Erasmus University Rotterdam
Erasmus School of Economics
Master Thesis Data Science and Marketing Analytics

The Causal Impact of Google's Gemini Release on Artificial Intelligence-Related Cryptocurrencies' Price Movements

Zhang, M. (531180)

Supervisor:	Dr. Zhang, X.
Second assessor:	Dr. Avagyan, V.
Date final version:	July 19, 2024

Contents

Abstract

This research investigates the causal impact of Alphabet's Gemini AI technology release on the price movements of AI-narrative and Big Data-related cryptocurrencies. To conduct this analysis, multiple control and treatment groups were established, using two reputable sources: *CoinMarketCap* and *CoinGecko*. Building on existing literature, this study uncovers latent structures in price movements using four statistical techniques: Lagged Dependent Variable (LDV), Difference-In-Difference (DID), Synthetic Difference-In-Difference (SDID), and Synthetic Control (SC). Previous research identified a significant impact of OpenAI's ChatGPT launch on cryptocurrency prices. Building on this, this study examines the effects of Alphabet's Gemini release in late 2023, investigating whether these effects differ from those of ChatGPT due to the inherent similarities between the two technologies. The results suggest that both the LDV and SDID display positive effects of Gemini's release on AI-coin price fluctuations, ranging from 0.2% to 0.3%. Consequently, this research provides evidence supporting the hypothesis that the release of Gemini influenced the price fluctuations of AI and Big Data cryptocurrencies between October 1, 2023, and March 31, 2024. By doing so, this study contributes to the growing body of knowledge on the interplay between technological advancements and cryptocurrency markets, highlighting the significance of adjacent technology release impacts on cryptocurrency price volatility.

1 Introduction

The testimony of OpenAI's CEO, Sam Altman, before the US Senate on May 16th, 2023, stands as a pivotal moment in the narrative surrounding artificial intelligence (AI) and its implications. This event transcended mere corporate affairs, capturing the attention of millions worldwide and sparking profound discussions about the future trajectory of AI technology. ChatGPT is an advanced language model with natural language processing abilities, enabling real-life applications along a plethora of fields. This model has been trained on a vast constellation of textual inputs, enabling it to process human-like inputs, commonly referred to as prompts (OpenAI, 2022). Its usability in domains has been proven by the numbers. The user count stands at approximately 180 million, which marks an 80% increase since its monumental achievement of 100 million users in January 2023 (Meer, 2024). Alphabet's Gemini (Google's parent company), which will be referred to as *Gemini* in this paper, exhibits similar characteristics as this model was built to be multidimensional, enabling seamless understanding and cross information type, including text, image, audio, video and code analysis (Sundar Pichai, 2023). Additionally, Alphabet claims that Gemini has outperformed state-of-the-art academic benchmarks used in large language models' (LLM) research and development in 30 out of 32 cases (Sundar Pichai, 2023). As Gemini was released as, coined by Marr (2024), *an attempt at dethroning ChatGPT as benchmark large language model*, comparisons are inevitable and hence this sparked the upcoming research design. To be more precise, the focus will be on the AI narrative and its inevitable impact on financial assets across the board.

Therefore, given the heightened scrutiny and attention surrounding AI-related technologies, investors and market participants are likely to closely monitor events such as Altman's testimony for insights into regulatory trends, industry dynamics, Gemini's release, and technological advancements. Consequently, understanding the impact of such pivotal moments on AI-related cryptocurrency prices can provide valuable insights into the interplay between technological innovation, regulatory environments, and market dynamics. In light of this, this research aims at **evaluating the causal impact of Gemini's release on the trading of AI-related narrative crypto tokens and their subsequent price movements**.

The significance of the aforementioned testimony and competitor's launch cannot be understated. This lies in its reflection of broader societal concerns regarding the implications of AI-advancement. Altman's appearance before the Senate underscored the need for regulatory frameworks and ethical considerations to guide the development and deployment of AI systems such as ChatGPT and Gemini. The discussion encompassed the potential benefits of AI, such as enhanced efficiency and productivity, alongside the accompanying risks, including job displacement and ethical dilemmas. This series of events was followed up by a consequent layoff of Altman as leads man of OpenAI in November of 2023, which led to all time high prices of Microsoft, one of Altman's and OpenAI's greatest benefactors. Shortly after this news, Alphabet undertook action into releasing its own AI model, namely: *Gemini*. This turns to the question of the existence of influence of AI-related events on AI-related financial assets, in particular: cryptocurrencies.

In the context of cryptocurrency markets, this event assumes additional importance due to its association with one of the most influential narratives of recent times – the intersection of AI and blockchain technology. Cryptocurrencies linked to AI projects represent a niche market that is particularly sensitive to developments in the AI landscape. The testimony, therefore, may have served as a catalyst for shifts in investor sentiment and subsequent price movements within this subset of coins. Prior research on the launch effects and spillovers of ChatGPT introduced by OpenAI have been studied by Saggu & Ante (2023). Keeping Saggu & Ante (2023) as a reference for this work, this literary piece will instead focus on the release of Gemini alongside the use of alternate research methods (including the Lagged Dependent Variable (LDV) among others), which will be introduced thoroughly in the methods section after some brief introductory remarks. Additionally, this paper will be focusing on a different time window. This will enable literature to assess time-series related differences in unison with technology advancements in the AI-landscape.

On this note, Ante & Demir (2023) demonstrate promising results in the light of financial *bear markets*, where positive average treatment effects were associated with the launch of ChatGPT, namely: 10.7 % to 15.6 % in one month post-launch. Bear markets are characterized by declining sentiments in financial markets, accompanied by falling asset price valuations and negative expectations (Bouri et al., 2018), whereas the opposite holds true for bull market conditions. The presence of a bear market environment made this finding more relevant as in inauspicious market conditions, this release continued to drive cryptocurrency AI-assets' prices positively. However, currently, the crypto market finds itself in a bullish state, in which the performance of Gemini can be assessed based on new market conditions. This has not been studied by either Ante & Demir (2023) or Saggu & Ante (2023), representing a notable shift in perspective. Consequently, the underlying hypothesis is that the release of Gemini will have a substantial causal impact on the price volatility of cryptocurrency assets within the predefined window of this research.

In uncovering a causal effect of Gemini's release on AI-related assets' price movements, two main methods will be utilized alongside several benchmark methods. Firstly, the Synthetic Difference-in-Difference (SDID), popularized by Arkhangelsky et al. (2021); Clarke et al. (2023) will be used to unveil the relative performance of a group of AI-tokens relative to control variables. The reason for choosing this method as opposed to the traditional and commonly used Difference-In-Difference (DID) model is based on the results found by both Saggu & Ante (2023); O'Neill et al. (2016), involving both respective methods. In line with DID and SDID methodology, DID heavily relies on the parallel trends assumption, which the SDID does not depend on. This will be explained in detail further in this paper. The aforementioned reasons, therefore, provide sufficient evidence to adopt more sophisticated and profound methods for uncovering the trend-altering causal effects of technology releases.

In addition to this method, a Lagged Dependent Variable (LDV) approach will be utilized (O'Neill et al., 2016) to find causal effects claims based on the launch of Gemini. O'Neill et al. (2016) demonstrated feasible and promising results in light of a context concerning best price tariffs (BPT) for hip fractures. The results were supportive in favor of LDV, having the most efficient and least biased estimators (O'Neill et al., 2016) compared to the SDID, DID and Synthetic Control (SC).

Lastly, for bench marking purposes, the DID will be used alongside the SC. These methods will serve as comparison to the main methods employed within the confines of this dissertation.

Apart from the means for analysis, theoretical frameworks were used to elaborate on potential causal effects. Firstly, market efficiency theories support the notion of perceived quality signals, in accordance to signaling theory Connelly et al. (2011). The release of Gemini may, as discussed by Ante & Demir (2023), serve as a quality indicator, given its significant media attention rate and adjacent retail investor valuations. Even though, investors, both retail and institutional, are prone to cognitive influences and heuristics (Ante, 2023), presence of large numbers of institutional investors prevents such overvaluation effects, given good information environments (Hu et al., 2016). Secondly, correlation neglect is present in investors (Kallir & Sonsino, 2009), in which investors treat correlation among variables as uncorrelated. This is potentially problematic for AI-assets' price movements in the long term, as investors possibly neglect the probability of correlation of heightened prices and Gemini's respective launch. Thirdly, networks effects and the existence thereof cannot be neglected. According to De Giorgi et al. (2020), consumption network effects can results in having magnified multiplier effects at the macro-level. This implies that the launch of Gemini may have sparked a self-supporting cycle of growing interest into the AI-space and also the AI-related assets' space. This may have driven continuous increases in users, demand, and valuations. Consequently, the price may also have been influenced by these factors.

The interplay of these factors underscores the rising significance of advancements in both AI and cryptocurrency realms. Consequently, research in this field becomes increasingly pivotal, particularly with the ongoing transition of traditional financial markets towards decentralized environments.

The emphasis of this research paper is to determine whether the results differ from the findings of Saggu & Ante (2023). In other words, the results of this prior study may not display the same causal effects as those relevant to Gemini's release due to time-bound factors and other covariates that significantly influence the price movements of AI-coins. Namely, Saggu & Ante (2023) found a significant increase in AI crypto turnover by 35.3% – 41.3% in the two months adjacent to the release of ChatGPT. Although similar, this research examines a different window, along with different controls and different methods as aforementioned. Moreover, tests will be conducted based on differing economic conditions, commonly referred to in the crypto space as the bull and bear market states, for which both have been present in recent years, allowing for a comprehensive overview of Gemini's release.

Lastly, despite the growing interest in both research and finance, empirical applications remain stagnant. especially considering that a limited number of firms are solely focused on AI developments. Most apply mere fractions of AI-related tools into their respective business models. Currently, this number of AI-centered firms is increasing, however research on the consequences of AI-related news, especially AI-model releases, on cryptocurrency market spaces has been limited. This dissertation aims to bridge this gap and it attempts to find potential causal effects of said releases on price movements, both now and in future applications.

2 Literature Review

AI-assets are closely interconnected to AI-progress itself. This will be discussed first to lay a fundamental understanding of the technology and why events related to AI will have influence over the price movements of AI-assets. AI-tokens, generally speaking, have seen many advancements in its blockchain technology. Blockchain has the ability to administer connections via smart contracts, after which AI assists in human-like capabilities (Salah et al., 2019). In essence, what this implies is that cryptocurrencies are based on the notion of cryptography, in which sensitive information is concealed from unwanted third parties. Therefore, cryptocurrencies are digital assets, each possessing unique keys, and developed by specialized programmers. AI is used to create an advanced machine learning program to emulate human-like intelligence. This will then be used in the development of AI-tokens, in which the parameters and structure are set by AI-machine learning. This includes the addition, deletion and verification of data on the stream of transactions, recorded on the blockchain (Ganapathy et al., 2020). To further corroborate these recent and past findings, Jeon et al. (2022) also speak of the fusion of AI and blockchain and the powerful implications thereof, in the ever growing Metaverse, in which the Metaverse is commonly referred to as a post-reality environment or universe that is perpetual and establishes multi-user environments that blend physical and digital realities (Mystakidis, 2022).

Having understood the essence of blockchain and AI technology, previous research on ChatGPT's respective effects must be covered before ultimately linking both and inducing Gemini's potential outcomes. Previous studies have examined the effects of launching ChatGPT to the public, most prominently in financial research (Dowling & Lucey, 2023), computer engineering (Sobania et al., 2023) and education (Adeshola & Adepoju, 2023), among others. Research on the effects of Gemini, essentially, a family of multi-faceted AI-models (Team et al., 2023), have been limited if not non-existent as of now. As the developments and the release of the model have been short-lived, research as such has not been conducted yet. Similar effects, comparable to the literature mentioned before is yet to be unveiled, in which this work will greatly benefit the existing literature on causal effect inference of AI-events and cryptocurrency fluctuations. A study by Ante & Demir (2023) has unveiled positive abnormal returns in 90% of all tokens within the research portfolio, for which the returns increased up to 41% after the launch of ChatGPT. The aim of this research is to find whether such effects exist for technological advancements in the AI-realm, and whether such technologies can exhibit similar effects to ChatGPT's launch.

In light of this, Haleem et al. (2022) discuss multiple applications of ChatGPT, which has seen multi-level real applications. Moreover, similarities between ChatGPT and Gemini have been discussed contextually by Rane et al. (2024). They argue that ChatGPT takes prevalence in creative text arrangements and natural language fluency, whereas Gemini shows powerful promise in factual accuracy and search integration by means of Google's infrastructure. Both, however have unique capabilities and qualities, thus no conclusive verdict can be drawn. Hence, although ChatGPT has proven to be a significant indicator of AI-related tokens' price movements (Saggu & Ante, 2023; Ante & Demir, 2023), few papers have been published on the potential effects of alternative AI-based models on AI-related tokens' price performance.

Adding to that notion and accounting for the growing importance and interest of AI, AI-related assets have seen momentum in recent times. Namely, multi-faceted releases of natural language processing tools with a plethora of features have been released, prompting interest in both AI and cryptocurrencies alike. Main cryptocurrencies such as Bitcoin, Ethereum and Solana have been prevalent for years, and for good reason. These coins are commonly referred to as layer-1 tokens. This analysis on upcoming narratives and corresponding tokens is important to unveil the inner workings of future strongholds, apart from the layer 1s. Jareño & Yousaf (2023) found a significant relationship between AI-backed stocks and the AI-token space. This research, now more than ever, can be revolutionary as an initial step towards understanding developments within the developer space of tokens, technology and innovation and the role it plays in financial (abnormal) returns. Furthermore, increased narrative attention has proven to be significant in the amount of returns expected by respective cryptocurrency tokens. This was proven by Nguyen et al. (2023) as part of research unveiling the causal effect of increased attention using Google Search Trends on perceived returns. The corresponding importance of narratives is also emphasized by Reijers & Coeckelbergh (2018) as a means to understand cryptocurrency tokens from a societal perspective.

Moreover, to evaluate the potential correlation between the launch of Gemini and subsequent changes in the prices of AI-assets, again we draw on the foundations of market efficiency theories, as popularized by Fama (1970, 1991) among others. These theories emphasize the significance of the perceived quality of public information. The introduction of a sophisticated AI-model through Gemini's launch can serve as a proxy measure for this quality. This proxy may be interpreted positively by both retail and institutional investors, aligning with signaling theory (Connelly et al., 2011). This proposition is consistent with existing literature, which explores how perceptions, whether positive or negative, can influence market dynamics (Howe, 1986).

Lastly, as aforementioned, causal effects of ChatGPT's launch on AI-assets' price movements have been uncovered and proven in the past, however the existence of such effects for the release of Gemini have not been studied before. This provides opportunity in the literary space, unveiling similar effects of Gemini to ChatGPT, from which the effects can be extended to many more applications. The implications, examined and presented in this literary work can be used for future employment into causal effects studies related to AI-news effects on metrics other than price movements. The possibilities are infinite and this research will be a considerable contribution to existing literature and perhaps it can even prompt additional research into this field of study.

3 Data

Daily AI-asset price data will be collected from CoinGecko (CG) and CoinMarketCap (CMC), respectively. All price data will be denoted in United States Dollar (USD). The AI coins, chosen to be in the cohort of CG denote the top 30 in terms of market capitalization. This sub sample approach was taken by Pessa et al. (2023) as an appropriate range in the study of the effect of market capitalization on large price fluctuations of cryptocurrency assets. Additional data will be uncovered from S&P

Cryptocurrency BDM Ex-Mega Capitalization Index (SPCBXM) and S&P Cryptocurrency BDM Ex-Large Capitalization Index (SPCBXL) to get a comparative measure and ensure robustness to discrepancies in data sources regarding selection biases. The SPCBXL excludes large capitalization tokens, whereas the SPCBXM excludes the mega capitalization tokens, which will be highlighted in the following section. The data spans from approximately 2 months before (October 1st, 2023) Gemini's launch (December 6th, 2023) to approximately 5 months after its initial release (March 31st 2024). This timeframe was deemed feasible as this period marks a transitory switch between bullish and bearish market conditions, allowing for an alternate perspective as compared to Ante & Demir (2023). This data will contain information on the following groups, distinguished as follows:

- 1. AI-Assets (treatment group)
- 2. Non-AI-Coins (excluding stable-coins) (control group 1)
- 3. Non-Stable Comparison Indexes (control group 2)

These groups include coins established at least as long as the defined timeframe, indicating equal windows considered for analysis, with exception for the index variables. This is especially pivotal for the Synthetic Difference-In-Difference Analysis (SDID) since this approach requires balanced data. Therefore, 183 periods (days) will be considered for both the AI treatment group and the non-AI assets control group. Conversely, the indexes will be considering a period of 130 days, emphasizing the factual trading of indexes on weekdays only. Furthermore, it must be mentioned that the indexes taken are suggested and corroborated by Saggu & Ante (2023). The separation between the control groups and treatment groups are established based on the ground of the narratives inhibiting the cryptocurrency mainstream world. Coins non-related to AI-technology are therefore considered to be unrelated to AI-events as well, in particular: the Gemini release. This was an approach similar to the Ante & Demir (2023) study. Thus, the indexes and both the control and treatment group contain several constituents, denoted in Table 4. A comprehensive list of all the tickers and their respective names in full will be provided in the Appendix.

On this note, taking Saggu & Ante (2023) as an inspiration to unveiling causal effects of Gemini on AI-coins' price movements, they found that AI-assets display resilience among a generally down trending market sentiment, which is unequivocally remarkable for an asset type as short lived as this AI-narrative related tokens. This is further established by initial descriptive statistics as the mean returns are 1.2% approximately over the period considered in this research.

The dependent variable will be denoted by AI-assets price fluctuations, whereas the independent variables will be constituted of multiple panels, of which stable coins represent subgroups and multiple indexes to account for additional variation. Furthermore, market conditions will be considered, for which bull and bear market conditions will be accounted for by means of a binary variable, marking a value of 1 for bull markets and 0 otherwise. Additionally, exogenous factors will also be introduced in formulation as covariates, which will be elaborated on in the next section.

The AI and big data related assets will strictly be greater than the 50 million USD threshold. This threshold was carefully chosen after deliberation on influence factors regarding cryptocurrencies, a similar approach and threshold was also taken by Saggu & Ante (2023).

Table 1: Descriptive Statistics Log Returns of CoinGecko AI-assets and CoinMarketCap AI-Assets

Ticker	Obs.	N	Mean	St. Dev.	Min	Max	Skew.	JB
GAI	5,490	30	0.011	0.081	−0.506	1.067	1.919	0.000***
CAI	10,065	55	−0.009	0.073	-0.748	0.445	-1.577	0.000***

*GAI denotes AI-assets related to CoinGecko, whereas CAI denotes AI-assets related to CoinMarketCap. JB represents the Jarque-Bera Test for Normality. *, **, *** denote the 0.1, 0.05 and 0.01 confidence levels, respectively.*

Table 2: Descriptive Statistics Log Returns of CoinGecko and CoinMarketCap Control Cohorts

Ticker	Obs.	N	Mean	St. Dev.	Min	Max	Skew.	JB
GCon	5,490	30	0.006	0.047	−0.456	0.452	1.492	0.000***
CCon	10,065	55	−0.006	0.058	−1.172	0.454	−2.793	0.000***

*GCon denotes control coins related to CoinGecko, whereas CCon denotes control cohorts related to CoinMarketCap. JB represents the Jarque-Bera Test for Normality. *, **, *** denote the 0.1, 0.05 and 0.01 confidence levels, respectively.*

Table 3: Descriptive Statistics Log Returns of the S&P500 Large Capitalization and Mega Capitalization Cryptocurrency Indexes

Ticker	Obs.	N	Mean	St. Dev.	Min	Max	Skew.	JB
SPCBXM	130	500+	0.008	0.037	−0.109	0.118	0.067	0.256
SPCBXL	130	500+	0.009	0.035	−0.104	0.121	0.212	0.182

*SPCBXL denotes the S&P500 Ex-Large Market Capitalization and where SPCBXM denotes the S&P500 Ex-Mega Market Capitalizations. JB represents the Jarque-Bera Test for Normality. *, **, *** denote the 0.1, 0.05 and 0.01 confidence levels, respectively.*

3.1 Descriptive Statistics

Table 1, Table 2 and Table 3 show the distributions of the separate datasets and their logarithmic returns (please refer to Equation 5 for the formulation). From this, it can be deducted that Table 1 showcases both a negative and positive effect ranging from −0.9 % to 1.1% positive returns for the period. Noticeably, CG pure AI-assets perform positively whereas the AI and Big Data coin sentiment seems to be negative on average. The skewness is negative for the CMC cohort and positive for the GC cohort, indicating left-skewness and right-skewness and henceforth implying more extreme positive returns than negative ones respectively for CG and the the other way around for the CMC cohort. Additionally, Table 2 shows that the overall performance of the control variables (largest coins in terms of market capitalization) performed positively for the top 30 coins on CG, but negatively for the top 55 coins on CMC. This is surprising as a similar effect was to be expected, however this may be due to the size of both samples. Next, Table 3 showcases the performance of a variety of coins, where both indexes (Ex-Large and Ex-Mega market capitalizations) are exhibiting positive average effects over a period of 130 days respectively. In addition to this, the skewness is positive for both indexes, indicating larger extreme positive returns than negative ones. As a concluding remark, it must be noted that the distributions from which the returns are drawn, are normally distributed, showing no evident favoritism for extreme negative and positive ends of the distribution when comparing the sample quantiles to the theoretical quantiles drawn from a normal distribution with mean 0 and standard error 1 (Ford, 2015). This can be concluded based on the visual representations in Figure 1 and Figure 2. Lastly, the Jarque-Bera Test for normality are all significant for Tables 1 and 2, implying non-normal distributions for the data concerning these cohorts (Thadewald & Büning, 2007). The opposite holds true for the indexes displayed in Table 3.

Additional descriptive statistics on the attributes used in this paper will be

Table 4: List of Tickers by Group

Group	**Ticker(s)**
CoinGecko Control Group	BTC, ETH, BNB, XRP, ADA, DOGE, MATIC, SOL, DOT, SHIB, LTC, AVAX, TRX, UNI, TON, LINK, NEAR, ICP, LEO, PEPE, KAS, APT, XMR, STX, HBAR, FIL, MNT, XLM, CRO, VET
CoinGecko Artificial Intelligence Group	RNDR, TAO, FETCH, AGIX, GRT, AKT, AIOZ, GLM, OCEAN, TRAC, NOS, RSS3, X0X, ARKM, PAAL, ORAI, AGI, RLC, IQ, POND, NMR, CQT, CGPT, CUDOS, LMWR, ZIG, PHB, HEART, AITECH, FORT
CoinMarketCap Control Group	AAVE, ADA, ALGO, APT, AR, ARB, ATOM, AVAX, BEAM, BGB, BNB, BONK, BTC, CHZ, CORE, CRO, DOGE, DOT, ENA, ETH, FIL, FLOKI, FLOW, FTM, GALA, HBAR, ICP, IMX, JUP, KAS, LDO, LEO, LTC, MATIC, MKR, MNT, NEO, OKB, ONDO, OP, PENDLE, PEPE, QNT, RUNE, SEI, SHIB, SOL, STX, SUI, TON, TRON, UNI, VET, W, WIF, XLM, XMR, XRP
CoinMarketCap Artificial Intelligence and Big Data Group	ABT, AGIX, AGI, AGRS, AIOZ, AITECH, AKT, ALEPH, ALI, ARKM, CQT, CTXC, CUDOS, DATA, DIA, DKA, DMTR, FETCH, FLUX, FORT, GLM, GPU, GRT, HOOK, INJ, IQ, KDA, LAT, LMWR, NEAR, NFP, NMR, NUM, OASIS, OCEAN, ORAI, OZONE, PAAL, PHA, PHB, POND, PRIME, RENDER, RLC, RSS3, SDAO, SIDUS, SURE, TAO, TFUEL, THETA, TOKEN, TRAC, VAI, VR, VRA, ZIG, HAI, MINA
S&P Ex-Mega Capitalization Group	BTC, ETH
S&P Ex-Large Capitalization Group	BTC, ETH, BNB, SOL, TON, ADA, SHIB, AVAX, TRON, BCH

provided in in the Appendix. The corresponding tickers to each subgroup are shown in Table 4, for which the full cryptocurrency names of all tickers will also be provided in the Appendix.

3.2 Normality Plots

The normality plots exhibit bell-shaped curves, which may suggest the presence of more extreme values compared to a normally distributed data source. This is in line with Ante & Demir (2023); Ford (2015) and corresponding current and past research on cryptocurrency price movements. The plots, showcasing the presence or exclusion of normality are represented in Figures 1 and 2. This does not apply to the respective indexes, for which the returns and residuals are normally distributed.

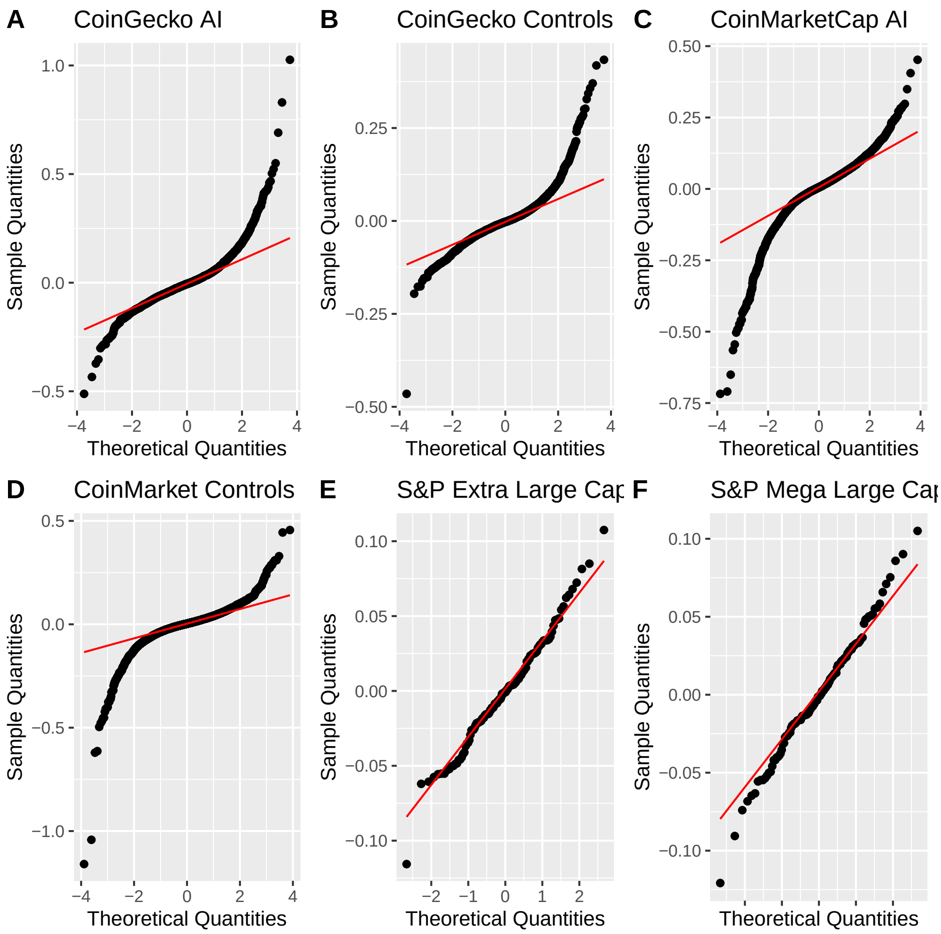

Figure 1: QQ-Normality Plots of All Data Panels

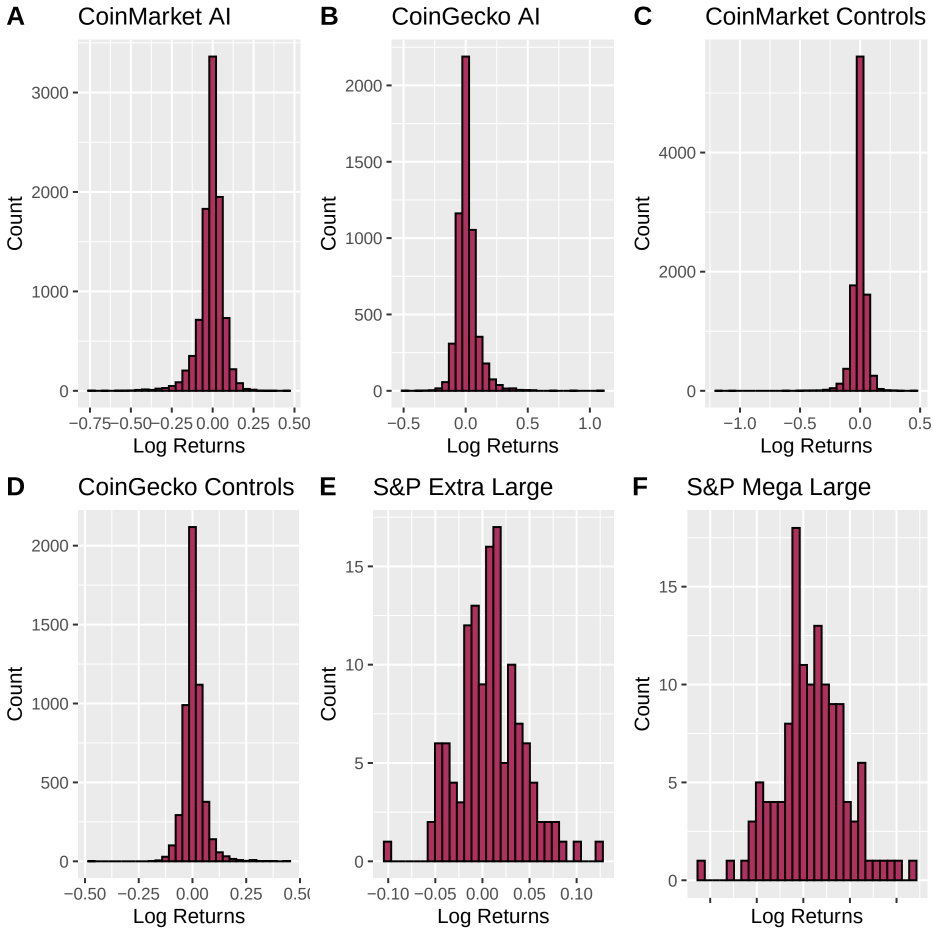

Figure 2: Histograms of All Logarithmic Returns

Figure 2 shows the skewness of the logarithmic returns in panels A, B, and C, whereas the indexes in panel D, E and F are relatively normally distributed.

3.3 Bullish and Bearish Markets

Market conditions provide flexibility in interpreting causal effects that lead to abnormal returns. Specifically, bullish and bearish markets have played a crucial role in influencing price movements in the cryptocurrency space, both historically and currently. Therefore, adding to the literature and to the controls of this research, bull and bear market conditions will be taken into account for the panel data, in which, again, 1 implies a coin confined within bullish market conditions and where 0 defines a period of relative bearish calmness and less volatility of the market. A bull-market is typically referred to as a sustained period of upward price movements of cryptocurrencies, led by Bitcoin respectively (Bouri et al., 2018). During these bullish markets, investors' confidence rises, driving prices even more upwards, resulting in all-time-high prices and increased attention from both institutional and private investors alike, sparking additional interest. This self-sustained state is not ever-lasting, meaning that prices and cryptocurrency movements consolidate at some point in time (Jain, 2024). Following this revelation, bullish market conditions are expected for periods after an all-time-high price record of Bitcoin, which happened in March 2024. Therefore, the data will be accounted for such that periods before March 2024 belong to bearish markets, whereas periods after the price ceiling of Bitcoin marks the respective start of bullish market conditions in this research. This approach is not without limitations and assumptions, however, this was deemed to be best for interpretability purposes. Additionally, Zhang et al. (2020) found that during bear markets, Bitcoin and similarly sized coins experienced increased liquidity. This suggests that positive price fluctuations may be driven by increased demand and investor confidence, which in turn reduce liquidity. As demand rises, investors are more willing to buy, driving prices up. This dynamic implies that lower liquidity, resulting from heightened demand, positively influences price movements. Thus, utilizing the all-time-high date as a threshold was therefore deemed feasible. The specific date for the all-time-high was March 14th, 2024, reaching a price of \$ $73,750.07$, respectively.

4 Methods

4.1 Potential Outcomes Framework

In this paper, the potential outcomes framework (POF) will be used for all methods, popularized by Rubin (1974); O'Neill et al. (2016). This framework consists of several elements which will be highlighted in this section. First, suppose $i = 1, ..., n$ crypto cohorts, and T time periods where T_0 is pre-treatment and T_1 to T_n are post-treatment. The pre-treatment window is comprised of returns ranging from October 6th 2023 to December 6th 2024. This window was chosen after careful deliberation based on Saggu & Ante (2023) and in accordance with event study methodology for the calculation of abnormal returns, in which windows outside of 12 months were considered unlikely, although no specific time trend consensus was agreed upon (Kothari & Warner, 2007; McWilliams et al., 1999). Likewise, the post-treatment window was chosen to be an extended period of 6 months rather than a more commonly used window of approximately 2 months (Saggu & Ante, 2023). The reason for this is that this study focuses on longer-horizon returns, which will be enabled and researched by scrutinizing a prolonged period of time after the respective release. The potential outcomes in the presence and absence of treatment are denoted by Y_{it}^1 and Y_{it}^0, where D_{it} represents whether a unit i was treated or not at time t. Based on these elements, a generic model can be operationalized when no treatment is present:

$$Y_{it}^0 = \mathbf{X_{it}^T}\beta + \lambda_t\mu_i + \delta_t + \epsilon_{it} \ , \tag{1}$$

followed by the following equation, regarding the presence of treatment:

$$Y_{it}^1 = \mathbf{X_{it}^T}\beta + \lambda_t\mu_i, + \delta_t + \tau + \epsilon_{it} \ , \tag{2}$$

where $\mathbf{X_{it}^T}$ represents a vector of time-varying features, μ_i represents time-invariant unobserved characteristics from which the effects denoted by λ_t differ across time but no across units. Furthermore, δ_t are common time effects and ϵ_{it} represent timely idiosyncratic potential shocks.

Next, by assuming the treatment is of influence, only in periods after treatment, the observed outcome is denoted as:

$$Y_{it} = D_{it}Y_{it}^1 + (1 - D_{it})Y_{it}^0 \ , \tag{3}$$

this outcome Y_{it} represents the price movements of the cohort of treated AI-assets and where D_{it} represents a dummy variable indicating placement in the treatment group denoted by 1 and 0 otherwise. Building upon Equation 1 and Equation 2, an estimand of the treatment can then be described as follows:

$$\tau = E[Y_{it}^1 - Y_{it}^0 | D_{it} = 1] \ , \tag{4}$$

Equation 4 holds, as time-common effects, denoted by δ_t are negated based on Equation 1 and Equation 2. Then, only τ remains as single effect. As a final note, if μ_i is imbalanced and $\lambda_t \neq 0$ (i.e. both treatment assignment and outcome are influenced), then μ_t is a potential confounder, possibly leading to biasedness in the average treatment effect on the treated (ATT). The methods used in this paper will make use of this framework, and therefore rely on the underlying assumptions presumed in accordance with this generalized model. SDID however, will not rely on the assumption of parallel trends (unlike DID), where the outcome of the control is independent of treatment assignment. The SC Method will also be based on re-weighing of unexposed cohorts to closely resemble the treatment group prior to being treated. The LDV approach conversely, will rely on the assumption of independence (conditional on past outcomes), which in essence implies that units in pre-treatment exhib similar outcomes, and therefore have similar potential post-treatment outcomes given the absence of treatment. This happens after conditioning on observed covariates X_{it} (O'Neill et al., 2016). The specifics of these respective methods will be expanded upon in separate sections.

4.2 (Abnormal) Returns

$$R_{it} = \log\left(\frac{P_{it}}{P_{i,t-1}}\right) , \tag{5}$$

where:

- P_{it} denotes the price of an AI-asset i at day t,
- $P_{i,t-1}$ denotes the price of an AI-asset i at day $t-1$.

Returns will be at the root of uncovering potential causal effects of Gemini's release and AI-asset performance, especially in the final SDID model equation (Saggu & Ante, 2023; Arkhangelsky et al., 2021) and the LDV approach. Therefore, the returns are visualized in Equation 5. Log returns are estimated since logarithmic values are typically more inclined to be normally distributed. Additional reasons for choosing this are the cumulative properties of logarithmic returns, which can be aggregated and extended to represent a longer period (Strong, 1992). However, Hudson & Gregoriou (2015) emphasize the exertion of caution as higher variance of the set of returns results in lower expected returns, ceteris paribus. The use of logarithmic transformation or alternative transformations will be assessed based on the nature of the data.

4.3 Difference-In-Difference (DID)

The Difference-In-Difference (DID) method will serve as a benchmark but will not be applicable if the assumption of parallel trends cannot be met or if independence from past outcomes does not hold. In the case of this research, where multiple unit and time periods are considered for evaluation, the following baseline model is then estimated based on a two-way fixed effects regression: (Jones & Rice, 2011):

$$Y_{it} = \mathbf{X_{it}}\beta + \lambda\mu_i + \delta_t + \tau D_i + \epsilon_{it}, \tag{6}$$

where:

- Y_{it}: represents the price of cohort i at time t,
- $\mathbf{X_{it}}$: represent a vector of covariates for cohort i at time t,
- β: the vector of coefficients for $\mathbf{X_{it}}$,
- μ_i: representing cohort fixed effects, capturing time-invariant charateristics of cohort i,
- λ: the coefficients for the cohort fixed effects,

- δ_t: representing time-fixed effects, displaying factors that affect cohorts similarly, but differ over time,
- τ: the variable of interest, denoting the ATT,
- D_i: the indicator treatment, which equals 1 if cohort i is in the treatment group and 0 otherwise,
- ϵ_{it}: and the error term, capturing latent factors influencing the outcome variable Y_{it}.

In this formula, the unobserved confounders denoted by μ_i and their respective effects (λ) do not vary over time, effecitvely satisfying the parallel trends assumption imposed by DID-methodology (Keele & Kelly, 2006). The same unobserved confounders are in turn controlled by the inclusion of dummy variables (D_i) for each cohort (*cohort fixed effects*). Moreover, the aggregate shocks denoted by δ_t are controlled for by introducing dummy variables for each time t (*time fixed effects*). Lastly, the estimate for the treatment denoted by τ can then be interpreted as the ATT for the post-treatment time window.

4.4 Synthetic Differene-in-Difference (SDID)

As introduced by Kattenberg et al. (2023), DID in its most primal form, consists of one treatment, two groups and two periods. This exact setting will also be applicable to this research design. This relies on the assumption that by taking the difference between the trend of outcomes in the treatment and control group, the average effect can be estimated as such. The baseline model (DID) assumes parallel trends in pre-treatment of the groups in order for valid results regarding ATT Saggu & Ante (2023) found mixed results when testing for homogeneity in trends pre-treatment, which issues caution when employing this method. Therefore, as employed by Saggu & Ante (2023), this paper will utilize a similar measure of finding ATT by utilizing the framework popularized by Arkhangelsky et al. (2021); Clarke et al. (2023). This framework is commonly referred to as the Synthetic Difference-in-Difference (SDID) model in which a binary assignment of treatment (1 = the launch of Gemini, 0 = no-event treatment) on an outcome variable, which in the context of this research comes down to the (potential) causal effects on price movements of AI-assets. This approach is free of the assumption of parallel trends, hence it renders operational even if this assumption is not satisfied. The ATTs will be estimated as a two-way fixed effects panel regression as seen in the following formula:

$$(\hat{\tau}^{\text{sdid}}, \hat{\mu}, \hat{\alpha}, \hat{\beta}) = \underset{\tau,\mu,\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} (R_{it} - \mu - \alpha_i - \beta_i - \text{GEMINI}_{it}\tau)^2 \hat{\omega}_i^{\text{sdid}} \hat{\rho}_t^{\text{sdid}} \right\} , \tag{7}$$

the estimand (quantificiation of the treatment) (Pohl et al., 2021) is generated based on the optimal unit output ($\hat{\omega_i}$) and time weights, represented by ($\hat{\rho_t}$). Exclusion of the optimal output and time weights parameter yields a traditional DID framework (Saggu & Ante, 2023). Thus, weight and time optimization ensure comparison ability of groups. The weighing is adjusted such that greater matches of pre and post treatment receive higher weights. This happens in order to justify constant difference between control groups' post-treatment average and pre-treatment averages of all controls. The overarching goal here is to achieve balance among pre and post treatment trends, which again are unaccounted for and not assumed in this SDID approach. First, the derivation of $\hat{\omega_i}^{sdid}$:

$$(\hat{\omega}_0, \hat{\omega}_i^{sdid}) = \underset{\hat{\omega} \in \mathbb{R}, \omega \in \Omega}{\operatorname{argmin}} \ \ell_{unit}(\omega_0, \omega) \ , \tag{8}$$

where:

- $\ell_{\text{unit}}(\omega_0, \omega) = \sum_{t=1}^{T_{\text{pre}}} \left(\omega_0 + \sum_{i=1}^{N_{\text{co}}} \omega_0 R_{it} - \frac{1}{N_{\text{tr}}} \sum_{i=N_{\text{co}+1}}^{N} R_{it} \right)^2 + \zeta^2 T_{\text{pre}} ||\omega||_2^2$,
- $\Omega = \{ w \in \mathbb{R}_+^N : \sum_{i=1}^{N_{\text{co}}} w_i = 1, \ w_i = N_{\text{tr}}^{-1} \ \forall \ i = N_{\text{co}} + 1, ..., N \}$

Second, the time weights, $\hat{\lambda_t}^{sdid}$, are not dependent on a regularization parameter, thus observations may be correlated. These time weights are implemented as follows:

$$(\hat{\lambda_0}, \hat{y_i}^{sdid}) = \underset{\hat{\lambda_0} \in \mathbb{R}, \lambda \in \Lambda}{\operatorname{argmin}} \ \ell_{time}(\lambda_0, \lambda) \ , \tag{9}$$

where:

- $\ell_{\text{time}}(\lambda_0, \lambda) = \sum_{i=1}^{N_{\text{co}}} \left(\lambda_0 + \sum_{t=1}^{T_{\text{pre}}} \lambda_0 R_{it} - \frac{1}{N_{\text{tr}}} \sum_{i=N_{co+1}}^{N} R_{it} \right)^2$,
- $\Lambda = \{ \lambda \in \mathbb{R}_+^T : \sum_{i=1}^{T_{\text{pre}}} \lambda_i = 1, \ \lambda_i = T_{\text{post}}^{-1} \ \forall \ i = T_{\text{pre}} + 1, ..., T \}$

For both of the weights, $\mathbb{R}_+$ denotes the real positive line, whereas ζ only represents a regularization parameter matching the magnitude of a standard outcome change of one period for unexposed units, multiplied using a scalar in Equation 8 (Saggu & Ante, 2023). ω_0, acting as an intercept, allows for more flexibility in the determination of weights, relaxing the prerequisites for pre and post treatments trends to be parallel. This is enabled through fixed-effects, denoted by α_i, which absorb constant differences in alternate units. An additional penalty term, introduced by Doudchenko & Imbens (2016) and corroborated by Saggu & Ante (2023) allow for further dispersion possibilities and greater uniqueness of corresponding weights.

The SDID approach and corresponding framework thus provide a flexible and robust approach for evaluating AI-assets ATT by incorporating temporal aggregate factors and unit-specific factors (Saggu & Ante, 2023; Clarke et al., 2023).

The final SDID model can be justified as follows, accounting for exogenous and time-varying covariates (e.g. market capitalization and liquidity):

$$R_{it}^{res} = R_{it} - Cap_{it}\hat{\beta}_1 - Vol_{it}\hat{\beta}_2 \ , \tag{10}$$

where $\hat{\beta}$ was obtained from regressing R_{it} on Cap_{it} and Vol_{it}. In unison, the hatted values represent the fitted (estimated) values of the regression. In the fully realized models, more independent variables will be implemented, where this formulation shown in Equation 10 represents a baseline, simplified model (Saggu & Ante, 2023).

4.5 Lagged Dependent Variable (LDV)

$$Y_{it} = \mathbf{X_{it}}\beta + \sum_{k=1}^{T_0} \theta_k Y_{i,t=k} + \tau D_i + \epsilon_{it} \quad \forall t > T_0 \ , \tag{11}$$

where:

- Y_{it}: represents the outcome variable for cohort i at time t,
- $\mathbf{X_{it}}\beta$: represents a vector of explanatory variables for cohort i at time t, containing coefficients β,
- $\sum_{k=1}^{T_0} \theta_k Y_{i,t=k}$: This term represents the summation of lagged dependent variables for cohort i at time t, where $Y_{i,t-k}$ is the value of the dependent variable k periods before time t. The coefficients θ_k represent the impact of each lagged value on the current outcome,
- τD_i: represents the treatment effect for cohort i at time t, where τ varies across cohorts denoted by dummy variable D,
- ϵ_{it}: is the error term for for cohort i at time t.

Estimation of the model defined in Equation 11 is done by using ordinary least squares (OLS) in post-treatment periods exclusively. This implies that D_i denotes the factual and counterfactual outcome of the log price fluctuations of AI and non AI-assets. Therefore, the time period considered is solely reliant on post-treatment outcomes, conditional on equal pre-treatment outcomes. Therefore, no subscripts will be used for the estimation of the ATT denoted by τ. This model is inherently constrained by its assumptions and reliance on post-treatment outcome estimation, which prevents the estimation of time-varying treatment effects. However, it allows for the estimation of unit-varying (i.e., cohort-varying) effects. (Keele & Kelly, 2006). Though this serves as a limitation to the study, this method will be helpful in establishing a benchmark of comparison for more sophisticated models and methods. Thus, it should be noted that the inclusion of past outcomes here does not create a fully dynamic model since we only condition on a fixed vector of pre-treatment outcomes Y_{it}, and not on any post-treatment lagged outcomes. This is because in a dynamic model, the vector of past outcomes must be relative to the period being considered. However, in this case, the vector of past outcomes is equal regardless of the period considered. If the equation then qualifies as identifying the true data generating process, then the assumption of independence conditional on past outcomes is satisfied. Any $t > T_0$ implies that for all $T > 0$ a causal effect can be observed. $D_i = 0$, then represents the outcome if no treatment occurs. τ is then able to quantify the difference between the treated effect and the counterfactual, namely: the ATT (O'Neill et al., 2016). One drawback of this approach is its reliance on a vector composed solely of pre-treatment outcomes, lacking the inclusion of lagged values post-treatments, as aforementioned. Consequently, this LDV method demonstrates optimal performance when applied to datasets with a reasonably and notably extensive pre-treatment period. And as a final note, the LDV approach has raised concern in light of recent literature, as the inclusion of lagged dependent variables as explanatory variables possibly lead to bias whenever idiosyncratic shocks are serially correlated, however this has not been proven in the context of ATT-estimation, and this will thus not be considered as a limitation for this study.

4.6 Synthetic Control (SC)

Finally, the Synthetic Control (SC) method aims to construct a counterfactual treatment-free outcome for the treated unit by appropriately weighting the outcomes of the control units. These weights are selected to ensure that the treated unit and the SC exhibit similar outcomes and covariates during the extensive pre-treatment period. Similar to the LDV approach, the SC method depends on the assumption of independence conditional on past outcomes (Angrist & Pischke, 2009). However, it employs a semi parametric approach to adjust for pre-treatment outcomes and covariates by re-weighting treated observations. In essence, a synthetic control for a single treated unit is created by finding the vector of weights W^* that minimizes $(X_1 - X_0W)'V(X_1 - X_0W)$, subject to the constraints that the weights in W are positive and aggregate to 1. Here, X_1 and X_0 represent the pre-treatment outcomes and covariates for the treated unit and control units, respectively, while V captures the relative importance of these variables as predictors of the outcome of interest (O'Neill et al., 2016). The outcome of interest in this case, is the ATT on the price fluctuations of cryptocurrency AI-assets.

Based on the nature of this study, containing multiple treated units, one must reweigh the designated disaggregated control units into a single and aggregated synthetic control unit. Therefore, considering multiple treated units, $\mathbf{X_1}$ represents the vector of covariates that is averaged across the treatment group. Then, the optimized set of weights is then used to comprise a synthetic control group, mimicking the average pre-treatment outcomes $\bar{Y}_{it}$ and observed covariates $\bar{X}_{it}$ of the treated units (O'Neill et al., 2016). Subsequently, this can be explained by the following formulas:

$$\sum_{j \in \text{Control}} w_j Y_{jt} = \bar{Y}_{it}, \quad \forall t \leq T_0, \tag{12}$$

where:

- $\sum_{j \in \text{Control}} w_j Y_{jt}$: represents the weighted combination of outcomes Y_{jt} from the control group,
- w_j represent the weights assigned to each control unit j, where the optimal weights are designated based on the minimization of distance between control and treatment group, pre-treatment,
- Y_{jt} determines the outcome of control unit j at time t,
- $\bar{Y}_{it}$ represents the mean outcome of treated unit i at time t,
- $\forall t \leq T_0$: ensuring that the parallel trends assumption holds for characteristics depicting both treatment and control groups.

$$\sum_{j \in \text{Control}} w_j \mathbf{X_{jt}} = \mathbf{\bar{X}_{it}}, \quad \forall t \leq T_0, \tag{13}$$

where:

- $\sum_{j \in \text{Control}} w_j \mathbf{X_{jt}}$: represents the weighted combination of covariates $\mathbf{X_{jt}}$ from the control group,

- w_j: represent the weights assigned to each control unit j, where the optimal weights are designated based on the minimization of distance between control and treatment group, pre-treatment,
- $\mathbf{X_{jt}}$: defines the vector of covariates for control unit j at time t,
- $\bar{\mathbf{X}}_{\mathbf{it}}$: represents the mean vector of covariates for treated unit i at time t,
- $\forall t \leq T_0$: ensuring that the covariates are balanced for all time periods up to the treatment period, maintaining comparability standards between controlled units and treated units.

The discussed methods will be used to establish an analysis into the causal effects of Gemini's release on cryptocurrency prices, with the findings presented in the Results and Analysis section.

5 Results & Analysis

5.1 Lagged Dependent Variable (LDV) Results

As mentioned before, the LDV approach will be estimated using an Ordinary Least Squares (OLS) approach using the lagged values of the logarithmic returns of cryptocurrency tokens. This approach was deemed feasible in light of the proxying of latent variables and in light of ATT-theory, considering multiple treatment and control groups. The results are presented in Table 5. The LDV utilizes an approach where post-treatment outcomes are regressed on the treatment indicator D and the corresponding treatment covariates.

Creating lagged dependent variables requires significant and deliberate consideration for the data available. Therefore, in this study, it was determined that defining pre-treatment and post-treatment periods were essential. An additional variable was then added to account for exposure to the treatment since both the control group and the treatment group were required to reside in the same dataset. If kept separate, singularities would arise for the treatment, and therefore, results would be non-applicable or indeterminate. Hence, after controlling for subsetting the data for post-treatment outcomes, exposure to the treatment was analyzed using the aforementioned treatment indicator (denoted by *Gemini*) , and corresponding covariates (*Market Capitalization, Date, Bull Market and Lagged Log Returns.* Taking pre-treatment outcomes was deemed unfeasible due to insignificance of results. For the control indexes, a similar approach will be taken. However, exposure to the treatment will not be considered, and only the OLS estimates and results will be presented as a baseline reference for circumstances where the treatment was not applied.

Table 5: Lagged Dependent Variable Regression Results CoinGecko and CoinMarketCap Control and Treatment Groups

	Dependent variable:	
	Log Returns	
	(1)	(2)
Date	0.000***	-0.000***
	(0.000)	(0.000)
Market Capitalization	-0.000	0.000
	(0.000)	(0.000)
Bull	-0.020***	0.016***
	(0.003)	(0.002)
Lagged Log Returns	0.303***	0.226***
	(0.011)	(0.008)
Gemini	0.003**	-0.001
	(0.002)	(0.002)
Intercept	-3.571***	2.849***
	(0.541)	(0.414)
Observations	7,199	13,199
R^2	0.108	0.059
Adjusted R^2	0.107	0.059

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$; (1) denotes the results for the CoinGecko Cohort and (2) denotes the results for the CoinMarketCap Cohort.

Table 6: Regression Results S&P Indexes Large and Mega Capitalization for Cryptocurrencies

	Dependent variable:
	Log Index Returns
Date	-0.000
	(0.000)
Capitalization Index	0.000
	(0.000)
Mega	0.012
	(0.011)
Bull	0.001
	(0.004)
Lagged Log Returns	0.371***
	(0.058)
Intercept	1.227
	(1.453)
Observations	259
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5 show the LDV results for both CG and CMC, respectively. The models both exhibit characteristics of a log-linear and log-log model depending on the features assessed. For CG, the variables of significance on the 5% confidence level are *Date, Bull, Lagged Log Returns, Gemini and the Intercept.* Indicatively, this implies that bull market conditions exhibit a negative effect on the logarithmic returns of AI-related assets throughout the study period. More precisely, pertaining to bull market conditions, this yields a decrease of -2% approximately in returns for the CG cohort. The opposite holds true for the AI and Big Data cohort, in which the consequences of having bull market conditions affect the log returns positively, significant on the 1% confidence level. Again, more precisely, an expected increase of 1.6% is expected in log price fluctuations in the CMC cohort. Furthermore, noticeably, AI-coins retrieved from CG exhibit positive influences of the Gemini release on its subsequent price movements in the given window. More precisely, the release of Gemini resulted in having an approximate 0.30% increase in prices, following the technology's release, significant on the 5% confidence level. Remarkably, this does not hold true for the larger cohort retrieved from CMC. Both intercepts or constants are significant, however its sole interpretation serves no purpose as the features within the model are inherently non-zero and non-negative. Lastly, the lagged returns are significant on the 1% level, marking an increase of 0.303% in the current log returns if the previous log returns with a lag of 1 are equal to 1%. The same can be said for the AI-coins and big data cohort (CMC), as these exhibit similar behavior, however differing in magnitude. The CMC cohort, exhibits an increase of 0.226% whenever the previous period prices increase by 1%, respectively. The reported R^2 and Adjusted R^2 are deemed significant according to a threshold set by Ozili (2023), as this recent research paper suggests that an R^2 of 0.1 or greater is suitable for circumstances in which the majority of the predictors within the confines of a regression are significant. Thus, the CG LDV results satisfy this assumption.

Next, the indexes' results are reported in Table 6. Indexes are traditionally believed to be stable and therefore, the returns are also assumed to holds table over time. This enables this research to control for bias in the selection of coins across the board. For instance, Stosic et al. (2018) argue that the minimum spanning tree of crypto tokens cross correlations show the existence of distinct community structures that are co-moving and stable over time.

As for the interpretation of the coefficients, the only noticeable significant feature remains the lagged log returns of the indexes. This implies that 1-periods lag exhibit significant influence over the log returns in subsequent periods. Thus, an increase of 1% in the period prior to a price measure, is expected to gain a 0.371% increase in the next period. The results in Table 6 are omitting the release of Gemini as a potential confounder and therefore, these results serve as a baseline for the changes in log returns for larger crypto cohorts. Next, a baseline DID and SDID approach will be taken to compare the results of the LDV OLS to more sophisticated model approaches.

5.1.1 Baseline Difference-In-Difference (DID)

5.1.2 CoinGecko DID

Firstly, plotting the differences of logarithmic returns over time, on specific dates will showcase the fluctuations of returns over time and thus be insightful in scrutinizing the difference between AI and non-AI assets in terms of mean logarithmic returns. This is visualized in Figure 4. Generally, Figure 4 visualizes the log returns such that the variation and therefore the fluctuations in prices for the AI-Assets (in pink) are greater than those of the non-AI related assets.

Secondly, the parallel trends assumption must hold for the CG cohort. After testing for a difference in coefficients for both the control and treatment group, based on the interaction term, the reported p-value was 0.099 and thus not ensuring statistical significance in the difference between trends of both groups pre-treatment (Gemini Release). Therefore, the parallel trends assumption holds and this is also confirmed by the findings visualized in Figure 3. Although the returns are not equal in magnitude, the trends appear visually similar.

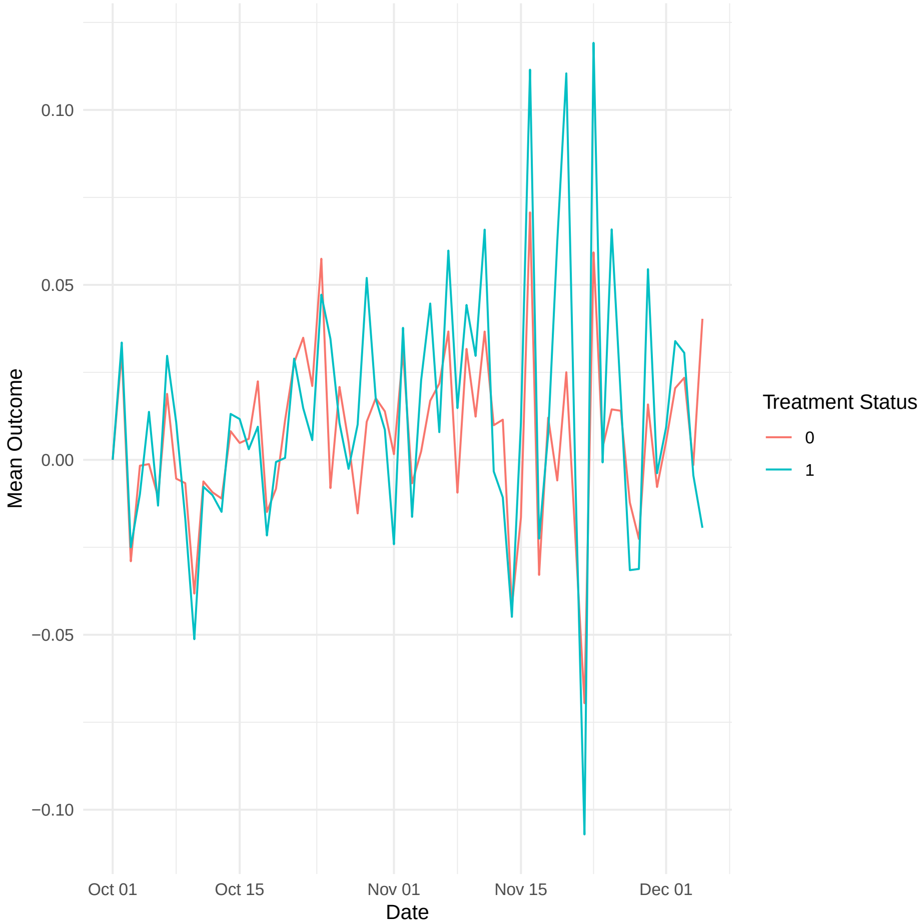

Figure 3: Pre-Treatment CoinGecko Cohort AI and Control Assets' Trends

Figure 4 represents the mean log returns for both AI-assets and non-AI assets for the CG cohort per day. One can observe that for positive returns (whenever the bar exceeds the horizontal axis) the AI-tokens outperform the non-AI assets. While the same applies for negative returns, the plot shows volatility of both types of assets, however also showing greater potential for AI-assets.

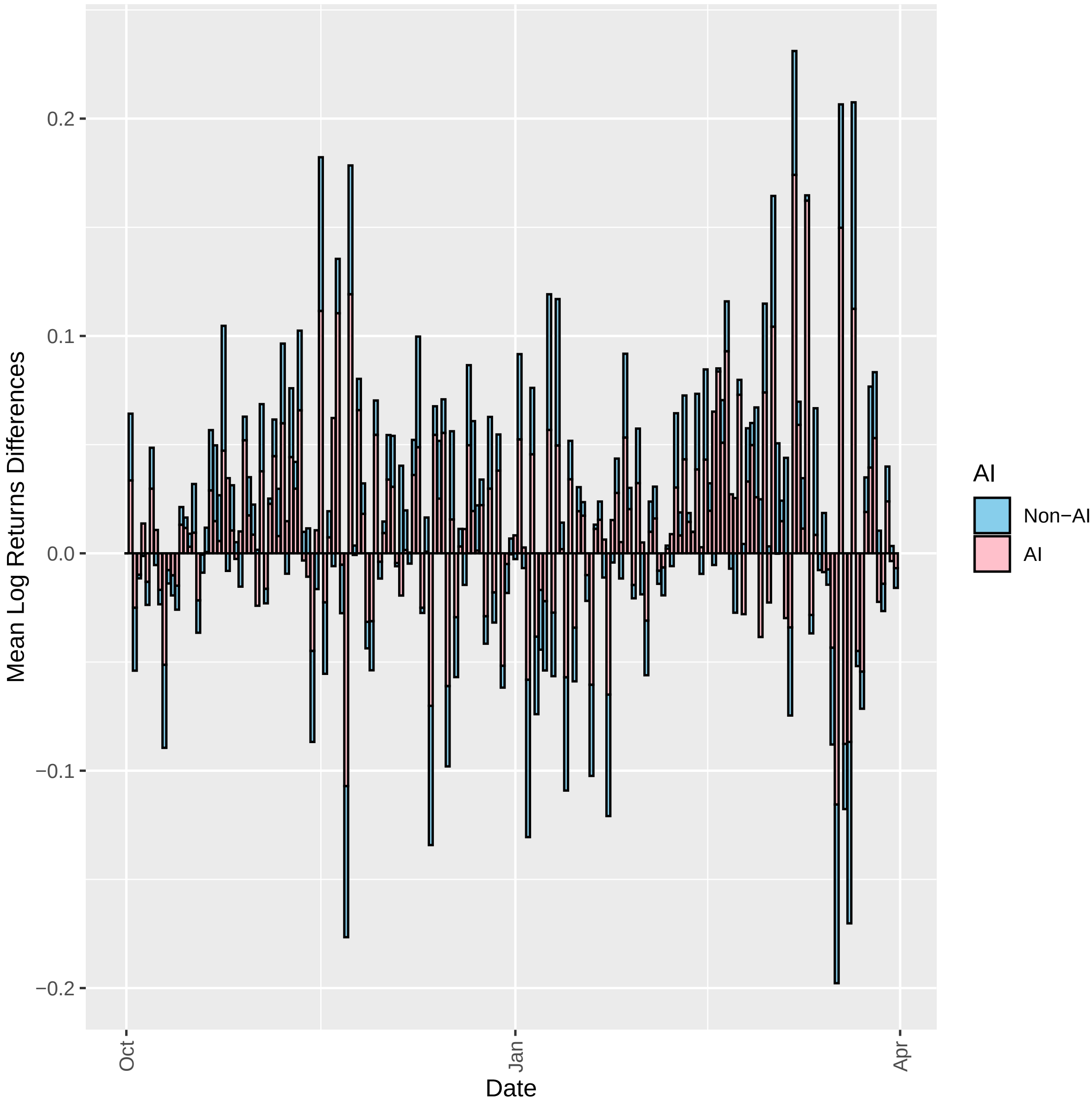

Figure 4: Mean Log Returns Differences per Date and Asset Nature for CoinGecko

Next, the baseline DID results will be visualized in Table 7 for the CG cohort consisting of both treatment and control group. The results show that the DID-estimator is non-significant at the 5% level. Interpretation of the coefficients is then realized as follows: *Gemini Exposure* represents the time before and after treatment, marked by Gemini's December 6th, 2023 release date. This effect is non-significant, but normally it would imply the mean difference had the release never occurred. Furthermore, *AI* represents the treatment group, more specifically the mean difference between the control and the treatment group, which in this case is statistically significant at the 5% level. The results imply a 0.73% increase in the log returns of the exposed cohort compared to the control cohort. This in turn implies that there is a significant difference between the control and treatment group, in favor of the treatment group. Next, the constant signifies the mean value of the outcome before the treatment and for the control group. Thus, a 0.005 or 0.50% increase would be expected for the control group, disregarding any treatment. Lastly, the DID-estimator denoted by *Gemini Exposure * AI* is non-significant at the 5% level, therefore no significant mean change in the log returns before and after Gemini's release for AI-tokens and non-AI tokens was recorded, according to the results.

Table 7: CoinGecko Difference-In-Difference Results

	Dependent variable:
	Log Returns
Gemini Exposure	0.0001
	(0.002)
AI	0.0073***
	(0.002)
Gemini Exposure * AI	-0.0025
	(0.003)
Constant	0.0056***
	(0.002)
Observations	10,980
Note:	*p<0.1; **p<0.05; ***p<0.01

5.1.3 CoinMarketCap DID

Following the CG cohort, the CMC larger cohort of AI and big-data tokens will be analyzed in a similar fashion. First, the parallel trends assumption must hold.

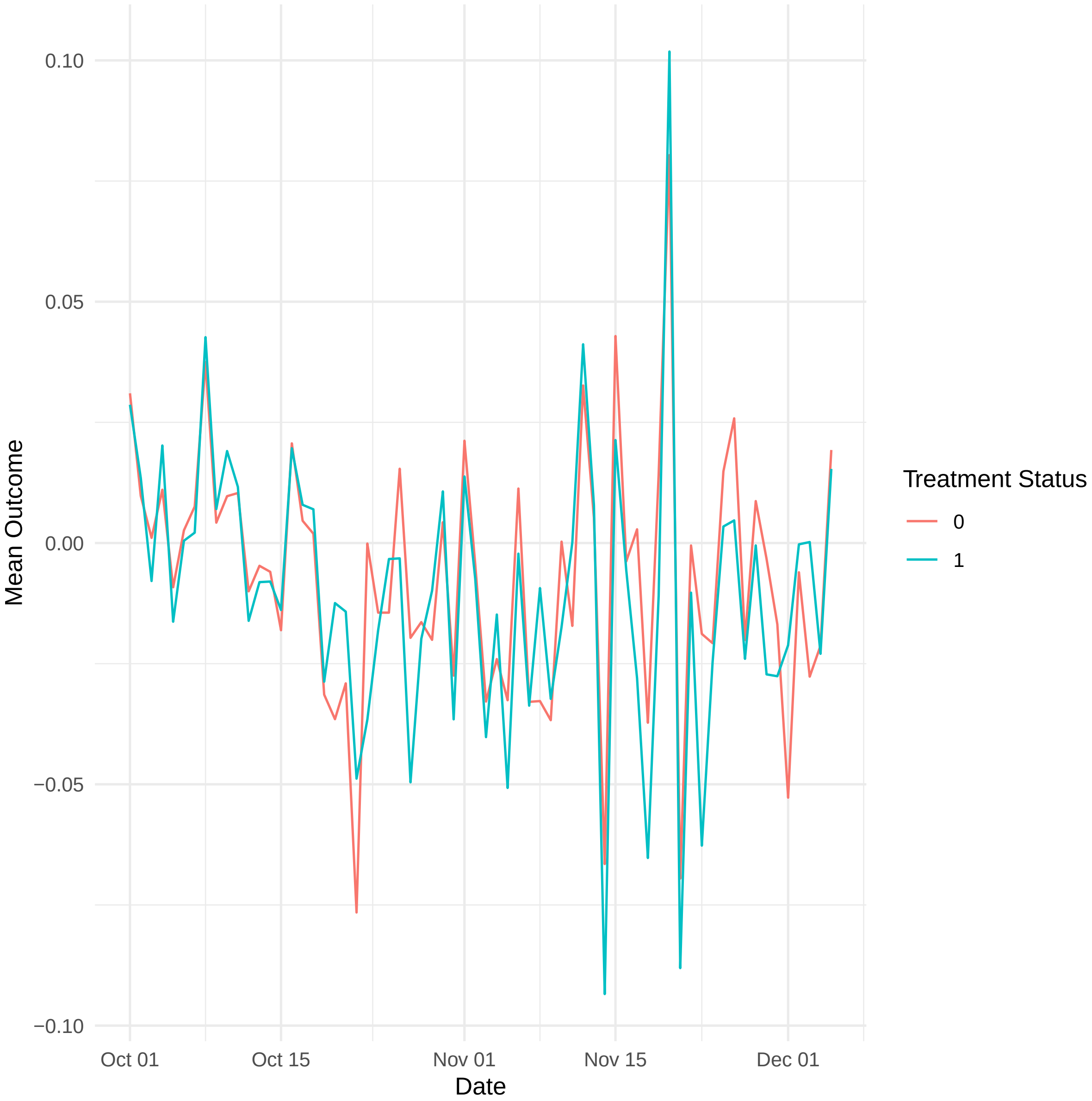

Figure 5: Pre-Treatment CoinMarketCap Cohort AI and Control Assets Trends

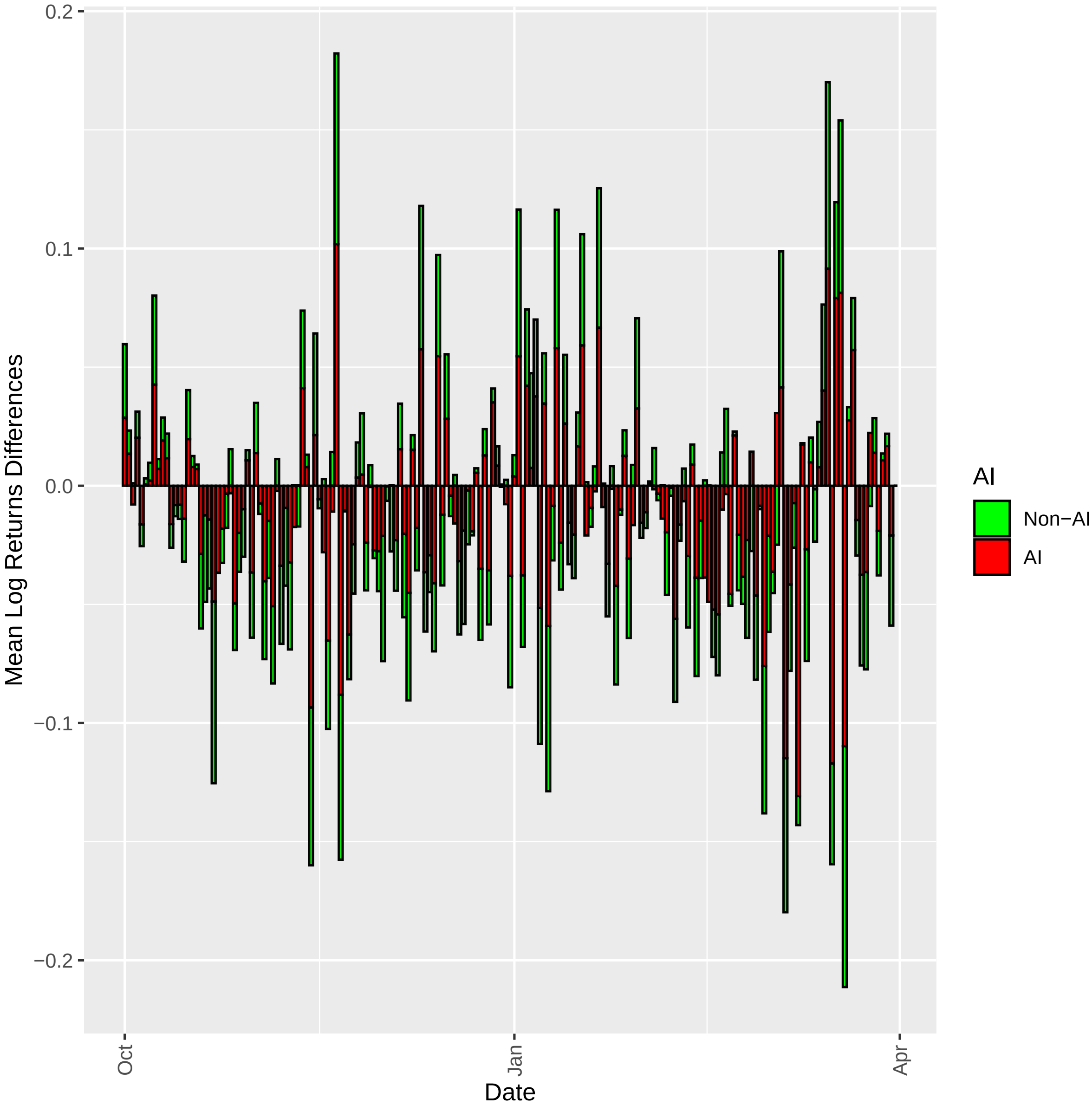

Figure 6: Mean Log Returns Differences per Date and Asset Nature for CMC

For Figure 5, the p-value for the difference in coefficients for both control and treatment group is 0.077, marking a non-significant relationship, hence parallel trends can be assumed for the CMC cohort. This is also supported by the visual evidence displayed in 5, where the overall trend is similar, except for a few spikes in late October and late November.

Furthermore, Figure 6 shows the average log returns for both AI and non-AI tokens over the predetermined research window, indicating no strong evidence for exceedingly high mean prices for either one of the cohorts over time. This is in contrast to the CG cohort and this may be due to the nature of the assets compared and the size of the datasets, respectively. Moreover, the returns per date are mostly negative, which is also in sharp contrast with the CG cohort. This may also be due to the bearish market conditions present back then.

As for the DID-regression results, according to Table 8, the DID-estimator is non-significant, alongside the treatment variable *Gemini Exposure*. *AI* and the intercept are once again significant at the 5% level, indicating a mean log returns of the control group before launch (intercept) and the mean difference of log returns pre and post treatment for the control group (AI). Both are negative, implying negative log returns prior to treatment for the control group, as well as a negative difference in log returns for the control group when comparing post and pre-treatment periods. Namely, being part of the treatment group yields an approximate decline in log prices of AI-assets of -0.30%. Whereas the control group, ceteris paribus, reports lower log returns on average of -0.70%.

Since both methods are inconclusive on the effects of the DID-regressors and generally non-imposing on the existence of an effect induced by Gemini's release, more sophisticated methods will be employed to unveil potential causal effects, similar to the release of ChatGPT and the research by Saggu & Ante (2023). These methods will be highlighted in the next section.

Table 8: CoinMarketCap Difference-In-Difference Results

	Dependent variable:
	Date
AI	-0.003**
	(0.002)
Gemini Exposure	0.001
	(0.001)
Gemini Exposure * AI	0.001
	(0.002)
Constant	-0.007***
	(0.001)
Observations	20,130
Note:	*p<0.1; **p<0.05; ***p<0.01

5.1.4 Synthetic Difference-In-Difference (SDID) and Synthetic Control (SC)

Relying on a more sophisticated and novel technique introduced by Arkhangelsky et al. (2019, 2021), one is potentially enabled to uncover results that differ from the ambiguous ones found by the LDV and the traditional baseline DID, in which only the LDV found a treatment effect of the release of Gemini, only for the CG cohort of pure AI-assets. Additional papers have shown the application, methods and integrities of the Synthetic Difference-In-Difference Method (SDID) Clarke et al. (2023). These will then be used as guidance in interpreting and conveying the results of this analysis. Alongside the SDID, Synthetic Control (SC) will be used to compare the results, in addition to reporting DID results using an alternate statistical package. For the former statistical technique, both CG and CMC cohorts have been examined by using *bootstrapping*, with replacement to simulate multiple scenarios. *Jackknife* resulted in having similar results, whereas *Placebo* was non-feasible due to multiple units being treated (Arkhangelsky et al., 2019; Torres-Reyna, 2015), whereas *Placebo* requires a single treated unit (Pierce et al., 1998). For a full description of these three options employed in the SDID algorithm, please refer to Arkhangelsky et al. (2019).

5.1.5 CoinGecko SDID, SC and DID

First, the results for the CG cohorts are represented in Table 9. These results report the alleged SC method. The estimates for SC and SDID are similar, but differ in magnitude, whereas the DID differs in sign. Further scrutiny of the results is necessary by means of Figure 7 and Figure 8. The first, shows that the trends are close to being parallel for all three methods. This was confirmed in the baseline DID and this holds once again, even though this assumption is non-essential for SDID. Furthermore, in Figure 8, one can see an improvement going from the DID dot plot to the SC dot plot, with an additional improvement of the SDID in the third panel. The conformation of the dots representing individuals coins visualizes the satisfaction of uniform pre-treatment trends, whereas the dots and the respective magnitudes represent contribution to the SDID-estimates. Therefore, the SDID and the SC both outperformed the baseline DID for the CG sample. Thus, based on the SDID analysis, the release of Gemini marks positive effect of 0.20% and a positive 0.6% effect by means of the SC. The improved fit of both the SC and SDID is partly caused by the weighing defined in Equation 7 in which the local fit is improved. The SC attempts to reweigh the unexposed AI-coins (the control group) so that the pre-treatment log returns of the controls match those of the AI-coins as closely as possible. SDID intends to perform a similar feat by reweighing the again unexposed control coins to obtain parallel time trends to the AI-coin cohort, however the aim is not necessarily trying to obtain identical trends. After, the SDID applies a DID on the reweighed panel. The plot presented in Figure 8 shows the difference between the adjusted weighted log return observation and the trend. Since no particular weights were attributed in the SDID analysis, the assumption of parallel trends is then satisfied. Lastly and noticeably, in both the SC and DID, larger market capitalized coins (both AI and non-AI related) show great influence in the point estimates of both methods. This discovery, in addition to the non-uniformness of the individual coin estimates provides evidence for non-parallel trends, allowing for the SDID to be the best performing model among the three. Thus, the SDID results will be used for the final interpretation.

Table 9: Estimation Output for Three Methods: Difference-In-Difference, Synthetic Control and Synthetic Difference-In-Difference for the CoinGecko AI-Cohort

	DID	SC	SDID
Estimate	-0.001	0.006	0.002
Standard Error	0.002	0.003	0.002

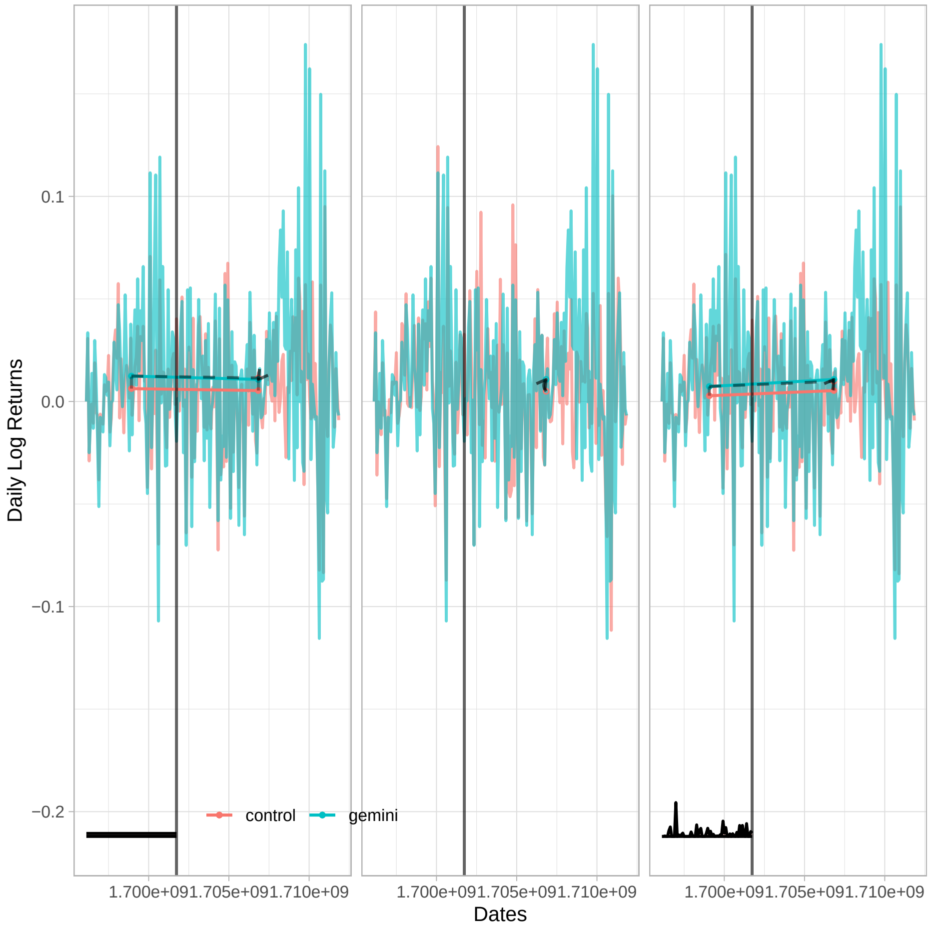

Figure 7: Trend Plot Representing the Trends of the 3 Methods: Difference-In-Difference, Synthetic Control and Synthetic Difference-In-Difference for the CoinGecko AI-cohort

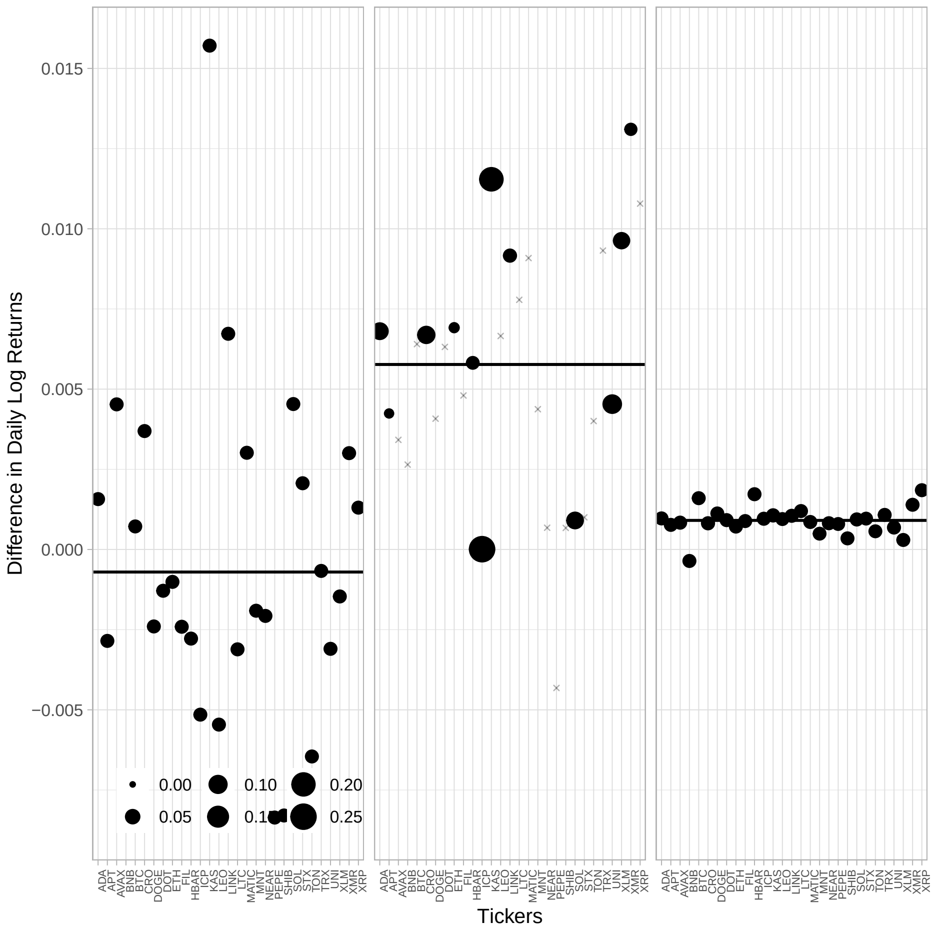

Figure 8: Dot Plot Representing the Performance of the 3 Methods: Difference-In-Difference, Synthetic Control and Synthetic Difference-In-Difference for the CoinGecko Cohort

5.1.6 CoinMarketCap SDID, SC and DID

Next, the CMC cohort will be examined on the same basis as the CG group. As for CMC, the trends are close to parallel; however, they display an upward trend for the DID and a downward trend for the SDID, as can be seen in Figure 9. Furthermore, Figure 10 displays the weights (denoted by larger dots in size) of each specific token in the CMC cohort and its corresponding effect on the log returns. One can tell that the SDID outperforms both the SC and DID, wihtout displaying the need to sustain irregular weights for observations. Therefore, based on the SDID approach, the results in Table 10 suggest that there was an overall positive effect of 0.20% on the Big Data and AI-cohort as per CMC standards. The results of the SDID for the CMC cohort coincide with the verdict provided by the analysis on the CG cohort, implying unity in the sign of the corresponding effect of Gemini's release on AI and Big-data related assets. As for treatment group constituted of pure AI-assets, both the LDV and SDID agree on the release having a positive effect. Lastly, since the past-trend independence cannot be assumed by means of the DID and SC, the SDID once again perseveres, allowing for the interpretation of Gemini's release having a significant and positive effect of approximately 0.20% on log returns of AI-assets, respectively.

Table 10: Estimation Output for Three Methods: Difference-In-Difference, Synthetic Control and Synthetic Difference-In-Difference for the CoinMarketCap Big-Data and AI-Cohort

	DID	SC	SDID
Estimate	0.001	-0.002	0.002
Standard Error	0.002	0.001	0.004

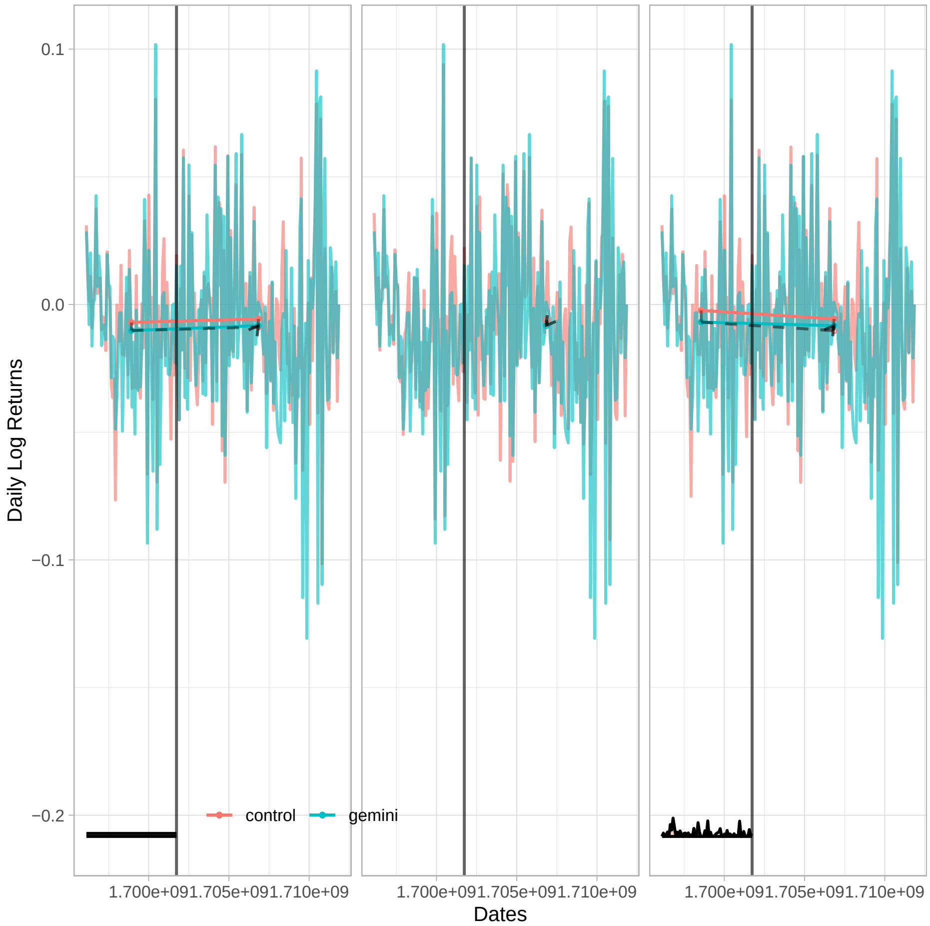

Figure 9: Trend Plot Representing the Performance of the 3 Methods: Difference-In-Difference, Synthetic Control and Synthetic Difference-In-Difference for the CoinMarketCap Cohort

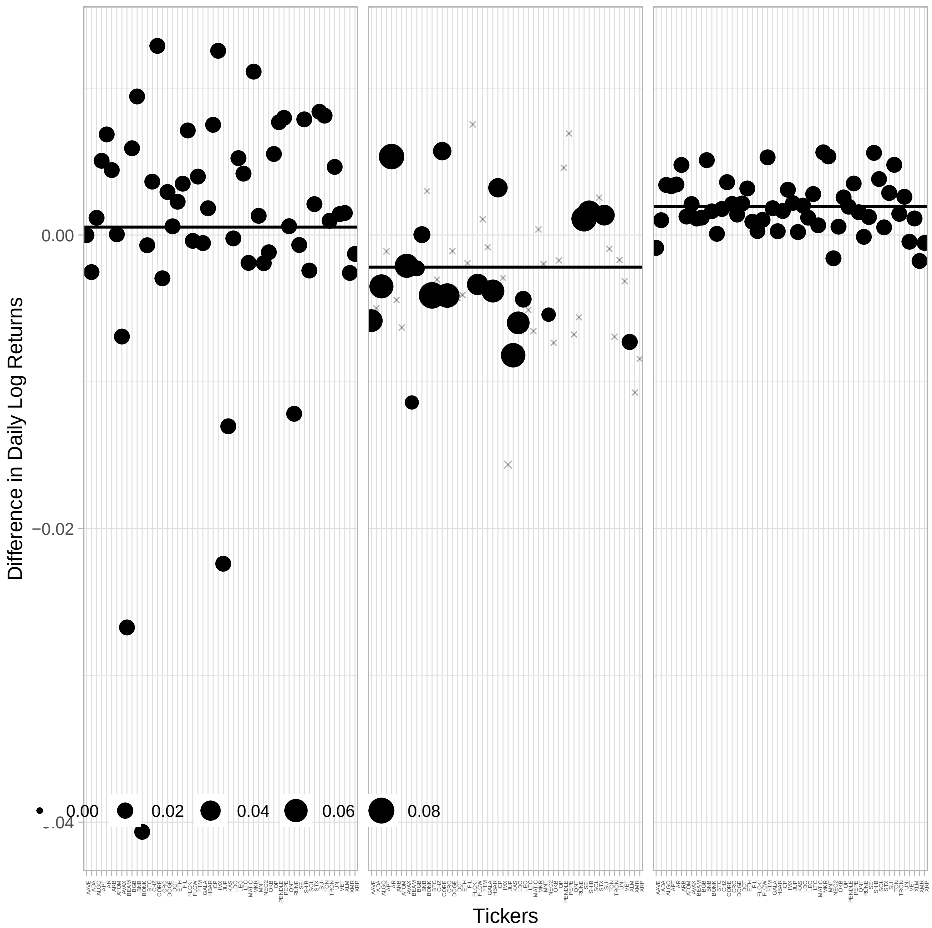

Figure 10: Dot Plot Representing the Performance of the 3 Methods: Difference-In-Difference, Synthetic Control and Synthetic Difference-In-Difference for the CoinMarketCap Cohort

6 Conclusion & Discussion

In this research, a plethora of statistical methods has been employed, both novel and traditional. However, the results sustained ambiguity compared to similar research conducted by Saggu & Ante (2023). The prospect of Gemini's release did not capture the same expectations as OpenAI's ChatGPT release back in November 2022. Nevertheless, tests on whether similar results would pertain for AI and broader Big Data cohorts remained evident, and therefore this research was carried out. Even though methods were not uniformly conclusive, a main trend has been established. The LDV concluded that the CG cohort exhibited positive significant returns for pure AI-assets, whereas the DID and SDID provided opposing results. Namely, the DID imposed both negative and positive results of Gemini's release on subsequent AI-asset prices, albeit being small in magnitude. The SDID, however, provided positive effects for both CG and CMC. Based on the results by these methods and the corresponding statistical tests for significance and robustness, it can be concluded that the DID and SC methods agree on the non-significance of the influence of Gemini on cryptocurrency price fluctuations, whereas the results by the LDV and SDID suggest positive significant effects (although the LDV suggests this for the CG cohort only). This duality serves as a sign of caution for interpretation; however, as the LDV and SDID served as the main methods in this research design, positive significant effects can be concluded based on both respective applications.

Even though positive effects were discovered, many limitations still apply to the structure of this research and its setting. For instance, the results may depend on the window chosen for this research (October 1st to March 31st) or on other limitations imposed by the methods used and the analysis following those. More specifically, different windows could potentially yield different results, affecting the inherent robustness. Additional limitations include, but are not limited to: variable selection, time window, market conditions, cryptocurrency selection both in the control and treatment group, statistical and empirical methods, and the use of a time-invariant treatment effect. To elaborate on the first, variable selection and its interplay may affect the outcomes by either being incorrectly chosen or unaccounted for, which may lead to biases and latent factors. Secondly, market conditions could significantly influence the extension to other periods due to differences in market and investor sentiments associated with bull market conditions. Thirdly, the selection of cryptocurrencies, albeit based on market capitalization, can impact the response to treatment, altering the conclusive effect of such treatment. Fourthly, as the statistical methods have their own set of limitations, these might not be able to fully grasp the intricate complexities of financial market assets, such as cryptocurrencies. Lastly, since the release of Gemini represents a single moment in time, future updates to the technology could enhance its capabilities, thereby improving its public image and increasing its usage. This, in turn, could have a more profound effect on cryptocurrency prices as the technology gains popularity. Employing a time-variant treatment would therefore be more suitable for assessing these potential effects in real-world scenarios. Additional limitations include the use of S&P indexes as controls, along with non-AI coins, which could lead to different conclusions regarding the results. Furthermore, by utilizing only 2 sources of data (namely: CG and CMC, respectively) the results are subject to the capabilities and measurements of these two sources. Finally, the depth of the data included in the analysis could have a significant and lasting influence on the final results, raising pressing considerations about the comprehensiveness and reliability of the dataset.

Given these concerns, future research could benefit from employing different control groups. This approach may yield varied results compared to this study and previous findings, enhancing the robustness and applicability of the conclusions. Additionally, focusing on more advanced and state-of-the-art technology releases, comparable to the highly anticipated release of OpenAI's ChatGPT, could provide deeper insights. This could involve analyzing multiple releases and conducting trend analyses to identify commonalities in exposure and publicity. Such an approach might include employing a treatment that varies both across different groups and over time, offering a nuanced perspective on the technology's impact.

Expanding upon this, future research may also employ alternative techniques to facilitate causal inference analysis. An interesting area of study could involve analyzing the impact of technology or information releases by examining the anticipation garnered both online and through word of mouth (WOM) instances. Although this research, given its limited capacity and resources, has delved into a potentially novel field where cryptocurrency tokens and their respective price movements may see increased scrutiny (particularly regarding the release of adjacent AI-technologies), which could then influence investor behavior, beliefs, and expectations.

Adding to these theoretical implications for future research, actionable insights enable the direct application of this work's insights in various areas of interest. The timing of technology releases is crucial for managers and investors alike. Both sides must strategically plan and consider the release of novel technologies, given their significant, causal impact on market prices. Furthermore, coordinating timely and relevant updates and marketing campaigns can alleviate the subsequent impact of major technology releases on price volatility.

A second implication for managers is the use of diversified portfolios. This approach helps mitigate the risks associated with rapid technological advancements that can significantly impact the volatility of AI-assets within both private and public portfolios. As more firms engage in AI-related businesses, such diversification becomes increasingly important.

Moving forward, policymakers should remain vigilant to prevent insider trading practices. The findings suggest that major technological releases, like those of Gemini or ChatGPT, can have a disruptive effect on the price volatility of AI-related assets. This apprehensiveness should extend to other areas of the financial markets as well, ensuring fair trading practices. Thus, developing and enforcing a regulatory framework that addresses the volatility in cryptocurrency markets following major technological advancements is pivotal. This can be realized by establishing guidelines, effectively enforcing transparency and therefore increase market integrity for both private and institutional investors.

On an individual level, private investors should implement measures to counteract sudden price movements of AI-assets. Strategies such as stop-loss orders, and again, portfolio diversification can help protect investments from unexpected market fluctuations triggered by new technology announcements, given a sufficiently large scale. Investors should also stay up to date with the latest technological developments, as this is fundamental to understanding market trends and making informed investment decisions. In today's world, there are numerous channels for staying informed about technological developments, including social media platforms and professional news outlets, such as X and the New York Times.

Returning to the research question and its imminent answer, the SDID results indicate significant empirical and statistical evidence for a causal impact of Gemini's release on AI-narrative and broader Big Data cohort-related cryptocurrency assets' price movements within the scope of this research. However, this is contradicted by the DID and SC methods, while the LDV method agrees to some extent with the CG cohort. Specifically, the LDV analysis showed a significant positive effect for the CG cohort of approximately 0.30%, whereas the SDID analysis indicated a positive effect of about 0.20% for both cohorts. This confirms the hypothesis and its corresponding notion that the release of Gemini would have a significant, albeit short-lived, causal impact on AI-related cryptocurrency assets. Furthermore, the longevity of these effects should be scrutinized and tested in future scientific research on this topic.

The polarization between these techniques highlights the potential for predicting and inferring causality based on cryptocurrency token trend behavior. Future research should explore these possibilities to deepen the understanding of how technology releases influence AI-cryptocurrency price volatility. This dissertation, therefore, serves as a foundation for more advanced and comprehensive research into the effects of AI-technology releases on AI-related assets.

7 Appendix

The code used in this dissertation can be accessed through the following Github Repository Link: ***https://github.com/MellaBI/DSMA-Thesis***.

Table 11: List of Tickers with Full Names

Ticker	**Full Name**
ABT	Arcblock
ADA	Cardano
AGI	SingularityNet
AGIX	SingularityNET
AGRS	Agoras: Currency of Tau
AIOZ	Aioz Network
AITECH	Solidus AI Tech
AKT	Akash Network
ALEPH	Aleph.im
ALI	Artificial Liquid Intelligence
ALGO	Algorand
APT	Aptos
ARB	Arbitrum
ARKM	Arkham
AR	Arweave
ATOM	Cosmos
AVAX	Avalanche
AAVE	Aave
BEAM	Beam
BCH	Bitcoin Cash
BGB	Bitget Token
BNB	Binance Coin
BTC	Bitcoin
CQT	Covalent
CGPT	ChainGPT
CHZ	Chiliz
CRO	Cronos
CTXC	Cortex
CUDOS	Cudos
DATA	Streamr
DIA	DIA

Continued on next page

Table 11 – continued from previous page

Ticker	Full Name
DKA	dKargo
DOGE	Dogecoin
DOT	Polkadot
DMTR	Dimitra
ETH	Ethereum
ENA	Ethena
FIL	Filecoin
FLOKI	Floki Inu
FETCH	Fetch.ai
FLUX	Flux
FORT	Forta
FTM	Fantom
GALA	Gala
GLM	Golem
GRT	The Graph
GPU	Node AI
HAI	Hacken Token
HBAR	Hedera
HOOK	Hooked Protocol
ICP	Internet Computer
IMX	Immutable
INJ	Injective
IQ	IQ
JUP	Jupiter
KAS	Kaspa
KDA	Kadena
KEY	SelfKey
LAT	PlatON
LEO	UNUS SED LEO
LDO	Lido DAO
LTC	Litecoin
LMWR	LimeWire
LINK	Chainlink
MATIC	Polygon
MKR	Maker
MNT	Mantle

Continued on next page

Table 11 – continued from previous page

Ticker	Full Name
NEO	NEO
NEAR	NEAR Protocol
NMR	Numeraire
NUM	Numbers Protocol
NOS	Nosana
NFP	NFPrompt
OASIS	Oasis Network
OKB	OKB
OCEAN	Ocean Protocol
ONDO	Ondo Finance
OP	Optimism
ORAI	Oraichain
OZONE	Ozone Chain
PAAL	Paladin
PEPE	Pepe Coin
PENDLE	Pendle
PHA	Phala Network
PHB	Phoenix Global
POND	Marlin
PRIME	Echelon Prime
QNT	Quant
RNDR	Render Token
RLC	iExec RLC
RUNE	THORChain
RSS3	RSS3
SEI	Sei Network
SHIB	Shiba Inu
SIDUS	Sidus
SOL	Solana
STX	Stacks
SURE	InSure DeFi
SUI	Sui
TAO	Bittensor
TFUEL	Theta Fuel
THETA	Theta Network
TON	Toncoin

Continued on next page

Table 11 – continued from previous page

Ticker	Full Name
TRAC	OriginTrail
TRX	TRON
UNI	Uniswap
VAI	VAIOT
VET	VeChain
VR	Victoria VR
VRA	Verasity
W	Wormhole
WETH	WETH
WIF	dogwifhat
X0X	WETH
XLM	Stellar
XMR	Monero
XRP	Ripple
ZIG	Zignaly

Table 12: Descriptive Statistics CoinGecko Control Group

Statistic	N	Mean	St. Dev.	Min	Max
Price	5,490	1,593.863	8,316.557	0.00000	73,097.770
Market Capitalization	5,490	47,566,665,745.000	168,809,146,693.000	263,405,179.000	1,436,631,290,572.000
Total Volume	5,490	1,885,014,149.000	6,477,707,918.000	77,611.560	96,403,762,013.000
Log Returns	5,490	0.006	0.047	−0.456	0.452
Gemini Exposure	5,490	0.656	0.475	0	1
AI	5,490	0.000	0.000	0	0
Gemini	5,490	0.000	0.000	0	0
Bull	5,490	0.098	0.298	0	1

Note: *Gemini* and *AI* represent dummy variables for being exposed to Gemini's release or not, denoted by 1 if exposed and 0 if not. *AI* (Artificial Intelligence) denotes the control group status, equal to 0 if it is part of the control group. Lastly, *Bull* denotes the presence of bull market conditions, defined by having a value of 1 and 0 otherwise.

Table 13: Descriptive Statistics CoinGecko Treatment Group

Statistic	N	Mean	St. Dev.	Min	Max
Price	5,490	12.692	70.887	0.002	728.385
Market Capitalization	5,490	314,937,312.000	616,397,854.000	0.000	4,999,869,338.000
Total Volume	5,490	32,785,138.000	96,604,927.000	35,125.300	1,338,093,577.000
Log Returns	5,490	0.011	0.081	−0.506	1.067
Gemini Exposure	5,490	0.656	0.475	0	1
AI	5,490	1.000	0.000	1	1
Gemini	5,490	0.639	0.480	0	1
Bull	5,490	0.098	0.298	0	1

Note: Gemini and *AI* represent dummy variables for being exposed to Gemini's release or not, denoted by 1 if exposed and 0 if not. *AI* (Artificial Intelligence) denotes the treatment group status, equal to 1 if it is part of the treatment group. Lastly, *Bull* denotes the presence of bull market conditions, defined by having a value of 1 and 0 otherwise.

Table 14: Descriptive Statistics CoinMarketCap Control Group

Statistic	N	Mean	St. Dev.	Min	Max
Price	10,065	912.911	6,225.964	0.00000	73,083.500
Total Volume	10,065	1,002,194,826.000	4,581,884,076.000	0.000	102,802,940,877.000
Market Capitalization	10,065	26,456,982,603.000	127,509,304,174.000	0.000	1,436,271,822,606.000
Log Returns	10,065	−0.006	0.058	−1.172	0.454
Gemini Exposure	10,065	0.656	0.475	0	1
AI	10,065	0.000	0.000	0	0
Gemini	10,065	0.000	0.000	0	0
Bull	10,065	0.098	0.298	0	1

Note: Gemini and *AI* represent dummy variables for being exposed to Gemini's release or not, denoted by 1 if exposed and 0 if not. *AI* (Artificial Intelligence) denotes the control group status, equal to 0 if it is part of the control group. Lastly, *Bull* denotes the presence of bull market conditions, defined by having a value of 1 and 0 otherwise.

Table 15: Descriptive Statistics CoinMarketCap Treatment Group

Statistic	N	Mean	St. Dev.	Min	Max
Price	10,065	13.960	74.382	0.001	730.255
Total Volume	10,065	29,186,500.000	88,201,872.000	0.000	1,642,844,689.000
Market Capitalization	10,065	367,038,073.000	815,708,349.000	0.000	9,271,213,900.000
Log Returns	10,065	−0.009	0.073	−0.748	0.445
Gemini Exposure	10,065	0.656	0.475	0	1
AI	10,065	1.000	0.000	1	1
Gemini	10,065	0.639	0.480	0	1
Bull	10,065	0.098	0.298	0	1

Note: Gemini and *AI* represent dummy variables for being exposed to Gemini's release or not, denoted by 1 if exposed and 0 if not. *AI* (Artificial Intelligence) denotes the control group status, equal to 0 if it is part of the control group. Lastly, *Bull* denotes the presence of bull market conditions, defined by having a value of 1 and 0 otherwise.

Table 16: Descriptive Statistics Ex-Large and Ex-Mega S&P500 Indexes for Cryptocurrencies

Statistic	N	Mean	St. Dev.	Min	Max
Index	260	4,420.383	2,591.777	1,270.530	11,460.250
Mega	260	0.500	0.501	0	1
Log Returns	260	0.009	0.036	−0.109	0.121
Bull	260	0.338	0.474	0	1
Lagged Log Returns	259	0.009	0.036	−0.109	0.121

Note: Mega denotes the Ex-Mega Large Capitalization coins, denoted by 1 if part of this group and 0 otherwise. *Bull* denotes the presence of bull market conditions, equalling 1 if present and 0 otherwise.

References

Adeshola, I., & Adepoju, A. P. (2023). The opportunities and challenges of chatgpt in education. *Interactive Learning Environments*, 1–14.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Ante, L. (2023). How elon musk's twitter activity moves cryptocurrency markets. *Technological Forecasting and Social Change*, *186*, 122112.

Ante, L., & Demir, E. (2023). The chatgpt effect on ai-themed cryptocurrencies. *SSRN 4350557*.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2019). *Synthetic difference in differences* (Tech. Rep.). National Bureau of Economic Research.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, *111*(12), 4088–4118.

Bouri, E., Das, M., Gupta, R., & Roubaud, D. (2018). Spillovers between bitcoin and other assets during bear and bull markets. *Applied Economics*, *50*(55), 5935–5949.

Clarke, D., Pailanir, D., Athey, S., & Imbens, G. (2023). Synthetic difference in differences estimation. *arXiv preprint arXiv:2301.11859*.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, *37*(1), 39–67.

De Giorgi, G., Frederiksen, A., & Pistaferri, L. (2020). Consumption network effects. *The Review of Economic Studies*, *87*(1), 130–163.

Doudchenko, N., & Imbens, G. W. (2016). *Balancing, regression, difference-in-differences and synthetic control methods: A synthesis* (Tech. Rep.). National Bureau of Economic Research.

Dowling, M., & Lucey, B. (2023). Chatgpt for (finance) research: The bananarama conjecture. *Finance Research Letters*, *53*, 103662.

Fama, E. F. (1970). Efficient capital markets. *The Journal of Finance*, *25*(2), 383–417.

Fama, E. F. (1991). Efficient capital markets: Ii. *The Journal of Finance*, *46*(5), 1575–1617.

Ford, C. (2015). Understanding qq plots| university of virginia library research data services+ sciences. *Retrieved July*, *27*, 2023.

Ganapathy, A., Redwanuzzaman, M., Rahaman, M. M., & Khan, W. (2020). Artificial intelligence driven crypto currencies. *Global Disclosure of Economics and Business*, *9*(2), 107–118.

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of chatgpt as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *2*(4), 100089.

Howe, J. S. (1986). Evidence on stock market overreaction. *Financial Analysts Journal*, *42*(4), 74–77.

Hu, S.-y., Lin, Y.-H., & Lai, C. W. (2016). The effect of overvaluation on investment and accruals: The role of information. *Journal of Empirical Finance*, *38*, 181–201.

Hudson, R. S., & Gregoriou, A. (2015). Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns. *International Review of Financial Analysis*, *38*, 151–162.

Jain, A. (2024). *Has the crypto bull run of 2024 officially started? insights and predictions.* Retrieved from https://crypto.news/has-the-crypto-bull-run-of-2024-officially-started-insights-and-predictions/ (Accessed: May 5, 2024)

Jareño, F., & Yousaf, I. (2023). Artificial intelligence-based tokens: Fresh evidence of connectedness with artificial intelligence-based equities. *International Review of Financial Analysis*, *89*, 102826.

Jeon, H., Youn, H., Ko, S., & Kim, T. (2022). Blockchain and ai meet in the metaverse. *Advances in the Convergence of Blockchain and Artificial Intelligence*, *73*(10.5772), 73–82.

Jones, A. M., & Rice, N. (2011). Econometric evaluation of health policies.

Kallir, I., & Sonsino, D. (2009). The neglect of correlation in allocation decisions. *Southern Economic Journal*, *75*(4), 1045–1066.

Kattenberg, M., Scheer, B., & Thiel, J. (2023). *Causal forests with fixed effects for treatment effect heterogeneity in difference-in-differences* (Tech. Rep.). CPB Netherlands Bureau for Economic Policy Analysis.

Keele, L., & Kelly, N. J. (2006). Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political analysis*, *14*(2), 186–205.

Kothari, S. P., & Warner, J. B. (2007). Econometrics of event studies. In *Handbook of empirical corporate finance* (pp. 3–36). Elsevier.

Marr, B. (2024). *Ai showdown: Chatgpt vs. google's gemini – which reigns supreme?* (https://www.forbes.com/sites/bernardmarr/2024/02/13/ai-showdown-chatgpt-vs-googles-gemini–which-reigns-supreme/)

McWilliams, A., Siegel, D., & Teoh, S. H. (1999). Issues in the use of the event study methodology: A critical analysis of corporate social responsibility studies. *Organizational Research Methods*, *2*(4), 340–365.

Meer, D. V. (2024). *Number of chatgpt users and key stats (march 2024).* (https://www.namepepper.com/chatgpt-users)

Mystakidis, S. (2022). Metaverse. *Encyclopedia*, *2*(1), 486–497.

Nguyen, K. Q., Nguyen, T. H., & Do, B. L. (2023). Narrative attention and related cryptocurrency returns. *Finance Research Letters*, *56*, 104174.

OpenAI, T. (2022). *Chatgpt: Optimizing language models for dialogue. openai.*

Ozili, P. K. (2023). The acceptable r-square in empirical modelling for social science research. In *Social research methodology and publishing results: A guide to non-native english speakers* (pp. 134–143). IGI global.

O'Neill, S., Kreif, N., Grieve, R., Sutton, M., & Sekhon, J. S. (2016). Estimating causal effects: Considering three alternatives to difference-in-differences estimation. *Health Services and Outcomes Research Methodology*, *16*, 1–21.

Pessa, A. A., Perc, M., & Ribeiro, H. V. (2023). Age and market capitalization drive large price variations of cryptocurrencies. *Scientific Reports*, *13*(1), 3351.

Pierce, J. P., Gilpin, E. A., Emery, S. L., White, M. M., Rosbrook, B., & Berry, C. C. (1998). Has the california tobacco control program reduced smoking? *Jama*, *280*(10), 893–899.

Pohl, M., Baumann, L., Behnisch, R., Kirchner, M., Krisam, J., & Sander, A. (2021). Estimands—a basic element for clinical trials: Part 29 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*, *118*(51-52), 883.

Rane, N., Choudhary, S., & Rane, J. (2024). Gemini versus chatgpt: Applications, performance, architecture, capabilities, and implementation. *Performance, Architecture, Capabilities, and Implementation (February 13, 2024).*

Reijers, W., & Coeckelbergh, M. (2018). The blockchain as a narrative technology: Investigating the social ontology and normative configurations of cryptocurrencies. *Philosophy & Technology*, *31*, 103–130.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688.

Saggu, A., & Ante, L. (2023). The influence of chatgpt on artificial intelligence related crypto assets: Evidence from a synthetic control analysis. *Finance Research Letters*, *55*, 103993.

Salah, K., Rehman, M. H. U., Nizamuddin, N., & Al-Fuqaha, A. (2019). Blockchain for ai: Review and open research challenges. *IEEE access*, *7*, 10127–10149.

Sobania, D., Briesch, M., Hanna, C., & Petke, J. (2023). An analysis of the automatic bug fixing performance of chatgpt. In *2023 ieee/acm international workshop on automated program repair (apr)* (pp. 23–30).

Stosic, D., Stosic, D., Ludermir, T. B., & Stosic, T. (2018). Collective behavior of cryptocurrency price changes. *Physica A: Statistical Mechanics and its Applications*, *507*, 499–509.

Strong, N. (1992). Modelling abnormal returns: A review article. *Journal of Business Finance & Accounting*, *19*(4), 533–553.

Sundar Pichai, D. H. (2023). *Introducing gemini: Our largest and most capable ai model.* (https://blog.google/technology/ai/google-gemini-ai/sundar-note)

Team, G., Anil, R., Borgeaud, S., Wu, Y., Alayrac, J.-B., Yu, J., … others (2023). Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Thadewald, T., & Büning, H. (2007). Jarque–bera test and its competitors for testing normality–a power comparison. *Journal of applied statistics*, *34*(1), 87–105.

Torres-Reyna, O. (2015). Differences-in-differences (using r). *Princeton University, http://www princeton edu/˜ otorres/DID101R pdf, last downloaded*, *23*, 2017.

Zhang, Y., Chan, S., Chu, J., & Sulieman, H. (2020). On the market efficiency and liquidity of high-frequency cryptocurrencies in a bull and bear market. *Journal of Risk and Financial Management*, *13*(1), 8.