Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of Business | Lee Kong Chian School of Business

2-2023

The financialization of cryptocurrencies

Lei HUANG

Tse-Chun LIN

Fangzhou LU

Jian SUN
Singapore Management University, jiansun@smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the Finance Commons, and the Finance and Financial Management Commons

Citation

HUANG, Lei; LIN, Tse-Chun; LU, Fangzhou; and SUN, Jian. The financialization of cryptocurrencies. (2023). 1-68.

Available at: https://ink.library.smu.edu.sg/lkcsb_research/7447

This Working Paper is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

The Financialization of Cryptocurrencies *

Lei Huang[a] Tse-Chun Lin[a] Fangzhou Lu[a] Jian Sun[b]

February 25, 2023

Abstract

Through the lens of cryptocurrency financialization, we show that change in Grayscale Bitcoin Trust premium is the most significant predictor of Bitcoin daily return. Using K-means clustering and LDA analysis, we find that this predictability is especially significant when there is a large variation in bullish and bearish market sentiment, innovation regarding CBDC, and regulations on crypto exchanges, but not when there is innovation regarding blockchain technology or Bitcoin mining. These findings suggest that indexing serves as a channel for information transmission, and Bitcoin prices react with a delay to the information contained in the sentiment of traditional investors.

JEL Codes: E42, G12, G30, G41.

Keywords: Cryptocurrency Predictability, Bitcoin Closed-End Funds and ETFs, GBTC, Investor Sentiment.

*[a]HKU Business School, The University of Hong Kong. [b]Lee Kong Chian School of Business, Singapore Management University. Please send correspondence to: hlei123@connect.hku.hk, tsechunlin@hku.hk, lufz@hku.hk, jiansun@smu.edu.sg. This research is previously circulated under the title "When Early Adopters Learn From the Followers: The Cryptocurrency Return Predictability of GBTC Discount and Premium". We thank Urban Jermann, Antoinette Schoar, and Robert Whitelaw for providing invaluable comments. We also thank seminar participants at HKU.

Electronic copy available at: https://ssrn.com/abstract=3948407

1 Introduction

As cryptocurrencies gradually become an acceptable asset class for both retail investors and institutional investors in recent years, there has been a trend of financialization of cryptocurrencies. The CME group lists Bitcoin futures and options on futures starting from 2021. Various closed-end mutual funds such as Grayscale Bitcoin Trust and Bitcoin ETFs are currently listed in the OTC market or Canadian stock markets that give non-crypto experts channels to gain cryptocurrency exposure. Similar to other non-sovereign stores of value such as Gold, anecdotal evidence suggests that this trend of indexing combined with Bitcoin custody service lead to an increase in the price level of Bitcoin and other cryptocurrencies (Jermann, 2021). Financialization does not only establish channels that transmit shocks but also transmit information between investors (Goldstein and Yang, 2022). Through this lens, our research empirically confirms the predictability of cryptocurrency returns, particularly for Bitcoin, and identifies the factors that determine cryptocurrency value in the market equilibrium.

Our focus is on the Grayscale Bitcoin Trust (GBTC), a closed-end public-traded fund that only holds Bitcoin and has been listed on the OTC market since 2015. GBTC is the longest-traded and financialized cryptocurrency product and holds between 2% to 3% of all Bitcoins in circulation, making it the largest Bitcoin fund.[1] Individual investors can easily trade GBTC through their brokerage or 401K accounts, allowing them to invest in Bitcoin without having to deal with the technicalities of cryptocurrency ownership.[2] For institutional investors or those in countries that ban Bitcoin, investing in GBTC is almost the

[1]https://www.etftrends.com/alternatives-channel/institutional-investors-piling-into-Bitcoin/. On October 19th, 2021, the first U.S. Bitcoin futures exchange-traded (ETF) fund launches, a milestone for the cryptocurrency and blockchain communities. The Bitcoin price has risen 8.6% to 61862 USD since the announcement on October 15th. Before the new ETF fund, most U.S. traditional investors could only get exposure to Bitcoin via investing in Grayscale Bitcoin Trust (GBTC) or investing in Bitcoin through Robinhood. GBTC reaches the highest AUM in Nov 2021 with 41 billion USD under management while all the other Bitcoin ETFs' combined AUM is less than 10% of that of GBTC.

[2]See https://www.nytimes.com/2021/01/12/technology/Bitcoin-passwords-wallets-fortunes.html. Figure 1 shows that GBTC has been traded at a premium since its inception. Only after February 2021, the premium turns into a discount. Therefore, we use GBTC premium to generically refer to both GBTC discount and premium in this paper.

Electronic copy available at: https://ssrn.com/abstract=3948407

only legitimate option to include Bitcoin in their portfolios while complying with government regulations. Since its inception, GBTC has been primarily traded at a premium that partly reflects the high demand from investors who prefer not to hold and store Bitcoin directly. In this paper, we hypothesize that the GBTC premium reflects the hyper sentiment of traditional investors that are new to cryptocurrency investment.[3] This hyper sentiment may be seen as a sign of recognition and endorsement among cryptocurrency communities, which could boost the price of Bitcoin. Given the market segmentation and limited attention of traditional investors, their sentiment towards cryptocurrency could transmit new and valuable information to investors in the Bitcoin spot markets.

We finds that the GBTC premium is the most powerful predictor of daily Bitcoin returns, outperforming previously studied predictors in the literature (e.g., Google search volume, Twitter search, and lagged Bitcoin return). A one-standard-deviation increase in the lagged GBTC premium change leads to on average a 80 basis point increase in Bitcoin returns the following day. Moreover, GBTC premium also predicts value-weighted cryptocurrency market return (both with and without Bitcoin and Ethereum). A one-standard-deviation increase in lagged GBTC premium change leads to on average a 1.25% increase in value-weighted cryptocurrency market return. This return predictability holds for both proof-of-work and proof-of-stake cryptocurrencies.

These findings support our hypothesis that the GBTC premium, determined by traditional investors, drives the spot price of Bitcoin, which is set by traders in specialized cryptocurrency markets. We confirm this relationship by using a vector autoregression (VAR) and show that changes in GBTC premium precede changes in Bitcoin returns, rather than the other way around. This suggests that Bitcoin specialists learn signals from traditional investors regarding the value of Bitcoin, such as the likelihood of Bitcoin and other cryptocurrencies' future adoption.

Unlike Bitcoin return predictability, stock return predictability is well established in the

[3]The GBTC premium/discount is also empirically uncorrelated with the closed-end fund discount documented in Lee, Shleifer, and Thaler (1991) and Baker and Wurgler (2006) at the daily level.

Electronic copy available at: https://ssrn.com/abstract=3948407

literature.[4] However, compared to the stock market, one obvious difference between Bitcoin and the stock market is the difference in the fundamentals. There are clear shocks to either cash flow or discount rates for traditional assets like stocks and bonds. For Bitcoin, there is only sentiment or belief regarding how widely Bitcoin will be adopted. Therefore, to gain a better understanding of the specific information contained in the GBTC premium, we use textual analysis to categorize news related to Bitcoin and blockchain. We collect comprehensive coverage of Blockchain News from Coindesk, a leading news source in the field, and apply K-means clustering and Latent Dirichlet allocation methods to categorize the news.

We find that the predictability of bitcoin returns through the GBTC premium is more pronounced in response to news about central bank digital currencies (CBDCs) and bans on crypto exchanges, but not to news about blockchain technology and bitcoin mining. The predictability of the GBTC premium is also influenced by trading sentiment, as indicated by keywords such as "bullish," "bearish," "bitcoin price," and "Chart."[5] These results suggest that the predictability of bitcoin returns is less tied to the underlying technology of bitcoin and more connected to investor sentiment, traditional investors' concerns about regulation, and the competition posed by CBDCs.

Furthermore, our findings show that after the GBTC indexing, there was a 6.3% increase in the daily return correlation with the S&P 500 for bitcoin relative to other cryptocurrencies not undergoing financialization. Similar increases were observed in the correlation with the returns of the Dow Jones and Nasdaq indices. There was also a 2% decrease in return volatility and a 3.2% decrease in the price delay R^2 measure, suggesting that price informativeness had increased (Hou and Moskowitz, 2005). Additionally, there was a rise in correlation with the SMB portfolio, and a decrease in correlation with the HML, RMW, and

[4] E.g., Fama and Schwert (1977), Lo and MacKinlay (1988), Breen, Glosten, and Jagannathan (1989), Badrinath, Kale, and Noe (1995), Sias and Starks (1997), Chordia and Swaminathan (2000), Ahn, Boudoukh, Richardson, and Whitelaw (2002), Hong, Torous, and Valkanov (2007), Driesprong, Jacobsen, and Maat (2008), Pástor and Stambaugh (2009), Menzly and Ozbas (2010), and Rapach, Strauss, and Zhou (2013).

[5] Chart is related to fundamental analysis such as bar chart used by technical investors.

Electronic copy available at: https://ssrn.com/abstract=3948407

CMA portfolios, indicating that traditional investors may view cryptocurrencies as similar to small, high-growth, less profitable firms with less investment. Overall, these results suggest that the financialization process is transmitting the views of traditional investors to the cryptocurrency market and affecting its equilibrium price.

Next, we investigate whether GBTC premium reflects the sentiment of blockchain communities. Liu and Tsyvinski (2021) show that address growth, transaction growth, wallet number growth, and payment growth are the four factors that are most related to the technological fundamental of Bitcoin. Easley, O'Hara, and Basu (2019) and Cong, He, and Li (2021a) show that supply factors of Bitcoin, such as mining pool competition, also matter for the success of Bitcoin. However, most traditional investors do not possess the technical expertise needed to manage a Bitcoin wallet, run nodes, or store Bitcoin on hard drives.[6] If the GBTC premium reflects the sentiment from blockchain specialists, then it should predict the technological fundamentals and supply factors of Bitcoin. Our result shows that the GBTC premium does not predict these variables, thereby suggesting that the GBTC premium does not reflect the sentiment of bitcoin specialists.

Baker and Wurgler (2006) find that growth, no-dividend-paying, and hard-to-value stocks are more affected by investor sentiment. Liu, Tsyvinski, and Wu (2021) show that size and momentum are two crucial factors in pricing cryptocurrencies in the cross-section. Therefore, we conjecture that GBTC premium may also predict factor premium in the cross-section. We find that cryptocurrencies with younger "age," smaller market capitalization, and lower prices are more likely to be influenced by GBTC premium. For example, the size premium in Liu et al. (2021) is higher at the daily level when the lagged change in GBTC premium is larger. A one-standard-deviation increase in lagged GBTC premium change leads to a 53 basis point daily return difference between cryptocurrencies in the lowest price quintile and those in the highest price quintile.

Furthermore, given that the GBTC premium contains valuable information for pricing

[6] An alternative explanation is that the attention cost prevents traditional investors from maintaining their own wallets or nodes (Gabaix, 2014).

Electronic copy available at: https://ssrn.com/abstract=3948407

bitcoin and other cryptocurrencies, we hypothesize that the impact of the GBTC premium will be greater for cryptocurrencies that are harder to value. We measure the difficulty of valuing a cryptocurrency by the readability of its whitepaper.[7] Specifically, we use the (i) Gunning-Fog index, (ii) length of the cryptocurrency whitepaper (Loughran and McDonald, 2014, 2016), (iii) the frequency of cryptocurrency technical words, (iv) the percentage of weak modal words, and (v) the percentage of uncertainty words to measure the difficulties of arbitraging and valuing these cryptocurrencies. We find that cryptocurrencies that are more difficult to be understood and valued are more influenced by the GBTC premium. On average, a one-standard-deviation increase in lagged GBTC premium change leads to a 1.05% return difference between hard-to-value cryptocurrencies and easy-to-value ones.

Makarov and Schoar (2020) show that there are significant price dispersions in cryptocurrency exchanges worldwide. Suppose there is a lead-lag relationship of Bitcoin returns between two cryptocurrency exchanges. One might concern that the Bitcoin return predictability of GBTC premium can be purely driven by such a lead-lag relationship. We address this concern by showing that the change in GBTC premium predicts Bitcoin return across major cryptocurrencies worldwide in different time zones, even though Bitcoin price is decentralized and dispersed across different exchanges. A one-standard-deviation increase in lagged GBTC premium change leads to a 0.5% to 3.2% increase in Bitcoin return among different exchanges and exchanges with various fiat currency denominations.

We also find that the sensitivity of cryptocurrency returns to GBTC premium is higher when there is more news about Bitcoin and a high trading volume of Bitcoin. When the trading volume is above the median Bitcoin dollar volume, a one-standard-deviation increase in lagged change of GBTC premium leads to a 1.1% higher Bitcoin return. In contrast, when the trading volume is below the median volume, a one-standard-deviation increase in lagged change of GBTC premium leads to a 0.35% higher Bitcoin return. The difference of 76 bps

[7] A cryptocurrency whitepaper is essentially the business plan that explains the technical details of the cryptocurrency, such as whether it is proof-of-stake or proof-of-work, its token distribution mechanism, or initial offering price.

Electronic copy available at: https://ssrn.com/abstract=3948407

is economically and statistically significant. This result indicates that market microstructure issues, e.g., illiquidity, do not drive Bitcoin return predictability. Finally, we find that the Bitcoin return predictability of GBTC premium changes holds for both positive and negative Bitcoin returns. The result is stronger when changes in GBTC premium changes are negative. Because GBTC is a de facto closed-end fund where redemptions of shares are not possible, this suggests that our result is not primarily driven by price impact from GBTC fund flow.

Our paper contributes to the literature by providing the first evidence that change in GBTC premium is the most significant predictor of Bitcoin return at the daily level. Makarov and Schoar (2020) show that large price dispersion exists in cryptocurrency exchanges, and capital control seems to be the primary reason that causes this price discrepancy. Liu and Tsyvinski (2021) show that only factors about Bitcoin per se are related to future Bitcoin prices. Griffin and Shams (2020) show that Bitcoin prices can be subjected to manipulation. We add to these studies by showing that traditional investors' sentiment toward Bitcoin could lead the investment decisions made by cryptocurrency specialists.

Through the lens of this financialization process, our research shows what determines cryptocurrency value in the equilibrium, and how shocks and information are transmitted between different clientele after indexing. Cheng and Xiong (2014) provide a comprehensive literature summary and show that previous literature finds that commodity financialization pushes up the commodity price, and facilitates or distorts the price discovery process. Henderson, Pearson, and Wang (2015) find that prices, volatilities, and correlations go up with financialization, but more so for index futures than for non-index futures. Goldstein and Yang (2022) find that financial speculators help improve price informativeness, whereas financial hedgers reduce informativeness, and the overall effect depends on the ratio and interaction between these two groups of investors. We find that cryptocurrency financialization increases price informativeness, return correlation with the stock market, and transmits traditional investors' information to crypto investors.

Our paper also contributes to the literature on investor sentiment. Baker and Wurgler

Electronic copy available at: https://ssrn.com/abstract=3948407

(2006) show that investor sentiment predicts stock return in the cross-section. Baker, Wurgler, and Yuan (2012) further find that relative sentiment levels predict the divergence of twin shares. Huang, Jiang, Tu, and Zhou (2015) show that after removing the noise component in sentiment, investor sentiment can predict stock return with strong economic and statistical significance. Our daily-level Bitcoin return predictability is in line with Griffin and Shams (2020) that most of the return dynamic of Bitcoin return happens at a higher frequency than the traditional asset classes.

2 Data and institutional details

Grayscale Bitcoin Trust (GBTC) is a trust fund that provides institutional investors with a formal channel to hold Bitcoin with custodian service since May 2015.[8] It is a product of Coinbase, which was listed on the Nasdaq on April 14th, 2021. The Grayscale Bitcoin trust is a fund that only holds Bitcoin. Compared to holding Bitcoin directly, holding GBTC has the feature of built-in security and storage service provided by Coinbase. For each share, it represents 0.001 Bitcoin at its inception.[9] As of 2021 January, it holds close to 3% of all Bitcoin in circulation.

Although GBTC may seem like an ETF at first, redemptions of shares are not possible, so it is a de facto closed-end fund. Different from closed-end funds that typically trade at a discount (Lee, Shleifer, and Thaler, 1991), Figure 1 shows that GBTC has been traded at a premium since its inception due to excess demand from traditional investors. Only after February 2021, the premium turns to a discount potentially caused by rumors of other Bitcoin ETFs as competitors. Only institutional investors can create new shares at par, but they are restricted from selling for a period of six months. GBTC also rarely sells Bitcoin, and it only sells Bitcoin occasionally to collect management fees.[10] Figure 2 shows

[8]Similarly, Coinbase also lists an Ethereum closed-end fund, ETHE in June 2019.

[9]Since GBTC charges a 2% management fee per year directly from the Bitcoin owned by GBTC shareholders, it represents 0.00094861 Bitcoin as of January 2021.

[10]For more information, also see https://blog.bybit.com/en-us/insights/why-does-grayscale-seldom-sell-

Electronic copy available at: https://ssrn.com/abstract=3948407

the time-series of the number of Bitcoins held by GBTC and a steady upward trend can be observed. Therefore, most of the fluctuation in GBTC premium reflects trading among existing shareholders of GBTC.

The Grayscale Bitcoin Trust fund is traded on the OTCQX, initially a pink-sheet market unregulated by SEC. GBTC can be traded through a brokerage firm in a similar manner to stocks, and it is also available through tax advantage accounts like 401Ks and IRAs. Due to its monopoly position in the U.S., it is one of the few ways institutional investors can purchase Bitcoin at the fund level because the SEC restricts hedge funds from buying assets without a custodian account. We obtain GBTC data from May 2015 to June 2021 from Bloomberg. We obtain BTC and other cryptocurrencies' return data from Coinmarketcap and different crypto exchanges' APIs. Despite the fact that GBTC only holds Bitcoin, its return correlation with Bitcoin was only 79.1% during our sample period at the daily level. This "low" correlation despite the same underlying suggests a de-facto market segmentation between the GBTC market and the Bitcoin spot market.

In addition to Grayscale Bitcoin Trust, there are also several newly listed Bitcoin ETFs, including Purpose Bitcoin ETF and CI Galaxy Bitcoin ETF trading on the Toronto stock exchange (February and March 2021). On October 19th, 2021, the first U.S. Bitcoin futures exchange-traded fund (ETF) BITO was launched. However, none of the above ETFs has a long historical return and high AUM level as the Grayscale Bitcoin Trust.

Lee, Shleifer, and Thaler (1990), Lee et al. (1991), and Chopra, Lee, Shleifer, and Thaler (1993) find that closed-end fund discount reflect retail investors' level of sentiment. Baker and Wurgler (2006) show that sentiment predicts stock return in the cross-section. Growth stocks, no-dividend paying stocks, and stocks that are hard to be valued generally are more affected by sentiments. Among the six measures that predict stock return, closed-end fund discount accounts for half of the sentiment measure variation. Motivated by these observations, we construct the closed-end fund discount/ premium based on the GBTC price premium. For

Bitcoin/

Electronic copy available at: https://ssrn.com/abstract=3948407

institutional investors and investors in countries that ban Bitcoin, GBTC is almost the only way that allows institutional investors to hold Bitcoin with custodian service legitimately and comply with government regulations. Figure 1 shows that after 2018.7.2, or the date Coinbase started to directly offer Bitcoin custodian service to institutions, there is a significant drop in GBTC premium.[11] Therefore, the graph implies that before that date, the premium partially reflects institutional investors' demand. However, Figure 3 shows that the total institutional holding of GBTC is less than 2% as of October 2021. Therefore, the GBTC premium should also reflect the demand from retail investors.

Moreover, the GBTC could also reflect the demand from foreign investors who face risks in their countries when holding Bitcoin directly. Figure 1 shows that after China banned cryptocurrency exchanges and ICOs in February 2017, there was a significant jump in GBTC premiums. Last but not least, the GBTC premium could also reflect the diversification demand from equity investors. Overall, it reflects the hyper sentiment of traditional investors that are new to cryptocurrency investment.

In Table 1, we show that the GBTC premium at a monthly level is negatively correlated with the sentiment measure in Baker and Wurgler (2006). This evidence suggests that the GBTC premium is a different sentiment measure from the aggregate economy sentiment in general. We also construct the closed-end fund discount at the daily level from Bloomberg's closed-end fund database. From May 2015 to June 2021, Bloomberg records 613 closed-end mutual funds at the daily level. However, changes in GBTC premium have a less than one percent correlation with changes in equal-weighted or value-weighted closed-end mutual fund discount at the daily level.

Besides the closed-end fund discount, is it possible to construct similar measures such as first-day returns on IPOs (RIPO), IPO volume (NIPO), equity share, value-weighted dividend premium (PDND, Baker and Wurgler (2004))? For equity shares and PDND shares,

[11] With Coinbase and other Fintech firms' Bitcoin custodian service in place, institutional investors don't have to hold GBTC to hold Bitcoin. Instead, they can buy Bitcoin on their own and adopt the custodian service offered by firms such as Coinbase.

Electronic copy available at: https://ssrn.com/abstract=3948407

there is no close analogy in the cryptocurrency market. However, as for RIPO and NIPO, we can obtain similar measures from the initial coin offering market (ICOs). This market is trendy between 2016 and the end of 2018. By the end of 2018, the market lost it vibe and was replaced by the decentralized finance (DEFI) market. Therefore, we don't have a particularly long time series. Using the monthly level of initial return and number of ICOs, or the dollar amount of ICOs at a monthly level similar as RIPO and NIPO, we don't find a particular strong predictability relationship between the number of ICOs, or the dollar amount of ICOs and Bitcoin return, probably since this ICO market is too short-lived. Moreover, unlike the closed-end fund discount, we cannot appropriately use daily data because the initial return and number of ICOs or the dollar amount of ICOs are slow-moving variables that can only be measured at a monthly level.

There are multiple methods for calculating the return of Bitcoin and constructing the GBTC premium. GBTC is traded during U.S. eastern time from 9:00 am to 4:00 pm, while BTC is traded continuously. To calculate the GBTC premium, we use the 4:00 pm eastern time GBTC price and divide it by the 4:00 pm eastern time BTC price. We also compute the return of BTC based on its price at 4:00 pm eastern time. Our robustness checks indicate that the results remain similar if we use any hour between 9:00 am to 4:00 pm eastern time to calculate the GBTC premium and BTC return.

3 Conceptual Framework

In this section, we provide a theoretical motivation of why GBTC premium could potentially lead the return of Bitcoin and the cryptocurrency market return in general. We build a simple and standard model in Appendix A.1 based on Goldstein and Yang (2017). The model considers three assets traded in the financial market: GBTC, BTC, and a risk-free asset. GBTC and BTC have the same random terminal value $\tilde{v}$. We assume that $\tilde{v} \sim N\left(0, \tau_v^{-1}\right)$ with precision $\tau_v > 0$. GBTC and BTC are traded at endogenous prices p_t and p_s at time 0,

Electronic copy available at: https://ssrn.com/abstract=3948407

respectively. In our model, p_t and p_s may not be the same due to the market frictions that we will discuss later.

There are three types of traders in the market: traditional traders, sophisticated traders and noise traders. For simplicity, traditional traders only trade in the GBTC market, and they have CARA utility and risk aversion γ_t. They represent investors who have limited knowledge about BTC, and face high costs to open and maintain a BTC wallet. So they will choose to have some BTC exposure through a more convenient financial tool, which is GBTC in our model. Sophisticated traders only trade in the BTC (spot) market and have risk aversion γ_s.[12] Compared to traditional traders, they have more knowledge about BTC, and thus prefer to open and maintain their BTC wallet themselves. Noise traders exist in both GBTC market and BTC market, and they are trading for pure liquidity reasons. We assume that noise traders demand $\tilde{x}_t$ and $\tilde{x}_s$ in the GBTC market and BTC market, respectively, where $\tilde{x}_t \sim N\left(0, \tau_{x,t}^{-1}\right)$ and $\tilde{x}_s \sim N\left(0, \tau_{x,s}^{-1}\right)$. $\tilde{x}_t$ and $\tilde{x}_s$ are independent of each other, and both of them are independent of all other random variables in the model.

Each traditional trader i can receive a private signal $s_{t,i}$ about terminal value $\tilde{v}$:

$$s_{t,i} = \tilde{v} + \epsilon_{t,i},$$

where $\epsilon_{t,i} \sim N\left(0, \tau_{t,\epsilon}^{-1}\right)$ is the i.i.d error term. And each sophisticated trader j can receive a private signal

$$s_{s,j} = \tilde{v} + \epsilon_{s,j},$$

where $\epsilon_{s,j} \sim N\left(0, \tau_{s,\epsilon}^{-1}\right)$ is the i.i.d error term. In addition to the private signals, the endogenous prices p_t and p_s are also informative about terminal value $\tilde{v}$, and can possibly be used by traders in their investment decisions. We assume that sophisticated trader can process

[12]Restricting sophisticated traders from only trading in BTC market is a simplifying assumption. But in practice, there are many reasons to rationalize this result: first, trading GBTC may incur significant loss from management fees; second, there is a significant premium in GBTC over a long time period (most of the time in our data sample), so it is very costly for sophisticated traders to gain BTC exposure by trading GBTC; third, BTC is traded 24 hours a day while the GBTC is only traded from 9:00 am to 4:00 pm U.S. eastern time.

Electronic copy available at: https://ssrn.com/abstract=3948407

all information available to them, including their private signals, GBTC price, and bitcoin price, and thus their information set is $\{s_{s,j}, p_t, p_s\}$. However, traditional traders have limited attention and are not able to process all information available in their investment decisions. For example, they may not be full-time Bitcoin traders or are trading various financial assets and thus can only allocate limited attention to BTC research. Specifically, we assume that traditional traders can always learn GBTC price p_t as they are trading in the GBTC market. But they can only access one of their private signals or BTC price p_s. So for each traditional trader i, his information set can either be $\{s_{t,i}, p_t\}$ or $\{p_t, p_s\}$.

3.1 Equilibria

For each traditional trader i, his information choice is between $s_{t,i}$ and p_s, both are noisy measures of terminal value $\tilde{v}$ in equilibrium. In equilibrium, if

$$\text{Var}\left(\tilde{v}|s_{t,i}\right) < \text{Var}\left(\tilde{v}|p_s\right),$$

then he will choose $s_{t,i}$, otherwise, he will choose p_s. Since the private signals are symmetric for all traditional traders, there are only two possible equilibria: either all traditional traders choose to learn their private signals, or all of them choose to learn BTC spot price p_s. We call the former equilibrium the informative equilibrium, and the latter the uninformative equilibrium.

Then we obtain the following proposition.

Proposition 1. *When $\tau_{t,\epsilon} > \left(\frac{\tau_{s,\epsilon}}{\gamma_s}\right)^2 \tau_{x,s}$, the equilibrium is the informative equilibrium, otherwise, the equilibrium is the uninformative equilibrium.*

We are especially interested in the informative equilibrium, under which GBTC price aggregates information about traditional traders' private signals, and sophisticated traders learn some information about terminal value $\tilde{v}$ from the GBTC price. Since sophisticated traders make investment decisions based on GBTC prices, our model suggests that the BTC

Electronic copy available at: https://ssrn.com/abstract=3948407

spot price can be partially predicted by the GBTC price. Besides, this predictability is sustained only when condition

$$\tau_{t,\epsilon} > \left(\frac{\tau_{s,\epsilon}}{\gamma_s}\right)^2 \tau_{x,s}$$

holds. Specifically, the informative equilibrium is more likely to hold when (1) the GBTC market is more price informative, (2) risk aversion in the Bitcoin spot market is high, (3) the Bitcoin spot market is less price informative, (4) the volatility of noisy trader demand is high in the Bitcoin spot market. In the empirical analysis, we test Proposition 1.

4 Empirical analysis

Bitcoin or cryptocurrencies, in general, are untraditional assets that cannot be priced by the traditional asset pricing model because there is no cash flow. According to the network theory, the more adoption of a certain cryptocurrency, the higher the value of the cryptocurrency, such as Bitcoin (Cong, Li, and Wang, 2021b). The belief or likelihood of Bitcoin adoption or any currencies in general naturally follows a self-fulfilling prophecy that depends on investors' confidence in the currency. As GBTC premium contains traditional investors' view of Bitcoin, this traditional investor sentiment toward cryptocurrency becomes crucial to Bitcoin's success, and likely contains useful signals for sophisticated investors in the Bitcoin spot market as we hypothesize in Proposition 1.

4.1 Investor sentiment and GBTC predictability

Baker and Wurgler (2006) and Baker and Wurgler (2007) show that higher investment sentiment leads to lower future stock return. Closed-end fund premium or discount is one of the most important factors that predict lower returns. However, since Bitcoin has no cash flow and belief in higher current value leads to a higher likelihood of future adoption, we hypothesize that higher sentiment is likely to drive Bitcoin price up in the short run. Therefore, we regress Bitcoin's daily return on lagged change in GBTC premium. We use changes in

Electronic copy available at: https://ssrn.com/abstract=3948407

GBTC premium instead of GBTC return as the main independent/state variable because the change in GBTC to BTC ratio identifies the part of the GBTC return that cannot be explained by contemporaneous Bitcoin return despite the same underlying asset. It is the excess information/noise from the traditional investor that has not been included in the current period Bitcoin spot price.

Table 2 panel A shows that changes in lagged GBTC Premium lead BTC return at the daily level. A one standard deviation increase in lag GBTC premium change leads to a 80 basis point increase in Bitcoin return. This is significant after we control for lagged Bitcoin return and lagged Google search volume of the keyword Bitcoin, which predicts Bitcoin return in Liu and Tsyvinski (2021) at the weekly level.[13] The lagged Bitcoin return controls for concern raised in Ahn et al. (2002) that auto-correlation in Bitcoin drives return predictability. While lagged Bitcoin return and Google search can barely predict future Bitcoin return, lagged change in GBTC premium can predict BTC return with a R^2 of close to 1%, far larger than the predictability other factors.[14] This is evidence that GBTC premium can explain return variation better than other variables.

A natural question that follows our result is, while GBTC premium leads Bitcoin return, can Bitcoin also lead GBTC? Therefore, we also conduct Vector autoregression (VAR) analysis similarly as in Da, Engelberg, and Gao (2011) at the daily level. We include both a constant and a time trend in the VAR. Figure 4 and Table 2 panel B shows that changes in GBTC premium indeed lead Bitcoin instead of the other way around. These results are consistent with our hypothesis that the GBTC premium determined by traditional investors leads the price of Bitcoin determined by the early adopters of cryptocurrencies.

[13] Similarly, the keyword "Bitcoin Hack" also does not have predictability at the daily level. Following the methodology in Da, Engelberg, and Gao (2011), we also attempt to construct Google search volume indexes for the keyword "Bitcoin" and "Bitcoin Hack" after demeaning the search volume either using the weekly mean search volume or monthly mean search volume. The demeaned search volume indexes also cannot predict Bitcoin return at the daily level. This suggests that at the daily frequency versus the weekly frequency, the predictive variables for cryptocurrency return may be different (Liu and Tsyvinski, 2021).

[14] While some research finds positive autocorrelation between Bitcoin daily return, in Appendix Table A.1, we actually find close to zero autocorrelation of Bitcoin at the daily level during our sample period, regardless the starting hour used to calculate the daily return.

Electronic copy available at: https://ssrn.com/abstract=3948407

Moreover, if we divide Bitcoin return into five groups based on the value of lagged changes in GBTC premium, Table 2 panel C shows that during the days with the biggest lagged changes in GBTC premium, on average, Bitcoin has a daily return of 1.10% with a t-statistics of 3.80. During the days with the smallest lagged changes in GBTC premium, on average, Bitcoin has a daily return of negative 23 basis points. This evidence shows that indeed lagged GBTC premium leads BTC returns.

According to Proposition 1, the GBTC predictability is likely to be stronger when the GBTC market is more price informative, risk aversion in the Bitcoin spot market is high, the Bitcoin spot market is less price informative, and the volatility of noisy trader demand is high in the Bitcoin spot market. We measure price informativeness at the monthly level separately for GBTC and BTC as the price delay R^2 measure in Hou and Moskowitz (2005) with four days of lagged daily return as independent variables. For investor risk aversion, Campbell and Cochrane (1999) and Guiso, Sapienza, and Zingales (2018) find that investors' risk aversion is usually higher after a financial crisis. Therefore, we use cryptocurrency volatility as a proxy for the aggregate investor risk aversion.[15] Bloomfield, O'hara, and Saar (2009) show that noise traders' influence is limited by the number of arbitrageurs. Makarov and Schoar (2020) show that there is strong time-series variation in Bitcoin price across different exchanges. Therefore, we use the median Bitcoin spread (in percentage) between all exchanges as a proxy for noise trading demand.

Therefore, we divide our sample into two groups based on the above characteristics in the time series at the monthly level. High (Low) Std is the sample with higher (lower) Bitcoin daily return volatility measured at the monthly level (above median). The High (Low) INF_G is the GBTC sample with a lower (higher) delay R^2 measure (above median). The High (Low) INF_B is the BTC sample with a lower (higher) delay R^2 measure (above median). High (Low) $Noise$ is the sample with a higher Bitcoin arbitrage spread (above median).

[15] Moreover, in unreported table, we can find similar results if we use crisis events such as the collapse of Terra Luna or FTX.

Electronic copy available at: https://ssrn.com/abstract=3948407

In Table 3 column (1) and column (2), we show that when the Bitcoin volatility is higher than the median, a one standard deviation increase in lagged GBTC premium change leads to a 92 bps increase in Bitcoin return. When the volatility is lower than the median, a one standard deviation increase in lagged GBTC premium change leads to a 33 basis point increase in Bitcoin return. The difference between the coefficients estimated from these two subsamples is also statistically significant. Moreover, the predictability is larger when there is larger Bitcoin price dispersion across different exchanges (column 3 and 4), higher price informativeness in the GBTC market (column 5 and 6), and lower price informativeness in the Bitcoin spot market (column 7 and 8). This evidence is consistent with the predictions of our theoretical framework.

Since the GBTC premium is a sentiment for the entire Bitcoin market, it should also be able to predict the return of other cryptocurrencies. For example, Table 4 shows that it can predict the return of Ethereum. A one standard deviation increase in lagged GBTC premium change leads to a 79 basis point increase in Ethereum return. Since Bitcoin and Ethereum together account for 80% of the cryptocurrency market, does GBTC premium predicts the return of other "smaller" cryptocurrencies? We show that GBTC also predicts the return of the whole cryptocurrency market, excluding Bitcoin and Ethereum. We show that a one standard deviation increase in lagged GBTC premium change leads to a 125 basis point increase in the whole cryptocurrency market. It leads to a similar price movement of the value-weighted returns of the top 50, 100, 200, and 300 cryptocurrencies, ranked by their market capitalization. Moreover, suppose the predictability differs for proof-of-work and proof-of-stake cryptocurrency. In that case, the predictability is more likely to reflect the view of crypto experts because they can tell the differences between various cryptocurrencies better than traditional investors.[16] However, we show that the predictability is similar for proof-of-stake and proof-of-work cryptocurrencies.

Is the hyper sentiment of traditional investors measured by GBTC premium the same as

[16]Fanti, Kogan, and Viswanath (2019), Cong, He, and Tang (2022), and Jermann (2023) show that there are fundamental differences between the pricing model of proof-of-stake and proof-of-work cryptocurrencies.

Electronic copy available at: https://ssrn.com/abstract=3948407

traditional investors' sentiment regarding the common equity market? Since the correlation between equity market sentiment and our sentiment measure is close to zero at the daily level, there is no evidence that GBTC premium/discount is the retailer's sentiment regarding the overall equity market or the whole economy.[17]

To further test this, we construct $CEFD$, the equal-weighted equity closed-end mutual fund discount at the daily level from Bloomberg. We also construct $CEFD_{Diff}$, the first difference of $CEFD$ at the daily level. In Appendix Table A.2, we show that $CEFD_{Diff}$ does not affect the predictability of changes in GBTC premium. Actually, a one standard deviation increase in $CEFD_{Diff}$ leads to a 29 bps decrease in Bitcoin return. This evidence suggests that there is a substitution effect between demand for traditional investment and cryptocurrencies from traditional investors.

4.2 Predictability by news category

To further differentiate the reason that drives the predictability of GBTC premium, we want to test the specific kind of news or new information that is most relevant for GBTC predictability. In the asset pricing literature, return predictability mostly happens around earning announcements. For Bitcoin, there are no earnings or earnings announcements. However, news related to the sentiment or fundamentals of Bitcoin, such as regulations related to cryptocurrency, new blockchain technology, and applications, or its competitor: central bank digital currency (CBDC), is reported on a daily basis.

We collect news related to blockchains and cryptocurrencies from Coindesk. CoinDesk is a news site specializing in Bitcoin, cryptocurrencies, and blockchains. It provides one of the most comprehensive coverage in terms of cryptocurrency news.[18] Table 1 shows that CoinDesk releases on average 14.8 pieces of articles per day (or News) regarding cryptocurrencies and blockchain in general.

[17]The correlation is also below 5% at the weekly and monthly levels.

[18]While Coindesk is not the only news site that provides news related to cryptocurrency, our results to robust when other crypto news sites are used.

Electronic copy available at: https://ssrn.com/abstract=3948407

We categorize news into groups and then determine the kind of news that is most relevant for GBTC predictability for a given day. To achieve this, we apply two methods, K-means clustering, and Latent Dirichlet allocation, when categorizing information into different categories. Previous literature has also used these methods in classifying text (e.g., Hanley and Hoberg, 2019; Lowry, Michaely, and Volkova, 2020). These two unsupervised machine learning methods categorize texts in different ways. K-means clustering assigns documents directly into a number of N topics based on the ex-ante input of the number N. LDA assigns each document a mixed number of N topics based on the ex-ante input of the number N. However, LDA also generates the percentage coverage of each topic in each document. For LDA analysis, we assign each news report one topic based on the topic with the highest percentage of material coverage.

Based on K-means clustering, if we first categorize news on Coindesk into two categories, the News can be categorized into trading sentiment and other news. Trading sentiment news includes keywords such as "bullish", "bearish", "Bitcoin price", and "Chart". If we categorize news into four groups, the news can be categorized into trading sentiment, crypto exchange, blockchain technology, and noise. For news related to blockchain technology, it includes keywords such as "blockchain technology", "distributed ledger", and "smart contract". If we further categorize the news into six groups, we can also find news related to Central Bank Digital Currency (CBDC) and Bitcoin Mining. For CBDC, it includes keywords such as "Libra", "Facebook", "Central Bank". Suppose we categorize news based on LDA analysis and divide it into six groups, in that case, similarly, we find six topics related to trading sentiment, bank & exchange, blockchain network, blockchain technology, user & wallet, mining & payment.

In order to determine the optimal number of classifications, We adopt K-means Silhouette Score and LDA Coherence Score. Figure 5 shows that K-means Silhouette Score suggests that seven topics are optimal while six topics are the second best. LDA Coherence Score suggests that six topics are optimal. Combining the two scores, we adopt six groups as the

Electronic copy available at: https://ssrn.com/abstract=3948407

optimal classification.

Figure 6 shows the word cloud for these different categorizations with K-means. Figure 7 shows the six categories with LDA. The categorization based on the LDA analysis is similar to the categorization based on K-means clustering. One difference is that the categorization based on LDA merges the crypto exchange and central bank into one category.

Figure 8 shows an example for each kind of news when K-means six groups of clustering are adopted (The Noise category has been taken out). News related to CBDC is about central banks' attitude toward digital currencies. News related to crypto exchanges seems to be related to regulation or ban on crypto exchanges, which echoes the possibility that custodian service and legal issues in countries that ban Bitcoin contribute to GBTC premium. News related to Bitcoin mining seems to be related to regulations on mining.

Then we proceed to conduct regression analysis based on different topics in the subsamples and check the predictability. We use the following specification to test how GBTC predictability depends on each category of news separately.

$$R_{BTC,t+1} = \alpha + \beta_1 GC_t + \beta_2 News_{t+1} + \beta_3 GC_t * News_{t+1} + \beta_4^{\top} X_i + \epsilon_i$$

Where GC_t is the lagged changes in GBTC premium, $News$ is the number of news articles within each category each day on Coindesk. E.g. when we categorize news into two groups, trading sentiment, and others, we conduct two regressions; while the GC_t remains the same in both regressions, the number of news varies depending on the K-means clustering's classification. The coefficient of central interest is β_3, which measures the interaction effect of GC_t and the number of news. If β_3 is especially significant for a specific kind of news category, it suggests that GBTC premium's predictability is particularly significant when there is more of that type of news.

In Table 5, using the K-means clustering, we find that GBTC's predictability is more significant when there is variation in news related to trading sentiment, CBDC, and crypto

Electronic copy available at: https://ssrn.com/abstract=3948407

exchanges, but not when there is news regarding blockchain technology and Bitcoin mining. This also suggests that the predictability is unrelated to Bitcoin's technological fundamentals, such as the underlying technology and mining, but more directly related to traditional investors' sentiment and Bitcoin's regulations that traditional investors can understand and process earlier than Bitcoin early adopters.

4.3 The Financialization of Cryptocurrency

The GBTC and other Bitcoin ETFs issuances represent a financialization process similar to the one for commodities as documented in Cheng and Xiong (2014), Henderson et al. (2015), Brogaard, Ringgenberg, and Sovich (2019), and Goldstein and Yang (2022). Therefore, we hypothesize that the financialization process, especially the GBTC issuance should lead to an increase in correlation with the stock market, a decrease in the volatility of cryptocurrency, and potentially an increase in price informativeness if investors in GBTC are speculators and a decrease in price informativeness if those investors are hedgers. We use a staggered difference-in-difference regression to test our hypothesis where we use the introduction of GBTC as a shock that represents cryptocurrency financialization:

$$Y_{i,t} = \alpha + \lambda_1 FinanceCurrency_t + \lambda_2 FinanceCurrency_t * Post_t + \lambda_3 FE_t + \epsilon_{i,t}$$

Where $Y_{i,t}$ is the dependent variables that are measured at the monthly level using daily data. R_{SP500} is the daily S&P 500 index return correlation with the cryptocurrency measured at the monthly level. R_{Dow} is the Dow Jones Industrial average index return. R_{Nasdaq} is the Nasdaq composite index return. R_{SMB}, R_{HML}, R_{CMA}, R_{RMW} are the Fama French five-factor return from the Kenneth French library. R_{MOM} is the daily Momentum portfolio return. Volatility is the standard deviation of the daily return measured at the monthly level for the cryptocurrency. Price Informativeness is the price delay R^2 measured as in Hou and Moskowitz (2005) with four days of lagged daily return as independent variables.

Electronic copy available at: https://ssrn.com/abstract=3948407

$FinanceCurrency_t$ equals one if the cryptocurrency has been included in the Grayscale portfolio. The two cryptocurrencies are Bitcoin and Ethereum. *Post* equals one if the cryptocurrency has been included as a closed-end fund at the time. This is months after May 2015 for Bitcoin, and June 2019 for Ethereum. The control group includes ten never treated cryptocurrencies (with the largest market capitalizations other than Bitcoin and Ethereum) as of May 2015. the FE_t is the time (month) fixed effect. We also include a linear time-trend term to control for any time trends in the dependent variables. The coefficient of central interest is λ_2, which measures the effect of being included in the closed-end fund.

Table 6 shows that after the GBTC and ETHE indexing, there is an increase of 6.3% of daily return correlation with S&P 5000 for Bitcoin and Ethereum relative to cryptocurrencies that are not included in the financialization process. There is a similar increase in correlation also with the Dow Jones index return and the Nasdaq index return. There is a 2% decrease in return volatility and a 3.2% decrease in the price delay R^2 measure which suggests an increase in price informativeness (Hou and Moskowitz, 2005). There has also been an increase in correlation with the SMB portfolio, and a decrease in correlation with the HML, RMW, and CMA portfolios. This may have an implication that traditional investors believe cryptocurrencies to be similar to a small, high-growth, and less profitable firm with less investment. Overall, this result suggests that the financialization process indeed transmits conventional or equity investors' view of cryptocurrency to the equilibrium price of the cryptocurrency.

4.4 Bitcoin fundamental and supply factors

Most traditional investors do not possess the technical knowledge to open their Bitcoin wallets, run nodes, or store Bitcoin on hard drives. Traditional investors also cannot spend Bitcoin easily as the transaction cost (gas fee) is high, and at least it takes ten minutes for a transaction to be confirmed (Foley, Karlsen, and Putniņš, 2019). Most traditional investors invest in Bitcoin due to speculative motives or their belief that blockchain technology will

Electronic copy available at: https://ssrn.com/abstract=3948407

be successful. To further show that the traditional investor sentiment drives our results, we hypothesize that changes in GBTC premium cannot predict actual Bitcoin usage and Bitcoin supply factors such as the cost of mining.

Liu and Tsyvinski (2021) find that address growth, transaction growth, wallet number growth, and payment growth are the four factors that are most related to the actual usage of Bitcoin. Address growth is the active Bitcoin address growth. Transaction growth is the percentage of Bitcoin transaction number growth. Payment growth is the number of Bitcoin payment growth. Wallet growth is the wallet user growth. PC real growth is the first principal component of these four Bitcoin fundamentals. Table 7 panel A shows that GBTC premium cannot predict the Bitcoin usage variables related to Bitcoin. This evidence suggests that the GBTC premium reflects the sentiment of traditional investors instead of the blockchain communities.

When our sentiment measure cannot predict Bitcoin usage variables, is it possible that GBTC premium is related to supply factors such as mining pool competition, electricity prices, or prices of mining machines such as ASIC or GPU? Cong et al. (2021a) show that mining pools of Bitcoin can be highly centralized yet still reach consensus. Easley et al. (2019) show that mempool size captures the waiting time of Bitcoin blockchain users. However, we cannot find any correlation between GBTC premium and the Herferfindal Index of the mining pool or the mempool size. The dynamic of fees charged by Bitcoin miners is a lagged factor instead of a leading factor of investor sentiment. Moreover, change in block reward is an essential factor for the supply of Bitcoin. We also don't find evidence that block-reward changes the predictability of GBTC.

For electricity price and mining equipment price, since they are slow-moving factors that do not change at high frequency, we use the utility sector stock index returns (MSCI China Utilities ETF) in China, and the US semiconductor index returns (iShares Semiconductor ETF) to proxy for supply factors. However, Table 7 panel B shows that changes in GBTC premium are also not correlated with these factors. Therefore, we don't find evidence that

Electronic copy available at: https://ssrn.com/abstract=3948407

our sentiment measure is casually driven by or contemporaneously correlated with supply factors at a daily frequency.

4.5 Bitcoin factor return in the cross-section

Baker and Wurgler (2006) show that highly volatile stocks, hard to value and hard to arbitrage stocks, are more subject to sentiment's influence. Therefore, we hypothesize that similar results hold for cryptocurrencies, i.e., hard-to-value cryptocurrencies are more subject to the influence of investors' sentiment. Liu et al. (2021) show that cryptocurrency size and momentum are the most critical factors that price cryptocurrencies in the cross-section. Therefore, we dynamically sort cryptocurrencies into quintiles on a daily basis and calculate the value-weighted return for each portfolio.[19] Then, we regress each portfolio's return on lagged changes in GBTC premium. Following our initial hypothesis, the return of the group which is harder to be valued should have a higher loading on lagged Bitcoin return.

We use four factors to test the above hypothesis. The first one is "age", or the number of days since a cryptocurrency was first listed on a crypto exchange. According to our hypothesis, "older" cryptocurrency should be easier to be valued, while "younger" cryptocurrency should be harder to be valued. The second one is the size. A larger cryptocurrency should be easier to be valued when a smaller cryptocurrency should be harder to be valued. Liu et al. (2021) use total market capitalization, price, and maximum price during last week as proxies for size. Similarly, we also use these three variables to proxy for size. In our sample, we exclude cryptocurrencies with less than one year of age and less than one million dollars in terms of total market capitalization.

Table 8 shows that younger cryptocurrencies indeed have a higher sensitivity to returns of GBTC premium. For the return difference between the portfolio with the highest and lowest quintile, the coefficient on lagged changes in GBTC is 0.037. This suggests that a one standard deviation movement of GBTC premium on a daily level generates a 35 basis point

[19] The results remain similar when we sort on a weekly basis.

Electronic copy available at: https://ssrn.com/abstract=3948407

return difference between the group of cryptocurrencies with younger age and the group of older cryptocurrencies. Table 8 also shows that cheaper cryptocurrencies have a higher sensitivity to lagged changes in GBTC premium. For the return difference between the portfolio with the highest and lowest quintile ranked by its current price, the coefficient on lagged changes in GBTC is 0.072. This suggests that a one standard deviation movement of lagged changes in GBTC premium generates a 68 basis point return difference between the group of cryptocurrencies with lower prices and the group of cryptocurrencies with higher prices. Using total market capitalization as a proxy for size, we obtain similar results but less statistical significance. The overall results suggest that changes in GBTC premium indeed measure investor sentiment.

4.6 Measuring difficulties of valuation using textual analysis

While Size and Age are proxies for difficulties of valuation and arbitrages, with textual analysis, we can construct alternative measures to corroborate our results further. Most cryptocurrencies have a "whitepaper" that describes their technical details regarding blockchains. Specifically, we use textual analysis on cryptocurrency whitepapers to measure each cryptocurrency's easiness for valuation and arbitrage.

We use five textual analysis variables to measure the difficulty of valuing a cryptocurrency. We group cryptocurrencies into terciles based on their whitepaper's readability. Low refers to the value-weighted daily return of the group that has the easiest readability. High refers to the value-weighted daily return of the group that has the hardest readability. Following the methodology of Gunning (1969), Loughran and McDonald (2011), and Loughran and McDonald (2014), the Gunning-Fog index measures the readability based on the whitepaper's total length, word complexity, and sentence length. Doc Length is the MB size (excluding Graphs) of the cryptocurrency whitepaper. We also use the frequencies of weak modal words (could, might, possible, depending) and uncertainty words (approximate, depend, fluctuate, variability) in the whitepaper as measures of the easiness of arbitrage. Cryptocurrency

Electronic copy available at: https://ssrn.com/abstract=3948407

technical words (Lu, 2018) measure the frequency that blockchain technical words show up in the whitepaper.

In Table 9, we show the predictability of GBTC premium separately for the group with low readability and high readability. We find that a one standard deviation increase in lagged GBTC premium change leads to a 2.15% increase in value-weighted return of cryptocurrencies with low readability, measured by the Gunning Fog index. However, a one standard deviation increase in lagged GBTC premium change only leads to a 1.27% increase in value-weighted return of cryptocurrencies with high readability. We also get similar results when we use the MB size of the whitepaper, the frequency of uncertainty words, weak modal words, and technical words. Therefore, our results are robust to alternative measures of readability, and suggest that GBTC premium is indeed a measure of investor sentiment.

4.7 Predictability across different cryptocurrency-exchanges

Makarov and Schoar (2020) show that the Bitcoin market is highly decentralized. Bitcoin price levels go up and down disproportionally in some markets but not in other markets. An alternative hypothesis is that there exists a positive and contemporaneous correlation between changes in GBTC premium and Bitcoin returns in some markets that react to information first. Therefore, the GBTC premium mechanically predicts BTC returns in other markets, which respond to news more slowly. If this alternative hypothesis is true, we expect to find that GBTC premium can only predict market return in some markets but not in others.

In Table 10, We show that lagged changes in GBTC premium predict BTC returns in all major markets around the globe: in Asian, European, and North American cryptocurrency markets with different currency denominations. This evidence suggests that our result is not caused by a potential autocorrelation structure among Bitcoin exchanges. For the economic magnitude, a one standard deviation increase in lagged GBTC premium change leads to a 0.5% to 3.2% increase in Bitcoin return in different crypto exchanges.

Electronic copy available at: https://ssrn.com/abstract=3948407

4.8 Market microstructure

An alternative explanation of our results is that could our results be driven by the market impact of GBTC buying and selling Bitcoin. We find this unlikely to be true. First, since GBTC is structured as a closed-end fund, it rarely sells Bitcoin. The changes in GBTC premium typically reflect the secondary market trading of GBTC shares, holding the number of Bitcoin in GBTC constant. If the declines in GBTC premium predict negative Bitcoin returns, then the results cannot be driven by price impact. Therefore, we test positive and negative GBTC predictability separately. Second, we can use the changes in BTC holding by GBTC to directly test the market impact explanation. We obtain GBTC's BTC holding from the daily Twitter announcement of GBTC's AUM, divided by the corresponding Bitcoin price. Our data of GBTC BTC holding dates back to March 2018.

Table 11 shows that for a one standard deviation decrease in lagged GBTC premium, the Bitcoin price drops by 94 basis points. During the 2018 to 2021 subsample period, a one standard deviation decrease in lagged GBTC premium leads to a 1.49% drop in Bitcoin price. The corresponding upside predictability of the GBTC premium is two-thirds of the downside predictability. This result suggests that our results are unlikely to be driven mechanically by market impact. Moreover, we show that overall changes in BTC holding by GBTC cannot predict BTC return. Neither positive changes in BTC holding nor negative changes can predict BTC return. This suggests that the market impact of GBTC is quickly absorbed and cannot be used to predict future BTC returns.

Next, to further show that our result is not driven by market microstructure friction, we divide our sample into two groups based on the following characteristics in the time series at the daily level. High (Low) Vol is the sample with higher (lower) Bitcoin dollar trading volume (above median). The High (Low) move is the sample with a higher (lower) absolute value of GC_t (above median). High (Low) N_News is the sample with a higher (lower) number of News regarding Bitcoin or Blockchain on CoinDesk (above median). High (Low) Abnormal is the sample with a higher abnormal number of News (above median). The

Electronic copy available at: https://ssrn.com/abstract=3948407

Abnormal number of News is defined as the number of News regarding Bitcoin or Blockchain on CoinDesk minus the one-week average number of News before that day.

If the predictability remains when the volume of trading is very high, then our result should not be caused by market microstructure-based explanations. Similarly, suppose the predictability remains when the number of news is high, or when the expected move of Bitcoin is largely due to the large absolute value of changes in lagged GBTC premium. In that case, our result should not be caused by market microstructure-based explanations.

In Table 12 column (1) and column (2), we show that when the trading volume is higher than the median of the dollar volume of Bitcoin trading between 2015 and 2021, a one standard deviation increase in lagged GBTC premium change leads to a 1.1% increase in Bitcoin return. When the trading volume is lower than the median, a one standard deviation increase in lagged GBTC premium change leads to a 35 basis point increase in Bitcoin return. Moreover, the predictability is larger when the absolute value of changes in GBTC premium, or the absolute value of expected changes in BTC return is larger.

In Table 12 columns (5) to (8), We also show that the predictability of GBTC premium becomes more significant when there is more news regarding Bitcoin.[20] When the number of news is high than the median of the number of news regarding Bitcoin between 2015 and 2021, a one standard deviation increase in lagged GBTC premium change leads to a 1.05% increase in Bitcoin return. When the number of news is low than the median, a one standard deviation increase in lagged GBTC premium change leads to a 0.49% increase in Bitcoin return. These patterns in the trading volume, variation in news, and expected Bitcoin return suggest that liquidity or stale prices cannot explain our results.

[20]When there is more "fundamental" news regarding Bitcoin, it is more likely that there is a higher proportion of informed trading and a lower proportion of noisy trading (a lower value of $\tau_{x,s}$). Therefore, this result also supports the empirical predictions derived from Proposition 1.

Electronic copy available at: https://ssrn.com/abstract=3948407

4.9 Trading strategy

We also construct a portfolio that buys Bitcoin when the lag GBTC premium is positive and sells short Bitcoin when the lag GBTC premium is negative. More specifically, the strategy initializes a buy of Bitcoin when changes in GBTC premium is above the 55 percentile of the past rolling one-year changes in GBTC premium. The strategy initializes a sell when changes in GBTC premium are below the 45 percentile of the past rolling one-year changes in GBTC premium. In Table 13, we show that this strategy has a significant alpha of 38 basis points on a daily basis, with close to zero and slightly negative loading for Bitcoin. Moreover, this strategy has no exposure to the Fama-French five-factor model. It does not have exposure to the momentum and has slight negative exposure to the reversal strategy. When we regress this strategy's return on Fama-French 50 industries, it has slight negative exposure to the Gun industry and positive exposure to the real estate industry.

What happens when this strategy incurs transaction costs? To avoid daily re-balancing, if the consecutive signal is buying or selling Bitcoin, we assume the portfolio hold on the position. Therefore, the strategy only incurs a transaction cost when the long/short position has been closed when the opposite signal appears. We assume four transaction cost levels: 10 bps, 20 bps, 30 bps, and 40 bps. Figure 9 shows the corresponding buy-and-hold return from January 1st 2016, to July 1st 2021. One dollar invested in this buy-and-hold strategy generates 13.5 dollars if there is no transaction cost, 7 dollars when the transaction cost is 10bps, 3 dollars when the transaction cost is 20bps, and does not make a profit when the transaction cost is 40 bps. Therefore, we find that although the strategy has a very high daily alpha, the high round-trip trading cost of Bitcoin is also usually large enough to offset the profit from this strategy.

4.10 Pump and dump scheme and price manipulation

Gandal, Hamrick, Moore, and Oberman (2018); Hamrick, Rouhi, Mukherjee, Feder, Gandal, Moore, and Vasek (2018); Kamps and Kleinberg (2018); Xu and Livshits (2019); Li, Shin,

Electronic copy available at: https://ssrn.com/abstract=3948407

and Wang (2020) show that pump and dump scheme can drive the price of cryptocurrencies. Does the GBTC premium coincide with any price manipulation?

However, we do not find a correlation between the Bitcoin pump and dump time period and the high sentiment time period. Moreover, our results still persist after ruling out trading days when there is documented potential manipulation of Bitcoin price either through a pump and dump scheme or Tether manipulation. Moreover, the pump and dump scheme and Tether manipulation happen at an hourly level and usually quickly reverse (Griffin and Shams, 2020; Gandal et al., 2018). Therefore, price manipulations are unlikely to drive Bitcoin prices at a daily level over a long time series, and our results are unlikely to be caused by the pump and dump scheme.

5 Robustness Check

Since Bitcoin's return dynamic could happen at the hourly level, we also check GBTC's predictability on Bitcoin at an hourly level. If the positive predictability is entirely caused by the market impact of the GBTC, we would expect the positive effect to concentrate in the first few hours or clustered around a few hours when GBTC is buying the Bitcoin. Figure 10 shows the GBTC Premium's predictability of Bitcoin price at the hourly level. The X-axis is the number of hours from the measure of GBTC premium change. The Y-axis is the coefficient of lagged changes in GBTC premium on Bitcoin hourly return multiplied by one thousand. Instead, we find that the predictability is scattered around the whole day, with reversals happening in between. Therefore, the hour-level analysis does not support a market microstructure interpretation.

To test the investor base that drives the Grayscale premium, we test the predictability of the Grayscale premium for each weekday. Appendix Table A.3 shows that the effect is stronger on Monday, Thursday, and Friday, this is different from the patterns documented for Bitcoin in Liu and Tsyvinski (2021) where the Bitcoin return is highest on Tuesday and

Electronic copy available at: https://ssrn.com/abstract=3948407

lowest on Friday, but more consistent with the pattern we see in the stock market. This is evidence that this sentiment reflects the influence of equity investors on the cryptocurrency market despite the market segmentation documented in the previous literature. This result is also consistent with the fact that GBTC itself is trading on an equity OTC market.

The Bitcoin market becomes relatively more liquid as time evolves. And it becomes more liquid relative to the GBTC market. Therefore, if the result is entirely driven by liquidity-based explanations, we would expect the predictability to become smaller and smaller over time. However, the predictability actually grows over time. Appendix Table A.3 shows that in 2016, a one standard deviation increase in lagged GBTC premium change leads to a 36 basis point increase in Bitcoin return. In 2020, a one standard deviation increase in lagged GBTC premium change leads to a 1.38% increase in Bitcoin return. The predictability coefficient has been growing steadily over the years. Therefore our results cannot be entirely driven by liquidity-related issues.

Another robustness check we conduct is to calculate the Bitcoin return using the difference between the ask price and lagged bid price, all divided by the lagged bid price. This approach considers the bid-ask spread and avoids any correlation due to the bid-ask spread structure. We collect bid-ask prices from Bitfinex. However, the results remain similar, and the lagged change in GBTC premium can still predict Bitcoin return after accounting for the bid-ask spread.

6 Conclusion

We show that Bitcoin's daily return can be predicted by closed-end Bitcoin fund premiums and discount changes. The sentiment measure's predictability is more significant than the lagged Bitcoin return and sentiment measures like Google or Twitter Search. A one standard deviation increase in lag GBTC premium change leads to a 0.8% increase in Bitcoin return. Even though Bitcoin price is decentralized and there is large price dispersion for Bitcoin

Electronic copy available at: https://ssrn.com/abstract=3948407

across different exchanges, we show that changes in GBTC premium predict Bitcoin return across major cryptocurrencies around the world. The GBTC premium is likely to reflect the hyper sentiment of traditional investors that are new to cryptocurrency investment. Such hyper sentiment may be viewed as a shot in the arm among cryptocurrency communities regarding societal recognition and endorsement and boosts the Bitcoin price.

To further shed light on the source of this predictability, we apply the K-means clustering and Latent Dirichlet allocation method when categorizing news into different kinds to overcome the problem that Bitcoin has no earning announcements like in the stock market. We find that GBTC's predictability is more significant when there is a lot of variation in news related to trading sentiment, including keywords such as "bullish", "bearish", "Bitcoin price", and "Chart". GBTC's predictability is also sensitive to news regarding central bank digital currencies (CBDC) and crypto exchange regulations, but not to news regarding blockchain technology and Bitcoin mining. This suggests that the predictability is more related to traditional investors' sentiment, their concerns about government regulation on cryptocurrencies, and Bitcoin's competitor: CBDC.

This paper is also related to the large literature regarding stock return predictability.[21] Our results are ex-post surprising because followers' or traditional investors' sentiment leads the return of Bitcoin determined on crypto exchanges by Bitcoin early adopters and specialists. Our results reveal the distinct nature of Bitcoin and cryptocurrencies in general as an investment category. We also show the importance of investor sentiment beyond traditional asset classes.

[21] More recently, Hong, Torous, and Valkanov (2007) show that some industry return leads the whole U.S. stock market return. Driesprong et al. (2008) show that oil price predicts aggregate stock market return. Menzly and Ozbas (2010) show that stocks with supplier relationships can predict other stocks. Rapach, Strauss, and Zhou (2013) show that the lagged U.S. stock return predicts stock return in other markets.

Electronic copy available at: https://ssrn.com/abstract=3948407

References

Ahn, D.-H., Boudoukh, J., Richardson, M., Whitelaw, R. F., 2002. Partial adjustment or stale prices? Implications from stock index and futures return autocorrelations. The Review of Financial Studies 15, 655–689.

Badrinath, S. G., Kale, J. R., Noe, T. H., 1995. Of shepherds, sheep, and the cross-autocorrelations in equity returns. The Review of Financial Studies 8, 401–430.

Baker, M., Wurgler, J., 2004. A catering theory of dividends. The Journal of Finance 59, 1125–1165.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The Journal of Finance 61, 1645–1680.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. Journal of economic perspectives 21, 129–152.

Baker, M., Wurgler, J., Yuan, Y., 2012. Global, local, and contagious investor sentiment. Journal of Financial Economics 104, 272–287.

Bloomfield, R., O'hara, M., Saar, G., 2009. How noise trading affects markets: An experimental analysis. The Review of Financial Studies 22, 2275–2302.

Breen, W., Glosten, L. R., Jagannathan, R., 1989. Economic significance of predictable variations in stock index returns. The Journal of finance 44, 1177–1189.

Brogaard, J., Ringgenberg, M. C., Sovich, D., 2019. The economic impact of index investing. The Review of Financial Studies 32, 3461–3499.

Campbell, J. Y., Cochrane, J. H., 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of political Economy 107, 205–251.

Cheng, I.-H., Xiong, W., 2014. Financialization of commodity markets. Annu. Rev. Financ. Econ. 6, 419–441.

Chopra, N., Lee, C. M., Shleifer, A., Thaler, R. H., 1993. Yes, discounts on closed-end funds are a sentiment index. Journal of Finance pp. 801–808.

Chordia, T., Swaminathan, B., 2000. Trading volume and cross-autocorrelations in stock returns. The Journal of Finance 55, 913–935.

Cong, L. W., He, Z., Li, J., 2021a. Decentralized mining in centralized pools. The Review of Financial Studies 34, 1191–1235.

Cong, L. W., He, Z., Tang, K., 2022. Staking, token pricing, and crypto carry. SSRN Working Paper .

Cong, L. W., Li, Y., Wang, N., 2021b. Tokenomics: Dynamic adoption and valuation. The Review of Financial Studies 34, 1105–1155.

Electronic copy available at: https://ssrn.com/abstract=3948407

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. The Journal of Finance 66, 1461–1499.

Driesprong, G., Jacobsen, B., Maat, B., 2008. Striking oil: Another puzzle? Journal of financial economics 89, 307–327.

Easley, D., O'Hara, M., Basu, S., 2019. From mining to markets: The evolution of bitcoin transaction fees. Journal of Financial Economics 134, 91–109.

Fama, E. F., Schwert, G. W., 1977. Asset returns and inflation. Journal of Financial Economics 5, 115–146.

Fanti, G., Kogan, L., Viswanath, P., 2019. Economics of proof-of-stake payment systems. SSRN Working Paper .

Foley, S., Karlsen, J. R., Putniņš, T. J., 2019. Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? The Review of Financial Studies 32, 1798–1853.

Gabaix, X., 2014. A sparsity-based model of bounded rationality. The Quarterly Journal of Economics 129, 1661–1710.

Gandal, N., Hamrick, J., Moore, T., Oberman, T., 2018. Price manipulation in the bitcoin ecosystem. Journal of Monetary Economics 95, 86–96.

Goldstein, I., Yang, L., 2022. Commodity financialization and information transmission. The Journal of Finance 77, 2613–2667.

Griffin, J. M., Shams, A., 2020. Is bitcoin really untethered? The Journal of Finance 75, 1913–1964.

Guiso, L., Sapienza, P., Zingales, L., 2018. Time varying risk aversion. Journal of Financial Economics 128, 403–421.

Gunning, R., 1969. The fog index after twenty years. Journal of Business Communication 6, 3–13.

Hamrick, J., Rouhi, F., Mukherjee, A., Feder, A., Gandal, N., Moore, T., Vasek, M., 2018. The economics of cryptocurrency pump and dump schemes. CEPR Working Paper .

Hanley, K. W., Hoberg, G., 2019. Dynamic interpretation of emerging risks in the financial sector. The Review of Financial Studies 32, 4543–4603.

Henderson, B. J., Pearson, N. D., Wang, L., 2015. New evidence on the financialization of commodity markets. The Review of Financial Studies 28, 1285–1311.

Hong, H., Torous, W., Valkanov, R., 2007. Do industries lead stock markets? Journal of Financial Economics 83, 367–396.

Electronic copy available at: https://ssrn.com/abstract=3948407

Hou, K., Moskowitz, T. J., 2005. Market frictions, price delay, and the cross-section of expected returns. The Review of Financial Studies 18, 981–1020.

Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. The Review of Financial Studies 28, 791–837.

Jermann, U. J., 2021. Non-sovereign stores of value. SSRN Working Paper .

Jermann, U. J., 2023. A macro finance model for proof-of-stake ethereum. SSRN Working Paper .

Kamps, J., Kleinberg, B., 2018. To the moon: defining and detecting cryptocurrency pump-and-dumps. Crime Science 7, 1–18.

Lee, C., Shleifer, A., Thaler, R. H., 1990. Anomalies: Closed-end mutual funds. Journal of economic Perspectives 4, 153–164.

Lee, C. M., Shleifer, A., Thaler, R. H., 1991. Investor sentiment and the closed-end fund puzzle. The Journal of Finance 46, 75–109.

Li, T., Shin, D., Wang, B., 2020. Cryptocurrency pump-and-dump schemes. Available at SSRN 3267041 .

Liu, Y., Tsyvinski, A., 2021. Risks and returns of cryptocurrency. The Review of Financial Studies 34, 2689–2727.

Liu, Y., Tsyvinski, A., Wu, X., 2021. Common risk factors in cryptocurrency. The Journal of Finance .

Lo, A. W., MacKinlay, A. C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. The Review of Financial Studies 1, 41–66.

Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures. the Journal of Finance 69, 1643–1671.

Loughran, T., McDonald, B., 2016. Textual analysis in accounting and finance: A survey. Journal of Accounting Research 54, 1187–1230.

Lowry, M., Michaely, R., Volkova, E., 2020. Information revealed through the regulatory process: Interactions between the SEC and companies ahead of their IPO. The Review of Financial Studies 33, 5510–5554.

Lu, F., 2018. The numeraire effect in initial coin offerings. SSRN Working Paper .

Makarov, I., Schoar, A., 2020. Trading and arbitrage in cryptocurrency markets. Journal of Financial Economics 135, 293–319.

Menzly, L., Ozbas, O., 2010. Market segmentation and cross-predictability of returns. The Journal of Finance 65, 1555–1580.

Electronic copy available at: https://ssrn.com/abstract=3948407

Pástor, L., Stambaugh, R. F., 2009. Predictive systems: Living with imperfect predictors. The Journal of Finance 64, 1583–1628.

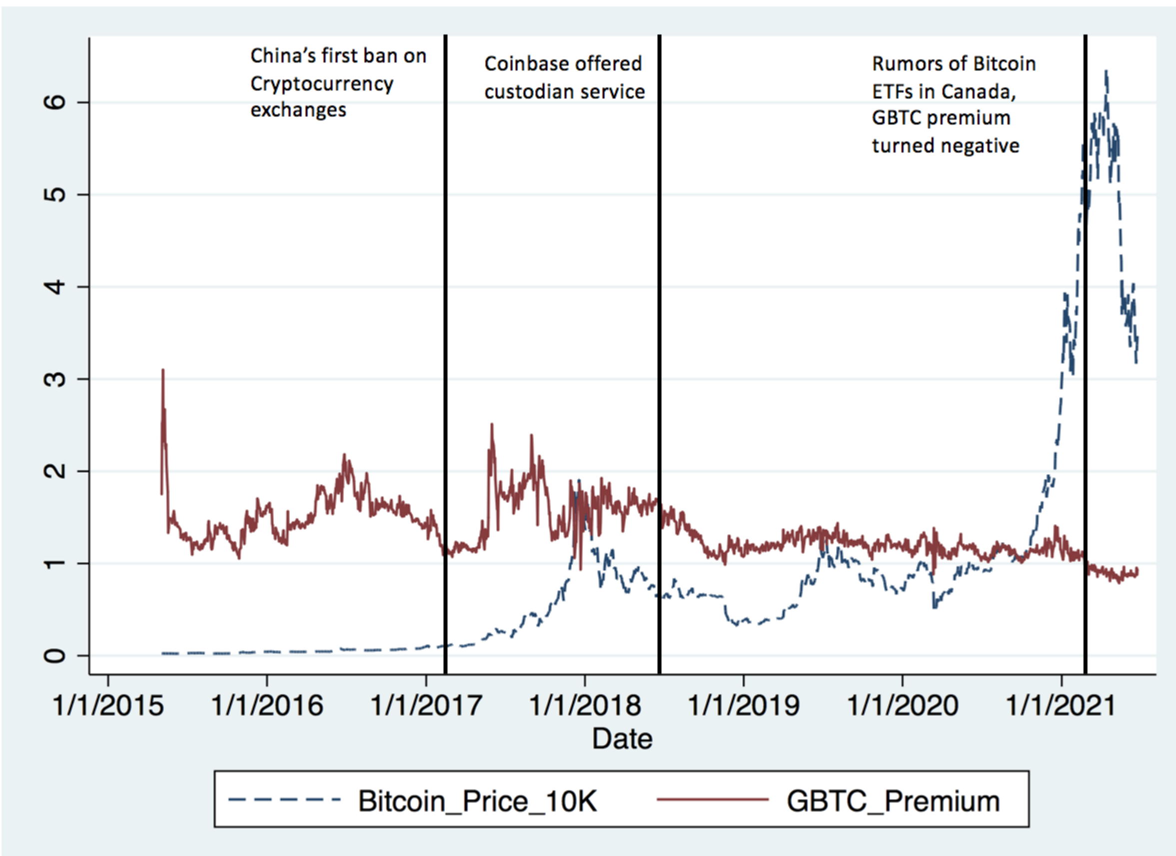
Rapach, D. E., Strauss, J. K., Zhou, G., 2013. International stock return predictability: What is the role of the united states? The Journal of Finance 68, 1633–1662.

Sias, R. W., Starks, L. T., 1997. Return autocorrelation and institutional investors. Journal of Financial Economics 46, 103–131.

Xu, J., Livshits, B., 2019. The anatomy of a cryptocurrency pump-and-dump scheme. In: *28th {USENIX} Security Symposium ({USENIX} Security 19)*, pp. 1609–1625.

Electronic copy available at: https://ssrn.com/abstract=3948407

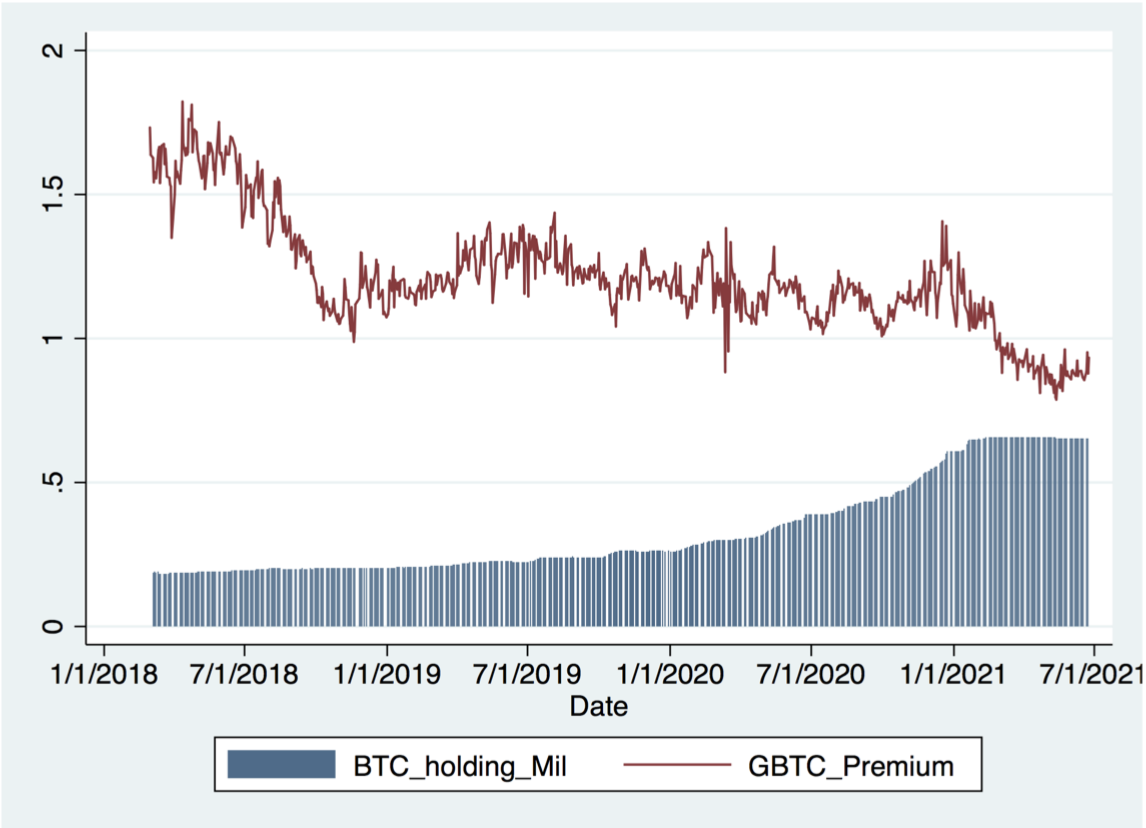
Figure 1: GBTC Premium/Discount and Bitcoin Price



Note: This figure shows the GBTC Premium/Discount (solid red line) and Bitcoin Price (dashed blue line).

Electronic copy available at: https://ssrn.com/abstract=3948407

Figure 2: GBTC Premium and GBTC Bitcoin holding



Note: This figure shows the GBTC Premium/Discount (solid red line) and GBTC Bitcoin holding in millions (blue bar) between March 2018 and July 2021.

Electronic copy available at: https://ssrn.com/abstract=3948407

Figure 3: GBTC top ten investors

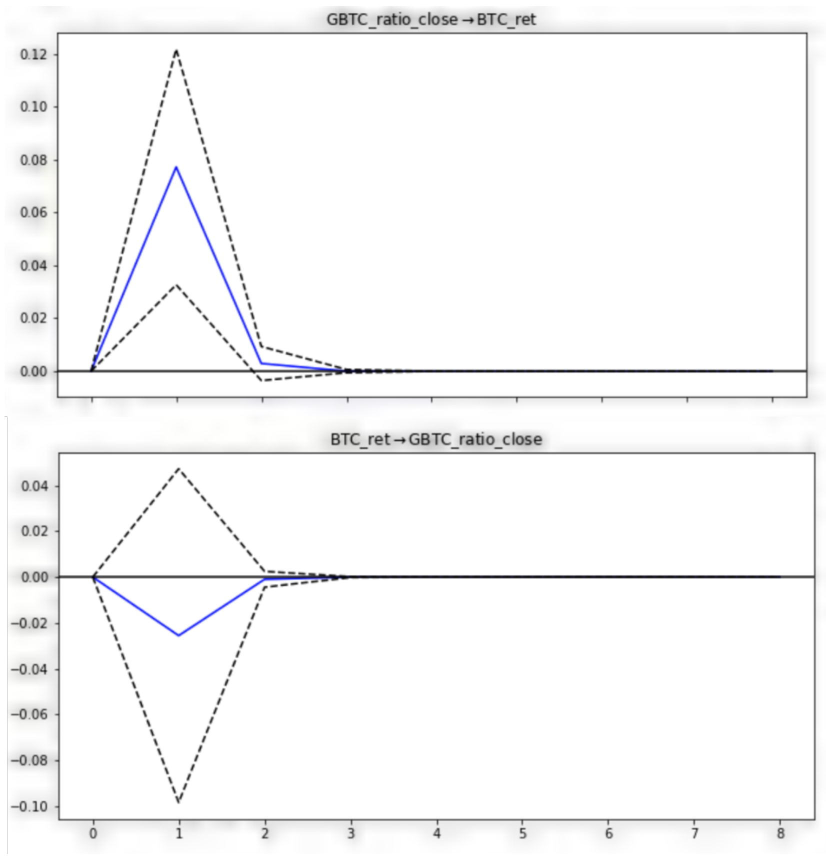
Top 10 Owners of Grayscale Bitcoin Trust (BTC)

Stockholder	Stake	Shares owned	Total value ($)
ARK Investment Management LLC	1.25%	8,659,237	327,059,381
Horizon Kinetics Asset Management...	0.34%	2,316,550	87,496,094
Miller Value Partners LLC	0.22%	1,500,000	56,655,000
Simplify Asset Management, Inc.	0.05%	309,907	11,705,187
Rg Liquid Alts LP	0.04%	255,185	9,638,337
Scopus Asset Management LP	0.03%	179,103	6,764,720
Rothschild Investment Corp.	0.02%	141,405	5,340,867
tru Independence Asset Management...	0.02%	124,378	4,697,757
Parkwood LLC	0.02%	125,000	4,721,250
IFP Advisors, Inc.	0.02%	120,896	4,566,242

Note: This figure shows the GBTC's top ten investors as of 2021 Oct 8th (Source: CNN).

Electronic copy available at: https://ssrn.com/abstract=3948407

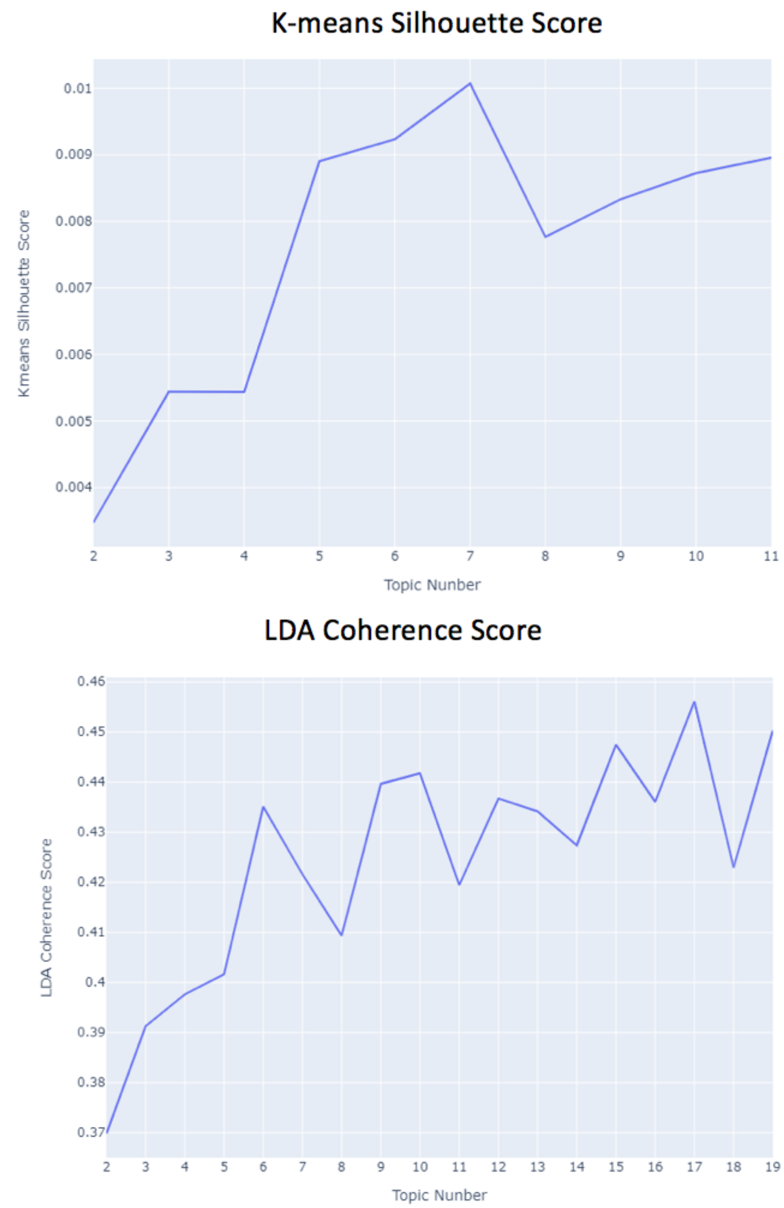
Figure 4: VAR analysis



Note: This figure shows the impulse responses of Bitcoin return (BTC_ret) to changes in GBTC premium (GBTC_ratio_close), and changes in GBTC premium to Bitcoin return at the daily level.

Electronic copy available at: https://ssrn.com/abstract=3948407

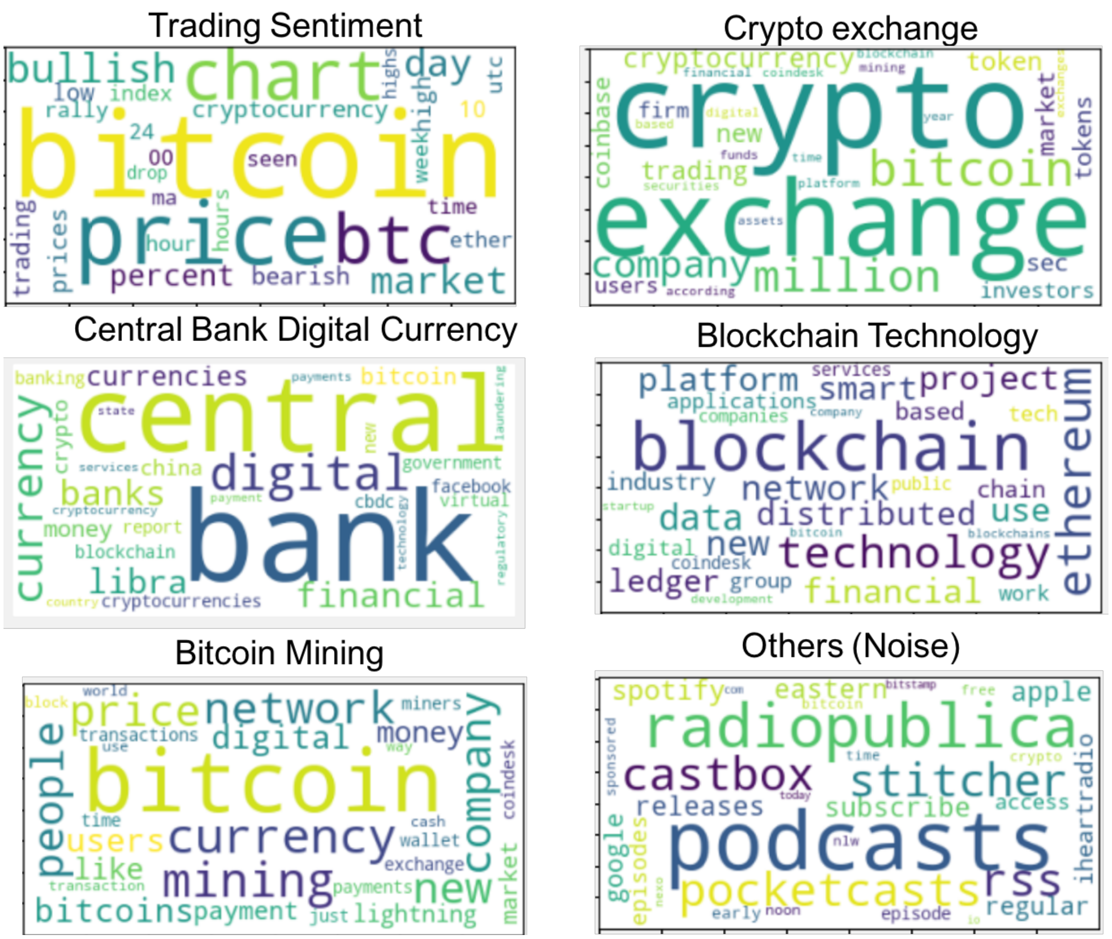
Figure 5: Optimal number of news categories



Note: This figure shows the K-means Silhouette Score for the K-means clustering and the LDA Coherence Score for the LDA methods. The X-axis shows the number of topics and the Y-axis shows the score.

Electronic copy available at: https://ssrn.com/abstract=3948407

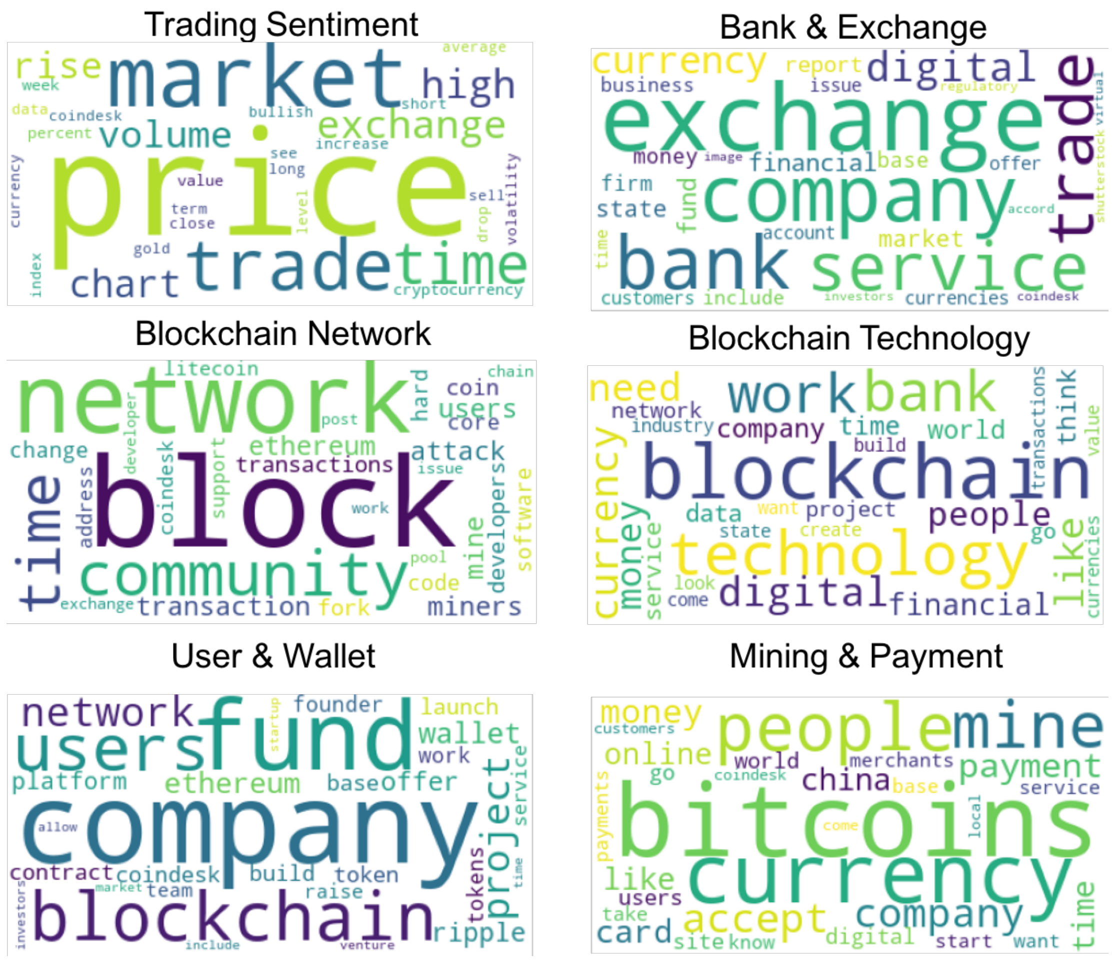
Figure 6: Word Cloud for different news category (K-means)



Note: This figure shows the word cloud (keywords) for six different kinds of News related to cryptocurrency, Bitcoin, and blockchain from Coindesk. The News is categorized using the K-means clustering into News regarding trading sentiment, crypto exchange, central bank digital currency, blockchain technology, Bitcoin mining, and others (noise).

Electronic copy available at: https://ssrn.com/abstract=3948407

Figure 7: Word Cloud for different news category (LDA)



Note: This figure shows the word cloud (keywords) for six different kinds of News related to cryptocurrency, Bitcoin, and blockchain from Coindesk. The News is categorized using the LDA into News regarding trading sentiment, bank & exchange, blockchain network, blockchain technology, user & wallet, and mining & payment.

Electronic copy available at: https://ssrn.com/abstract=3948407

Figure 8: Sample of news in different categories of K-means clustering

Trading Sentiment

Market Wrap: Bitcoin Bullish Sentiment Fades as Selling Abates

Analysts expect sentiment to become more normal for September as bitcoin's price consolidates.

CBDC related News

Bank of England: Any UK CBDC Will Be 'Tens of Thousands' Times More Efficient Than Bitcoin

The fintech lead at U.K.'s central bank urged eco-conscious citizens not to "throw the blockchain baby out with the bitcoin bathwater."

Crypto Exchanges

US Sanctions Enforcer Blacklists a Crypto Exchange for First Time

The Treasury's Office of Foreign Assets Control labeled Suex.io a "specially designated national," putting the exchange in a category with suspected terrorists.

Blockchain Technology

Sure, Bitcoin's Price Is Cool, but Bitcoin's Technology Is Hot

Getting bitcoin to the moon requires some serious technology.

Bitcoin Mining

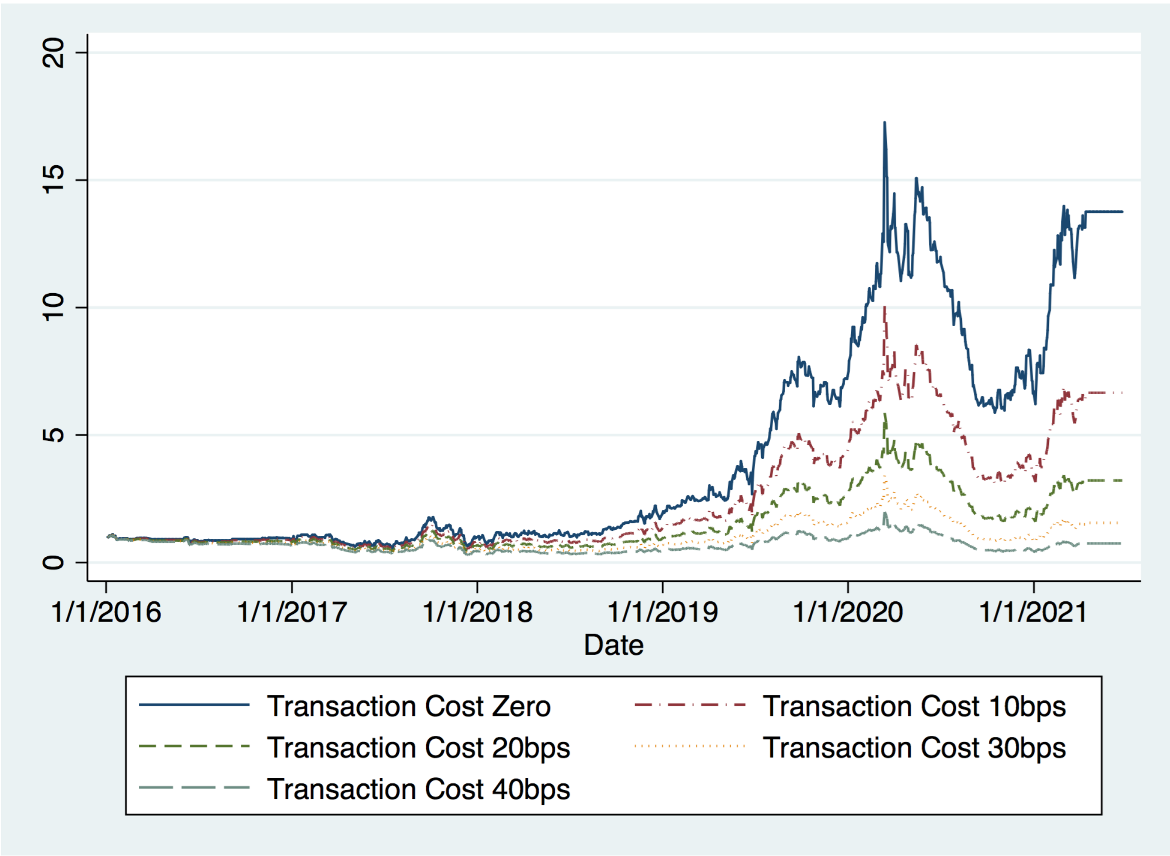
Alibaba to Stop Selling Crypto Mining Machines

The company said it was following Friday's PBoC guidelines, but also taking note of global crypto regulation instability.

Note: This figure shows five samples of news that are being categorized into five different categories of News using the K-means clustering. The news that is related to trading sentiment, crypto exchange, central bank digital currency, blockchain technology, and Bitcoin mining.

Electronic copy available at: https://ssrn.com/abstract=3948407

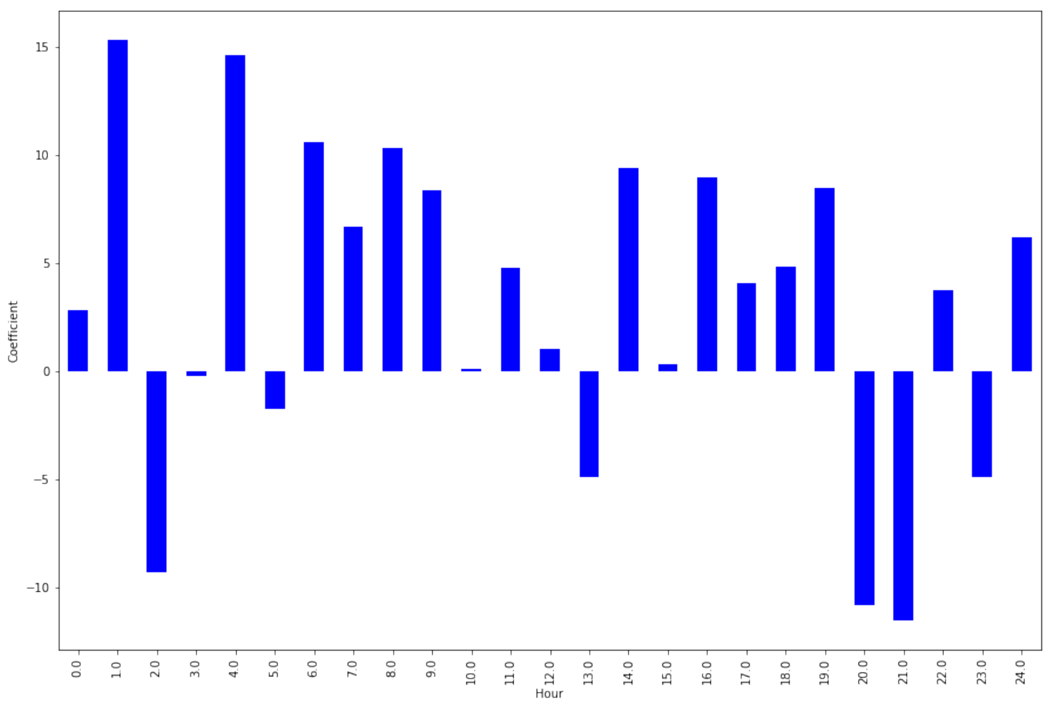
Figure 9: GBTC Strategy with different transaction costs



Note: This figure shows the GBTC (BTC) strategy with four different transaction costs, 10, 20, 30, and 40 basis points. The strategy initializes a buy of Bitcoin when changes in GBTC premium is above the 55 percentile of the past rolling one-year changes in GBTC premium. The strategy initializes a sale of Bitcoin when changes in GBTC premium are below the 45 percentile of the past rolling one-year changes in GBTC premium. The strategy only incurs a transaction cost when the long/short position closes when the opposite signal (GBTC value) appears.

Electronic copy available at: https://ssrn.com/abstract=3948407

Figure 10: GBTC hourly predictability



This figure shows the GBTC_Premium's predictability of Bitcoin price at the hourly level. The X-axis is the number of hours after 4 pm US eastern time. The Y-axis shows the regression coefficients of Bitcoin hourly return on changes in GBTC premium multiplied by one thousand.

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 1: Summary statistics

This table includes Grayscale Bitcoin Trust Fund (GBTC) data from 2015 May to 2021 June. $GBTC$ is the daily unit price of GBTC. $GBTC_{Premium}$ is the dollar price of GBTC divided by the dollar price of BTC per GBTC share. Each GBTC share contains 0.001 BTC at its inception. Since Coinbase deducts a 2% management fee from GBTC's holding of Bitcoin directly, each share of GBTC contains less BTC over the years. GC is the first difference of $GBTC_{Premium}$ at the daily level. News is the number of news collected from Coindesk. N is the number of trading days in our sample. In the cross-section, we collect all cryptocurrencies from Coinmarketcap with a market capitalization larger than one million USD that has listed for at least one year. MCAP is the natural logarithm of the dollar value of a cryptocurrency. PRC is the dollar price of a cryptocurrency. MAXDPRC is the maximum price of a cryptocurrency from the week before. AGE is the number of days that a cryptocurrency's price becomes available on Coinmarketcap. We first measure the cross-sectional summary statistics at a daily level and then take the average value over the time series. Number is the number of cryptocurrencies in the final sample. SENT is the monthly sentiment index from Baker and Wurgler (2006). CEFD is from Baker and Wurgler (2006) and is the monthly closed-end fund discount from Neal and Wheatley (1998) for 1934 to 1964 ("domestic stock funds"); Lakonishok, Shleifer, Vishny (1991) for 1965 to 1985 (general equity funds only); CDA/Wiesenberger for 1986; Herzfeld from 1987-2010; Morningstar from 2011. The data only includes un-levered general equity aggregated on an equal-weighted basis. We construct the daily level CEFD from Bloomberg. Standard Errors are reported in the parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	N	Mean	SD	Min	Max	P25	P50	P75
$GBTC$	1538	7.060	6.593	0.210	47.85	1.040	6.850	11.010
$GBTC_{Premium}$	1538	1.374	.282	.711	3.099	1.172	1.295	1.562
GC	1538	0.002	0.095	-0.774	0.859	-0.352	-0.001	0.333
News	1538	14.873	12.275	2	41	9	14	19
Cross-Section	Number	Mean	SD	Min	Max	P25	P50	P75
MCAP	1378	18.768	1.996	16.372	25.119	17.368	18.277	19.661
PRC	1378	0.718	2.533	-3.896	8.469	-0.852	0.285	1.776
MAXDPRC	1378	0.772	2.525	-3.840	8.495	-0.792	0.343	1.830
AGE	1378	955.024	424.814	412.051	1981.210	645.582	914.836	1156.481
Autocorrelation regression	GC_{t+1}	GC_{t+2}	GC_{t+3}	GC_{t+4}	GC_{t+5}	GC_{t+6}	GC_{t+7}	
GC_t	0.058	-0.025	0.002	-0.019	-0.008	-0.039	0.011	
	(0.029)	(0.028)	(0.031)	(0.033)	(0.031)	(0.032)	(0.033)	
Sentiment correlation	SENT	CEFD	CEFD (Daily)					
GC	-0.214	-0.201	0.018					

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 2: The predictability of Bitcoin at daily level

This Table presents the predictability of Bitcoin at the daily level. GC_t is the first difference of $GBTC_{Premium}$ at the daily level. $R_{BTC,t+1}$ is Bitcoin's return during day t+1. $Search_t$ is the level of Google search for the keyword "Bitcoin" or "BTC". N is the number of trading days in our sample. Panel A presents the OLS results. Panel B presents the vector-autoregression (VAR) results. In Panel C of the Table, we group the daily Bitcoin return into five groups based on the value of GC_t. Panel C shows the average return, t-statistics and the Sharpe ratio of these returns. Standard Errors are reported in the parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: OLS				
	(1)	(2)	(3)	(4)
	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$
	b/se	b/se	b/se	b/se
GC_t	0.080***	0.081***		
	(0.021)	(0.021)		
$R_{BTC,t}$	-0.010		-0.025	
	(0.038)		(0.031)	
$Search_t$	0.026			0.053
	(0.293)			(0.292)
R^2	0.011	0.011	0.000	0.000
N	1538	1538	1538	1538

Panel B: Vector Autogression		
	$R_{BTC,t+1}$	$GC_{,t+1}$
	b/se	b/se
GC_t	0.077***	0.057**
	(0.021)	(0.030)
$R_{BTC,t}$	-0.019	-0.025
	(0.029)	(0.037)
R^2	0.010	0.003
N	1538	1538

Panel C: BTC return by GC_t quintiles			
Rank	Formation Return	T-statistics	Sharpe
Low	-0.236	-0.606	-0.037
2	0.494	2.170	0.130
3	0.291	1.239	0.073
4	0.509	1.997	0.120
High	1.106	3.804	0.228

Electronic copy available at: https://ssrn.com/abstract=3948207

Table 3: Predictability by subsamples

This Table presents the predictability of Bitcoin at the daily level. We divide our sample into two groups based on the following characteristics in the time series. High (Low) Std is the sample with higher (lower) Bitcoin daily return volatility measured at the monthly level (above median). The High (Low) INF_G is the sample with a lower (higher) GBTC delay R^2 measure (above median). The High (Low) INF_B is the sample with a lower (higher) BTC delay R^2 measure (above median). High (Low) $Noise$ is the sample with a higher bitcoin arbitrage spread (above median). Lagged Bitcoin return is included as a control variable, but its coefficient has been omitted for brevity. Standard Errors are reported in the parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$
	High Std	Low Std	High Noise	Low Noise
	b/se	b/se	b/se	b/se
GC_t	0.097***	0.035***	0.135***	0.048*
	(0.02)	(0.02)	(0.02)	(0.03)
Constant	0.003**	0.002***	0.002***	0.004***
	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.017	0.006	0.013	0.005
N	769	769	769	769
$High - Low$	0.062**		0.087***	
	(0.03)		(0.03)	

	(5)	(6)	(7)	(8)
	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$
	High INF_G	Low INF_G	High INF_G	Low INF_B
	b/se	b/se	b/se	b/se
GC_t	0.087***	0.029	0.048**	0.091***
	(0.02)	(0.02)	(0.02)	(0.02)
Constant	0.003**	0.002***	0.005***	0.004***
	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.015	0.010	0.015	0.009
N	769	769	769	769
$High - Low$	0.058**		-0.043	
	(0.03)		(0.03)	

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 4: The predictability of whole cryptocurrency market with GBTC premium

This Table presents lagged changes in GBTC premium (GC_t)'s predictability of the overall cryptocurrency market and cryptocurrencies other than Bitcoin. ETH is the return of Ethereum. CMKT is the value-weighted return of the overall cryptocurrency market, including Bitcoin. CMKT_ex_BTC_ETH is the value-weighted return of the overall cryptocurrency market excluding Bitcoin and Ethereum. Top_50, 100, 200, and 300 are the value-weighted return of the top market capitalization cryptocurrency excluding Bitcoin and Ethereum. PoW is the value-weighted return of all proof-of-work cryptocurrencies excluding Bitcoin and Ethereum (The definition of PoW is from Coinmarketcap). PoS is the value-weighted return of all proof-of-stake cryptocurrencies excluding Bitcoin and Ethereum (The definition of PoS is from Coinmarketcap). Lagged Bitcoin return is included as a control variable, but its coefficient has been omitted for brevity. Standard Errors are reported in the parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Predictability of the whole crypto market		
Cryptocurrencies	Coefficient	Standard Error
ETH	0.084***	(0.031)
CMKT	0.076***	(0.020)
CMKT_ex_BTC_ETH	0.132***	(0.026)
Top_50	0.133***	(0.026)
Top_100	0.133***	(0.026)
Top_200	0.132***	(0.026)
Top_300	0.132***	(0.026)
PoW	0.127***	(0.030)
PoS	0.123***	(0.040)

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 5: Textual analysis and predictability by news category

This Table presents lagged changes in GBTC premium (GC_t)'s predictability of Bitcoin at the daily level interacted with the number of different categories of news using the following specification:

$$R_{BTC,t+1} = \alpha + \beta_1 GC_t + \beta_2 News_{t+1} + \beta_3 GC_t * News_{t+1} + \beta_4^\top X_i + \epsilon_i$$

The News is categorized using the k-means clustering first into two groups, then into four groups, and six groups. The News is categorized using the LDA into six groups. Lagged Bitcoin return is included as a control variable, but its coefficient has been omitted for brevity. Standard Errors are reported in the parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1) $R_{BTC,t+1}$ b/se	(2) $R_{BTC,t+1}$ b/se	(3) $R_{BTC,t+1}$ b/se	(4) $R_{BTC,t+1}$ b/se	(5) $R_{BTC,t+1}$ b/se	(6) $R_{BTC,t+1}$ b/se
News category (K-means)	Trading Sentiment	Others				
GC*News	0.033**	0.007				
	(0.014)	(0.005)				
GC	0.015	0.002				
	(0.033)	(0.051)				
News	0.805	0.267				
	(0.875)	(0.208)				
News category (K-means)	Trading Sentiment	Crypto exchange	Blockchain Technology	Others		
GC*News	0.035**	0.010**	0.001	-0.011		
	(0.016)	(0.004)	(0.011)	(0.014)		
GC	0.018	0.016	0.071	0.096		
	(0.033)	(0.034)	(0.047)	(0.030)		
News	0.326	0.074	1.564	1.148		
	(0.997)	(0.242)	(0.645)	(0.807)		
News category (K-means)	Trading Sentiment	Crypto exchange	Central Bank Digital Currency	Blockchain Technology	Bitcoin Mining	Others
GC*News	0.035**	0.011**	0.035**	-0.014	-0.024	0.049
	(0.016)	(0.005)	(0.016)	(0.011)	(0.015)	(0.040)
GC	0.016	0.020	0.028	0.123***	0.114***	0.070***
	(0.033)	(0.032)	(0.030)	(0.041)	(0.030)	(0.020)
News	0.000	0.000	0.002**	0.001*	0.001	-0.000
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)
News category (LDA)	Trading Sentiment	Bank & Exchange	Blockchain Network	Blockchain Technology	Users & Wallet	Mining & Payment
GC*News	0.030***	0.022***	-0.005	-0.015	-0.008	0.031
	(0.012)	(0.007)	(0.023)	(0.011)	(0.013)	(0.022)
GC	0.008	-0.012	0.081***	0.129***	0.092***	0.056**
	(0.033)	(0.034)	(0.027)	(0.042)	(0.033)	(0.024)
News	0.001	0.000	0.001	0.001	0.000	0.001
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 6: Cryptocurrency Financialization

This Table presents the regression coefficients (λ_2) from the following staggered difference-in-differences regression specification:

$$Y_{i,t} = \alpha + \lambda_1 FinanceCurrency_t + \lambda_2 FinanceCurrency_t * Post_t + \lambda_3 FE_t + \epsilon_{i,t}$$

$Y_{i,t}$ is the dependent variables that measured at the monthly level using daily data. R_{SP500} is the daily S&P 500 index return correlation with the cryptocurrency measured at the monthly level. R_{Dow} is the Dow Jones Industrial average index return. R_{Nasdaq} is the Nasdaq composite index return. R_{SMB}, R_{HML}, R_{CMA}, R_{RMW} are the Fama French five-factor return from the Kenneth French library. R_{MOM} is the daily Momentum portfolio return. Volatility is the standard deviation of the daily return measured at the monthly level for the cryptocurrency. Price Informativeness is the price delay R^2 measured as in Hou and Moskowitz (2005) with four days of lagged daily return as independent variables with the following specification:

$$r_{i,t} = \alpha_i + \sum_{j=1}^{4} \Gamma_j r_{i,t-j} + \varepsilon_{i,t}$$

$FinanceCurrency_t$ equals one if the cryptocurrency has been included in the Grayscale portfolio. The two cryptocurrencies are Bitcoin and Ethereum. *Post* equals one if the cryptocurrency has been included as a closed-end fund at the time. This is months after May 2015 for Bitcoin, and June 2019 for Ethereum. The control group includes ten never treated cryptocurrencies (with the largest market capitalizations) as of May 2015. *expanded* is the sample that includes 30 never treated cryptocurrencies (with the largest market capitalizations) as of May 2015. Standard Errors are reported in the parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Correlation (Monthly)	R_{SP500} b/se	$R_{SP500,expanded}$ b/se	R_{Dow} b/se	$R_{Dow,expanded}$ b/se	R_{Nasdaq} b/se	$R_{Nasdaq,expanded}$ b/se
Financialization	0.063***	0.065***	0.051***	0.056***	0.075***	0.078***
	(0.011)	(0.014)	(0.014)	(0.013)	(0.021)	(0.023)
	R_{SMB} b/se	$R_{SMB,expanded}$ b/se	R_{HML} b/se	$R_{HML,expanded}$ b/se	R_{RMW} b/se	$R_{RMW,expanded}$ b/se
Financialization	0.045***	0.041***	-0.039**	-0.033*	-0.033***	-0.031***
	(0.013)	(0.022)	(0.019)	(0.030)	(0.008)	(0.009)
	R_{CMA} b/se	$R_{CMA,expanded}$ b/se	R_{MOM} b/se	$R_{MOM,expanded}$ b/se	Volatility b/se	Price Informativeness b/se
Financialization	-0.031**	-0.035**	-0.001	-0.002	-0.019***	-0.032***
	(0.015)	(0.014)	(0.005)	(0.007)	(0.004)	(0.006)

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 7: GBTC premium predictability of bitcoin-related fundamental variables and supply factors

This table presents lagged changes in GBTC premium (GC_t)'s predictability of bitcoin's technical fundamental variables and supply factors. Similar to Liu and Tsyvinski (2021), Address_growth is the active bitcoin address growth. Transaction is the percentage of bitcoin transaction number growth. Payment growth is the percentage of bitcoin payment growth. Wallet growth is the wallet user growth. PC real growth is the growth of the principal component of these variables. Mining HHI is the daily Herfindahl-Hirschman Index of the mining pools of Bitcoin. Electricity stock is the daily portfolio return of China's utility sector. Chip stock is the daily portfolio return of the US's semiconductor sector. Similar to Easley et al. (2019), Mempool is the Mempool Size (Bytes), or the aggregate size in bytes of transactions waiting to be confirmed.

	(1) Coefficient	(2) Standard Error
Panel A: Bitcoin Fundamental		
Address_growth	-0.640	(0.417)
Transaction_growth	-0.620	(0.418)
Payment_growth	0.115	(0.418)
Wallet_growth	-0.314	(0.824)
PC_real_growth	-0.768	(0.574)
Panel B: Supply Factors		
Mining HHI	0.859	(1.035)
Mempool	0.386	(0.502)
Electricity stock	0.005	(0.009)
Chip stock	-0.003	(0.007)

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 8: Investor sentiment and factor return

This Table presents the regression coefficients of cryptocurrency cross-sectional factors' value-weighted quintile portfolio returns on lagged changes in GBTC premium. High indicates the group with the quintile highest value of AGE, MCAP, PRC, and MAXDPRC similar to the factors in Liu et al. (2021). Standard Errors are reported in the parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Low	2	3	4	High	High-Low
AGE	0.121***	0.103***	0.097***	0.119***	0.104***	-0.037***
	(0.030)	(0.033)	(0.032)	(0.027)	(0.023)	(0.008)
MCAP	0.104***	0.153***	0.117***	0.124***	0.079***	-0.031
	(0.033)	(0.029)	(0.028)	(0.028)	(0.020)	(0.027)
PRC	0.124***	0.165***	0.080***	0.125***	0.071***	-0.072***
	(0.034)	(0.037)	(0.031)	(0.028)	(0.020)	(0.024)
MAXDPRC	0.131***	0.163***	0.084***	0.128***	0.071***	-0.065**
	(0.034)	(0.036)	(0.031)	(0.028)	(0.020)	(0.027)

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 9: GBTC's predictability of cryptocurrency return in the cross-section by whitepaper readability

This Table presents the predictability of cryptocurrencies at the daily level using lagged changes in GBTC premium. We group cryptocurrencies into terciles based on their whitepaper's readability. Low refers to the value-weighted daily return of the group that has the easiest readability. High refers to the value-weighted daily return of the group that has the hardest readability. Following the methodology of Gunning (1969), Loughran and McDonald (2011) and Loughran and McDonald (2014), the Gunning-Fog index measures the readability based on the whitepaper's total length, word complexity, and sentence length. Doc Length is the MB size (excluding Graphs) of the cryptocurrency whitepaper. We also use the frequencies of weak modal words (could, might, possible, depending) and uncertainty words (approximate, depend, fluctuate, variability) in the whitepaper as measures of the easiness of arbitrage. Cryptocurrency technical words (Lu, 2018) measures the frequency that blockchain technical words show up in the whitepaper.

(1) Readability measure	(2) Coefficient	(3) Standard Error
Gunning Fog		
Low	0.227***	(0.07)
High	0.134***	(0.03)
Doc Length		
Low	0.391***	(0.12)
High	-0.064	(0.28)
Uncertainty		
Low	0.241***	(0.07)
High	0.103*	(0.06)
Weak Modal words		
Low	0.224***	(0.04)
High	0.042	(0.08)
Technical Words		
Low	0.426***	(0.12)
High	-0.152	(0.51)

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 10: Predictability of Bitcoin by exchange

Panel A presents lagged changes in GBTC premium (GC_t)'s predictability of Bitcoin at the daily level for different exchanges across the globe. Panel B shows the predictability of Bitcoin separated by different currency types (different currency denominations). E.g., Euro is the value-weighted daily return of bitcoin aggregated across different exchanges with the Euro as Bitcoin's buy-in currency. Lagged bitcoin return is controlled but omitted from the Table for brevity. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A

Exchange	Coefficient	Standard Error	Exchange	Coefficient	Standard Error
AAX	0.204***	(0.057)	HitBTC	0.087***	(0.019)
AscendEX	0.157***	(0.041)	HollaEx	0.203***	(0.063)
Bequant	0.087***	(0.019)	KuCoin	0.128***	(0.028)
BigONE	0.186***	(0.032)	NDAX	0.334***	(0.050)
Binance	0.191***	(0.058)	NovaDAX	0.071***	(0.021)
BitBay	0.249***	(0.088)	OKCoin	0.341***	(0.050)
bitcoin.com	0.087***	(0.019)	OKEX	0.042***	(0.019)
Bitfinex	0.149***	(0.048)	Poloniex	0.082***	(0.019)
BitMart	0.306***	(0.017)	ProBit	0.178***	(0.044)
BitMEX	0.100***	(0.020)	STEX	0.073***	(0.021)
Bittrex	0.093**	(0.044)	TideBit	0.153***	(0.020)
BTCTurk	0.177***	(0.037)	TimeX	0.156***	(0.063)
Bybit	0.176***	(0.044)	Upbit	0.109***	(0.019)
Coinbase	0.081***	(0.020)	VCC Exchange	0.158***	(0.048)
Currency.com	0.096***	(0.019)	WhiteBit	0.185***	(0.056)
Delta Exchange	0.163***	(0.048)	Xena Exchange	0.106***	(0.038)

Panel B

Currency	Coefficient	Standard Error
USD	0.173***	(0.032)
Euro	0.139***	(0.049)
HKD	0.115**	(0.045)
Korean Won	0.231***	(0.051)
Japanese Yen	0.209**	(0.083)
Tether	0.163***	(0.042)

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 11: GBTC market impact

This Table presents the predictability of Bitcoin at daily level, separately for positive and negative GC. $\triangle Holding$ is daily changes in Bitcoin holding of GBTC. $\triangle Holding_{positive}$ equals $\triangle Holding$ if $\triangle Holding$ is positive, zero otherwise. $\triangle Holding_{negative}$ equals $\triangle Holding$ if $\triangle Holding$ is negative, zero otherwise. $GC_{positive}$ is equal to GC if GC is positive, zero otherwise. $GC_{negative}$ is equal to GC if GC is negative, zero otherwise. Lagged Bitcoin return is included as a control variable, but its coefficient has been omitted for brevity. Standard Errors are reported in the parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$
	Whole Sample	2018-2021 Sample	2018-2021 Sample	2018-2021 Sample
	b/se	b/se	b/se	b/se
$GC_{positive}$	0.063*	0.089**		
	(0.034)	(0.041)		
$GC_{negative}$	0.099***	0.157***		
	(0.034)	(0.052)		
$\triangle Holding$			-0.003	
			(0.008)	
$\triangle Holding_{positive}$				0.006
				(0.012)
$\triangle Holding_{negative}$				-0.015
				(0.034)
R^2	0.011	0.011	0.000	0.000
N	1538	813	813	813

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 12: Robustness Check: Predictability by Bitcoin volume, number of news regarding blockchain, and expected return of Bitcoin

This Table presents the predictability of Bitcoin at the daily level. We divide our sample into two groups based on the following characteristics in the time series. High (Low) Vol is the sample with higher (lower) Bitcoin dollar trading volume (above median). High (Low) move is the sample with a higher (lower) absolute value of GC_t (above median). High (Low) N_News is the sample with a higher (lower) number of News regarding Bitcoin or Blockchain on CoinDesk (above median). High (Low) Abnormal is the sample with a higher abnormal number of News (above median). Abnormal number of News is defined as the number of News regarding Bitcoin or Blockchain on CoinDesk minus the one-week average number of News before that day. Lagged Bitcoin return is included as a control variable, but its coefficient has been omitted for brevity. Standard Errors are reported in the parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$
	High Vol	Low Vol	High move	Low move
	b/se	b/se	b/se	b/se
GC_t	0.117***	0.038***	0.067***	0.086*
	(0.02)	(0.01)	(0.01)	(0.05)
Constant	0.004**	0.005***	0.005***	0.003***
	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.017	0.006	0.013	0.005
N	769	769	769	769
$High - Low$	0.079***		-0.019	
	(0.02)		(0.05)	
	(5)	(6)	(7)	(8)
	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$
	High N_News	Low N_News	High Abnormal	Low Abnormal
	b/se	b/se	b/se	b/se
GC_t	0.114***	0.056***	0.149***	0.075***
	(0.02)	(0.02)	(0.04)	(0.02)
Constant	0.004**	0.006***	0.003***	0.005***
	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.015	0.010	0.015	0.009
N	769	769	769	769
$High - Low$	0.058**		0.074*	
	(0.02)		(0.04)	

Electronic copy available at: https://ssrn.com/abstract=3948407

Table 13: Alpha of GBTC strategic return

This Table presents the alpha of the strategic portfolio (given GBTC signal) daily return on daily Bitcoin return, Fama-French five-factor factor, momentum return, and reversal return (From Kenneth French Library). The strategy initializes a buy of bitcoin when changes in GBTC premium is above the 55 percentile of the past rolling one-year changes in GBTC premium. The strategy initializes a sale of bitcoin when changes in GBTC premium are below the 45 percentile of the past rolling one-year changes in GBTC premium.

	(1) Strategy b/se	(2) Strategy b/se	(3) Strategy b/se	(4) Strategy b/se
Alpha	0.380***	0.390***	0.390***	0.410***
	(0.132)	(0.132)	(0.132)	(0.134)
R_{BTC}	-0.112***	-0.106***	-0.107***	-0.106***
	(0.028)	(0.028)	(0.028)	(0.028)
Mkt-RF		-0.060		-0.002
		(0.111)		(0.114)
SMB		-0.293		-0.220
		(0.212)		(0.216)
HML		0.037		0.286
		(0.189)		(0.242)
RMW		-0.339		-0.282
		(0.331)		(0.338)
CMA		0.550		0.332
		(0.407)		(0.417)
Mom			0.083	0.199
			(0.108)	(0.166)
ST_Rev			-0.283**	-0.278**
			(0.116)	(0.129)
R^2	0.014	0.018	0.019	0.023

Electronic copy available at: https://ssrn.com/abstract=3948407

A Online Appendix

A.1 A simple model

Suppose there are two dates, time 0 and time 1, and no discounting. At time 0, three assets are traded in the financial market: GBTC, Bitcoin (BTC) and a risk-free asset. The risk-free asset has a constant value 1 and is in unlimited supply. GBTC and BTC have the same random terminal value $\tilde{v}$ which pays off at time 1. We assume that $\tilde{v} \sim N\left(0, \tau_v^{-1}\right)$ with precision $\tau_v > 0$. GBTC and BTC have fixed supply Q_G and Q_B, are traded at endogenous prices p_t and p_s at time 0, respectively. In our model, p_t and p_s may not be the same due to the market frictions that we will discuss later.

Traders. There are three types of traders in the market: traditional traders, sophisticated traders and noise traders. For simplicity, we assume that the measures of traditional traders and sophisticated traders are the same, normalized to be 1.[22] Traditional traders only trade in the GBTC market, and they have CARA utility over their time 1 wealth and risk aversion γ_t. They represent investors who have limited knowledge about BTC, and opening and maintaining a BTC wallet can be very costly for them. So they will choose to have some BTC exposure through a more convenient financial tool, which is GBTC in our model. Sophisticated traders only trade in the BTC (spot) market, and they have CARA utility over their time 1 wealth and risk aversion γ_s.[23] Compared to traditional traders, they have more knowledge about BTC, and thus would like to open and maintain their BTC wallet themselves. Noise traders exist in both GBTC market and BTC market, and they are trading for pure liquidity reasons. We assume that noise traders demand $\tilde{x}_t$ and $\tilde{x}_s$ in the GBTC market and BTC market, respectively, where $\tilde{x}_t \sim N\left(0, \tau_{x,t}^{-1}\right)$ and $\tilde{x}_s \sim N\left(0, \tau_{x,s}^{-1}\right)$.[24]

[22] Assuming the measures to be the same is not important in our model.

[23] Restricting sophisticated traders from only trading in BTC market is a simplifying assumption. But in practice, there are many reasons to rationalize this result: first, trading GBTC may incur significant loss from management fees; second, there is a significant premium in GBTC over a long time period (most of the time in our data sample), so it is very costly for sophisticated traders to gain BTC exposure by trading GBTC; third, BTC is traded 24 hours a day while the GBTC is only traded from 9:00 am to 4:00 pm U.S. eastern time.

[24] Assume $\tau_{x,t} > 0$ and $\tau_{x,s} > 0$.

Electronic copy available at: https://ssrn.com/abstract=3948407

$\tilde{x}_t$ and $\tilde{x}_s$ are independent of each other, and both of them are independent of all other random variables in the model.

Information. Each traditional trader i can receive a private signal $s_{t,i}$ about terminal value $\tilde{v}$:

$$s_{t,i} = \tilde{v} + \epsilon_{t,i},$$

where $\epsilon_{t,i} \sim N\left(0, \tau_{t,\epsilon}^{-1}\right)$[25] is the i.i.d error term. And each sophisticated trader j can receive a private signal

$$s_{s,j} = \tilde{v} + \epsilon_{s,j},$$

where $\epsilon_{s,j} \sim N\left(0, \tau_{s,\epsilon}^{-1}\right)$[26] is the i.i.d error term. The endogenous prices p_t and p_s are also informative about terminal value $\tilde{v}$, and can possibly be used by traders in their investment decisions. We assume that sophisticated trader can process all information available to them. Specifically, sophisticated trader j's information set is $\{s_{s,j}, p_t, p_s\}$. However, traditional traders have limited attention and are not able to process all information available in their investment decisions. For example, they may not be full-time traders or are trading various financial assets and thus can only allocate limited attention to BTC research. Specifically, we assume that traditional traders can always learn GBTC price p_t as they are trading in the GBTC market. But they can only learn one from their private signals and BTC price p_s. So for each traditional trader i, his information set can either be $\{s_{t,i}, p_t\}$ or $\{p_t, p_s\}$.

A.2 Equilibria

For each traditional trader i, his information choice is between $s_{t,i}$ and p_s, both are noisy measures of terminal value $\tilde{v}$ in equilibrium. In equilibrium, if

$$\mathrm{Var}\left(\tilde{v}|s_{t,i}\right) < \mathrm{Var}\left(\tilde{v}|p_s\right),$$

[25] $\tau_{t,\epsilon} > 0$
[26] $\tau_{s,\epsilon} > 0$

Electronic copy available at: https://ssrn.com/abstract=3948407

then he will choose $s_{t,i}$, otherwise, he will choose p_s.[27] Since the private signals are symmetric for all traditional traders, there are only two possible equilibria: either all traditional traders choose to learn their private signals, or all of them choose to learn BTC spot price p_s. We call the former equilibrium the informative equilibrium, and the latter the uninformative equilibrium.

Informative equilibrium.

In this equilibrium, traditional traders choose to learn their private signals. Following the literature (Goldstein and Yang 2017), we conjecture that the equilibrium prices are

$$p_t^{in} = \alpha_0 + \alpha_v \tilde{v} + \alpha_x \tilde{x}_t, \tag{A-1}$$

$$p_s^{in} = \beta_0 + \beta_v \tilde{v} + \beta_t p_t^{in} + \beta_x \tilde{x}_s, \tag{A-2}$$

where the information contained in both prices can be represented by

$$s_{t,p}^{in} = \frac{p_t^{in} - \alpha_0}{\alpha_v} = \tilde{v} + \frac{\alpha_x}{\alpha_v}\tilde{x}_t = \tilde{v} + \rho_{t,in}^{-1}\tilde{x}_t$$

and

$$s_{s,p}^{in} = \frac{p_s^{in} - (\beta_0 + \beta_t p_t^{in})}{\beta_v} = \tilde{v} + \frac{\beta_x}{\beta_v}\tilde{x}_s = \tilde{v} + \rho_{s,in}^{-1}\tilde{x}_s.$$

Note that both $s_{t,p}^{in}$ and $s_{s,p}^{in}$ are normally distributed, with the same mean $\tilde{v}$ and precision $\rho_{t,in}^2 \tau_{x,t}$ and $\rho_{s,in}^2 \tau_{x,s}$, respectively.

Following the literature, we can compute the demand function for any traditional trader i:

$$D_t^{in}\left(s_{t,i}, p_t^{in}\right) = \frac{\tau_{t,\epsilon} s_{t,i} + \rho_{t,in}^2 \tau_{x,t} s_{t,p}^{in} - \left(\tau_v + \tau_{t,\epsilon} + \rho_{t,in}^2 \tau_{x,t}\right) p_t^{in}}{\gamma_t}.$$

[27] We assume that in the break-even case $\mathrm{Var}\left(\tilde{v}|s_{t,i}\right) < \mathrm{Var}\left(p_s|s_{t,i}\right)$, the traditional trader i chooses to learn p_s.

Electronic copy available at: https://ssrn.com/abstract=3948407

The market clear condition is

$$\int_i D_t^{in}\left(s_{t,i}, p_t^{in}\right) di + \tilde{x}_t = Q_G.$$

Similarly, we can compute the demand function for any sophisticated trader j:

$$D_s^{in}\left(s_{s,j}, p_t^{in}, p_s^{in}\right) = \frac{\tau_{s,\epsilon} s_{s,j} + \rho_{t,in}^2 \tau_{x,t} s_{t,p}^{in} + \rho_{s,in}^2 \tau_{x,s} s_{s,p}^{in} - \left(\tau_v + \tau_{t,\epsilon} + \rho_{t,in}^2 \tau_{x,t} + \rho_{s,in}^2 \tau_{x,s}\right) p_s^{in}}{\gamma_s}.$$

The market clear condition is

$$\int_j D_s^{in}\left(s_{s,j}, p_t^{in}, p_s^{in}\right) dj + \tilde{x}_s = Q_B.$$

We can solve all coefficients in (A-1) and (A-2) from the above market clear conditions. Specifically, we can show that (Goldstein and Yang 2017)

$$\rho_{t,in} = \frac{\tau_{t,\epsilon}}{\gamma_t}$$

and

$$\rho_{s,in} = \frac{\tau_{s,\epsilon}}{\gamma_s}.$$

Uninformative equilibrium.

In this equilibrium, traditional traders choose to learn the BTC spot price. We conjecture that the equilibrium prices are

$$p_t^{un} = a_0 + a_s p_s^{un} + a_x \tilde{x}_t, \tag{A-3}$$

$$p_s^{un} = b_0 + b_v \tilde{v} + b_s \tilde{x}_s, \tag{A-4}$$

Electronic copy available at: https://ssrn.com/abstract=3948407

where the information contained in p_s^{un} can be represented by

$$s_{s,p}^{un} = \frac{p_s^{un} - b_0}{b_v} = \tilde{v} + \frac{b_s}{b_v}\tilde{x}_s = \tilde{v} + \rho_{s,un}^{-1}\tilde{x}_s.$$

Note that $s_{s,p}^{un}$ is normally distributed, with mean $\tilde{v}$ and precision $\rho_{s,un}^2\tau_{x,s}$. In this setting, the traditional traders' information set is equivalent to $\{p_s^{un}\}$, as the GBTC price p_t^{un} is just a noisy measure of p_s^{un}.

Following the literature, we can compute the demand function for any traditional trader i:

$$D_t^{un}\left(p_t^{un}, p_s^{un}\right) = \frac{\rho_{t,un}^2\tau_{x,t}s_{t,p}^{un} - \left(\tau_v + \rho_{t,un}^2\tau_{x,t}\right)p_t^{un}}{\gamma_t}.$$

The market clear condition is

$$\int_i D_t^{un}\left(p_t^{un}, p_s^{un}\right)di + \tilde{x}_t = Q_G.$$

Similarly, we can compute the demand function for any sophisticated trader j:

$$D_s^{un}\left(s_{s,j}, p_t^{un}, p_s^{un}\right) = \frac{\tau_{s,\epsilon}s_{s,j} + \rho_{s,un}^2\tau_{x,s}s_{s,p}^{un} - \left(\tau_v + \tau_{t,\epsilon} + \rho_{s,un}^2\tau_{x,s}\right)p_s^{un}}{\gamma_s}.$$

The market clear condition is

$$\int_j D_s^{un}\left(s_{s,j}, p_t^{un}, p_s^{un}\right)dj + \tilde{x}_s = Q_B.$$

We can solve all coefficients in (A-3) and (A-4) based on the above market clear conditions. Specifically, we can show that

$$\rho_{s,un} = \frac{\tau_{s,\epsilon}}{\gamma_s}.$$

Equilibrium Comparison

What are the conditions that sustain both equilibria? Suppose we are in the informative equilibrium. As we discussed earlier, traditional traders will choose their private signals if

Electronic copy available at: https://ssrn.com/abstract=3948407

and only if

$$\text{Var}\left(\tilde{v}|s_{t,i}\right) < \text{Var}\left(\tilde{v}|p_s^{in}\right),$$

which is equivalent to

$$\tau_{t,\epsilon} > \rho_{s,in}^2 \tau_{x,s} = \left(\frac{\tau_{s,\epsilon}}{\gamma_s}\right)^2 \tau_{x,s}.$$

If we are in the uninformative equilibrium, traditional traders will choose to learn BTC spot price if and only if

$$\text{Var}\left(\tilde{v}|s_{t,i}\right) \geq \text{Var}\left(\tilde{v}|p_s^{un}\right),$$

which is equivalent to

$$\tau_{t,\epsilon} \leq \rho_{s,in}^2 \tau_{x,s} = \left(\frac{\tau_{s,\epsilon}}{\gamma_s}\right)^2 \tau_{x,s}.$$

Thus, we obtain Proposition 1.

Electronic copy available at: https://ssrn.com/abstract=3948407

Table A.1: BTC's autocorrelation at the daily level using different starting hour

This Table presents the BTC's autocorrelation regression coefficients at the daily level using different starting hours. The Bitcoin return is calculated at a daily level (24 hours) for different starting hours, where zero refers to UTC time hour zero.

UTC Hour	**coef**	**std**	**t-stat**	**p**	R^2
0	-0.061	0.024	-2.484	0.013	0.004
1	-0.078	0.025	-3.186	0.001	0.006
2	-0.107	0.025	-4.342	0.000	0.011
3	-0.073	0.025	-2.901	0.004	0.005
4	-0.078	0.025	-3.157	0.002	0.006
5	-0.082	0.025	-3.287	0.001	0.007
6	-0.092	0.025	-3.746	0.000	0.009
7	-0.089	0.025	-3.581	0.000	0.008
8	-0.084	0.024	-3.432	0.001	0.007
9	-0.073	0.025	-2.956	0.003	0.005
10	-0.047	0.025	-1.892	0.059	0.002
11	-0.037	0.025	-1.489	0.137	0.001
12	-0.024	0.025	-0.976	0.329	0.001
13	0.021	0.025	0.852	0.394	0.000
14	0.002	0.025	0.064	0.949	0.000
15	0.015	0.025	0.609	0.543	0.000
16	0.006	0.024	0.253	0.800	0.000
17	-0.009	0.024	-0.355	0.723	0.000
18	-0.007	0.025	-0.272	0.786	0.000
19	-0.019	0.025	-0.781	0.435	0.000
20	-0.032	0.025	-1.287	0.198	0.001
21	-0.025	0.025	-1.011	0.312	0.001
22	-0.030	0.025	-1.229	0.219	0.001
23	-0.021	0.025	-0.828	0.408	0.000

Electronic copy available at: https://ssrn.com/abstract=3948407

Table A.2: GBTC premium, closed-end fund premium/discount, and sentiment index

$CEFD$ is the equal-weighted equity closed-end mutual fund discount at the daily level from Bloomberg. $CEFD_{Diff}$ is the first difference of $CEFD$ at the daily level. Lagged Bitcoin return is included as a control variable, but its coefficient has been omitted for brevity. Standard Errors are reported in the parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Daily CEF discount and Bitcoin return			
	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$
	b/se	b/se	b/se
$CEFD$	-0.081		
	(0.094)		
$CEFD_{Diff}$		-0.926**	-0.944**
		(0.394)	(0.392)
GC_t			0.077***
			(0.020)
R^2	0.001	0.004	0.015

Panel B: Monthly SENT and Bitcoin return		
	$R_{BTC,t+1}$	$R_{BTC,t+1}$
	b/se	b/se
$SENT$	-0.001	-0.001
	(0.004)	(0.003)
GC_t		0.077***
		(0.020)
R^2	0.000	0.011

Electronic copy available at: https://ssrn.com/abstract=3948407

Table A.3: GBTC predictability for each weekday and year

This Table presents the GBTC predictability of Bitcoin at daily level for different weekday and year. Lagged Bitcoin return is included as a control variable, but its coefficient has been omitted for brevity. Standard Errors are reported in the parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$
	Mon	Tue	Wed	Thurs	Fri
	b/se	b/se	b/se	b/se	b/se
GC_t	0.127***	0.076**	0.097***	0.103**	0.137***
	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)
Constant	0.006**	0.005**	0.001	-0.001	0.003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.009	0.011	0.013	0.012	0.016
N	306	306	306	306	306
	(1)	(2)	(3)	(4)	(5)
	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$	$R_{BTC,t+1}$
	2020	2019	2018	2017	2016
	b/se	b/se	b/se	b/se	b/se
GC_t	0.148***	0.124***	0.123***	0.079**	0.041*
	(0.04)	(0.05)	(0.03)	(0.03)	(0.03)
Constant	0.007***	0.003	-0.004	0.013***	0.004**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.238	0.099	0.196	0.107	0.038
N	253	252	251	251	252

Electronic copy available at: https://ssrn.com/abstract=3948407