Fundamental Sentiment and Cryptocurrency Risk Premia

Ilias Filippou* My T. Nguyen† Ganesh Viswanath-Natraj‡

§

This version: October 18, 2024

Abstract

This paper examines the cross-sectional predictive ability of text-based fundamental factors in the cryptocurrency market. Using BERT topic modeling, we construct a Fundamental Sentiment Index (FSI) from news articles related to cryptocurrencies. We find that cryptocurrencies with high exposure to fundamental sentiment, typically payment and platform tokens, earn a risk premium, while governance tokens hedge against fundamental risks. Currency betas correlate with blockchain value metrics, such as the ratio of users and transactions to market cap. A long-short strategy based on FSI exposure delivers statistically significant returns, highlighting the importance of fundamental sentiment in explaining cryptocurrency returns beyond traditional factors.

Keywords: cryptocurrency, fundamentals, media coverage, textual analysis.

JEL Classifications: G11, G12, G14, G32.

*ifilippou@fsu.edu, Florida State University, USA

†my.nguyen@fulbright.edu.vn, Fulbright University Vietnam, Vietnam.

‡Ganesh.Viswanath-Natraj@wbs.ac.uk, University of Warwick, Warwick Business School, UK

§We would like to thank Yukun Liu and seminar participants at INQUIRE 2023, Wolfe Research Annual Global Quantitative and Macro Investment Conference 2023, Global AI Finance Research Conference 2023, the Gillmore Center for Financial Technology 2023, and the MARBLE 2023 conference for helpful comments. All errors remain those of the authors. The views expressed in this paper are those of the authors and not necessarily those of Washington University in St. Louis, University of Warwick, or any other institution with which the authors may be affiliated or associated.

1 Introduction

This paper investigates the cross-sectional predictive ability of text-based fundamental trading measures in the cryptocurrency market. By November 2021, the market capitalization of cryptocurrencies surpassed 3 trillion USD, with spot and futures contract trading volumes reaching 8.8 trillion USD in the first quarter of 2020 (e.g., Helms, 2020). Cryptocurrencies have become a significant asset class for both retail and institutional investors (e.g., Harvey et al., 2022), experiencing substantial price fluctuations in recent years. While speculative forces often drive these price movements, recent research shows that blockchain fundamentals play a crucial role in explaining cryptocurrency prices and the cross-section of returns (Liu et al., 2021; Cong et al., 2021; Liu et al., 2022; Bhambhwani et al., 2023).

In this study, we develop a novel text-based factor-pricing framework that captures news related to blockchain fundamentals. We collect news articles from Factiva that mention the top 43 cryptocurrencies by market capitalization as of December 2021. Using Bidirectional Encoder Representations from Transformers (BERT) topic modeling, we identify the key topics discussed in these articles, such as fundamentals, technical trading, regulation, lending, payments, derivatives, social media, and hedging. Among these, we find that text-based measures capturing blockchain fundamentals are particularly important for explaining the cross-section of cryptocurrency returns.

To frame our analysis, we introduce a simple model that explains how cryptocurrency fundamentals influence both the supply and demand for digital currencies. On the *supply side*, key factors include the hash rate, which reflects the computational power of the blockchain, and transaction costs, such as gas fees on the Ethereum network. These factors are critical for the efficiency of mining activities and overall blockchain operations. On the *demand side*, elements such as the number of active addresses and institutional liquidity needs play a central role in determining the utility of tokens for payments and platform services.

Our model, based on Biais et al. (2023), provides a structured framework to test hypotheses related to these fundamental drivers. It explores how transaction costs and benefits, both influenced by blockchain characteristics, impact cryptocurrency valuation. Cryptocurrencies generate transactional benefits over time, and these benefits, alongside transaction costs, help determine their market prices. This framework allows us to investigate two main hypotheses: first, that cryptocurrencies with higher exposure to fundamental sentiment are riskier and command higher expected returns, and second, that the sensitivity of cryptocurrencies to blockchain fundamentals varies based on their use case. Payment and platform tokens are more sensitive to these fundamentals, while governance tokens are less sensitive and provide a hedge against fundamental risks.

To test our hypotheses, we use sentiment extracted from articles that discuss blockchain fundamentals. Articles classified as fundamental by the BERT model are used to construct the Fundamental Sentiment Index (FSI). This index measures sentiment by calculating the difference in frequency between positive and negative words, based on the Loughran and McDonald (2011) approach, which captures net positive sentiment (optimism) on fundamental topics.

Using the FSI, we estimate rolling betas to measure each cryptocurrency's exposure to fundamental sentiment, while controlling for risk factors including market, size, momentum, volatility, and liquidity. We classify cryptocurrencies by their token type and find that those with high betas are generally platform or payment tokens. These tokens exhibit positive comovement with fundamental sentiment, suggesting that shifts in blockchain congestion or transaction benefits influence their utility. In contrast, cryptocurrencies with low betas are often governance tokens, which serve as a hedge against changes in fundamental sentiment. Our findings also show that these betas are related to various value indicators, such as the ratio of transactions, users, and addresses to market capitalization, as demonstrated in Cong et al.

(2021). Cryptocurrencies with higher value metrics are typically used more for payments or platform activities, whereas governance tokens, which are less frequently traded, tend to have lower value metrics.

Next, we construct a sentiment-based factor using FSI. To evaluate the predictive ability of this factor, we create long-short portfolios based on cryptocurrency exposure to FSI, sorting them into quartiles each week according to their 60-week rolling betas. We then form long-short portfolios by buying cryptocurrencies with high FSI exposure and selling those with low exposure (HML_{FSI}). This fundamental-based strategy delivers statistically significant returns, with a Sharpe ratio of 1.24, outperforming the broader cryptocurrency market. These results are robust across various sample groups, sentiment proxies, portfolio constructions, and factor beta specifications.

Our analysis of the fundamental sentiment factor yields three key findings. First, we show that conventional cryptocurrency risk factors cannot fully explain the returns associated with text-based fundamental factors. Regressions of our FSI factor on market, size, momentum, liquidity, and volatility reveal that the alpha remains statistically and economically significant. Following Cong et al. (2021), we also examine value factors such as the number of transactions, cumulative addresses, and addresses with a balance. The FSI factor generates positive and significant alphas, indicating that text-based fundamental sentiment provides additional explanatory power beyond existing factor models.

Second, we perform Fama and MacBeth (1973) cross-sectional regressions to assess the pricing power of the fundamental factor, controlling for various determinants of cryptocurrency returns. Our findings indicate that cryptocurrencies with positive exposure to fundamental sentiment are riskier, and investors demand a risk premium for holding them. In a baseline model that includes market and our fundamental factor, the price of risk for the fundamental factor is 1.2 percent per week. These results hold even when alternative factor models, incorporating volatility and momentum, are considered, as outlined by Liu et al. (2022) and Cong et al. (2021).

Third, we demonstrate that the fundamental sentiment factor offers significant diversification benefits when combined with traditional asset pricing factors, such as market, size, illiquidity, volatility, and momentum. The inclusion of the fundamental sentiment factor substantially improves risk-adjusted returns. For instance, the Sharpe ratio of the market portfolio increases from 0.08 to 0.71, and similar enhancements are observed for the size, illiquidity, volatility, and momentum factors. An equally weighted portfolio of all factors sees its Sharpe ratio rise from 0.91 to 1.41 per annum, underscoring the value of fundamental sentiment as a diversifying component in cryptocurrency portfolios.

Finally, we examine sentiment related to other topics, including regulation, lending, payments, derivatives, social media, hedging, and technical trading. We find that none of these sentiment factors predict the cross-section of cryptocurrency returns, except for technical sentiment, which negatively predicts returns but is unrelated to fundamental sentiment.

Literature review. Our paper contributes to the emerging literature on the cross-section of cryptocurrency returns (Bianchi and Babiak, 2021; Cong et al., 2021; Liu et al., 2021, 2022; Filippou et al., 2023; Bhambhwani et al., 2023; Schwenkler and Zheng, 2020; Kogan et al., 2022; Bianchi et al., 2022; Han et al., 2022; Luo et al., 2021).

The seminal work by Liu et al. (2022) establishes that cryptocurrency return factors, particularly those based on market, momentum, and volatility, exhibit pricing power for the cross-section of cryptocurrency returns. Building on this, other studies such as Bhambhwani et al. (2023) and Cong et al. (2021) demonstrate that value and network-based factors also possess significant explanatory power, and Liu et al. (2021) show that innovations to the number of new blockchain addresses explain a significant portion of cryptocurrency return variations. Specifically, blockchain characteristics, such as hash rate and the number of active addresses, are positively correlated with cryptocurrency prices. Greater exposure to these blockchain metrics is associated with higher expected returns, offering investors a risk premium.

Our approach is distinct in that we derive fundamental factors directly from cryptocurrency-related news articles. This method allows us to disentangle competing theories of cryptocurrency pricing more precisely, such as whether pricing dynamics are driven primarily by retail trading (e.g., Kogan et al., 2022), or by news about blockchain characteristics.

Prior research has used textual analysis in cryptocurrency markets (Filippou et al., 2023; Schwenkler and Zheng, 2020; Liu et al., 2021). For example, Schwenkler and Zheng (2020) document a substitution effect where negative news about a peer currency drives investors toward others with similar network characteristics. Liu et al. (2021) construct a Tech Index from whitepapers to measure technology sophistication, finding that while high-tech cryptocurrencies see early success, they experience lower long-term returns, particularly after market shocks like the Luna crash. Filippou et al. (2023) apply various news sources to machine learning models, concluding that fundamental factors best predict cryptocurrency returns.

Relative to these studies, the novelty of our approach lies in the application of the BERT model to extract text-based factors from cryptocurrency news. By measuring sentiment in specific topics, we construct indices that capture optimism about fundamental news. Using standard asset pricing tests, we find that these sentiment-driven factors are priced in the cross-section of cryptocurrency returns, complementing models that rely on value factors and blockchain characteristics.

The remainder of this paper is structured as follows. Section 2 outlines the theoretical motivation and testable hypotheses. Section 3 describes the data and methodology, including the use of BERT to identify cryptocurrency-related topics and the construction of sentiment measures based on fundamental news. Section 4 presents our main empirical results, while

Section 5 concludes the paper.

2 Testable Hypotheses

Fundamental analysis offers a framework that helps investors identify the intrinsic value of an asset by examining a variety of related economic and financial factors. In contrast to the equities market, which adheres to Generally Accepted Accounting Principles (GAAP) for financial measurements, the cryptocurrency market lacks a standardized accounting framework. This absence complicates efforts for traders and regulators to determine the fundamental value of cryptocurrencies (Liu et al., 2021).[1]

Despite this, a wealth of publicly available information about economic activities on the blockchain provides a promising alternative for establishing the intrinsic value of cryptocurrencies. This data, verifiable through the public ledger, serves as a valuable resource. For example, Liu et al. (2021) apply accounting and finance valuation methods to the cryptocurrency market, finding that information about new addresses is highly value-relevant for cryptocurrencies. Additionally, Bhambhwani et al. (2023) show that both the number of addresses and the hash rate are robust predictors of cryptocurrency returns, while Cong et al. (2021) argue that these blockchain characteristics can be used as value-based factors.

To motivate our analysis of fundamental blockchain characteristics, we introduce a simplified model based on Biais et al. (2023).[2] This model is set in an overlapping generations framework where young consumers have access to both cryptocurrency and fiat money as mediums of

[1]In the equities market, traders can analyze financial statements to estimate firm value and compare it with market price. For instance, Lev and Thiagarajan (1993) and Abarbanell and Bushee (1997) demonstrate that signals capturing information on firm fundamentals (such as inventory changes, gross margins, selling expenses, capital expenditures, and labor force sales productivity) are commonly used by financial analysts to forecast future firm performance. Abarbanell and Bushee (1998) further build a trading strategy using these fundamental analysis signals to generate abnormal returns.

[2]For a detailed derivation of the model, please refer to the Appendix A. Our model simplifies the original analysis by omitting discussions of crash risk and the role of hackers.

exchange. Cryptocurrencies incur transaction costs, ψ_t, which represent fees from exchanges or the costs of validating transactions via mining. In the next period, these cryptocurrencies generate benefits, captured by the parameter θ_{t+1}. These benefits could include transactional advantages such as the use of cryptocurrency in cross-border payments or its programmability through smart contracts.

The model yields the following Euler equation, which relates the current price of cryptocurrency, p_t, to its perceived future transaction benefits and costs:

$$p_t = \frac{1}{1+r_t}\mathbb{E}_t\left[\frac{u'(c^o_{t+1})}{\mathbb{E}_t u'(c^o_{t+1})}\frac{1+\theta_{t+1}}{1+\psi_t}p_{t+1}\right] \tag{1}$$

Proof: See Appendix.

Defining the net transactional benefits of holding a currency as $1+\mathcal{T}_t = \frac{1+\theta_{t+1}}{1+\psi_t}$ and iterating forward, we obtain the expression that the price of cryptocurrency is the net present value of its future stream of transactional benefits:

$$p_t = \sum_{j=1}^{\infty}\left[\prod_{k=0}^{j-1}\frac{1}{1+r_{t+k}}\mathbb{E}_t\left[\frac{u'(c^o_{t+k+1})}{\mathbb{E}_t[u'(c^o_{t+k+1})]}\mathcal{T}_{t+k+1}p_{t+k+1}\right]\right] \tag{2}$$

Using this framework, we derive two key hypotheses that guide our empirical analysis:

H1. *Cryptocurrencies that exhibit positive exposure to fundamental sentiment are riskier.*

(a) Fundamental sentiment positively predicts the cross-section of cryptocurrency returns.

(b) Investors demand a risk premium for holding these cryptocurrencies.

To test this hypothesis, we create a news-based sentiment measure focused on cryptocurrency fundamentals. Our sentiment analysis draws on expert discussions about blockchain economics, supply-demand dynamics, and technological advancements. Cryptocurrencies with returns that co-move positively with fundamental sentiment are considered riskier, as they are more exposed to shifts in these fundamentals. We hypothesize that investors demand a risk premium for this exposure. Our empirical tests utilize standard asset pricing methodologies to measure this risk premium.

H2. *The sensitivity of cryptocurrencies to blockchain fundamentals depends on whether they are used for payments, platform operation, or governance.*
(a) Payment and platform tokens are more sensitive to fundamental sentiment and thus riskier.
(b) Governance tokens that are not used for payments are less sensitive to blockchain fundamentals, providing a hedge against fundamental risks.

This hypothesis builds on our model's prediction that the sensitivity of a cryptocurrency to blockchain fundamentals is conditional on its function as a medium of exchange. We classify tokens into three categories: general payment tokens, platform tokens, and governance tokens (see Section 3.4 for more details). Payment and platform tokens typically compensate miners for transaction verification, making them more sensitive to blockchain congestion and technological shocks (e.g., hash rate fluctuations). Conversely, governance tokens, which are used for voting rights rather than payments, tend to be less sensitive to such factors. We hypothesize that governance tokens provide a hedge against fundamental risks, while payment and platform tokens are more exposed to these risks and thus command a risk premium.

3 Data and Definitions

This section discusses cryptocurrency data. We provide a detailed description of our corpus, the topic modeling approach, and the construction of the fundamental sentiment indexes.

3.1 Cryptocurrency Data

3.1.1 Cryptocurrency characteristics

We collect daily cryptocurrency data from CoinMetrics, which includes prices and other cryptocurrency characteristics data. CoinMetrics provides quality data on cryptocurrency characteristics. We begin with 50 cryptocurrencies with the highest market capitalization as of January 2022. Then we eliminate five stablecoins and two coins that are pegged to bitcoin.[3] Therefore our sample contains 43 cryptocurrencies. The data spans the period of June 2017 to December 2021. We convert our data to a weekly series by setting Friday as the end of the week to be consistent with the Fama and French factors convention. Therefore we construct weekly returns by calculating the difference between the closing price on the Friday of a week and the closing price on the Friday of the previous week.[4]

In Appendix C, we report the total number of cryptocurrencies per year, the total market capitalization at the end of the year (in Billion $), the ratio of the total market capitalization of our sample to the total market capitalization of the cryptocurrency market, the average volatility and the average number of accounts. Our sample of cryptocurrencies varies by year. The total number of cryptocurrencies increased from 20 in 2017 to 43 in 2021. Our sample covers at least 78% of the total market capitalization every year. Therefore it covers most of the representative

[3]We remove the following cryptocurrencies: Tether (USDT), USD Coin (USDC), Binance USD (BUSD), DAI (DAI), Paxos Standard (PAX), Wrapped Bitcoin (WBTC), renBTC (RENBTC).

[4]We construct returns at the weekly frequency to avoid outliers and day-of-the-week effect as in Biais et al. (2023).

cryptocurrencies in the market.

3.1.2 Newspapers

We collect newspaper articles from Factiva mentioning the top 43 cryptocurrencies by market capitalization as of December 2021. In particular, our search keywords are both the name and abbreviation of cryptocurrencies.[5] Our data span the period from June 2017 to December 2021. During this sample period, 27,382 articles satisfy our search criteria.

3.2 BERT Topic Modeling Approach

Our objective is to extract the most prominent topics from news articles in order to reduce noise and derive factors that provide valuable insights into the cross-section of cryptocurrency returns. Conventional topic modeling methods, such as Latent Dirichlet Allocation (LDA) and Latent Semantic Indexing (LSI), are widely used in the literature but exhibit certain limitations.[6] Specifically, these methods overlook the importance of word order and context in capturing the full meaning of text. In contrast, the Bidirectional Encoder Representations from Transformers (BERT) model (Devlin et al., 2018) overcomes these limitations by incorporating both word order and semantics, making it more effective for our analysis of complex financial news data. Therefore, we adopt the BERT model to explore the topics within our corpus of cryptocurrency-related articles.

[5]Articles from Factiva are collected from the following 47 publications from around the world: *The Cointelegraph*, *CoinDesk.com*, *Blockonomi*, *Dow Jones Newswires*, *express.co.uk* (UK), *PR Newswire*, *CE NoticiasFinancieras* (Latin America), *Investing.com*, *Financial Times*, *Reuters*, *iCrowdNewswire*, *The Wall Street Journal*, *M2 Presswire*, *The Independent*, *Blockchain.News*, *The Times* (UK), *Investor's Business Daily* (US), *The Telegraph* (UK), *MarketWatch*, *Brave New Coin*, *Sputnik News Service* (Russia), *Benzinga.com*, *Mondaq Business Briefing*, *Business Insider*, *CNN*, *Forbes*, *Business Wire*, *City AM* (London), *South China Morning Post*, *GlobeNewswire* (US), *Investment Weekly News*, *The Economic Times*, *ACCESSWIRE*, *Postmedia Breaking News* (Canada), *Hedge Week*, *Daily Mail*, *The Australian*, *Financial News* (Europe), *Exchange News Direct*, *Korea Times* (South Korea), *The Globe and Mail*, *Agence France Presse*, *Institutional Asset Manager*, *The Canadian Press*, *Barron's*, *Times of India*, *The New York Times*.

[6]LDA and LSI rely on the bag-of-words representation of documents, which ignores word order and semantics, often leading to a loss of context. See Blei et al. (2003) for a detailed discussion of LDA.

BERT's strength lies in its ability to pre-train deep bidirectional representations from unlabeled text by considering both the preceding and succeeding context of each word in all layers.[7] This context-aware structure enables BERT to generate more accurate embeddings compared to traditional word-level models. Once pre-trained, BERT can be fine-tuned with an additional output layer for specific tasks such as topic modeling.[8] Additionally, BERT enhances its precision by tokenizing words into subwords, which allows for more granular interpretation of text. Its ability to process sequences up to 512 tokens also makes it particularly well-suited for the analysis of long documents, further distinguishing it from earlier models.

Topic Modeling procedure. The input for BERT in our topic modeling approach is the corpus of cryptocurrency news articles. The steps in our procedure are as follows, and is illustrated in panel A of Figure 1:

1. **Document Embeddings**: Using Sentence Transformers, we extract document embeddings from the news articles. The pre-trained RoBERTa model, developed by Liu et al. (2019), is used to create these embeddings. Documents are transformed into vector representations, capturing their semantic meaning.[9]

2. **Dimensionality Reduction and Clustering**: We apply the Uniform Manifold Approximation and Projection (UMAP) algorithm (McInnes et al., 2018) to reduce the dimensionality of the document embeddings. UMAP reduces the vectors to five dimensions, optimizing

[7]BERT is pre-trained on large datasets such as BooksCorpus (800 million words) and Wikipedia (2.5 billion words), enabling it to capture the semantics of words effectively through pre-training.

[8]We use the BERT base model for embeddings, which consists of 12 layers, 768 hidden states, 12 attention heads, and 110 million parameters.

[9]There exist several methods for generating word embeddings, such as Word2Vec (Mikolov et al., 2017), GloVe (Pennington et al., 2014), and FastText (Joulin et al., 2016). These methods typically generate word-level embeddings, which often misinterpret context due to their limited capacity to encode entire sentences (Perone et al., 2018). In contrast, BERT focuses on contextual embeddings, where the input is a sentence rather than a single word. This allows BERT to account for both preceding and succeeding context, making it a bidirectional model. This property differentiates BERT from earlier models, such as ELMo and ULMFit, which are unidirectional.

the balance between local and global structure by using a neighborhood size of 15. Similar documents are clustered together.

3. **Topic Creation**: To create topics, we apply class-based TF-IDF (c-TF-IDF) to the clusters of documents. This approach aggregates the documents in each cluster and calculates the importance of terms within the cluster. The c-TF-IDF score is calculated as follows:

$$\text{c-TF-IDF}_i = \frac{t_i}{w_i} \times \log \frac{m}{\sum_j^n t_j} \tag{3}$$

where t_i is the frequency of term t in cluster i, w_i is the total number of words in the cluster, m is the total number of documents across all clusters, and n is the sum of occurrences of term t in all documents. The highest c-TF-IDF scores help us label the clusters based on the most relevant keywords.

The output of BERT topic modeling for our dataset includes 20 topics and the top 30 keywords associated with each topic. The topic identified as containing Fundamental content is characterized by keywords such as "mining", "hash", "operations", "network", and "technology". We provide a visual summary of these keywords in Panel B of Figure 1. Other topics include lending, regulation, payment, derivatives, social media, hedging, and technical trading, and word clouds for these topics are provided in Appendix E.[10]

[FIGURE 1 ABOUT HERE]

3.3 Fundamental Sentiment Index

BERT provides us with a sample of news articles classified as having fundamental content. Panel A of Figure 2 shows the raw number of these fundamental news articles over time. We

[10]Of the 20 topics, our classification identifies 1 fundamental topic, 4 derivatives topics, 4 regulation topics, 1 lending topic, 1 payment topic, 1 hedging topic, 6 technical trading topics, and 2 miscellaneous topics. For further details on the word clouds for each topic, we refer readers to Appendix E.

observe spikes in the number of fundamental news articles around significant events, such as the cryptocurrency mining malware in North Korea, the bitcoin mining blackout in China, and China's crackdown on cryptocurrency mining.

We calculate the sentiment of articles with fundamental trading content by counting the number of positive and negative words using the Loughran and McDonald (2011) dictionary. To reduce noise, we compute the sentiment only for sentences that mention the specific cryptocurrencies in our dataset. The sentiment measure is calculated as follows:

$$FSI = \frac{\text{Number of positive words-Number of negative words}}{\text{Total number of words}} \tag{4}$$

Here, FSI represents the fundamental sentiment index. An increase in the sentiment measure indicates higher optimism about fundamentals in the cryptocurrency market. For instance, the sentence *"coinhive reportedly had to shut down its services amidst a 50 percent decline in hash rate following the last Monero hard fork."* has a sentiment measure of -0.2.[11]

Summary statistics of the FSI are reported in Table 1. The FSI exhibits negative skewness and excess kurtosis, and it is stationary according to the augmented Dickey-Fuller test. Correlations between the FSI and other prominent risk factors in cryptocurrency pricing are also presented. Our FSI index shows a weak positive correlation with size, momentum, and volatility factors, and a weak negative correlation with the illiquidity factor. All correlations are below 0.11 and statistically significant. These results suggest that the FSI captures different dimensions of risk compared to conventional risk factors in the literature.

[TABLE 1 ABOUT HERE]

Panel B of Figure 2 displays the time-series of the FSI index. It further confirms the sta-

[11]For more details on the types of articles classified as fundamental, see Appendix B, where we provide additional examples of fundamental article sentences and their sentiment scores.

tionarity of the FSI and captures periods of both positive and negative fundamental sentiment. For example, the FSI spiked at the end of 2020 during Bitcoin's appreciation, and dropped in January 2022 following news of China's cryptocurrency ban.

[FIGURE 2 ABOUT HERE]

3.4 Token Classification

To categorize the various types of tokens, we adopt the classification framework from Cong and Xiao (2021), with an extension that distinguishes between platform tokens and governance tokens.[12] A detailed classification of the 43 cryptocurrencies included in this study is provided in Appendix C.

General Payment Tokens. General payment tokens are designed to facilitate transactions and act as a medium of exchange. Their primary function is to enable everyday transactions. An example is Bitcoin (BTC), which is used as a decentralized digital currency for peer-to-peer transactions.

Platform Tokens. Platform tokens are native to specific blockchain platforms and serve various purposes, including the payment of transaction fees, the execution of smart contracts, and the provision of access to platform resources. These tokens typically act as the primary medium of value transfer within their respective ecosystems. For instance, Ethereum (ETH), the native token of the Ethereum platform, is used to settle transaction fees, support the execution of smart contracts, and enable participation in decentralized applications (DApps) built on the platform.

[12] This distinction is essential for the analysis of the cross-sectional characteristics of currency betas, as discussed in Section 4.1.

Product Tokens. Product tokens are issued by specific projects or companies to represent value within their respective ecosystems. These tokens are often tied to a particular product, service, or utility. For example, Basic Attention Token (BAT) is used within the Brave browser ecosystem to compensate users for viewing advertisements and engaging with content, thereby supporting a more efficient and equitable digital advertising model.

Governance Tokens. Governance tokens confer decision-making rights to their holders, allowing participation in the governance of a blockchain platform or protocol. Holders of these tokens can propose and vote on changes to the system, including protocol upgrades and governance matters, with the aim of decentralizing control and ensuring community participation in the platform's evolution. MakerDAO's governance token (MKR) is a prominent example, granting holders voting power over decisions regarding the stability and governance of the MakerDAO platform, which issues the DAI stablecoin.

4 Empirical Results

In this section, we examine the pricing ability of the text-based fundamental factor for the cross-section of cryptocurrency returns. We also provide a comparison with other fundamental factors and augment existing cryptocurrency asset pricing models with the fundamental sentiment factor to explore its role in improving existing models.

4.1 Fundamental Sentiment: Beta Determinants

Estimating Betas. To measure the exposure of each cryptocurrency to the fundamental sentiment index (FSI), we regress individual cryptocurrency excess returns at time t on the FSI, controlling for additional cryptocurrency risk factors. These risk factors include the cryptocurrency market factor (MKT), size factor (SMB), momentum factor (MOM), volatility factor

(VOL), and liquidity factor ($ILLIQ$).[13] The estimated time-varying slope coefficient from this regression is $\beta_{i,t}^{FSI}$, as specified in the model below:

$$rx_{i,t} = \alpha_{i,t} + \beta_{i,t}^{FSI} FSI_t + \beta_{i,t}^{MKT} MKT_t + \beta_{i,t}^{SMB} SMB_t + \beta_{i,t}^{MOM} MOM_t + \beta_{i,t}^{VOL} VOL_t + \beta_{i,t}^{ILLIQ} ILLIQ_t + \epsilon_{i,t}, \tag{5}$$

where $rx_{i,t}$ is the cryptocurrency return at time t, and FSI_t represents the fundamental sentiment index at time t. Other factors are included to control for additional determinants of cryptocurrency returns.

Economic Interpretation of the Betas. Panel A of Figure 3 presents the time-series average of the betas with respect to the fundamental sentiment factor. Cryptocurrencies are sorted based on their beta values, with those on the left exhibiting the most negative exposure to the fundamental sentiment index, while those on the right show the most positive betas.

[FIGURE 3 ABOUT HERE]

Based on the token classification in Section 3.4, we observe significant cross-sectional differences in beta values across token types. Cryptocurrencies with high betas with respect to the fundamental sentiment index are typically platform or general payment tokens. These token returns co-move positively with our measure of fundamental sentiment, meaning that an increase in blockchain congestion or transaction benefits directly impacts their utility. As a result, these tokens are riskier, primarily because they serve as a medium of exchange or facilitate value transfers on blockchain platforms.

Conversely, cryptocurrencies with low (negative) betas are often governance tokens. Unlike platform and general payment tokens, governance tokens are not used for payments; instead, they are staked in protocols to vote on governance proposals. Thus, news captured in the

[13] Further details on risk factors and variable definitions are provided in Appendix C.

fundamental sentiment index, such as blockchain characteristics like the hash rate and mining-related costs, has a limited impact on the valuation of these tokens. As a result, governance tokens provide a hedge against changes in fundamental sentiment.

Fundamental Sentiment Betas and Value Factors. Next, we explore characteristics that explain the cross-sectional variation in fundamental sentiment betas. We perform Fama and MacBeth (1973) cross-sectional regressions, using various measures of cryptocurrency value to predict β^{FSI}.[14] The model is as follows:

$$\hat{\beta}_{i,t}^{FSI} = \lambda_{0,t} + \lambda_{1,t} Value_{i,t} + \epsilon_{i,t},$$

where $\hat{\beta}_{i,t}^{FSI}$ represents the estimated FSI betas, and $Value$ denotes different measures of cryptocurrency value (Cong et al., 2021). These value measures include the transaction-to-market cap ratio, user-to-market cap ratio, and address-to-market cap ratio, which capture the network effects of cryptocurrencies with higher levels of transactions, users, and addresses.

Cryptocurrencies with higher value measures, such as the transaction-to-market cap ratio, user-to-market cap ratio, and address-to-market cap ratio, tend to have higher betas with respect to the fundamental sentiment index. These tokens are often used for payments or as part of platform services, making them more sensitive to changes in blockchain activity and transaction benefits. In contrast, governance tokens, which are primarily used for staking and voting on proposals, are less frequently traded or used for payments. As a result, they exhibit lower value metrics and tend to have lower exposure to fundamental sentiment.

Table 2 presents the average coefficients of contemporaneous cross-sectional regressions, estimated on a weekly basis. We find that value factors are strong positive predictors of fundamental sentiment betas, highlighting the connection between value factors and the fundamental

[14] See Appendix C for variables used in the analysis.

sentiment index. The cross-sectional R^2 values range from 9% to 12%, demonstrating the ability of the BERT model to capture meaningful topics that explain cryptocurrency risk premia.

[TABLE 2 ABOUT HERE]

4.2 Fundamental Sentiment Portfolio

Construction. We construct portfolios based on each cryptocurrency's exposure to the fundamental sentiment index (FSI). At time t, we sort cryptocurrencies into quartiles based on their previous week's ($t-1$) beta with FSI. These portfolios are rebalanced weekly. The first portfolio (P_1) includes cryptocurrencies with the lowest FSI betas, while the fourth portfolio (P_4) consists of those with the highest FSI betas. A zero-cost portfolio (HML_{FSI}) is created by going short on P_1 and long on P_4.

Summary Statistics. If FSI is a valid pricing factor for the cross-section of cryptocurrency returns, there should be a notable difference in excess returns between low-beta and high-beta portfolios. Thus, the HML_{FSI} portfolio should generate statistically significant excess returns. Table 3 presents summary statistics for portfolios sorted by their exposure to FSI (β_{FSI}).

The results in Table 3 indicate that long positions in cryptocurrencies with the highest FSI exposure (β_{FSI}) and short positions in cryptocurrencies with the most negative FSI exposure yield positive excess returns. The average portfolio returns increase monotonically with FSI beta. The HML_{FSI} portfolio achieves an annualized average excess return of 65% with a Sharpe ratio of 1.24 per annum. Additionally, the fundamental sentiment strategy exhibits positive skewness and excess kurtosis, which further supports its profitability.

[TABLE 3 ABOUT HERE]

Panel B of Figure 3 displays cumulative returns of the HML_{FSI} strategy compared to the market factor. The HML_{FSI} strategy outperforms the market portfolio over the sample period, showing that it is consistently profitable and less susceptible to major downturns in the cryptocurrency market. Notably, the market portfolio performs poorly during the 2019-2020 period and the November 2021 cryptocurrency crash. However, between the end of 2020 and April 2021, the market portfolio shows positive performance, a period marked by strong gains in Bitcoin and other major cryptocurrencies.

Robustness Tests. Several robustness tests were conducted to ensure the validity of our results in Appendix D. First, we analyzed the portfolio turnover for the HML_{FSI} strategy and found that certain cryptocurrencies, such as GNO, NEO, ADA, and DOGE, consistently appear in either the low or high FSI portfolios. This suggests that the strategy's performance is driven by specific currencies with persistent fundamental sentiment exposures.

We also re-ran the analysis using only the top 15 cryptocurrencies by market capitalization, and our results remained robust. The HML_{FSI} portfolio for this subset still achieved a strong excess return with a Sharpe ratio of 1.04, indicating that our findings are not biased by smaller cryptocurrencies.

In addition, we tested alternative factor specifications to estimate β^{FSI}, including models that control only for market, size, and momentum factors. Across all specifications, the HML_{FSI} portfolio consistently generated positive and statistically significant excess returns, further confirming the robustness of our results.

Lastly, we explored alternative sentiment proxies and different numbers of portfolios (e.g., terciles and quintiles). These tests showed that the HML_{FSI} strategy remained profitable, yielding annualized returns of up to 62% with Sharpe ratios exceeding 1.2 per annum, regardless of the sentiment measure or portfolio grouping used.

4.2.1 FSI and Other Cryptocurrency Factors

Traditional Cryptocurrency Factors. We first test whether our sentiment factor offers significant alphas after controlling for traditional risk factors, such as market, size, momentum, liquidity, and volatility. Column 1 of Table 4 presents the results of a contemporaneous regression of the spread portfolio based on the fundamental strategy (HML_{FSI}) on the market factor. The market factor coefficient is positive but statistically insignificant, while the alpha is 67.6% annually and statistically significant, with a Newey and West (1987) t-statistic of 2.66.

Table 4 also examines the relationship between HML_{FSI} and other conventional investment strategies. In a two-factor model that includes the market and size factors, the FSI strategy yields an alpha of 78%, which is statistically significant. When a momentum factor is added to the model, the FSI strategy offers an alpha of 72.8%. Further augmenting the model with liquidity and volatility factors results in a robust alpha of 72.8%, statistically significant at the 1% level. These results suggest that the HML_{FSI} strategy generates a positive and statistically significant alpha even when controlling for traditional asset pricing models.

[TABLE 4 ABOUT HERE]

Value-Based Factors. Next, we explore the relationship between HML_{FSI} and value-related risk factors in the cryptocurrency literature. Specifically, we test whether HML_{FSI} can be explained by fundamental risk factors constructed in Cong et al. (2021). To do this, we regress HML_{FSI} on three value factors to see if they can account for the returns generated by the FSI strategy.

Table 5 presents the results, using three independent variables: the transaction-to-market ratio (T/M), the user-to-market ratio (U/M), and the address-to-market ratio (A/M). The coefficients for all three value factors are positive and strongly significant, with t-statistics of 3.90, 3.76, and 3.63, respectively. This supports the hypothesis that our fundamental sentiment measure reflects over- or under-valuation of cryptocurrencies. Importantly, the alphas remain positive and statistically significant across all regressions, indicating that while value factors are correlated with HML_{FSI}, they do not fully explain its returns. This suggests that HML_{FSI} captures a different dimension of fundamental cryptocurrency characteristics beyond these value factors.

[TABLE 5 ABOUT HERE]

4.2.2 Asset Pricing Tests

We now present the framework for our asset pricing tests. Under standard conditions, a stochastic discount factor (SDF), M_t, can price the excess returns of any asset i, $rx_{i,t}$. This relationship is expressed as:

$$\mathbb{E}[M_t rx_{i,t}] = 0 \tag{6}$$

Following Bhambhwani et al. (2023), we assume that the SDF is a linear function of observable factors F_t, where μ_F is the mean of the factors, f_t is the set of factors centered around their means, and b is a vector of parameters:

$$M_t = 1 - b'(F_t - \mu_F) \tag{7}$$

Using this SDF equation, we can express expected returns as a linear function of factor betas:

$$\mathbb{E}[rx_{i,t}] = \lambda' \beta_i \tag{8}$$

where β_i represents the exposure of returns to factor i, and λ is the risk price associated with factor i.[15]

Test Assets. In our first method, we use individual cryptocurrencies as test assets rather than portfolios. This decision is motivated by Ang et al. (2018), who argue that grouping assets into

[15] $\beta_i = \mathbb{E}[(f_t - \mu_F)(f_t - \mu_F)']^{-1}\mathbb{E}[(f_t - \mu_F)' rx_{i,t}]$ is the vector of factor betas for cryptocurrency i, and $\lambda = \mathbb{E}[(f_t - \mu_F)(f_t - \mu_F)']b$ is the vector of risk prices.

portfolios reduces the cross-sectional dispersion of betas and results in less efficient estimates of factor risk premia. Therefore, we focus on the relationship between betas and returns at the individual cryptocurrency level.

Cross-Sectional Regressions. After estimating β^{FSI} from Equation (5), we examine how these betas relate to expected excess returns at the cryptocurrency level. To do so, we run weekly cross-sectional regressions of the form:

$$rx_{i,t+1} = \lambda_{0,t} + \lambda_{1,t}\hat{\beta}_{i,t}^{FSI} + \lambda_{2,t}X_{i,t} + \epsilon_{i,t+1}, \tag{9}$$

where $X_{i,t}$ includes control variables such as β^{MKT}, β^{Size}, $\beta^{Momentum}$, and $\beta^{Volatility}$, all of which are estimated from Equation (5). We then compute the time-series average of the slope coefficients, $\lambda_{1,t}$, and report the Newey and West (1987) t-statistics alongside the average adjusted R^2.

Table 6 presents the estimated risk prices of β^{FSI}. In the univariate regression (Column 1), we find a strong positive relationship between β^{FSI} and future cryptocurrency returns. The coefficient on β^{FSI} is 0.004, with a t-statistic of 2.61, indicating that cryptocurrencies with higher β^{FSI} tend to have higher expected returns. This result is consistent with the portfolio sorting results in Table 3, where a long position in high-β^{FSI} portfolios predicts higher future returns.

To assess the economic significance, we compute the difference in average β^{FSI} between portfolios P_1 and P_4 from Table 3. The difference is 2.76 [= 1.46 - (-1.30)], implying that moving from P_1 to P_4 increases expected returns by 1.10% per week [= 2.76 × 0.004]. Even after controlling for market, volatility, and momentum factors, the coefficient on β^{FSI} remains robust, increasing to 0.005 with a t-statistic of 2.00. These results confirm that the fundamental sentiment index is a robust predictor of cryptocurrency returns, even when accounting for other risk premia.

[TABLE 6 ABOUT HERE]

Having demonstrated that β^{FSI} has strong predictive power for next-week cryptocurrency returns, we now explore whether these sentiment factors can predict returns over longer horizons. We regress excess returns from $t+2$ weeks to $t+12$ weeks on β^{FSI} at time t, along with the same control variables used in the previous analysis.

Table 7 shows that the coefficient on β^{FSI} is positive and statistically significant for up to 5 weeks ahead, after which it gradually diminishes and is no longer significant after week 10. This pattern suggests that the strategy tends to reverse between weeks 5 and 11.

[TABLE 7 ABOUT HERE]

Fama-MacBeth Two-Pass Regressions. Next, we employ the Fama and MacBeth (1973) two-pass regression methodology. In the first pass, we estimate factor loadings using time-series regressions, and in the second pass, we estimate risk prices using cross-sectional regressions. For this analysis, we use 24 test assets, which include portfolios sorted by size, momentum, liquidity, volatility, value, and fundamental sentiment.[16]

The first stage involves time-series regressions of excess returns on factor exposures:

$$rx_{i,t} = \alpha_{i,t} + \beta_i^{FSI} HML_{FSI,t} + \beta_i^{MKT} MKT_t + \epsilon_{i,t} \tag{10}$$

In the second stage, cross-sectional regressions of average portfolio returns on the estimated factor loadings yield the factor risk prices:

$$\overline{rx}_i = \lambda_{0,i} + \lambda_i^{FSI} \hat{\beta}_i^{FSI} + \lambda_i^{MKT} \hat{\beta}_i^{MKT} + \epsilon_i \tag{11}$$

We augment the standard factor models with the fundamental sentiment factor (HML_{FSI}) and assess its performance across the 24 test assets. Table 8 reports the estimated risk prices,

[16] This broader cross-section of 24 test assets helps account for the possibility of spurious factors, as noted by Lewellen et al. (2010) and Harvey et al. (2022).

t-statistics, root mean square error (RMSE), and cross-sectional R^2.In Panel A, we begin with a two-factor model that includes the market (MKT) and size factors. The results show that the market factor is not significantly priced in the cross-section of cryptocurrency returns, consistent with prior literature, while the size factor is positively priced. Adding the HML_{FSI} factor to this model significantly improves the fit, increasing the cross-sectional R^2 from 15% to 36%. The HML_{FSI} factor has a statistically significant risk premium of 1.2% per week.

In Panel B, we extend the model by adding a momentum factor alongside the market and size factors. The inclusion of HML_{FSI} increases the cross-sectional R^2 from 17% to 38%, with the price of risk for HML_{FSI} remaining positive and statistically significant.

Finally, in Panel C, we consider a five-factor model that includes the market, size, momentum, liquidity, and volatility factors. Incorporating HML_{FSI} raises the cross-sectional R^2 from 25% to 39%, with the price of risk for the fundamental sentiment factor being statistically significant at the 5% level.

In summary, adding the HML_{FSI} factor significantly improves the explanatory power of standard asset pricing models for cryptocurrencies. Our risk premium estimate for β^{FSI} is 1.2% per week in the baseline model from Panel A, and the results remain robust when incorporating alternative factor models. These findings confirm that fundamental sentiment is a key driver of cryptocurrency returns.

[TABLE 8 ABOUT HERE]

4.2.3 Diversification Benefits

Table 9 evaluates the diversification benefits provided by the fundamental sentiment factor when combined with other well-established factors from the literature. Specifically, we assess whether adding the fundamental sentiment factor enhances the Sharpe ratios of the market, size, illiquidity, volatility, and momentum factors. To do this, we create equally weighted portfolios by blending each factor with the fundamental sentiment factor.

Panel A presents summary statistics for each individual factor. Panel B reports the summary statistics for the blended portfolios that combine each factor with the HML_{FSI} factor. The final row of Panel B displays the weight of each factor in the blended portfolio. Additionally, the last column in Panel B shows the results for an equally weighted portfolio of all factors, where the fundamental sentiment factor carries a weight of 16%.

Our findings demonstrate that the fundamental sentiment factor provides significant diversification benefits across all strategies considered. For example, the annualized Sharpe ratio of the market portfolio improves from 0.08 to 0.71, for size from 1.33 to 2.00, for illiquidity from 0.45 to 1.47, for volatility from 1.17 to 1.70, and for momentum from 0.02 to 1.02. Similarly, the Sharpe ratio of the equally weighted portfolio that includes all factors increases from 0.91 to 1.41 per annum.

Overall, these results indicate that incorporating the fundamental sentiment factor into a variety of existing factor strategies significantly enhances risk-adjusted performance, offering strong diversification benefits.

[TABLE 9 ABOUT HERE]

4.3 Sentiment of Other Topics

To address concerns that our findings may be the result of data mining, we construct sentiment indices for additional topics identified by BERT topic modeling. This allows us to examine the role of other types of sentiment in the cryptocurrency market. The topics include lending, regulation, payment, derivatives, social media, hedging, and technical trading. The results are presented in Table 10, with word clouds for these topics provided in Appendix E.

For most of these topics, the sentiment does not predict cross-sectional cryptocurrency returns, and the HML (high-minus-low) portfolios constructed using these factors are statistically insignificant. The only exception is the sentiment associated with technical trading. We find that a strategy that goes long on cryptocurrencies with low technical sentiment (TSI) and short on those with high technical sentiment yields an annualized return of 71% and a Sharpe ratio of 1.30.

[TABLE 10 ABOUT HERE]

The reliance on price patterns may be more pronounced in markets like cryptocurrencies, which lack standardized accounting frameworks for assessing fundamentals.[17] We observe that media discussions about price movements significantly impact technical trading in the cryptocurrency market. This is especially relevant in the absence of a standardized accounting framework that provides reliable financial measurements. For example, Detzel et al. (2021), among others, highlight the predictive power of 1- to 20-week moving averages of daily bitcoin prices, both in-sample and out-of-sample. They show, through an equilibrium model, that in the presence of uncertainty about fundamental growth, rational learning by investors with differing priors can lead to strong predictability of returns via moving average rules. We hypothesize that sentiment surrounding price movements, particularly in media discussions by experts, provides valuable information for the cross-section of cryptocurrency returns.

To test the robustness of this finding, we conduct Fama and MacBeth (1973) regressions in Appendix E, where we regress cryptocurrency returns at time $t + 1$ on TSI betas and various control factors (size, momentum, liquidity, and volatility) at time t. We find that technical

[17]Several studies, including Treynor and Ferguson (1985), Brown and Jennings (1989), Hong and Stein (1999), Cespa and Vives (2012), Edmans et al. (2015), Han et al. (2016), and Keloharju et al. (2019), demonstrate that in imperfect markets, past prices can provide valuable information about future price movements. This implies that technical indicators based on price history could be influential trading signals. For instance, Brock et al. (1992) and Lo et al. (2000) present empirical evidence that technical indicators are profitable in stock markets.

sentiment remains a strong negative predictor of the cross-section of cryptocurrency returns, even after controlling for these factors. In Appendix E, we extend the analysis by including FSI betas in the regression to determine whether one factor subsumes the predictive power of the other. Both factors are found to be priced in the cross-section of cryptocurrency returns, though with opposite signs, indicating that they provide distinct information. This result is not surprising, as the two factors exhibit a low correlation of just 0.04.

In summary, technical sentiment plays a significant role in predicting the cross-section of cryptocurrency returns. Investors demand a risk premium for holding cryptocurrencies with high pessimism regarding technical trading. Additionally, this factor appears to be independent of fundamental factors, which offer unique information in the cryptocurrency market.

5 Conclusion

This paper investigates the cross-sectional predictive ability of text-based fundamental sentiment in the cryptocurrency market. Using news articles covering the top 43 cryptocurrencies, we apply BERT topic modeling to focus on fundamentals and construct a Fundamental Sentiment Index (FSI) that captures the balance of positive and negative sentiment in blockchain-related news. Our study examines how this sentiment influences cryptocurrency returns and risk premia.

We test two key hypotheses: first, that cryptocurrencies with higher exposure to fundamental sentiment are riskier and thus demand higher expected returns, and second, that sensitivity to blockchain fundamentals differs based on the token's use case. Our findings show that platform and payment tokens, which are more exposed to fundamental sentiment, are riskier, while governance tokens tend to serve as a hedge against fundamental risks.

By estimating betas with respect to the FSI, we find that cryptocurrencies with higher FSI exposure tend to co-move positively with blockchain fundamentals, such as transaction benefits and blockchain congestion. A long-short portfolio strategy based on this exposure generates significant returns, with a Sharpe ratio of 1.24, outperforming the broader cryptocurrency market. These results are robust across various portfolio constructions, alternative sentiment proxies, and sample groups.

We also perform Fama and MacBeth (1973) cross-sectional regressions to assess the pricing of the FSI factor. Our results indicate that cryptocurrencies with higher exposure to fundamental sentiment are riskier, and investors demand a risk premium for holding these assets. In our baseline model, the price of risk for the FSI factor is 1.2 percent per week. This result remains robust even when controlling for traditional risk factors such as market, size, momentum, volatility, and liquidity.

Lastly, we demonstrate that the FSI factor provides meaningful diversification benefits when combined with traditional asset pricing factors. Adding the fundamental sentiment factor to strategies based on market, size, illiquidity, volatility, and momentum significantly improves risk-adjusted returns. For instance, the Sharpe ratio of an equally weighted portfolio of all factors increases from 0.91 to 1.41 per annum.

In conclusion, our findings underscore the importance of text-based fundamental sentiment as a key determinant of cryptocurrency returns. Investors should consider fundamental sentiment as part of their investment strategies, as it offers valuable insights beyond those captured by traditional factor models.

References

Abarbanell, J. S. and B. J. Bushee (1997). Fundamental analysis, future earnings, and stock prices. *Journal of accounting research 35*(1), 1–24.

Abarbanell, J. S. and B. J. Bushee (1998). Abnormal returns to a fundamental analysis strategy. *Accounting Review*, 19–45.

Ang, A., J. Liu, and K. Schwarz (2018). Using stocks or portfolios in tests of factor models. In *AFA 2009 San Francisco Meetings Paper*.

Bhambhwani, S. M., S. Delikouras, and G. M. Korniotis (2023). Blockchain characteristics and cryptocurrency returns. *Journal of International Financial Markets, Institutions and Money 86*, 101788.

Biais, B., C. Bisiere, M. Bouvard, C. Casamatta, and A. J. Menkveld (2023). Equilibrium bitcoin pricing. *The Journal of Finance 78*(2), 967–1014.

Bianchi, D. and M. Babiak (2021). A factor model for cryptocurrency returns. *Available at SSRN 3935934*.

Bianchi, D., M. Babiak, and A. Dickerson (2022). Trading volume and liquidity provision in cryptocurrency markets. *Journal of Banking & Finance 142*, 106547.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of machine Learning research 3*(Jan), 993–1022.

Brock, W., J. Lakonishok, and B. LeBaron (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of finance 47*(5), 1731–1764.

Brown, D. P. and R. H. Jennings (1989). On technical analysis. *The Review of Financial Studies 2*(4), 527–551.

Cespa, G. and X. Vives (2012). Dynamic trading and asset prices: Keynes vs. hayek. *The Review of Economic Studies 79*(2), 539–580.

Cong, L. W., G. A. Karolyi, K. Tang, and W. Zhao (2021). Value premium, network adoption, and factor pricing of crypto assets. *Network Adoption, and Factor Pricing of Crypto Assets (December 2021)*.

Cong, L. W. and Y. Xiao (2021). Categories and functions of crypto-tokens. *The Palgrave Handbook of FinTech and Blockchain*, 267–284.

Detzel, A., H. Liu, J. Strauss, G. Zhou, and Y. Zhu (2021). Learning and predictability via technical analysis: evidence from bitcoin and stocks with hard-to-value fundamentals. *Financial Management 50*(1), 107–137.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Edmans, A., I. Goldstein, and W. Jiang (2015). Feedback effects, asymmetric trading, and the limits to arbitrage. *American Economic Review 105*(12), 3766–3797.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy 81*(3), 607–636.

Filippou, I., D. Rapach, and C. Thimsen (2023). Cryptocurrency return predictability: A machine-learning analysis. *Available at SSRN*.

Han, W., D. Newton, E. Platanakis, C. Sutcliffe, and X. Ye (2022). On the (almost) stochastic dominance of cryptocurrency factor portfolios & implications for cryptocurrency asset pricing.

Han, Y., G. Zhou, and Y. Zhu (2016). A trend factor: Any economic gains from using information over investment horizons? *Journal of Financial Economics 122*(2), 352–375.

Harvey, C. R., T. Abou Zeid, T. Draaisma, M. Luk, H. Neville, A. Rzym, and O. Van Hemert (2022). An investor's guide to crypto. *The Journal of Portfolio Management 49*(1), 146–171.

Helms, K. (2020). $8.8 trillion traded in cryptocurrency spot and futures markets in q1: Reports. *Bitcoin. com News*.

Hong, H. and J. C. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance 54*(6), 2143–2184.

Joulin, A., E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov (2016). FastText.zip: Compressing text classification models. Technical report, arXiv:1612.03651.

Keloharju, M., J. T. Linnainmaa, and P. M. Nyberg (2019). Are return seasonalities due to risk or mispricing? evidence from seasonal reversals. *Journal of Financial Economics (JFE), Forthcoming*.

Kogan, S., I. Makarov, M. Niessner, and A. Schoar (2022). Are cryptos different? evidence from retail trading. *Evidence from Retail Trading (November 30, 2022)*.

Lev, B. and S. R. Thiagarajan (1993). Fundamental information analysis. *Journal of Accounting research 31*(2), 190–215.

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial economics 96*(2), 175–194.

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Liu, Y., J. Sheng, and W. Wang (2021). Technology and cryptocurrency valuation: Evidence from machine learning. *Available at SSRN 3577208*.

Liu, Y., A. Tsyvinski, and X. Wu (2021). Accounting for cryptocurrency value. *Available at SSRN 3951514*.

Liu, Y., A. Tsyvinski, and X. Wu (2022). Common risk factors in cryptocurrency. *The Journal of Finance 77*(2), 1133–1177.

Lo, A. W., H. Mamaysky, and J. Wang (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The journal of finance 55*(4), 1705–1765.

Loughran, T. and B. McDonald (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of finance 66*(1), 35–65.

Luo, D., T. Mishra, L. Yarovaya, and Z. Zhang (2021). Investing during a fintech revolution: Ambiguity and return risk in cryptocurrencies. *Journal of International Financial Markets, Institutions and Money 73*, 101362.

McInnes, L., J. Healy, and J. Melville (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Mikolov, T., K. Chen, G. S. Corrado, and J. A. Dean (2017, August 22). Computing numeric representations of words in a high-dimensional space. US Patent 9,740,680.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Pennington, J., R. Socher, and C. D. Manning (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543.

Perone, C. S., R. Silveira, and T. S. Paula (2018). Evaluation of sentence embeddings in downstream and linguistic probing tasks. Technical report, arXiv:1806.06259.

Schwenkler, G. and H. Zheng (2020). News-driven peer co-movement in crypto markets. *Available at SSRN 3572471*.

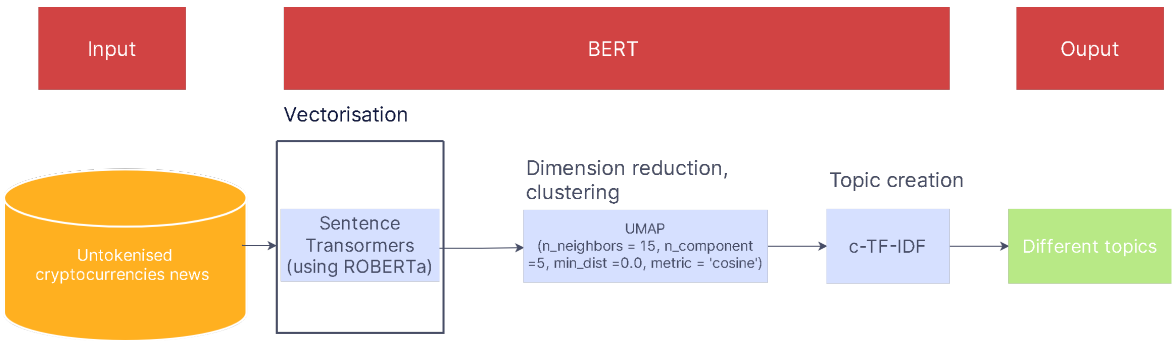
Shanken, J. (1992). On the estimation of beta-pricing models. *The review of financial studies 5*(1), 1–33.

Treynor, J. L. and R. Ferguson (1985). In defense of technical analysis. *The Journal of Finance 40*(3), 757–773.

Figure 1: BERT Topic Modeling and Fundamental Topic Analysis

This figure shows the results from BERT topic modeling and fundamental topic analysis. Panel A displays a summary of the BERT algorithm used for topic modeling, while Panel B highlights the keywords for the Fundamental topic. The data is weekly from June 2017 to December 2021.

Panel A: BERT Model Summary



Panel B: Fundamental Topic from BERT Topic Modeling

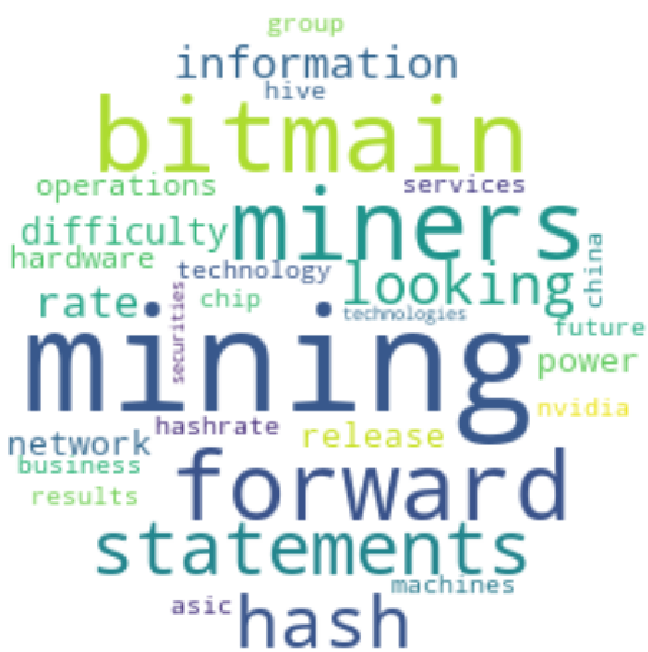
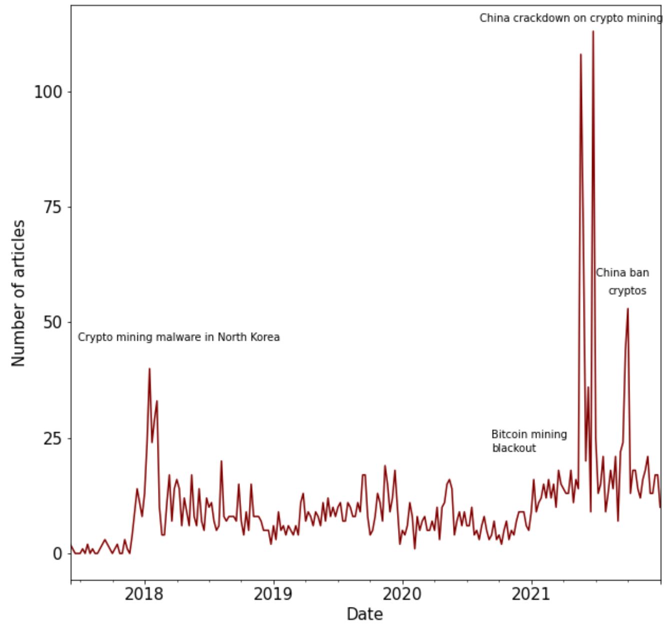


Figure 2: Fundamental News Articles and Sentiment Index

This figure illustrates the number of fundamental news articles and the Fundamental Sentiment Index. Panel A presents the number of news articles, while Panel B tracks the sentiment based on news data. The data is weekly from June 2017 to December 2021.

Panel A: Fundamental News Articles



Panel B: Fundamental Sentiment Index

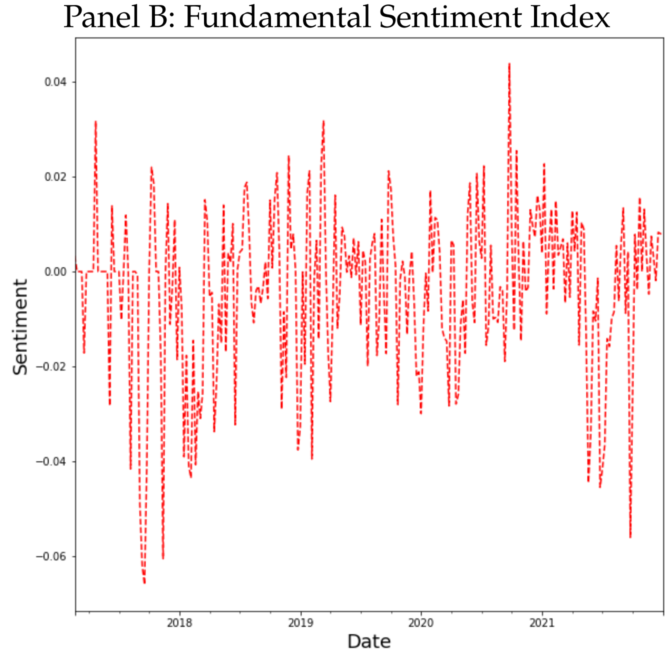
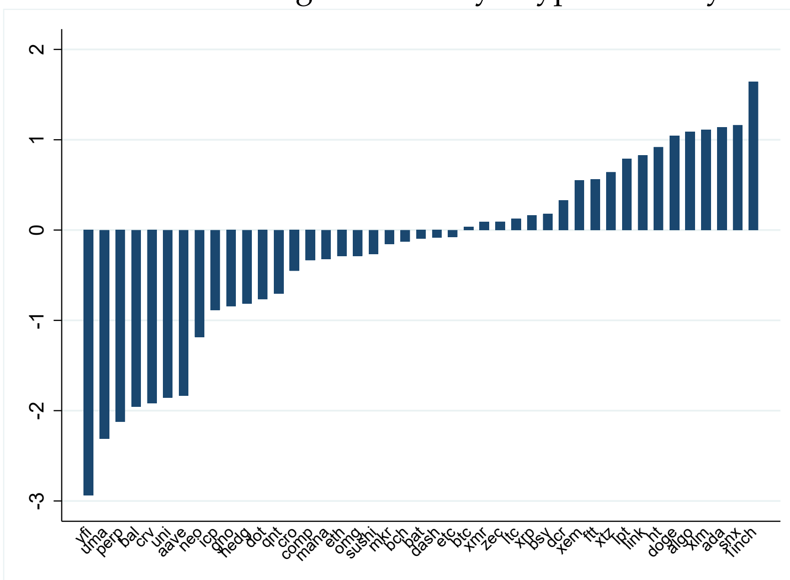


Figure 3: Average FSI Beta and Cumulative Returns

This figure presents the average FSI beta and cumulative returns. Panel A shows the average FSI beta by cryptocurrency, while Panel B presents the cumulative returns of the Fundamental Sentiment Index strategy (HML_{FSI}) and the market portfolio (MKT). The data is weekly from June 2017 to December 2021.

Panel A: Average FSI Beta by Cryptocurrency



Panel B: Cumulative Returns of Fundamental Sentiment Index Strategy

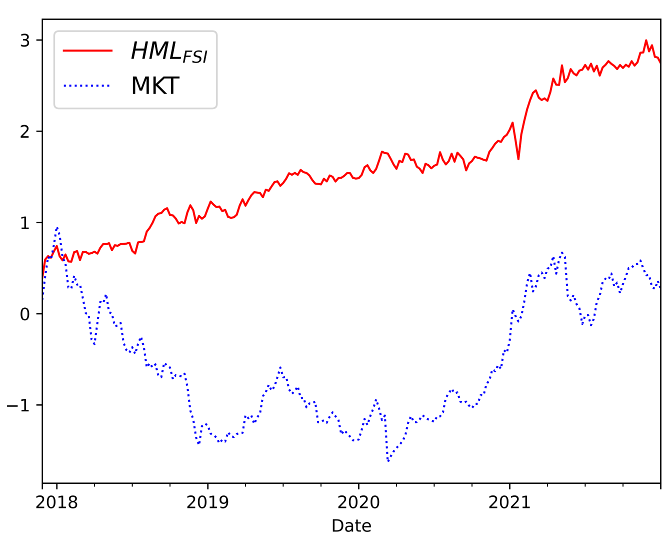


Table 1: Summary Statistics and Correlations with Existing Cryptocurrency Risk Factors

This table reports summary statistics of the Fundamental Sentiment Index (FSI) in Panel A. Correlations between portfolio ranking of the beta of the Fundamental Sentiment Index (FSI), and the portfolio rank of size, momentum, liquidity, and volatility are reported in Panel B. p-values are reported in brackets. ** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. Weekly data are between June 2017 and December 2021.

Panel A: Summary Statistics of FSI

	Mean	SD	Min	Max	Skewness	Kurtosis	Dickey-Fuller t-statistics
FSI	-0.004	0.018	-0.066	0.044	-0.849	4.021	-7.344***

Panel B: Correlations of Portfolio Ranks

Variables	β_{FSI}	Size	Momentum	Volatility	Liquidity
FSI	1.00				
Size	0.11	1.00			
	(0.00)				
Momentum	0.07	0.11	1.00		
	(0.00)	(0.00)			
Volatility	0.04	-0.29	0.19	1.00	
	(0.00)	(0.00)	(0.00)		
Liquidity	-0.07	-0.63	-0.07	0.31	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	

Table 2: Cross-Sectional Regressions

This table reports Fama and MacBeth (1973) cross-sectional regressions for Fundamental Sentiment Index betas β^{FSI}. We run the following model:

$$\hat{\beta}_{i,t}^{FSI} = \lambda_{0,t} + \lambda_{1,t} Value_{i,t} + \epsilon_{i,t}$$

where $\hat{\beta}_{i,t}^{FSI}$ denotes the 60-week rolling betas with the FSI index. $Value$ represents different measures of cryptocurrency value, as in Cong et al. (2021). t-statistics are reported in squared brackets, where *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. The data are weekly from June 2017 to December 2021.

Dependent variable: β^{FSI}			
	(1)	(2)	(3)
Value (T/M ratio)	0.308***		
	(3.70)		
Value (U/M ratio)		0.032***	
		(4.27)	
Value (A/M ratio)			2.379***
			(3.79)
Constant	-0.065**	-0.054**	-0.038**
	(-3.72)	(-2.57)	(-1.99)
Observations	5,966	5,751	5,966
R^2	0.12	0.09	0.09

Table 3: Portfolios sorted on Fundamental Sentiment Index

This table reports summary statistics for the excess returns of 4 cryptocurrency portfolios sorted on exposure to the Fundamental Sentiment Index β^{FSI} (Panel B). Portfolio 1 (P_1) contains cryptocurrencies with the lowest β^{FSI}, and Portfolio 4 (P_4) contains cryptocurrencies with the highest β^{FSI}. HML represents the portfolio that has a long position in the high beta portfolio (P_4) and a short position in the low beta portfolio (P_1). For each portfolio, we report annualized mean and its t-statistics (reported in squared brackets), standard deviation (Std), skewness, kurtosis, and Sharpe ratios (SR). The data are weekly from June 2017 and December 2021.

Fundamental Sentiment Index Portfolio					
	P_1	P_2	P_3	P_4	HML_{FSI}
Mean	-0.64	-0.47	-0.21	0.01	0.65
					[2.52]
Std	0.95	0.98	1.00	0.99	0.52
Skewness	-0.74	-0.52	-0.10	-0.42	0.67
Kurtosis	4.71	5.20	5.64	4.98	6.86
β	-1.30	-0.27	0.33	1.46	2.76
SR					1.24

Table 4: Fundamental Sentiment Sorted Portfolio Profit and other Risk factors

This table reports contemporaneous time-series regressions of HML_{FSI} on the market factor, size factor, momentum factor, liquidity factor, and volatility factor. The alphas are annualized. t-statistics are reported in squared brackets, where *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. The data are weekly from June 2017 and December 2021.

	Dependent variable: HML_{FSI}				
	(1)	(2)	(3)	(4)	(5)
Constant	0.676***	0.780***	0.728***	0.728***	0.728***
	(2.66)	(3.31)	(3.30)	(3.34)	(3.34)
Market factor_t	0.079	0.073	0.079	0.070	0.070
	(1.56)	(1.39)	(1.47)	(1.34)	(1.34)
Size factor_t		-0.153	-0.109	-0.064	-0.069
		(-1.49)	(-1.11)	(-0.67)	(-0.78)
Momentum factor_t			-0.236**	-0.226**	-0.225**
			(-2.55)	(-2.40)	(-2.38)
Liquidity factor_t				0.414**	0.423**
				(2.16)	(2.17)
Volatility factor_t					-0.041
					(-0.24)
Observations	214	214	214	214	214
Adj R^2	0.01	0.03	0.10	0.13	0.12

Table 5: Fundamental Sentiment Sorted Portfolio Profit and Value Risk factors

This table reports contemporaneous time-series regressions of HML_{FSI} on value factors as in Cong et al. (2021). The alphas are annualized. t-statistics are reported in squared brackets, where *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. The data are weekly from June 2017 and December 2021.

Dependent variable: HML_{FSI}			
	(1)	(2)	(3)
Value factor (T/M ratio)	0.309***		
	(3.90)		
Value factor (U/M ratio)		0.276***	
		(3.76)	
Value factor (A/M ratio)			0.264***
			(3.63)
Constant	0.572**	0.468**	0.572**
	(2.57)	(2.07)	(2.39)
Observations	214	214	214
R^2	0.14	0.13	0.12

Table 6: Cross-Sectional regressions

This table reports Fama Macbeth cross-sectional regressions for Fundamental Sentiment Index betas β^{FSI}. We run the model below:

$$rx_{i,t+1} = \lambda_{0,t} + \lambda_{1,t}\hat{\beta}^{SI}_{i,t} + \lambda_{2,t}X_{i,t} + \epsilon_{i,t+1}$$

where $rx_{i,t+1}$ is the individual cryptocurrency return, FSI. We report t-statistics are reported in squared brackets, where *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. The data are weekly from June 2017 and December 2021.

	Fundamental Sentiment Index betas β^{FSI}					
	(1)	(2)	(3)	(4)	(5)	(6)
β_t^{FSI}	0.004***	0.004**	0.004***	0.005**	0.005**	0.005**
	(2.61)	(2.45)	(2.65)	(2.23)	(2.28)	(2.00)
β_t^{MKT}		-0.005	-0.005	-0.002	-0.007	-0.006
		(-0.63)	(-0.66)	(-0.30)	(-0.80)	(-0.73)
$Size_t$			-0.002*	-0.002	-0.002*	-0.002*
			(-1.69)	(-1.60)	(-1.79)	(-1.71)
$Momentum_t$				0.002	0.002	0.006
				(0.15)	(0.23)	(0.51)
$Liquidity_t$					0.228	0.263
					(1.10)	(1.23)
$Volatility_t$						-0.156
						(-1.13)
Constant	0.001	0.006	0.043*	0.040	0.048*	0.056**
	(0.12)	(0.56)	(1.73)	(1.61)	(1.94)	(2.02)
Observations	6,138	6,138	6,138	6,138	6,138	6,138
R^2	0.05	0.11	0.15	0.23	0.30	0.35

Table 7: Long-term predictive power of Fundamental Sentiment Index

This table reports Fama Macbeth cross-sectional regressions for Fundamental Sentiment Index betas (β^{FSI}). We run the model below:

$$rx_{i,t+n} = \lambda_{0,t} + \lambda_{1,t}\hat{\beta}^{FSI}_{i,t} + \lambda_{2,t}X_{i,t} + \epsilon_{i,t+1}$$

where $rx_{i,t+n}$ is the individual cryptocurrency return in week $t+n$. We consider an n of 1 to 12 weeks. We report t-statistics in parenthesis, where *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. The data are weekly from June 2017 and December 2021.

	Fundamental Sentiment Index betas β^{FSI}											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	n = 1	n = 2	n = 3	n =4	n = 5	n = 6	n = 7	n = 8	n = 9	n = 10	n = 11	n = 12
β_t^{FSI}	0.005**	0.004**	0.003*	0.003*	0.004**	0.003	0.004*	0.004***	0.004***	0.004*	0.003	0.001
	(2.00)	(2.16)	(1.80)	(1.92)	(2.54)	(1.61)	(2.32)	(2.61)	(2.53)	(1.97)	(1.54)	(0.76)
β_t^{MKT}	-0.006	-0.008	-0.003	-0.003	-0.000	-0.000	0.000	-0.002	-0.006	-0.011	-0.010	-0.006
	(-0.73)	(-0.92)	(-0.39)	(-0.36)	(-0.01)	(-0.02)	(0.06)	(-0.26)	(-0.86)	(-1.53)	(-1.27)	(-0.86)
$Size_t$	-0.002	-0.002	-0.002	-0.003*	-0.003*	-0.003*	-0.003*	-0.003*	-0.002	-0.002	-0.001	-0.001
	(-1.71)	(-1.76)	(-1.73)	(-2.53)	(-2.21)	(-2.14)	(-2.58)	(-2.42)	(-1.53)	(-1.44)	(-0.94)	(-0.66)
$Momentum_t$	0.006	-0.006	-0.012	-0.004	-0.014	0.002	0.011	0.028**	0.017	0.007	0.011	-0.007
	(0.51)	(-0.51)	(-1.03)	(-0.34)	(-1.28)	(0.19)	(1.19)	(2.95)	(1.59)	(0.72)	(1.17)	(-0.74)
$Liquidity_t$	0.263	0.458	-0.883	0.807	0.104	0.532	0.713	-0.788	0.914	0.218	0.154	0.567
	(1.23)	(0.56)	(-0.70)	(0.67)	(1.10)	(0.47)	(0.72)	(-0.64)	(0.69)	(1.37)	(1.84)	(0.75)
$Volatility_t$	-0.156	-0.172	-0.149	-0.253	-0.224	-0.323*	-0.399**	-0.383**	-0.252	-0.119	-0.0890	0.153
	(-1.13)	(-1.29)	(-0.96)	(-1.89)	(-1.49)	(-2.29)	(-2.77)	(-2.69)	(-1.57)	(-0.73)	(-0.56)	(0.97)
Constant	0.056**	0.056*	0.050*	0.075**	0.067**	0.065**	0.074**	0.070**	0.058*	0.053	0.038	0.022
	(2.01)	(1.83)	(1.75)	(2.51)	(2.22)	(2.03)	(2.54)	(2.41)	(1.76)	(1.62)	(1.23)	(0.72)
Observations	5,911	5,869	5,827	5,786	5,744	5,703	5,661	5,619	5,578	5,537	5,496	5,455
R^2	0.35	0.34	0.31	0.31	0.31	0.32	0.32	0.33	0.34	0.33	0.34	0.32

Table 8: Fundamental text-based factor: asset pricing tests

This table reports regressions results for the asset pricing tests. Test assets used are four size portfolios, four momentum portfolios, four liquidity portfolios, four volatility portfolios, four value portfolios, and four *FSI* portfolios. Portfolios are rebalanced weekly. Newey and West (1987) (NW) and Shanken (1992) (SH) t-statistics are reported in squared brackets, where *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. We also report R^2, Root Mean Squared Error (RMSE). The data are weekly from June 2017 and December 2021.

Panel A: Two-factor model								
	λ_{MKT}	λ_{Size}		RMSE	R^2			
FMB	-0.005	0.011***		0.005	0.15			
(NW)	[-0.52]	[2.52]						
(SH)	[-0.57]	[2.26]						
	λ_{MKT}	λ_{Size}	λ_{FSI}	RMSE	R^2			
FMB	-0.006	0.015***	0.012**	0.005	0.36			
(NW)	[-0.58]	[3.13]	[2.39]					
(SH)	[-0.65]	[3.05]	[2.25]					
Panel B: Three-factor model								
	λ_{MKT}	λ_{Size}	λ_{MOM}		RMSE	R^2		
FMB	-0.005	0.012***	-0.004		0.005	0.17		
(NW)	[-0.53]	[2.61]	[-0.71]					
(SH)	[-0.58]	[2.44]	[-0.59]					
	λ_{MKT}	λ_{Size}	λ_{MOM}	λ_{FSI}	RMSE	R^2		
FMB	-0.006	0.015***	0.000	0.014***	0.005	0.38		
(NW)	[-0.58]	[3.08]	[0.22]	[2.49]				
(SH)	[-0.66]	[3.03]	[0.02]	[2.42]				
Panel C: Five-factor model								
	λ_{MKT}	λ_{Size}	λ_{MOM}	$\lambda_{Liquidity}$	$\lambda_{Volatility}$		RMSE	R^2
FMB	-0.001	0.012***	-0.001	-0.008	0.004		0.005	0.25
(NW)	[-0.11]	[2.59]	[-0.38]	[-2.94]	[1.68]			
(SH)	[-0.14]	[2.41]	[-0.31]	[-2.73]	[1.29]			
	λ_{MKT}	λ_{Size}	λ_{MOM}	$\lambda_{Liquidity}$	$\lambda_{Volatility}$	λ_{FSI}	RMSE	R^2
FMB	-0.004	0.015***	0.000	-0.007	-0.001	0.013**	0.005	0.39
(NW)	[-0.39]	[3.00]	[0.06]	[-2.62]	[-0.28]	[2.41]		
(SH)	[-0.45]	[2.97]	[0.05]	[-2.32]	[-0.26]	[2.32]		

Table 9: Diversification Benefits of FSI Strategy

This table reports the benefits of adding HML_{FSI} strategy to conventional currency strategies. HML_{FSI} is the strategy that goes sell the lowest quartile portfolio sorted by FSI Index beta while buying the top quartile portfolio sorted by FSI Index beta. For each portfolio, we report annualized mean, standard deviation (Std) and Sharpe ratios (SR), all in percentage points. We also report skewness and kurtosis. We report the portfolio performance of individual trading strategies (Panel A), portfolio performance including FSI to each individual strategy and the equally weighted (EW) portfolio (Panel B). The bottom row of *Panel B* shows the weight of the HML_FSI portfolio. The data are weekly between June 2017 and December 2021.

Panel A: Excluding the FSI Strategy						
	MKT	$Size$	$Illiqudity$	$Volatility$	$Momentum$	EW
Mean	0.06	0.65	0.09	0.31	0.04	0.23
Std	0.82	0.49	0.21	0.26	0.64	0.25
Skewness	-0.64	0.40	0.38	0.58	-0.21	0.39
Kurtosis	5.30	4.56	5.43	4.87	5.49	4.23
SR	0.08	1.33	0.45	1.17	0.06	0.91

Panel B: Including the FSI Strategy						
	$MKT + HML_{FSI}$	$Size + HML_{FSI}$	$Illiquidity + HML_{FSI}$	$Volatility + HML_{FSI}$	$Momentum + HML_{FSI}$	$EW + HML_{FSI}$
Mean	0.37	0.66	0.38	0.49	0.35	0.31
Std	0.51	0.33	0.26	0.29	0.35	0.22
Skewness	0.12	0.96	0.88	1.14	0.76	0.85
Kurtosis	4.83	6.12	6.39	6.49	5.19	5.94
SR	0.71	2.00	1.47	1.70	1.02	1.41
$w_{HML_{FSI}}(w_F)$	0.50(0.50)	0.50(0.50)	0.50(0.50)	0.50(0.50)	0.50(0.50)	0.16(0.84)

Table 10: Portfolios sorted on other Topics

This table reports summary statistics for the excess returns of four portfolios sorted on exposure to Lending (Panel A), Regulation (Panel B), Payments (Panel C), Derivatives (Panel D), Social Media (Panel E), Hedging (Panel F), and Technical Trading (Panel G). Portfolio 1 (P_1) contains currencies with the lowest betas, and Portfolio 4 (P_4) contains currencies with the highest betas. HML represents the portfolio that has a short position in the high beta portfolio (P_4) and a long position in the low beta portfolio (P_1). For each portfolio, we report annualized mean and its t-statistics (reported in squared brackets), standard deviation (Std), and Sharpe ratios (SR). The data are weekly from June 2017 and December 2021.

Panel A: Lending Sentiment Portfolio

	P_1	P_2	P_3	P_4	HML
Mean	-0.14	0.34	-0.12	0.00	0.14
					[0.19]
Std	1.00	0.99	0.89	0.97	0.46
SR					0.31

Panel B: Regulation Sentiment Portfolio

	P_1	P_2	P_3	P_4	HML
Mean	-0.19	0.19	0.02	0.09	0.28
					[1.20]
Std	0.98	0.99	0.92	1.01	0.56
SR					0.49

Panel C: Payment Sentiment Portfolio

	P_1	P_2	P_3	P_4	HML
Mean	-0.09	-0.25	0.41	0.03	0.12
					[1.05]
Std	0.95	0.95	1.00	0.96	0.45
SR					0.28

Panel D: Derivatives Sentiment Portfolio

	P_1	P_2	P_3	P_4	HML
Mean	-0.09	-0.019	0.20	0.18	0.26
					[0.93]
Std	0.96	0.99	0.96	0.97	0.49
SR					0.54

Panel E: Social Media Sentiment Portfolio

	P_1	P_2	P_3	P_4	HML
Mean	-0.22	0.16	-0.02	0.20	0.42
					[1.31]
Std	0.97	0.98	0.92	1.00	0.52
SR					0.80

Panel F: Hedging Sentiment Portfolio

	P_1	P_2	P_3	P_4	HML
Mean	-0.02	-0.13	0.24	-0.05	0.01
					[0.05]
Std	1.00	0.93	0.98	0.96	0.51
SR					0.03

Panel G: Technical Sentiment Portfolio

	P_1	P_2	P_3	P_4	LMH
Mean	-0.01	-0.38	-0.28	-0.72	0.71
					[2.75]
Std	1.03	0.94	0.95	0.99	0.51
SR					1.30

Internet Appendix to

"Fundamental Sentiment and Cryptocurrency Risk Premia"

(Not for publication)

A Model derivation

The model is an overlapping generations framework and is a simplified version of Biais et al. (2023). The young generation consume c_t^y, subject to a budget constraint that includes their endowment e_t, net of savings s_t, and their holdings of money. The two types of money they can hold are fiat currency at price $\hat{p}_t$ with quantity $\hat{q}_t$, and holdings of cryptocurrency p_t with quantity q_t. In addition, users have to pay a transaction cost ψ_t per unit of cryptocurrency. This can be due to costs of transacting on exchanges, and the fees required to validate transactions by miners.

In the next period, they consume their savings which earn the risk-free rate r_t, and their money balances, which are now evaluated at prices p_{t+1} and $\hat{p}_{t+1}$. Finally, users can also obtain transaction benefits θ_{t+1} per unit of cryptocurrency transactions. These benefits can be accrued due to the ease of conducting cross-border payments, and the additional programmability features such as smart contracts that cryptocurrencies can provide.

Formally, we maximize utility in Equation (12) subject to the budget constraints in Equations (13) and (14).

$$\max_{q_t, s_t, \hat{q}_t} \qquad u(c_t^y) + \beta \mathbb{E}_t[u(c_{t+1}^o)] \tag{12}$$

subject to:

$$c_t^y = e_t - s_t - q_t p_t - \hat{q}_t \hat{p}_t - \psi_t q_t p_t, \tag{13}$$

$$c_{t+1}^o = s_t(1 + r_t) + q_t p_{t+1} + \hat{q}_t \hat{p}_{t+1} + \theta_{t+1} q_t p_{t+1}. \tag{14}$$

First order conditions:

$$-p_t - \psi_t p_t u'(c_t^y) + \beta \mathbb{E}_t u'(c_{t+1}^o)\,(p_{t+1} + \theta_{t+1} p_{t+1}) = 0 \tag{15}$$

$$-u'(c_t^y) + (1 + r_t)\beta \mathbb{E}_t u'(c_{t+1}^o) = 0 \tag{16}$$

$$-\hat{p}_t u'(c_t) + \beta \mathbb{E}_t u'(c_{t+1})\hat{p}_{t+1} = 0 \tag{17}$$

Solving the first order conditions yields a Euler equation for the cryptocurrency price p_t, the fiat currency price $\hat{p}_t$, and the discount factor β.

$$p_t = \beta \mathbb{E}_t \left[\frac{u'(c^o_{t+1})}{u'(c^y_t)} \frac{1+\theta_{t+1}}{1+\psi_t} p_{t+1} \right] \tag{18}$$

$$\beta = \frac{1}{1+r_t} \frac{u'(c^y_t)}{\mathbb{E}_t u'(c^o_{t+1})} \tag{19}$$

$$\hat{p}_t = \beta \mathbb{E}_t \left[\frac{u'(c^o_{t+1})}{u'(c^y_t)} \hat{p}_{t+1} \right] \tag{20}$$

Substituting the formula for β in the Euler equation for the cryptocurrency price p_t yields the equation in Section 2.

B Examples of Fundamental Analysis Articles

B.1 Sample of Fundamental Articles

Some articles identified as Fundamental articles are listed

Fundamental Article 1

"Cryptocurrencies have been a winning bet this year, but the chip makers who play a key role in the market are still playing their hands very cautiously. The exploding value of cryptocurrencies this year has created a strong incentive for "miners" who use high-end computers that match and update cryptocurrency transactions in return for rewards. Mining for many of the fastest-rising currencies, including ethereum, is powered by graphics processors from companies like Nvidia and Advanced Micro Devices. These chips, also called GPUs, are the same type used in high-end gaming PCs. Cryptocurrency mining seems to have created a decent market for both companies. Nvidia credits about $220 million in revenue over its last two quarters to cryptocurrency demand, which is a little less than 5% of the company's total sales. AMD CEO Lisa Su estimates the market will account for a mid-single digit percentage of the company's projected 23% growth this year, which suggests revenue around $50 million for the year. But neither company wants to bake cryptocurrency into their outlooks, and with good reason. Cryptocurrencies are highly volatile. Changes to the underlying technology can sharply affect the economic value of mining. Joseph Moore of Morgan Stanley says an expected shift by ethereum in the next year or so will render GPU-based mining for the currency "obsolete." Still, there were 26 cryptocurrencies with total market values over $1 billion as of Thursday. Only bitcoin and ethereum were in that range a year ago. Mitch Steves of RBC Capital notes that several of those rising fast are mined with GPUs. Cryptocurrencies may be unpredictable, but they are likely here to stay. Which is ultimately good news for those with chips in the game. Write to Dan Gallagher at dan.gallagher@wsj.com (END) Dow Jones Newswires"

Fundamental Article 2

"The Bitcoin (BTC) hash rate reached a new all-time high today, according to data from monitoring resource Blockchain.com on July 7. The previous record was broken in the second half of June, when bitcoin hash rate reached 65.19 TH/s and growth has steadily continued since then. Hash rate is the number of calculations that a given hardware or network can perform every second. It is a very important parameter for miners, as a higher hash rate will increase their chances of solving the mathematical problem, sealing off the block and collecting their reward. A higher network hash rate also increases the amount of resources needed for performing a 51% attack, making the network safer."

Fundamental Article 3

"Aave, the DeFi platform, has announced that it will be implementing Polygon to offer more scalability and lower fees amid increasing congestion on the Ethereum Network. The platform was originally launched on Ethereum L1 and quickly became one of the most important Decentralized Finance (DeFi) projects during the DeFi Summer of 2020, a period in which DeFi took the cryptocurrency ecosystem by storm in what would become one of the biggest bull runs seen by the cryptocurrency market. However, despite Ethereum occupying the spot as the leading blockchain network at this time, the network has seen its block space supply grow increasingly scarce and limited, which has resulted in increased congestion and gas prices, which have affected the projects it initially helped succeed. Aave Sees Polygon as a Solution Now, Aave integration with Polygon will allow users to enjoy more scalability, faster transactions, and lower gas prices that will boost the platform to new levels as the cryptocurrency market continues to grow. The move is the "first wave in Aave Protocol. "New Frontiers exploration mission, which is aimed to allow it to build synergies with other projects and expand to a multi-market approach to secure the future growth of the protocol. Using Sidechains with Polygon This first wave will see the implementation of a scalable sidechain on Ethereum by using Polygon, increasing throughput and reducing fees, as well as allowing the collaboration with other DeFi protocols and projects by facilitating

communication. Polygon partnership with Chainlink will also allow the Avee protocol to provide better quality on price feeds by taking advantage of one of the best Oracle Networks in the current cryptocurrency ecosystem, improving the protocol's current standards. Aave users will also have access to MATIC, Polygon cryptocurrency, being able to use it as collateral in addition to other assets such as USDC, USDT, DAI, WETH, AAVE, and WBTC. Many Fresh Features This will be possible once the Smart Contract Bridge is deployed, with users who make use of it receiving part of transaction fees used in MATIC to cover part of their transaction fees on the Polygon blockchain. The bridge can also be used to transfer assets from Ethereum to Polygon, which will prove useful for users wanting to migrate their assets. The recent rise in popularity experienced by Polygon has also made the process of transferring assets to Polygon easier than ever before, with popular wallets like Metamask deploying one-click solutions. Transforming Ethereum Into a Multichain System Matic rebranded to Polygon earlier this year as it aimed to become a solution to Ethereum growing congestion problem by transforming it into a multi-chain network and offering integration with other Layer-2 solutions. With the rebranding, Polygon said it would extend the scope of the Matic Platform by allowing Ethereum to integrate scalation solutions like zkRollups, Optimistic Rollups, and Validium, as well as interchain communication protocols to become the internet of blockchain. Polygon, originally launched in 2019, has become increasingly relevant in the cryptocurrency ecosystem as the congestion on the Ethereum network increased. However, it would not be until early 2021 when the project would become one of the top 100 projects in the cryptocurrency market by market capitalization. The announcement of the integration saw MATIC's value increase by over 10% in a matter of minutes, a similar trend to the one experienced by AAVE. Polygon also saw DeFi platform Zapper announced that it will be integrating the network, which is expected to be the first of many sidechains as xDAI, Optimism, and Binance Chain will also be covered in the future. These moves show an increasing interest from cryptocurrency projects to find alternatives to the Ethereum network at a time when its future is still uncertain as competition in the blockchain industry continues to increase. The post Aave Will Integrate With Polygon Sidechains for

Much Lower Fees appeared first on Blockonomi."

B.2 Sample of Fundamental Sentences and Their Sentiment Score

Some sentences in Fundamental articles with their sentiment score are listed

Fundamental Sentence 1

"the suspension appears to have plunged the bitcoin mining power as much as 30%." (Sentiment Score -0.2)

Fundamental Sentence 2

"dr. sivakumar arumugam concluded,"the striking divergence between the global hash rate and bitcoin prices suggests that mining is becoming increasingly unprofitable, the review of publicly available data reveals that the global hash rate has been increasing at a steady exponential rate in recent months." (Sentiment Score -0.04)

Fundamental Sentence 3

"coinhive reportedly had to shut down its services amidst a 50 percent decline in hash rate following the last monero hard fork." (Sentiment Score -0.19)

Fundamental Sentence 4

"ethereum gas fees have exploded in 2021, which has been a hindrance to both inexpensive nfts, and also defi platforms that were designed to deal with small amounts of value." (Sentiment Score -0.09)

Fundamental Sentence 5

the scaling woes of ethereum are well-documented and came to a head when transaction costs soared in gas fees, and many dapps became prohibitively cumbersome to use and remain so today." (Sentiment Score -0.08)

C Summary Statistics and Variable Descriptions

Table A1: Summary Statistics of Full Sample

This table reports summary statistics of our cryptocurrency data per year. We present the number of cryptocurrencies, the total market capitalization at the end of the year (in Billion $), the ratio of the total market capitalization of our sample to the total market capitalization of the cryptocurrency market, the average volatility, and the average number of accounts. Our sample contains weekly data from June 2017 to December 2021.

Year	Number of coins	Total Market capitalization ($B)	Sample/Total Market capitalization ratio	Volatility	Number of accounts
2017	20	661	0.87	0.91	73,957.8
2018	25	145	0.78	0.91	64,978.62
2019	30	195	0.83	0.91	58,252.67
2020	40	654	0.91	0.91	58,126.36
2021	43	1,750	0.82	0.91	71,446.09

Table A2: Variable Descriptions

This table provides descriptions of variables used in the paper.

Variable	Description
MKT	Value-weighted returns of cryptocurrencies in the sample, based on the market capitalization ratio.
Size	The difference between average returns of cryptocurrencies in the low (Small) and high (Big) portfolios, based on market capitalization.
Momentum	The difference between average returns of cryptocurrencies in the high (Winner) and low (Loser) portfolios, based on the 6-week cumulative return.
Liquidity	The difference between average returns of cryptocurrencies in the high (Liquid) and low (Illiquid) portfolios, based on the Amihud ratio.
Volatility	The difference between average returns of cryptocurrencies in the high (High volatility) and low (Low volatility) portfolios, based on idiosyncratic volatility.
Value (T/M ratio)	The difference between average returns of cryptocurrencies in the high and low portfolios, based on the transaction-to-market ratio.
Value (U/M ratio)	The difference between average returns of cryptocurrencies in the high and low portfolios, based on the user-to-market ratio.
Value (A/M ratio)	The difference between average returns of cryptocurrencies in the high and low portfolios, based on the address-to-market ratio.
Network 1 (BA growth)	The difference between average returns of cryptocurrencies in the high and low portfolios, based on the growth rate of total addresses with balance.
Network 2 (TA growth)	The difference between average returns of cryptocurrencies in the high and low portfolios, based on the growth rate of total addresses.
Network 3 (Volgrowth)	The difference between average returns of cryptocurrencies in the high and low portfolios, based on the growth rate of transaction volume.
Network 4 (VolUSDgrowth)	The difference between average returns of cryptocurrencies in the high and low portfolios, based on the growth rate of transaction volume in USD.

Table A3: Token Classification

Ticker	Token Type	Description
yfi	Governance	Yearn Finance is a decentralized finance (DeFi) platform that enables yield aggregation and lending.
uma	Governance	UMA is a decentralized financial contracts platform that allows users to create synthetic assets.
perp	Governance	Perpetual Protocol provides decentralized perpetual contracts for crypto assets, allowing traders to go long or short.
bal	Governance	Balancer is an automated portfolio manager and liquidity provider, functioning as a decentralized exchange.
crv	Governance	Curve is a decentralized exchange optimized for stablecoin swaps with low slippage and minimal fees.
uni	Governance	Uniswap is a leading decentralized trading protocol and liquidity provider on the Ethereum blockchain.
aave	Governance	Aave is a decentralized non-custodial liquidity protocol for earning interest on deposits and borrowing assets.
neo	Governance	NEO is a blockchain platform designed for building digital assets and smart contracts with a focus on a decentralized economy.
icp	Governance	Internet Computer is a blockchain platform aimed at extending the internet's functionality by hosting smart contracts and decentralized applications.
gno	Governance	Gnosis is a prediction market platform built on Ethereum, enabling users to speculate on the outcome of future events.
hed	Governance	Hedera Hashgraph is a public distributed ledger technology that offers fast, fair, and secure applications.

Table A3: Token Classification (continued)

Ticker	Token Type	Description
gdot	Platform	Polkadot is a multi-chain blockchain platform enabling interoperability between blockchains for asset transfers and smart contracts.
qnt	Platform	Quant Network facilitates seamless interoperability between multiple blockchains and networks.
cro	Product	Crypto.com provides cryptocurrency-related services such as payments, trading, and financial products.
comp	Governance	Compound is a decentralized lending and borrowing protocol where users can earn interest or borrow assets.
mana	Product	Decentraland is a virtual reality platform powered by Ethereum, allowing users to create, experience, and monetize content and applications.
eth	Platform	Ethereum is a decentralized platform for building decentralized applications and smart contracts using blockchain technology.
omg	Platform	OMG Network is a layer-2 scaling solution for Ethereum, enabling fast and secure payments and asset transfers.
sushi	Governance	SushiSwap is a decentralized exchange and automated market maker protocol for trading crypto assets.
mkr	Governance	MakerDAO is the organization behind the DAI stablecoin, allowing users to generate DAI by locking collateral.
bch	General Payment	Bitcoin Cash is a peer-to-peer electronic cash system that enables fast, low-cost payments.
bat	Product	Basic Attention Token (BAT) is used to reward users and content creators within the Brave browser ecosystem.

Table A3: Token Classification (continued)

Ticker	Token Type	Description
dash	Platform	Dash is a cryptocurrency focused on fast, low-cost digital payments and privacy-enhanced transactions.
etc	Platform	Ethereum Classic is a continuation of the original Ethereum blockchain, maintaining the unaltered history of the chain.
btc	General Payment	Bitcoin is a decentralized digital currency that allows peer-to-peer transactions on a trustless, decentralized network.
xmr	General Payment	Monero focuses on privacy and anonymity in transactions, making all transactions confidential and untraceable.
zec	Platform	Zcash is a privacy-focused cryptocurrency offering users the option of sending public or shielded transactions.
ltc	General Payment	Litecoin is a peer-to-peer cryptocurrency that offers faster transaction confirmation times than Bitcoin.
xrp	Platform	XRP is the native cryptocurrency of the Ripple network, optimized for cross-border payments.
bsv	General Payment	Bitcoin SV (Satoshi's Vision) aims to restore Bitcoin's original protocol and vision, focusing on large block sizes for scalability.
dcr	General Payment	Decred is a cryptocurrency with a strong emphasis on decentralized governance and community involvement.
xem	Platform	NEM (New Economy Movement) is a blockchain platform for managing assets and data with built-in smart contract functionality.
ftt	Platform	FTX Token (FTT) is the native utility token of the FTX cryptocurrency exchange, offering fee discounts and other benefits.

Table A3: Token Classification (continued)

Ticker	Token Type	Description
xtz	Platform	Tezos is a self-amending blockchain platform focused on security, upgradability, and smart contract functionality.
lpt	Platform	Livepeer is a decentralized video streaming network built on the Ethereum blockchain, allowing users to contribute computing resources.
link	Platform	Chainlink provides decentralized oracle services, connecting smart contracts with real-world data.
ht	Platform	Huobi Token (HT) is the native utility token of the Huobi cryptocurrency exchange, offering fee discounts and other benefits.
doge	General Payment	Dogecoin, originally created as a joke, has gained widespread popularity and is used for microtransactions.
algo	Platform	Algorand is a blockchain platform focused on speed, security, and scalability, using a pure proof-of-stake consensus mechanism.
xlm	Platform	Stellar is a blockchain platform designed for cross-border payments and remittances, focusing on financial inclusion.
ada	Platform	Cardano is a third-generation blockchain platform focused on scalability, sustainability, and interoperability with its native cryptocurrency, ADA.
snx	General Payment	Synthetix is a protocol for creating and trading synthetic assets on the Ethereum blockchain.
inch	Platform	1inch is a decentralized exchange aggregator and liquidity provider that helps users find the best prices across decentralized exchanges.

D Robustness Tests on FSI Portfolio

D.1 Portfolio Turnover

Figure A1 presents the portfolio turnover of the FSI strategy. Specifically, it reports the frequency with which each cryptocurrency appears in the low and high fundamental sentiment portfolios. Panel A shows the turnover for the low FSI portfolio, while Panel B shows the high FSI portfolio.

We find that the FSI strategy is predominantly driven by certain cryptocurrencies. In the low FSI portfolio, BAT, GNO, and NEO are prominent, while ADA, DOGE, LINK, and XLM are key drivers in the high FSI portfolio. For example, GNO appears in the low beta portfolio in almost 70% of the weeks in which the portfolio is rebalanced, while NEO appears in 60% of the holding periods. Similarly, XLM is present in nearly 70% of the weeks in the high FSI portfolio, and DOGE, ADA, and LINK typically appear in 60% of the sample weeks in the high beta portfolio.

D.2 Top 15 Cryptocurrencies

To ensure that smaller cryptocurrencies do not disproportionately influence our results, we replicate the FSI strategy using only the top 15 cryptocurrencies, ranked by their average market capitalization over the sample period.[18] The portfolio sorting results are reported in Table A4.[19]

The results show that, when sorting by β^{FSI}, the average returns increase in a monotonic fashion from portfolio 1 to portfolio 3. The HML_{FSI} portfolio achieves an annualized excess return of 77%, with a Newey and West (1987) t-statistic of 2.06, and a Sharpe ratio of 1.04 per

[18]The sample includes Bitcoin, Bitcoin Cash, Cronos, Stellar, Dogecoin, Chainlink, Ethereum, Cardano, Ripple, Polkadot, Litecoin, Uniswap, Internet Computer, Algorand, and FTX Token.

[19]Due to the smaller number of cryptocurrencies available at the beginning of the sample, we limit the number of portfolios to three.

annum. This provides evidence that our results are robust even when focusing only on the largest cryptocurrencies.

D.3 Alternative Specifications for Estimating β^{FSI}

We also estimate β^{FSI} using alternative specifications. Equation (5) includes five factors: market, size, momentum, volatility, and liquidity, in addition to the fundamental sentiment factor. We explore two alternative specifications:

In the first, we control only for the market (MKT) factor:

$$rx_{i,t} = \alpha_{i,t} + \beta_{i,t}^{SI} SI + \beta_{i,t}^{MKT} MKT_t + \epsilon_{i,t} \tag{21}$$

In the second specification, we control for the market (MKT), size (SMB), and momentum (MOM) factors:

$$rx_{i,t} = \alpha_{i,t} + \beta_{i,t}^{FSI} FSI + \beta_{i,t}^{MKT} MKT_t + \beta_{i,t}^{SMB} SMB_t + \beta_{i,t}^{MOM} MOM_t + \epsilon_{i,t} \tag{22}$$

We then construct long-short portfolios based on past β^{FSI} estimates. Table A7 presents summary statistics for these portfolios. Constructing long-short portfolios using alternative specifications for β^{FSI} consistently generates positive and statistically significant returns, confirming that our fundamental sentiment factor is robust to different factor models.

D.4 Alternative Sentiment Proxies

We construct alternative sentiment measures for fundamental trading factors, as outlined in Equation (4). Two alternative proxies are considered: one is based on the proportion of negative words over the total number of words (Equation (23)), while the other measures net negative sentiment as the difference between positive and negative words, normalized by the total number of sentiment words (Equation (24)).

$$Sent = 1 - \frac{\text{Number of negative words}}{\text{Total number of words}} \tag{23}$$

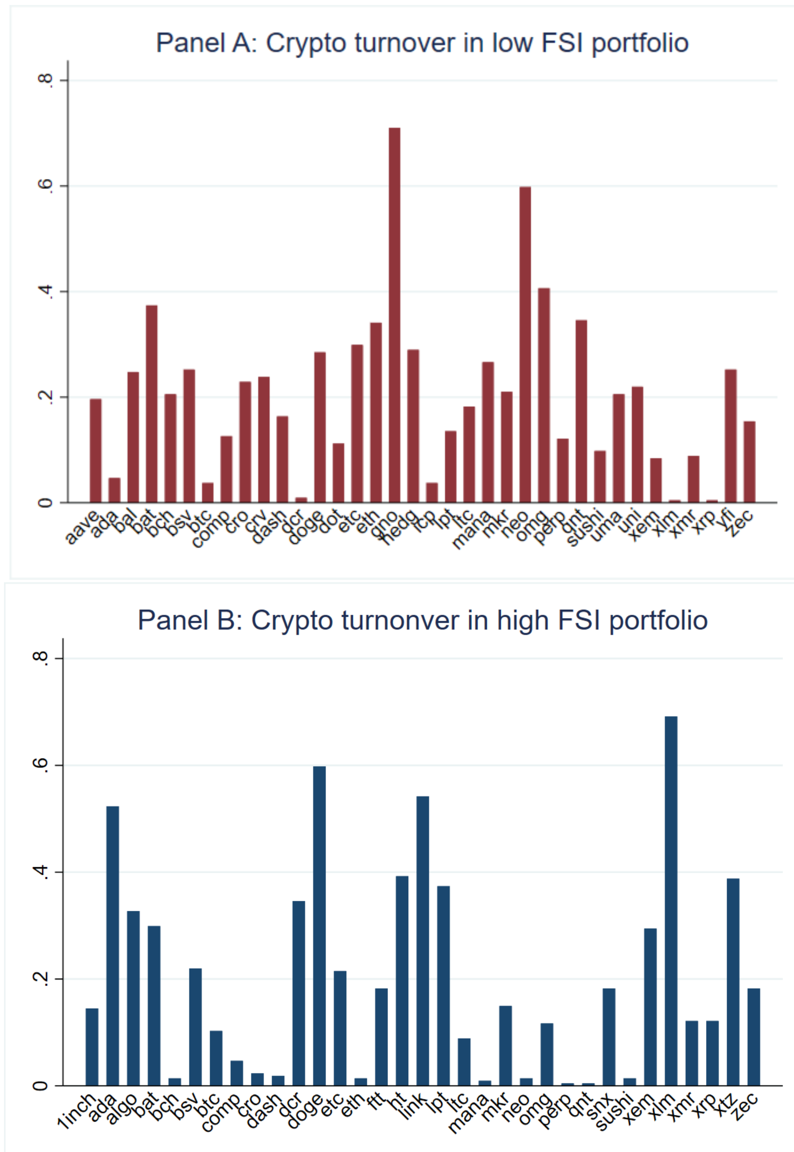
$$Sent = \frac{\text{Number of positive words} - \text{Number of negative words}}{\text{Number of positive words} + \text{Number of negative words}} \tag{24}$$

Table A5 reports the results for these alternative sentiment proxies. Panel A presents results for the first measure (negative sentiment), while Panel B shows results for the second measure (net sentiment). Both alternative measures of sentiment generate robust factors in predicting cryptocurrency returns. Long-short portfolios based on high and low sentiment cryptocurrencies yield annualized returns of 62% and 60%, with Sharpe ratios of 1.21 and 1.16, respectively. Thus, our results are robust to alternative sentiment specifications.

D.5 Different Number of Portfolios

Our main analysis uses quartile portfolios. In Table A6, we show that the choice of the number of portfolios does not affect our results. Panel A reports the results for tercile portfolios, while Panel B reports quintile portfolios. The annual returns for the fundamental sentiment strategy are 55% and 62%, respectively.

Figure A1: FSI Portfolio Turnover



The figure shows cryptocurrency turnover for low beta FSI portfolios (Panel A) and high beta FSI portfolios (Panel B). The data covers the period between June 2017 and December 2021.

Table A4: Portfolios sorted on Fundamental Sentiment Index (Top 15 cryptocurrencies by market capitalization)

This table reports summary statistics for the excess returns of three currency portfolios sorted on exposure to the Fundamental Sentiment Index FSI for the top 15 cryptocurrencies by market capitalization. Portfolio 1 (P_1) contains currencies with the lowest Fundamental Sentiment Index betas, and Portfolio 3 (P_3) contains currencies with the highest Fundamental Sentiment Index betas. HML represents the portfolios that have a long position in the high beta portfolio (P_3) and a short position in the low beta portfolio (P_1). For each portfolio, we report annualized mean and its t-statistics (reported in squared brackets), standard deviation (Std), and Sharpe ratios (SR). The data are weekly from June 2017 to December 2021.

Fundamental Sentiment Index Portfolio				
	P_1	P_2	P_3	HML_{FSI}
Mean	-0.08	0.48	0.69	0.77
				[2.06]
Std	0.88	1.14	1.16	0.74
Skewness	-0.74	0.97	0.29	1.38
Kurtosis	5.53	10.24	4.94	7.47
β	-0.59	0.24	1.56	2.15
SR				1.04

Table A5: Portfolios sorted on Fundamental Sentiment Index - Alternative proxy for sentiment

This table reports summary statistics for the excess returns of 4 cryptocurrencies portfolios sorted on exposure to the Fundamental Sentiment Index β^{FSI} based on the following 2 specifications to estimate sentiment:

$$Sent = 1 - \frac{\text{Number of negative words}}{\text{Total number of words}}$$

$$Sent = \frac{\text{Number of positive words-Number of negative words}}{\text{Number of positive words+Number of negative words}}$$

Portfolio 1 (P_1) contains cryptocurrencies with the lowest or β^{FSI}, and Portfolio 4 (P_4) contains cryptocurrencies with the highest β^{FSI}. HML represents the portfolio that has a long position in the high beta portfolio (P_4) and a short position in the low beta portfolio (P_1). For each portfolio, we report annualized mean and its t-statistics (reported in squared brackets), standard deviation (Std), skewness, kurtosis, and Sharpe ratios (SR). The data are weekly from June 2017 and December 2021.

Panel A: Alternative proxy 1					
	P_1	P_2	P_3	P_4	HML_{FSI}
Mean	-0.23	-0.11	0.34	0.39	0.62
					[2.41]
Skewness	-0.91	-0.22	-0.38	-0.38	0.66
Kurtorsis	5.43	4.97	5.17	5.30	5.59
Std	0.96	0.92	0.98	1.03	0.51
β	-0.16	-0.03	0.05	0.18	0.34
SR					1.21

Panel B: Alternative proxy 2					
	P_1	P_2	P_3	P_4	HML_{FSI}
Mean	-0.14	-0.08	0.15	0.46	0.60
					[2.41]
Skewness	-0.81	-0.31	-0.57	-0.42	0.43
Kurtorsis	5.41	4.74	5.55	4.99	5.59
Std	0.96	0.93	0.97	1.01	0.51
β	-0.17	-0.02	0.05	0.19	0.36
SR					1.16

Table A6: Portfolios sorted on Fundamental Sentiment Index - Terciles and Quintiles

This table reports summary statistics for the excess returns of 3 cryptocurrencies portfolios (Panel A) or 5 cryptocurrencies portfolios (Panel B) sorted on exposure to the Fundamental Sentiment Index β^{FSI}. Portfolio 1 (P_1) contains cryptocurrencies with the lowest β^{FSI}, and Portfolio 3 (P_3) in Panel A (or Portfolio 5 (P_5) in Panel B) contains cryptocurrencies with the highest β^{FSI}. HML represents the portfolio that has a long position in the high beta portfolio and a short position in the low beta portfolio. For each portfolio, we report annualized mean and its t-statistics (reported in squared brackets), standard deviation (Std), skewness, kurtosis, and Sharpe ratios (SR). The data are weekly from June 2017 and December 2021.

Panel A: Terciles

	P_1	P_2	P_3	HML_{FSI}
Mean	-0.61	-0.32	-0.06	0.55
				[2.51]
Skewness	-0.79	-0.42	-0.36	0.63
Kurtorsis	4.92	5.17	5.01	5.51
Std	0.95	0.97	0.98	0.45
SR				1.22

Panel B: Quintiles

	P_1	P_2	P_3	P_4	P_5	HML_{FSI}
Mean	-0.56	-0.63	-0.29	-0.20	0.07	0.62
						[2.03]
Skewness	-0.66	-0.74	-0.26	-0.44	-0.20	0.72
Kurtorsis	4.53	5.17	4.77	4.93		6.15
Std	0.96	0.98	1.02	0.93	1.07	0.62
SR						1.00

Table A7: Portfolios sorted on Fundamental Sentiment Index - Alternative specification to estimate β

This table reports summary statistics for the excess returns of 4 cryptocurrencies portfolios sorted on exposure to the Fundamental Sentiment Index β^{FSI} (Panel B) based on the following specification:

$$rx_{i,t} = \alpha_{i,t} + \beta^{FSI}_{i,t} FSI + \beta^{MKT}_{i,t} MKT_t + \epsilon_{i,t}$$

$$rx_{i,t} = \alpha_{i,t} + \beta^{FSI}_{i,t} FSI + \beta^{MKT}_{i,t} MKT_t + \beta^{SMB}_{i,t} SMB_t + \beta^{MOM}_{i,t} MOM_t + \epsilon_{i,t}$$

Portfolio 1 (P_1) contains cryptocurrencies with the lowest β^{FSI}), and Portfolio 4 (P_4) contains cryptocurrencies with the highest β^{FSI}. HML represents the portfolio that has a long position in the high beta portfolio and a short position in the low beta portfolio. For each portfolio, we report annualized mean and its t-statistics (reported in squared brackets), standard deviation (Std), skewness, kurtosis, and Sharpe ratios (SR). The data are weekly from June 2017 and December 2021.

Panel A: Alternative specification 1					
	P_1	P_2	P_3	P_4	HML_{FSI}
Mean	-0.07	0.18	0.19	0.47	0.54
					[1.98]
Skewness	-0.53	-0.62	-0.52	-0.22	0.22
Kurtorsis	4.90	5.82	5.44	5.32	4.32
Std	1.00	0.94	0.94	1.02	0.58
β	-1.31	-0.23	0.32	1.56	2.87
SR					0.93

Panel B: Alternative specification 2					
	P_1	P_2	P_3	P_4	HML_{FSI}
Mean	-0.16	0.13	0.35	0.45	0.61
					[2.16]
Skewness	-0.52	-0.65	-0.25	-0.25	0.43
Kurtorsis	5.12	5.59	5.22	5.22	6.06
Std	0.97	0.95	0.96	1.03	0.60
β	-0.17	-0.02	0.05	0.19	0.36
SR					1.02

E Alternative Topics

The figure shows keywords for various cryptocurrency-related topics generated from BERT topic modeling. These topics include two related to Derivatives (Panel A), two related to Social Media (Panel B), four related to Regulation (Panel C), and several related to Lending, Payment, Hedging, and Technical Trading (Panel D). The data span from June 2017 to December 2021.

(a) Derivatives Topic 1

(b) Derivatives Topic 2

(c) Social Media Topic 1

(d) Social Media Topic 2

(e) Regulation Topic 1

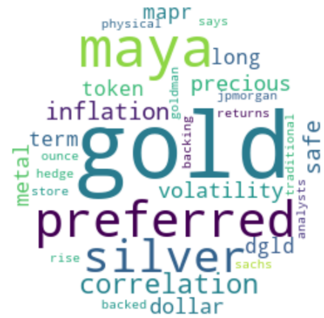
(f) Regulation Topic 2

(g) Regulation Topic 3

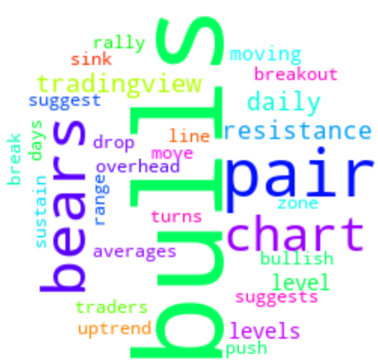
(h) Regulation Topic 4

(i) Lending Topic

(j) Payment Topic



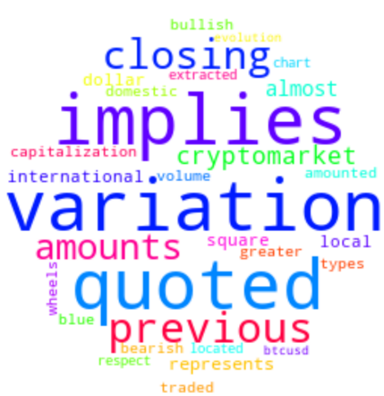
(k) Hedging Topic



(l) Technical Topic 1

(m) Technical Topic 2

(n) Technical Topic 3



(o) Technical Topic 4

(p) Technical Topic 5

(q) Technical Topic 6

Table A8: Cross-Sectional regressions: Technical Sentiment

This table reports Fama Macbeth cross-sectional regressions for Technical Sentiment Index betas β^{TSI}. We run the model below:

$$rx_{i,t+1} = \lambda_{0,t} + \lambda_{1,t}\hat{\beta}^{TSI}_{i,t} + \lambda_{2,t}X_{i,t} + \epsilon_{i,t+1}$$

where $rx_{i,t+1}$ is the individual cryptocurrency return . We report t-statistics are reported in squared brackets, where *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. The data are weekly from June 2017 and December 2021.

	Technical Sentiment Index betas β^{TSI}					
	(1)	(2)	(3)	(4)	(5)	(6)
β^{TSI}_t	-0.005***	-0.005***	-0.005***	-0.006***	-0.006***	-0.006***
	(3.43)	(3.52)	(3.47)	(3.75)	(3.61)	(3.55)
β^{MKT}_t		-0.004	-0.004	-0.003	-0.008	-0.006
		(-0.57)	(-0.52)	(-0.41)	(-0.97)	(-0.82)
$Size_t$			-0.001	-0.001	-0.002*	-0.001
			(-1.41)	(-1.27)	(-1.66)	(-1.20)
$Momentum_t$				0.006	0.006	0.011
				(0.69)	(0.67)	(1.02)
$Liquidity_t$					0.210	0.241
					(1.05)	(1.22)
$Volatility_t$						-0.150
						(-0.95)
Constant	0.002	0.006	0.035	0.030	0.044*	0.045
	(0.20)	(0.53)	(1.42)	(1.25)	(1.73)	(1.40)
Observations	6,138	6,138	6,138	6,138	6,138	6,138
R^2	0.06	0.11	0.15	0.23	0.29	0.34

Table A9: Cross-Sectional regressions: Fundamental and Technical Sentiment

This table reports Fama Macbeth cross-sectional regressions for Technical Sentiment Index betas β^{TSI} and Fundamental Sentiment Index betas β^{FSI}. We run the model below:

$$rx_{i,t+1} = \lambda_{0,t} + \lambda_{1,t}\hat{\beta}^{FSI}_{i,t} + \lambda_{2,t}\hat{\beta}^{TSI}_{i,t} + \lambda_{3,t}X_{i,t} + \epsilon_{i,t+1}$$

where $rx_{i,t+1}$ is the individual cryptocurrency return . We report t-statistics are reported in squared brackets, where *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. The data are weekly from June 2017 and December 2021.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
β^{FSI}_t	0.004***	0.004**	0.004**	0.004**	0.004**	0.005**	0.005**
	(2.71)	(2.35)	(2.30)	(2.46)	(2.29)	(2.12)	(2.01)
β^{TSI}_t		-0.005***	-0.005***	-0.005***	-0.006***	-0.005***	-0.005***
		(-3.22)	(-3.48)	(-3.39)	(-3.44)	(-2.94)	(-3.00)
β^{MKT}_t			-0.005	-0.005	-0.005	-0.008	-0.007
			(-0.71)	(-0.65)	(-0.68)	(-0.96)	(-0.89)
$Size_t$				-0.001	-0.001	-0.002*	-0.002**
				(-1.54)	(-1.43)	(-1.86)	(-2.43)
$Momentum_t$					-0.003	0.001	0.002
					(-0.76)	(0.14)	(0.33)
$Liquidity_t$						0.213	0.283
						(0.91)	(1.18)
$Volatility_t$							-0.171
							(-1.36)
Constant	-0.007	-0.007	-0.002	0.031	0.027	0.030	0.057**
	(-0.69)	(-0.68)	(-0.14)	(1.23)	(1.16)	(1.22)	(2.00)
Observations	6,138	6,138	6,138	6,138	6,138	5,911	5,911
R^2	0.05	0.11	0.16	0.20	0.26	0.33	0.38