in order to better understand this market. It has been shown that the price of many of the major cryptomoney systems is largely based on the Bitcoin price. Since all these currencies have been fluctuating along with Bitcoin for years, we decided to first calculate some correlation comparisons to see exactly how much of a hold BTC commands.

Following this analysis assessing the correlation of BTC with the other cryptocurrencies, we focused in a second step on the influence that trades in BTC have on other cryptocurrencies. In order to compute this quantity, we used cross-response functions, introduced by Henao-Londono et al. 2020. Notably, we took a closer look at the behaviour of these latter during BTC's massive flash crash of March 2020.

Figure 1: The major cryptocurrencies logos

Contents

In the following article, we will show several graphs - feel free to read the corresponding Python Notebooks where we have handled the data. The Github repository is freely accessible and can be found **here**.

1 Data introduction

1.1 Which cryptos ?

To compare cryptocurrencies, we used data available on CoinMarketCap. On Figure 2, we can see the 10 largest cryptos according to their market capitalization.

We can notably see there that Bitcoin is an unchallenged leader in the cryptomoney industry for the moment, with a capitalization 5 times higher than its first challenger Ethereum, and more than 100 times bigger than the 10th bigger cryptocurrency, named Chainlink.

For our analysis, we did not focus on the details and similarities of blockchain implementations and other functionalities of cryptosystems, but instead we relied on the size (represented here by market capitalization) and type of currency. For example, currencies like the Tether (USDT) or the USD Coin (USDC) are stable coins, which are supposed to replicate the US dollar rate. They are therefore not correlated with Bitcoin. Thus, we will not compare them to the evolution of Bitcoin but instead, we will use one of them, the USDT, as a reference exchange rate to compare Bitcoin with other cryptos.

#	Name	Price	24h	7d	Market Cap	Volume	Circulating Supply	Last 7 Days
1	Bitcoin BTC	$31,136.87	▼5.49%	▲15.22%	$582,496,453,178	$82,228,047,815 2,624,341 BTC	18,590,606 BTC	
2	Ethereum ETH	$1,016.78	▲7.57%	▲37.46%	$116,092,156,808	$59,049,698,989 58,044,491 ETH	114,115,911 ETH	
3	Tether USDT	$1.01	▲0.52%	▲0.06%	$21,797,659,016	$130,208,671,140 130,274,243,828 USDT	21,808,636,250 USDT	
4	XRP XRP	$0.232700	▼0.54%	▼19.08%	$10,580,937,903	$6,268,841,468 26,900,323,978 XRP	45,404,028,640 XRP	
5	Litecoin LTC	$151.29	▼1.25%	▲14.91%	$10,050,987,545	$14,313,078,055 94,329,981 LTC	66,240,780 LTC	
6	Polkadot DOT	$9.27	▼4.70%	▲48.97%	$8,286,200,612	$3,277,370,096 354,673,795 DOT	896,724,547 DOT	
7	Bitcoin Cash BCH	$404.86	▼0.57%	▲9.67%	$7,459,287,795	$8,579,652,434 21,394,661 BCH	18,600,863 BCH	
8	Cardano ADA	$0.211423	▲8.19%	▲22.81%	$6,558,008,115	$3,260,608,138 15,468,968,451 ADA	31,112,484,646 ADA	
9	Binance Coin BNB	$39.63	▼3.04%	▲11.06%	$5,737,946,966	$805,919,669 20,282,531 BNB	144,406,561 BNB	
10	Chainlink LINK	$13.46	▲2.60%	▲5.60%	$5,351,003,176	$3,027,851,197 226,061,441 LINK	399,509,556 LINK	

Figure 2: Top 10 coins according to their market capitalisation [3]

In the beginning we took 9 coins to explore. We chose first the 3 biggest classical cryptos after Bitcoin:

- Ethereum (**ETH**) #2
- Ripple (**XRP**) #4
- LiteCoin (**LTC**) #5

Then, we had chosen 3 ”average” coins:

- Cardano (**ADA**) #8
- Binance Coin (**BNB**) #9 because the data we will use (see 1.2) comes from the Binance API.
- ChainLink (**LINK**) #10

Finally, we took 3 ”alternative coins”, which have smaller market capitalisation:

- Theta (**THETA**) #18 with a market cap of $1,895,383,337 on 04/01/2021
- Dash (**DASH**) #36 with a market cap of $690,832,977 on 04/01/2021
- Nano (**NANO**) #78 with a market cap of $114,535,560 on 04/01/2021

1.2 Fetching data

In order to find Cryptocurrencies Big Data, we decided to use data from a large crypto trading platform. Among the two biggest exchanges platform currently in use, Coinbase and Binance, we chose the latter. By accessing this data via the API [1], we have been able to get tick by tick data from the launch of our cryptocurrencies on the platform until January 10th. This fully up-to-date data in huge quantity is especially interesting for our future analysis. With this range we will be able to focus on interesting moments of crypto markets as the early 2018 rise and the last one which happened just few days ago. Historical prices can be visualised on Figure 9.

However, after extensive research, we find out that BBO data for the cryptomarket is very difficult to obtain. They are either sold at quite high prices or their historical records are not kept on the API servers because they represent a huge amount of data. This was the case with the Binance API where BBO data is only available at the present time. To fetch such data, we would have had to leave our computer running continuously which is not feasible. Instead, we switched to trade history. The binance API allows only 1000 trades per request from a given date, so we had to program several methods to combine all the requests without forgetting any trades or having duplicate ones. Once the data was downloaded (it took more than 4 full days), we saved each daily crypto trades dataframe in a hdf5. Then we merged all daily files for each cryptos into big parquet files (the hdf5 format didn't support such a dataframe, it crashed for our largest dataset, the bitcoin with more than 500 millions trades).

As mentioned above we fetched, the exchange rate between our cryptos and Tether, the stable coin which replicates dollars. Here is what the data looks like before we processed it to keep only what we are interested in:

timestamp	trade_id	price	volume	first_id	last_id	is_buyer_maker
2017-11-24 00:00:01.170	666367	405.03	7.17000	709881	709881	True
2017-11-24 00:00:01.170	666368	405.00	19.93873	709882	709884	True
2017-11-24 00:00:01.196	666369	405.00	0.69427	709885	709885	False
2017-11-24 00:00:02.649	666370	405.03	0.35481	709886	709886	False
2017-11-24 00:00:03.427	666371	405.03	1.20269	709887	709888	True

Figure 3: **ETH/USDT** 5 first trades on Binance platform

Note that initially, there was one last feature that was removed. It showed whether the current trade was made at the best price or not. We eliminated all the trades that were not at the best price (it was a very small part of the dataframe $\pm 0.01\%$).

While downloading data, the `trade_id` was used to find the following trade in a new query. The USDT trade value and the volume are represented respectively in the `price` and `volume` columns. The time at which a trade occurs is in the `timestamp` feature. When a trade is splitted into several chunks we can track them thanks to the `first_id` and `last_id` features. Finally, the last one `is_buyer_maker` indicates if the trade was buyer initiated or not. More precisely, when `is_buyer_maker == True`, the trade price is the bid price, otherwise, the trade price is the seller price, also known as the ask price.

Below are some statistics about the amount of raw data we had. Due to computational power limitations, we had to aggregate it.

Table 1

Raw Data Stats			
Cryptocurrency	Start Date	End Date	Number of Trades
ADA	2018-07-25	2021-01-10	41,772,850
BTC	2017-11-24	2021-01-10	500,317,769
BNB	2018-02-13	2021-01-10	91,584,907
DASH	2019-07-05	2021-01-10	12,053,267
ETH	2017-11-24	2021-01-10	208,098,931
LINK	2019-04-25	2021-01-10	57,312,498
LTC	2018-03-22	2021-01-10	60,731,167
NANO	2019-07-11	2021-01-10	2,582,072
THETA	2019-07-18	2021-01-10	9,394,721
XRP	2018-08-11	2021-01-10	101,056,031

1.3 Preprocessing and Aggregation

The preprocessing part was relatively trivial to carry out. From the prices, all was needed was to create the daily continuously simple return, defined by:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{1}$$

On the other hand, the aggregation part was more specific. Pandas has difficulty managing dataframes as big as these and its calculations are much less efficient. Hence, aggregation was essential.

There are two ways of aggregating data:

- Binning in physical time: price return every X seconds
- Binning in activity time (e.g. trade)

In order to compute correlations, we need to bring synchronicity to our data. For this reason, we decided initially to aggregate in physical time, and aggregated data with a frequency of 30 seconds. This number was meticulously chosen to compress our data just enough for it to be managable, all the while maintaining a high granularity. Note that is it possible to aggregate it with a larger frequency if needed.

Table 2

Compression Stats			
Cryptocurrency	Nb Trades Raw	Nb Trades Agg	Compression ratio
ADA	41,772,850	3,945,102	10.59
BTC	500,317,769	6,536,042	76.55
BNB	91,584,907	5,489,119	16.68
DASH	12,053,267	1,648,314	7.31
ETH	208,098,931	6,473,357	32.15
LINK	57,312,498	2,786,520	20.57
LTC	60,731,167	5,044,320	12.04
NANO	2,582,072	575,820	4.48
THETA	9,394,721	1,082,832	8.68
XRP	101,056,031	4,813,842	20.99

The data is aggregated by the trade time as mentioned earlier but also on the feature `is_buyer_maker`. As this category is evenly split between True and False values (49.89% for ETH/USDT in this case), it allows us to approximate the value of the `mid_price`. To compute it, we first compute the average price for buyer initiated trades and for non buyer initiated trades during the slot of 30 seconds, and then we compute the mean between the two values that have the same `timestamp`.

The Figure 4 shows us such aggregated data. We can even notice that sometimes we might obtain only buyer initiated trades during a time slot. Fortunately, this doesn't happen anymore with more recent data because there are many more trades.

timestamp	is_buyer_maker	nb_trade	min_price	max_price	mid_price	volume_tot
2017-11-24 00:00:30	False	13	404.99	405.03	404.943269	34.13454
2017-11-24 00:00:30	True	8	404.00	405.03	404.943269	31.66771
2017-11-24 00:01:00	True	3	404.01	404.03	404.020000	0.40172
2017-11-24 00:01:30	False	1	407.00	407.00	406.002500	4.74070
2017-11-24 00:01:30	True	2	405.00	405.01	406.002500	1.58246

Figure 4: **ETH/USDT** 5 first aggregated trades

To switch from the raw data represented by the DataFrame `df` (Figure 3) to the compressed data represented by `df_agg` (Figure 4), we use the following piece of code:

```
df_agg = df.groupby([pd.Grouper(freq='30s',origin=0,label='right'),
                     'is_buyer_maker'])
                      .agg({'trade_id': ['count'],
                            'price': ['min', 'max', 'mean'],
                            'volume':'sum'})
                      .reset_index()
df_agg.columns = df_agg.columns.get_level_values(0)
df_agg.index = df_agg.timestamp
df_agg.drop(columns='timestamp', inplace=True)
df_agg.columns = ['is_buyer_maker', 'nb_trade', 'min_price',
                  'max_price', 'mid_price', 'volume_tot']
df_agg['mid_price'] = df_agg['mid_price'].groupby('timestamp').mean()
```

For correlation computations, we also tried the Hayashi-Yoshida method to get rid off the asynchronous problem of trades (Section 2.3). This method consists in multiplying the return of a first asset A by the last known return of the other asset, B. It is a way to bring synchronicity to our data without binning it.

A simple way to process two time series of returns, to put them in the form required by the Hayashi-Yoshida method is to concatenate them vertically. This will add NaN values when there is no trade in one time series but in the other. Then we just replaced the NaN values by the last known value below it (Fig 5). Note that it creates bigger dataframe, then we will observe Hayashi-Yoshida correlation over a shorter period of time.

```
MERGED = pd.concat([df_returns, btc_returns], axis=1)
MERGED.columns = ['return_ETH', 'return_BTC']
MERGED.fillna(method='ffill', inplace=True)
MERGED.dropna(inplace=True)
```

timestamp	return_XRP	return_BTC
2020-01-01 00:00:06.566	NaN	0.000207
2020-01-01 00:00:13.945	NaN	-0.000207
2020-01-01 00:00:21.171	0.000000	NaN
2020-01-01 00:00:30.249	NaN	0.000156
2020-01-01 00:00:32.393	-0.000518	NaN

timestamp	return_XRP	return_BTC
2020-01-01 00:00:06.566	0.000207	0.000207
2020-01-01 00:00:13.945	-0.000207	-0.000207
2020-01-01 00:00:21.171	0.000000	0.000000
2020-01-01 00:00:30.249	0.000156	0.000156
2020-01-01 00:00:32.393	-0.000518	-0.000518

Figure 5: Hayashi-Yoshida Aggregation

1.4 Basic introductory charts

The analysis of traded volumes can allow us to understand both the daily habits of buyers, but also in a more global way, the evolution of the world interest in cryptos. Although the study of the correlation between the volumes of these cryptos is interesting, we have decided to focus on the study of prices. However here are the graphs concerning the volumes related to Bitcoin and Ethereum. The other graphs are available directly on the github repository.

1.4.1 Volumes

The first interesting chart concerning volumes is the historical record of all transactions. Here we aggregate the transaction each minutes in order to obtain a meaningful graph.

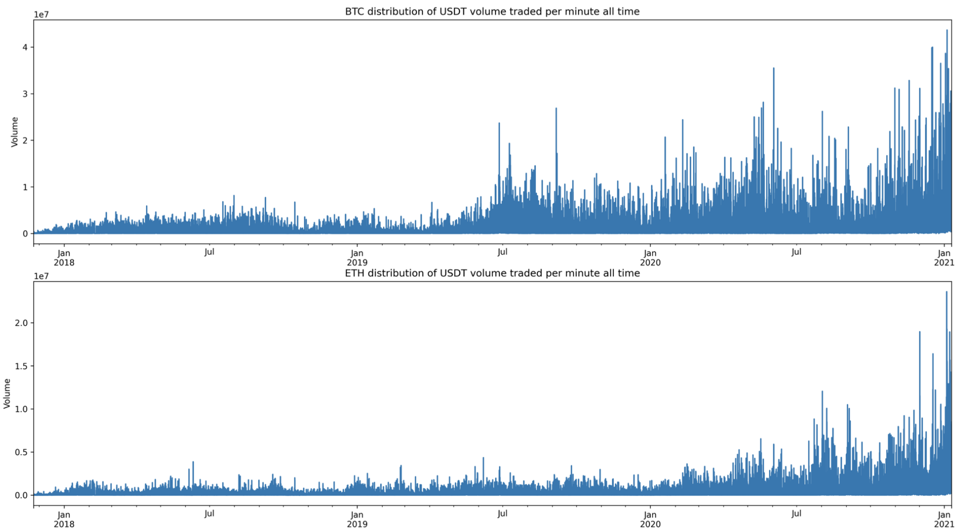

Figure 6: Historical Volumes traded - **BTC/ETH** comparison

With this plot, we can see the growing interest in cryptomoney. The quantities exchanged on the platform are almost 10 times greater than two years ago. This can be explained by the ease of access to these markets thanks to platforms that are increasingly simpler to use now. But it is obvious that the world is becoming more and more attracted to this market.

Over the day the number of exchanges and the amount exhanged vary a lot. Contrary to the classical financial markets, the cryptocurrencies market exchanges are open 24/7. We will therefore not find the kind of distribution we saw in the course of the day for the daily distribution of volumes. However some patterns are worth seeing.

The left side of Figure 7 represents the sum of all exchanges per minute since the end of 2017. We notice that at the beginning of each hour, there is a peak of trades. There is also a global peak at 16:00 UTC, and that for all other currency cryptos. If we take into account that there have been 1143 days studied, we find that on average, about 330 trades take place every minute for Bitcoin against about 130 for Ethereum.

On the right side we can observe the distribution of amount traded. As we could assume, and as seen in the classical financial markets, there are exponentially more small trades than large ones.

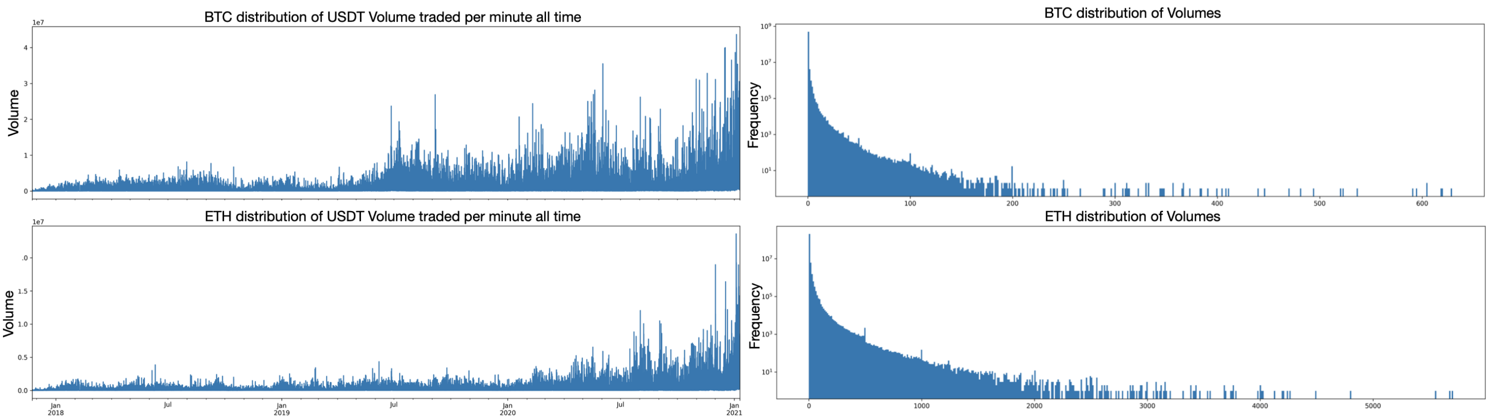

Figure 7: Daily Number of Trades Distribution and Amount Traded Distribution - **BTC/ETH** comparison

1.4.2 Prices

As mentioned above, the correlation study will be based on prices. The first chart that will be useful to spot the periods when the price correlation analysis between cryptos will be interesting is the historical price trend. Thanks to this plot, we will be able to focus on key events.

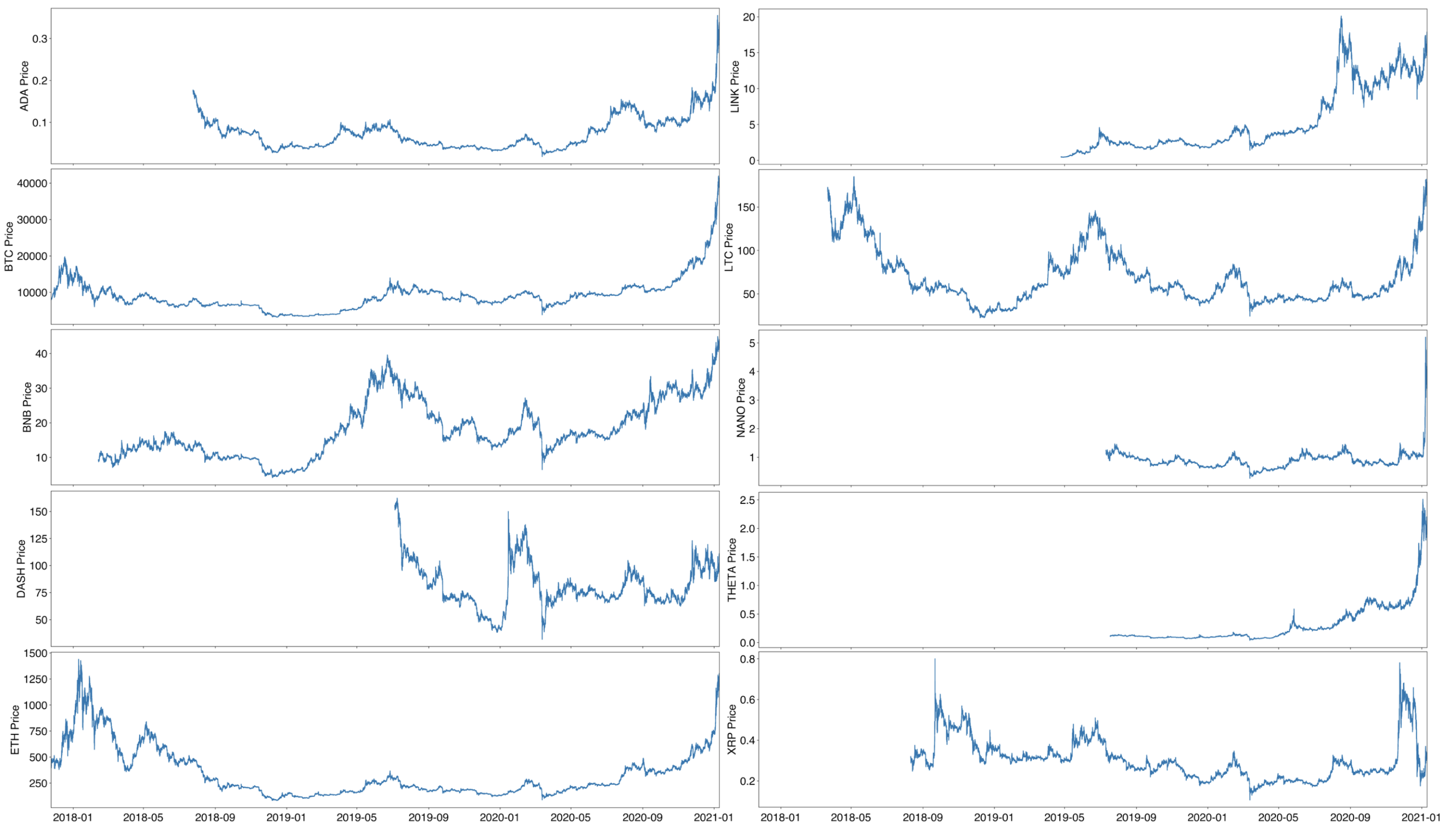

Figure 8: Historical Prices for the 10 studied cryptos

At first glance, we can see interesting patterns that seem to occur even between some "families" of cryptos. This is notably the case of the BNB, DASH and LTC which seem to be strongly correlated since the beginning of 2019. We can see three distinct increases, in summer 2019, spring 2020 and late 2020/early 2021. Another group of crypto that seem to be highly correlated are ADA, BTC, ETH. Finally, others seem to be correlated with the latter at certain times and otherwise seem to follow their own path.

For our correlation analysis, it is relevant to analyze three kinds of periods: one where the cryptos all undergo a common crash, see how they evolve after this crash, and finally a period where all the values are rising like in this end of year.

1.4.3 Spreads

Finally, we also looked at the evolution of the spread of each cryptocurrency through time. To retrieve a spread within a given time interval, we chose to compute it as the difference between the maximum ask price and the minimum bid price within that time interval.

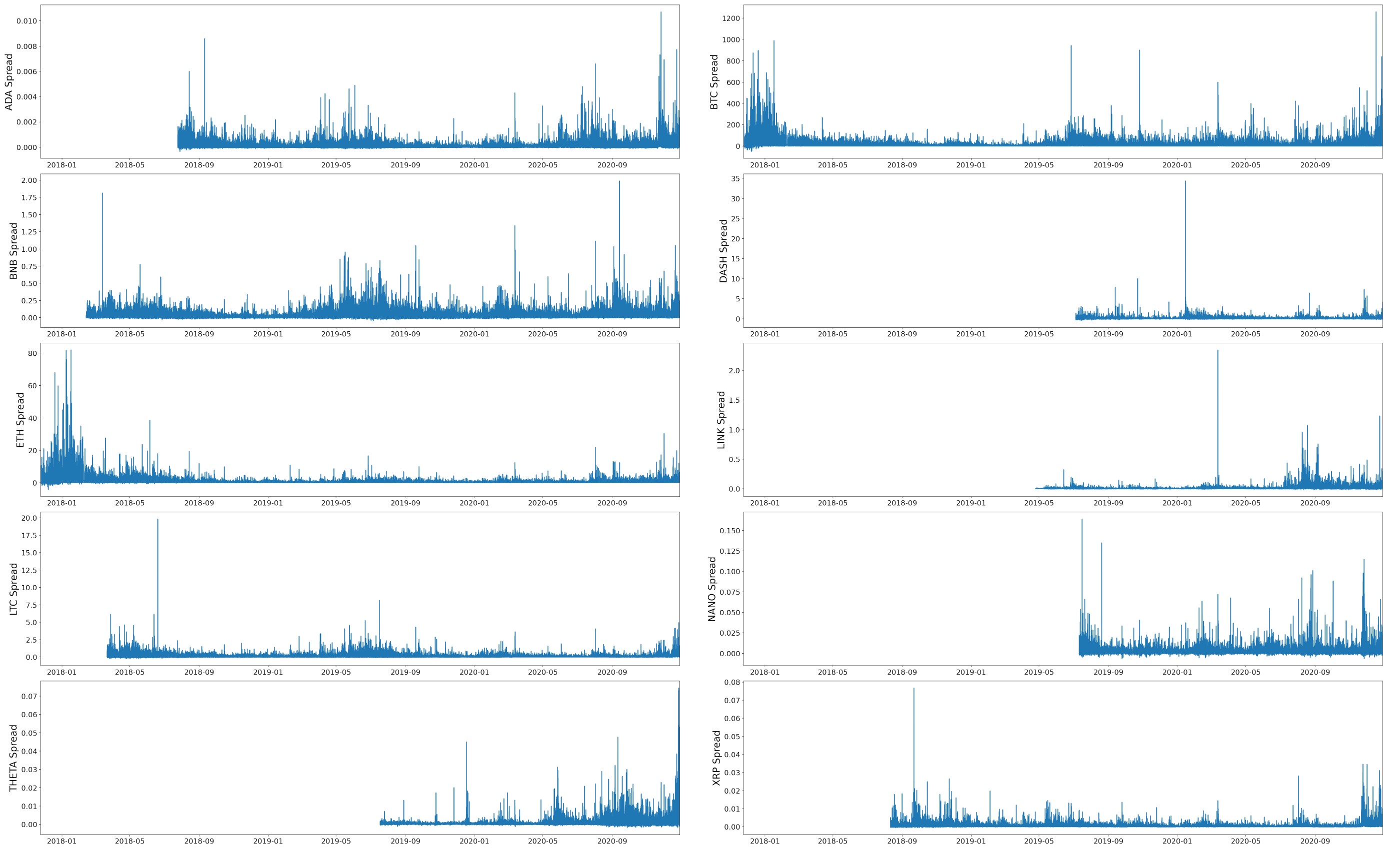

Figure 9: Historical spreads for the 10 studied cryptos

We can observe on these plots several spreads that stand out at the same time for different cryptocurrencies. In fact, these correspond to various flash crashes that occured in the world of crypto markets, which will be commented in Section 3.

2 Correlation Analysis

Correlation is one of the most common statistics for analyzing financial markets. A correlation is a number that describes the degree of relationship between two assets. It can be used to measure how one market evolves relative to another market, or in our case, how crypto-currencies evolve relative to Bitcoin. The measurement scale is from -1 to +1. A perfect positive correlation between two assets will be in the form of +1, and the markets will then move in a similar way. A perfect negative correlation will be in the form of -1, the markets move in the opposite direction. Obviously some of these values fluctuate over time and deviations may appear. First of all, as we have seen previously, in order to gain precision and not be biased by differences in currency values (Bitcoin is currently around $35,000 while other cryptomoney systems such as Ripple have much lower values, around $1), we calculate the correlation on price returns.

Correlation between asynchronous timeseries is tricky. One of the major problems is to have values to compare. But with tick-by-tick data, exchanges between cryptos do not take place at the same time. First of all, we aggregated the data to group it in 30-second intervals. Next, we will study a more advanced way to analyze the correlation at higher frequencies. For this part we get focused on five cryptos which are ADA, ETH, LTC NANO and XRP.

2.1 Different sliding window size

It is possible to calculate the correlation over different periods. If we want a general correlation between a timeserie x and another y on the whole history of two cryptos, it is enough to take a computing window as large as the total period for which we have data for both currencies. But if we want to see an evolution of the correlation, we have to calculate the correlation $R_{t,w}$ at a time t according to a sliding window of size w:

$$R_{t,w}(x,y) = \frac{\sum_{i=t-w+1}^{t}(x_i-\overline{x_i})(y_i-\overline{y_i})}{\sqrt{\sum_{i=t-w+1}^{t}(x_i-\overline{x_i})^2 \cdot \sum_{i=t-w+1}^{t}(y_i-\overline{y_i})^2}} \tag{2}$$

$$\overline{x_i} = \frac{1}{w}\sum_{i=t-w+1}^{t} x_i \qquad \overline{y_i} = \frac{1}{w}\sum_{i=t-w+1}^{t} y_i$$

We first want to analyze the evolution of the correlation as a function of the w size of this window. To start and have a global but not necessarily high frequency overview of the correlation of our currencies, we use the daily return. We calculated it by keeping only the close price each day. The close price being the last price of the day for the markets still open.

Figure 10 represents the correlation with Bitcoin of the daily return of the 5 cryptos mentioned above according to the size of the window used.

On these plots we notice that the larger the window, the more stable the correlation is, the narrower the window, the more sensitive the correlation is to small differences in price returns.

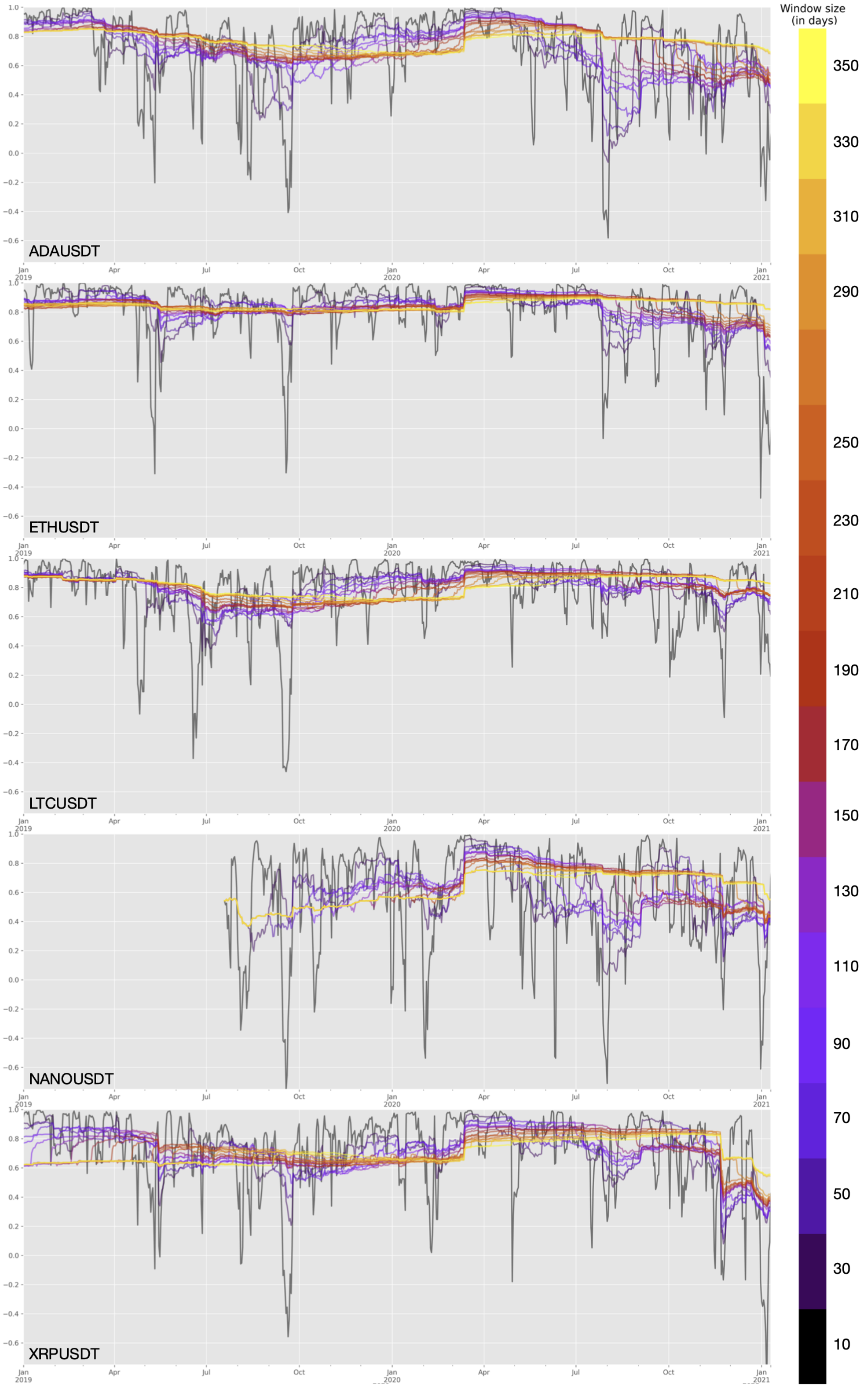

Figure 10: Correlation of daily returns with BTC in function of the window size (in days)

The bright curves are used to determine if a currency is highly correlated to Bitcoin on average. We can see that Ethereum is the most correlated currency - always above 0.8 since January 2019 if we look at the correlation with a 550 days window. On the other hand, the least correlated on average until the end of 2020 was the Nano coin.

These curves with large window size also point out that there was a jump in correlation of all the cryptos at the beginning of March 2020. All correlations jumped by about 0.1. So it will be interesting to focus on this period. (2.2)

The darker curves represent the correlation over a smaller period of time, they are more accurate but more sensitive to changes as well. These curves show peaks of decorrelation at certain times. What is interesting is that these peaks often occur in several cryptos at the same time. A common external factor has then impacted the price of these prices but not Bitcoin's price. This is notably the case in September 2019 or end of July 2020. The most pronounced negative peak occurred recently, in early January 2021, for Ripple. The U.S. Securities and Exchange Commission (SEC) has filed a lawsuit against Ripple for the sale of unregistered securities for $1.38 billion. As a direct result, the withdrawal of cryptos began on certain platforms such as Coinbase and the price of XRP collapsed while those of other cryptos rose in parallel with Bitcoin.

The size of the window used is therefore different from what we are interested in calculating. For the next part, we will focus on more specific periods.

2.2 Data fetched with different intervals

To begin with, let's take the correlations on the daily returns seen earlier, but with a fixed window size that we have set at 31 days, as a benchmark example.

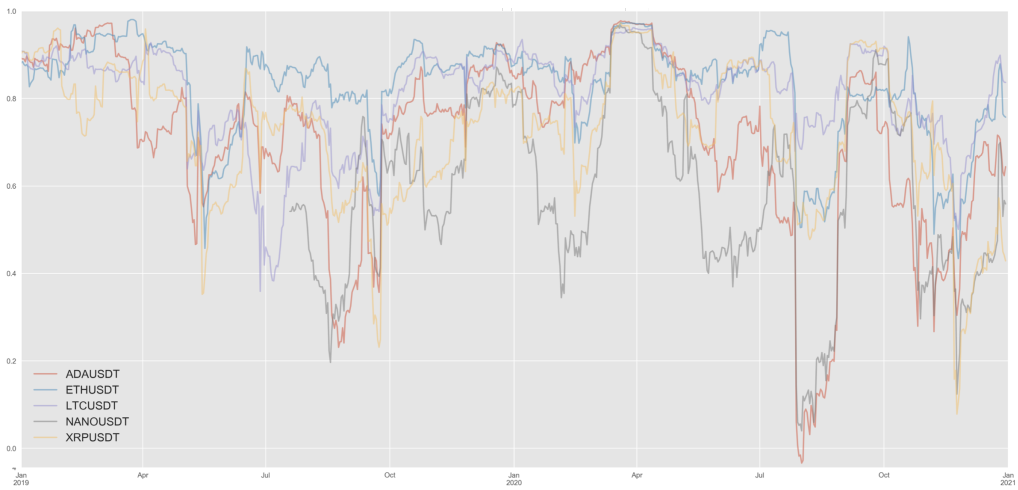

Figure 11: Correlation with BTC of daily return with a 31 days window

This graph allows us to precisely locate the two areas of interest to be analyzed. The first one, the important correlation zone of all cryptos with Bitcoin between the beginning of March 2020 and the end of April 2021, that is to say the period of the first wave of Covid-19 in Occidental countries. Then the crash period of all the correlations of the cryptos with Bitcoin ends around July 26, 2020. We are going to analyze what happened to the price curve at those times. But first, let's have a look on a more granular correlation graph: the correlation with BTC of hourly price returns, on the same period.

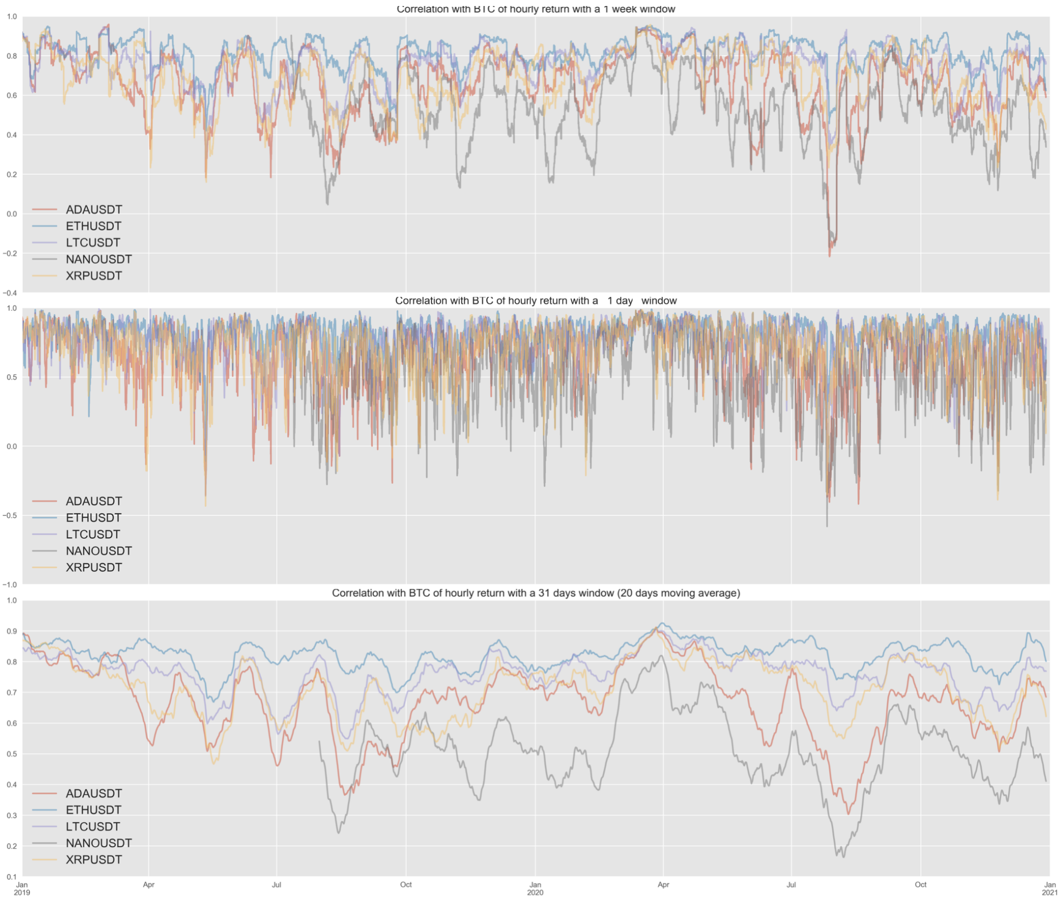

Figure 12: Correlation with BTC of hourly return with a 1 week window (2nd with a moving average)

The three graphs use different sized windows. The first uses a one-week window - so $24 \cdot 7 = 168$ times the sampling frequency. We still recognize the periods referred above: the stable period and the crash but the window size is huge. As soon as we use a smaller window, comparable to the smaller ones used in the graph 10, the data is much too noisy and detailed to fully grasp it at this scale. The last graph represents the moving average of this noisy graph.

To be able to understand and analyze higher frequency correlations, such as hourly price returns, we need to focus on more precise areas. Let's start by analyzing the spring 2019 lock-down period, which ran from early March to early May in most western countries of the world. This has necessarily impacted the cryptos market.

First of all, Figure 13 shows the exchange rate during this period.

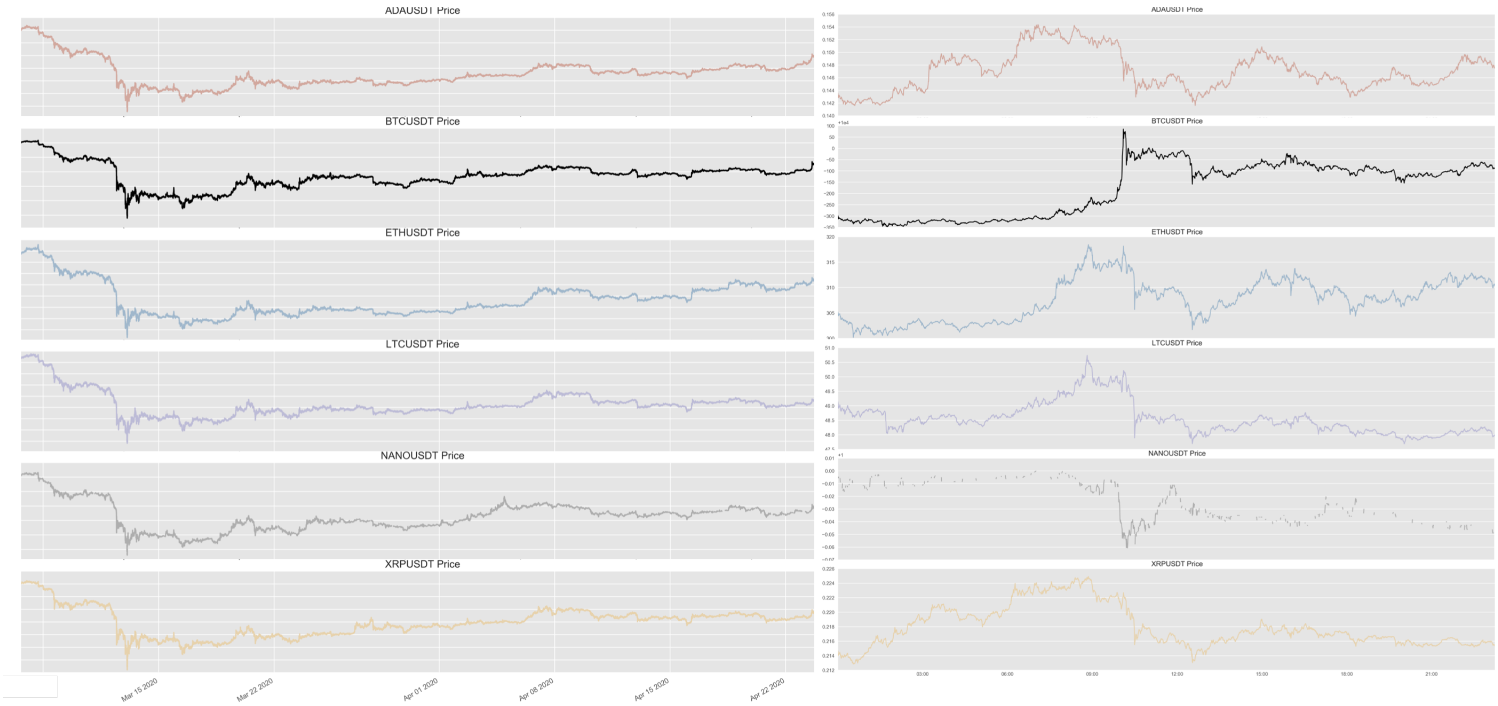

Figure 13: Prices of coins during the first world lock-down period (spring 2019) on the left and during the decorrelation period (July, 26th) on the right

We can observe that all prices move in the same way. We therefore expect to have a correlation close to 1 for all coins towards BTC in crisis periods. Correlations graphs are on Figure 14.

The first two graphs show respectively the hourly correlation with a 1 day window and its 4 days moving average. There are peaks of NANO decorrelations that we had already noticed in Figure 10. The period when the currencies are most correlated is the period when the price of Bitcoin collapsed at the very beginning of the confinement (13). But globally all correlations are above or around 0.8.

The last two graphs show respectively the correlation minutes by minutes with a 2 hours window and its 4 days moving average. What is striking on these two graphs is the sudden drop in correlation when we take a higher sampling frequency. While the correlation of the NANO with the BTC was around 0.55 using hourly returns, the correlation is around 0.2 using minute by minute returns. This comes from the Epps effect. In econometrics and time series analysis, this estimation bias, named after T. W. Epps, is the phenomenon that the empirical correlation between the returns of two different stocks decreases with the length of the interval for which the price changes are measured. The phenomenon is caused by non-synchronous/asynchronous trading and discretization effects [2].

We have identified three possible causes of this effect highlighted in the literature:

- lead/lag relationships
- asynchronous trading

It is worth saying that two assets have a lead/lag relationship if the price of one of them is statistically ahead of the other. The follower asset needs some time to embed

Figure 14: Correlations of coins during the first world lock-down period (spring 2019

the information included in the lead price. If we estimate the correlation on a smaller time scale than the characteristic time scale of the lead/lag relationship, then part of the joint movement of the two assets is not taken into account. Obviously, this pulls the high-frequency estimated correlation downwards.

Asynchronous trading refers to the randomness of market order times. Therefore, there is no chance that the prices of two assets will change exactly at the same time. Thus, the probability that the prices of two assets will change over a given period of time tends towards zero when the duration of this period becomes shorter and shorter. This leads to summing a large number of zeros when estimating the covariance of two high-frequency sampled assets. These zeros result from the nullity of at least one of the two multiplied returns. This phenomenon of asynchronous trading, which is intrinsic to order-driven markets, biases the estimated covariance towards zero without impacting the realized volatilities, resulting in an underestimation of the correlation.

Numerous solutions have been suggested in the literature to overcome these estimation biases. We will try to remove it in part 2.3. by using the Hayashi-Yoshida approach.

But first of all we always have the same concern about this noisy information. Correlations are not easy to evaluate because the data is too sensitive. So we have to focus on an even smaller period to detect high frequency correlation patterns. The analysis of the correlation crash of July 26, 2020 may be of interest.

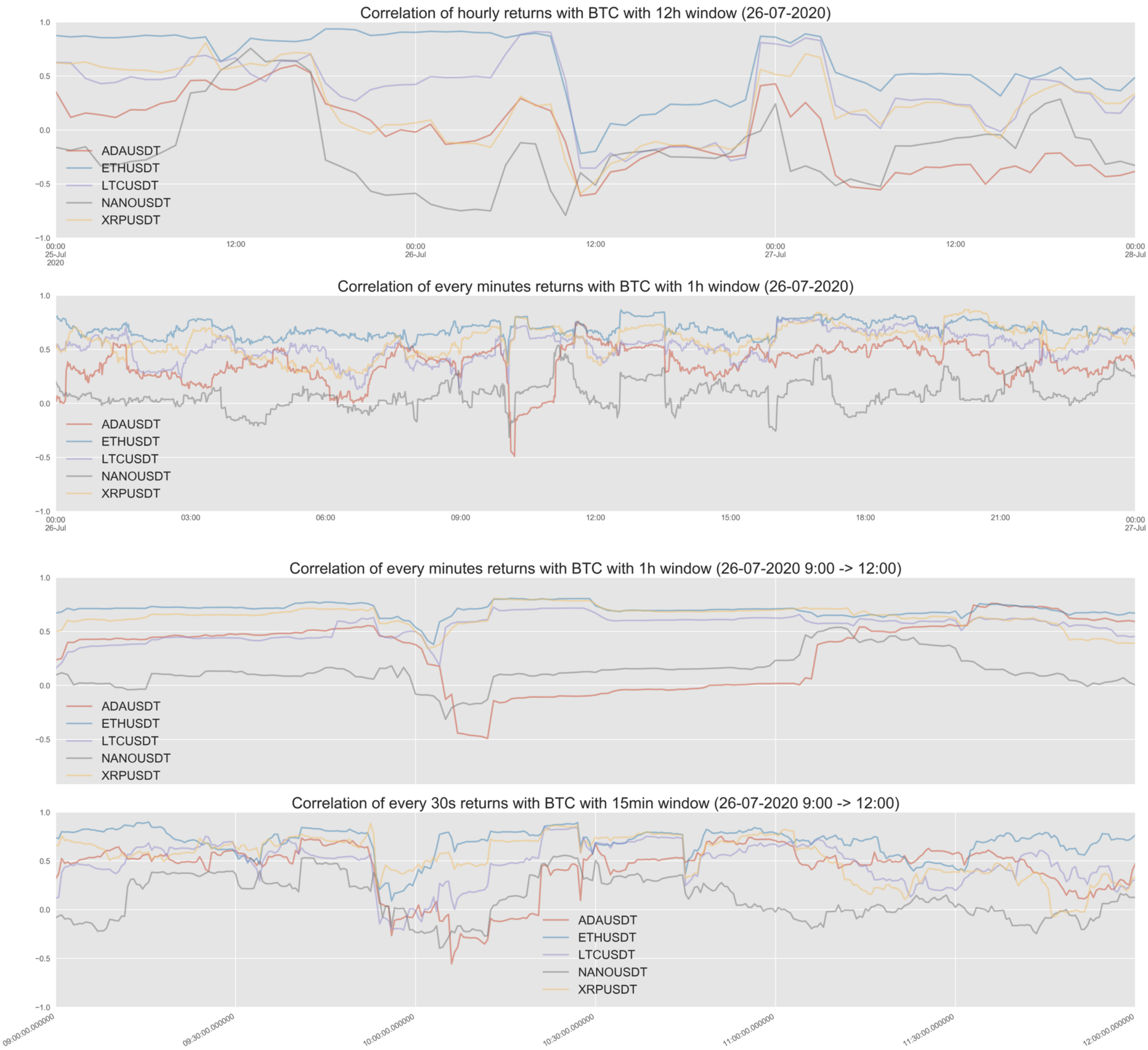

Figure 15: Correlations of coins during a decorrelated period (July 2019)

These 4 plots still depict the correlations of the 5 coins studied with respect to Bitcoin, but at different periods and with different aggregation of the data. The first graph runs from July 25 to July 28 and represents the hourly correlation.

The next two show the minute-by-minute correlation with a 1-hour window. The third graph is just a zoom of the second one, to be able to establish where exactly the most uncorrelated moment takes place.

Finally, the last one is the most precise. Here we have shown the correlation by taking a sampling frequency of only 30 seconds and a 15 minutes window. We can see that the decorrelation started at 9:50 am, going from values around 0.8 for the most correlated coins to negative values. In order to further increase our accuracy and overcome the Epps effect, we will try Hayashi-Yoshida's method to calculate the correlation at this point in time and compare the results obtained with this graph.

2.3 Hayashi-Yoshida method

Hayashi-Yoshida's estimator was presented in [4]. It is designed to solve the problem of asynchronous trading. We saw duting lectures that this estimator is both unbiased for covariance and consistent for correlation. We defined it in the course in two different ways. A first more mathematical one:

$$\hat{C}_{HY} = \sum_{i}^{n_1}\sum_{j}^{n_2}(p_{1,t_{1,i}} - p_{1,t_{1,i-1}})(p_{2,t_{2,j}} - p_{2,t_{2,j-1}})K_{i,j}$$

$$K_{i,j} = \begin{cases} 1 & \text{if } \max(t_{1,i-1}, t_{2,j-1}) < \min(t_{1,i}, t_{2,j}) \\ 0 & \text{otherwise} \end{cases}$$

$$\lim_{\Delta t \to 0} \hat{C}_{HY} = C \text{ for Gaussian returns}$$

Figure 16: Cumulated covariance estimator: overlapping time windows

And we have also seen that this is equivalent to multiplying the return on the first asset A by the latest return on the second asset B (see aggregation of data on Fig. 5):

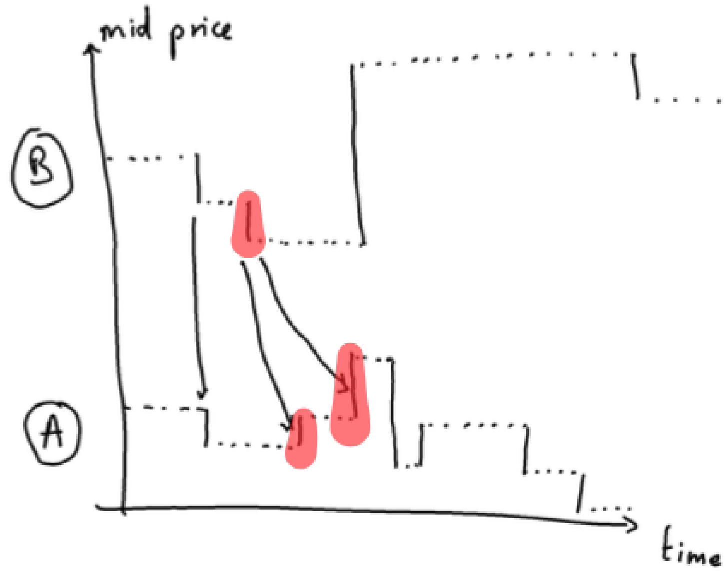

Figure 17: Hayashi Yoshida way to handle asynchonous returns

It is this last method that we used to make the following graphs. Note that as the number of trades in Bitcoin is much more important than the number of trades in other cryptos, we have aggregated it in activity time: Bitcoin trades are grouped by 10 so that there is no over-representation of Bitcoin trades compared to other trades. This also allows us to obtain higher returns for Bitcoin which initially has very low returns because of its high value.

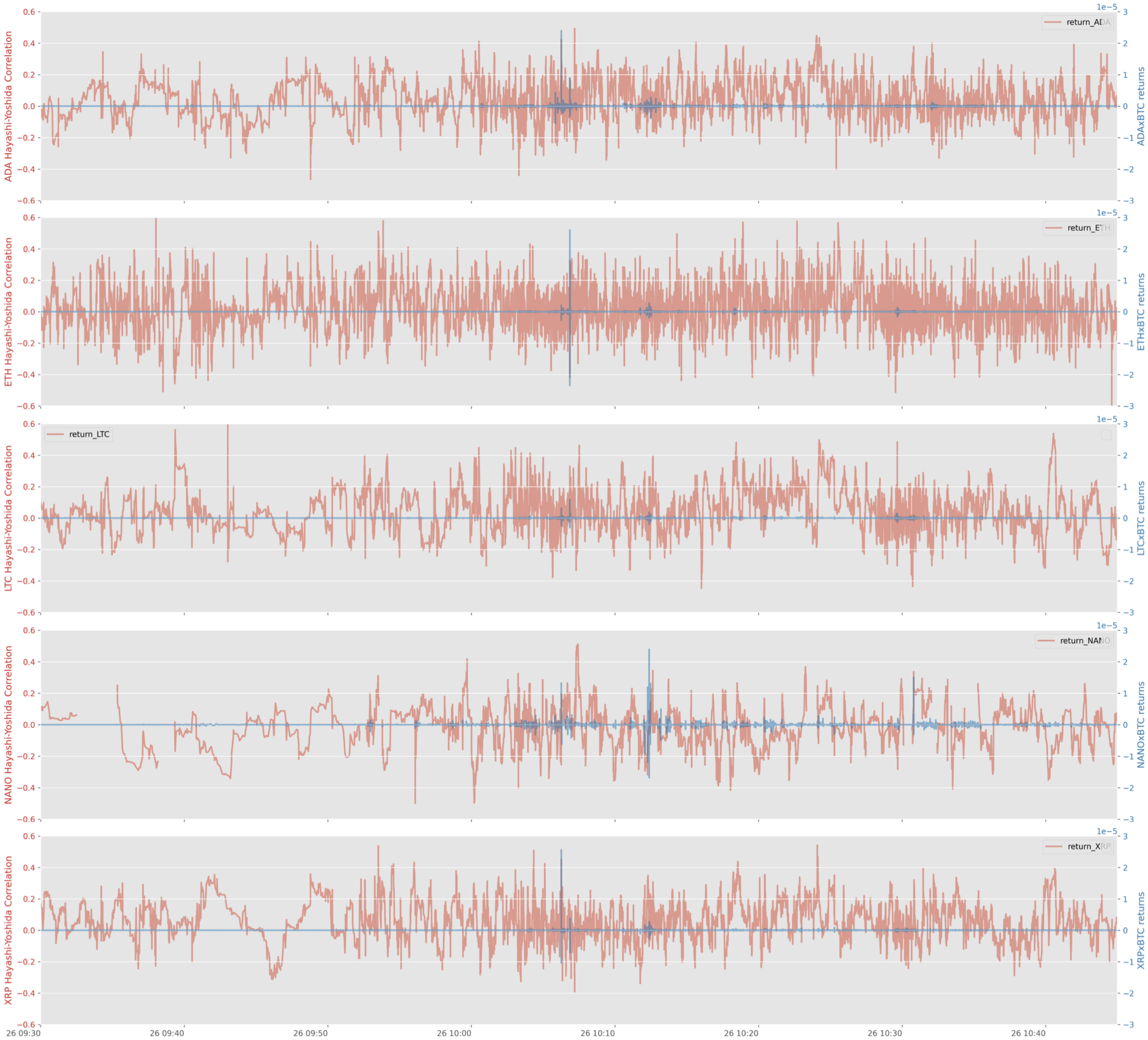

Figure 18: Hayashi Yoshida correlations (red) and aggregated multiplication of tick-by-tick returns (blue)

This figure represents the uncorelated phase that took place on July 26th, 2020 and that we observed in Fig.15. As we can see, Hayashi-Yoshida method resolves the problem of asynchronous trades. It can be noticed that some correlations (depicted in red) are more granular than others. It is due to the abundance of tick by tick trades for these coins. Moreover, all curves seem to be more granular between 9:50 and 10:20. This is about the time we noticed the decrease in correlation between the currencies and the Bitcoin Fig.15.

The markets therefore freaked out in the face of this unknown behavior, and much more trades were exchanged.

The blue plot represents the multiplication of each tick-by-tick data shown in Table 5. This time, the anomaly between the returns of the coins and the BTC is more precisely identified. We can spot that these irregular moves started, not around 9:50 as we assumed earlier, but between 10:06:30 and 10:07:00. Fig.?? just shows a enlargement of the ETH graph (the second one at the top). It shows that the anomaly occurred at 10:06:52 for this crypto.

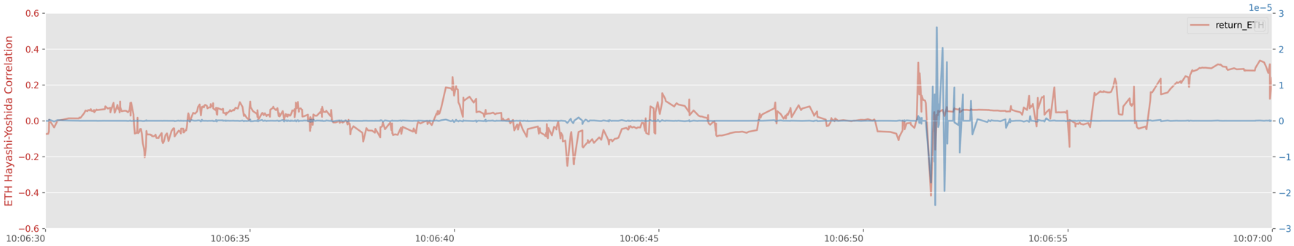

Figure 19: Hayashi Yoshida ETH correlation (red) and aggregated multiplication of tick-by-tick returns (blue)

However, the correlation is not as meaningful as earlier, since it is too noisy and much closer to 0 than with the previous method. We cannot observe, for example, a drop from a positive value to a negative value or to 0 of the correlation, as we did before. Indeed the correlation obtained through this approach seems surprisingly suffering more from the Epps effect, although the problem of asynchronous trading is solved.

In the next section, we will study the influence of trades between different crypto systems.

3 Response Functions Analysis

Response functions are frequently used in physics in order to quantify the response to a certain event. For instance, pushing a jelly to one side will result it to bounce back. In that context, response functions allows us to quantify the response value of the jelly with respect to how much we push it to one side.

The same way we can quantify the response of the bouncing jelly with respect to a push, we can quantify the response of a market with respect to a trade. If a trade in a given cryptocurrency is buyer-initiated, we consider that the price of that same cryptocurrency should increase. On the contrary, if a trade is seller-initiated, we consider that the price should decrease. We will define the response function as being positive in both of these scenario. Hence, in theory, the response function should always be positive. This type of response funcion is called a self-response function, as it measures one's own response to own action.

As mentioned in the abstract, by using response functions, we can also evaluate the impact that a trade in a given cryptocurrency has on another cryptocurrency. The idea is similar to what is described above, with a little twist : the response function of a cryptocurrency i to a cryptocurrency j's trade is defined as being positive (resp. negative) if i's price increases (resp. decreases) following j's trade when this latter is buyer-initiated (resp. seller-initiated). This type of response function is called a cross-response function.

Using this function, we are hence capable of evaluating the impact of a trade in BTC over the price of other cryptocurrencies. To this end, we have first analyzed this influence in average over time, before quantifying over more specific periods such as flash crashes. Through this analysis, we would like to estimate to what extent dominant cryptocurrencies' trades such as BTC have influence on other cryptocurrencies' price, given that they present a high value of correlation in price.

As specified in Section 1.3, we aggregated our data by binning in physical time. For this reason, we conducted our analysis of response function on the physical time scale.

Our work in this section has been greatly inspired from Henao-Londono et al. 's work [5] from 2020 regarding price response functions in correlated financial markets.

3.1 Data Wrangling

To define response functions for a cryptocurrency, we need for each time step :

- The trade sign (whether it is buyer-initiated or seller-initiated)
- The price return of the cryptocurrency with respect to time lags

We needed one trade sign per time interval. Since we are conducting the research on the physical time scale, we first aimed to conduct the research using the following trade sign definition :

$$\varepsilon(t) = \begin{cases} \operatorname{sgn}\left(\sum_{n=1}^{N(t)} \varepsilon(t,n) V(t,n)\right), \text{ If } N(t) > 0 \\ \varepsilon(t-1), \text{ If (total volume of buyer trades(t) = total volume of seller trades(t)} \\ 0, \text{ If } N(t) = 0 \end{cases}$$

where :

- t corresponds to a time step.
- n corresponds to a trade index.
- $\varepsilon(t,n)$ corresponds to the sign of the trade n with respect to the time step t.
- $N(t)$ correponds to the number of trades with respect to the time step t.

Moreover, we excluded the response functions where $\varepsilon(t,n) = 0$ like in the paper, as the price responses including them are weaker than the excluding ones (cf section 4.2 [5]).

$\varepsilon(t,n)$ can be equal to 0 when either there are no trades within the time step, or when there is the same number of buyer trades and seller trades within a time step.

The latter case occurs frequently. As showed in the notebook (which can be found in the githug repository), some cryptocurrencies showed such cases for over 4% of the time steps.

As this leads to a partial loss of information regarding the mid-price within such time steps, we proposed an alternative trade sign definition. Indeed, a better alternative may be to consider that the trade sign for a given time step corresponds to the sign of the type of trades that produced the most volume within the given time step. This way, $\varepsilon(t) = 0$ when either there are no trades within the time step, or when the volume associated with buyer trades is equal to the volume associated with seller trades (which is not likely).

Moreover, to prevent loss of information regarding the mid-price, for cases where the volume associated with buyer trades is equal to the volume associated with seller trades, we assumed that $\varepsilon(t) = \varepsilon(t-1)$ as financial markets tend to be highly self-correlated.

For time steps containing no trades, $\varepsilon(t)$ are equal to 0 and weren't considered for what follows.

This led to the following definition of the trade sign:

$$\varepsilon(t) = \begin{cases} \operatorname{sgn}\left(\sum_{n=1}^{N(t)} \varepsilon(t,n)V(t,n)\right), \text{ If } N(t) > 0 \\ \varepsilon(t-1), \text{ If (total volume of buyer trades(t) = total volume of seller trades(t)} \\ 0, \text{ If } N(t) = 0 \end{cases}$$

where :

- $V(t,n)$ corresponds to the volume of the trade n with respect to the time step t.
- $\varepsilon(t,n)$ corresponds to the sign of the trade n with respect to the time step t.
- t corresponds to a time step.
- $N(t)$ corresponds to the number of trades with respect to the time step t.

Not only is this definition more coherent (it doesn't make sense to say that the trade sign within a time step is positive in the case where there are more buyer trades than seller trades but more total volume for the latter than the former), it reduced the number of cases where *eps(t)* = 0 since there are less cases where the volume associated to buyer trades is equal with the volume associated to seller trades.

In compliance with this trade sign definition, our data achieved the form showed in Figure 20:

timestamp	trade_sign	mid_price
2018-07-25 00:00:30	1.0	0.173046
2018-07-25 00:01:00	1.0	0.172807
2018-07-25 00:01:30	1.0	0.173655
2018-07-25 00:02:00	-1.0	0.173963
2018-07-25 00:02:30	-1.0	0.173849

Figure 20: **5 first time steps of ADA following the data wrangling process**

Regarding the price returns, we computed them on-the-fly (following the definition specified in Section 1.3) since we had to do it for different range of values of time lag.

3.2 Self-response function

In this second subsection, we will analyze the self-response function for each cryptocurrency. We will use the following definition, as suggested in [5], to compute this quantity :

$$R(\tau) = E\left[\varepsilon(t)\left(r(t-1,\tau)\right)\right]$$

where :

- t corresponds to a time step.
- τ corresponds to a time lag.
- $\varepsilon(t)$ corresponds to the trade sign with respect to the time step t.
- $r(t,\tau)$ corresponds to the price return with respect to the time step t and the time lag τ.

First, we computed the self-response functions of each cryptocurrency over 1000 time steps spaced by 30 seconds, which represents 8 hours. The results are shown on Figure 21.

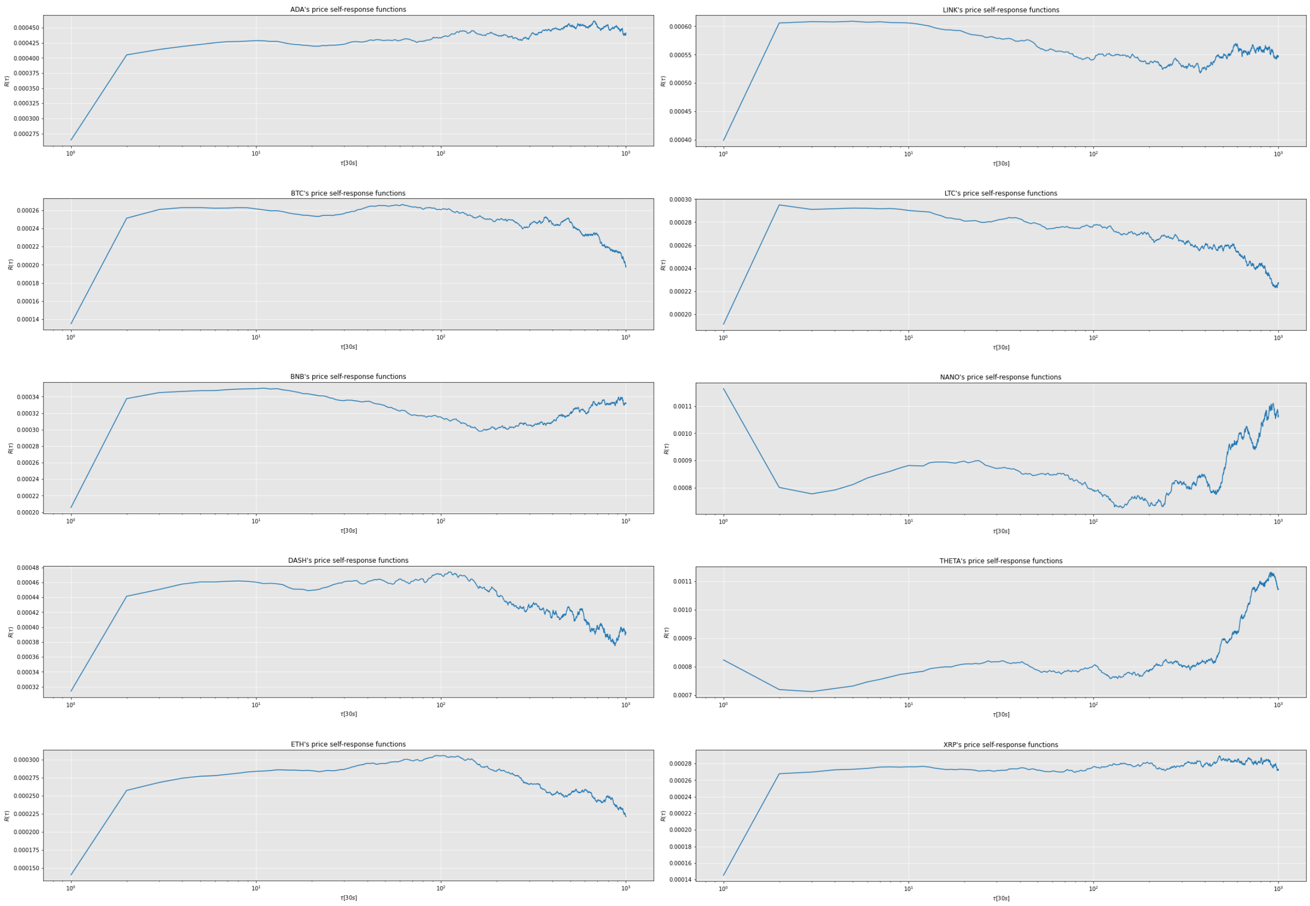

Figure 21: All the cryptocurrencies' self-reponse functions to price over 8 hours

As expected, we can observe on each of these plots that the response function is positive, and decreases over time : a trade has notable influence on the currency's price up to 8 hours after the event. Moreover, it is the smallest coins (DASH, THETA NANO) that achieve the highest response value.

In order to see the behaviour of these functions over a more extended period of time, we computed the self-response functions of each cryptocurrency over 1000 time steps, but this time spaced by 3000 seconds, which represents approximately a month. Due to limitated computational power, more granularity unfortunately couldn't be achieved. To cover a wider range of time, we were constrained to choose a wider time step. We show in Figure 22 that the results are different.

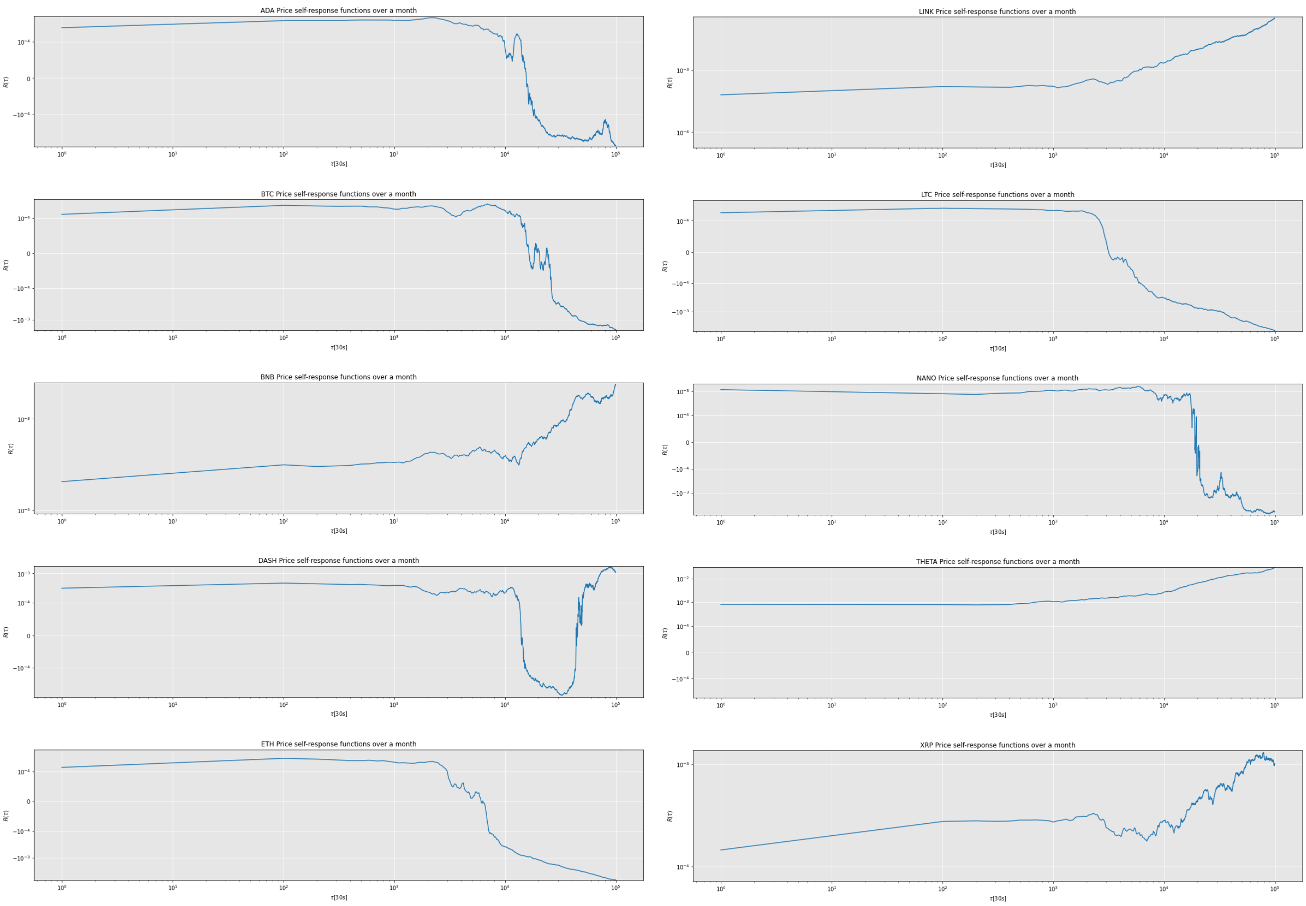

Figure 22: All the cryptocurrencies' self-reponse functions to price over a month

As showed on Figure 22, some cryptocurrencies' trades, such as BNB, THETA or Link, have a long lasting influence on its price. However, we can also observe that after a month, BTC's self-response function reaches a negative value in average, namely 10e-3. This behaviour is extremely intriguing, as it would mean that buyer-initiated trades make the price decrease in average, and that seller-initiated trades make the price increase in average, over an extended period of time. In fact, other major cryptocurrencies, such as ETH and LTC, have similar behaviour as BTC. We would expect the response functions to simply converge to 0 over an extended period of time, but that doesn't seem to be the case.

For what follows, we will only focus on the analysis of response functions over a month, as they seem to comprise most of the information that the plots over 8 hours have, above the fact that the difference in the response functions between cryptocurrencies is more noticeable when measured over a month.

3.3 Cross-response function

In this third subsection, we will analyze the cross-response function between different cryptocurrencies. We will use the following definition :

$$R_{ij}(\tau) = E\left[\varepsilon_j(t)\left(r_i(t-1,\tau)\right)\right]$$

where :

- i and j each correpond to a cryptocurrency.
- t corresponds to a time step.
- τ corresponds to a time lag.
- $\varepsilon_j(t)$ corresponds to the trade sign of cryptocurrency j with respect to the time step t.
- $r_i(t, \tau)$ corresponds to the price return of the cryptocurrency i with respect to the time step t and the time lag τ.

We refer to this function as *cryptocurrency i's cross-response function to cryptocurrency j's trades.*

Figure 23 shows all the cryptocurrencies' price cross-response functions to other cryptocurrencies over a month.

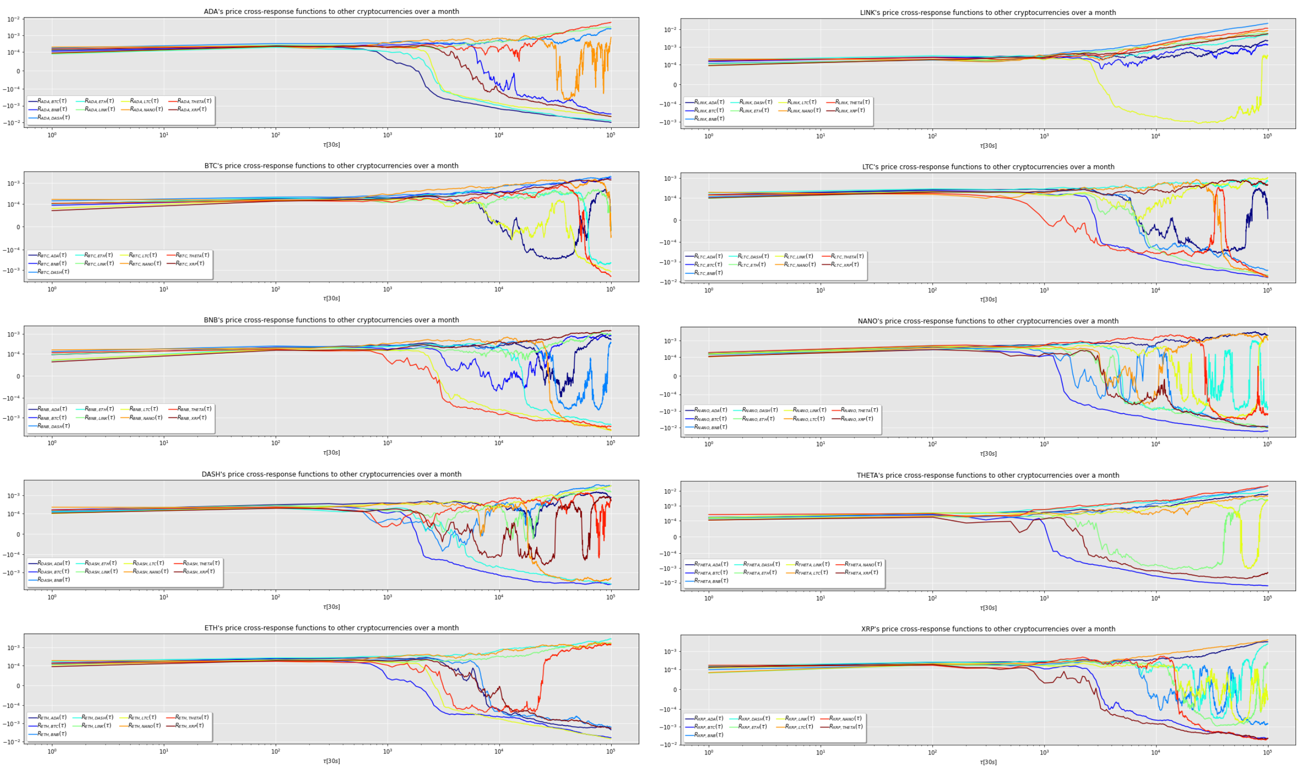

Figure 23: All cryptocurrencies' price cross-reponse functions to other cryptocurrencies over a month

Interestingly enough, all of the cryptocurrencies have on average positive price cross-response functions to all of the other cryptocurrencies' trades over 8 hours. However, we can observe that passed the period of 8 hours, this uniformity is completely broken. We can observe 3 types of curves : curves that maintain a positive value, some that oscillates around 0, and others that drops and maintain a negative value.

Curves that oscillate around 0 aren't a surprise : it simply means that a given cryptocurrency's price isn't influenced by another cryptocurrency's trade after an extended period of time.

Curves that maintain a positive value aren't a surprise either : it wouldn't be unseen for trades in a given cryptocurrency to have a long lasting impact on other cryptocurrencies, moreso if these latter are governed by the former.

However, curves that drop below 0 and maintain a negative value are really intriguing : as discussed in the previous subsection, that would mean that some cryptocurrencies are negatively influenced by other cryptocurrencies' trades. Particurlaly, ETH's price cross-response function to trades by BTC, LTC, THETA, XRP, ADA and BNB is in average negative after a week, and continues decreasing exponentially as a function of time. Similar behaviour can be seen on ADA's plot for BTC, ETH, LTC, BNB and XRP.

In particular, we observe that a consequent number of cryptocurrencies' price cross-response function to trades by BTC are in average negative and decrease exponentially as a function of time after few days. Figure 24 illustrates this matter in particular.

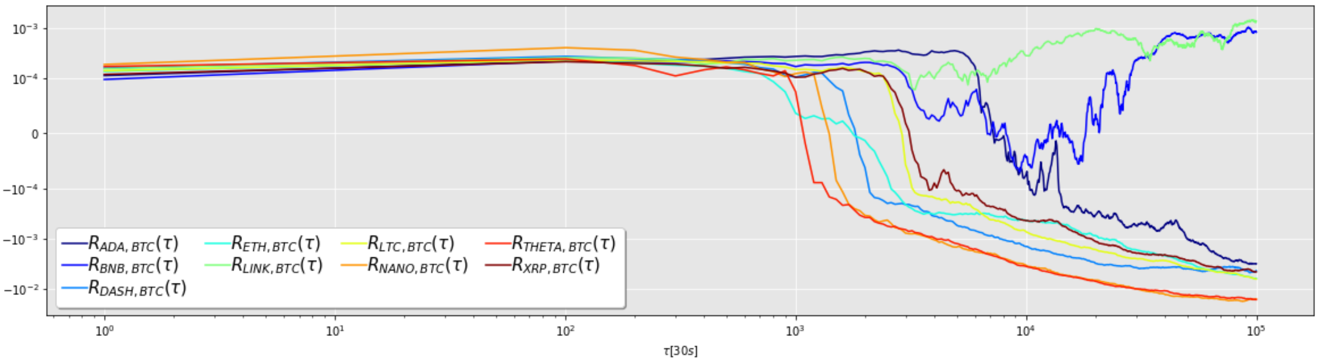

Figure 24: All cryptocurrencies' price cross-reponse functions to BTC over a month

BTC does have a strong correlation with other cryptocurrencies, and its trades do have a strong influence on other cryptocurrencies' price, but, on the contrary to our expectations, the values of the cryptocurrencies' price cross-response functions to BTC's trades are negative and decrease over time.

3.4 Cross-response function during flash crashes

In this last section, we will analyze the behaviour of cryptocurrencies' price cross-response functions to BTC's trades, but this time, by focusing specifically on a famous BTC's flash crash that occured in March 2020. Studied cryptocurrencies' price as a function of time is illustrated on Figure 25.

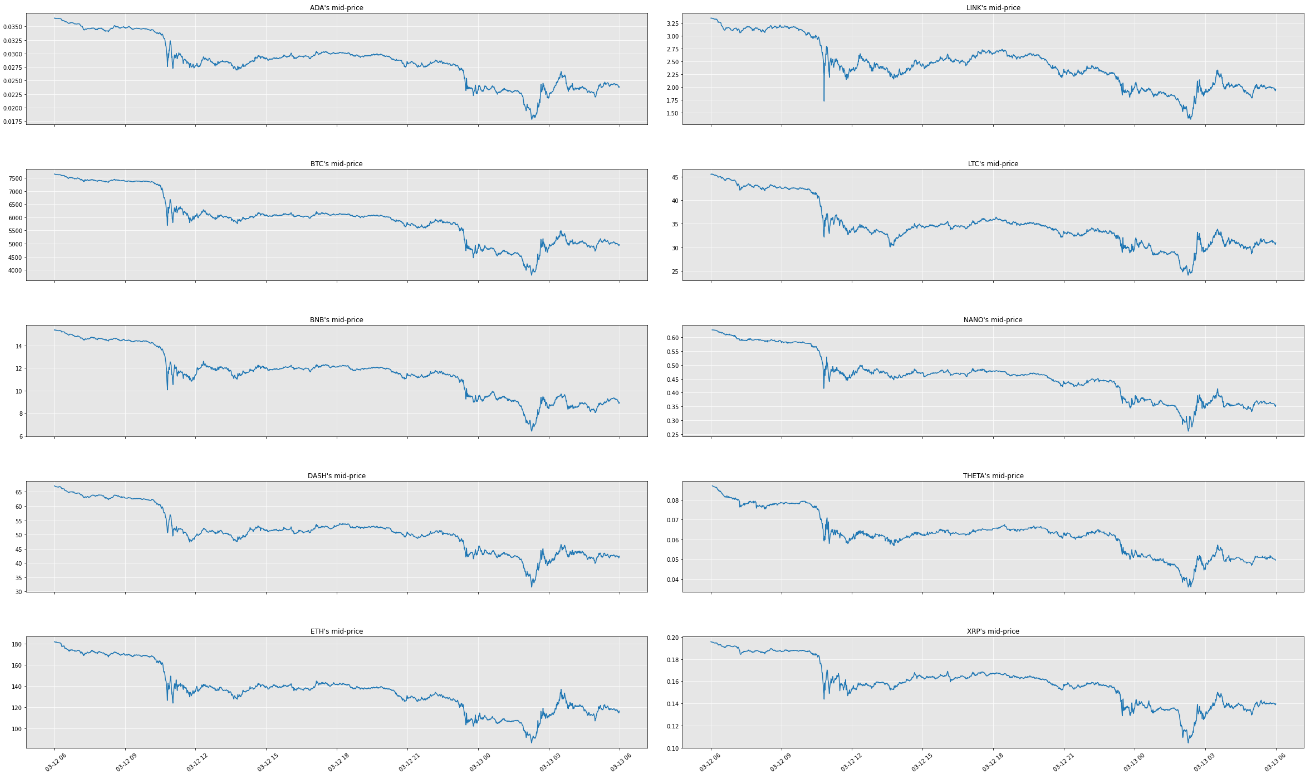

Figure 25: All cryptocurrencies' price on March 2020's flash crash

This flashcrash was caused by BTC, and caused all the other cryptocurrencies' price to drop at the same time. In such extreme conditions, what are the behaviour of price cross-response functions of cryptocurrencies ?

We computed these latter over 3000 time steps spaced by 30 seconds each, representing 24 hours.

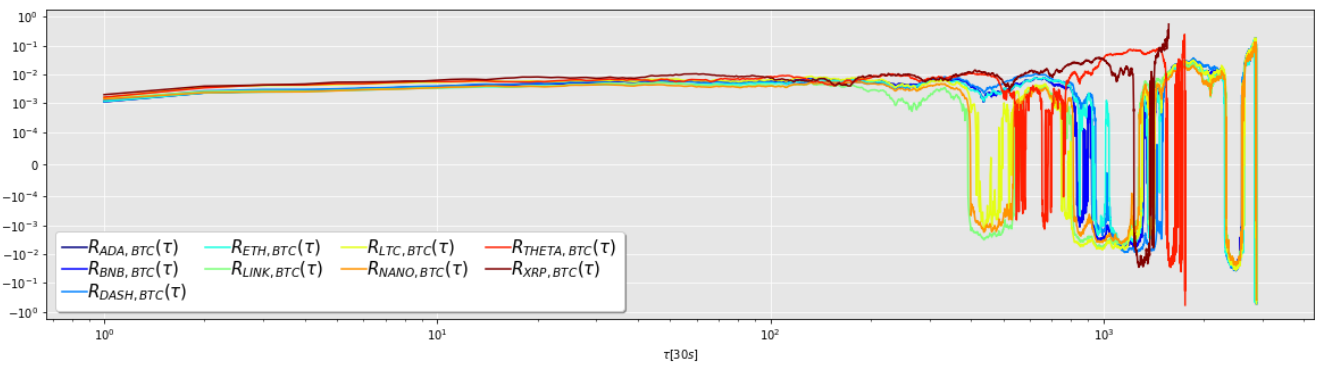

Figure 26: Cryptocurrencies' price cross-response functions to BTC's trades over 24 hours on March 2020's flash crash

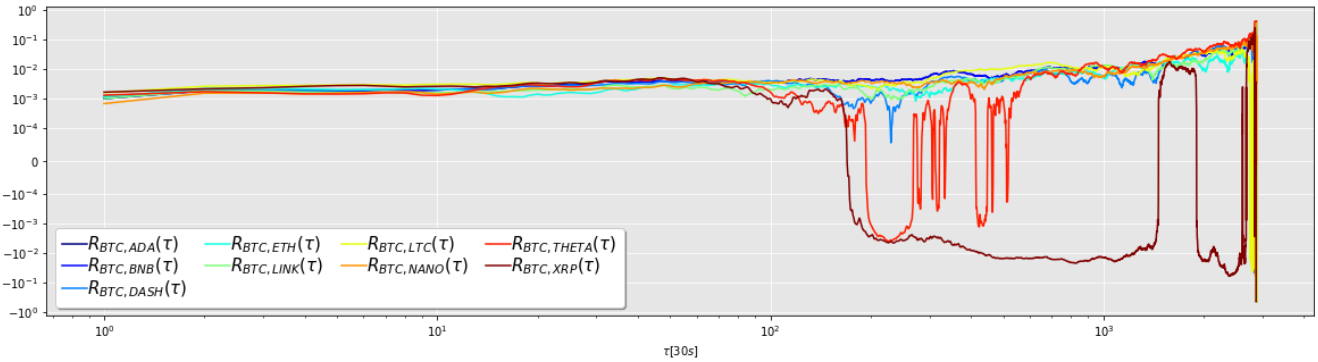

Figure 27: BTC's price cross-response functions to other cryptocurrencies' trades over 24 hours on March 2020's flash crash

Figure 26 shows us results that are different to results obtained in Section 3.3 while computing the same functions over a month : we can observe that most of the curves are oscillating around 0, before reaching a response value of 1 or -1 - values never seen in any of the previous plots. These values are reached in average after 8 to 10 hours.

On the contrary, Figure 27 shows us that during the flash crash, most cryptocurrencies's trades have an influence on BTC's price.

4 Conclusion and Future Work

In the first part of the analysis where we used two different methods to analyze the correlation between cryptos and Bitcoin, we have convincing results showing a clear correlation using low frequency data. We have also seen that some external factors such as a worldwide lockdown period must have had an impact on the cryptos market because they were found to be more correlated than usual. Then we analyzed a period when cryptos were decorrelated with Bitcoin. Using Hayashi Yoshida's method, which allowed us to analyze the correlation at the highest possible frequency - tick-by-tick data - by getting rid of the problem of asynchronous trading, we were able to highlight anomalies in the returns that caused this decorrelation. Unfortunately, we cannot overcome the Epps Effect in this high-frequency analysis.

In the second part of the analysis, we quantified cryptocurrencies' trades influence over their own price and others' using response functions. The results weren't as concluding as we would have thought, but still, judging from our analysis and observations, we can affirm that cryptocurrencies do influence each other's price by trading. In particular, BTC's trades do have an influence over most of the studied cryptocurrencies, even in the long run. Finally, we have seen that extreme conditions such as flash crashes in the world of cryptocurrencies is reflected by extreme behaviour of these functions.

In the future, we could run the same analysis with data that is less aggregated, which could lead to results that are more granular and potentially a better reflection of the reality. Moreover, regarding the work done in Section 3, we could introduce the number of trades into the response functions in order to analyze response functions not only as functions of the cryptocurrency's price but also as functions of the cryptocurrency's activity. This may lead to new results, and meaningful insights.

References

[1] URL: https://python-binance.readthedocs.io/en/latest/binance.html.

[2] URL: https://en.wikipedia.org/wiki/Epps_effect.

[3] *CoinMarketCap Ranking of Cryptocurrencies according to their market capitalisation.* URL: https://coinmarketcap.com/.

[4] *On covariance estimation of non-synchronously observed diffusion processes. Bernoulli.*

[5] *Price response functions and spread impact in correlated financial markets.* URL: https://arxiv.org/abs/2010.15105.