M.Sc Economics and Finance, Major in Finance

Chair of Asset Pricing

Behavioural biases in Crypto Markets: an empirical analysis

BORRI NICOLA

SUPERVISOR

CYBO OTTONE ALBERTO

CO-SUPERVISOR

DURANTE MATTEO - 762031

CANDIDATE

Academic Year 2023/2024

Table of contents

Introduction

In mainstream economic literature theories, agents are assumed to behave with perfect rationality when it comes to managing their finances and make economic decisions. Vilfredo Pareto argued that humankind had different human extensions, among which: *homo oeconomicus homo ethicus and homo religious;* and each of its governing choices on its own sphere[1]. *Homo oeconomicus* translates as "economic man" or "rational economic agent" that represents a hypothetical individual as such all agents in economy are assumed to behave. *Homo oeconomicus* is assumed to make decisions that maximize their own utility or well-being. They have access to complete information about all available choices and assess outcomes, and they consistently choose the option that yields to the highest expected utility. This utility is often measured in terms of monetary wealth, and choices are gauged on a rational calculation of costs and benefits. *Homo oeconomicus* is motivated solely by self-interest and is not influenced by altruism, empathy, or social considerations. The concept of *homo oeconomicus* is rooted in the fundaments of economic theories. particularly in neoclassical economics, which dominated economic thought for much of the 20th century. It is used to build economic models that provided valuable insights into various aspects of economics, such as consumer behaviour, market equilibrium, and resource allocation. This argument is still considered a staple in academy, as economists and professionals rely on literature to build more advanced and tailored framework. It is even possible to argue that the wider area of economic research depends on past economic research by analysing economic papers citations, as the biggest

[1] Pareto, V. (1906). Manuale di Economia Politica [Critical Edition, Eds. A. Montesano, A. Zanni, & L. Bruni]

proportion of them is directed towards other economists or research within that field for hundreds of years [2].

My work wants instead to have ground on a different view of economic choices, choosing a Behavioural Economics approach to explain a real market phenomenon. Behaviourists argue that real-world individuals often deviate from the assumptions of perfect rationality, consistent preferences, and purely self-interested behaviour. This rising field of research relies on bounded rationality, that acknowledges individuals' cognitive limitations that do not allow them to always make a perfectly rational choice.

The final aim of this work is to understand whether a Behavioural Finance perspective could shed a light on the unusual reactions observed in cryptocurrencies markets and investigate the efficiency of these assets' prices. Focussing on the FTX exchange crash on November 2022 and the subsequent market panic, this work wants to inquire if behavioural components could be priced on the markets and if this finding could help explaining the trading behaviours observed and understanding their reliability.

[2] Aistleitner, M., Kapeller, J., & Steinerberger, S. (2019). Citation patterns in economics and beyond. Science in Context, 32(4), 361-380. https://doi.org/10.1017/S0269889720000022

Chapter 1

Academic review and presentation of the events

1.1 Rational choices theory

Mainstream academy believes that investors, as economic agents, make their choices in a perfectly rational way. The concept of perfect rationality may be summarised with the definition of *homo oeconomicus*. It is not straightforward to identify the pioneer of this concept, but we can indeed state that it set grounds to further economic theories and research. In this work, I will follow the definition of *homo oeconomicus* used by the Italian economist Vilfredo Pareto in his "Manuale di economia politica, con una introduzione alla scienza sociale"[3] (Economic policy manual with an introduction to the social science). The *homo oeconomicus* is a rational individual who is assumed to take all the economic related decision. Pareto argues that our personality could be divided according to specific work fields, and that a different extension of ourselves would arise when making a choice related to the task we are undertaking. Among these, we *find homo oeconomicus*; as well as *homo ethicus* and *homo religious*. The "economic personality" described by the economist shows some characteristics believed to be shared by all the humankind.

The first characteristic is utility maximization. The insight is that when people make economic decisions, they follow the primary goal of maximising their utility or satisfaction. According to this axiom, an economic decision is primary based on the assessment of costs and benefits analysed according to a utility function, which is assumed to be based on financial wealth in most of economics frameworks. What the agent performs is then a rational calculation that compares a monetary sacrifice and the

[3] Pareto, V. (1906). Manuale di Economia Politica [Critical Edition, Eds. A. Montesano, A. Zanni, & L. Bruni]

financial gains achieved as of result. The utility function can be seen as a tool that allow agents to gauge their preferences, and their objective is to maximize it. An agent would choose the option who gives the most utility given a cost and its preferences. Investing is an example of economic choice. According to this axiom, using personal wealth as measure of utility, the investor will decide to invest if they believe the wealth arising from this choice would be higher than the wealth sacrificed to take it.

The second characteristic of a rational agent is consistency. People are expected to exhibit perfect consistency in their preferences and choices. A choice is consistent if it is transitive. Following an example, it means that a preference of A over B together with a preference of B over C implies a preference of A over C.

The third characteristic of *homo oeconomicus* is self-interest. In this framework, agents are solely motivated by their self-interest. Altruism, empathy, and social considerations are not taken into account: the utility maximisation is related to the unique personal sphere of the person. As of this definition, the same agent that uses personal wealth as measure of utility would not be much interested in ESG investing if its returns were not higher than a traditional approach, as their function does not take into account that value. In order to account for such preferences, the personal utility function should be changed.

The last trait defining rationality is perfect information processing. It is assumed that people have access to a vast amount of information that are not biased nor erratic, that this access is instantaneous and that they are able to assess them perfectly without making mistakes, yielding to an optimal decision making. Therefore, any news, any speech, any document release, is instantly captured and process by investors, that exactly know how to react.

This framework is a building block for finance and economics in general, and mainstream academy did not deviate from these claims. This is not something peculiar to this specific theory. It can be inferred that economic academy mostly relies on past economic papers and does not take into account the broader development in academy. As reported by

Aistleitner et al (2019)[4] “economists are comparatively less inclined to import findings from other disciplines and also have less trust towards interdisciplinary approaches and research strategies”; to which is added an average old age of imported references.

1.2 Daniel Kahneman and Amos Tversky's groundbreaking research

The intersection of economy and other subject is then an area mostly unexplored, but it is still possible to find thriving research fields that base their conclusions on different sources. One of the most successful papers that shows the brilliancy of merging economy with other disciplines is "Prospect Theory: An Analysis of Decision under Risk" by Daniel Kahneman and Amos Tversky[5]. To understand the calibre of this pioneering work, this study not only granted the Nobel Prize in Economic Sciences to Kahneman in 2002 but it also is the most cited economic paper with over 80000 citations![6]

Their work, using an experimental approach based on surveys and tests, had the aim to question the assumption of economic rationality in decision making situations under risk. Their work, rooted on psychology, has shown that people are not as rational as *homo oeconomicus* and indeed lack some of its characteristics. The early cited paper focussed on analysing complex decision situation when the future consequences are uncertain: when agents face risks. In their experiments, people do not appear to perfectly understand the law of probabilities, nor behave according to a utility maximisation process. What instead is inferred is the reliance on some heuristics or rule of thumb, easier to apply rather than complex calculation that *homo oeconomicus* was assumed to handle fast paced and without any difficulty.

These findings opened to a new lively research area called “Behavioural Finance”. The idea is to apply insights from psychology to study and deeper understand the function of economy and financial market. Interpreting and analysing data is key for a sound investing activity, and evidence of poor performance in performing these tasks could help

---

[4] Aistleitner, M., Kapeller, J., & Steinerberger, S. (2019). Citation patterns in economics and beyond. Science in Context, 32(4), 361-380. https://doi.org/10.1017/S0269889720000022
[5] Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263-291. https://doi.org/10.2307/1914185
[6] Prospect theory: An analysis of decision under risk. JSTOR https://www.jstor.org/stable/1914185

explain real market phenomena such as bizarre fluctuations like the one observed in the crypto market in the month on November 2022.

1.3 Rational pricing model

Traditionally, behavioural finance models are developed as expansion to existing models, therefore an introduction to rational models is needed to successfully explain their rationale. Investors look after financial assets as they produce cash flows. Investors are willing to sacrifice part of their wealth to obtain future cash flows. These cash flows are uncertain as they depend on several factors that are related in a complex manner: interest rates, energy prices, product demands, competition, company projects and events, political developments, etc. Assessing a financial asset becomes more challenging as the uncertainty grows. As of that, academy and practitioners have developed a vast amount of models, ranging from elementary to more articulated ones, to specify the pricing process of an asset. To proceed with the analysis, it is useful to introduce a first pricing model:

$$\boldsymbol{P_0} = \sum_{\boldsymbol{t=1}}^{\infty} \frac{\boldsymbol{E[CF_t|I_0]}}{\boldsymbol{(1+r)^t}}$$

According to this simple yet universal pricing model, investors forecast cash flows (CF_t) conditional on news (I_0), and discount them according to a proper risk factor (r); the estimated value arose from this process will be the price that conducts their trades. Trades of different investors are pooled together in a financial market that will accommodate every trade, and therefore opinion. Prices built according to this process will then reflect investors' perception about future cash flows and risk conditional on the current available information. As trades are matched in the market, an equilibrium asset price will be reached: a price that will satisfy both the buyer and seller of the asset; and will allow movement of wealth. Financial markets are successful when they satisfy their duty of allocating money, driving them from agents that have excess wealth to agents that need liquidity and can better use it; price is a key element for their functions.

The idea behind this simple model is that expectations are rational: agents understand the environment and base their buying and selling decision on a mathematical model. By observing a price is then possible to extract what the investors beliefs and risk aversion are. If investors acted following this framework, it could then be assumed that choices satisfy the logical axioms that underpin Expected Utility Theory (EUT).

A different dimension of complexity is added when considering information updating. Information regarding financial asset is continuously provided to investors as of public policies and news. Neoclassical academical finance assumes Bayesian Updating: when new evidence becomes available, prices are statistically optimally updated based on the new expectations. According to Bayesian Updating, when investors receive new information, this is synthesized in a signal. This signal is needed to update the precedent price according to the quality of the signal: an unclear signal will not contribute much to the price as a clear one does. According to EUT, investors do not have any bias toward news (not pessimist nor optimistic); and they are able to correctly understand the quality the signal.[7]

1.4 Efficient market theory

Before exploring the behavioural finance realm, one last building block of financial theory must be introduced: efficient market hypothesis. This idea was first introduced by Eugene Fama[8] in 1970. The idea is that financial markets are highly efficient, they incorporate information in a fast-paced way and prices reflect that information. Three hypothesis of market efficiency were presented: Weak; Semi-strong; and Strong. Fama discusses a weak-form efficient market where the only information the prices reflect is the one about past trading information such as historical prices, return and volumes. The Semi-strong form asserts that asset prices consider and incorporate all information that is public, like news, financial statements, and speeches. The Strong hypothesis concerns all possible information about a specific asset, even insider information that is not known to the public and therefore should not be used by traders. In his review, the economist gives robust evidence for the first two forms in traditional financial markets but has not enough evidence for a Strong form. The major conclusion that comes with this work is that is extremely difficult to beat the market (stressed in case

[7] The implications of this updating will be further discussed in Chapter 2

[8] Fama, Eugene. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." Journal of Finance. 25:2, pp. 383-417.

of transaction costs) with active trading strategies, and that diversification is key for a sound portfolio.

The present work does not want to challenge these last statements but wants to highlight another finding already discussed by the researcher now more than 50 years ago. Fama's work, even is strongly supporting the efficient market hypotheses, acknowledges that there are some events and patterns where this property does not seem to be consistent. He even discusses some possible explanations for these anomalies: one is information asymmetry; the second, the pillar of this work, is behavioural biases.

Following the years, many critics have been directed towards EMH, backing the idea that prices are at least predictable and do not follow the so known "random walk". One of the salient critics backing a behavioural explanation of the market is moved by Burton G. Malkiel, author of the best seller "A random walk down Wall Street". In one of his papers, he investigates anomalies of the markets and "will describe the major statistical findings in the stock market, as well as their behavioural underpinnings"[9]. Phenomena such as Short-term momentum, seasonal patterns, and specific event of irrational prices like the dotcom bubble of late 90s are discussed under a behavioural aspect and a sound explanation is given.

It is then possible to say that a behavioural component helps understanding the evolution of prices and market itself by explaining its inefficiencies; and investors may not be all *homo oeconomicus* as previously assumed.

1.5 When is a market efficient?

It has already been discussed about the importance of efficiency for the correct achievement of financial market purpose, and the presence of forces that do not allow financial markets to be fully efficient. Irrational prices could threat the willingness of investors to finance a project and therefore not fulfil the main purpose of financial market. A quote of the renowned economist John Maynard Keynes says, "Markets can remain irrational longer than you can remain solvent". Efficiency implies that the securities are

---

[9] Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. Journal of Economic Perspectives, 17(1), 59-82.

always correctly priced and there is no room for arbitrage: prices are stable, and money revolves around.

As information leads to updates of prices, a coherent evolution must then be observed to allow the market to fulfil its social goal. Academy has then defined criteria that can help identifying an efficient market, a market sounds where investors will not be threatened by frequent bizarre events. Academy defines a market efficient if either:1) all investors never make any mistakes in their calculations; 2) people can make mistakes, but the mistakes are random, and therefore they cancel out in the marketplace; 3) Some investors make correlated mistakes due to behavioural biases, but sophisticated and unbiased investors (arbitrageurs) trade to exploit any mispricing, thus bring prices back to the correct levels swiftly.[10]From the precedent paragraph it should be clear that the first two statement cannot be accepted, the attention is therefore stressed towards the last one.

To understand if financial markets are efficient, evidence of "price correcting" traders needs to be found. An interesting case study that backs this view and shows the relevance of academy at the same time is "The January effect" analysed by Haugen and Lakonishok.[11] They discovered a surprising seasonal anomaly related to the month of January in which return were observed to be higher when compared to other months of the year. The idea is that this finding is related to tax incentives or some kind of behavioural bias, therefore from this work we can then affirm that there was some kind of common mistake made by investors that led to stocks sold in December to be rebought in January. What is interesting to assess is the reaction of markets after this discovery: quoting Burton G. Malkiel[12], "The so-called "January effect" in which stock prices rose early January, seems to have disappeared soon after it was discovered". It can be inferred that even if some investors would continue with their practices to save taxes, some kind of unbiased investor, spotting the opportunity thanks to academy, acted in correction removing the noise affecting the market.

---

[10] Bodie, Z., Kane, A., & Marcus, A. J. (2021). Investments (12th ed.). McGraw-Hill Education.
[11] Haugen, R. A., & Lakonishok, J. (1988). The Incredible January Effect: The Stock Market's Unsolved Mystery. Homewood, Ill.: Dow Jones-Irwin.
[12] Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. Journal of Economic Perspectives, 17(1), 59-82.

With that example, it should be clear that prices in financial markets are close to efficient, as whenever a common anomaly is discovered another participant of the market would move to exploit that opportunity.

1.6 Behavioural Finance

If markets behaved according to a rational model, we would not observe unique events like the one just shown. As of the last example, it is indeed clear that the market can sometimes deviate from pure rationality, but nonetheless be efficient even if some correlated mistakes are made in the market. Behavioural Finance is that branch of finance that studies fallacies of *homo oeconomicus* decision making that is shown in the market. Patterns related to excitement, fear or even football matches[13] are analysed under a psychological perspective to understand market reactions to news or even exogenous events. Behavioural Finance then offers explanations to events that cannot be catch by a rational model, leveraging the essence of human beings whose choices may not be perfect; it assesses the attitude and reaction of investors as human beings, using a psychological point of view. It does not have to be framed as a completely different explanation of prices, but as a way to account tweaks for investors' nature. Meanwhile some biases are not harmful for the markets, other errors in human judgment could lead to moments of panic and disasters. Some biases are relevant and have been shown to have an impact on prices, other are not even significant when taking into account trading commissions. In a world where financial markets are living in symbiosis, even a short-framed downturn could lead to a catastrophe. Having a behavioural model that helps explaining event not covered by traditional econometric tools is a useful resource for sound professionals to avoid spiralling up or down into irrationality. Behavioural Finance is a rising branch of academic research as both the insights and implications that it can give to financial markets, and after decades of studies we can indeed state that different extensions of human thinking are reflected in the market, and that decision making is not limited to the sole the rational sphere.

[13] Edmans, A., Garcia, D., & Norli, Ø. (2007). Sports Sentiment and Stock Returns. Journal of Finance, 62(4), 1967-1998 shows market underperformance on days following a National team loss

1.7 Biases in cryptocurrencies market

Having defined what behavioural finance is and stated that financial markets are on average efficient, the aim of this work is to assess whether cryptocurrencies markets show the same properties as the traditional financial markets and investigate their performance under a behavioural framework. Cryptocurrencies market offers a big opportunity for behavioural finance scholars for different reasons, first of which is the volume of the market. As previously mentioned, market efficiency can be preserved even in presence of behavioural biases thanks to unbiased sophisticated investors. This is achievable in traditional financial markets as their scale allows for a multitude of this investors and their interest to join. The smaller crypto market could show increased significance of movements due to behavioural biases as the increased a single trade has more weight when compared with the total market. Another important feature of cryptocurrencies market is its players. In one of his paper Al-Mansour[14] shows that "investors' choices of selecting the types of digital currencies are affected by other investors' choices and therefore, will significantly affect their investment decisions" and "investors behave as speculators in the cryptocurrency market". The rather qualitative work of Al-Mansour set ground for a prolific behavioural finance analysis. His analysis wants to assess the driver of the investment decision in a cryptocurrency environment regressing the three decision's common behavioural factors: Herding, Heuristics and Prospect. The findings then show that: other investors' decision have an impact on the own investment decision (herding); investors base their decision on previous trades schemes and following to their (thought) superior knowledge of the market (heuristics); investors are more likely to make riskier trades when they have just made a profitable one (Prospect). This analysis, ran through a sample of Arab crypto investors, shows that cryptocurrencies market may be led by noise traders, traders that focus their operations on news rather than fundamentals. The aim of this work is then to find some evidence of

---

[14] Al-Mansour, Bashar (2020). Cryptocurrency Market: Behavioural Finance Perspective. Journal of Asian Finance, Economics and Business , 7(12), 159-168

noise investors biases, supporting the idea of markets deeply influenced by behavioural biases.

1.8 The FTX events

To allow the reader to understand the upcoming analysis, a background of the event object of this thesis needs to be given. During the month of November 2022, the cryptocurrency market was shocked by the downfall of one of its major exchanges: FTX Trading Ltd.

FTX Trading Ltd. was a cryptocurrency exchange founded in 2019 by Sam Bankman-Fried and Gary Wank. FTX's operations began within a trading firm founded by Sam Bankman-Fried, Caroline Ellison, and other former employees of Jane Street in 2017 called "Alameda Research". Later, they extended their presence in the cryptocurrency market with various acquisitions. For instance, in August 2020, FTX acquired Blockfolio (crypto currency tracking app) for $150 million. In July 2021, they were able to raise $900 million at an $18bn valuation. [15]

FTX Trading Ltd faced a severe crisis supposedly triggered by a CoinDesk article and the leak of Alameda Research's balance sheet[16]. An analysis marked insufficient liquidity, prompting customers to withdraw $650 million on November 7th. The value of the exchange's profit-sharing mechanism, ftxToken, plummeted by 90%. Despite initial talks of acquisition by Binance on November 8th, the deal collapsed after Binance scrutinized FTX's financial condition[17]. FTX's founder, Sam Bankman-Fried, disclosed an $8 billion shortfall to investors, foreseeing bankruptcy without additional capital. FTX restricted customers from withdrawing funds on Nov. 8 by removing the option online, leaving hundreds of thousands of customers unable to access their money. The aftermath affected the broader crypto industry, with Bitcoin, the major cryptocurrency, experiencing a 19%

---

[15] Ferreira, P. (2023, May 9). The FTX Full Story: All You Need to Know. Finance Magnates. https://www.financemagnates.com/cryptocurrency/the-ftx-full-story-all-you-need-to-know/

[16] Nelson, D., & Baker, N. (2023, September 20). Breaking Down the Infamous Alameda Balance Sheet. CoinDesk. https://www.coindesk.com/policy/2023/09/20/breaking-down-the-infamous-alameda-balance-sheet/

[17] Keller, L. (2022, November 10). Bitcoin Hits Lowest Price in 10 Years Following Binance, FTX Acquisition. Forkast News. https://forkast.news/bitcoin-lowest-price-years-binance-ftx-acquisition/

drop since November 8th. JPMorgan Chase warned of a potential industry-wide "cascade" effect, anticipating deleveraging and company failures. [18] The Securities Commission of the Bahamas froze FTX's assets, and despite assurances about FTX US's liquidity, voluntary bankruptcy proceedings were initiated on November 11 for most affiliated entities, including the U.S. exchange.

FTX crash was due to mismanagement of funds and lack of liquidity and the large volume of withdrawals. Subsequent fears of FTT (FTX's token used to inflate their balance sheet that raised the first doubts that sparked the collapse) price decline led to increased customer withdrawals, resulting in a liquidity crisis.

1.9 The impact of a single piece news

It is easy to understand the impact of a single piece of news on FTX, as its asset were shown to not be backed with any liability, but this does not explain the reaction of the broader market. It is true that FTX was one of the biggest players of the market, but there was no evidence that its collapse would have impacted the value of competitors and other digital assets. As a matter of fact, at the time of writing, cryptocurrencies markets are still operating and have got the attention of the wider public as well[19]. Looking back at it then, there was no direct connection between a single market player and the market as a whole. As reported by JP Morgan, there could have been troubles related to a spiral of margin calls, margin calls linked to Bitcoin, a different cryptocurrency. A disaster was then foiled, but if the market was sound and rational, there would not even have been a threat. Following a behavioural finance view, we could see cryptocurrencies market crash as a reaction to new facts that were only relevant for a specific business and should have not influenced the whole market as it is,overreacting towards news that had international media coverage. Having introduced the fundaments, the rationale, and the background of behavioural finance in cryptocurrencies markets, in the next chapters, using a quantitative

[18] Bambrough, B. (2022, November 13). JPMorgan Reveals Shock Cascade Bitcoin Price Prediction After Stunning FTX Meltdown. Forbes. https://www.forbes.com/sites/billybambrough/2022/11/13/jpmorgan-reveals-shock-cascade-bitcoin-price-prediction-after-stunning-ftx-meltdown/

[19] Chipolina, S. (2024, January 23). Bitcoin price falls 15% following launch of ETFs. Financial Times. https://www.ft.com/content/99e406f0-c5af-4a67-9937-afb1ea885565. The impact of the event is beyond the scope of this work even if it shows a similar pattern

approach, this work wants to present evidence of specific biases toward news for crypto investors using the FTX events as a sample, shedding a light on the forces moving cryptocurrencies prices.

Chapter 2

Presentation of data and methodology

2.1 News under a behavioural focus: strength and weight framework

News is a crucial driver of the market. Flows of information are constantly priced in financial market as prices, even according to the simple pricing model reported in the previous chapter[20], are updated to reflect this new information available. The different efficient market hypothesis accounts for different levels of information (trading information, publicly available information and all the information) but they do not question Bayesian updating and rationality: every information is updated according to the expected utility theorem. By this process, a signal is extracted from the new information, and the price of the asset is adjusted according to its quality. As of that, information that gives a weaker signal should be weighted less than information that gives a reliable signal about the new asset value.

Let the investor have a common prior so that $\boldsymbol{CF_{t+1}} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\sigma_p^2})$. The information constructs a signal perceived by investors as $\boldsymbol{S_i} = \boldsymbol{CF_{t+1}} + \boldsymbol{\epsilon_i}$; where i represents the single investor and ϵ_i is a forecast error made by the investor. We can define the forecast error as $\boldsymbol{\epsilon_i} \sim \boldsymbol{N}(\boldsymbol{\alpha_i}, \boldsymbol{\sigma_{s,i}^2})$; where α_i is the bias of the investor and $\sigma_{s,i}$ is the quality of the signal perceived by the single investor.

According to Bayesian updating, the agents should update their expectations, and therefore the price of the asset, following:

[20] See 1.3 Rational pricing model

$$E(\boldsymbol{CF_{t+1}}|\boldsymbol{S_i}) = \boldsymbol{\mu}\left(\frac{\mathbf{1}/\boldsymbol{\sigma_p^2}}{\mathbf{1}/\boldsymbol{\sigma_p^2} + \mathbf{1}/\boldsymbol{\sigma_{s,i}^2}}\right) + \boldsymbol{S_i}\left(\frac{\mathbf{1}/\boldsymbol{\sigma_{s,i}^2}}{\mathbf{1}/\boldsymbol{\sigma_p^2} + \mathbf{1}/\boldsymbol{\sigma_{s,i}^2}}\right)$$

The investor should average the signal from the new information with the prior distribution based on its quality to capture the new statistical efficient price. Rational theory also presents another assumption to this process: investors are always rational and are perfectly able to assess the information. As of that, the investor's bias should always be 0 as they do not make mistake and are always rational, $\alpha_i = 0$; and the quality of the signal perceived by the single investor should always be the true quality of the signal, as they are perfectly capable of understanding the relevance of the information, $\boldsymbol{\sigma_{s,i}^2} = \boldsymbol{\sigma_s^2}$.

The behavioural approach that gives foundation to the analyses of this work contrasts the perfect rationality and the implications just presented. The empirical exploration presented in this study wants to support a different information updating process presented by Griffin & Tversky[21]. Their paper investigates the human process of evaluating evidence and their confidence in doing do, showing that people are often more confident about their assessments than what really the facts back. The researchers propose a process now commonly referred as *salience heuristic*, process that may have affected the price formation during the FTX events.

According to this theory, salient information could drive people decision making, even if it should not have an impact on the fundamental value of the assets, as people focus more on their "first impression" rather than the actual relevance of the information. The theory suggests that the relationship between strength and weight found in statistics and calculus is not shown in human judgment. In probability theory, any evidence coming from a sample should always be compared with the sample size to shed a light on the true probability. Therefore, the results found should always be tested to account for their predictive validity. The researchers propose that people tend to stress more the strength of the evidence, as they leave the weight for a marginal role to adjust the response.

[21] Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. Cognitive Psychology, 24(3), 411–435. https://doi.org/10.1016/0010-0285(92)90013-R

As of that, there is evidence that people are more likely to be overconfident when the results emerging from the sample are strong, even if the sample is not large enough and therefore as little weight; and they are more likely to be underconfident when the sample is very explicative but the results coming from it are not salient.

We analyse signal coming from news as evidence coming from a sample. This work finds ground in the application of strength and weight framework to explain crypto investors behaviour. The idea is that investors may have overreacted to some salient information that should not have impacted the price of the whole cryptocurrencies markets and that this could have driven the chaotic response after the FTX crash: shocking information that is not completely relevant might able to affect investors judgment in crypto market. Practically speaking; the strength of the information would be the illiquidity of FTX, an information definitely relevant; the weight instead would be the single business of FTX and its coin FTT, size that should have circumscribed the reaction toward a single market. The empirical analysis presented in this work wants to assess which information is labelled as salient for crypto investors and if this can affect the price formation of cryptocurrencies.

2.2 Presentation of the datasets

The research was carried out on a dataset kindly offered by a cryptocurrency market making company[22]. A market maker is an agent playing a critical role in pricing assets by discovering their fair value. Behind these "fair" prices, a market maker is always present and prepared to both buy and sell the asset. They bridge the gap between buyers and sellers, providing continuous buy and sell orders which not only makes trading smoother, but also contributes to the overall liquidity of the asset. This liquidity is essential because it ensures that participants can enter or exit their positions without causing significant price movements, facilitating growth for the market.

The dataset consists in the volume of trades and orderbooks, minute by minute, executed on the major cryptocurrency exchanges for the major cryptocurrencies. The dataset

[22] I would like to thank Jan Gobeli and the Keyrock Research Team

presents trades carried on Binance, Bitstamp, FTX, Gate-io and Okex[23] for different cryptocurrencies during the first 20 days of November 2022[24], month of the FTX collapse. The coverage of cryptocurrencies varies among the exchanges, but the research is limited to the largest cryptocurrencies by market cap back at that time, specifically: Bitcoin (BTC), Ethereum (ETH), FTT, and Solana (SOL). The dataset also presents entries relative to the best bid and best ask price available in the market on an aggregate level. The dataset includes information on both spot and future markets.

As usual for large datasets, it is found that data is not complete; therefore, it is necessary to specify the manipulation of data to avoid misrepresentation. Two main issues arise: duplicate and missing values. For duplicate values about a specific entry in a specific time[25], the most populated entry is kept, as it is presumed to be more informative. For missing values, two different approaches are taken. For the orderbooks datasets, it is assumed that missing values are caused by unchanged state. As of that, missing values are filled using by propagating the last valid observation over the next valid observation[26]. For the trades' datasets, the missing value could be imputed to both miss management or recording of data, as well as the presence of no events at that specific time (no trades are actually carried out at that moment). This issue is crucial when taking into account the FTX exchange collapse, as the access could have been limited or denied. As of that, it is decided that missing values are not filled, assuming there were no trades at the missing value time.

2.3 A first look at the data

Having such a voluminous dataset, we are able to extract precise information about the price formation of assets and even discriminate among different exchanges. To understand the impact of the crash, a first look at the time series of the asset prices is key. The datasets allow to compute the VWAP, volume-weighted average price. The VWAP is a common measure used to benchmark trading costs. Investors can evaluate their dealer

---

[23] FTT data in the Okex exchange is not available as the asset was not listed
[24] FTX data is missing after Novembre 14th as of its collapse and impossibility of trades
[25] This is mostly found in the orderbooks datasets at 23:59:00 time for each day. It is possible to argue that the problem arose due to the day change.
[26] Carried out using "ffill" method

performance when comparing their offered price with the VWAP. As of that, VWAP is an explicative synthesis of the price and its dynamics for a specific exchange[27].

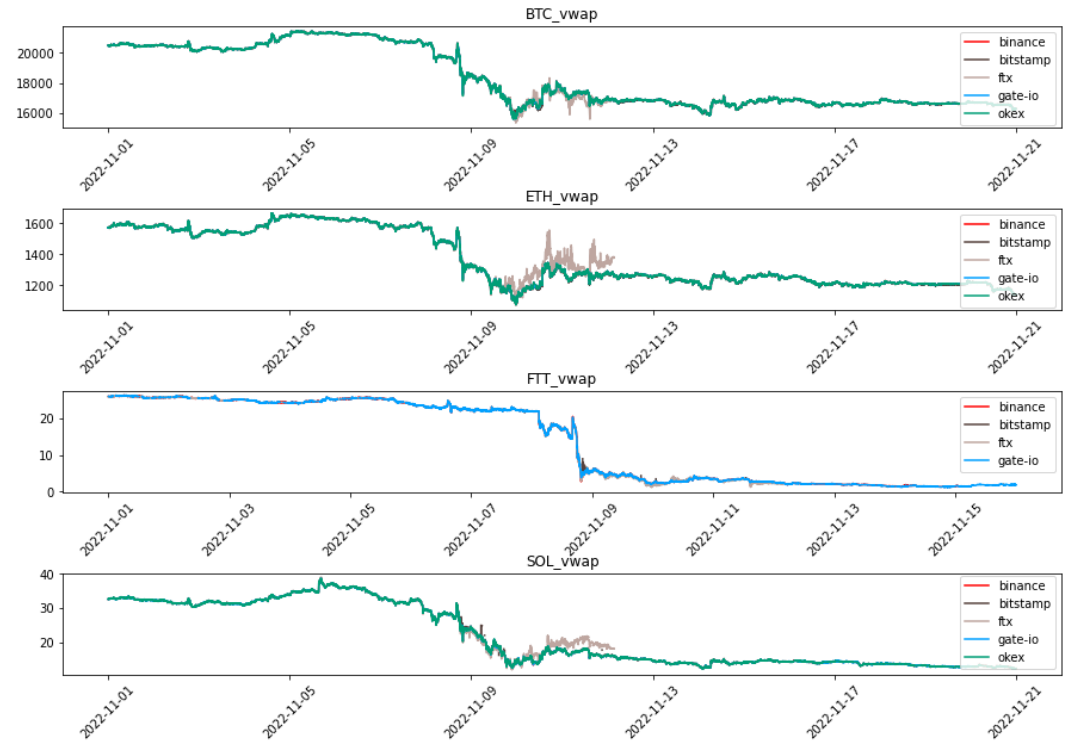

Figure 1 Volume-weighted average prices

From the analysis of Figure 1, several important facts arise. Starting from the key one object of the present work, it can be noted that starting around November 8th 2022, all the cryptocurrencies showed a decline in price. The dynamic of this dump is different for each asset, but it presents common factors. Three of the assets, not taking into account FTT whose value nearly reached 0 as of the bankrupt of the "issuing" firm, show a major down around the 9th of November, but subsequently an upward correction can be noted. This pattern is consistent with a behavioural point of view in which overconfidence impact prices, as prices hit a lower value than they should have to be subsequently realigned. Following the model presented on the first section of this chapter: investors may, at first, not have captured the true signal of the information stream; or the signal extracted could have been biased negatively. We know that salient information is prone

[27] Foucault, T., Pagano, M., & Röell, A. (2013). Market Liquidity: Theory, Evidence, and Policy. Oxford University Press.

to generate overaction, therefore evidence of this behaviour may be possibly found in the dataset.

The second point is relative to the efficiency of the cryptocurrencies market. Each plot refers to assets priced on different exchanges, but they represent the same object. The law of one price should be observed: asset priced on different markets should have the same price. Two major deviation may be observed on FTX and Bitstamp. For FTX we observe major deviations starting from the 9th of November. This deviation is not shocking due to the real events happened at time, as the fairness of the exchange was questioned, and the correct functioning of the platform was undermined. What raises more questions about the efficiency of crypto markets is the prices behaviour on Bitstamp, exchange that as of the day of writing of this work is still running. A major deviation can be spotted around the 9th of November for SOL and some minor deviations around the 10th of November for BTC. This deviation to do threat the correct functioning of the markets but rise some questions about its efficiency and its market structure: this misalignment creates arbitrage opportunities that should not be observed in a sound financial market.

Overall, the law of one price seems to hold in the market. This allows to use the different assets and exchanges as different samples and expand the analysis to the wide dataset, to increase the evidence that supports an empirical analysis.

Having assessed the prices allows the assessment of a key metric for financial markets: volatility. To present an evolution of the volatility of the prices, the daily variance of returns is compute.

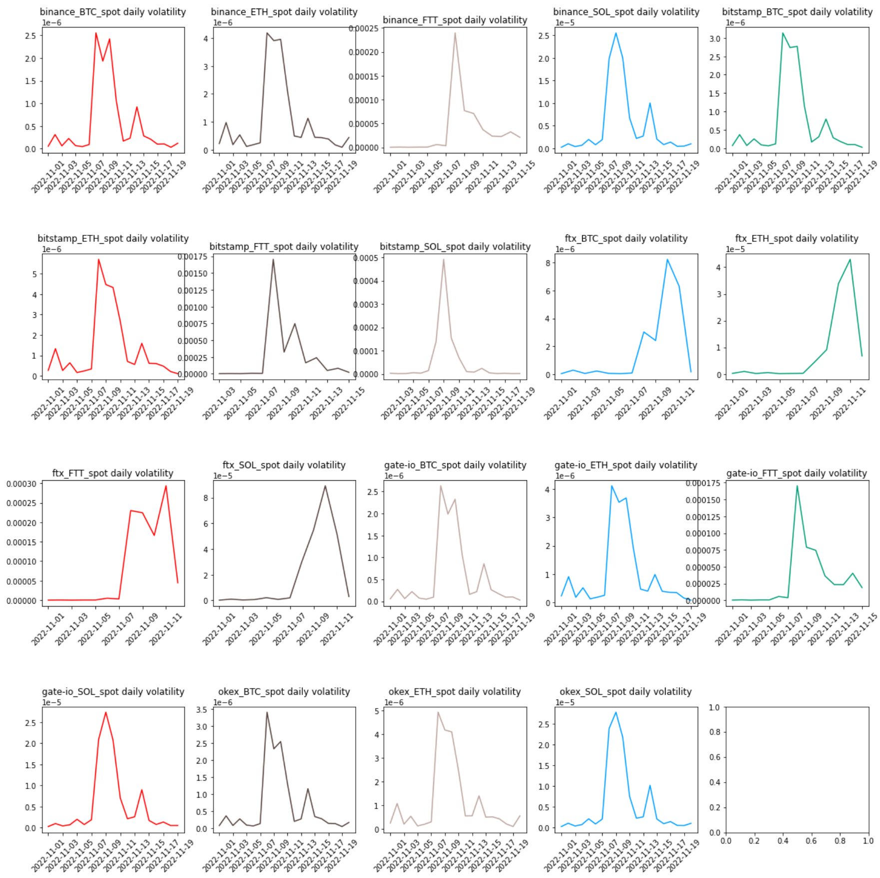

Figure 2 Daily volatility

As of Figure 2, we can see a common pattern among the markets: volatility spiked around the 7th of November. This finding highlights a pattern found in Figure 1. It can be noted that prior to the effective crash, a first decline in the price was followed by an initial recover before tumbling to an all-time low. This instability in prices is reflected in the volatility metrics. Volatility is a direct result of panic in the market and insecurities. These spikes shows that prices were hardly consistent throughout the days, implying a continuous release of inconsistent news regarding this markets for the EUT to hold.

2.4 Assessing market liquidity

Before exploring the behavioural analysis core of this work, an assessment of the liquidity of the markets is presented. Liquidity is a benchmark of market performance. Investors and intermediaries look after liquid market, as it increases their performance as trading strategies are designed to minimise effect of liquidity on returns. Illiquidity is usually gauged by the cost of trading: the higher the costs, the less liquid is the market. It is then clear that player look after liquid market to ensure lower trading costs and liquidity can have impact on prices and affect the price formation process. The presented dataset allows for a liquidity comparison, useful to understand how market execution solidity on a technical side is affect during stressful instances, and therefore which markets are sounder.

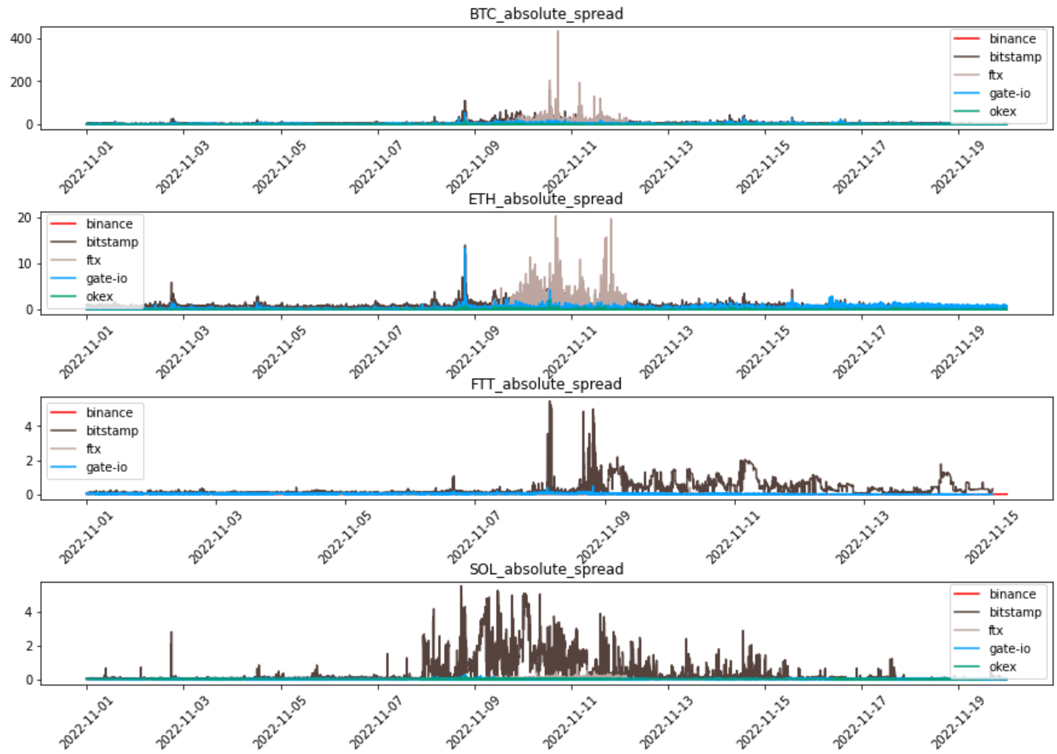

Figure 3 Absolute bid-ask spreads

Figure 3 presents the absolute bid-ask spreads of the different markets. This metrics highlights the difference between the best available price at which investors can buy the asset (bid) and the best available price at which investors can sell the asset (ask). This measure explicates the cost of trading as bid and ask offers need to be matched to perform a successful transaction. Therefore, investors look after the lowest bid-ask spread, to minimise the price difference for buying and selling. Several information can be extracted from Figure 3. Focussing on the different exchange, it is possible to assess that Binance showed the lowest bid-ask spread throughout the whole-time horizon and for every asset, even during days of market panic. We can therefore infer that Binance is the most efficient exchange for cryptocurrency. A similar behaviour is found for Okex, that shows some noticeable increases only during the market crash for ETH. The other exchanges instead showed some problematic evolution. Prior to the collapse, FTX showed big inefficiencies for BTC and ETH, the first two coins for market capitalization. This finding should again not be shocking as the platform normal functioning was undermined. What is more relevant is the reaction of Bitstamp and Gate-io, exchanges that are still operating as the time of writing. It is possible to see the poor performance of Bistamp for Solana: during the days of the crash the spread peaked 4$, expressing a big inefficiency. This situation persisted in the market for over a week, creating poor conditions for market players. Gate-io instead showed a poor performance for the ETH coin at the beginning of the crash.

With this analysis is possible to infer that the broad cryptocurrencies market liquidity and feasibility of execution were not threated during the crash considering the general market that kept showing the same prices across different exchanges. Some exchanges continued with their regular business and processed the trades in a sound way; therefore, prices decline may only be imputed to liquidity for small, isolated parts of our dataset.

A key metric to assess liquidity and its impact on prices is the "illiquidity ratio" proposed by Amihud[28], commonly referred as Amihud ratio. This measure shows the impact of orders on the asset prices and is therefore referred as illiquidity. The insight is that a strong net positive demand may bring to an upward correction of prices, as trades are not settled due to the low amount of sell offers (and vice-versa).

[28] Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets, 5, 31–56.

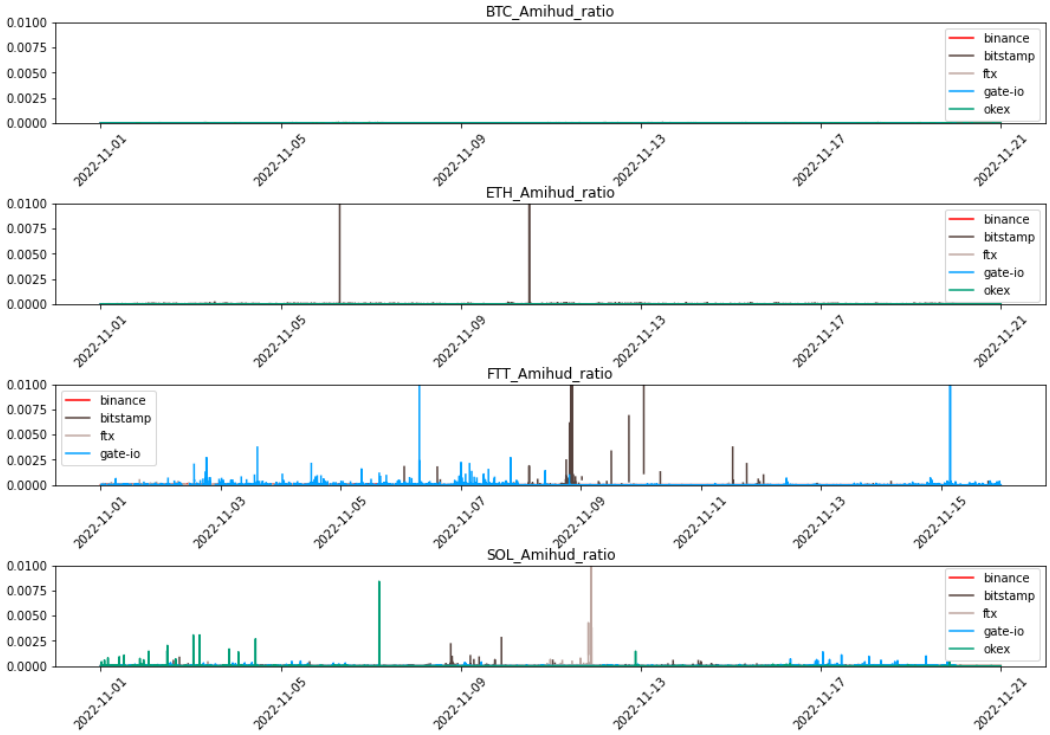

Figure 4 Amihud ratios

Figure 4 reports the Amihud ratios found in the dataset. Starting from Bitcoin, it can be seen that the orders flow did not impact the price formation at all, as the line is flat and close to 0. However, there is evidence of risen market microstructure impact on prices, especially for Solana. The data suggest that order volumes may have contributed to the price movements of Sol in the Okex exchange as well as in Bitstamp, where an impact on ETH was present as well.

After this analysis, two major fact arises: first that Bitcoin and Binance are respectively the most efficiency coin and the most efficient exchange; and that lack liquidity was not present in the broad and cannot explain the crash by itself. The first finding can be related to the larger volume of this market, backing the idea that a larger scale is related to a greater market efficiency. The second finding instead gives ground to the soon-to-be behavioural analysis, given that no strong evidence of movements solely due to market microstructure is found.

2.5 Financial Markets and keywords: a sentiment index

Understood that liquidity and market microstructure cannot explain the price forming process and the bizarre reaction observed on cryptocurrency markets, this work wants to try to explain the crash followed by the FTX event via a behavioural approach. A qualitative explanation of the event can be rather accepted or not by the reader, that is why a robust empirical, quantitative analysis is performed. The methodology of this works follows the intuition brought by Da et al[29]. In their paper, the researchers investigate the relationship between market performance and investor sentiment using an index based on internet search volumes. Their study appears to be relevant as their FEARS index has been shown to predict short term return reversal and temporary increase in volatility, phenomena observed in the presented dataset.

Their rationale is rooted in figure of humans and how relevant sentiment could impact decision making, deviating from the *homo oeconomicus* assumptions. With this work, the researchers build a list of words that "reveal sentiment toward economic conditions"[30] They start with an initial list of words based on text analytic literature that are labelled as negative, namely the Harvard IV-4 Dictionary and the Lasswell Value dictionary. A second filter is imposed on "economic" word, as the sentiment investigated should the sole relative to economic conditions. To build the index following an empirical approach, they extract the words' Google trend data during the sample period. They are able to assess how the research volume changes during the sample period, and therefore if there is any common pattern in research and price changes.

The FEARS Index is build performing a backward expanding rolling regression of the Google trend data on market returns, to assess the evolution of the return's variability explained by the single word. The 30 most relevant words are then pooled together to form a sentiment index that represent the words associated with households' negative economic sentiment.

---

[29] Da, Zhi & Engelberg, Joseph & Gao, Pengjie. (2015). The sum of all FEARS investor sentiment and asset prices. Review of Financial Studies. 28. 1-32. 10.1093/rfs/hhu072.
[30] Da, Zhi & Engelberg, Joseph & Gao, Pengjie. (2015). The sum of all FEARS investor sentiment and asset prices. Review of Financial Studies. 28. 1-32. 10.1093/rfs/hhu072. Pag.6

The idea behind this approach is that when the economic condition gets worse, non-professional use their tools to get information about the current state of economy. As Google is the most used search engine, owing more than 85% of the search traffic, evidence is likely to be found on the platform. The behavioural component would be present in form of overconfidence: non-professional reading about the poor economic conditions would prefer to exit the market, causing a down pressure on market return.

This view is backed by the empirical analysis carried out by Da et al, that find significant negative regression coefficients that can explain 6% of the variation of the returns.

2.6 Creating a tailored index

The FEARS Index is a first candidate to create a sentiment index for cryptocurrency market, but limiting the analysis to this index may be misleading. As discussed on Chapter 1, crypto investors show different behaviours when compared to investors in traditional financial markets. As of that, the keywords driver of sentiment could be different; and the FEARS index may be obsolete, as investors may learn about the markets and their dynamics.

Then, to proceed with the analysis, a new index may need to be built. A new index is built starting from the Loughran-McDonald Master Dictionary. This dictionary presents a list of words labelled with a sentiment component. The dictionary is continuous updated, therefore issues about obsolescence of the labels is avoided, and the version used for the present analysis is the 2021 one. The words found in this list are proper to the financial word: three-fourths of the words labelled as negative by the Harvard Dictionary are words that typically do not show a negative meaning in financial jargon[31]. The list is trained on financial speeches and balance sheets which makes it a strong tool for financial textual analysis.

The construction starts from the negative words found in the dictionary. To filter out words that are extremely relevant for investors, we only keep words that are present in

[31] Loughran, Tim and McDonald, Bill, When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks (March 4, 2010). Journal of Finance, Forthcoming, Available at SSRN: https://ssrn.com/abstract=1331573

more 100000 analysed documents. We are then left with 251 words commonly used and certified to have financial relevance. We then continue with a similar methodology used by Da et al. The Google trend data about the 251 words for the sample period is extracted, then an expanding rolling regression for each word on market returns[32] is performed. After that, each word has a series of regression coefficients associated with the market return. This index is used to look for evidence of traditional financial sentiment in cryptocurrency markets.

To make sure that no component is overlook, a third index is built. The need of a third index arises due to the fact that crypto markets and their investors are not driven by the same news and sentiment of traditional financial markets. To account for that, a list of the most common words related to the crypto world and crash is extracted. This list is based on news and posts published after the sample period that tried to explain the facts of November 2022.

Each of three indexes will be used in the next chapter as benchmark of investors' market sentiment to test whether market sentiment can partially explain the odd evolution of return observed in the market after the FTX events.

2.7 Testing for herding behaviour

It has been discussed that crypto investors show some common traits, mainly heuristics, prospect, and herding behaviours. The previously proposed methodology would test if overconfidence towards news is a driver of cryptocurrency markets. Finding evidence of the relationship between news volume and returns would confirm one of the presented findings as well, as the "buy on the rumour, sell on the news" common crypto heuristic would find ground there, and an accentuation of this behaviour during salient news would underwrite the finding.

Prospect theory is something that needs to be assessed at personal level: the dataset does not reveal the identity of the trader nor gives any information about the source of the trade; therefore, it would not be possible to state an increasing risk taking (each trades

[32] Bitcoin prices on Binance are used as market benchmark, as they have been shown to be the most liquid coin and the most efficient exchange

could have been placed by a different investors, and we do not even know the final positive or negative result of the operation)

Nonetheless, the dataset allows to test for herding behaviour. As previously introduced, herding behaviour is the phenomenon by which individuals in a group follow the same direction, influenced by each other. The name calls the herd movements of animals, that may aggregate in herds, packs, and flocks; to move together towards a same destination. This attitude is not uncommon in human life, think for example at behaviours in sporting events or religious gatherings.

Herding behaviour in financial markets has given grounds to several research, mostly related to market microstructure and order flows. Following Toth et al[33], equity orders flow may be imputed to two different phenomena: order splitting and herding behaviour. The idea behind the first element is that investors may want to time the market to complete trades at a better price: liquidity may be insufficient to place a large order at the best price, therefore large investors may decide to divide (split) their order in several micro trades to obtain the best price possible. The other argument is that order flows may be serially correlated for the discussed herding behaviour: a sell trade will follow a sell trade due to the emulation of the first trader by the second one. The researchers have a more detailed dataset than the one available, therefore the same methodology cannot be applied. However, a different test for herding may be executed, starting form Toth et Al's intuition. Assuming their intuition is correct, the serial correlation is explained by either order splitting or herding behaviour. In moments of market panic, we should observe a lower proportion of order splitting, as the uncertainty about the future bid-ask spread should not yield to any sound plan of trading cost cutting; finding backed by the illiquidity spike observed in the dataset. We can therefore infer that a movement of serial correlation could be driven by investors decision making, and therefore the sentiment index introduced should be able to track the evolution of order flow: as metrics of investors sentiment, they

[33] Toth, Bence and Palit, Imon and Lillo, Fabrizio and Farmer, J. Doyne, Why Is Equity Order Flow so Persistent? (June 15, 2014). Available at SSRN: https://ssrn.com/abstract=2460731 or http://dx.doi.org/10.2139/ssrn.2460731

could help explain the patterns of orders flow explaining the herding behaviour component.

The available dataset allows the computation of aggregate signed trade autocorrelations: it can be assessed if a market that showed positive buy volume is followed by positive buy volume (and vice-versa). In an efficient market, serial autocorrelation should not be observed, as an upward pressure on the price should be followed by a downward pressure, to make sure the “fair” price is still observed in the market and the changes balance out. Observing serial autocorrelation would mean that order splitting or herding behaviour are present on the market, and an evolution of serial autocorrelation harmonic with our sentiment indexes would support the idea that herding behaviour is monitored by the presented index in crypto markets, as this attitude is magnified by sentiment in moments of market panic.

Chapter 3

Results and interpretations

3.1 Defining the Indexes

As reported in Chapter 2, this work will be carried out on three different indexes: the FEARS Index, an index based on the Loughran-McDonald Master Dictionary, and a third index based on crypto specific words. To clearly understand the results as well as the limitations about the upcoming analysis, a discussion about the index building process outcome must be debated. The building process proposed by Da et al[34] is based on market returns and volume of Google research. As of this process, the words that are not even a candidate to explain the returns are not taken into account in the analysis, they will not be included in the final word list. Then, the final list of word included in the index are already empirical relevant for the dataset available. We decide to keep the words whose regression coefficients average absolute t-statistic is greater than 1.645, to ensure that the regression is significant at least a 10% alpha level confidence. We are left with 31 words, a number similar to the FEARS index that would likely not overfit the data. By the nature of this construction then, it is reasonable to expect a greater significance of the analysis: the words included are tailored to the specific dataset that only covers a monthly a time series (even if it covers different exchanges and assets), therefore some words that address the specific event happened during the time period may arise. When having this in mind, the results from the index constructions are not surprising. In the final version of the index, we find words like “fraud” and “insolvency” that were not present in the FEARS indexes, that was more related to macroeconomic drivers like “inflation rate” and “GDP”.

A further step for the analysis of behavioural components in the evolution of cryptocurrencies asset prices can be completed already by looking at these two indexes. It is possible to say that the traditional financial words that are more related to crypto

[34] Da, Zhi & Engelberg, Joseph & Gao, Pengjie. (2015). The sum of all FEARS investor sentiment and asset prices. Review of Financial Studies. 28. 1-32. 10.1093/rfs/hhu072.

asset prices are not words that related to macroeconomy, wealth or technology; but are more related to words that relate to "corporate" events and activisms in the market. As of that, it is possible to infer that crypto investors' trading activity is not driven by traditional financial variables as it happens in traditional financial markets; but the activity is instead justified by local salient events that the investors perceive, acting as noise traders consistently with Al-Mansour[35].

The third index is instead built following the rationale outlined in Chapter 1 with the methodology presented in Chapter 2.6. According to the simple models presented, salient news could imply a mispricing of asset, due to the inability of agents to extract the signal from the information and their bias in doing so. As of that, the index build gathers the most common Googles that includes words specific from the crypto world, like "blockchain" and "crypto" itself, paired with capital market jargon. The outcome will then represent what an investor could have searched for during our time horizon and therefore represent the salient news they were exposed to because of this action.

This will test whether the strength and weight framework introduced can help explaining the reaction observed in the markets: market investors that look for news using these keywords are exposed to common information that they may label as relevant due to this framework.

Given that the Google trends database is standardized on the peak of research, we construct the Indexes values based on the daily change in the search for each term of the word lists. Having extracted the Search Volume Index of each word, we define the daily change of each word as:

$$\Delta SVI_{j,t} = \ln\left(SVI_{j,t}\right) - \ln\left(SVI_{j,t-1}\right)$$

Where *j* is the single search item. We then define the Index on day *t* as

$$INDEX_t = \frac{1}{n}\sum_{j=1}^{n} \Delta SVI_{j,t}$$

[35] Al-Mansour, Bashar (2020). Cryptocurrency Market: Behavioural Finance Perspective. Journal of Asian Finance, Economics and Business , 7(12), 159-168. See 1.7 Biases in cryptocurrencies market for details.

3.2 Presentation of results: returns

Stated the characteristics of the Indexes, the proper results of the empirical analysis are hereby presented. The methodology follows again the rationale of Da et al. To understand the relationship between returns and volume of Google research, a linear regression is performed. As we are interested in understanding if the evolution of market prices, the crypto currencies returns are used as dependent variables. We want to assess if research volume has explainability power, therefore we will use them as independent variables. Each market return is regressed on each different Index[36] finding:

FEARS INDEX:

Regression on FEARS Index

y	Coeff	T-stat	Adj_R^2
binance_BTC_spot_vwap_mean	-0.027099717676390384	-0.602343888959426	-0.03669806557557331
binance_ETH_spot_vwap_mean	-0.018306255615368586	-0.28082723557839423	-0.053934270288475084
binance_FTT_spot_vwap_mean	-0.08753694547718804	-0.30999381460484415	-0.052871943757097695
binance_SOL_spot_vwap_mean	-0.05203013143779955	-0.3701672648117151	-0.05035741362881785
bitstamp_BTC_spot_vwap_mean	-0.026027420291891888	-0.5747058155774549	-0.03864409367584609
bitstamp_ETH_spot_vwap_mean	-0.010479074453731384	-0.16088971009342523	-0.05721373500684668
bitstamp_SOL_spot_vwap_mean	-0.04626761488514629	-0.33934901868489026	-0.051699333193844677
ftx_BTC_spot_vwap_mean	-0.023354598460643756	-0.5317052340806181	-0.04150330157295312
ftx_ETH_spot_vwap_mean	-0.011751126585711636	-0.1789304659026809	-0.05683319473324033
ftx_FTT_spot_vwap_mean	0.05411278693063888	0.1951691727653492	-0.05645638315134427
ftx_SOL_spot_vwap_mean	-0.08815144336261105	-0.5803136724447426	-0.03825604612049571
gate-io_BTC_spot_vwap_mean	-0.026734404735970548	-0.5931075671955408	-0.037357754724004044
gate-io_ETH_spot_vwap_mean	-0.016461138759225016	-0.25204611755733547	-0.0548815459522336
gate-io_FTT_spot_vwap_mean	-0.08176902269614408	-0.26922470122109315	-0.05432824960810145
gate-io_SOL_spot_vwap_mean	-0.05209338731739785	-0.37055956789204014	-0.050339603209829775
okex_BTC_spot_vwap_mean	-0.027137969824720658	-0.6040461687208837	-0.03657546305677539
okex_ETH_spot_vwap_mean	-0.018415724476826174	-0.28298285235149234	-0.05385927602567531
okex_SOL_spot_vwap_mean	-0.051766921564450355	-0.3705388888587955	-0.050340542486434936

Figure 5 Return regression on FEARS Index

[36] Due to an inconsistency in the dataset, the analysis does not include the FTT returns in the bitstamp exchange

Dictionary INDEX:

Regression on Dictionary Index

y	Coeff	T-stat	Adj_R^2
binance_BTC_spot_vwap_mean	-0.032867205428702605	-0.8824383540425595	-0.012447624577301086
binance_ETH_spot_vwap_mean	-0.05554050542285872	-1.0469208448254161	0.00530741742682761
binance_FTT_spot_vwap_mean	-0.037142963556939086	-0.1566857625833687	-0.057296642041608115
binance_SOL_spot_vwap_mean	-0.04999911127089228	-0.4251683753817968	-0.047683074793020364
bitstamp_BTC_spot_vwap_mean	-0.03173734117277503	-0.8457830588180653	-0.01606804459207667
bitstamp_ETH_spot_vwap_mean	-0.0544213984314746	-1.0271186068329452	0.0030447364070486094
bitstamp_SOL_spot_vwap_mean	-0.07736008125633859	-0.6842616411439254	-0.030443066855927903
ftx_BTC_spot_vwap_mean	-0.02994399418963704	-0.823025250734984	-0.018250985659686414
ftx_ETH_spot_vwap_mean	-0.06388012537475793	-1.2088453213312398	0.024987776360143266
ftx_FTT_spot_vwap_mean	0.054965999219675724	0.2367752824352727	-0.055343222606355935
ftx_SOL_spot_vwap_mean	-0.06721728671550137	-0.5273269771950418	-0.04178279766304982
gate-io_BTC_spot_vwap_mean	-0.03307367572642306	-0.8867870684551433	-0.012009670290079155
gate-io_ETH_spot_vwap_mean	-0.05597307000413726	-1.0539908198171584	0.006123200325893752
gate-io_FTT_spot_vwap_mean	-0.04509989800121125	-0.17704341728208606	-0.056874877711483585
gate-io_SOL_spot_vwap_mean	-0.052304481757618634	-0.4449214597193506	-0.04663606165788603
okex_BTC_spot_vwap_mean	-0.0331943141347275	-0.892892591398926	-0.011391801454400996
okex_ETH_spot_vwap_mean	-0.055974047236997865	-1.0574831187070872	0.006527695126745114
okex_SOL_spot_vwap_mean	-0.05193813408583534	-0.4445626573175075	-0.046655484844458917

Figure 6 Return regression on Dictionary Index

Crypto INDEX:

Regression on Crypto Index

y	Coeff	T-stat	Adj_R^2
binance_BTC_spot_vwap_mean	-0.02370085601576005	-2.106125496736665	0.16028187801247473
binance_ETH_spot_vwap_mean	-0.04127944587848592	-2.72483321299014	0.26304158576325054
binance_FTT_spot_vwap_mean	-0.10420049908257908	-1.3980428858698342	0.05035863286393816
binance_SOL_spot_vwap_mean	-0.07067998119839075	-2.0036029137068985	0.14344549936795836
bitstamp_BTC_spot_vwap_mean	-0.02325053911589875	-2.041578375307444	0.14966156167081424
bitstamp_ETH_spot_vwap_mean	-0.03976860615296299	-2.5935014744773714	0.24134660650990292
bitstamp_SOL_spot_vwap_mean	-0.06503953230119694	-1.8801081995033229	0.12343952691461224
ftx_BTC_spot_vwap_mean	-0.019478221975598776	-1.7137721109927657	0.0971567139385684
ftx_ETH_spot_vwap_mean	-0.03843350882505457	-2.4460685054551914	0.2168209843098461
ftx_FTT_spot_vwap_mean	-0.08654246888187753	-1.1658369453261153	0.019563829407756028
ftx_SOL_spot_vwap_mean	-0.06672957754100392	-1.691860708045827	0.09376476881400486
gate-io_BTC_spot_vwap_mean	-0.023374067291493293	-2.0656607799827635	0.15361646936411066
gate-io_ETH_spot_vwap_mean	-0.03990210941561349	-2.591758201759561	0.24105746154144325
gate-io_FTT_spot_vwap_mean	-0.1447641235954703	-1.8810616568322844	0.12359257913427057
gate-io_SOL_spot_vwap_mean	-0.07021406601533986	-1.9869122353482174	0.14072205119544512
okex_BTC_spot_vwap_mean	-0.023746637475620523	-2.1148675464898052	0.16172490845334497
okex_ETH_spot_vwap_mean	-0.041219441006735295	-2.7255941572869955	0.26316670472791204
okex_SOL_spot_vwap_mean	-0.07075021634614276	-2.021195452392824	0.14632189936483198

Figure 7 Return Regression on Crypto Index

The three regressions present different results, worth exploring to shed a light on sentiment drivers of cryptocurrency markets. Starting from the FEARS Index, we can observe that all the regression coefficients are negative, implying a negative relationship between cryptocurrencies prices and search volume of the used words, and presents an adjusted R^2 around 0.05 for most of the coins. This would mean that 5% of the variation in the returns can be explained by a variation of the Index, and that a surge in the index volume is related to a downfall in crypto prices. This result is similar to the findings of Da et al. However, we find that the T-statistic of the coefficients is sensibly lower than an acceptable threshold like 1.65, a significance at 10% alpha level. We cannot therefore state that the FEARS Index is a successful predictor of crypto returns. This could mean that either the Index does not relate to crypto investors sentiment, or that this sentiment does not have an effect on cryptocurrencies prices.

When assessing the Dictionary Index instead, we find again negative regression coefficients, but the R^2, and therefore the degree of explicability of returns attributable to a variation in the search volume, are lower than the FEARS Index. The regressions are again not significant at an acceptable level, but there is an improvement when compared to the FEARS. This latter finding is not shocking, as it is embedded in the construction of the Index: the words kept are the one that had a stronger relationship with the return. However, this relationship is not able to explain the variation of the returns. This means that the Dictionary Index is able to track and be driver of some kind of sentiment for crypto investors, however we do not have enough evidence to say that this sentiment is a mirror of their decision making.

After the analysis of these two indices, one relying on successful past academical analysis for traditional financial markets, and one built according to the same underlying methodology, we cannot say that macroeconomic indicators and traditional words related to financial pessimism drive an altered decision-making process for crypto investors.

The more interesting facts arise looking at the third Index. The Crypto Index should express a sentiment proper to the specific crypto market, and therefore highlight information to which crypto investors were exposed at the time of trades. We find again that the regression coefficients are negative, but the significance and magnitude of the regression is extremely improved using this Index. The variance explained ranges from as little as 5% to more than 26%. These high value coefficients are also backed by high T-statistics, that allow us to reject a null hypothesis of the coefficient be null at 5% for Bitcoin in Binance, and at 1% for Ethereum in Binance; which we showed to be the most efficient coins and the most efficient exchange.

After this finding we are therefore able to state that crypto investors are affected by the volume of news exposure, but that the information they are looking for and which triggers their action is not related to the one affecting traditional financial markets and that academy refers as sentiment charged words. We can say that crypto investors react towards some sentiment that is specific to crypto market and orthogonal to the ones found in traditional financial markets.

3.3 Presentation of results: autocorrelation

As introduced in Chapter 2, following the rationale of Toth et al[37], equity order flows can give evidence of herding behaviour and order splitting. By observing serial autocorrelation, it is possible to assess the presence of these phenomena in the market. The investigation performed is an autocorrelation analysis on the whole dataset, this time expanding the research to both Spot and Futures prices, as we will mainly value the behaviour of market players rather than their impact on the market.

[37] Toth, Bence and Palit, Imon and Lillo, Fabrizio and Farmer, J. Doyne, Why Is Equity Order Flow so Persistent? (June 15, 2014). Available at SSRN: https://ssrn.com/abstract=2460731 or http://dx.doi.org/10.2139/ssrn.2460731

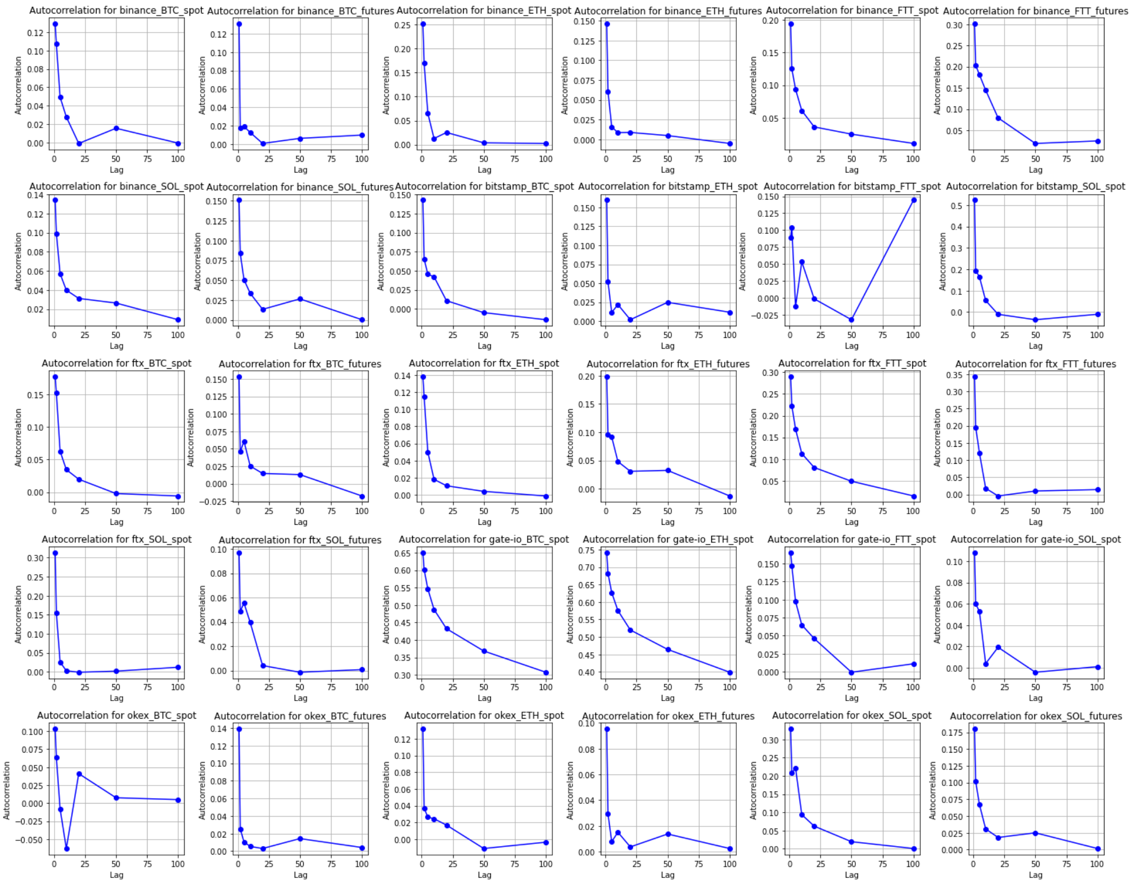

Figure 8 Cryptocurrencies autocorrelation

Figure 8 reports the findings of such analysis. The results appear to be consistent with the analysis carried out in the cited paper for most of the exchanges. We observe positive autocorrelation in shorter time horizons, reducing over time. This implies that the order splitting and herding behaviour are short time phenomena that are not constantly present in the markets as there is no evidence of their spread in a long-time window. For most of the cryptocurrencies and exchanges, we see that the autocorrelation value declines to 0 as the lag increases, confirming such attribute.

However, this property is not shown in some specific markets, mainly three: Gate-io BTC spot, Gate-io ETH spot, and Bitstamp FTT spot. For the first two markets. We observe a positive autocorrelation and a decline over time, however the reduction does not yield to a close to null autocorrelation value, but a rather strong positive value. This finding casts some shadows on the suitability of Gate-io as a sound market and its execution of orders. In Chapter 2 it has been shown that the Gate-io exhibited a poor performance in terms of liquidity during the crash, especially for the ETH cryptocurrency. As of this new finding, it can be inferred that the assets' orders execution could have been mitigated, and liquidity issues could have been even magnified without such practices, as some market players may have chosen to split their orders to get better execution conditions. Had these orders been placed in a single point in time, assuming the participation of non-experienced professionals that can easily participate the cryptocurrencies markets, we can guess that the exchange could have shown further signs of suffering, questioning the capability of handling times of market panic.

The other interesting case is the Bitstamp FTT spot one. In this specific market we see that the autocorrelation value does not decline over time, as we see positive value for a higher time horizon when compared to the lower lags. A sound interpretation of this finding is not found, but this bizarre finding could be imputed to data mining rather than a relevant market behaviour. Indeed, this example highlight the shortcoming that may arise when focussing on a single sample and the need of findings to be consistent in order to construct a proof.[38]

This first piece of evidence of autocorrelation incorporates both the herding and order split, and due to the nature of the dataset used for this work, it is not possible to isolate the two phenomena using the same methodology as the cited paper. However, a different analysis can be performed. Assuming that the order split activity is constant throughout the dataset, we should be able to assess the evolution of herding behaviour in the market by performing a similar regression as the one carried out on the returns. The finding of autocorrelation linked with one of our indexes would then support the idea that the herding behaviour is stressed during times when the returns are sentiment driven.

[38] We can consider this case as the proverbial exception that proves the rule

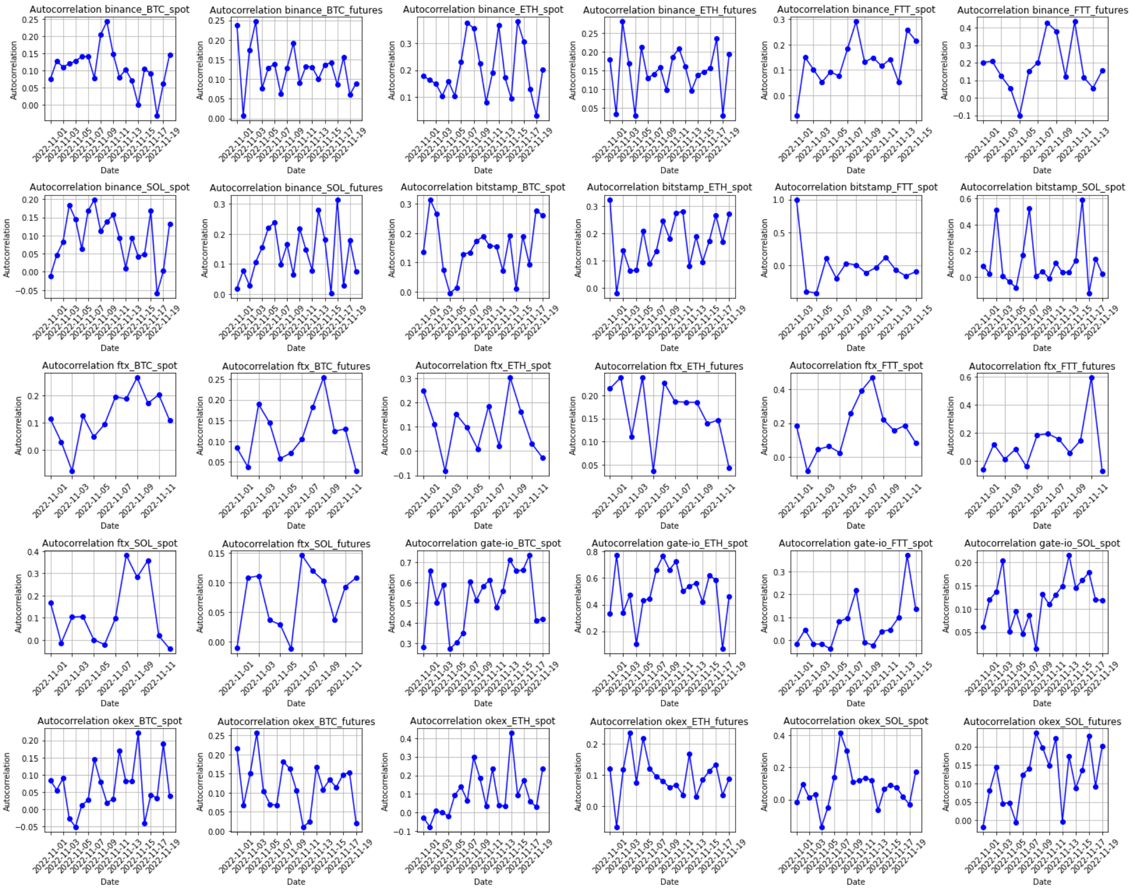

Figure 9 Cryptocurrencies Daily autocorrelation

Figure 9 presents the evolution of daily one lag autocorrelation in the markets analysed. This operation not only gives a higher detail to the process later presented but allows an analysis of the order splitting and herding phenomena. It can be noted that the daily autocorrelation does not appear to be stable, its value changes drastically from one day to another, highlighting the fact that herding and order splits are event driven activities not always present, but at the same time their presence is on average positive and significant. A clear example of this property is, again, the daily autocorrelation for FTT in the Bitstamp exchange: in the first day analysed the autocorrelation is at its max value, but over the days the behaviours changes, and the presence of this phenomena is less present, with the value stabilising around 0.

Understood that these phenomena are event driven, the same analysis performed using the returns can be carried out, seeking for evidence of magnified herding behaviour during time rich sentiment.

Regression on Dictionary Index

y	Coeff	T-stat	Adj_R^2
binance_BTC_spot	-87.96020134127681	-0.41325636266117655	-0.04829245608285859
binance_BTC_futures	3.880461991820432	0.6319062783900469	-0.034524047743692776
binance_ETH_spot	-0.23170859701657498	-0.12346622133956484	-0.057874932473698903
binance_ETH_futures	3.9553684888703398	1.302160436936481	0.03720773830542179
binance_FTT_spot	2.398830360491541	1.6315716721682236	0.08452974841748506
binance_FTT_futures	2.2523138565377154	1.5400274041920787	0.07080873180349934
binance_SOL_spot	5.400171527479981	0.33047660280315233	-0.05206464622681817
binance_SOL_futures	16.10008388409382	0.6694432446733237	-0.031627743198873404
bitstamp_BTC_spot	1.0084188427807574	0.16087386734860995	-0.05721405154096426
bitstamp_ETH_spot	-0.1429355794101751	-0.05571372548378914	-0.05863023478928264
bitstamp_SOL_spot	9.369422697735022	1.337189255204584	0.04194549467480946
ftx_BTC_spot	1.2296377597225128	0.9030561228661385	-0.010355562945722374
ftx_BTC_futures	2.0228826642946296	1.7556079870674084	0.10368204745022225
ftx_ETH_spot	8.719063173296743	1.2916403093625781	0.03580044497339552
ftx_ETH_futures	2.311907677629745	1.528615470114966	0.06912594482113943
ftx_FTT_spot	4.187840910572023	1.5550635447611056	0.07303565601220785
ftx_FTT_futures	-2.290283642025694	-0.7653417836967159	-0.023556117888608563
ftx_SOL_spot	-20.992287993863016	-1.4736357630247225	0.0611113427004768
ftx_SOL_futures	-5.462409762102188	-1.1201576490711203	0.01395544253909109
gate-io_BTC_spot	0.7510161353350667	1.5545178915253917	0.07295465132056012
gate-io_ETH_spot	-0.26591052432599643	-0.1347766914365977	-0.05769336808584158
gate-io_FTT_spot	-0.12415785506627955	-0.05867264072188777	-0.058609162378443935
gate-io_SOL_spot	-1.3655184679725594	-0.5385264054243074	-0.041063555333632484
okex_BTC_spot	-2.6223696828493983	-1.0455970763595328	0.005155133889980457
okex_BTC_futures	1.8246895353714123	0.9912831562755445	-0.0009652476982395797
okex_ETH_spot	16.31002724394233	1.2230837717603826	0.026813131744696794
okex_ETH_futures	0.4977448377789673	0.2791241803911721	-0.053993121833222135
okex_SOL_spot	2.272983167827968	0.7203383972066608	-0.027462508454246803
okex_SOL_futures	-48.16171042897797	-2.6383271799131847	0.2487720649870615

Regression on Crypto Index

y	Coeff	T-stat	Adj_R^2
binance_BTC_spot	53.97920526840757	0.773903822269246	-0.022789620696918478
binance_BTC_futures	-0.04100961166313	-0.01989745589692059	-0.05879887130947781
binance_ETH_spot	0.9483398877107729	1.638271950308719	0.08554869667108034
binance_ETH_futures	0.12059191349807252	0.11415520686018014	-0.05801250611304232
binance_FTT_spot	0.21722880671047745	0.4162124119004287	-0.04814278354956536
binance_FTT_futures	0.2651623532027648	0.5159471423501548	-0.04249913035408759
binance_SOL_spot	4.8342702647415745	0.9103061256142617	-0.009610525085553379
binance_SOL_futures	-8.901194937208496	-1.1426999484212323	0.016703109793725357
bitstamp_BTC_spot	2.905085942481759	1.4834986130174026	0.06253750508276334
bitstamp_ETH_spot	0.09604019856516254	0.11286960906611848	-0.058030656880746934
bitstamp_SOL_spot	0.4575693980057445	0.18743371558452102	-0.05663992720304556
ftx_BTC_spot	0.3702258553579811	0.8161199773762932	-0.01890338520331003
ftx_BTC_futures	0.2720654608743739	0.6632482798382404	-0.03211611158943417
ftx_ETH_spot	5.203520779826082	2.6299323735035567	0.2473829085698337
ftx_ETH_futures	-0.09728215907244296	-0.18196852703696634	-0.05676516565149825
ftx_FTT_spot	-0.20293284852816745	-0.21280745362277426	-0.056010379114924014
ftx_FTT_futures	-0.009776525237228356	-0.009682221431588266	-0.05881769062938336
ftx_SOL_spot	1.7164219074100038	0.34318475182191915	-0.051538484775747406
ftx_SOL_futures	-1.2998688014204025	-0.7894568779356269	-0.02137846341860672
gate-io_BTC_spot	0.014895829933573516	0.08698123621781771	-0.058352516755363126
gate-io_ETH_spot	-0.4642392439095915	-0.7195876669840163	-0.027525911534000524
gate-io_FTT_spot	0.37935596074526484	0.545010495346923	-0.04064069370487733
gate-io_SOL_spot	-0.9814292165511255	-1.2052555471803388	0.02452987274979912
okex_BTC_spot	-1.0758282350714863	-1.3155607715835018	0.039010829180029516
okex_BTC_futures	0.8369512514026614	1.4081263145466716	0.051774169082026034
okex_ETH_spot	-1.1554807798737974	-0.25086035018008296	-0.054918412865851085
okex_ETH_futures	0.049253765881552725	0.08308152937952742	-0.05839378788169114
okex_SOL_spot	-0.16591174089259228	-0.15623535940358818	-0.05730539512614219
okex_SOL_futures	-5.425171532556146	-0.7675216083922487	-0.023361674436594093

Figure 10 Daily autocorrelation Regression on Crypto Index

Figure 11 Daily autocorrelation Regression on Dictionary Index

Figure 10 and 11 presents the results of a regression of daily autocorrelation on the Indexes[39]. We find that the results of the regression are highly inconsistent in different exchanges and different cryptocurrencies, even for the Crypto Index, that better explained the returns. It is possible to observe that only a few markets autocorrelation's can we well explained by our Index, like the Okex-SOL futures with the Dictionary Index and the FTX-ETH spot with the Crypto one. In such specific markets, we see that more than 24% of the variation is explained by the surge of the Indexes with a confidence level of 1%. However, the sign of the correlation in the two markets is different. All these findings together do not form enough evidence to support a sound behavioural relationship between investors sentiment and order flows. This could either be due to the fact that investor's herding behaviour is not affected by market sentiment; or that possible reduction of order splits occurred during the crash compensated the herding effected.

After this analysis we can therefore say that investors did not alter their decision making based on other investors reaction toward some news, but they rather focussed on (or we

[39] To avoid redundancies, the results for the FEARS Index are not presented. The computation can be found on the code presented on the Appendix

may better rather say tunnel visioned) the signal they personally extracted from the news on their own.

3.4 Shortcomings and further improvements

For a complete analysis and a comprehensive work, some weaknesses of the presented elaborations must be pointed out, the most crucial of which is overfitting and bias toward a specific time. Both the regressions and the Indexes are tailored to the specific time when the crash happened, therefore there may be some lack of generalisation towards the more extended market. This issue had been mitigated by expanding the results to different cryptocurrencies and exchanges, but they all refer to the same time period and have the same web research as underlying driver. The words chosen could then be biased as of a backword-looking process: the empirical analysis is based on the facts happened and the specific news posted, and therefore it is possible that results may be related too closely to the specific event. This backward-looking bias may threat the capability of the proposed tools to explain future behaviours. Nonetheless, the present methodology successfully managed to explain a driver of the unorthodox prices' evolution over time. This sets grounds to the construction of a sentiment specific Index that takes incorporates different time periods to better understand the attitudes of crypto investors as well as their evolution over time: as the market grows and therefore becomes more efficient, these irrational behaviours should be less harmful.

A second problem may arise due to the dataset used to build the sentiment Index. Even if Google trends is indeed widely used, we have no evidence supporting the idea that crypto investors would carry out their research using that specific engine, even if it reasonable to assume it due to its large scale and dominant market position. Other platforms such as X or Discord are examples of other communities that may have traces of investors' sentiment, however, due to the limited computation tools and accessibility, an analysis of these socials is not presented, but it leaves space for further analysis.

More sources leave room for a further exploration of the problem, using different alternatives dataset to test the same hypothesis. A further exploration of the problem can be carried out using high frequency search volume, information that accessible to the public only few weeks after the event date. Using a more detailed research volume, other interesting information could be assessed, for example the reaction speed to a specific news and the moment of market or news reading peak; useful to adjust the short-term liquidity needed in a sound, efficient market.

One last point that needs to be discussed is the effectiveness of a regression. Even if the results are relevant and significant, the relationship between the variables analysed is not linked by causality. We can statistically say that the two variables are correlated, but this statistical tool does not grant the hypotheses proposed: the evidence does not imply that the change in the prices is due to a change in the research volume; but that a prince change is tied to a change in the research volume.

Conclusions

The presented work aimed to evaluate the presence of behavioural biases in crypto currencies market to shed a light on the efficiencies of their exchanges as well as prices. As results of this data driven empirical analysis, we can say that the assumption of a *homo oeconomicus*, always rational and capable of correct and consistent choices, is not supported by the evidence. The analysis firstly highlighted the technical excellence of some crypto currencies exchanges, namely Binance, that even in turbulent times were able to offer competitive prices to its investors, and therefore questioned a view by which the bizarre reaction observed in the markets was accountable to lack of liquidity and market microstructure. It is instead possible to state that there is evidence of the behavioural attitudes of crypto investors reported by Al-Mansour[40]. The data show a tendency to overreaction to salient news, in line with the strength and weight framework of Griffin & Tversky[41], highlighting the market players' main conduct as speculative noise traders following heuristics, one of which is "buy on the rumour, sell on the news". These investors are not moved by the same drivers as traditional markets investors, and even seem to show different sentiment triggers, that are uncorrelated to traditional financial jargon, but more closely related to the crypto technological and business environment. We also find that in time of market panic investors do not appear to alter their trading decisions based on market direction, but they rather extract signals from news at a personal level, implying a lower relevance of herding behaviours in tumultuous periods, phenomenon however still present in the market. As results of this analysis, we can state that there is strong evidence of behavioural biases in the crypto currencies market and that they threat the asset prices' efficiency due to the lack of professional, unbiased

[40] Al-Mansour, Bashar (2020). Cryptocurrency Market: Behavioural Finance Perspective. Journal of Asian Finance, Economics and Business , 7(12), 159-168. See 1.7 Biases in cryptocurrencies market for details.
[41] Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. Cognitive Psychology, 24(3), 411–435. https://doi.org/10.1016/0010-0285(92)90013-R

investors that act as price correctors found in traditional financial markets. The methodology presented in this work allows for replicability and therefore scalability of the research, allowing the expansion of the Indexes used incorporating new words when new data is collected to further explore the decision-making process of crypto investors and understand the human bounded rationality when operating in domains of uncertainty. The results of this work underline the need of a better financial literacy in the tested markets and that price efficiency in crypto markets is yet to be achieved, questioning the reliability for long term stable returns due to its speculative drivers. Nevertheless, a further growth in scale as well as an increased participation of unbiased investors may help to mitigate the recurrence of similar nefarious events and reactions.

Bibliography

Al-Mansour, B. (2020). Cryptocurrency Market: Behavioural Finance Perspective. *Journal of Asian Finance, Economics and Business,* 7(12), 159-168.

Aistleitner, M., Kapeller, J., & Steinerberger, S. (2019). Citation patterns in economics and beyond. *Science in Context,* 32(4), 361-380. https://doi.org/10.1017/S0269889720000022

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5, 31–56.

Bambrough, B. (2022, November 13). JPMorgan Reveals Shock Cascade Bitcoin Price Prediction After Stunning FTX Meltdown. *Forbes.* https://www.forbes.com/sites/billybambrough/2022/11/13/jpmorgan-reveals-shock-cascade-bitcoin-price-prediction-after-stunning-ftx-meltdown/

Bodie, Z., Kane, A., & Marcus, A. J. (2021). Investments (12th ed.). *McGraw-Hill Education.*

Chipolina, S. (2024, January 23). Bitcoin price falls 15% following launch of ETFs. *Financial Times*. https://www.ft.com/content/99e406f0-c5af-4a67-9937-afb1ea885565

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies*, 28(1), 1-32. https://doi.org/10.1093/rfs/hhu072

Edmans, A., Garcia, D., & Norli, Ø. (2007). Sports Sentiment and Stock Returns. *Journal of Finance*, 62(4), 1967-1998.

Ferreira, P. (2023, May 9). The FTX Full Story: All You Need to Know. *Finance Magnates*. https://www.financemagnates.com/cryptocurrency/the-ftx-full-story-all-you-need-to-know/

Foucault, T., Pagano, M., & Röell, A. (2013). Market Liquidity: Theory, Evidence, and Policy. *Oxford University Press.*

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 25(2), 383-417.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology,* 24(3), 411–435. https://doi.org/10.1016/0010-0285(92)90013-R

Haugen, R. A., & Lakonishok, J. (1988). The Incredible January Effect: The Stock Market's Unsolved Mystery. *Homewood, Ill.: Dow Jones-Irwin*.

Keller, L. (2022, November 10). Bitcoin Hits Lowest Price in 10 Years Following Binance, FTX Acquisition. *Forkast News*. https://forkast.news/bitcoin-lowest-price-years-binance-ftx-acquisition/

Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. Journal of Economic Perspectives, 17(1), 59-82.

Nelson, D., & Baker, N. (2023, September 20). Breaking Down the Infamous Alameda Balance Sheet. *CoinDesk*. https://www.coindesk.com/policy/2023/09/20/breaking-down-the-infamous-alameda-balance-sheet/

Pareto, V. (1906). Manuale di Economia Politica *[Critical Edition, Eds. A. Montesano, A. Zanni, & L. Bruni]*.

Prospect theory: An analysis of decision under risk. *JSTOR* https://www.jstor.org/stable/1914185

Toth, B., Palit, I., Lillo, F., & Farmer, J. D. (2014, June 15). Why Is Equity Order Flow so Persistent? *SSRN*. https://ssrn.com/abstract=2460731

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291. https://doi.org/10.2307/1914185

Appendix A

PYTHON CODE

DATA EXPLORATION

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import os

exchange_list = ['binance', 'bitstamp', 'ftx', 'gate-io', 'okex']
symbol_list = ['BTC', 'ETH', 'FTT', 'SOL']
contract_type = ['spot', 'futures']

orderbooks = {}
trades = {}
for exchange in exchange_list:
    orderbooks[exchange] = {}
    trades[exchange] = {}
    for symbol in symbol_list:
        orderbooks[exchange][symbol] = {}
        trades[exchange][symbol] = {}

os.listdir('FTX_data')
orderbooks_paths = []
trades_paths = []
for filename in os.listdir('FTX_data'):
    path = 'FTX_data/' + filename
    if 'csv' not in filename: continue
    df = pd.read_csv(path)
    df['time'] = df['time'].astype('datetime64[ns]')
    if 'orderbooks' in filename or 'quotes' in filename:
        for exchange in exchange_list:
            for symbol in symbol_list:
                if (exchange in filename) and (symbol in filename):
                    if 'futures' in filename or 'PERP' in filename or 'swap' in filename:
                        orderbooks[exchange][symbol]['futures'] = df
                    else:
                        orderbooks[exchange][symbol]['spot'] = df
    else:
        for exchange in exchange_list:
            for symbol in symbol_list:
                if (exchange in filename) and (symbol in filename):
                    if 'futures' in filename or 'PERP' in filename or 'swap' in filename:
                        trades[exchange][symbol]['futures'] = df
                    else:
```

```
                trades[exchange][symbol]['spot'] = df

orderbooks['ftx']['ETH']['spot'].head()

orderbooks_triples = []
trades_triples = []
for exchange in exchange_list:
    for symbol in symbol_list:
        for contract in contract_type:
            if contract in orderbooks[exchange][symbol].keys():
                orderbooks_triples.append((exchange, symbol, contract))
            if contract in trades[exchange][symbol].keys():
                trades_triples.append((exchange, symbol, contract))

orderbooks_triples

#check for duplicates
orderbooks['bitstamp']['BTC']['spot'][orderbooks['bitstamp']['BTC']['spot'].duplicated(subset='ti
me') == True]

#account missing values, example
df_miss_example = orderbooks['ftx']['BTC']['spot']
t_index = pd.date_range(df_miss_example['time'].min(), df_miss_example['time'].max(),
freq='min')
df_miss_example_2 = df_miss_example.set_index('time').reindex(t_index)
df_miss_example_2[df_miss_example_2.isnull().any(axis=1)]
#I fill with ffill
df_miss_example_2.fillna(method='ffill').loc['2022-11-09']

#set the time as index, and filter records earlier than 2022-11-01
for exchange, symbol, contract in orderbooks_triples:
    df = orderbooks[exchange][symbol][contract]
    df.drop_duplicates(subset='time', keep='last', inplace=True)
    t_index = pd.date_range(df['time'].min(), df['time'].max(), freq='min')
    df = df.set_index('time').reindex(t_index)
    df = df.loc['2022-11-01':]
    orderbooks[exchange][symbol][contract] = df
for exchange, symbol, contract in trades_triples:
    trades[exchange][symbol][contract].drop_duplicates(subset='time', keep='last', inplace=True)
    df = trades[exchange][symbol][contract]
    df.drop_duplicates(subset='time', keep='last', inplace=True)
    t_index = pd.date_range(df['time'].min(), df['time'].max(), freq='min')
    df = df.set_index('time').reindex(t_index)
    df = df.loc['2022-11-01':]
    trades[exchange][symbol][contract] = df

orderbooks['binance']['FTT']['spot']
orderbooks['binance']['BTC']['spot'].tail()
```

```
orderbooks_summary = pd.DataFrame(columns=['exchange', 'symbol', 'contract_type', 'len',
'num_miss'])
for exchange, symbol, contract in orderbooks_triples:
    df = orderbooks[exchange][symbol][contract]
    orderbooks_summary.loc[len(orderbooks_summary)] = [exchange, symbol, contract, len(df),
                                              len(df[df.isnull().any(axis=1)])]
orderbooks_summary

trades_summary = pd.DataFrame(columns=['exchange', 'symbol', 'contract_type', 'len',
'num_miss'])
for exchange, symbol, contract in trades_triples:
    df = trades[exchange][symbol][contract]
    trades_summary.loc[len(trades_summary)] = [exchange, symbol, contract, len(df),
                                        len(df[df.isnull().any(axis=1)])]
trades_summary

#use ffill for orderbooks
for exchange, symbol, contract in orderbooks_triples:
    orderbooks[exchange][symbol][contract].fillna(method='ffill', inplace=True)

#check
orderbooks_summary = pd.DataFrame(columns=['exchange', 'symbol', 'contract_type', 'len',
'num_miss'])
for exchange, symbol, contract in orderbooks_triples:
    df = orderbooks[exchange][symbol][contract]
    orderbooks_summary.loc[len(orderbooks_summary)] = [exchange, symbol, contract, len(df),
                                              len(df[df.isnull().any(axis=1)])]
orderbooks_summary

orderbooks_summary[orderbooks_summary['contract_type'] == 'spot'].sort_values(
    by=['symbol', 'exchange', 'contract_type'])
#FTT is the only coin with different lenght as of the total crash

for exchange, symbol, contract in orderbooks_triples:
    df = orderbooks[exchange][symbol][contract]
    print(exchange, symbol, contract)
    print(list(df.columns))

#the code presents two different parts as FTX data has different labelling
for exchange, symbol, contract in orderbooks_triples:
    df = orderbooks[exchange][symbol][contract]
    if exchange == 'ftx':
        df['mid'] = (df['ask_price'] + df['bid_price']) / 2
        df['relative_spread'] = (df['ask_price'] - df['bid_price']) / df['mid']
        df['absolute_spread'] = df['ask_price'] - df['bid_price']
    else:
        df['mid'] = (df['top_ask'] + df['top_bid']) / 2
        df['relative_spread'] = (df['top_ask'] - df['top_bid']) / df['mid']
        df['absolute_spread'] = df['top_ask'] - df['top_bid']
    orderbooks[exchange][symbol][contract] = df
```

```
orderbooks['ftx']['ETH']['spot'].head()


# columns of trades file are identical
for exchange, symbol, contract in trades_triples:
   df = trades[exchange][symbol][contract]
   print(exchange, symbol, contract)
   print(list(df.columns))

for exchange, symbol, contract in trades_triples:
   df = trades[exchange][symbol][contract]
   df['vwap'] = df['volume'] / df['amount']
   df['abs_return']=abs(df['vwap'].pct_change())
   df['Amihud_ratio']=df['abs_return']/df['volume']
   df['number_trades_sell'] = df['number_trades'] - df['number_trades_buy']
   df["Roll_measure"] = 2 * (-
(df['vwap'].pct_change().shift(1).rolling(window=2).cov(df['vwap'].pct_change()))**(1/2))
   df['amount_sell'] = df['amount'] - df['amount_buy']
   df['volume_sell'] = df['volume'] - df['volume_buy']
   trades[exchange][symbol][contract] = df



trades['binance']['BTC']['spot'].head()

colors = ['#FF0000', '#56423D', '#BEA6A0', '#00A0FF', '#009E76']

#plot the orderbooks mid prices
fig, axes = plt.subplots(4, 1, figsize=(15, 10))
plt.subplots_adjust(hspace=0.8)
for index, symbol in enumerate(symbol_list):
   ax = axes.flat[index]
   ax.set_title(symbol + '_mid')
   for index, exchange in enumerate(exchange_list):
      if 'spot' not in orderbooks[exchange][symbol].keys(): continue
      book = orderbooks[exchange][symbol]['spot']
      ax.plot(book.index, book['mid'], c=colors[index % len(colors)], label=exchange)
   ax.xaxis.set_tick_params(rotation=45)
   ax.legend()

#plot the trades vwap
fig, axes = plt.subplots(4, 1, figsize=(15, 10))
plt.subplots_adjust(hspace=0.8)
for index, symbol in enumerate(symbol_list):
   ax = axes.flat[index]
   ax.set_title(symbol + '_vwap')
   for index, exchange in enumerate(exchange_list):
      if 'spot' not in trades[exchange][symbol].keys(): continue
      trade = trades[exchange][symbol]['spot']
```

```
        ax.plot(trade.index, trade['vwap'], c=colors[index % len(colors)], label=exchange)
    ax.xaxis.set_tick_params(rotation=45)
    ax.legend()

#plot the trades Amihud ratio
fig, axes = plt.subplots(4, 1, figsize=(15, 10))
plt.subplots_adjust(hspace=0.8)
for index, symbol in enumerate(symbol_list):
    ax = axes.flat[index]
    ax.set_title(symbol + '_Amihud_ratio')
    for index, exchange in enumerate(exchange_list):
        if 'spot' not in trades[exchange][symbol].keys(): continue
        trade = trades[exchange][symbol]['spot']
        ax.plot(trade.index, trade['Amihud_ratio'], c=colors[index % len(colors)], label=exchange)

    ax.set_ylim(0, 0.01)
    ax.xaxis.set_tick_params(rotation=45)
    ax.legend()

#plot the trades roll's measure
fig, axes = plt.subplots(4, 1, figsize=(15, 10))
plt.subplots_adjust(hspace=0.8)
for index, symbol in enumerate(symbol_list):
    ax = axes.flat[index]
    ax.set_title(symbol + '_Roll_measure')
    for index, exchange in enumerate(exchange_list):
        if 'spot' not in trades[exchange][symbol].keys(): continue
        trade = trades[exchange][symbol]['spot']
        ax.plot(trade.index, trade['Roll_measure'], c=colors[index % len(colors)], label=exchange)

    ax.set_ylim(0, -0.25)
    ax.xaxis.set_tick_params(rotation=45)
    ax.legend()

#plot the orderbooks absolute spreads
fig, axes = plt.subplots(4, 1, figsize=(15, 10))
plt.subplots_adjust(hspace=0.8)
for index, symbol in enumerate(symbol_list):
    ax = axes.flat[index]
    ax.set_title(symbol + '_absolute_spread')
    for index, exchange in enumerate(exchange_list):
        if 'spot' not in orderbooks[exchange][symbol].keys(): continue
        book = orderbooks[exchange][symbol]['spot']
        ax.plot(book.index, book['absolute_spread'], c=colors[index % len(colors)], label=exchange)
    ax.xaxis.set_tick_params(rotation=45)
    ax.legend()


#Dictionary for volatilities
Dailiy_Volatilities = {}
```

```
for exchange, symbol, contract in trades_triples:
    df = trades[exchange][symbol]["spot"]

    # Calculate daily volatility using resample
    df["daily_volatility"] = df['abs_return'].resample('1D').var()
    Dailiy_Volatilities[exchange+"_"+symbol+"_"+"spot"] = df["daily_volatility"].dropna()



# Get the list of keys (symbols or assets) from the dictionary
vol_list = list(Dailiy_Volatilities.keys())

fig, axes = plt.subplots(4, 5, figsize=(20, 20))
plt.subplots_adjust(hspace=0.8)

for index, vol in enumerate(vol_list):
    row_index = index // 5
    col_index = index % 5
    ax = axes[row_index, col_index]
    ax.set_title(vol + " daily volatility")
    ax.plot(Dailiy_Volatilities[vol], c=colors[index % len(colors)])
    ax.xaxis.set_tick_params(rotation=45)
plt.show()

#Dictionary for market_direction
market_direction = {}

for exchange, symbol, contract in trades_triples:
    df = trades[exchange][symbol][contract]
    df["market_direction"] = df["amount_buy"]-df["amount_sell"]
    market_direction[exchange+"_"+symbol+"_"+contract] = df["market_direction"]

aut_list = list(market_direction.keys())

lags = [1, 2, 5, 10, 20, 50, 100]

autocorrelation_results = {}

for asset in aut_list:
    autocorrelation_results[asset] = {}
    for lag in lags:
        autocorr_value = market_direction[asset].autocorr(lag=lag)
        autocorrelation_results[asset][f'Autocorrelation at Lag {lag}'] = autocorr_value


fig, axes = plt.subplots(5, 6, figsize=(20, 16))
plt.subplots_adjust(hspace=0.8)

for index, asset in enumerate(aut_list):
```

```
    row_index = index // 6
    col_index = index % 6
    ax = axes[row_index, col_index]
    autocorrelation_data = autocorrelation_results[asset]
    lags = [int(key.split()[-1]) for key in autocorrelation_data.keys()]
    autocorrelation_values = list(autocorrelation_data.values())
    ax.plot(lags, autocorrelation_values, marker='o', linestyle='-', color='b')
    ax.set_xlabel('Lag')
    ax.set_ylabel('Autocorrelation')
    ax.grid(True)
    ax.set_title(f'Autocorrelation for {asset}')
plt.tight_layout()
plt.show()

daily_data_dict = {}
for asset in aut_list:
    daily_data = market_direction[asset].groupby(market_direction[asset].index.date)
    daily_data_dict[asset] = {}
    for date, daily_group_df in daily_data:
        daily_data_dict[asset][date] = daily_group_df.dropna()

date_list=list(daily_data_dict["binance_BTC_futures"])


daily_autocorrelation_results = {}
for asset in aut_list:
    daily_autocorrelation_results[asset] = {}
    for day in date_list:
        if asset in daily_data_dict and day in daily_data_dict[asset]:
            autocorr_value = daily_data_dict[asset][day].autocorr(lag=1)
            daily_autocorrelation_results[asset][day] = autocorr_value
        else:
            # Handle missing data here, such as setting a default value
            daily_autocorrelation_results[asset][day] = None


fig, axes = plt.subplots(5, 6, figsize=(20, 16))
plt.subplots_adjust(hspace=0.8)

for index, asset in enumerate(aut_list):
    row_index = index // 6
    col_index = index % 6
    ax = axes[row_index, col_index]
    autocorrelation_data = daily_autocorrelation_results[asset]
    time = list(autocorrelation_data.keys())
    autocorrelation_values = list(autocorrelation_data.values())

    # Plot the autocorrelation data
    ax.plot(time, autocorrelation_values, marker='o', linestyle='-', color='b')
```

```
    ax.set_xlabel('Date')
    ax.set_ylabel('Autocorrelation')
    ax.grid(True)
    ax.set_title(f'Autocorrelation {asset}')
    ax.xaxis.set_tick_params(rotation=45)

plt.tight_layout()
plt.show()

dir_path = './output/'

path = dir_path + 'trades_spot_vwap_daily_mean' + '.csv'
df = pd.concat([trades[exchange][symbol]['spot']['vwap'].resample('1D').mean().rename(
    exchange + '_' + symbol + '_spot_' + 'vwap_mean')
    for exchange, symbol, contract in trades_triples if contract == 'spot'], axis=1)
df.to_csv(path)


path = dir_path + 'daily_autocorrelation' + '.csv'
df = pd.DataFrame(daily_autocorrelation_results)
df.to_csv(path)
```

TREND EXTRACTION

```
import time
import pandas as pd
import pytrends

keywords = pd.read_csv("Words_crypto.csv")

from pytrends.request import TrendReq
pytrends = TrendReq(hl='en-US', tz=0, retries=10)

# Perform searches
wait = 6 # in seconds
print('Number of queries to do: ', len(keywords))

trends =pd.DataFrame()

#keywords.drop(keywords.index[0:1], inplace=True) #USE THIS LINE TO REMOVE THE WORDS EXTRACTED IN CASE THE LOOP STOPS

for k in keywords["WORD - crypto"]:
    time.sleep(wait)
    pytrends.build_payload([k], timeframe='2022-11-01 2022-11-20',gprop="")
    trends[k] = pytrends.interest_over_time()[k]
    print(k," successfully extracted!")
```

```
trends.to_csv('Trends_crypto.csv')

Results = pd.read_csv("Trends_crypto.csv")

#CODE FOR FEARS

#keywords = pd.read_csv("FEARS.csv")

#for k in keywords["FEARS index"]:
   #time.sleep(wait)
   #pytrends.build_payload([k], timeframe='2022-11-01 2022-11-20',gprop="")
   #trends[k] = pytrends.interest_over_time()[k]
   #print(k," successfully extracted!")

#trends.to_csv('Trends_FEARS.csv')

#Results = pd.read_csv("Trends_FEARS.csv")
```

INDEX BUILDING

```
import pandas as pd
import numpy as np
import statsmodels.api as sm

vwap_prices = pd.read_csv("output/trades_spot_vwap_daily_mean.csv")

vwap_prices.head()

list=vwap_prices.columns.values.tolist()

list=list[1:]

log_returns=pd.DataFrame()

for k in list:
   log_returns[k]=np.log(vwap_prices[k])-np.log(vwap_prices[k].shift(1))

trends= pd.read_csv("Trends_dictonary.csv")

list_words=trends.columns.values.tolist()

list_words=list_words[1:]

daily_research_change=pd.DataFrame()

for k in list_words:
   daily_research_change[k]=np.log(trends[k])-np.log(trends[k].shift(1))
```

```
#try the expanding regression

Y = log_returns["binance_BTC_spot_vwap_mean"]
X = trends

results=pd.DataFrame()

for k in np.arange(4,21):
   for j in list_words:
      X=trends[j]
      Y_subset= Y.iloc[1:k]
      X_subset= X.iloc[1:k]
      X_subset= sm.add_constant(X_subset)
      model= sm.OLS(Y_subset, X_subset).fit()

      t_stats= model.tvalues.iloc[1]

      results_temp= {
      "Word":j,
      "Observations": k,
      "Coefficients": model.params.iloc[1],
      "intercepts": model.params.iloc[0],
      "T-stats": t_stats
      }
      results_temp=pd.DataFrame([results_temp])
      results=pd.concat([results,results_temp])

results=results.set_index(results["Observations"])
results= results.drop(columns=["Observations"])

Statistics=pd.DataFrame()

for word in list_words:

   Filter= results[results["Word"] == word]
   Stats=Filter.describe()
   Keep={
     "Word": word,
     "Mean T-stat":Stats.iloc[1,2],
     "SD T-stat":Stats.iloc[2,2],
     }
   Keep=pd.DataFrame([Keep])
   Statistics=pd.concat([Statistics,Keep])

Refine = Statistics[ (Statistics["Mean T-stat"] < -1.6)]

final_list=Refine["Word"].values.tolist()


Final_trends_dictionary = pd.DataFrame()
```

```
for word in final_list:
    a = pd.DataFrame(trends[word])
    Final_trends_dictionary = pd.concat([Final_trends_dictionary, a], axis=1)

Final_trends_dictionary = pd.concat([ trends["date"], Final_trends_dictionary], axis=1)

Final_trends_dictionary.to_csv('Trends_dictionary_final.csv', index=False)
```

REGRESSIONS

```
import pandas as pd
import numpy as np
import statsmodels.api as sm
import matplotlib.pyplot as plt

vwap_prices = pd.read_csv("output/trades_spot_vwap_daily_mean.csv")
vwap_prices = vwap_prices.rename(columns = {"Unnamed: 0" : "date"})
vwap_prices = vwap_prices.set_index("date")

returns = vwap_prices.pct_change()

trends_fears = pd.read_csv("Trends_FEARS.csv")
trends_crypto = pd.read_csv("Trends_crypto.csv")
trends_dictionary = pd.read_csv("Trends_dictionary_final.csv")

trends_fears = trends_fears.set_index('date')
log_trends_change_fears = pd.DataFrame()


for col in trends_fears.columns :
    res = np.log1p(trends_fears[col]) - np.log1p(trends_fears[col].shift(1))
    log_trends_change_fears[col] = res

fears_index = pd.DataFrame(log_trends_change_fears.mean(axis = 1), columns = ['fears_index'])

y = returns['binance_BTC_spot_vwap_mean'].dropna()
x = fears_index.dropna()
x = sm.add_constant(x)

model = sm.OLS(y,x).fit() #linear regression
print(model.summary())

#I drop the column bitstamp_FTT_spot_vwap_mean because nan in the beginning make the size
of X and y different
returns = returns.loc[:, returns.columns != 'bitstamp_FTT_spot_vwap_mean']

results_FEARS = pd.DataFrame(columns=['y',  'Coeff', 'T-stat', 'Adj_R^2'])
for col in returns.columns:
```

```
    y = returns[col].dropna()
    x = fears_index.dropna()
    X = sm.add_constant(x)

    model = sm.OLS(y,X).fit() #linear regression
    coefficient = model.params[1]
    t_stat = model.tvalues[1]
    adjusted_R_squared = model.rsquared_adj
    dic = {'y' : col,
        'Coeff' : coefficient,
        'T-stat':t_stat,
        'Adj_R^2': adjusted_R_squared}
    df = pd.DataFrame(dic, index = [0])
    results_FEARS = pd.concat([results_FEARS, df], ignore_index=True)

trends_crypto= trends_crypto.set_index('date')
log_trends_change_crypto = pd.DataFrame()

for col in trends_crypto.columns :
   res = np.log1p(trends_crypto[col]) - np.log1p(trends_crypto[col].shift(1))
   log_trends_change_crypto[col] = res

crypto_index = pd.DataFrame(log_trends_change_crypto.mean(axis = 1), columns = ['crypto_index'])

results_crypto = pd.DataFrame(columns=['y',  'Coeff', 'T-stat', 'Adj_R^2'])

for col in returns.columns:
    y = returns[col].dropna()
    x = crypto_index.dropna()
    X = sm.add_constant(x)


    model = sm.OLS(y,X).fit() #linear regression
    coefficient = model.params[1]
    t_stat = model.tvalues[1]
    adjusted_R_squared = model.rsquared_adj
    dic = {'y' : col,
        'Coeff' : coefficient,
        'T-stat':t_stat,
        'Adj_R^2': adjusted_R_squared}
    df = pd.DataFrame(dic, index = [0])
    results_crypto = pd.concat([results_crypto, df], ignore_index=True)

trends_dictionary= trends_dictionary.set_index('date')
log_trends_change_dictionary = pd.DataFrame()

for col in trends_dictionary.columns :
    res = np.log1p(trends_dictionary[col]) - np.log1p(trends_dictionary[col].shift(1))
    log_trends_change_dictionary[col] = res
```

```
dictionary_index = pd.DataFrame(log_trends_change_dictionary.mean(axis = 1), columns =
['dictionary_index'])

results_dictionary = pd.DataFrame(columns=['y', 'Coeff', 'T-stat', 'Adj_R^2'])

for col in returns.columns:
   y = returns[col].dropna()
   x = dictionary_index.dropna()
   X = sm.add_constant(x)


   model = sm.OLS(y,X).fit() #linear regression
   coefficient = model.params[1]
   t_stat = model.tvalues[1]
   adjusted_R_squared = model.rsquared_adj
   dic = {'y' : col,
      'Coeff' : coefficient,
      'T-stat':t_stat,
      'Adj_R^2': adjusted_R_squared}
   df = pd.DataFrame(dic, index = [0])
   results_dictionary = pd.concat([results_dictionary, df], ignore_index=True)

autocorrelation = pd.read_csv("output/daily_autocorrelation.csv")
autocorrelation = autocorrelation.rename(columns = {"Unnamed: 0" : "date"})
autocorrelation = autocorrelation.set_index("date")

autocorrelation_change= autocorrelation.pct_change()

#I drop the column bitstamp_FTT_spot_vwap_mean because nan in the beginning make the size
of X and y different
autocorrelation_change = autocorrelation_change.loc[:, autocorrelation_change.columns !=
'bitstamp_FTT_spot']

results_autocorrelation_FEARS = pd.DataFrame(columns=['y', 'Coeff', 'T-stat', 'Adj_R^2'])
for col in autocorrelation_change.columns:
   y = autocorrelation_change[col].dropna()
   x = fears_index.dropna()
   X = sm.add_constant(x)

   model = sm.OLS(y,X).fit() #linear regression
   coefficient = model.params[1]
   t_stat = model.tvalues[1]
   adjusted_R_squared = model.rsquared_adj
   dic = {'y' : col,
      'Coeff' : coefficient,
      'T-stat':t_stat,
      'Adj_R^2': adjusted_R_squared}
   df = pd.DataFrame(dic, index = [0])
```

```
    results_autocorrelation_FEARS = pd.concat([results_autocorrelation_FEARS, df], ignore_index=True)

results_autocorrelation_crypto = pd.DataFrame(columns=['y',  'Coeff', 'T-stat', 'Adj_R^2'])
for col in autocorrelation_change.columns:
    y = autocorrelation_change[col].dropna()
    x = crypto_index.dropna()
    X = sm.add_constant(x)

    model = sm.OLS(y,X).fit() #linear regression
    coefficient = model.params[1]
    t_stat = model.tvalues[1]
    adjusted_R_squared = model.rsquared_adj
    dic = {'y' : col,
       'Coeff' : coefficient,
       'T-stat':t_stat,
       'Adj_R^2': adjusted_R_squared}
    df = pd.DataFrame(dic, index = [0])
    results_autocorrelation_crypto = pd.concat([results_autocorrelation_crypto, df], ignore_index=True)

results_autocorrelation_dictionary = pd.DataFrame(columns=['y',  'Coeff', 'T-stat', 'Adj_R^2'])
for col in autocorrelation_change.columns:
    y = autocorrelation_change[col].dropna()
    x = dictionary_index.dropna()
    X = sm.add_constant(x)

    model = sm.OLS(y,X).fit() #linear regression
    coefficient = model.params[1]
    t_stat = model.tvalues[1]
    adjusted_R_squared = model.rsquared_adj
    dic = {'y' : col,
       'Coeff' : coefficient,
       'T-stat':t_stat,
       'Adj_R^2': adjusted_R_squared}
    df = pd.DataFrame(dic, index = [0])
    results_autocorrelation_dictionary= pd.concat([results_autocorrelation_dictionary, df], ignore_index=True)


#results_FEARS
fig, ax = plt.subplots(figsize=(6, 4.5))
ax.axis('tight')
ax.axis('off')
ax.set_title('Regression on FEARS Index')
ax.table(cellText=results_FEARS.values, colLabels=results_FEARS.columns, loc='center')
plt.savefig('results_FEARS.png', dpi=500, bbox_inches='tight', pad_inches=0.05)

#results_dictionary
fig, ax = plt.subplots(figsize=(6, 4.5))
```

```
ax.axis('tight')
ax.axis('off')
ax.set_title('Regression on Dictionary Index')
ax.table(cellText=results_dictionary.values, colLabels=results_dictionary.columns, loc='center')
plt.savefig('results_dictionary.png', dpi=500, bbox_inches='tight', pad_inches=0.05)

#results_crypto
fig, ax = plt.subplots(figsize=(6, 4.5))
ax.axis('tight')
ax.axis('off')
ax.set_title('Regression on Crypto Index')
ax.table(cellText=results_crypto.values, colLabels=results_crypto.columns, loc='center')
plt.savefig('results_crypto.png', dpi=500, bbox_inches='tight', pad_inches=0.05)

#results_autocorrelation_FEARS
fig, ax = plt.subplots(figsize=(6, 6.6))
ax.axis('tight')
ax.axis('off')
ax.set_title('Regression on FEARS Index')
ax.table(cellText=results_autocorrelation_FEARS.values,
colLabels=results_autocorrelation_FEARS.columns, loc='center')
plt.savefig('results_autocorrelation_FEARS.png', dpi=500, bbox_inches='tight',
pad_inches=0.05)

#results_autocorrelation_dictionary
fig, ax = plt.subplots(figsize=(6, 6.6))
ax.axis('tight')
ax.axis('off')
ax.set_title('Regression on Dictionary Index')
ax.table(cellText=results_autocorrelation_dictionary.values,
colLabels=results_autocorrelation_dictionary.columns, loc='center')
plt.savefig('results_autocorrelation_dictionary.png', dpi=500, bbox_inches='tight',
pad_inches=0.05)

#results_autocorrelation_crypto
fig, ax = plt.subplots(figsize=(6, 6.6))
ax.axis('tight')
ax.axis('off')
ax.set_title('Regression on Crypto Index')
ax.table(cellText=results_autocorrelation_crypto.values,
colLabels=results_autocorrelation_crypto.columns, loc='center')
plt.savefig('results_autocorrelation_crypto.png', dpi=500, bbox_inches='tight',
pad_inches=0.05)
```

Summary

Traditional mainstream financial literature is based on the assumptions that economic agents are perfectly rational and are always able to assess financial decisions in their best interest pursing a utility maximisation approach. This concept is often synthesized under the definition of *Homo oeconomicus.* This element has been a building block of financial literature. However, the groundbreaking research of Daniel Kahneman and Amos Tversky, using insights from other disciplines like psychology, questioned this assumption defining a new research area called Behavioural Economics, in which the deviation from rationality are observed and analysed to explain real world phenomena. Following the path of this new research field, the present work wants to explain the bizarre reaction observed on crypto markets during the crash of FTX, a major player in the crypto environment, observed during the month of November 2022.

To understand the rationale of this work, a first simple rational pricing model is introduced. Prices can be seen as present value of future cash flows forecasts, that are conditional on the information available at the current time.

$$\boldsymbol{P_0} = \sum_{\boldsymbol{t=1}}^{\infty} \frac{\boldsymbol{E[CF_t|I_0]}}{\boldsymbol{(1+r)^t}}$$

According to rational theories, agents are able to perfectly analyse the current information, create forecasts of the prices, and assess them according to a mathematical model. The market generated from this process will then reflect investor's opinion on a specific asset.

Information reflected are key to assess the soundness of a market, as it is possible to refer to the concept of Efficient Market Theory. The idea is that: information is incorporated in a fast way, prices are a spectrum that information, and it is difficult to beat the market with active trading strategies. Many critics have been casted towards this theory, as markets presents exploitable inconsistencies. These anomalies were justified by researchers by information asymmetry and behavioural biases, key element of this work.

Granted the possible presence of deviations, academia has then started to define a market as efficient if, even in cases of common mispricing errors from investors, these are corrected by unbiased investors that therefore make the market efficient. Behavioural Finance is that field of economic research that investigates biases in the markets, contributing to market efficiency by formalizing common mispricing that, if significant, are exploited by this latter kind of investors.

Recent academic research highlighted, in a rather qualitative way, the vulnerability of crypto markets to behavioural biases due to the attitude shown by its player that show herding behaviour, use of heuristics and risk-taking behaviour according to Prospect Theory. This work wants then to cast a light using a quantitative, data driven approach and test these findings.

The analysis is carried out on a market making dataset, that includes a list of all trades and orderbooks of several cryptocurrencies markets during November 2022, time when the FTX business passed from being a market leader to filing a bankruptcy claim, generating panic and irrational pricings all over the crypto markets. A single piece of news started a spiral that could have wiped out investors capital even if there were no big signals of reduced future cash flows.

The idea that news that should not have any implications on market prices but instead have an impact is summarised in the strength and weight framework. According to this theory, investors do not update the prices according to a statistically optimal Bayesian updating, but they tend to take more into account salient news that have low relevance, rather than news that have low coverage that however share big implications on forecasts.

$$E(\boldsymbol{CF_{t+1}}|\boldsymbol{S_i}) = \boldsymbol{\mu}\left(\frac{\boldsymbol{1}/\boldsymbol{\sigma_p^2}}{\boldsymbol{1}/\boldsymbol{\sigma_p^2} + \boldsymbol{1}/\boldsymbol{\sigma_{s,i}^2}}\right) + \boldsymbol{S_i}\left(\frac{\boldsymbol{1}/\boldsymbol{\sigma_{s,i}^2}}{\boldsymbol{1}/\boldsymbol{\sigma_p^2} + \boldsymbol{1}/\boldsymbol{\sigma_{s,i}^2}}\right)$$

In such framework, $\boldsymbol{\sigma_{s,i}^2}$ is underestimated and $\boldsymbol{S_i}$ presents a biased error component. This framework gives ground to an empirical analysis, seeking evidence of this behaviour.

This work operates with a data driven approach. The dataset available for the tests includes the trades and orderbook of the 4 major crypto coins (Bitcoin, Ethereum, FTT and Solana) in 5 major exchanges at the time of the crash (Binance, Bitstamp, FTX, Gate-io and Okex).

After a refinement of the data, we find the following time series:

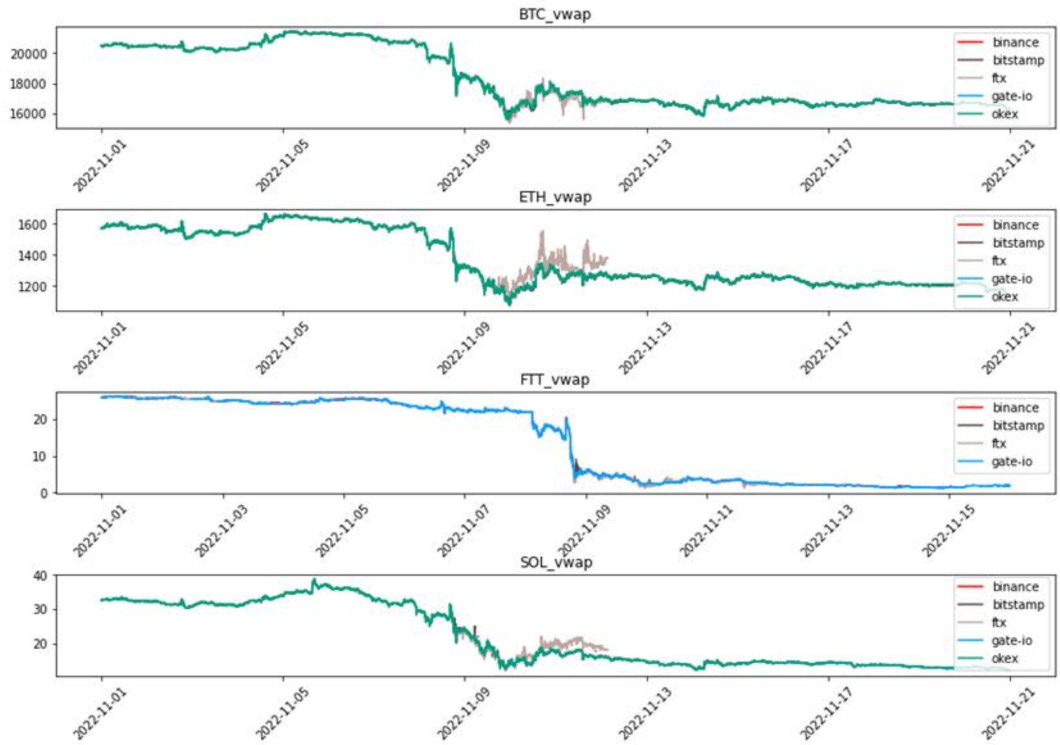

Figure 1 Volume-weighted average prices

It is found that the law of one price mostly holds for the 4 different coins, and that all the markets show a common reaction, mainly: a first decline in the price was followed by an initial recover; a crash starting around November 8th 2022, when all the cryptocurrencies showed a decline in price. The dynamic of this dump is different for each asset, but it presents common factors: one of which is a major low around the 9th of November, but subsequently an upward correction can be noted. This pattern is consistent with a behavioural argument in which overconfidence impact prices as investors may, at first, not have captured the true signal of the information stream, that therefore is later corrected; or the signal extracted could have been biased negatively, yielding to the same correction.

Before proceeding with the behavioural analysis, a report on liquidity using the Amihud ratio is presented to assess whether liquidity problems may have caused the unusual reaction. We do not observe relevant anomalies in the markets with the ratio, that expresses the impact of trades on market prices and magnitude needed to move prices, being relatively low, especially for the most technically efficient markets.

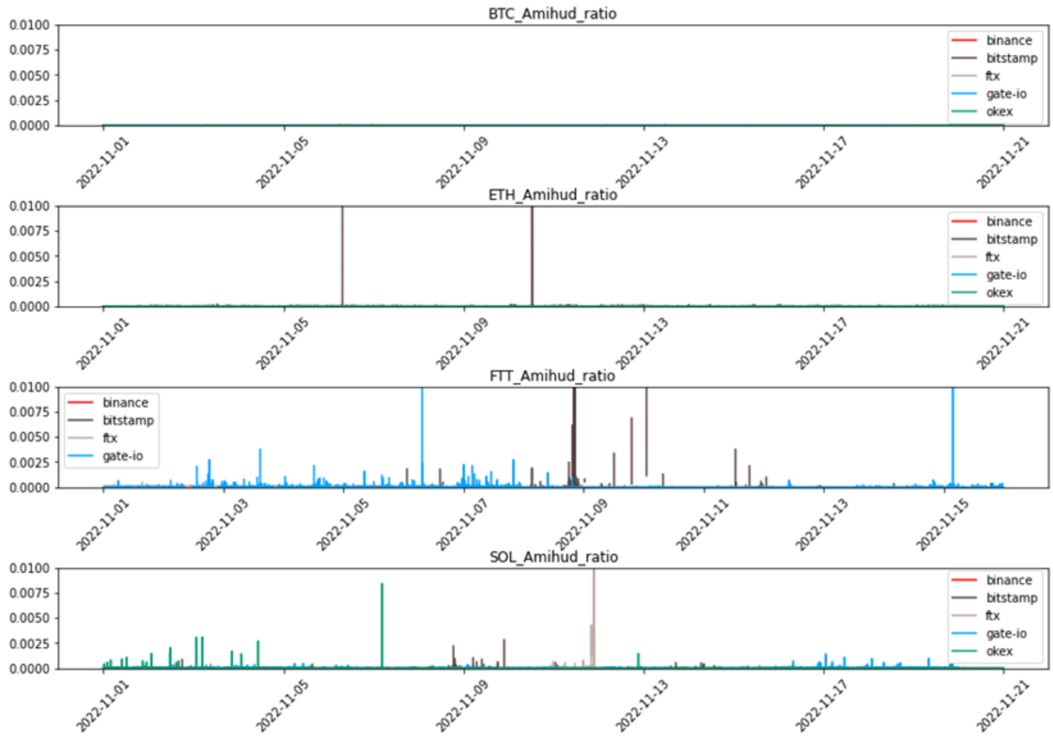

Figure 2 Amihud ratios

Understood that liquidity and market micro-structure cannot explain the crash, a sentiment analysis is performed to explore investors behaviours. Following the methodology of Da et al[1], three different word lists are constructed, words that should be able to track investors negative sentiment. The first list is based on successful academia, and it is defined as FEARS Index. This list of words included financial jargon, mostly related to macroeconomic drivers. The second index is built following the same construction methodology as the FEARS Index, refining the Loughran-McDonald Master Dictionary, a constantly updated dictionary that labels words relatively to their financial sentiment charge. A third list is built according to the most present keywords in posts and news crypto specific related to that period. These lists are used to build indexes based on the volume of Google trend research, metric that should represent the salience of the information on a specific day.

[1] Da, Zhi & Engelberg, Joseph & Gao, Pengjie. (2015). The sum of all FEARS investor sentiment and asset prices. Review of Financial Studies. 28. 1-32. 10.1093/rfs/hhu072.

Another analysis is set up in relation to the autocorrelation of order flows following the rationale that this phenomenon, if present, is due to order splitting or herding behaviour, biased reported to be significant in cryptocurrencies markets.

The results of this empirical study show that the Crypto Index is able to track market return and explain from 12 to 26% of the variation, with a confidence level ranging from 5 to 1% alpha level, whereas the two other Indexes fail in giving significant evidence.

Regression on Crypto Index

y	Coeff	T-stat	Adj_R^2
binance_BTC_spot_vwap_mean	-0.02370085601576005	-2.106125496736665	0.16028187801247473
binance_ETH_spot_vwap_mean	-0.04127944587848592	-2.72483321299014	0.26304158576325054
binance_FTT_spot_vwap_mean	-0.10420049908257908	-1.3980428858698342	0.05035863286393816
binance_SOL_spot_vwap_mean	-0.07067998119839075	-2.0036029137068985	0.14344549936795836
bitstamp_BTC_spot_vwap_mean	-0.02325053911589875	-2.041578375307444	0.14966156167081424
bitstamp_ETH_spot_vwap_mean	-0.03976860615296299	-2.5935014744773714	0.24134660650990292
bitstamp_SOL_spot_vwap_mean	-0.06503953230119694	-1.8801081995033229	0.12343952691461224
ftx_BTC_spot_vwap_mean	-0.019478221975598776	-1.7137721109927657	0.0971567139385684
ftx_ETH_spot_vwap_mean	-0.03843350882505457	-2.4460685054551914	0.2168209843098461
ftx_FTT_spot_vwap_mean	-0.08654246888187753	-1.1658369453261153	0.019563829407756028
ftx_SOL_spot_vwap_mean	-0.06672957754100392	-1.691860708045827	0.09376476881400486
gate-io_BTC_spot_vwap_mean	-0.023374067291493293	-2.0656607799827635	0.15361646936411066
gate-io_ETH_spot_vwap_mean	-0.03990210941561349	-2.591758201759561	0.24105746154144325
gate-io_FTT_spot_vwap_mean	-0.1447641235954703	-1.8810616568322844	0.12359257913427057
gate-io_SOL_spot_vwap_mean	-0.07021406601533986	-1.9869122353482174	0.14072205119544512
okex_BTC_spot_vwap_mean	-0.023746637475620523	-2.1148675464898052	0.16172490845334497
okex_ETH_spot_vwap_mean	-0.041219441006735295	-2.7255941572869955	0.26316670472791204
okex_SOL_spot_vwap_mean	-0.07075021634614276	-2.021195452392824	0.14632189936483198

Figure 3 Return Regression on Crypto Index

These results support the strength and weight framework introduced, as an increase in Google research is linked to a negative performance of the market: people may have a negative bias towards those specific words; and the reaction is stronger than it should be as the signal form the salient information is weighted more than it should. We also find that crypto investors sentiment does not appear to be influenced by macroeconomic and traditional financial jargon as the two other Indexes do not present significant results.

In relation to order flows correlation, we find positive autocorrelation decaying over time lags, consistent with financial literature. However, we do not find enough evidence that this phenomenon increases in high sentiment periods, suggesting that the herding behaviour component is not stressed during market turbulences or the reduced number order splits compensate this effect.

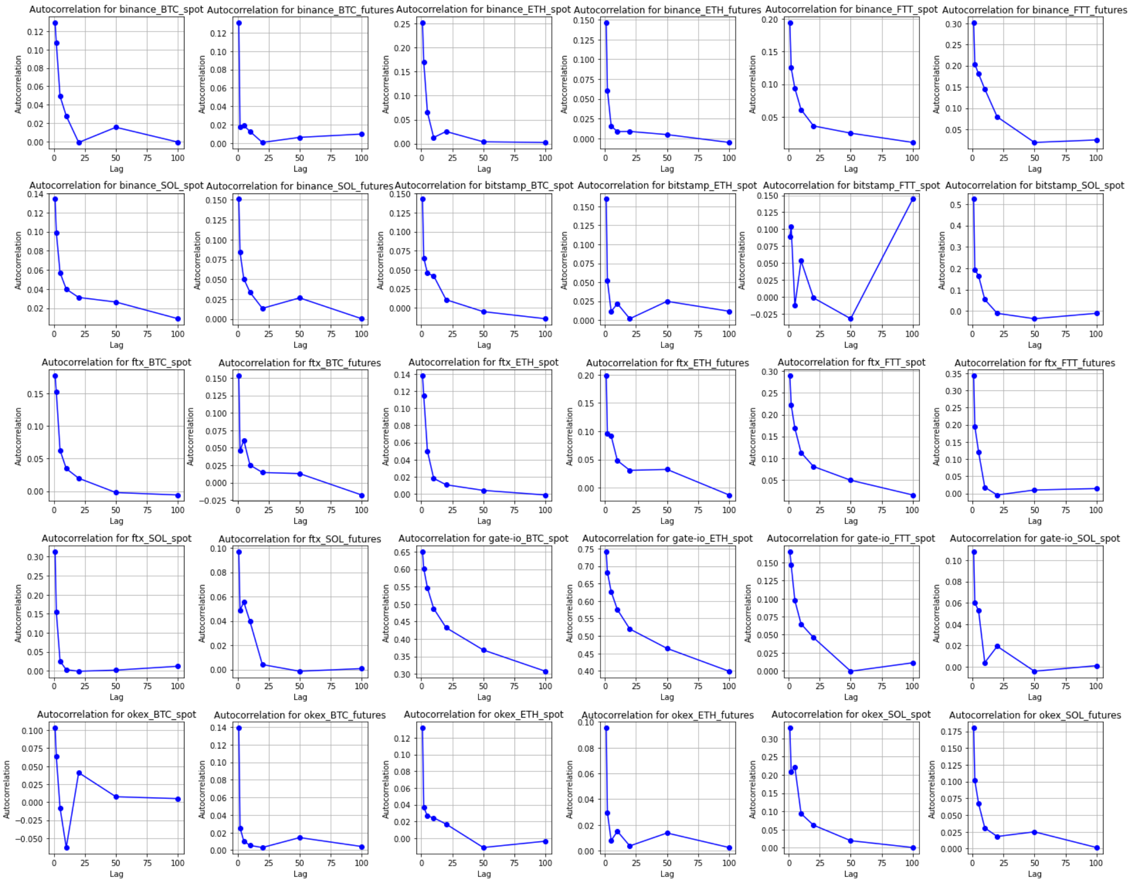

Figure 4 Cryptocurrencies autocorrelation

The arguments proposed in this work suggest that even if the cryptocurrencies market are technically efficient, crypto investors' behavioural biases undermine the sound price formation process of the assets. The incorrect assessment of news seems to originate at a personal level rather than being influenced by market direction, as evidence of stressed herding behaviour is not found. The methodology presented in the present work is scalable, allowing the expansion and refinement of the Indexes as the market records increase.

The results of this work underline requirements of a better financial literacy in the tested markets and that price efficiency in crypto markets is yet to be achieved. Nevertheless, a further growth in scale as well as an increased participation of unbiased investors may help to mitigate the recurrence of similar nefarious events and reactions.