Imperial College London

Imperial College London

Department of Computing

Determinants of Mispricing of Cryptocurrency Futures

Author:
Daniel Pelnar

Supervisor:
Dr Paul A. Bilokon

Second marker:
Prof William J. Knottenbelt

Submitted in partial fulfillment of the requirements for the MSc degree in Computing of Imperial College London

September 2022

Abstract

This paper examines the mispricing determinants of Bitcoin and Ethereum quarterly futures traded on Binance from February 2021 to June 2022 by employing contract–by–contract regression analysis. Additionally, we consider quadratic form specification of the regression models to determine whether futures trading volume has a diminishing effect on the mispricing term (or the basis) and whether a point of maxima can be found. Finally, we amend the general cost–of–carry model so it does not treat Bitcoin or Ethereum as a non–dividend paying stock or a zero–coupon bond. The results suggest that futures volume is the most important explanatory variable for both Bitcoin and Ethereum, and the estimated coefficients of the quadratic form specification are statistically significant for 50% and 66% of Bitcoin and Ethereum contracts respectively. The analysis of the point of maxima suggests that there are 14 days during which Bitcoin underwent a possible market manipulation (for Ethereum, it was 23 days). Our modification of the cost–of–carry model performs worse than the benchmark for Bitcoin contracts and better than the benchmark for Ethereum contracts.

Acknowledgments

This endeavour would not have been possible without my supervisor Paul A. Bilokon whose guidance and experience was invaluable. I would also like to thank Kamil Kurdziel for introducing me to futures contracts traded on cryptocurrency exchanges. Finally, I would like to express my gratitude to William Knottenbelt for providing valuable feedback during the initial stages of the project.

Contents

Chapter 1

Introduction

This chapter is structured as follows. Section 1.1 explains why research on the dynamics of cryptocurrency futures is important and what the aims and objectives are. Section 1.2 states the contributions. Section 1.3 introduces the outline of this paper. Section 1.4 discusses the legal, social, ethical and professional considerations related to this paper. Section 1.5 contains a statement of originality and publications.

1.1 Motivation, aims and objectives

Adding to the understanding of the dynamics of cryptocurrency futures is important to various entities. It can be a retail trader trying to leverage their positions; an arbitrageur trying to make riskless profit, and, in the process, induce market efficiency; a regulator trying to protect investors; or a central banker trying to maintain price stability.

The aim of this paper is to examine the determining factors of mispricing of Bitcoin and Ethereum quarterly futures traded on Binance from February 2021 to June 2022 by performing contract–by–contract regression analysis. Contracts are analysed both individually and in the aggregate. Furthermore, we run regressions with futures trading volume as quadratic term to find out whether a diminishing effect and turning point exist. The hypothesis states that with each additional trade, the impact on the mispricing term (or the basis) is still present, however lower than from the previous unit. Moreover, we hypothesize that the estimated coefficients of the quadratic regression turns out to be positive for the linear term and negative for the quadratic term (concave function with a point of maxima), indicating a possible manipulation taking place on days whose futures volume exceeds the point of maxima. The final objective is to alter the general cost–of–carry model by considering innate cryptocurrency–specific characteristics that reflect the cost associated with carrying (being long) spot.

1.2 Contributions

In general, the regression results from contract–by–contract analysis are consistent with Hattori *et al.* [1] and De Blasis *et al.* [2], providing further evidence that futures trading volume is the most important explanatory variable. Regarding the simple *f* model with the mispricing term as the dependent variable, futures volume is statistically significant for 66% and 83% of Bitcoin and Ethereum contracts respectively. All significant coefficients have the expected sign and *mean* $\bar{R}^2$ is 19.4% and 13.4% for Bitcoin and Ethereum contracts, respectively.

Our quadratic form specification improved *mean* $\bar{R}^2$ (44.3% for Bitcoin and 32.7% for Ethereum). Regarding Bitcoin, 50% of the estimates were jointly statistically significant and for Ethereum, it was 66%. The calculated points of maxima suggest that there are 14 days during which Bitcoin underwent a possible market manipulation. For Ethereum, this number increased to 23 days.

Our modification of the cost–of–carry model had mixed results: for Bitcoin, the altered model performed better for the first 2 contracts, but fell off for the remaining 4 in comparison to the benchmark (simple cost–of–carry model with the risk–free rate as the only cost). The results for Ethereum were more promising as the adjusted model performed better for all contracts, apart from the last 2 months of the last contract (Jun–22).

We also contributed by examining explanatory variables derived from datasets usually not used for these purposes. Namely, *scam fraud breach*, *insurance fund balance* and *marketcap domination*. Unfortunately, dummy variable *scam fraud breach* is not statistically significant for any of the contracts; *insurance fund balance* is not stationary for the majority of the contracts; and *marketcap domination* is estimated with positive sign, however, the expectation was that it would contribute to the decrease of the mispricing term (and the basis) towards zero.

1.3 Outline

The remainder of this paper is organized as follows:

Chapter 2 describes previous research in chronological order and explains how this study fills the research gap. A discussion on the evolution of the risk–free rate for the pricing of derivatives is also included.

Chapter 3 explains finance prerequisites, notably the cost–of–carry model that is used for calculating the theoretical futures price.

Chapter 4 is about data collection, formatting, as well as, suitability and limitations of data sources. The chapter also describes each explanatory and dependent variable by the means of descriptive statistics and graphical illustrations.

Chapter 5 discusses the econometric, statistical and machine learning techniques used for the contract–by–contract regression analysis, whose results are presented in chapter 6, which also talks about the results of hypotheses testing and the modification of the cost–of–carry model.

Chapter 7 concludes with limitations pointed out and further research suggested.

1.4 Legal, ethical and social considerations

Legal, social, ethical and professional considerations have been made. We said "no" to all points from the LSEP Checklist, apart from point number 2 from section 10, which is about legal issues:

"Will your project use or produce goods or information for which there are data protection, or other legal implications?"

The answer to the question above is a possible "yes". To collect the dates when cryptocurrency scam, fraud or breach was detected, see section 4.4, we scraped the introductory landing page of a report on the Crystal Blockchain website. This is potentially illegal (breach of contract claim) under the UK law if Crystal blockchain T&Cs prohibits data scraping or equivalent activities. Reading through Crystal Blockchain T&Cs, we did not find any mention of prohibition of this type of activity. We also contacted Crystal Blockchain via email directly, asking them for permission to scrape the landing page for academic research. No response has been provided as of writing.

1.5 Statement of originality and publications

I declare that this paper was written by myself, and that the work that it presents is my own except where otherwise stated.

Chapter 2

Literature review

Literature review is structured as follows. Section 2.1 chronologically introduces previous studies and explains how our paper builds on their findings and where the contribution lies. Section 2.2 clarifies ambiguous terminology and notation used in the past. Section 2.3 discusses the evolution of the risk–free rate used for pricing of derivatives.

2.1 Historical background

Since the first block was mined on January 9, 2009 at 3:54 AM GMT+1 [3], academics and practitioners have examined Bitcoin from many perspectives. The most prominent are the technical analysis perspective, the macroeconomic perspective, and the efficient market perspective.

Our research builds on—and is related to—the efficient market perspective, which examines whether new information is immediately reflected in prices. To what extent cryptocurrency markets are efficient is debatable since the results of many existing academic papers are mixed. Urquhart [4] was the first to study the market efficiency of Bitcoin. His findings reveal that returns are significantly inefficient, although some tests suggest more efficiency in the second half of the dataset. By contrast, Nadarajah *et al.* [5] found out that the Bitcoin market is efficient, which was later supported by other studies conducted by Brauneis *et al.* [6] and Tiwari *et al.* [7]. Conversely, in late 2018, more studies emerged arguing for inefficiency, for instance Yonghong *et al.* [8] or Cheah *et al.* [9].

Approximately a year or two after the introduction of the first regulated Bitcoin futures contract by CBOE, several academic studies evaluated the informational content of spot and futures prices from the price discovery point of view, for example [10, 11, 12].

Adding to Kapar *et al.* [10], Baur *et al.* [11], and Fassas *et al.* [12] research, Lee *et al.* [13] contributed by examining the deviations from no-arbitrage bounds between Bitcoin spot and futures prices using CME and CBOE futures contracts traded from January 2018 to March 2019. The results suggest that arbitrage opportunities are persistent and expand with the issuance of alternative cryptocurrencies, and with Bitcoin thefts, such as scams, frauds and hacks.

In 2020, Matsui *et al.* [14] contributed to the study area of speculative efficiency of Bitcoin markets. Leveraging non-overlapping data on Bitcoin spot and CME futures prices in a period from December 2017 to April 2020, the study found evidence that the market is inefficient as suggested by an analysis of futures contracts expiring in 1 month, whereas shorter contracts (2 weeks and 1 week until maturity) asserting the opposite. According to Matsui *et al.* [14], a possible explanation for the market inefficiency—emerging from the 1–month futures contracts—is a low liquidity.

In 2021, Hattori *et al.* [1] conducted research on market efficiency in the context of Bitcoin futures markets and arbitrage opportunities, using intraday data from CBOE and daily data from CME. The authors observed fewer arbitrage opportunities during "normal times" (normal times defined as periods without more than 40% drop within 1 trading month), and substantial arbitrage opportunities during high market volatility.

A few months later, De Blasis *et al.* [2] examined arbitrage opportunities of quarterly Bitcoin contracts on Binance. One of the objectives was to ascertain whether the frequency, duration and magnitude of arbitrage opportunities found on regulated futures markets, such as CME or CBOE, also exits on a much larger, less regulated cryptocurrency exchange such as Binance. The arbitrage analysis (cash-and-carry and reverse) was performed on quarterly Bitcoin contracts, specifically on two contracts maturing on June 25 and September 24, 2021. Findings agreed with Hattori *et al.* [1] results suggesting that Bitcoin arbitrage opportunities exists mostly during market dislocations not only on CME and CBOE exchanges, but also on Binance.

The methodological framework that Hattori *et al.* [1] and De Blasis *et al.* [2] utilized was developed during the 1980s, for instance [15]. The framework evaluates whether deviation between the theoretical and market–quoted futures price is less than the cost related to arbitrage, such as the bid-ask spread. Hattori *et al.* [1] found the determinants by estimating population parameters of the following equation (note that De Blasis *et al.* [2] used almost the same approach, though with different notation):

$$Deviation_t = \alpha + \beta \times Bidaskspread_t + \gamma \times Turnover_t + \epsilon_t \tag{2.1}$$

Before the estimates $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{\gamma}$ were obtained by Ordinary Least Squares estimation method, *deviation* needed to be calculated as follows:

$$Deviation_t = |G_t - H_t| \tag{2.2}$$

The market–quoted futures price, H_t is an observable variable, however, the theoretical forward price, G_t needed to be computed:

$$G_t = S_t \times e^{r_f \times T} \tag{2.3}$$

The estimation of the population model 2.1 allowed Hattori *et al.* [1] and De Blasis *et al.* [2] study the determinants of the deviation, or in other words, allow them to examine the sources that are associated with mispricing.

Even before the inception of cryptocurrencies, the computation of theoretical forward price, see equation 2.3, has been a common practice for many years. For instance MacKinlay *et al.* [16], Bhatt *et al.* [17] and Switzer *et al.* [18] utilized the formula in their research on futures with stock indices as the underlying assets. Based on the literature review, we observed that vast majority of studies have treated Bitcoin as a non-dividend paying stock or a zero-coupon bond. To the best of our knowledge, there have only been two cryptocurrency–related studies that attempted to incorporate cost–of–carry components (other than risk–free rate) into formula 2.3.

The first study, conducted by Lian *et al.* [19], recognizes that the cost–of–carry model can be used to value Bitcoin futures. The paper considers electricity expenses and equipment costs required to mine one Bitcoin in their amended futures pricing formula. After carefully examining their work, we were not able to determine the reasoning behind this choice. Intuitively, both of these costs should already be reflected in the spot price of Bitcoin. It is as if we included capital and human cost required to extract oil from underground reservoirs—the price of barrel of oil already reflects these expenses. Oil futures contracts have cost of carry associated with them, however, those are cost of carrying the barrel of oil until maturity, such as storage cost. The paper by Lian *et al.* [19] could have considered, for example, the electricity expense and equipment required to *validate* that a Bitcoin is carried (is in one's possession) by the means of running a full node. However, this type of cost would have been significantly lower, possibly even negligible.

The second study, done by Kim [20], employed the cost–of–carry model with convenience yield term, y, hypothesizing that there is a benefit of carrying (holding) Bitcoin. The findings suggest that there is a premium of 5.4% with holding Bitcoin rather than buying Bitcoin futures contract. Mentioned sources of this benefit include short-selling restrictions and voting rights. The estimates of convenience yield was determined by repeatedly solving the following equation for y for all time periods:

$$CME\ Futures\ Price = Spot\ Price \times e^{(TBill + 0 - y) \times maturity} \tag{2.4}$$

After that the estimates were averaged to arrive at one value. In our opinion, the limitation with this approach is that the determined yield is not strictly speaking merely convenience yield, but an aggregate of all the costs and benefits of holding Bitcoin.

Going back to the line of research conducted by Hattori *et al.* [1] and De Blasis *et al.* [2], the most recent study on this topic was written by Matsui *et al.* [21] in February 2022. Their research focused on the determinants of volatility (dependent variable) of gold, oil and Bitcoin futures prices, however, basis was also employed as the second dependent variable of interest. The period of study was from December 18, 2017 to November 30, 2021. Utilizing contract-by-contract analysis, Matsui *et al.* [21] run OLS regressions of all combinations of dependent variables (volatility, basis), assets (gold, oil, Bitcoin), explanatory variables (time to maturity, trading volume, open interest) and contracts. Regarding results for Bitcoin and basis, the explanatory variables—trading volume and open interest—were statistically significant in only 17.0% and 12.8% of all contracts, respectively (in the *mvo* model). Time to maturity was significant in most contracts (55.3% in the same model), and out of the significant coefficients, 84.6% of them were estimated with positive sign, suggesting that maturity affects the basis of Bitcoin positively.

Our research further builds on Hattori *et al.* [1] and De Blasis *et al.* [2] by considering other possible explanatory variables affecting the mispricing term (or the basis), using contract–by–contract analyses—recently applied by Matsui *et al.* [21]—with Bitcoin and Ethereum futures data from Binance, utilizing all available contracts, which is a period from February 2021 to June 2022. Furthermore, a different functional form specification, specifically the quadratic function, is tested for statistical significance and derived point of maxima cast light on whether a possible manipulation takes place in a given day—as it occurs for the gold futures markets [22]. Similarly to Lian *et al.* [19] and Kim [20], although with different methodological approach, we attempt to amend the general cost–of–carry model so it does not treat Bitcoin or Ethereum as a non-dividend paying stock or a zero-coupon bond.

2.2 Note on notation

Due to the fact that some of the aforementioned frameworks in the previous section 2.1 have been around for more than 40 years, authors have come up with many different, and occasionally contradictory, notations and namings. The following bullet points clarify some of the ambiguous terminology and notations.

- **Theoretical forward price** – For instance, Hattori *et al.* [1] used *theoretical forward price*, De Blasis *et al.* [2] used *theoretical futures price*, Lee *et al.* [13] used *expected futures price* and Cornell *et al.* [15] used *implied forward price.* All four are synonyms in the context they were used. There is a well known result e.g., [23], which conveys the fact that under certain assumptions, the forward price and futures price equals. This is an important result, which all of these studies, including ours, use. Unfortunately, some studies are not explicit about this, immediately presenting *theoretical or expected futures price*, which causes confusion. Without skipping any steps the ideal order should be to first present the computation of the theoretical forward price by setting up riskless portfolio and arguing that the return should be the risk–free rate. Then, under the assumptions found in [23], a conclusion can be drawn that the calculated *theoretical forward price* is the *theoretical futures price*. Finally, the *theoretical futures price* can be used in equation 2.2 as G_t to calculate the deviation.

- **Market–quoted futures price** – This is the price of the futures contract that can be directly observed on the futures exchange. It is determined by the demand and supply for those contracts during trading. Most researchers call *market-quoted futures price* simply *futures price* e.g., [2].

- **Deviation** – Deviation, which is usually defined as the distance between the theoretical and market–quoted futures price, see equation 2.2, is occasionally also called *spread* e.g., [2], *discrepancy* e.g., [1], or *mispricing term* e.g., [13]

The above clarifications were provided so that no confusion arises as of this moment. However, a more detailed explanation, including notations and terminology used in our paper, is provided in chapter 3: finance preliminaries.

2.3 Evolution of risk-free rate

In section 3.2, we will derive the general forward (futures) pricing formula, by setting up riskless portfolio, see Table 3.1, and argue that, under certain assumptions 3.3, the return that shall be earned is the risk–free rate, r. This section of the literature review unveils what proxy for the risk-free rate could be the most appropriate to use. As we shall see, there is quite a lot of room for maneuvering when it comes to the choice due to recent developments—and hence, the importance of this section.

Before the financial crisis that started in 2007–2008, life was quite simple for both practitioners and academics attempting to price derivatives. Academic papers tended to mostly use yields of US Treasury bills, notes and bonds as a proxy for the risk-free rate, for example studies [24, 25], whereas the industry practitioners used to deploy LIBOR rates (with appropriate term structures—usually 3 months) [26]. However, a paradigm shift had began after the crisis and the 2012 LIBOR Scandal.

To understand what happened, we should begin with an explanation of what LIBOR is. Since its inception in 1980s, *London Interbank Offered Rate* (LIBOR) was the prevailing interest rate referenced by most financial contracts. Initially comprising of 15 maturities and 10 currencies, LIBOR was compiled daily by asking 18 global AA-rated banks to provide rates at which they believe to be able to borrow funds from other banks. The highest and lowest four quotes for each borrowing period and currency combination were discarded and the rest was averaged to determine the LIBOR rates for the given day and combination. The most commonly used LIBOR rate is the 3–month US dollar LIBOR rate, which has been a reference rate to financial contracts worth hundreds of trillions of dollars [27].

The following are the main problems with LIBOR rates:

1. Only a handful of banks (albeit AA-rated) deciding what the rate should be. During the crisis, it became evident that these banks had understated their borrowing rates to appear in better financial condition than they were [27].

2. In the nature of how the LIBOR rate is determined, incentives for manipulation and collusion were established. In 2012, an international investigation revealed a widespread manipulation of LIBOR rates by banks dating as far as 2003. Regulators in many countries have fined these banks and many traders have been persecuted with mixed results. A high profile case was Tom Hayes, a former trader for UBS and Citigroup, who was convicted for LIBOR manipulation. The UK court sentenced him for 14 years in prison in August 2015. This whole incident, which profoundly influenced the evolution of risk–free rates, was later dubbed the LIBOR Scandal. [28, 29].

3. Other flaws with LIBOR rates: narrow range of counterparties, no transactions rates, and low volume. First, these rates do not take into account a broader range of wholesale counterparties such as insurance companies, money market funds, or investment funds. Second, LIBOR rates are merely quoted rates rather than rates used in real transactions, leading to no robustness checks by the market. Finally, for some days, there might not be enough interbank borrowing for the 18 banks to make accurate estimates of their quoted rates for all the combinations required and therefore some judgment is unavoidable [26, 27].

It is worth noting that some papers have already expressed concerns about the LIBOR rate even prior 2007, for example [30].

After the LIBOR Scandal, the flaws were too difficult to ignore. In March 2021, FCA (UK financial regulator) announced that no bank would be allowed to issue a new LIBOR contract (EUR, GBP, JPY, CHY and 1–week and 2–month USD) by December 2021, and that by June 2023 the remaining US dollar settings (mainly 3–month USD LIBOR) will be completely discontinued after 43 years of service. [31]. This initiated a search for a new proxy for the risk-free rate.

There was a brief moment in the period from 2012 to 2016 where some practitioners and textbooks suggested using *Overnight Indexed Swap (OIS) rate* with federal funds rate [26]. This rate is much closer to the theoretical risk-free rate than plain LIBOR rate in terms of credit risk, because it is continuously refreshed each day. In effect, a creditworthy bank would have to default in a single day. This risk was considered small enough to be discarded [32, 33]. Furthermore, the OIS rate was determined by market transactions as opposed to quotes, and according to some studies, was deemed to be appropriate not only for collateralized, but also for unsecured derivatives [32]. Despite these advantages, (cryptocurrency) derivatives–based papers written in the past few years have not been using OIS, perhaps due to the fact that new official standards were announced in the 2016–2018 period, which are supposed to replace the hole that LIBOR will leave behind after 2023. The following paragraph shed some light on these standards:

Overnight Risk-Free-Rates (RFRs), introduced in the period between 2016 and 2018, aim at replacing interbank offered rates (not only LIBOR, but also similarly constructed rates of other countries, e.g. *the Euro Overnight Index Average* (EONIA) or *the Tokyo Overnight Average Rate* (TONA)). The following are the main RFRs that had emerged [27]:

- *The Sterling Overnight Index Average (SONIA)* was recommended by a BOE working group in April 2017 as an alternative for GBP LIBOR. The aim was to establish SONIA as the main GBP interest rate benchmark in the sterling debt and derivatives markets by the end of 2021. At its January 2022 meeting, the working group concluded that it had met its objective of SONIA replacing GBP LIBOR [34].
- *The Euro short term rate* (ESTER) was selected by a ECB working group as the alternative to EONIA, which was an early replacement for EUR LIBOR, in September 2018 [35].
- *The Tokyo Overnight Average Rate* (TONA) was recommended by a working group in Japan called the Study Group on Risk-Free Reference Rates in December 2016 as a replacement for JPY LIBOR [36].
- *The Secured Overnight Financing Rate* (SOFR) was selected by the Alternative Reference Rates Committee (ARRC) in the United States as a replacement for USD LIBOR [37]. The New York Federal Reserve started publishing the rate in April 2018 [38].

RFRs are fundamentally different from LIBOR rates. The main distinction lies in the fact that they are based on actual market transactions. In essence, they were engineered to eliminate or limit the flaws that LIBOR rates had [27]. SONIA, EONIA and TONA are not collateralized like LIBOR, whereas SOFR, represents loans backed by Treasury bonds. Due to this reason and coupled with the fact that US has no risk premium and Aaa credit rating [39], SOFR is on a good trajectory to replace USD LIBOR. This is further supported by Fassas [40] which suggests that SOFR contribution to price discovery has been increasing lately, though LIBOR was still dominating price discovery in the US money market rates in 2020.

One disadvantage of all the new RFRs is that they are not officially reported with terms (1–month, 3–month and so on), which is problematic when we want to used them for futures pricing. Fortunately, in April 2021, the CME group announced *Term SOFR Rates* (1–month, 3–month, 6–month and 12–month) and three months later, ARRC officially recommended these term rates [41]. As of this moment, CME Term SOFR Rates are available directly via CME DataMine Query API [42, 43].

As of writing, there are still a lot of recent cryptocurrency related academic papers that use LIBOR rates or US Treasury rates. Sampling the literature from the past three years shows that most studies deploy 3–month US Treasuries, for instance [44, 45, 46]. There are also a few research papers that use 1–month US Treasury bills [47, 48] and even LIBOR [49, 50]. Rompolis [51], which was aware of the LIBOR Scandal when writing his paper, still chose LIBOR instead of US Treasuries, arguing that in his sample period, the correlation between term-structures of LIBOR and Treasury rates was 95%, meaning that the results should be robust regardless of the chosen rates. Furthermore, Rompolis [51] pointed out that US Treasury bills, notes and bonds have preferential tax and regulatory treatment, which causes them to be artificially low and hence not a good proxy for risk-free rates. This disadvantage of US treasuries is echoed in many studies and textbooks, for example [52, 26].

To the best of our knowledge, there are so far no papers about derivates and cryptocurrencies that use SOFR despite being arguably the most appropriate proxy for the risk-free rate on US dollar–denominated derivatives and loans as of this moment. We believe that there might be three reasons for this: 1) Academics might not be aware of RFRs since they were introduced quite recently. 2) The CME Term SOFR Rates have been reported for only about one year. 3) Using LIBOR rates or Treasuries might still be preferable in cases when a long time series of risk–free rates is required.

We would like to conclude this section of the literature review talking about what we believe the most appropriate proxy for the risk–free rate is based on what we have seen so far.

For neither practitioners, nor academics does it make sense to use any of the LIBOR rates in new papers and work unless historic data prior to 2021 are required. Depending on the contract and the currency in which it is denominated, any of the RFRs with an appropriate term structure can be used. For USD denominated contracts, Term SOFR Rate is very likely the closest proxy to the theoretical risk–free rate. Its disadvantage is that it has a short time series of observations, dating only about one year back. Perhaps, due to this reason, traditional Treasury rates can still be used, which also have an advantage of being reported with term structures shorter than 1 month. As suggested by Rompolis [51], their disadvantage of having artificially lower rates due to preferential tax and regulatory treatment tend to be negligible in most cases.

In Chapter 4: Data, we will choose a proxy for the risk-free rate, which is the most appropriate for our paper, considering the arguments provided in this section.

Chapter 3

Finance preliminaries

As we alluded to in section 2.2 of the first chapter, the notations and terminology in studies that deal with futures and forwards can be ambiguous and misleading. In order to make this paper more clear, we derive the theoretical futures price from the beginning in a two step process: First to formulate the theoretical forward pricing formula by setting up riskless portfolio and asserting that the return on this portfolio is the risk–free rate (see section 3.2). Second to argue that under a certain assumption the theoretical forward price can be treated as theoretical futures price (see section 3.4). From section 3.5 and onwards, we simply talk about theoretical futures price without explicitly mentioning these steps.

Note that this chapter is mostly common finance knowledge. Bibliographic references can be found in most finance textbooks focusing on derivatives, e.g., [26].

3.1 Terminology

Before going into the theory of forward pricing and valuation, there are a few definitions and notations that need to be stated.

Now is denoted by $t = 0$ and *expiration date* is denoted by $t = T$. We define *time to maturity* as $\tau = T - t$. Note that if we start from *now* ($t = 0$), the time to maturity is simply $\tau = T - 0$ or $\tau = T$.

The price of the underlying asset at time t, also called spot price at t, is denoted as $S(t)$. It follows that $S(0)$ is the price of the underlying *now* and $S(T)$ is the price of the underlying at the time when the contract matures.

The *forward price*, which is agreed on at time $t = 0$, is denoted as K, and is fixed for the whole duration of the contract. It can also be called *delivery price*. Furthermore, we need a notation for *forward price* that is not fixed but changing, we denote this notion as $F(t)$. Note that at time $t = 0$, $K = F(0)$, but for $t \in [1, T - 1]$, $F(t)$ can change, unlike K which stays the same. This will be discussed further in section 3.2.

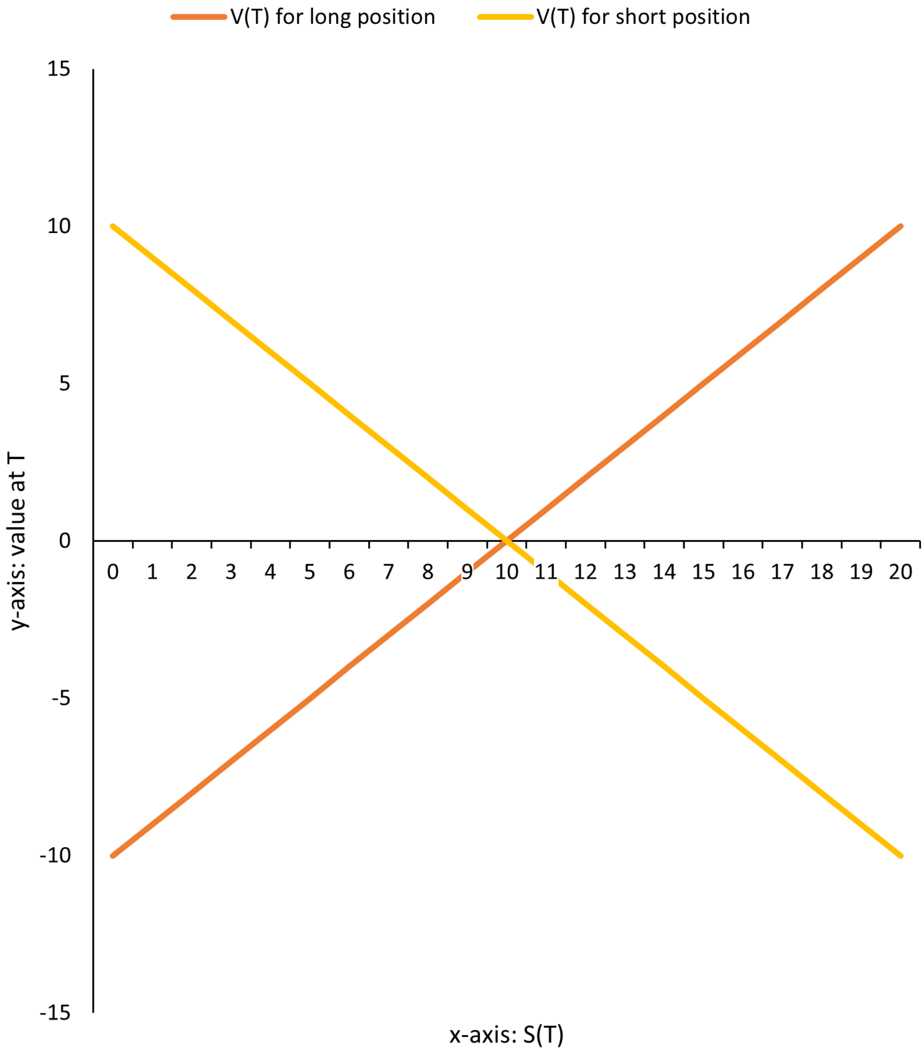

Figure 3.1: Value of forward contract on the delivery day

When trading forwards, we can either *go long* or *go short* a forward contract (or, in other words, *buy* or *short* forward contract). By going long, we enter a contract, which obliges us to buy the underlying asset for an agreed price, K (agreed at time $t = 0$) on the expiration date of the contract (at time $t = T$). Similarly, by going short, we enter a contract, which obliges us to sell the underlying asset for an agreed price, K (agreed at time $t = 0$) on the expiration date of the contract (at time $t = T$).

The value of the contract at time t is denoted as $V(t)$. We think of a forward contract as a contract that can be resold to a third party in this time interval: $t \in (0, T)$. $V(t)$ changes with price changes in the underlying asset over time. By construction, $V(0) = 0$ because no financial resources are exchanged initially. On the expiration date, the payoff is $V(T) = S(T) - K$ for the long position, and $V(T) = K - S(T)$ for the short position. This is intuitively appealing as the long position makes money if the underlying asset's price ends up higher than the agreed price K. The difference

Table 3.1: Value of the portfolio at time $t = 0$ and T; $\tau = T - 0$

	$t = 0$	$t = T$
Forward	0	$S(T) - K$
Stock	$-S(0)$	$-S(T)$
Risk-free Asset	$S(0)$	$e^{r\times\tau} \times S(0)$
Total	0	$e^{r\times\tau} \times S(0) - K$

is the profit for the long position. Similarly, the short position bets on the underlying asset to decrease below the agreed price K. On the expiration date, the underlying asset is sold for K and if the asset's price is below K, the short position's profit is positive. As an illustration of these dynamics, see Figure 3.1 where $K = 10$ at the expiration date. Formula for $V(t)$, $t \in [1, T-1]$ will be presented in Section 3.2.

3.2 Pricing forwards

Let's imagine a theoretical scenario where we are sitting in front of a computer screen on which we can see the price of the underlying asset, $S(0)$, the constant risk-free interest rate, r and the expiration date, T. What is the fair forward price, K of this about-to-be issued contract?

To determine the value of K, we can use a no-arbitrage argument. Consider constructing the following portfolio: Going long one forward contract, going short corresponding amount of the underlying asset, and investing the money from the short to a risk-free asset. The value of this portfolio can be found in Table 3.1.

By no arbitrage argument, we can derive K [26]:

$$\begin{gathered} e^{r\times\tau} \times S(0) - K = 0 \\ K = S(0) \times e^{r\times\tau} \end{gathered} \tag{3.1}$$

Should we observe formula 3.1 not holding (i.e. market-quoted K being higher or lower), then there is an opportunity for free profit. To see this, consider what would occur, should the value K be: $K > S(0) \times e^{r\times\tau}$. In this situation, a market participant could make a risk–free profit by shorting the forward contract and simultaneously buying the underlying asset, which would be financed by r. Now, what if the value K was $K < S(0) \times e^{r\times\tau}$? Then, the market participant could buy forward contract and short the underlying asset. This would again guarantee risk–free profit. Therefore, the fair, arbitrage–free, forward price is the one derived in formula 3.1. Interestingly, we do not even require to be able to *short* the underlying asset. This is because, as long as enough market participants hold the underlying asset for investment purposes, it is rational for them to simply sell the underlying at the market price, when they observe $K < S(0) \times e^{r \times \tau}$, to increase their returns.

Because of the way how we defined *forward price* in section 3.1, formula 3.1 can be rewritten as follows:

$$F(0) = S(0) \times e^{r \times \tau} \tag{3.2}$$

This representation will be useful when we draw the distinction between futures and forwards in section 3.4.

As we said before, the value of the forward contract at time $t = 0$ is 0 (i.e. $V(0) = 0$). This is for both long and short position. We also already know that at time T the forward contract will be worth: $S(T) - K$ (i.e. $V(T) = S(T) - K$) for the long position, and it will be worth: $K - S(T)$ (i.e. $V(T) = K - S(T)$) for the short position. The value of the contract for the **long position** at time $t \in [0, T]$ is the discounted difference between the forward price that would be applicable if we negotiated the forward contract at time t and the delivery price, K, at time $t = 0$:

$$V(t) = \frac{F(t) - K}{e^{r \times \tau}} \tag{3.3}$$

where $\tau = T - t$ and $F(t) = S(t) \times e^{r \times \tau}$. To get some intuition of why formula 3.3 is true, consider the following portfolio:

- Going short a forward contract, which obliges us to sell the underlying for $F(t)$ at the time of maturity.
- Going long a forward contract, which obliges us to buy the underlying for K at the time of maturity.

The payoff from the portfolio at maturity is $F(t) - S(T)$ and $S(T) - K$ from the first and second contract respectively. Therefore, the total payoff, which is known at time t, is $F(t) - S(T) + S(T) - K = F(t) - K$. Formula 3.3 then simply discounts this to obtain the present value of the forward contract.

Likewise, the value of the contract for the **short position** is:

$$V(t) = \frac{K - F(t)}{e^{r \times \tau}} \tag{3.4}$$

The general pricing formula 3.2 is sufficient for pricing forward contracts whose underlying asset has *simple* characteristics. We define *simple* as not having storage costs, not generating any income and not having convenience yield. Underlying assets with *simple* characteristics are, for example, non-dividend paying stocks or zero-coupon bonds. For underlying assets with more complicated characteristics, it is sometimes appropriate to amend the general pricing formula. As mentioned in the literature review, the formula then becomes cost–of–carry formula with not merely the risk–free interest rate as the cost that needs to be carried.

Known Income. Underlying assets such as stocks that pay constant dividends per share or coupon-bearing bonds have known income. To reflect this characteristic, we amend the general pricing formula 3.2 as follows:

$$F(0) = (S(0) - I) \times e^{r \times \tau} \tag{3.5}$$

where I is the *present value* of the income during the life of the contract (e.g. dividends per share) [26].

Known Yield. In this case, the underlying asset provides known yield rather than known income. This means that the income is known when it is expressed as a proportion of the underlying asset's price at the time when the income is paid. An example can be the S&P 500 stock index that has a dividend yield of $x\%$ per annum. The general pricing formula 3.2 can be amended as follows:

$$F(0) = S(0) \times e^{(r-\kappa) \times \tau} \tag{3.6}$$

where κ is the *average* annualised yield of the underlying asset during the life of the contract. (e.g. average annualized dividend yield of the S&P 500 index) [26].

Storage Cost. Underlying assets such as gold or oil incur storage cost for the entity that physically holds them. Intuitively, what this means is that the long position of the forward contract should pay extra for the privilege of not having to deal with costs associated with storing the asset and keeping it secured. Hence, we can treat storage cost as negative known income or negative known yield. The amendments 3.7 and 3.8 of the general pricing formula follows:

$$F(0) = (S(0) + C) \times e^{r \times \tau} \tag{3.7}$$

where C is the *present value* of all the storage costs during the life of the contract (e.g.

$\$x$ per ounce paid annually to store the gold) [26].

$$F(0) = S(0) \times e^{(r+\gamma)\times\tau} \tag{3.8}$$

where γ denotes the annualised storage cost as a proportion of the underlying asset's price (e.g. average annualized storage cost proportional to the price of one ounce of gold. So, for instance, $x\%$ of the price of one ounce of gold.) [26].

When it comes to cryptocurrencies, we could simplify the analysis—as was done in previous studies—by thinking of them as non–dividend–paying stocks or zero–coupon bonds, and hence use the general pricing formula 3.2. An alternative approach would be to amend the pricing formula to reflect each specific cryptocurrency characteristic. For instance, for cryptocurrencies that operate of the proof-of-stake consensus mechanism, formula 3.6 can be used. We could even consider to come up with a similar concept to *storage cost*, which could reflect the possibility of forgetting the wallet's PIN and recovery phrase (losing the cryptocurrency). We continue this discussion in a separate subsection 3.5 dedicated specifically to pricing of cryptocurrency forwards.

3.3 Assumptions

This section talks about assumptions needed for the derivation of the forward pricing formula in the previous section 3.2. Note that we do not need the assumptions to hold for all market participants. What is needed for determining the relationship between the forward and spot price is that the assumptions hold (to a reasonable extent) for *key* market participants such as large institutional dealers. The underlying assumption is that their cumulative power over the market is sufficient to arbitrage away any market inefficiencies and restore *true* forward prices in the long run [26].

Assumption 1: Arbitrage - Market participants are unrestricted in taking advantage of arbitrage opportunities as they happen.

Assumption 2: Constant, risk–free interest rate - short-term interest rates, either today or at all future dates, are known, constant and non-negative: $r(t) = r > 0$ for $\forall t >= 0$.

Assumption 3: Debit and credit risk–free interest rates equal - This assumption states that the "borrowing" interest rate is the same as the "lending" interest rate. This assumption rarely holds in practice for small dealers, but is true for larger ones under normal market conditions.

Assumption 4: Continuous compounding - This assumption is usually not realistic to make since interest payments are not given continuously but rather in certain discrete intervals. Nevertheless, this paper is working with continuous compounding to simplify the calculations and reasoning. To find out more about how continuous compound interest is derived, see appendix 8.1.

Assumption 5: Same tax rates - This assumption states that market participants are subject to the same tax rates on the profit.

Assumption 6: No transaction costs - To derive the general pricing formula 3.2, we assumed 0 transaction and trading fees.

3.4 Futures vs. Forward

Forward contracts are simpler to reason about than futures contracts due to the fact that they do not have daily settlements, but rather single payment at time $t = T$. For this reason, all the derivations done so far in this chapter were for forward contracts. Luckily, it can be proven that if the short term risk–free interest rate is constant, see assumption 2 in section 3.3, the futures and forward contract prices are in theory identical. [23] This is good news because we can therefore assume that the pricing formulas hold for both futures and forward contracts—and, indeed, we do that in the remaining sections of this paper.

Intuition of why constant interest rates are needed is not difficult to get by. Consider a situation where we observe a positive correlation between the interest rates and the price of the underlying asset. If the price of the underlying asset decreases, the long position incurs an immediate loss, and due to the positive correlation, the capital that has to finance the loss is less expensive. If the price of the underlying asset increases, the long position realizes an immediate gain, and due to the positive correlation, this profit can be invested for a higher rate. Therefore, when there is positive correlation between the interest rates and the spot price, the futures prices ought to be higher than forward prices. Similar argument can be made, should we observe a negative correlation. In that case, the futures prices ought to be lower than forward prices.

We do not observe constant interest rates in real life, however, this is often ignored in practise as the effect of changing interest rates on the forward and futures prices is negligible in most situations when the contract term is only a few months [26]. This paper works with quarterly futures contracts and that is the reason why it is reasonable to make assumption 2.

For completeness, Table 3.2 summarizes other differences between futures and forward contracts. The rest of this section will discuss the specific characteristics, such as *margin accounts, margin requirements, initial margin, maintenance margin and daily settlement (marking to market)*, which are not present in forward contracts but are essential concepts when we discuss futures contracts. We finish this section by giving an example of an operation of margin account for a **long position**.

As Table 3.2 conveys, futures contracts are standardized financial products that are

Futures	Forward
Daily settlement	Settlement only once at the end of the contract
No counter-party risk	Counter-party can default on the obligation
Standardised contracts	Tailor-made contracts
Contracts traded on exchanges	Contracts traded over-the-counter
Contracts usually closed out	Delivery usually takes place

Table 3.2: Differences between Futures and Forward contracts [26]

traded on organized futures markets. These markets ensure that neither long position, nor short position needs to be concerned about the counter-party risk, i.e., a risk that the party who took the opposite bet defaults on its promise to honour the contract. The way how the futures market manages this risk is by forcing both long position and short position to set some financial resource aside into their respective margin accounts. If the maintenance margin is low for either participant, extra collateral needs to be added or the position is closed by the broker. Futures are settled by daily margining (done by the clearing house). Therefore, on the expiration date (or when the contract is closed), the profit/loss is the amount of collateral in the margin account less the initial margin and less any extra collateral that might have been added to top up the margin account.

Problem description: Let's illustrate these concepts on a simplified example: we long one quarterly future contract, which requires us to buy 1 BTC in 90 days at FX rate to the $35,000. The initial margin is 50% ($17,500). And the maintenance margin is 50% of the initial margin ($8,750). Unlike in practice, we are going to assume that our broker does not give us any interest on the collateral that is locked up in the margin account. We further assume that we cannot withdraw any temporary profit from the margin account, which is in excess of the initial margin. In reality, we would be entitled to do so and be able to earn interest.

Scenario (Illustrating of marking to market): Let's elaborate on the example from the beginning (from $t = 0$). Both the long position and the short position would start by depositing $17,500 to their margin accounts. Then it would depend on the price movement of BTC from which account to which account the money flows at the end of each day. Let's see how the situation develops from the perspective of the long position:

At the end of day 0, the futures price decreased to $34,000. This means that the daily loss was $1,000, and hence the balance in the margin account was reduced by $1,000 to $16,500. This loss did not trigger margin call since the margin account had not fallen below $8,750. At the end of next day (day 1), the futures price dropped

Table 3.3: Illustration of daily settlements for a long position in one Bitcoin futures contract.

Day	Trade price	Settlement price	Daily gain/loss	Cumul. gain/loss	Margin acc. balance	Margin call
0	35,000	...	...	...	17,500	...
0	...	34,000	-1,000	-1,000	16,500	...
1	...	30,000	-4,000	-5,000	12,500	...
2	...	22,500	-7,500	-12,500	5,000	12,500
3	...	23,000	500	-12,000	18,000	...
4	...	6,000	-17,000	-29,000	1,000	16,500
5	...	50,000	44,000	15,000	61,500	...
6	90,000	...	40,000	55,000	101,500	...

Note: The futures contract is entered into on Day 1 at $35,000 and closed out on Day 7 at $90,000. The initial margin is 50% ($17,500), and the maintenance margin is 50% of the initial margin ($8,750). The units are dollars besides column: day.

again—this time by $\$4,000$ to $\$30,000$. Cumulative loss stood at $-\$5,000$ and margin account balance shrank to $\$12,500$. Day 2 caused so far the largest drop. The daily loss was $-\$7,500$ and margin call was issued by the broker, demanding to top up the margin account by $\$12,500$ to its initial value of $\$17,500$. The $\$12,500$ margin was provided by the end close of trading on day 3. What followed was a relatively calm day 3 where the trader experienced a daily gain of $500. Margin account balance stood at $\$18,000$ $(5000 + 12500 + 500)$. An extreme decline was experienced on day 4 where the daily loss was recorded at $-\$17,000$ and a margin call of $\$16,500$ was issued. This margin was paid by the close of the following day. By the end of day 5, the futures price climbed to $\$50,000$, a daily gain of $\$44,000$ and the cumulative gain had become for the first time positive, reaching $\$15,000$. On day 6, the trader decided to close out her position by instructing the broker to *sell* one Bitcoin contract. The futures price on day 6 was $\$90,000$ and the cumulative gain got to $\$55,500$. Note that the trader had excess margin on days 3 and 5.

3.5 Cryptocurrency futures pricing

We have seen in section 3.2 how pricing can be done for traditional forward (futures), and also how the general pricing formula 3.1 can be amended to better reflect the specific characteristics of the underlying commodity or financial asset. In this section, we attempt to amend the pricing formula to reflect some of the characteristics of cryptocurrencies.

Lost Coins Cost We define *lost coins cost* as losing private keys from the wallet where the cryptocurrency is stored (includes forgetting recovery phrase). There are two possible approaches, which can aid us in incorporating *lost coins cost* in our cryptocurrency futures pricing formula: *"Insurance" approach*, *"Not Moved Since" approach.*

"Insurance" Approach – The first option is to take insurance fees (measured as flat value per one coin of cryptocurrency per month) and discount them. Then the present value of these fees can be used in the futures pricing formula:

$$F(0) = (S(0) + PVI) \times e^{(r \times \tau)} \tag{3.9}$$

where PVI is the *present value of the insurance fees* during the life of the contract.

As an illustration, consider a 3–month futures contract on one BTC with the following parameters:

- $S(0) = \$25,000$ per BTC
- $r = 0.06$
- $\tau = \frac{1}{4}$
- Monthly insurance fee per one coin is \$100.

then the present value of the insurance fees is:

$$\begin{aligned} PVI &= 100 \times e^{(-0.06 \times \frac{1}{12})} + 100 \times e^{(-0.06 \times \frac{2}{12})} + 100 \times e^{(-0.06 \times \frac{3}{12})} \\ &= 297 \end{aligned} \tag{3.10}$$

and the theoretical futures price is:

$$F(0) = (25000 + 297.02) \times e^{(0.06 \times \frac{1}{4})} = 25679 \tag{3.11}$$

Note that the theoretical price would be \$25,378, should we not consider *lost coins cost.*

Before we talk about the second approach, it is important to point out the possible problems, but also advantages with this approach. As of this moment, cryptocurrency insurance is still very niche business and only a few companies are offering this service. We will not be able to get a representative average of insurance fees until larger insurance companies have entered this market. The benefit of this approach is that the insurance companies are the ones who do the calculations and appraisal. We simply assume that by the law of supply and demand for these insurance fees, the market arrives at fees that reflect the unobserved cost of losing coins.

"Not Moved Since" Approach – In this approach, the *lost coins cost* is incurred at any time and is proportional to the dynamically evolving price of the underlying cryptocurrency. This can be model by formula 3.12:

$$F(0) = S(0) \times e^{((r+m)\times\tau)} \tag{3.12}$$

where m denotes the proportion of total circulating supply of the given cryptocurrency which has not been moved for at least 5 years.

Parameter m has different advantages and disadvantages than insurance fees. The main disadvantage is that we are explicitly assuming that if a cryptocurrency has not been used in a transaction for 5 or more years, it must have been lost (meaning the wallet is not accessible). This, however, disregards wallets whose purpose is to merely *hold* the cryptocurrency for a long time (longer than 5 years). To mitigate this issue, we select 5 years or longer rather than a shorted time span. The main advantage is that this type of data can be obtained from the primary source—the cryptocurrency blockchain. Alternative approach is to query this data from data provides that specialize on cryptocurrency on-chain data, such as Messari [53]. Yet another approach how to get this parameter is through surveys. For instance, Krombholz *et al.* [54] conducted a survey of 990 Bitcoin users and found out that 22.5% of them lost their bitcoins. Approximately 11% stated a self-induced error as a reason. The benefit of the survey is that the questions asks directly about the required parameter—*lost coins*—so no proxy is necessary. However, surveys have their own disadvantages and limitations [55].

As an example, consider again 3–month futures contract on one BTC with the parameters as before, and $m = 0.2$ (meaning 20% of total circulating supply of BTC has not moved for at least 5 years. Then the theoretical futures price at time $t = 0$ can be calculated as:

$$F(0) = 25000 \times e^{(0.06+0.2)\times\frac{1}{4})} = 26679 \tag{3.13}$$

In the next chapter, we look at how data were collected and formatted, and limitations and suitability of data sources are discussed.

Chapter 4

Data

This chapter is structured into sections by **data sources**, namely Binance, FRED, Messari, and Crystal Blockchain. In each section, we present general information (data type, time period, suitability, limitations, reliability), collection and formatting, and variable description.

The outcome is 12 datasets—representing each contract (6 Bitcoin and 6 Ethereum contracts)—covering a period February 2021 to June 2022, and containing 15 variables (although most of which not utilized for the contract–by–contract analysis).

4.1 Binance

This section covers data from Binance.

4.1.1 General information

looking to conduct research into cryptocurrency futures, the main decision regarding data is to choose from which cryptocurrency exchange the quotes will come. As observed in literature review, the majority of studies utilized CME (e.g., [10, 11, 12, 13, 14, 1, 21]), which began offering Bitcoin futures in the fourth quarter of 2017 and Ethereum futures in the first quarter of 2021 [56, 57]. They are cash (USD) margined and settled by reference to final settlement price (using Bitcoin/Ethereum Reference Rate); contract unit is 5 BTC and 50 ETH; have central clearing; and trade at specific times [58, 59]. In comparison, Binance Bitcoin and Ethereum futures are either cash (USDT or BUSD) or coin margined and settled (using price index and mark price); contract unit is 1 BTC and 1 ETH; instead of central clearing, they have auto-deleveraging and insurance fund; and trade continuously until expiration [60, 61].

In a comparison to CME, one of the limitations of Binance is shorter time series for Bitcoin futures, which are only available from the first quarter of 2021. Despite this limitation, we decided to base our analyses on Binance data, because that allows us to compare results with previous studies—which used data from CME—and comment on the similarities and differences. For instance, Matsui *et al.* [21] studied the determinants of the Bitcoin basis with CME data. An advantage of Binance is its high trading volume [62, 63] in comparison to other futures exchanges.

There are four main types of cryptocurrency futures available on Binance [64]:

- Quarterly futures settled and margined in a stablecoin (BTCUSDT USDⓈ–M and ETHUSDT USDⓈ–M quarterly futures)
- Quarterly futures settled and margin in cryptocurrency (Coin–M quarterly futures)
- Perpetual futures settled and margined in a stablecoin (BTCUSDT USDⓈ–M and ETHUSDT USDⓈ–M perpetual futures, and others)
- Perpetual futures settled and margin in cryptocurrency (Coin–M perpetual futures)

In order to be able to compare our results with previous findings, we decided to use BTCUSDT USDⓈ–M and ETHUSDT USDⓈ–M quarterly futures. This is the list of all available futures contracts of this type for both Bitcoin and Ethereum (note that the symbol name conveys the expiration date, e.g., BTCUSDT_210326 expired on March 26, 2021):

- BTCUSDT_210326, ETHUSDT_210326
- BTCUSDT_210625, ETHUSDT_210625
- BTCUSDT_210924, ETHUSDT_210924
- BTCUSDT_211231, ETHUSDT_211231
- BTCUSDT_220325, ETHUSDT_220325
- BTCUSDT_220624, ETHUSDT_220624
- BTCUSDT_220930, ETHUSDT_220930

We decided not to use BTCUSDT_220930 and ETHUSDT_220930 (expiring September 30, 2022) due to the fact that, as of writing, not enough observations are available to ensure valid regression results (specifically, HAC standard errors could be influenced due to their reliance on asymptotic approximation). It is highly dependent on the data's population distribution, but the general consensus is that around 100 observations should be sufficient, and going below 50 is not recommended, e.g., [65, 66].

Binance has Bitcoin and Ethereum price index (BTCUSDT and ETHUSDT Price Index) as the underlying assets for the contracts mentioned above. These indices are derives from prices available on the following exchanges: Binance, Okex, Huobi, Bittrex, HitBTC, Bitmax, and FTX. [67].

The final part of this subsection is dedicated to a discussion on the number of observations within each contract. Previous studies often trimmed their datasets by omitting the first 30 observations and last 3 observations (or similar variations). For example, Matsui *et al.* [21] decided to work with the last 60 days, which had high trading volume for all their contracts—the first month tended to be characterized with very low volumes relative to the remaining months [21]. Looking at futures volume data we obtained, see figures 4.3 and 4.4, the market activity does not seem to be lower for Binance quarterly futures contracts in the first 30 days. It is, however, considerably lower on the expiration date for each contract. For this reason, coupled with the already–mentioned statement that more observations is better than fewer, we decided to only omit the last observation (the expiration date) from each contract.

In the next subsection, we look at how data collection and formatting was done for Binance data.

4.1.2 Collection and formatting

Most of the required data were queried via the Binance public API [68]. The only time series of interest that can't be obtained through the API is *insurance fund balance*, which had to be downloaded manually [69].

In terms of design, the program was written in Python and its functionally split into two main functions. Figure 4.1 depicts function *create_links* whose purpose is to create a list of parameterized links based on custom specification of which data, frequency, and time period is required.

Figure 4.2 shows a function *request_format_create_csv* that request data from Binance according to the list of links passed in, formats the obtain data into datasets representing each Bitcoin or Ethereum contract, and exports the datasets as csv files.

The main function simply passes symbols of quarterly futures contracts and symbols of index prices into the former function from Figure 4.1, and calls the function from Figure 4.2.

Pandas library was chosen to format the obtained data due to its high performance, as its critical paths are written in C or Cython programming language [70]. Python standard library *re* (regular expression) [71] was utilized to efficiently find a string pattern in each link, which was then used to name each of the datasets. This can be done with language primitives, however with much more lines of code and nested loops.

In the next subsection, we dive into the description of variables that were obtained from Binance, and present figures that show how futures trading volume and basis are changing over all contracts and time periods.

```python
### Binance
# querying:
# 1) # USD(s)-M delivery BTC and ETH futures close and volume (USDT marginned and settled)
# 2) Binance SPOT close and SPOT VOLUME for BTCUSDT and ETHUSDT
# 3) BTCUSDT and ETHUSDT: PRICE INDEX and MARK PRICE

def main():
    all_contracts_symbols = ["BTCUSDT_210326", "BTCUSDT_210625", "BTCUSDT_210924", "BTCUSDT_211231", "BTCUSDT_220325",
                             "BTCUSDT_220624", "BTCUSDT_220930", "ETHUSDT_210326", "ETHUSDT_210625", "ETHUSDT_210924",
                             "ETHUSDT_211231", "ETHUSDT_220325", "ETHUSDT_220624", "ETHUSDT_220930"]
    spot_and_index_symbols = ["BTCUSDT", "ETHUSDT"]


    # GET /fapi/v1/klines
    request_format_create_csv(create_links(all_contracts_symbols, "/fapi/v1/klines"), "klines")
    # GET /fapi/v1/indexPriceKlines
    request_format_create_csv(create_links(spot_and_index_symbols, "/fapi/v1/indexPriceKlines"), "indexPriceKlines")
    # GET /api/v3/klines
    request_format_create_csv(create_links(spot_and_index_symbols, "/api/v3/klines"), "klines")
    # GET /fapi/v1/markPriceKlines
    request_format_create_csv(create_links(all_contracts_symbols, "/fapi/v1/markPriceKlines"), "markPriceKlines")

def create_links(symbols, your_request):
    """
    Parameters
    ----------
    symbols : list
        A list of symbols such as BTCUSDT_210625 or BTCUSDT.
    your_request : set
        A set of api request strings, a specific format is needed.
        See VALID_REQUESTS, otherwise ValueError
    Returns
    -------
    all_links : list
        A list of all available links, which can be requested from Binance.
    """
    VALID_REQUESTS = {"/fapi/v1/klines", "/api/v3/klines", "/fapi/v1/indexPriceKlines", "/fapi/v1/markPriceKlines"}

    if your_request not in VALID_REQUESTS:
        raise ValueError("results: your_request must be one of {}".format(VALID_REQUESTS))

    if your_request == "/fapi/v1/klines":
        template_link = "https://fapi.binance.com/fapi/v1/klines?symbol={}&interval=1d"
    elif your_request == "/api/v3/klines":
        template_link = "https://api.binance.com/api/v3/klines?symbol={}&limit=1000&interval=1d"
    elif your_request == "/fapi/v1/indexPriceKlines":
        template_link = "https://fapi.binance.com/fapi/v1/indexPriceKlines?pair={}&limit=1000&interval=1d"
    elif your_request == "/fapi/v1/markPriceKlines":
        template_link = "https://fapi.binance.com/fapi/v1/markPriceKlines?symbol={}&limit=1000&interval=1d"

    all_links = []
    for symbol in symbols:
        link = template_link.format(symbol)
        all_links.append(link)
    return all_links

```

Figure 4.1: Python script – Part 1 – Obtaining data from Binance

4.1.3 Variable description

In the following next paragraphs, we define the variables and commend on descriptive statistics tables 4.1 and 4.2. Please note that in the following chapters we often use f and s to denote *futures volume* and *spot volume*, respectively. This was necessary to be able to display regression output in a reasonable format.

Futures close is the market–quoted futures closing price obtained in daily intervals from both Bitcoin and Ethereum contracts. The units of measurement is USDT. This variable is used to calculate the *mispricing term* and the *basis*. Over all contracts, the mean is slightly higher than the median for both Bitcoin and Ethereum, which indicates a positively skewed distributions. In fact, the skewness values are 0.15 and 0.26 for Bitcoin and Ethereum respectively. The range is quite large for both Bitcoin and

```python
def request_format_create_csv(all_links, data_type):
    """
    Parameters
    ----------
    all_links : list
        A list of all available links, which can be requested from Binance.
    data_type : set
        A set of strings, only a few strings are allowed.
        See VALID_DATA_TYPES, otherwise ValueError
    Returns
    -------
    None.

    This function requests data from Binance according to the links,
    then formats the data and creates csv files.
    """
    # Imports:
    import json
    import urllib.request
    import pandas as pd
    import re

    VALID_DATA_TYPES = {"klines", "indexPriceKlines", "markPriceKlines"}

    if data_type not in VALID_DATA_TYPES:
        raise ValueError("results: data_type must be one of {}".format(VALID_DATA_TYPES))

    if data_type == "klines":
        columns_queried = ['time_period', 'open', 'high', 'low', 'close',
                           'volume', 'close_time', 'quote_asset_volume', 'number_of_trades',
                           'taker_buy_base_asset_volume', 'taker_buy_quote_asset_volume', 'ignore' ]

        columns_not_needed = ['open', 'high', 'low', 'close_time', 'quote_asset_volume', 'number_of_trades',
                              'taker_buy_base_asset_volume', 'taker_buy_quote_asset_volume', 'ignore']
    elif data_type == "indexPriceKlines" or "markPriceKlines":
        columns_queried = ['time_period', 'open', 'high', 'low', 'close',
                           'Ignore1', 'Ignore2', 'Ignore3', 'Ignore4',
                           'Ignore5', 'Ignore6', 'Ignore7' ]

        columns_not_needed = ['open', 'high', 'low',
                              'Ignore1', 'Ignore2', 'Ignore3', 'Ignore4',
                              'Ignore5', 'Ignore6', 'Ignore7']

    dict_dataset = dict()

    for link in all_links:
        response = urllib.request.urlopen(link).read()
        data = pd.DataFrame(json.loads(response), columns = columns_queried)
        data['time_period'] = pd.to_datetime(data['time_period'], unit='ms')
        data.drop(columns_not_needed, axis=1, inplace=True)
        start_index = re.search("=", link)
        end_index = re.search("&", link)
        symbol_name = link[start_index.start()+1:end_index.start()]
        dict_dataset[symbol_name] = data
        data.to_csv("{}{}{}.csv".format(symbol_name, "_", data_type))

if __name__ == "__main__":
    main()
```

Figure 4.2: Python script – Part 2 – Obtaining data from Binance

Ethereum as can be observed from the maximum and minimum values—in the sample, the Bitcoin futures closing price reached as high as 70,341, and as low as 18,950. For Ethereum, the maximum value was 4,981 and the minimum value was 994. Ethereum was slightly more volatile (in relative terms) than Bitcoin and both have platykurtic distributions (light–tails), indicating lower likelihood of extreme events than if they were normally distributed.

Futures volume, measured in coins (either Bitcoin or Ethereum), represents the number of completed trades in a specific contract for a given day. Being one of the most important variable when it comes to measuring trading activity of futures contracts (e.g., [2, 21]), it is considered as one of the explanatory variables in our study as well. Its distribution is asymmetric, with heavy tails and skewness to the right. The tails are extremely heavy for Ethereum futures volume, indicating frequent occurrence of rare events. Considerably higher number (around 10 times) of Ethereum contracts

Table 4.1: Descriptive statistics for Bitcoin related variables

	Bitcoin						
	mean	**median**	**max**	**min**	**std. dev.**	**skewness**	**kurtosis**
futures close	45668	44522	70341	18950	10816	0.15	-0.54
futures volume	2198	1932	11581	383	1475	2.46	9.87
price index	44827	44175	67522	18976	10018	0.02	-0.53
spot volume	63132	54031	354347	15805	36998	2.60	11.35
insurance fund	450	460	709	127	169	-0.31	-1.10
tfp (close)	44832	44185	67526	18980	10017	0.02	-0.53
mispricing term	-836	-341	65	-6770	1175	-2.27	5.68
basis	-841	-351	64	-6773	1173	-2.27	5.69

Note: Insurance fund balance is in millions of USDT and BUSD, and its descriptive statistics was calculated from 3.2.2021 to 23.6.2022. *tfp (close)* stands for theoretical futures price (close). Kurtosis is reported as excess Kurtosis.

are traded on Binance than Bitcoin contracts. The mean value is 2,198 for Bitcoin, whereas for Ethereum, it is 24,848. The volatility, measured by standard deviation, is slightly larger for Ethereum.

Price index is the Binance reference price rate for spot price (both Bitcoin and Ethereum). As mentioned is subsection 4.1.1, this variable is derived from BTCUSDT and ETHUSDT prices from multiple exchanges. Price index is used to calculate the *theoretical futures price* and the *basis*. In terms of descriptive statistics, the *price index* mean is lower than *futures close* mean for both Bitcoin and Ethereum. *Price index* is also less volatile than *futures close*, indicating that futures are riskier than spot for both Bitcoin and Ethereum. Furthermore, their tails are lighter than normal distribution's tails.

Spot volume represents the number of complete trades of the underlying asset (Bitcoin or Ethereum) on Binance. This is one of the variables that De Blasis *et al.* [2] utilized to explain the distance (absolute value) between *theoretical futures price* and the *market–quoted futures price*. We will also consider it as an explanatory variable in our model. *Spot volume* is far from normally distributed; it is skewed to the right and is leptokurtic for both Bitcoin and Ethereum.

Insurance fund balance is a Binance specific time series variable, which records the end of the day balance of the insurance fund. At the end of 2019, the balance was around 10 million USDT (and 0 BUSD). In the absence of central clearing, Binance has been increasing its insurance fund balance. On August 17, 2021, BUSD was added to the mix and since then has become the main stablecoin in the fund. As of middle of August, 2022, the fund stands at 760 million dollars in stablecoins (160 million USDT

Table 4.2: Descriptive statistics for Ethereum related variables

	Ethereum						
	mean	**median**	**max**	**min**	**std. dev.**	**skewness**	**kurtosis**
futures close	2879	2834	4981	994	874	0.26	-0.65
futures volume	25	19	221	2	23	3.38	19.11
price index	2832	2806	4808	995	860	0.22	-0.74
spot volume	732	579	4310	154	499	2.65	10.57
insurance fund	451	461	709	128	169	-0.31	-1.10
tfp (close)	2832	2807	4809	996	860	0.22	-0.74
mispricing term	-46.7	-22.7	5.2	-254.9	55.3	-1.41	1.25
basis	-47.1	-23.0	5.2	-255.0	55.2	-1.41	1.25

Note: Insurance fund balance is in millions of USDT and BUSD, and its descriptive statistics was calculated from 3.2.2021 to 23.6.2022. Futures volume and spot volume are in thousands of coins. *tfp (close)* stands for theoretical futures price (close). Kurtosis is reported as excess Kurtosis.

and 600 million BUSD) [69]. *Insurance fund balance* was a candidate for one of the explanatory variables as its amount can, in theory, have an impact on the *basis* or the *mispricing term*. The rational is that as the balance increases, the *basis* or the *mispricing term* should decrease as the counter–party risk diminishes for arbitragers who are exploiting the mispricing. Unfortunately, *insurance fund balance* turned out to be $I(1)$, meaning it has be to differenced once to be stationary. We talk about this more in the next chapter 5. The descriptive statistics is the same for Bitcoin and Ethereum as the fund is shared between them. Calculated in a period 3.2.2021 to 23.6.2022, the average balance is 450 million USDT (BUSD) and the median is 460 million USDT (BUSD). Its distribution is relatively normal, although with lighter tails and a slight skewness to the left.

Theoretical futures price is the arbitrage–free, closing price of the futures contract, defined as $theoretical\ futures\ price = price\ index \times e^{4wTbill \times \frac{time\ to\ maturity}{360}}$, where $4wTbill$ is the 4–week US T-bill, which we formally introduce in the next section 4.2. Due to the proxies for the risk–free rate being minuscule for the majority of the contracts, see Figure 4.6, the *theoretical futures price* is almost equal to the *price index* as can be seen from the descriptive statistics tables. *Theoretical futures price* is used to calculate the *mispricing term.*

Mispricing term is one of the dependent variables, defined as the difference between *theoretical futures price* and *futures close*. We mentioned in the previous paragraph that *theoretical futures price* is almost the same as *price index* due to the low interest rates in the sample period. As a result, *basis*, which comes after this paragraph, is also almost identical to *mispricing term*. For this reason, the determinants of the

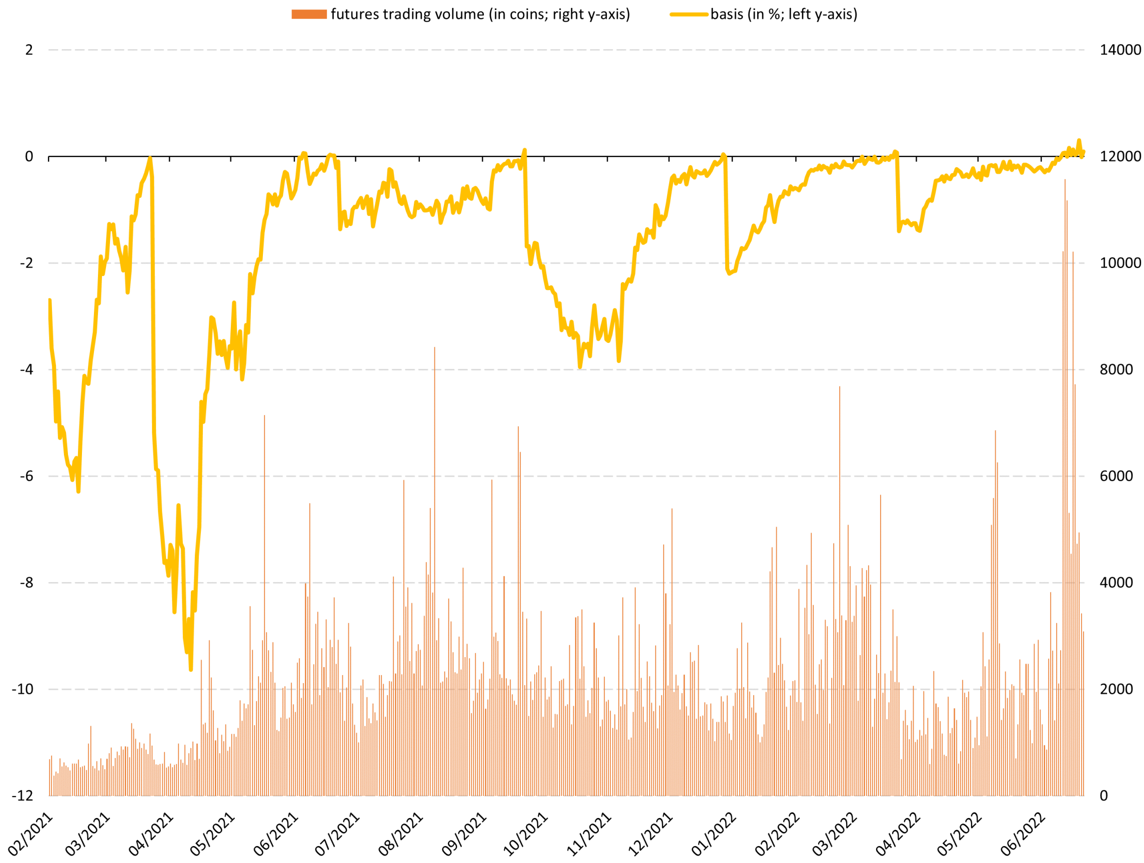

Figure 4.3: Bitcoin – futures volume and basis over all contracts

mispricing term and the determinants of the *basis* should be the same in terms of signs and significance, and only slightly differ in terms of magnitude. The descriptive statistics conveys the fact that the average value was -836 and -47.1 over all contracts for Bitcoin and Ethereum, respectively. The highest mispricing was recorded at value -6,770 for Bitcoin and -254.9 for Ethereum. It has a negatively skewed, leptokurtic distribution.

Basis is the second dependent variable, defined as the difference between *price index* and *futures close*.

We finish this section by examining figures 4.3 and 4.4, which depict futures volume and basis over time, aggregated using all contracts. There are a few interesting observations to note (within each contract period and also across all contracts):

1) Basis – Within each contract, it can be seen that the basis starts low, but approaches zero as each contract gets closer to expiration. Across all contracts, the trend seems to imply that more recent contracts have the starting value of the basis closer to zero; for instance, the most recent BTCUSDT_220624 contract started with only -1.39% basis, while BTCUSDT_220325 began with -2.11% basis. This suggests that both Bitcoin and Ethereum futures markets are maturing and becoming more efficient.

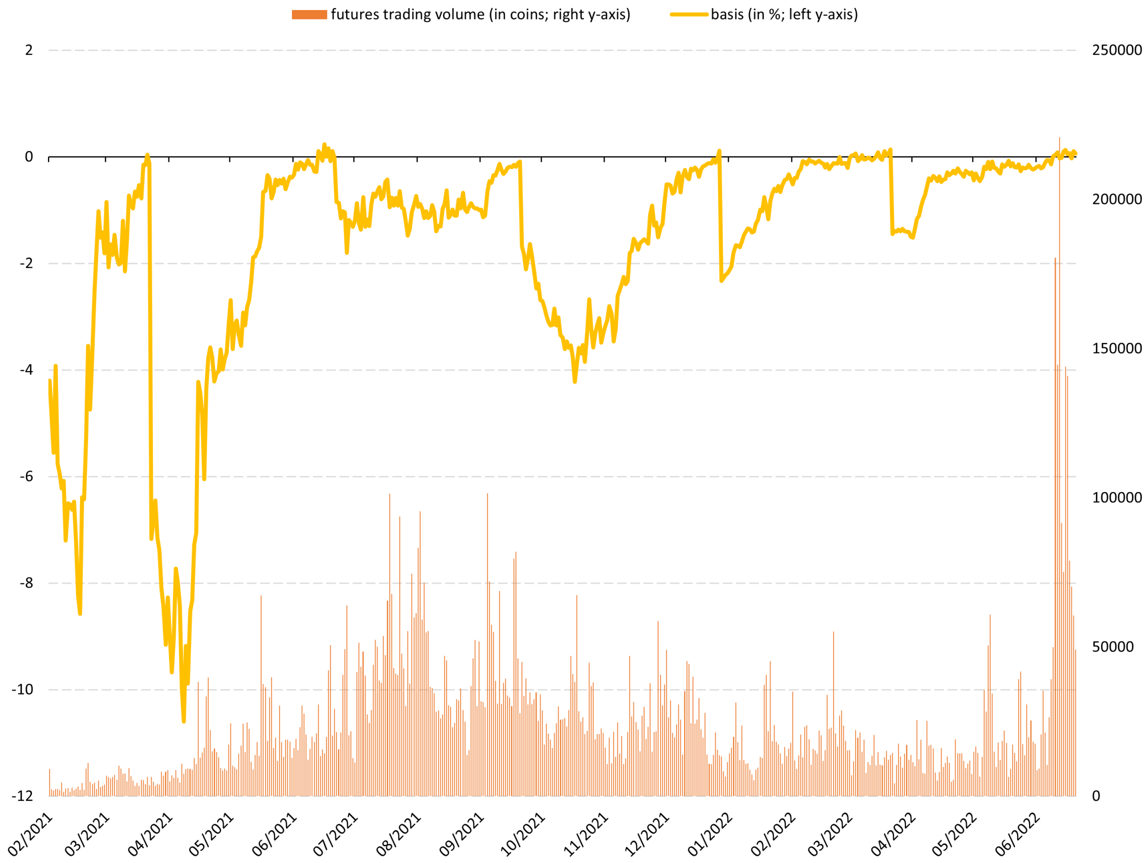

Figure 4.4: Ethereum – futures volume and basis over all contracts

2) Futures volume – Within each contract, it can be observed that volume is relatively stable, although gets higher in a week or two before expiration. Across all contracts, the trend is an increasing volume for both Bitcoin and Ethereum. This observation suggest a positive correlation between the futures volume and the basis, i.e., as the futures volume increases, the basis should shrink towards zero.

4.2 FRED

This section covers data that come from the Federal Reserve Economic Datasets.

4.2.1 General information

As discussed in section 2.3, there are options regarding the choice of risk–free rate. Despite CME Term SOFR rate being arguably a better proxy for the risk–free rate, its time series is not long enough for our analysis. Term SOFR rate is available from the end of April 2021, however, we require interest rate data from February 2021. In order to be consistent (having the same risk–free rates across all contracts), we decided to

Table 4.3: Descriptive statistics for interest rates

	Interest rates – 3.2.2021 to 23.6.2022						
	mean	**median**	**max**	**min**	**std. dev.**	**skewness**	**kurtosis**
4–week US T–bill	0.139	0.050	1.180	0.000	0.239	2.694	6.879
3–month US T–bill	0.253	0.050	1.780	0.010	0.386	1.931	2.993

Note: Units are percentages.

use the 4—week US T–bill and check our result with 3–month US T–bill.

4.2.2 Collection and formatting

Both rates were collected from the FRED website [72, 73]. As both contained missing values (weekends and public holidays), we averaged the preceding and following 5 values (when available) and the mean was substituted for the missing value. The resulting dataset was merged (by time period) with the contract datasets, described in the previous section.

4.2.3 Variable description

Descriptive statistics for both of the rates is available in Table 4.3. Overall, the rates are positively skewed with heavy tails. The 3–month US T–bill rate is closer to normal distribution than the 4–week US T–bill rate, although is more volatile. The average is 0.25% for the former and 0.14% for the latter. Median is equivalent for both, standing at 0.05%.

Figure 4.5 plots the the 4–week US T–bill rate against time with major world events highlighted in blue. It can be observed that since the beginning of the 2008 financial crisis, and until 2016, the rate was kept close to zero. Before the COVID–19 pandemic struck in Europe and the United States in March 2020, the rates managed to peak at approximately 2.5%. In 2022, the FED had to respond to the rising inflation by a series of hikes.

We can see how relevant this is to our analysis by looking at Figure 4.6, which also plots the 4–week US T–bill rate against the time period that covers all the contracts that we collected. As can be seen, the first 5 contracts' *theoretical futures prices* were calculated with a near–zero risk–free rate. As we mentioned, this is the reason behind the almost identical *price index* and *theoretical futures prices* for these contracts. The sixth contract expired on 24.6.2022 when the risk–free rate was 1.18%, however, the average, 0.55% during the whole duration of this contract is still relatively small. It would be interesting to see the possibly different behaviour of BTCUSDT_220930, ETHUSDT_220930 contracts, which are, at the moment, being traded in an interest rates environment above 1%.

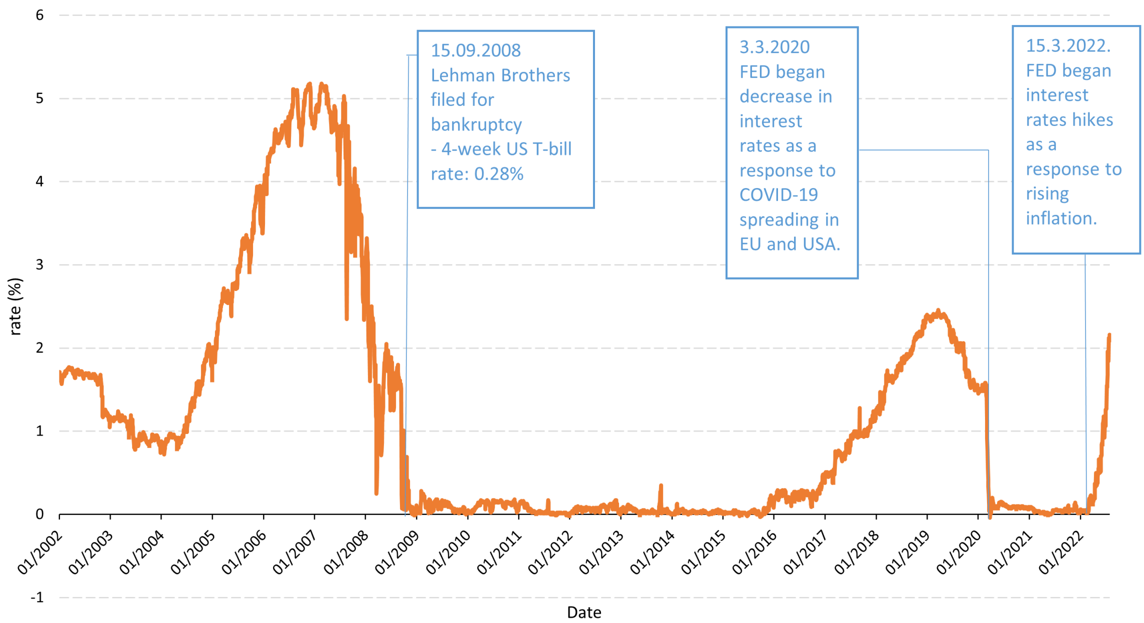

Figure 4.5: 4–week US T–bill rate and major global events

Next section covers data from Messari.

4.3 Messari

In this section, we dive into data, which are sourced from Messari, a cryptocurrency research, data and tools company.

4.3.1 General information

Reading paper from Lee *et al.* [13], which used a dummy variable *NewCoin* (1 if a new cryptocurrency was released and 0 otherwise) to explain the mispricing term, inspired us to look for an explanatory variable with similar effect. A promising candidate turned out to be market capitalization dominance of Bitcoin and Ethereum (relative to the whole cryptocurrency market). There are numerous websites that provide market capitalization of cryptocurrencies. We chose Messari due to its one–week free trial.

4.3.2 Collection and formatting

Data were collected using Messari API [53] and merged (by time period) with the contract datasets. The process was not difficult as the data were continuous without any missing values.

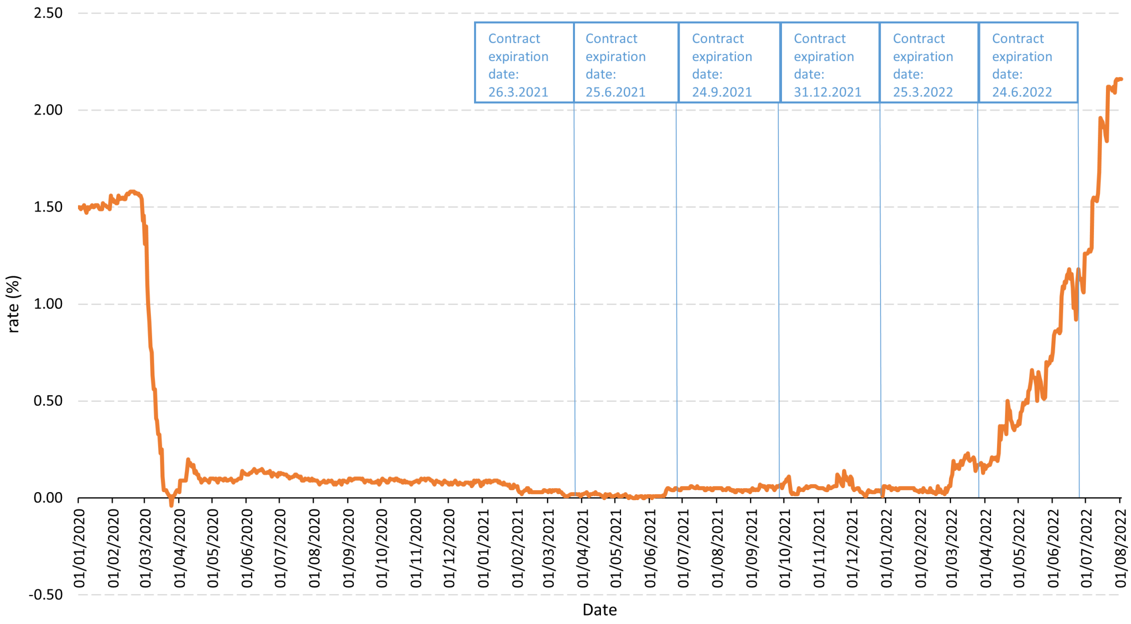

Figure 4.6: 4–week US T–bill rate and expiration dates of 6 contracts

4.3.3 Variable description

The descriptive statistics is provided in Table 4.4. The following are the resulting variables that were collected.

Marketcap dominance represents the Bitcoin's or Ethereum's percentage share of the whole market capitalization of all cryptocurrencies. The unit of measurement is percentage. We chose *marketcap dominance* to be one of the explanatory variables to discover whether or not popularity and dominance influences the *basis* and the *mispricing term*. We expect that as *marketcap dominance* increases, *basis* and the *mispricing term* should increase towards zero. Speculatively, as a cryptocurrency becomes more known, the arbitragers (and academic researchers) allocate more time to it, which could result in better understanding and a development of arbitrage strategies and software. The opposite can also be true: an unknown cryptocurrency might go unnoticed and with less arbitrage trading, the *basis* and the *mispricing term* can potentially be larger (be more negative). The descriptive statistics shows that the average value is above 50% for Bitcoin and about 14% for Ethereum. The highest value was 72.2% for Bitcoin and 25.6% for Ethereum. Bitcoin is about twice as volatile as Ethereum in terms of *marketcap dominance*. The distributions are light–tailed and slightly skewed to the right.

Outstanding supply is the sum of all coins ever created on the particular blockchain. We use this variable to calculate *lost coins proxy*.

Five year active supply represents the sum of all coins that transacted at least once in the past five years. In our analysis, this variable is used to calculate *lost coins proxy*.

Table 4.4: Descriptive statistics for Bitcoin and Ethereum – selected variables

	mean	**median**	**max**	**min**	**std. dev.**	**skewness**	**kurtosis**
Bitcoin							
marketcap dominance	53.9	55.2	72.2	35.0	10.2	0.01	-1.35
outstanding supply	18.0	18.2	19.1	16.5	0.8	-0.39	-1.15
five year active supply	14.2	14.3	14.6	13.6	0.3	-0.21	-1.13
lost coins proxy	3.8	3.9	4.6	2.9	0.5	-0.57	-0.97
Ethereum							
marketcap dominance	13.9	12.7	25.6	7.0	4.3	0.19	-1.37
outstanding supply	108.8	109.5	119.2	93.8	7.4	-0.36	-1.07
five year active supply	103.1	105.6	112.9	83.8	8.9	-0.60	-0.96
lost coins proxy	5.7	5.2	10.0	3.6	1.8	0.63	-0.74

Note: *Marketcap dominance* is in percentages. *Outstanding supply*, *five year active supply*, and *lost coins proxy* are in millions of coins. Time period 05/08/2017 to 01/08/2022.

Lost coins proxy is a variable that we defined for the purposes of improving the cost–of–carry model for Bitcoin and Ethereum. The idea is to obtain a proxy for coins that can not be accessed anymore, or in other words, coins that are lost. We define this proxy as the difference between *outstanding supply* and *five year active supply.* In effect, those are the coins that were not part of any transaction for more than five years. The limitation of this proxy is that those coins also include coins that people use for long–term investment. The value fluctuates, but on average 3.8 million (21%) Bitcoins are lost and 5.7 million (5%) Ethereums are lost.

Figure 4.7 depicts the market capitalization dominance of both Bitcoin and Ethereum over time. The ratio between them was historically very volatile until approximately the middle of 2021, after which it has remained relatively constant. From the figure, we can also see that before the COVID–19 pandemic begin, Bitcoin's market capitalization dominance was consistently in the 60% to 70% range. Now it is fluctuating between 35% and 50%.

Figure 4.8 shows the lost coins proxy over time in percentages. For Bitcoin, the trend is an upward sloping, which means that more coins are being lost than found. For Ethereum, the trend is U–shaped.

4.4 Crystal Blockchain

This section is about data, which are obtained from a report conducted by Crystal Blockchain.

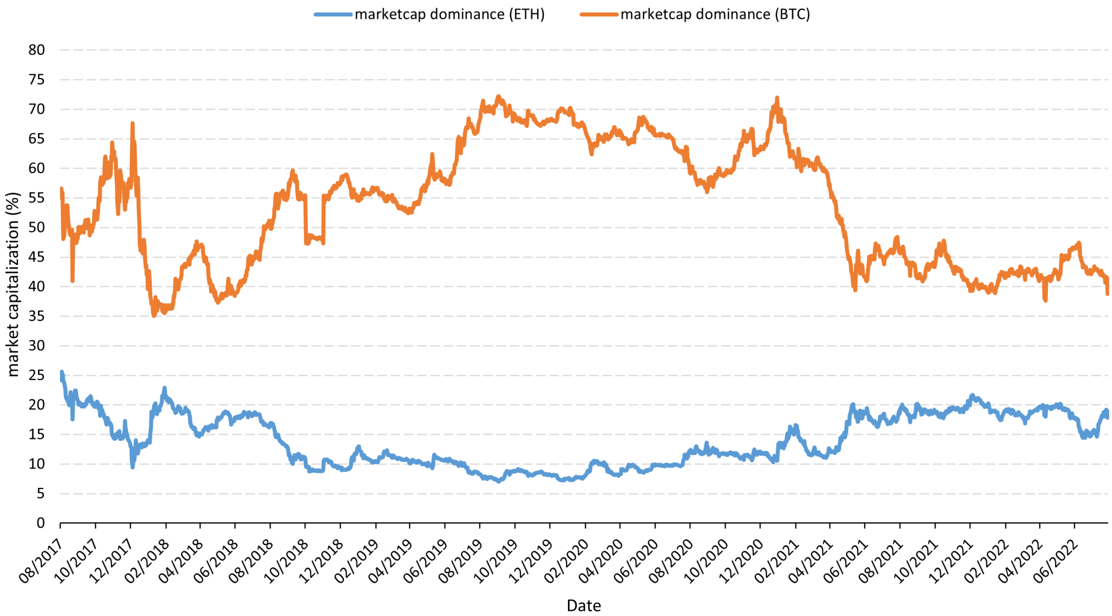

Figure 4.7: Market capitalization dominance of Bitcoin and Ethereum

4.4.1 General information

A Report called "Crypto & DeFi Security Breaches, Fraud & Scams" [74], which was conducted by Crystal Blockchain, studies the largest incidence of cryptocurrency funds over the last eleven years. We were intrigued by the possibility of these incidents having an impact on our dependent variables: *basis* and *mispricing term*.

The outcome was an collection of 364 dates where an incident occurred, and a creation of a time series dummy variable.

4.4.2 Collection and formatting

The dates were collected by scrapping the report summary on the Crystal Blockchain website [74]. For this purpose, Python libraries Beautiful Soup [75] and Selenium [76] were used. Selenium allows for an automatic control of web browsers and Beautiful Soup pulls data out of HTML.

Each date can be found as an innerHTML text wrapped by a span tag whose attribute's class value is "list-item_date". The span tags are nested in a list item element whose attribute's class value is "list-item list-item_country" (those are the rows depicted in figure 4.9). All the rows are nested inside a div whose attribute's class value is "simplebar_content".

Python program: 4.10 (part 1), 4.11 (part 2), iteratively goes inside the nested tags until it find a span whose attribute's class value is "list-item_date". The inner HTML text of this tag is pulled out.

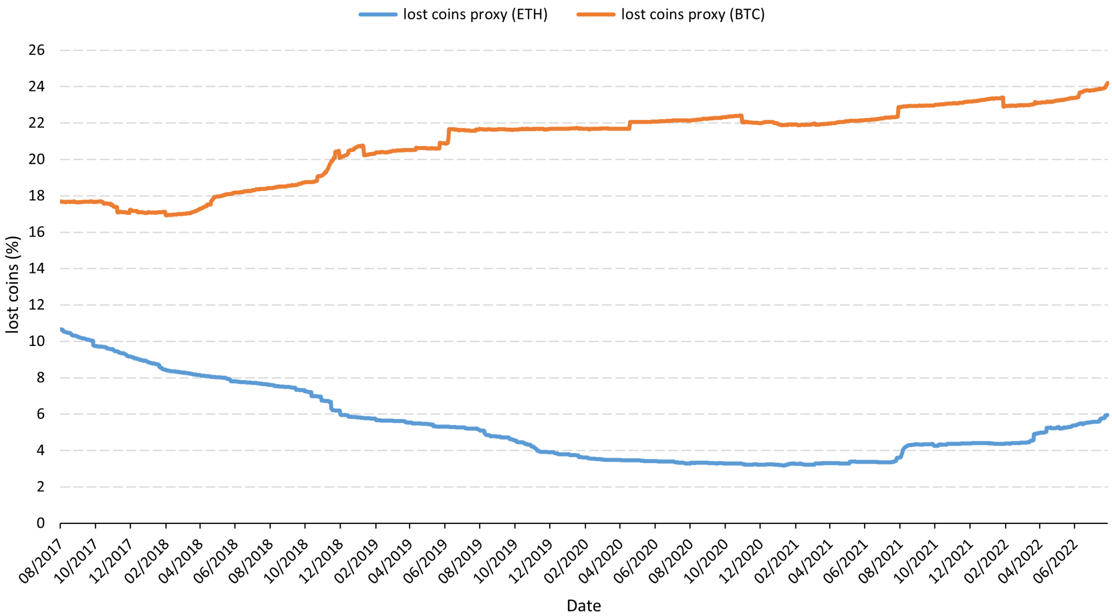

Figure 4.8: Bitcoin and Ethereum – Lost coins proxy

The program then removes all duplicates, creates a dummy variable with values 1 if there was an incident on that day and 0 otherwise, and exports the dataset by creating a csv file.

4.4.3 Variable description

Scam fraud breach is a dummy variable which equals 1 if a major scam, fraud or security breach occurred, and 0 otherwise. We hypothesize that if there is an incident on a given day, the *basis* and the *mispricing term* should increase (be more negative) as it might be more difficult or more risky for arbitragers to obtain the underlying cryptocurrency (the arbitrage requires to buy the underlying and sell the futures contract when *basis* or *mispricing term* are negative). In terms of descriptive statistics, the average is 0.32, which means that 32% of the observations are 1 and 68% are 0.

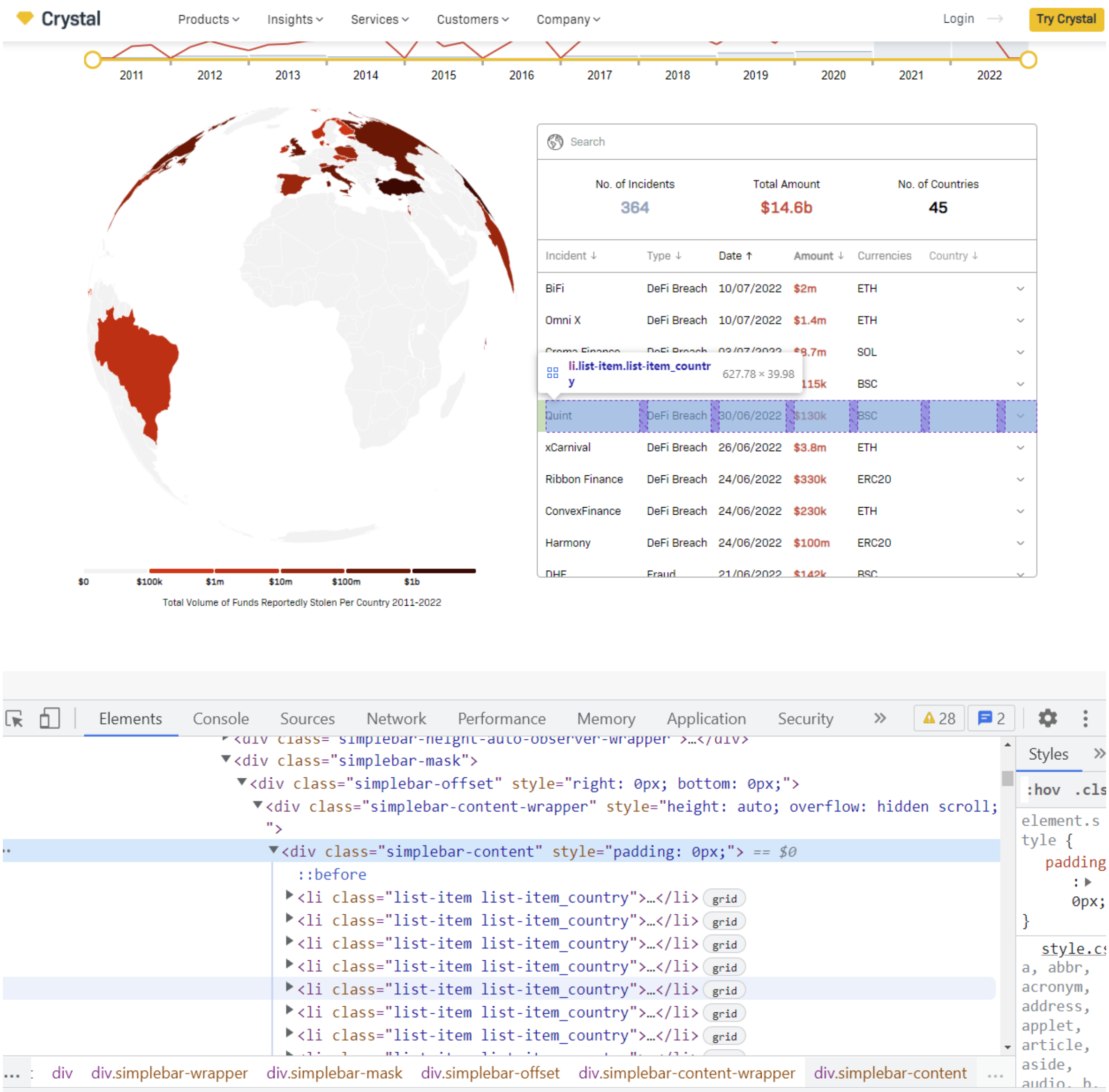

Figure 4.9: "Crypto & DeFi Security Breaches, Fraud & Scams" report [74]

```python
### Source: Crystal Blockchain --
# https://crystalblockchain.com/security-breaches-and-fraud-involving-crypto/
# scrapping dates of security breaches, fraud and scams

'''
Readme:
chromedriver.exe needs needs to be in the working directory
A appropriate version of chromedriver.exe needs to be downloaded on this website:
https://sites.google.com/chromium.org/driver/
'''

def main():
    import os

    cwd = os.getcwd()
    report_url = "https://crystalblockchain.com/security-breaches-and-fraud-involving-crypto/"

    scaped_dates = scrape_dates(cwd, report_url)
    format_and_export(scaped_dates)


def webpage_exists(url_path):
    """
    Checks if the website status is 200

    Parameters
    ----------
    url_path : string
        Url path of the website to be checked.

    Returns
    -------
    Bool
        Returns true or false
    """
    import requests
    response = requests.head(url_path)
    return response.status_code == requests.codes.ok

```

Figure 4.10: Python script – Part 1 – Obtaining data from Crystal Blockchain

```python
def scrape_dates(cwd, report_url):
    """
    Pulls out unique dates from the website
    and formats them as datetime objects

    Parameters
    ----------
    cwd : string
        Path of the current working directory
        where chromedriver.exe is located
    report_url : string
        Url path of the website

    Returns
    -------
    unique_dates : set
        Returns a set of unique datetime dates
    """
    from selenium import webdriver
    from selenium.webdriver.chrome.service import Service
    from bs4 import BeautifulSoup
    import datetime
    import time
    import sys

    if not webpage_exists(report_url):
        raise Exception("The website cannot be accessed.")

    try:
        option = webdriver.ChromeOptions()
        driver = webdriver.Chrome(service=Service(cwd+"{}".format("\chromedriver")), options=option)
    except BaseException:
        # exception selenium.common.exceptions.WebDriverException
        sys.exit("Exiting...")

    driver.get(report_url)
    time.sleep(3)  # let it load the website; wait 3 seconds
    website_content = driver.page_source
    website_content = BeautifulSoup(website_content, features="lxml")

    unique_dates = set()
    div_content = website_content.find("div", {"class":"simplebar-content"})
    print("\nPrinting dates when an incident occured:")
    for li_row in div_content.find_all("li"):
        date_tag = li_row.find("span", {"class":"list-item__date"})
        date = date_tag.get_text() # getting innerHTML text wrapped in the span tag
        print(date)
        unique_dates.add(datetime.datetime.strptime(date, "%d/%m/%Y"))

    driver.quit()
    return unique_dates

def format_and_export(scaped_dates):
    """
    Creates a dummy variable which equals to 1
    if the incident occured on a given day
    and 0 otherwise. Exports as a csv file.

    Parameters
    ----------
    scaped_dates : set
        A set of unique datetime dates

    Returns
    -------
    None.
    """
    import pandas as pd
    import numpy as np

    all_contracts_periods = pd.date_range(start="2021-02-03", end="2022-06-24", freq='D')
    dataset = pd.DataFrame({"time_period":all_contracts_periods}, index=all_contracts_periods)
    dataset["incidents_dummy"] = np.where(dataset.index.get_level_values(0).isin(scaped_dates), 1, 0)
    dataset.to_csv("incidents.csv", index=False)

if __name__ == "__main__":
    main()

```

Figure 4.11: Python script – Part 2 – Obtaining data from Crystal Blockchain

Chapter 5

Methodology

This chapter is structured as follows. Section 5.1 draws the distinction between cross–sectional and time–series regression analysis, and how standard errors need to be adjusted so that p–values are valid. Section 5.2 introduces population models and the OLS estimator. Section 5.3 explains how contract–by–contract analyses is utilized in this paper. Section 5.4 states the quadratic relationship hypothesis and how it is empirically tested.

5.1 Cross–sectional vs time–series regressions

Unlike for cross–sectional data where the independent and identically distributed (iid) assumption is often reasonable to make, time-series variables are, by definition, dependent (experience autocorrelation or serial correlation) and are often not identically distributed [66].

Without iid, the Law of Large Numbers (LLN) and the Central Limit Theorem (CLM) can not be apply for the derivation of robust standard errors, which are used for t–statistics, p–values, and ultimately to determine whether estimated coefficients are statistically significant [66].

Fortunately, it can be shown that stationarity and ergodicity can replace the iid assumption, allowing us to have asymptotic approximations (consistency and asymptotic normality) with time-series variables as well [66, 77]. Let's see what stationarity and ergodicity is in the next few paragraphs.

Informally, a variable is stationary (covariance weakly stationary) if the mean, variance and covariance do not depend on time. Intuitively, stationarity replaces the "independence" part in iid because it forces the variable to be "stable" over time. A formal definition is shown in formulas 5.1 [77].

$$\begin{aligned} E[x_t] &= \mu \\ Var[x_t] &= \sigma^2 \\ Cov[x_t, x_{t+h}] &= f(h) \\ &\neq h(t) \end{aligned} \tag{5.1}$$

The definition conveys the fact that the expectation and the variance of the process x_t has to be a constant, which is not a function of time, and the covariance is also not changing across time. If all three holds, then x_t is coming from some data generating process (DGP) for all periods [66]. Hence, x_t does not come from one DGP in one period and from another in another period. We are assuming that there is some underlying DGP, which is generating x_t for all periods.

We can test for the presence of unit root (this would mean non-stationary) by either Augmented Dickey–Fuller (ADF) test or Phillips–Perron (PP) test. The alternative hypothesis for both of these tests is that the data follow a stationary process [77, 78].

Ergodicity means that the covariance between x_t and x_{t-j} approaches 0 quickly enough as j gets large [77].

In summary, replacing iid assumption with stationarity and ergodicity allows us to estimate population model's coefficients consisting of time-series variables using OLS estimation method with the same methodological approach as if the data were cross–sectional.

To be able to determine statistical significance of the estimated model's coefficients—unlike for cross–sections—(robust) standard errors are not sufficient for time series data. They need to be adjusted to reflect the inherent autocorrelation. A popular option is Newey–West heteroscedasticity and autocorrelation consistent (HAC) standard errors [79] with number of lags equal to $T^{\frac{1}{4}}$ as suggested by Greene [80]. Together with ergodicity, this ensures that p–values are valid.

The bottom line is that we need to test all variables to determine if they are stationary. After that, we run the regressions, calculate HAC standard errors and corresponding p–values.

Results of tests are usually reported in the results chapter, however, we will briefly talk about it here as we need to know which variables we can use as explanatory variables when we define our models in the next section 5.2. Most variables that we considered for the models are stationary across the majority of contracts with an exception of *insurance fund balance*, for which we could not reject the null hypothesis of either ADF or PP test at 5% or lower for 5 out of the 6 contracts.

In the next section, we define the general model and introduce the estimation method.

5.2 Population models and OLS estimation

We start with the following general population model expressed in matrix form [66], see 5.2.

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu} \tag{5.2}$$

In model 5.2, the common terminology holds. $\boldsymbol{Y}_{t\times 1}$ is a $t \times 1$ vector, comprising observations of the dependent variable. $\boldsymbol{X}_{t\times(p+1)}$ is a $t \times (p+1)$ matrix, containing observations of p explanatory variables and a vector that contains ones. $\boldsymbol{\beta}_{(p+1)\times 1}$ is a $(p+1) \times 1$ vector, comprising p population slopes and the intercept. $\boldsymbol{\mu}_{t\times 1}$ is a $t \times 1$ vector of error terms.

Population model 5.2 can be estimated by Ordinary Least Squares (OLS). Expression 5.3 is the OLS estimator in matrix form [66].

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y} \tag{5.3}$$

Vector $\hat{\boldsymbol{\beta}}$ contains p estimated slopes and 1 estimated intercept. Note that there are many options of estimation methods, however we have chosen OLS as it is the most common one for a type of analysis where the number of observations is relatively small.

Finally, the estimated equation of population model 5.2 is depicted in formula 5.4, where $\hat{\boldsymbol{Y}}$ are the fitted values [66].

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{5.4}$$

We now apply the aforementioned generalizations to our specific dataset and research question. The population model of interest is shown below, see 5.5,

$$mispricing = \beta_0 + \beta_1 f + \beta_2 d + \beta_3 s + \beta_4 b + \mu \tag{5.5}$$

where *mispricing* is the dependent variable defined as *mispricing term* in previous chapter 4. Note that *basis* will also be used as a dependent variable. β_0, β_1, β_2, β_3 and β_4 are the population slopes and the intercept that we want to estimate. Explanatory variables f, d, s and b stand for *futures volume*, *marketcap dominance*, *spot volume* and *scam fraud breach* respectively; those were also defined in chapter 4. For scaling purposes, f was divided by a factor of 100 for Bitcoin contracts and by a factor of 10,000 for Ethereum contract. s was divided by a factor of 10,000 for Bitcoin contracts and by a factor of 1,000,000 for Ethereum contracts. d is measured in percentages and b is a dummy variable. Finally, the error term is denoted as μ.

Using the OLS estimator 5.3, we estimate model 5.5. The resulted formula 5.6 is depicted below

$$\widehat{mispricing} = \hat{\beta}_0 + \hat{\beta}_1 f + \hat{\beta}_2 d + \hat{\beta}_3 s + \hat{\beta}_4 b \tag{5.6}$$

where $\widehat{mispricing}$ are the fitted values of the *mispricing term*; and $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$ are the estimated coefficients.

The interpretation of model 5.6 is as follows: if there is a one unit change of a particular explanatory variable, for instance, s, the result is a $\hat{\beta}_3$ unit change in the dependent variable. Formally, we take the first partial derivative with respect to the variable of interest.

So far, we have only considered one contract regressions, however, we have 6 contracts for both Bitcoin and Ethereum. Next section sheds some light on contract–by–contract analysis, a popular option to analyse regression results from many contracts and choose a model that has the best explanatory power based on adjusted R^2.

5.3 Contract-by-contract analysis

The inspiration to perform contract–by–contract analysis comes from Matsui *et al.* [21] who used this type of analyses to study the determinants of volatility and basis for Bitcoin, gold and oil futures contracts using CME data.

To do this analysis, we need to run $(2^\alpha - 1) \times \beta \times \gamma \times \delta$ regressions, where α is the number of explanatory variables; β is the number of contracts; γ is the number of assets (cryptocurrencies); and δ is the number of dependent variables. The $2^\alpha - 1$ term represents all the possible combinations of explanatory variables and we subtract one, because we are not concerned with models that only have an intercept. For our specific case, the number of regressions to run is $(2^4 - 1) \times 6 \times 2 \times 2 = 360$.

Once all models are estimated, we divide them into groups: by the asset class (Bitcoin and Ethereum) and then by the dependent variables (*mispricing term* and *basis*), which leaves us with four groups, each comprising of 90 estimated models. Each of those groups are further divided into 15 groups by the particular combination of the explanatory variables. For example 6 estimated models which only have f as explanatory variable or 6 estimated models which only have f and d as explanatory variables.

At this point, we can start the comparison. First, we compute the adjusted R^2, which measures how much of the variation in the dependent variable is explained by the explanatory variables. Adjusted R^2 was chosen instead of the simple R^2 due to the fact that the former is penalized for adding more variables, whereas the latter is not [66]. Therefore, the best model can, in theory, be a model of only one variable if the other candidates have no explanatory power. Second, in each final group, we report the number of significant estimates of each coefficient and the average adjusted R^2.

As we have seen above, the power of contract–by–contract analysis lies in aggregating all the contracts and taking the average of the averages. The results are reported in chapter 6.

In the last section of this chapter, we state hypotheses of futures volume possibly having quadratic relationship with the *mispricing term* or the *basis*.

5.4 Quadratic relationship

It's a nice sunny day in London, almost 41° Celsius and we (the author) have a freezer stacked with chocolate cornettos. Despite being allergic to chocolate, we swallow the first one and our satisfaction goes through the roof. As the freezer is being emptied, we notice that although the happiness of gobbling up each new cornetto is still there, it's somehow diminishing, becoming less grandiose. Fifth, sixth, seventh, eighth cornetto down the throat and the allergy has kicked in. No more satisfaction from eating cornettos, only suffering.

This experience of extreme weather conditions during summer 2022, inspired us to hypothesize about a possibility of *futures volume* having a diminishing effect on the *mispricing term* (or the *basis*), and whether the point of maxima can be reached for some of our observations.

Here are the following reasons of why we believe the quadratic relationship is present. For the first few trades, it is reasonable to assume that those have a relatively large effect on the convergence of the *mispricing term* (or the *basis*) towards zero as the initial wave of arbitragers try to capitalize on the large discrepancy. However, with more trades, the discrepancy might get smaller (less attractive for arbitragers), and as a result, each additional trade contributes less to the convergence. Furthermore, we expect that if there are observations that are beyond the point of maxima, it is possible that those are days when a manipulation took place by the means of high number of (unprofitable) trades attempting to influence the futures price.

There is some evidence that major banks (JPMorgan Chase, HSBC, and Scotiabank) were manipulating the price of gold via futures markets on behave of the Federal Reserve, either by short–selling of uncovered Comex gold futures or via the Globex system [22, 81]. Roberts *et al.* [22] investigates and describes the process of gold price manipulation by the Federal Reserve. Historically, the purpose was to drive the price down to keep USD strong (in terms of gold) and to satisfy Asian's demand for physical delivery of gold during market turbulence.

A similar situation is possible to occur for the cryptocurrency futures markets (particularly Bitcoin) should investors start using Bitcoin as a safe haven during turbulent times. Although our methodological approach to investigate whether some manipulation has already taken place differs from Roberts *et al.* [22] or Abdullah *et al.* [81], the analysis of the point of maxima can be fruitful. As we said before, what we look for is relatively high spike of futures volume in a given day, which is accompanied by an expansion of the *mispricing* (or the *basis*). This condition occurs after the point of maxima (the turning point) is reached.

$$mispricing = \beta_0 + \beta_1 f + \beta_2 f^2 + \mu \tag{5.7}$$

Regression 5.7 will help us with testing the following two hypotheses 5.8 and 5.9 empirically.

$$\begin{aligned} &H_0 : \beta_2 = 0 \\ &H_a : \beta_2 \neq 0 \end{aligned} \tag{5.8}$$

$$\begin{aligned} &H_0 : \beta_2 = 0 \ and \ \beta_1 = 0 \\ &H_a : otherwise \end{aligned} \tag{5.9}$$

We determine whether we reject each null hypothesis by performing t–test on 5.8 and joint F–test on 5.9. Should the null hypothesis of 5.8 be rejected at 5% significance level or lower, we test the joint significance of both coefficients. If we also reject the second null hypothesis, we conclude an existence of a quadratic relationship.

Moreover, if we observe: $\hat{\beta}_1 > 0 \ and \ \hat{\beta}_2 < 0$, the relationship can be modeled as a concave function where the point of maxima can be computed according to formulas 5.10.

$$\begin{aligned} &\widehat{mispricing} = \hat{\beta}_0 + \hat{\beta}_1 f + \hat{\beta}_2 f^2 \\ &\frac{\partial \widehat{mispricing}}{\partial f} = 0 \\ &f^* = \left| \frac{\hat{\beta}_1}{2 \times \hat{\beta}_2} \right| \end{aligned} \tag{5.10}$$

Formulas 5.10 convey the fact that the point of maxima can be found where *futures* *volume* reaches f^* contracts traded during a given day.

In the next chapter, we discuss results to the contract–by–contract analysis, as well as, the stated hypotheses.

Chapter 6

Results and evaluation

This chapter reports results of contract-by-contract analysis 6.1, hypotheses testing 6.2 and the alteration of cost–of–carry model 6.3.

6.1 Contract-by-contract analysis

Prior to estimating all combinations of the regression models according to the contract-by-contract analysis, we first regressed each of the four explanatory variables separately to gauge the general significance. It turned out that explanatory variable *scam fraud breach (b)* was not significant at 5% level or lower for any of the 12 contracts. Hence, we decided to omit this variable from our population model 5.5.

This leaves us with *fds* model to analyse. In total, we run $(2^3 - 1) \times 6 \times 2 \times 2 = 168$ regressions, for a combination of 3 explanatory variables, 6 contracts, 2 assets and 2 dependent variables.

Table 6.1: Contract–by–contract models comparison with dependent variable: mispricing

	Bitcoin							**Ethereum**						
	f	(+)	d	(+)	s	(+)	*mean* $\bar{R}^2$	f	(+)	d	(+)	s	(+)	*mean* $\bar{R}^2$
f	**66**	100	-	-	-	-	0.194	**83**	100	-	-	-	-	0.134
d	-	-	**100**	33	-	-	0.415	-	-	**83**	40	-	-	0.373
s	-	-	-	-	**66**	75	0.100	-	-	-	-	**50**	100	0.069
fd	**66**	100	**100**	33	-	-	0.480	33	100	**66**	50	-	-	0.404
fs	**50**	100	-	-	33	0	0.257	**50**	100	-	-	33	50	0.163
ds	-	-	**83**	40	**50**	66	0.440	-	-	**83**	40	16	100	0.376
fds	**50**	100	**83**	40	16	0	0.504	**50**	100	**100**	33	33	50	0.424

Note: The dependent variable is the **mispricing term**. Rows denote 7 models (all combinations of 3 explanatory variables). The f, d and s columns depict the percentages of contracts for which f, d and s are statistically significant (at 5% or lower) out of the 6 contracts (values greater than 50% are in bold). The (+) columns represent the percentages of significant contracts with positive coefficients. $Mean\ \bar{R}^2$ was obtained by averaging six $\bar{R}^2$, which came from the same combination of explanatory variables, see Table 6.2.

Looking at Table 6.1, which was created by aggregating information from Table 6.2, the choice of the most optimal model is not straightforward. Should we merely choose based on the highest average of adjusted R^2 (*mean* $\bar{R}^2$), we would select the ***fds* model** for both Bitcoin and Ethereum as the *mean* $\bar{R}^2$ is 50.4% and 42.4% respectively. Nevertheless, for both of these assets, the estimated coefficient for s is not statistically significant for the majority of the contracts (only 16% for Bitcoin and 33% for Ethereum).

The ***fd* model** is more promising as the *mean* $\bar{R}^2$ for both Bitcoin and Ethereum drops only slightly, suggesting that s was not an important explanatory variable. The issue with the *fd* model is that d does not have the expected sign for the majority of the contracts. Out of all significant contracts, the positive sign turned out to be estimated only 33% of the time for Bitcoin and 50% for Ethereum. We hypothesized that the sign would be positive for most of the contracts as with more popularity, recognition and market adoption, the *mispricing* (and *basis*) should be increasing towards zero.

Unless there is an explanation for negative d, which we have missed, this leaves us with the ***f* model**, whose *mean* $\bar{R}^2$ is 19.4% and 13.4% for Bitcoin and Ethereum respectively. Not all contracts showed f to be statistically significant, but it was the majority (66% and 83%). Moreover, all significant coefficients have the expected sign. Matsui *et al.* [21] suggest that it is preferable to use the minimum number of explanatory variables required to describe the dependent variable.

Taking all this into consideration, we choose the f model to be the final model for both Bitcoin and Ethereum, though, for Ethereum, the *fd* model could be good as well.

Table 6.2[1] displays the details, which were utilized to create Table 6.1. There are 6 contracts for each asset, ranging from March 2021 to June 2022 expiration dates. Each contract has 7 combinations of the explanatory variables.

Table 6.2: Regression results – all combinations with dependent variable: mispricing

	Bitcoin				**Ethereum**			
	f	d	s	$\bar{R}^2$	f	d	s	$\bar{R}^2$
Mar-21	218.7*** (0.004)	-	-	0.264	95.71*** (0.008)	-	-	0.196
Mar-21	-	-497.6** (0.023)	-	0.132	-	-23.50*** (0.004)	-	0.451
Mar-21	-	-	-147.4** (0.012)	0.107	-	-	-26.11 (0.230)	0.015
Mar-21	199.1*** (0.005)	-404.6** (0.025)	-	0.350	51.42* (0.067)	-20.41*** (0.008)	-	0.495

[1]Statistical significance is denoted by $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. P-values are shown in parenthesis and were computed using Newey and West HAC standard errors [79, 82] with number of lags equal to $T^{\frac{1}{4}}$ as suggested by Greene [80].

Table 6.2 continued from previous page

	Bitcoin				Ethereum			
	f	d	s	$\bar{R}^2$	f	d	s	$\bar{R}^2$
Mar-21	260.7*** (0.000)	-	-202.3*** (0.000)	0.484	127.9*** (0.001)	-	-58.34*** (0.002)	0.334
Mar-21	-	-383.4* (0.089)	-102.7** (0.043)	0.169	-	-23.73*** (0.004)	3.007 (0.855)	0.440
Mar-21	246.7*** (0.000)	-184.8 (0.285)	-177.8*** (0.000)	0.491	69.34** (0.030)	-17.70** (0.030)	-21.86 (0.251)	0.501
Jun-21	116.2*** (0.000)	-	-	0.467	34.88*** (0.000)	-	-	0.284
Jun-21	-	-238.0*** (0.000)	-	0.685	-	18.62*** (0.000)	-	0.559
Jun-21	-	-	204.8*** (0.000)	0.237	-	-	41.06*** (0.002)	0.117
Jun-21	37.57** (0.043)	-194.1*** (0.000)	-	0.709	11.96** (0.020)	16.11*** (0.000)	-	0.579
Jun-21	125.2*** (0.000)	-	-28.93 (0.664)	0.464	46.45*** (0.000)	-	-25.74 (0.231)	0.295
Jun-21	-	-221.7*** (0.000)	48.16 (0.161)	0.693	-	18.66*** (0.000)	-0.355 (0.969)	0.554
Jun-21	39.56** (0.041)	-193.9*** (0.000)	-6.063 (0.881)	0.706	26.66*** (0.001)	16.50*** (0.000)	-33.90** (0.034)	0.610
Sep-21	0.871 (0.560)	-	-	-0.004	1.323*** (0.004)	-	-	0.059
Sep-21	-	-28.83** (0.040)	-	0.123	-	-2.237 (0.217)	-	0.030
Sep-21	-	-	-5.660 (0.286)	0.002	-	-	-0.540 (0.920)	-0.010
Sep-21	1.010 (0.397)	-29.08** (0.038)	-	0.123	1.734*** (0.001)	-3.293* (0.058)	-	0.131
Sep-21	1.027 (0.501)	-	-6.201 (0.279)	0.000	1.614*** (0.007)	-	-5.856 (0.391)	0.067
Sep-21	-	-29.27** (0.045)	0.807 (0.846)	0.114	-	-2.268 (0.120)	-1.290 (0.796)	0.021
Sep-21	1.003 (0.401)	-29.22** (0.047)	0.257 (0.954)	0.113	2.260*** (0.000)	-3.840** (0.015)	-9.260 (0.150)	0.163
Dec-21	-1.354 (0.908)	-	-	-0.010	0.536 (0.945)	-	-	-0.010
Dec-21	-	-292.9*** (0.000)	-	0.698	-	39.10*** (0.000)	-	0.499
Dec-21	-	-	-76.35 (0.278)	0.013	-	-	-4.970 (0.923)	-0.011
Dec-21	3.115 (0.530)	-293.5*** (0.000)	-	0.696	3.020 (0.553)	39.35*** (0.000)	-	0.498

Table 6.2 continued from previous page

	Bitcoin				Ethereum			
	f	d	s	$\bar{R}^2$	f	d	s	$\bar{R}^2$
Dec-21	17.52 (0.171)	-	-140.0* (0.055)	0.019	1.952 (0.857)	-	-14.57 (0.840)	-0.020
Dec-21	-	-302.6*** (0.000)	49.42* (0.073)	0.704	-	39.75*** (0.000)	29.43 (0.378)	0.501
Dec-21	-7.000 (0.318)	-306.5*** (0.000)	76.49* (0.065)	0.703	0.342 (0.950)	39.74*** (0.000)	27.74 (0.413)	0.500
Mar-22	16.10*** (0.001)	-	-	0.291	15.19** (0.027)	-	-	0.172
Mar-22	-	242.3*** (0.000)	-	0.675	-	-34.56*** (0.000)	-	0.581
Mar-22	-	-	68.88** (0.011)	0.103	-	-	64.23** (0.018)	0.162
Mar-22	7.856*** (0.006)	211.7*** (0.000)	-	0.733	5.883* (0.065)	-31.79*** (0.000)	-	0.600
Mar-22	40.21*** (0.000)	-	-184.8*** (0.000)	0.431	10.62 (0.561)	-	21.68 (0.754)	0.166
Mar-22	-	231.7*** (0.000)	33.34*** (0.009)	0.697	-	-32.51*** (0.000)	16.48 (0.145)	0.585
Mar-22	19.58** (0.011)	192.8*** (0.000)	-84.20* (0.057)	0.757	14.31 (0.151)	-33.00*** (0.000)	-41.65 (0.329)	0.606
Jun-22	3.745*** (0.001)	-	-	0.154	1.450*** (0.002)	-	-	0.010
Jun-22	-	41.95*** (0.003)	-	0.179	-	-3.956*** (0.000)	-	0.119
Jun-22	-	-	16.51*** (0.000)	0.130	-	-	10.05*** (0.001)	0.141
Jun-22	2.882*** (0.004)	33.99*** (0.008)	-	0.262	0.676 (0.254)	-2.746* (0.083)	-	0.121
Jun-22	3.640* (0.093)	-	0.543 (0.945)	0.145	-0.650 (0.367)	-	13.57** (0.013)	0.135
Jun-22	-	34.46*** (0.007)	12.03*** (0.001)	0.239	-	-2.143** (0.040)	6.990** (0.030)	0.156
Jun-22	3.583 (0.101)	34.35*** (0.007)	-3.680 (0.680)	0.255	-1.887* (0.066)	-3.300** (0.017)	15.55*** (0.006)	0.170

As a robustness check, we also performed the same analysis with dependent variable *basis*. As can be seen from Table 6.3, it is apparent that the same number of estimated coefficients is significant as in Table 6.2. This result was expected as the risk–free interest rate is still very low for all the contracts, including the last one, which expired in June 2022.

Table 6.3: Contract–by–contract models comparison with dependent variable: basis

	Bitcoin							Ethereum						
	f	(+)	*d*	(+)	*s*	(+)	*mean* $\bar{R}^2$	*f*	(+)	*d*	(+)	*s*	(+)	*mean* $\bar{R}^2$
f	**66**	100	-	-	-	-	0.196	**83**	100	-	-	-	-	0.117
d	-	-	**100**	33	-	-	0.418	-	-	**83**	40	-	-	0.376
s	-	-	-	-	**66**	75	0.100	-	-	-	-	**50**	100	0.071
fd	**66**	100	**100**	33	-	-	0.483	33	100	**66**	50	-	-	0.407
fs	**50**	100	-	-	33	0	0.259	**50**	100	-	-	33	50	0.165
ds	-	-	**83**	40	**50**	66	0.440	-	-	**83**	40	16	100	0.379
fds	**50**	100	**83**	40	16	0	0.508	**50**	100	**100**	33	33	50	0.427

The results of this section agree with findings of papers written by Hattori *et al.* [1] and De Blasis et al [2], which used CME, CBOE and Binance futures data. It differs from results by Matsui *et al.* [21], who found f not be to statistically significant for most Bitcoin contracts.

In the next section, we report the results of the two hypotheses of the quadratic relationship between the futures volume and the mispricing term.

6.2 Quadratic relationship

This section discusses the results of hypotheses stated in section 5.4.

Table 6.4 shows that 3 out of 6 Bitcoin contracts (50%) and 4 out of 6 Ethereum contracts (66%) can be modeled by quadratic function. For these 7 contracts, both hypotheses of zero coefficients (no significance) were rejected at 5% significance level or lower. The observed estimated signs indicate concave functional relationship. The points of maxima for these contracts were computed according to formula 5.10, and highlighted in bold in Table 6.4.

Looking at the Bitcoin futures datasets, it can be observed that **Jun–21 contract** has the daily futures trading volume higher than 4,360 (its point of maxima) on 2 different days: 19.5.2021 where the trading volume reached 7,145 trades, and on 10.6.2021 where the trading volume reached 5,497 trades.

The point of maxima for **Mar–22 contract** is 4,603, which was exceeded on 7 different days: 22.1.2022, 24.1.2022, 10.2.2022, 21.2.2022, 24.2.2022, 28.2.2022 and 16.3.2022. The highest trading volume for this contract was recorded on 24.2.2022 (7,694).

The point of maxima for **Jun-22 contract** is 7,429 traded futures. This point was exceeded on 5 different days, all close to expiration date: 13.6.2022, 14.6.2022, 15.6.2022, 18.6.2022 and 19.6.2022. The highest trading volume of 11,581 was recorded on 14.6.2022.

Table 6.4: Regression results – quadratic relationship – dependent variable: mispricing

	Bitcoin					**Ethereum**				
	f	f^2	$F-stats$	f^*	$\bar{R}^2$	f	f^2	$F-stats$	f^*	$\bar{R}^2$
Mar-21	218.7*** (0.004)	-	-	-	0.264	95.71*** (0.008)	-	-	-	0.196
Mar-21	704.8** (0.031)	-29.04 (0.109)	11.18*** (0.000)	-	0.289	520.0*** (0.000)	-368.7*** (0.000)	17.95*** (0.000)	**0.705**	0.409
Jun-21	116.2*** (0.000)	-	-	-	0.467	34.88*** (0.000)	-	-	-	0.284
Jun-21	294.8*** (0.000)	-3.381*** (0.000)	97.02*** (0.000)	**43.60**	0.658	82.83*** (0.000)	-9.675*** (0.001)	30.01*** (0.000)	**4.281**	0.367
Sep-21	0.871 (0.560)	-	-	-	-0.004	1.323*** (0.004)	-	-	-	0.059
Sep-21	1.810 (0.605)	-0.013 (0.790)	0.344 (0.710)	-	-0.014	2.095 (0.127)	-0.081 (0.537)	3.637** (0.030)	-	0.052
Dec-21	-1.354 (0.908)	-	-	-	-0.010	0.536 (0.945)	-	-	-	-0.010
Dec-21	-46.13 (0.170)	0.855 (0.103)	0.676 (0.511)	-	-0.007	13.01 (0.529)	-2.001 (0.483)	0.252 (0.778)	-	-0.015
Mar-22	16.10*** (0.001)	-	-	-	0.291	15.19** (0.027)	-	-	-	0.172
Mar-22	50.73*** (0.000)	-0.551*** (0.002)	35.95*** (0.000)	**46.03**	0.437	57.93*** (0.000)	-9.450*** (0.002)	28.03*** (0.000)	**3.065**	0.375
Jun-22	3.745*** (0.001)	-	-	-	0.154	1.450*** (0.002)	-	-	-	0.010
Jun-22	12.63*** (0.001)	-0.085*** (0.002)	15.14*** (0.000)	**74.29**	0.233	4.719*** (0.001)	-0.188*** (0.003)	9.585*** (0.000)	**12.55**	0.156

Note: Statistical significance is denoted by $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. P-values are shown in parenthesis and were computed using Newey and West HAC standard errors [79, 82] with number of lags equal to $T^{\frac{1}{4}}$ as suggested by Greene [80]. F-stats represents F-statistics value of joint F-test of f and f^2 and beneath it, in parenthesis, is the p-value of this test. f^* denotes the point of maxima (highlighted in bold), computed only for estimated regressions where both of hypotheses 5.8 and 5.9 were rejected at 5% significance level or lower. As stated in 5.5, f was scaled down by the factor of 100 for Bitcoin and by the factor of 10,000 for Ethereum; therefore f^* needs to be multiplied accordingly for interpretation purposes.

Pictorially, this is depicted in scatter plots 6.1, where quadratic curve was fitted according to estimates from Table 6.4. The above provides some evidence of Binance Bitcoin futures being manipulated by large entities, although further investigation needs to look at intraday data for the mentioned days. It is interesting to point out that 2022 contracts had more such days than 2021 contracts, which suggests that the number of manipulations is increasing over time, although it not possible to be conclusive as we only have limited number of observations—in this case 3 contracts.

Looking at Ethereum futures datasets, we can see some evidence of Ethereum futures manipulation as well. The **Mar–21 contract's** point of maximum is at 7,050 with 9 days going beyond this value: 4.2.2021, 22.2.2021, 23.2.2021, 8.3.20201, 10.3.2021 to 13.3.2021 and 15.3.2021. The highest trading volume of 11,120 was recorded on 23.2.2021.

Jun–21 contract, with the point of maxima equal to 42,810, contains 2 days above this value: 19.5.2021 (67,250) and 22.6.2021 (50,710).

Mar–22 contract has its point of maxima equal to 30,650 and 7 days which go beyond this number: 7.1.2022, 21.1.2022, 22.1.2022, 24.1.2022, 4.2.2022, 21.2.2022 and 24.2.2022. The highest trading volume was reached on 24.2.2022 (55,180).

Jun–22 contract has 5 days that are above the point of maxima of 125,500: 13.6.2022 to 15.6.2022, 18.6.2022 and 19.6.2022, highest of which is 15.6.2022 with trading volume equals to 220,900.

The results above indicate some form of manipulation of Ethereum futures contracts as well. In the next section, we report the results of our amended cost–of–carry model.

6.3 Cost-of-carry model

In section 3.5, we introduced two possible approaches of how the original cost–of–carry model 3.2 can be amended to better reflect cryptocurrency characteristics.

Unfortunately, the cryptocurrency insurance industry is still in its infancy and so we were not able to find suitable data to test the "Insurance" approach 3.9. Regarding the second method, "Not Moved Since" approach, see 3.12, we managed to find data, which can be used to calculate a proxy of *lost coins*, as a reminder see 4.3.3.

The results are mixed for Bitcoin and Ethereum. Regarding Bitcoin, the amended model gets closer to the market–quoted futures price for the first 2 contracts, however it is considerably worse than the original model for the remaining 4 contracts, see Figure 6.2. Regarding Ethereum, the findings are more promising. For the first 5 contracts, the amended model performs more optimally. The original model performs better only for the last two months of the last contract, see Figure 6.3.

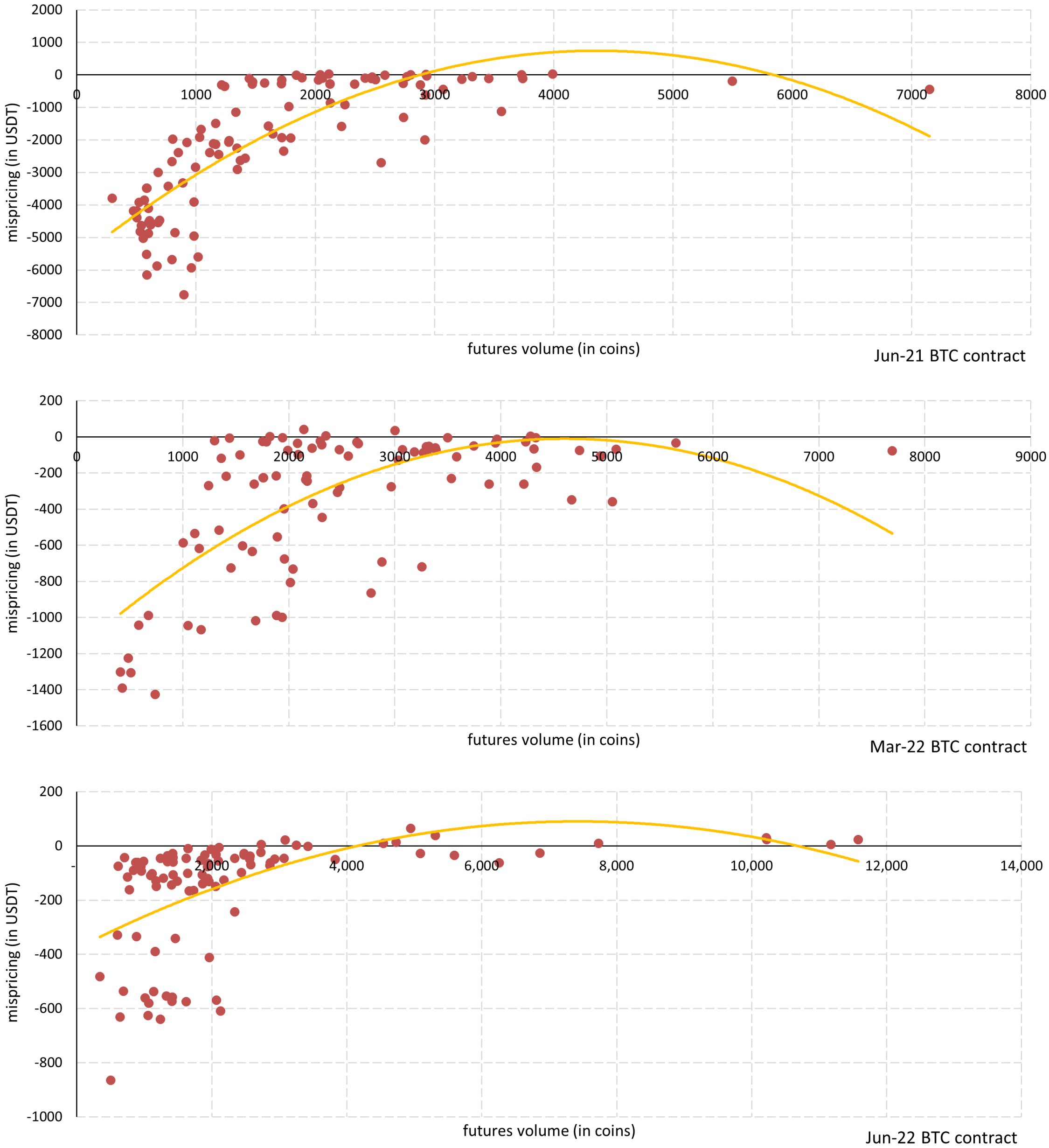

Figure 6.1: Scatter plots of mispricing term and futures volume with quadratic fitted curve (Bitcoin contracts)

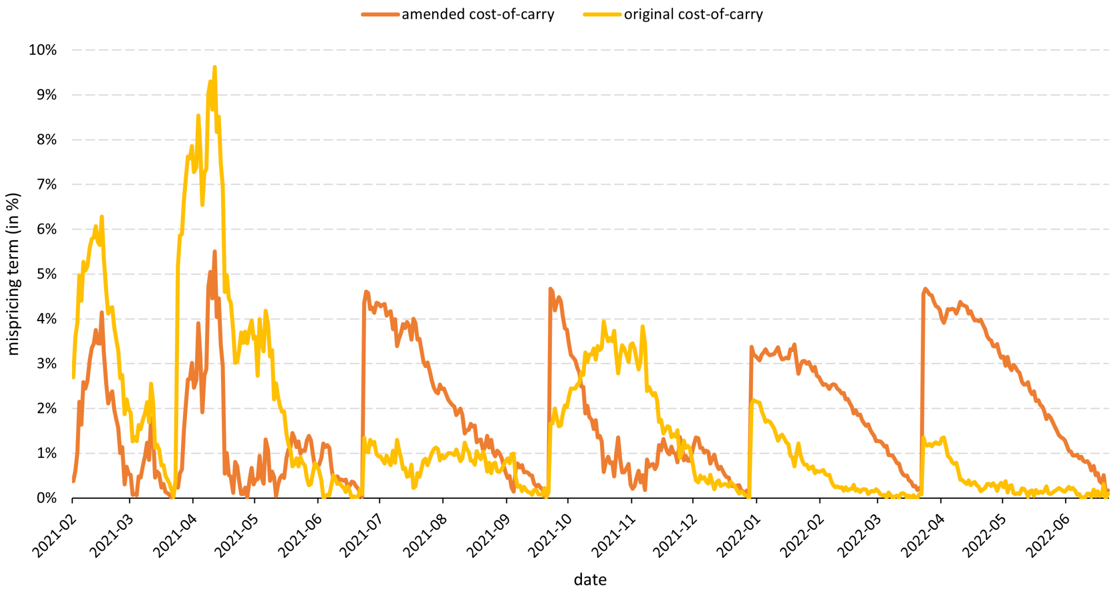

Figure 6.2: Comparison of Bitcoin mispricing term of amended and original cost-of-carry model

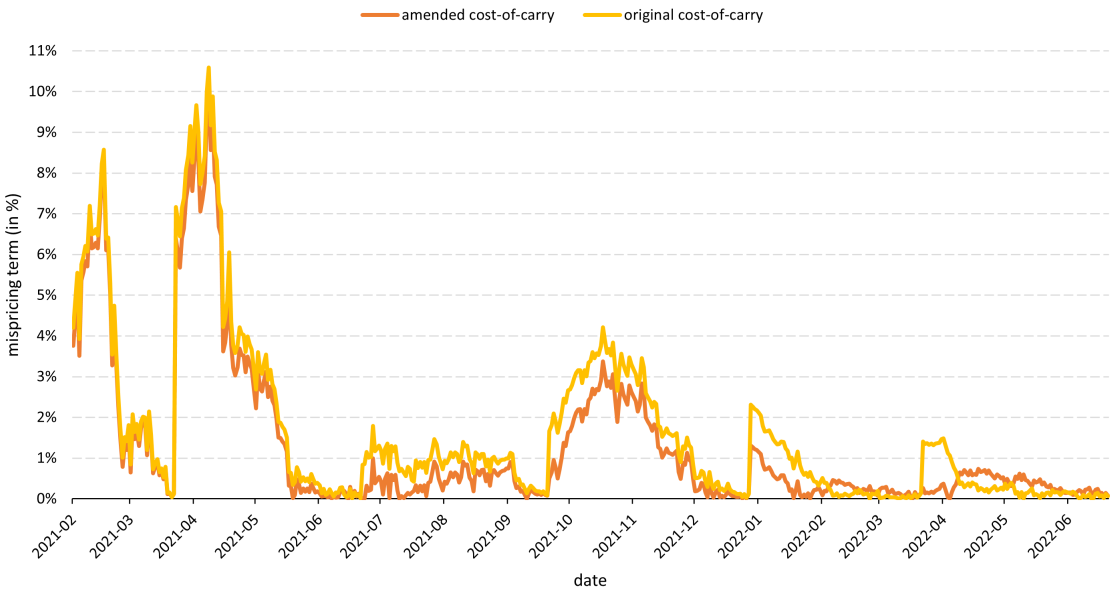

Figure 6.3: Comparison of Ethereum mispricing term of amended and original cost-of-carry model

Chapter 7

Conclusion

We collected data from Binance, FRED, Messari and Crystal Blockchain, and merged them by time period (daily frequency) to arrive at 12 datasets (representing 6 Bitcoin and 6 Ethereum contracts), which were then utilized in the contract–by–contract regression analysis. For Bitcoin, the f model transpired to be the most optimal, whereas for Ethereum, an argument can be made for either the f or fd model. In aggregates over all contracts, the f model's *mean* $\bar{R}^2$ was 19.4% for Bitcoin and 13.4% for Ethereum. Futures volume (f) turned out to be statistically significant for 66% of Bitcoin contracts and 83% for Ethereum contracts, with all significant coefficients having positive sign. The estimated coefficient of marketcap dominance (d) was highly statistically significant—100% of Bitcoin contracts and 83% of Ethereum contracts—however, mostly with a positive sign, which is not what we expected. The main limitation is the number of contracts available—6 contracts for each asset is not optimal for the contract–by–contract analysis. Further research could consider other cryptocurrencies and verify our results using data from other emerging exchanges. Specifically, it would be interesting to see whether the discovered dynamics are present in contracts, which are traded in high interest rate and inflationary environment.

Furthermore, two hypotheses were specified and later tested by running regressions with the quadratic form specification. The hypotheses asked a question of whether futures volume has a diminishing effect on the mispricing term (or the basis) and whether the linear and quadratic term are estimated with signs such as we obtain concave function or convex function, for which a point of maxima or a point of minima, respectively, can be found by taking the first partial derivative with respect to futures volume, equating it to zero, and solving for futures volume. We found out that 3 out of the 6 Bitcoin contracts and 4 out of the 6 Ethereum contracts can be modeled as a concave function. The points of maxima were computed and days whose futures volume exceeded this value were red–flagged as possible victims of market manipulation by large entities. In total, 14 and 23 days for Bitcoin and Ethereum, respectively, were red–flagged. Further research could investigate these days in more details, possibly with intraday data and a co–movement with other variables.

Finally, we amended the general cost–of–carry model, which is used to compute the theoretical, fair futures price by considering costs and benefits of carrying (holding) the spot asset. The benchmark was cost–of–carry model with only the risk–free rate as the cost. Our modified model included the cost associated with the possibility of losing the cryptocurrency (loss of wallet's PIN and recovery phrase). For these purposes, we computed a proxy called *lost coins proportion*, which acted as an additional cost that the spot investor indirectly needs to pay to hold the cryptocurrency. *Lost coins proportion* can be thought of as the probability of losing the coin, however it is defined as the difference between the total supply outstanding and the number of coins that have moved in the last 5 years. The results were mixed; for Bitcoin, the amended model performed worse than the benchmark in 4 out of 6 contracts; for Ethereum, the model performed better than the benchmark in 5 out of 6 contracts. The limitation of this approach is that the proxy is upward biased as it also includes coins that have not moved for 5 years because investors hold them as a long–term investment (instead of merely losing them). As a consequence, further research is needed on this topic. A promising candidate for a cost can be insurance fee to insure one coin (possibly paid as a proportion of the coins current market value). Unfortunately, this type of insurance product is not yet available for the cryptocurrency market so quotes can not be obtained.

Bibliography

[1] Hattori T, Ishida R. The relationship between arbitrage in futures and spot markets and Bitcoin price movements: Evidence from the Bitcoin markets. The journal of futures markets. 2021;41(1):105-14.

[2] De Blasis R, Webb A. Arbitrage, contract design, and market structure in Bitcoin futures markets. Journal of Futures Markets. 2022;42(3):492-524.

[3] Bitcoin Explorer: block 1;. Accessed: 2022-07-12. https://www.blockchain.com/btc/block/00000000839a8e6886ab5951d76f411475428afc90947ee320161bbf18eb6048.

[4] Urquhart A. The inefficiency of Bitcoin. Economics letters. 2016;148:80-2.

[5] Nadarajah S, Chu J. On the inefficiency of Bitcoin. Economics letters. 2017;150:6-9.

[6] Brauneis A, Mestel R. Price discovery of cryptocurrencies: Bitcoin and beyond. Economics letters. 2018;165:58-61.

[7] Tiwari AK, Jana RK, Das D, Roubaud D. Informational efficiency of Bitcoin—An extension. Economics letters. 2018;163:106-9.

[8] Jiang Y, Nie H, Ruan W. Time-varying long-term memory in Bitcoin market. Finance research letters. 2018;25:280-4.

[9] Cheah ET, Mishra T, Parhi M, Zhang Z. Long memory interdependency and inefficiency in Bitcoin markets. Economics Letters. 2018;167:18-25.

[10] Kapar B, Olmo J. An analysis of price discovery between Bitcoin futures and spot markets. Economics Letters. 2019;174:62-4.

[11] Baur DG, Dimpfl T. Price discovery in bitcoin spot or futures? The journal of futures markets. 2019;39(7):803-17.

[12] Fassas AP, Papadamou S, Koulis A. Price discovery in bitcoin futures. Research in International Business and Finance. 2020;52:101116.

[13] Lee S, Meslmani NE, Switzer LN. Pricing Efficiency and Arbitrage in the Bitcoin Spot and Futures Markets. Research in international business and finance. 2020;53:101200.

[14] Matsui T, Gudgeon L. The Speculative (In) Efficiency of the CME Bitcoin Futures Market. In: Mathematical Research for Blockchain Economy. Springer; 2020. p. 91-103.

[15] CORNELL B, REINGANUM MR. Forward and Futures Prices: Evidence from the Foreign Exchange Markets. The Journal of finance (New York). 1981;36(5):1035-45.

[16] MacKinlay AC, Ramaswamy K. Index-Futures Arbitrage and the Behavior of Stock Index Futures Prices. The Review of financial studies. 1988;1(2):137-58.

[17] Bhatt S, Cakici N. Premiums on stock index futures-some evidence. The journal of futures markets. 1990;10(4):367-75.

[18] Switzer LN, Varson PL, Zghidi S. Standard and Poor's depository receipts and the performance of the S&P 500 index futures market. Journal of Futures Markets: Futures, Options, and Other Derivative Products. 2000;20(8):705-16.

[19] Lian YM, Cheng CH, Lin SH, Lin JH, et al. A Cost of Carry-Based Framework for the Bitcoin Futures Price Modeling. Journal of Mathematical Finance. 2019;9(01):42.

[20] Kim ST. Is it worth to hold bitcoin? Finance research letters. 2022;44:102090.

[21] Matsui T, Al-Ali A, Knottenbelt WJ. On the Dynamics of Solid, Liquid and Digital Gold Futures. arXiv preprint arXiv:220209845. 2022.

[22] Roberts PC, Kranzler D. The hows and whys of gold price manipulation. Institute for Political Economy. 2014.

[23] Proof That Forward and Futures Prices Are Equal When Interest Rates Are Constant;. Accessed: 2022-06-09. `http://www-2.rotman.utoronto.ca/~hull/TechnicalNotes/TechnicalNote24.pdf`.

[24] Gorton G, Rouwenhorst KG. Facts and fantasies about commodity futures. Financial Analysts Journal. 2006;62(2):47-68.

[25] Hess AC, Kamara A. Conditional time-varying interest rate risk premium: Evidence from the treasury bill futures market. Journal of Money, Credit and Banking. 2005:679-98.

[26] Hull J. Options, futures, and other derivatives. Eleventh edition. ed. Harlow, UNITED KINGDOM: Pearson; 2021 - 2022.

[27] Schrimpf A, Sushko V. Beyond LIBOR: a primer on the new benchmark rates. BIS Quarterly Review March. 2019.

[28] Ashton P, Christophers B. On arbitration, arbitrage and arbitrariness in financial markets and their governance: unpacking LIBOR and the LIBOR scandal. Economy and Society. 2015;44(2):188-217.

[29] COUNCIL on FOREIGN RELATIONS: Understanding the Libor Scandal;. Accessed: 2022-06-24. https://www.cfr.org/backgrounder/understanding-libor-scandal.

[30] JOHANNES M, SUNDARESAN S. The Impact of Collateralization on Swap Rates. The Journal of finance (New York). 2007;62(1):383-410.

[31] FCA: Announcements on the end of LIBOR;. Accessed: 2022-06-24. https://www.fca.org.uk/news/press-releases/announcements-end-libor#:~:text=The%20FCA%20has%20confirmed%20that,month%20US%20dollar%20settings%3B%20and.

[32] Hull JC, White A. LIBOR vs. OIS: The derivatives discounting dilemma. The Journal of Investment Management, Forthcoming, Rotman School of Management Working Paper. 2013;(2211800).

[33] Hull J, White A. Valuing derivatives: Funding value adjustments and fair value. Financial Analysts Journal. 2014;70(3):46-56.

[34] Working Group on Sterling Risk-Free Reference Rates;. Accessed: 2022-06-24. https://www.bankofengland.co.uk/-/media/boe/files/markets/benchmarks/rfr-terms-of-reference.pdf.

[35] ECB: Private sector working group on euro risk-free rates recommends ESTER as euro risk-free rate;. Accessed: 2022-06-24. https://www.ecb.europa.eu/press/pr/date/2018/html/ecb.pr180913.en.html.

[36] BoJ: Interest Rate Benchmark Reform (Preparedness for the Discontinuation of LIBOR);. Accessed: 2022-06-24. https://www.boj.or.jp/en/finsys/libor/index.htm/.

[37] ARRC: Transition from LIBOR;. Accessed: 2022-06-24. https://www.newyorkfed.org/arrc/sofr-transition.

[38] ARRC: Secured Overnight Financing Rate Data;. Accessed: 2022-06-24. https://www.newyorkfed.org/markets/reference-rates/sofr.

[39] Aswath Damodaran: Country Default Spreads and Risk Premiums);. Accessed: 2022-06-24. https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/ctryprem.html.

[40] Fassas AP. Price discovery in US money market benchmarks: LIBOR vs. SOFR. Economics letters. 2021;204:109882.

[41] ARRC Formally Recommends Term SOFR;. Accessed: 2022-06-24. https://www.newyorkfed.org/medialibrary/Microsites/arrc/files/2021/ARRC_Press_Release_Term_SOFR.pdf.

[42] CME DataMine Query API for TERM SOFR RATES;. Accessed: 2022-06-24. https://datamine.cmegroup.com/#/datasets/cme.sofrt.

[43] CME Term SOFR Rates);. Accessed: 2022-06-24. https://www.cmegroup.com/market-data/cme-group-benchmark-administration/term-sofr.html.

[44] Shynkevich A. Impact of bitcoin futures on the informational efficiency of bitcoin spot market. Journal of Futures Markets. 2021;41(1):115-34.

[45] Nguyen H, Liu B, Parikh NY. Exploring the short-term momentum effect in the cryptocurrency market. Evolutionary and Institutional Economics Review. 2020;17(2):425-43.

[46] Venter PJ, Mare E, Pindza E. Price discovery in the cryptocurrency option market: A univariate GARCH approach. Cogent Economics & Finance. 2020;8(1):1803524.

[47] De Blasis R, Webb A. Arbitrage, contract design, and market structure in Bitcoin futures markets. Journal of Futures Markets. 2022;42(3):492-524.

[48] Nguyen KQ. Conditional beta and uncertainty factor in the cryptocurrency pricing model. arXiv preprint arXiv:201012736. 2020.

[49] Zhu P, Zhang X, Wu Y, Zheng H, Zhang Y. Investor attention and cryptocurrency: Evidence from the Bitcoin market. PLoS One. 2021;16(2):e0246331.

[50] Anyfantaki S, Topaloglou N. Diversification, integration and cryptocurrency market. Integration and Cryptocurrency Market (March 29, 2018). 2018.

[51] Rompolis LS. Risk-Free Rates and Variance Futures Prices. Journal of Futures Markets. 2016;36(10):943-67.

[52] Blanco R, Brennan S, Marsh IW. An empirical analysis of the dynamic relationship between investment-grade bonds and credit default swaps. The Journal of finance (New York). 2005;60(5):2255-81.

[53] Messari's Crypto Data API (1.0.0);. Accessed: 2022-08-16. https://messari.io/api/docs.

[54] Krombholz K, Judmayer A, Gusenbauer M, Weippl E. The other side of the coin: User experiences with bitcoin security and privacy. In: International conference on financial cryptography and data security. Springer; 2016. p. 555-80.

[55] Wright KB. Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. Journal of computer-mediated communication. 2005;10(3):JCMC1034.

[56] CME Group Announces Launch of Bitcoin Futures;. Accessed: 2022-08-15. https://www.cmegroup.com/media-room/press-releases/2017/10/31/cme_group_announceslaunchofbitcoinfutures.html.

[57] CME Group Announces Launch of Ether Futures;. Accessed: 2022-08-15. https://www.cmegroup.com/media-room/press-releases/2021/2/08/cme_group_announceslaunchofetherfutures.html.

[58] BITCOIN FUTURES - CONTRACT SPECS;. Accessed: 2022-08-15. https://www.cmegroup.com/markets/cryptocurrencies/bitcoin/bitcoin.contractSpecs.html.

[59] ETHER FUTURES - CONTRACT SPECS;. Accessed: 2022-08-15. https://www.cmegroup.com/markets/cryptocurrencies/ether/ether.contractSpecs.html.

[60] USDT-Margined Quarterly Contract Specification and Trading Rules;. Accessed: 2022-08-15. https://www.binance.com/en/support/faq/3ae441db4ae740e19af3fe9228eb6619.

[61] What Are USDT-Margined Futures and Coin-Margined Futures;. Accessed: 2022-08-15. https://www.binance.com/en/support/faq/85eac2bba0b342819122dc9bd4745e9b.

[62] Skew - dashboard: trading volume and open interest of major cryptocurrency exchanges;. Accessed: 2022-08-15. https://analytics.skew.com/dashboard/bitcoin-futures.

[63] Coinmarketcap - Top Cryptocurrency Derivatives Exchanges;. Accessed: 2022-08-15. https://coinmarketcap.com/rankings/exchanges/derivatives/.

[64] Binance - Trade Crypto Futures;. Accessed: 2022-08-15. https://www.binance.com/en/futures.

[65] Bartoszynski R, Niewiadomska-Bugaj M. Probability and statistical inference. John Wiley & Sons; 2020.

[66] Wooldridge JM. Introductory econometrics : a modern approach. 6th ed. Australia: Cengage Learning; 2016 - 2016.

[67] BTCUSDT Price Index;. Accessed: 2022-08-15. https://www.binance.com/en/futures/funding-history/3.

[68] Binance API documentation;. Accessed: 2022-08-15. https://binance-docs.github.io/apidocs/futures/en/#general-info.

[69] Binance - Insurance Fund Balance data;. Accessed: 2022-08-15. https://www.binance.com/en/futures/funding-history/2.

[70] About Pandas;. Accessed: 2022-08-15. https://pandas.pydata.org/about/.

[71] Regular expression operations;. Accessed: 2022-08-15. https://docs.python.org/3/library/re.html.

[72] 4-Week Treasury Bill Secondary Market Rate, Discount Basis;. https://fred.stlouisfed.org/series/DTB4WK.

[73] 3-Month Treasury Bill Secondary Market Rate, Discount Basis;. Accessed: 2022-08-16. https://fred.stlouisfed.org/series/TB3MS.

[74] Crypto & DeFi Security Breaches, Fraud & Scams Report;. Accessed: 2022-08-16. https://crystalblockchain.com/security-breaches-and-fraud-involving-crypto/.

[75] Beautiful Soup Documentation;. Accessed: 2022-08-17. https://beautiful-soup-4.readthedocs.io/en/latest/.

[76] Selenium 4.4.0;. Accessed: 2022-08-17. https://pypi.org/project/selenium/.

[77] Tsay RS. Analysis of financial time series. John wiley & sons; 2005.

[78] Hatanaka M. Time-series-based econometrics : unit roots and co-integrations /. Advanced texts in econometrics. Oxford: Oxford University Press; 1996.

[79] NEWEY WK, WEST KD. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica. 1987;55(3):703-8.

[80] Greene WH. Econometric analysis. 7th ed. Harlow: Pearson Education; 2011.

[81] Abdullah A, Bakar MJA. The Gold Market and the Value of the US Dollar. International Business Research. 2015;8(3):190.

[82] Zeileis A. Econometric Computing with HC and HAC Covariance Matrix Estimators. Journal of statistical software. 2004;11(1):1-17.

[83] Wilmott P. Paul Wilmott introduces quantitative finance. John Wiley & Sons; 2007.

Chapter 8

Appendix

8.1 Deriving continuous compound interest

Consider the following formula:

$$FinalValue = ValueNow \times (1 + r)^t \tag{8.1}$$

$ValueNow$ can be represented as *one unit of cash* as we have done so far in the paper to simplify the calculations. Then we are left with $(1 + r)^t$ which is discretely compounding the *one unit of cash* over t number of years with one payment per year.

Let's try to increase the frequency of payments over the period, where n represents the number payments per year:

$$FinalValue = (1 + \frac{r}{n})^{n \times t} \tag{8.2}$$

Next, we want to see what would occur should the number of payments approach infinity (continuous, "non-stop" payments). Note that Taylor series is used for the derivation:

$$\begin{aligned} \lim_{n\to\infty} (1 + \frac{r}{n})^{n \times t} &= \lim_{n\to\infty} e^{n \times t \times \ln{(1 + \frac{r}{n})}} \\ &= \lim_{n\to\infty} e^{t \times (r - \frac{1}{2} \times \frac{r^2}{n^1} + \frac{1}{6} \times \frac{r^3}{n^2} + \ldots)} \\ &= e^{t \times r} \end{aligned} \tag{8.3}$$

Therefore, under the assumption of continuous compounding, equation 8.2 can be rewritten as follows [83]:

$$FinalValue = e^{t \times r} \tag{8.4}$$