Fundamental Sentiment and Cryptocurrency Risk Premia

Ilias Filippou My T. Nguyen Ganesh Viswanath-Natraj

This version: April 4, 2024

Acknowledgements: 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. Filippou: Olin Business School, Washington University in St. Louis, USA. Email: iliasfilippou@wustl.edu; Nguyen: Olin Business School, Washington University in St. Louis, USA. Email: myn@wustl.edu; Viswanath-Natraj: The University of Warwick, Warwick Business School, UK. Email: Ganesh.Viswanath-Natraj@wbs.ac.uk.

Fundamental Sentiment and Cryptocurrency Risk Premia

Abstract

This paper investigates the cross-sectional predictive ability of text-based factors in the cryptocurrency market, an important asset class for retail and institutional investors. We build a text-based factor related to cryptocurrency fundamentals and find that the exposure to sentiment on fundamentals is priced. High beta currencies are more sensitive to fundamental news and earn a risk premium, and are typically payment and platform tokens. Currency betas correlate with measures of value on the blockchain, such as the ratio of the network of users and transactions to market cap. Our results demonstrate the importance of considering text-based factors when analyzing cryptocurrency returns.

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

JEL Classification: G11, G12, G14, G32.

1 Introduction

This paper examines the cross-sectional predictive ability of text-based fundamental trading measures in the cryptocurrency market. The market capitalization of this market exceeded 3 trillion USD in November 2021, with a total trading volume for spot and futures contracts of 8.8 trillion USD in the first quarter of 2020 (e.g., Helms, 2020). Cryptocurrencies are an important asset class for investors (e.g., Harvey, Abou Zeid, Draaisma, Luk, Neville, Rzym, and Van Hemert, 2022). Retail and institutional traders populate this market, which has experienced large price movements over the previous years. The lack of adequate market regulation and information quality has led many investors to link cryptocurrencies with market manipulation and fraud (e.g., Gandal, Hamrick, Moore, and Oberman, 2018). Other investors view this market as an important innovation, and its underlying blockchain technologies impact the financial system. However, the risk and return tradeoffs of cryptocurrencies are not well understood.

Building on the work of Liu, Tsyvinski, and Wu (2022) and Cong, Karolyi, Tang, and Zhao (2021), we develop a novel text-based factor-pricing framework that significantly improves our understanding of the cross-section of cryptocurrency returns. We collect news articles from Factiva that mention the top 43 cryptocurrencies by market capitalization as of December 2021. We then implement Bidirectional Encoder Representations from Transformers (BERT) topic modeling to identify the most prominent topics. We identify topics that are related to fundamentals, technical trading, regulation, lending, payments, derivatives, social media, and hedging. We find that the most important text-based measures are factors that capture fundamentals. This finding is in line with Filippou, Rapach, and Thimsen (2023), who find that fundamentals are the most important out-of-sample predictors of cryptocurrency returns.

Therefore our analysis focuses on the sentiment of articles that discuss issues related to fundamentals. We label the sentiment of the articles classified as fundamental by the BERT model as the Fundamental Sentiment Index (FSI). Our measure of sentiment is based on the difference in frequency between positive and negative words based on the Loughran and McDonald (2011) measure. This measure captures net positive sentiment (or optimism) on fundamental topics.

Cryptocurrency fundamentals affect the demand and supply of a currency. On the supply side, important factors include the hash rate, which measures the computational power of a blockchain network, factors that affect the technology of mining and the costs of executing transactions on the blockchain, such as gas fees on the Ethereum blockchain. On the demand side, fundamentals like the number of addresses and institutional factors such as demand for liquidity in exchanges matter for the utility of tokens in general payments and in offering platform services.

First, we estimate rolling betas to measure the exposure of different cryptocurrencies to the Fundamental Sentiment Index while controlling for various risk factors. These risk factors include cryptocurrency market factors, size, momentum, volatility, and liquidity. We categorize cryptocurrencies based on their token classification, indicating that those with high betas are typically platform or general payment tokens. These tokens exhibit positive co-movement with the fundamental sentiment, implying that changes in blockchain congestion or transaction benefits impact their utility. Conversely, low-beta currencies are often governance tokens, serving as a hedge against fundamental sentiment fluctuations. Supporting our token classification, we find that the betas relate to various value indicators, such as the ratio of transactions, users, and addresses to market cap, used in Cong et al. (2021). The analysis suggests that cryptocurrencies with higher value metrics are often used more for payments on blockchain platforms or as general payment currencies, while governance tokens, which are less traded in the secondary market, tend to have lower value metrics.

We now turn to asset pricing tests using the Fundamental-based sentiment (FSI) factor. We find that the factor is *positively* priced in the cross-section of cryptocurrency returns. The rationale behind this finding is that cryptocurrencies with positive exposure to fundamental analysis are riskier, so investors demand a risk premium for holding these cryptocurrencies. The intuition is that negative sentiment on fundamentals indicates the currency has weaker supply and demand fundamentals. This could be a lower hash rate, indicating lower security of the network, or alternatively, a reduction in the number of addresses using the currency as a medium of exchange. Therefore currencies that exhibit low returns when cryptocurrency fundamentals are weak are riskier and therefore command a risk premium.

To examine the predictive ability of fundamental sentiment (FSI), we form long-short portfolios based on the exposure of each cryptocurrency to these factors and sort cryptocurrencies into quartiles every week based on their 60-week rolling betas. Then we form long-short portfolios that buy cryptocurrencies with high exposure to FSI and sell cryptocurrencies with low exposure to this factor (HML_{FSI}). The fundamental-based strategy offers statistically significant returns that are higher than the cryptocurrency market return and has a Sharpe ratio of 1.24.

We provide four sets of results with our fundamental text-based factor. First, we show that conventional cryptocurrency risk factors cannot explain the returns of the text-based factors. We contemporaneously regress the fundamental spread portfolios on the market, size, momentum, liquidity, and volatility factors and find that the alpha of the regression is statistically and economically significant. Following Cong et al. (2021), we also consider different value factors, based on the number of transactions recorded, the cumulative number of addresses to date created, and the number of addresses with balance, and we find that the text-based fundamental factor provides positive and statistically significant alphas. We show that text-based fundamental sentiment factors are priced in the cross-section of cryptocurrency returns offering information over and above other existing factor models in the literature.

Second, we show that our fundamental factor correlates with value factors defined in Cong et al. (2021). This suggests that our factor captures value measures, such as the ratio of addresses, hash rate, and volume of transactions to market cap, and supports our theory that fundamental news is linked to the risk premia of currencies.

Third, we conduct Fama and MacBeth (1973) cross-sectional regressions to test the pricing ability of the fundamental factor after controlling for different determinants of cryptocurrency returns. We find that the fundamental factor is a strong predictor of cryptocurrency returns even after controlling for other characteristics. In a baseline model that includes a market factor and text-based factors, we find the price of risk for the fundamental factor is 1.2 percent. Our risk premia estimates are robust to adding alternative factor models that include volatility and momentum based on Liu et al. (2022) and Cong et al. (2021).

Finally, we conduct a number of additional robustness tests. Our results are robust when considering a smaller sample of a group of 15 cryptocurrencies with the highest market capitalization, alternative sentiment proxies, different numbers of portfolios, and different specifications to estimate factor betas. We also show that fundamental sentiment offers strong diversification benefits for other factors such as the market, size, illiquidity, volatility, and momentum portfolios. We also consider other types of sentiment for topics such as regulation, lending, derivatives, payments, social media, hedging, and technical trading. We find that they cannot explain the cross-section of cryptocurrency returns. The only exception is the technical sentiment which is a negative predictor of cryptocurrency returns, but it is unrelated to our fundamental sentiment factor.

Literature review. Our paper contributes to an emerging literature explaining the cross-section of cryptocurrency returns (Bianchi and Babiak, 2021; Cong et al., 2021; Liu, Tsyvinski, and Wu, 2021; Liu et al., 2022; Filippou et al., 2023; Bhambhwani, Delikouras, and Korniotis, 2021; Schwenkler and Zheng, 2020; Kogan, Makarov, Niessner, and Schoar, 2022; Bianchi, Babiak, and Dickerson, 2022; Han, Newton, Platanakis, Sutcliffe, and Ye, 2022; Luo, Mishra, Yarovaya, and Zhang, 2021).

The seminal work in Liu et al. (2022) establishes that cryptocurrency return factors based on market, momentum, and volatility have pricing power for the cross-section. However, in addition to return-based factors, Bhambhwani et al. (2021), and Cong et al. (2021) establish that value and network-based factors have sufficient explanatory power for cryptocurrency returns. In particular, blockchain characteristics relating to the hash rate and the number of addresses transacting with the network correlate positively with cryptocurrency prices. Higher exposure to these characteristics can, in turn, lead to higher expected returns, which provide investors with a risk premium. Our work is differentiated because we infer our fundamental factor directly from news on cryptocurrency articles. This decomposition allows us to disentangle alternative theories of cryptocurrency pricing more accurately and whether retail trading dominates it (e.g., Kogan et al., 2022) or by news on blockchain characteristics.

News and textual analysis in cryptocurrency markets has been used in prior work (Filippou et al., 2023; Schwenkler and Zheng, 2020; Liu, Sheng, and Wang, 2021). Schwenkler and Zheng (2020) use a textual analysis method to determine peer co-movement in cryptocurrency markets and document competition effects, where negative news about a peer can lead to substitution toward currencies that have similar network and blockchain characteristics. Filippou et al. (2023) use a variety of news sources (Factiva, Reddit comments, Google Trends) to develop return characteristics in a machine-learning model for forecasting cryptocurrency returns. The novelty of our paper in the textual analysis is using the BERT model to obtain text-based factors using cryptocurrency news. We measure the net sentiment in these topics to construct indices that measure optimism about fundamental news. Using standard asset pricing tests, we find both factors are priced in the cross-section of cryptocurrency returns and support alternative models that use value factors and blockchain characteristics. Liu et al. (2021) use reports from initial coin offerings to construct a technology index for cryptocurrencies. Cryptocurrencies with a higher technology index predict long-term positive performance.

The paper is organized as follows. Section 2 discusses the theoretical motivation and testable hypotheses of the paper. Section 3 outlines the data and definitions, including using BERT to identify cryptocurrency topics and constructing sentiment measures for fundamental news. Section 4 outlines our main empirical asset pricing tests. Section 5 concludes.

2 Testable Hypotheses

Fundamental analysis provides a framework that can help investors identify the intrinsic value of an asset by examining different related economic and financial factors.

In equities, traders can analyze the financial statements of a firm to estimate the underlying firm value and compare it with its market price. For example, Lev and Thiagarajan (1993) and Abarbanell and Bushee (1997) demonstrate that signals, which capture information on firm fundamentals such as inventory changes, gross margins, selling expenses, capital expenditures, effective tax rates, inventory methods, and labor force sales productivity, are associated with different rules of fundamental analysis used by financial analysts to forecast future firm performance. Abarbanell and Bushee (1998) build a trading strategy in the equities market using fundamental analysis signals and find that it offers abnormal returns.

In comparison to the equities market, which adheres to the Generally Accepted Accounting Principles (GAAP) for financial measurements, the cryptocurrency market lacks a standardized accounting framework. This absence poses a challenge for traders and regulators in determining the fundamental value of cryptocurrencies (Liu et al., 2021).

Nonetheless, there exists a wealth of publicly available information about economic activities within the blockchain. This real-time data, verifiable through the public ledger, holds promise in establishing the intrinsic value of a cryptocurrency. For instance, Liu et al. (2021) apply accounting and finance valuation methods to the cryptocurrency market, highlighting the significance of information related to new addresses as highly value-relevant for cryptocurrencies. Additionally, Bhambhwani et al. (2021) present evidence that both the number of addresses and the hash rate serve as robust predictors of cryptocurrency returns. Moreover, Cong et al. (2021) assert that these blockchain characteristics can be utilized as value-based factors.

To motivate our discussion of fundamental blockchain characteristics, we introduce a simple model based on Biais, Bisiere, Bouvard, Casamatta, and Menkveld (2023).[1] The model is an overlapping generations framework, where consumers in the young generation have access to cryptocurrency and fiat money as a medium of exchange. Cryptocurrencies have the feature that they incur transaction costs ψ_t: this could be fees incurred on exchanges and the costs of validating transactions through mining. In the next period, generations earn benefits from holding cryptocurrency, which we capture through the parameter θ_{t+1}. This can include transactional benefits such as cryptocurrency use in cross-border payments and its programmability and smart contract features. This model yields an intuitive Euler equation that relates the cryptocurrency price today, p_t, to its perceived 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 $1+\mathscr{T}_t = \frac{1+\theta_{t+1}}{1+\psi_t}$ and iterating forward, we obtain an expression which states that the cryptocurrency price is the net present value of its future stream of transactional benefits:

[1]For the derivation and more details on the model we refer readers to the Appendix. We simplify the analysis in Biais et al. (2023) as we do not discuss the role of crash risk and the role of hackers in our model setup.

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

We can use this simple framework to structure our hypothesis.

Hypothesis 1 (H1). *Cryptocurrencies that have positive exposure to fundamental sentiment are riskier.*

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

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

Our hypothesis is that exposure to cryptocurrency fundamentals, captured by our news sentiment measure, is priced in the cross-section of cryptocurrency returns. In particular, currencies that exhibit returns that co-move positively with the fundamental sentiment factor are riskier: and currencies that co-move less provide a risk hedge against fundamentals. To test this hypothesis in our empirical analysis, we construct a news-based measure of sentiment about cryptocurrency fundamentals. Our news sources include discussions from experts on the market about the economics of the blockchain and demand and supply dynamics.

Our model also shows that the sensitivity of a currency to blockchain fundamentals is conditional on being used as a medium of exchange. This is important as we have different cryptocurrency tokens, which we highlight in Section 3.5. While some token types (general payment and platform) are used for payments and compensate miners for authenticating transactions, other token types are used for governance and voting rights. There are differences in the sensitivity of these token types to our measure of cryptocurrency fundamentals, which typically capture factors related to blockchain congestion, such as the hash rate, and shocks to cryptocurrency mining. These differences in token use translate to differences in the sensitivity to fundamental news sentiment. Therefore, we hypothesize that general payment and platform tokens are more sensitive to fundamental news and investors demand a risk premium for exposure to fundamental risk. In contrast, we hypothesize governance tokens are less sensitive to fundamental news and provide a risk hedge against fundamental news-based sentiment.

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

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.[2] 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.[3]

Table A1 of the Internet Appendix offers summary statistics of the data per year. Specifically, 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

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

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

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 cryptocurrencies in the market.

3.2 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.[4] Our data span the period from June 2017 to December 2021. During this sample period, 27,382 articles satisfy our search criteria.

3.3 BERT topic modeling approach

Our goal is to extract the most prominent topic from the news articles. This way, we can reduce the noise in our estimates and obtain factors that provide useful information for the cross-section of cryptocurrency returns. More conventional topic modeling methods extensively used in the literature are Latent Dirichlet Allocation (LDA) and Latent Semantic Indexing (LSI). Despite their widespread usage, there are shortcomings associated with these methods. The most crucial is that LDA and LSI rely on the bag-of-word representation of documents, implying that word ordering and semantics are overlooked. On the other hand, the Bidirectional Encoder Representations from Transformers (BERT) modeling approach is a state-of-the-art topic modeling structure developed to overcome these shortcomings

[4] 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 Noticias-Financieras* (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*.

(e.g., Devlin, Chang, Lee, and Toutanova, 2018). It is, therefore, our choice of algorithm to explore the topics of our corpus.

There are many variants of embeddings, such as Word2Vec (Mikolov, Chen, Corrado, and Dean (2017)), GloVe (Pennington, Socher, and Manning (2014)), and FastText (Joulin, Grave, Bojanowski, Douze, Jégou, and Mikolov (2016)), that utilize the distributional hypothesis to capture relationships in the embedding space (Harris (1954)). This hypothesis suggests that semantically similar words exhibit similar distributions and thus appear in related linguistic contexts. Nevertheless, these methods have certain shortcomings. More precisely, they consider word-level embeddings because they usually demonstrate lower performance as sentence encoders meaning that they usually misinterpret context. Therefore, they use as an input to the model one word, and the output is a vector representation of that word (Perone, Silveira, and Paula (2018)). On the other hand, BERT focuses on contextual embeddings, so the input to the model is a sentence instead of a single word. This method is also directional, which implies that it takes into account both preceding and subsequent context to generate the embeddings of a word, in contrast to unidirectional models (like ELMo and ULMFit). To this end, BERT can interpret texts with more precision and has been increasingly used in other studies; for example, see Chava, Du, and Malakar (2021) and Gorodnichenko, Pham, and Talavera (2023).

BERT is built to pre-train deep bidirectional representations from the unlabeled text by joint conditioning on both the left and right context in all layers.[5] As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer for a wide range of tasks, including topic modeling.[6] BERT also reduces the number of unique words that are included in the model by partitioning each word into smaller tokens (such as subwords). It is also designed to encode entire sentences with a length of 512 tokens.

[5]The neural network of BERT is pre-electronic trained on 800 million BooksCorpus and 2,500 million Wikipedia words. Thus, the model is able to identify words that have similar meanings based on pre-training.

[6]There are different variants of BERT embeddings based on the training data and the architecture. We consider word embeddings that are created from the BERT base model (12 layers, 768 hidden states, 12 heads, and 110 million parameters).

Transfer learning is another advantage of this method compared to previous word vector models. In other words, a model created for a task is used as the starting point for a model on the following task.

The input required from the BERT topic modeling approach is the corpus, which is the set of cryptocurrency news articles in our case. In the first step, Sentence Transformers are used to extract document embeddings. The pre-trained model we use to extract document embeddings is RoBERTa, developed by Liu, Ott, Goyal, Du, Joshi, Chen, Levy, Lewis, Zettlemoyer, and Stoyanov (2019). In particular, documents are embedded to create representations in vector space that can be compared semantically. The next step in this process is to apply the UMAP algorithm (McInnes, Healy, and Melville, 2018) to the document embeddings. The purpose of this step is to reduce dimensions and cluster similar documents. UMAP is used to reduce the dimensionality of the vectors to 5 with the size of the neighborhood set to 15. The number of nearest neighbors optimizes the balance between the local and global structure in the new embedding, and this value gives the best results in preserving the local structure. Semantically similar documents are also grouped in different clusters. The last step is topic creation based on a class-based variant of TF-IDF (Term Frequency–Inverse Document Frequency) (i.e., c-TF-IDF). At this stage, all documents in the same cluster are treated as a single document. c-TF-IDF, which is a score indicating the importance of a word for a particular cluster, is constructed based on the following equation:

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

where t_i is the frequency of term t in cluster i, and it is divided by the total number of words in the cluster w_i. This is multiplied by the logarithmically scaled fraction of the total number of n documents across all clusters m divided by the sum of occurrences of term t in all those documents. Words with top c-TF-IDF in each cluster help us label that cluster. We summarize the process in Figure 1.

[FIGURE 1 ABOUT HERE]

The output generated from BERT topic modeling for our corpus is 20 topics and the top 30 keywords for each topic. We summarize the keywords for the topic we identify as having Fundamental content in Figure 2. We find that the most prominent words in this topic include words that describe fundamentals such as mining, hash, hash rate, operations, network, power, technology, securities, rate, hardware, and bitmain.

[FIGURE 2 ABOUT HERE]

3.4 Fundamental Sentiment Index

BERT gives us a sample of news articles classified as having fundamental content. We plot the raw number of fundamental news articles over time in Figure 3. We can see that the number of fundamental news articles spikes around events such as the cryptocurrency mining malware in North Korea, the bitcoin mining blackout in China, or the crackdown on cryptocurrency mining by China.

[FIGURE 3 ABOUT HERE]

We calculate the sentiment of articles with fundamental trading content. Specifically, we count the number of positive and negative words in Loughran and McDonald (2011) dictionary. We only compute the sentiment of the sentences that mention the specific cryptocurrencies in our dataset to reduce the noise in our measure. Therefore the sentiment takes the following form:

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

where FSI denotes the fundamental sentiment. Therefore an increase in the sentiment measure indicates higher optimism about fundamentals in the cryptocurrency market. An example of a sentence is "*coinhive reportedly had to shut down its services amidst a 50 percent decline in hash rate following the last monero hard fork.*", with a sentiment measure of -0.2.[7]

Summary statistics of FSI are reported in Table 1. FSI exhibits negative skewness and excess kurtosis. The FSI is stationary according to the augmented Dickey-Fuller test. Correlations between FSI and some other prominent risk factors in the cryptocurrency pricing literature are reported. Our FSI index exhibits 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 they are statistically significant. Overall results suggest that the fundamental sentiment index captures different dimensions of risks compared with other conventional risk factors in the literature.

[TABLE 1 ABOUT HERE]

Figure 4 displays the time-series of the FSI index. It further verifies that the FSI index is stationary and it captures different periods with positive and negative fundamental sentiment. For example, it spiked at the end of 2020 when Bitcoin appreciated in value, and it dropped around January 2022 in response to news on China's cryptocurrency ban.

[FIGURE 4 ABOUT HERE]

[7]For more details on the types of articles that are classified as fundamental, we refer readers to Appendix A and B of the Internet Appendix. We also provide more examples of fundamental article sentences and their sentiment scores.

3.5 Token classification

To understand the different types of tokens, we follow the nomenclature in Cong and Xiao (2021) and categorize our 43 currencies into different types of cryptocurrency tokens.[8] A full list of the token classification is provided in Appendix Table A3.

- **Governance Tokens**: Governance tokens are used to give holders the power to participate in decision-making processes related to the development and management of a blockchain platform or protocol. Token holders may propose and vote on changes, upgrades, and other governance matters. These tokens aim to create a decentralized governance model where the community has a say in the network's evolution. One example is MakerDAO's governance token (MKR) which allows holders to participate in decisions about the stability and governance of the MakerDAO platform, which issues the stablecoin DAI.

- **Platform Tokens**: Platform tokens are native to a specific blockchain platform and are used for various purposes, such as paying transaction fees, executing smart contracts, and accessing resources on that platform. They often serve as the primary means of value transfer within the ecosystem. For example, Ethereum (ETH) is the native cryptocurrency of the Ethereum platform and is used to pay for transaction fees, execute smart contracts, and participate in decentralized applications (DApps) built on the Ethereum blockchain.

- **Product Tokens**: Product tokens are issued by specific projects or companies as a form of value representation within their respective ecosystems. These tokens are often tied to a particular product, service, or utility offered by the issuing entity. For example, Basic Attention Token (BAT) - BAT is used within the Brave browser

[8] Relative to Cong and Xiao (2021), we further differentiate between platform tokens and governance tokens. This distinction is important when we discuss the cross-sectional characteristics of currency betas in Section 4.1

ecosystem to reward users for viewing advertisements and content. It aims to create a more equitable and efficient digital advertising ecosystem.

- **General Payment Tokens**: General payment tokens are designed primarily for facilitating transactions and serving as a means of exchange. They are intended to be used for everyday purchases and transactions. A common example is Bitcoin (BTC), which can be used as a medium of exchange.

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 Sensitivities to Fundamental Sentiment

Rolling Betas. To measure the exposure of each cryptocurrency to the fundamental sentiment index (FSI), we regress individual cryptocurrency excess returns at time t on a constant and FSI, controlling for other cryptocurrency risk factors. These risk factors include the cryptocurrency market factor (MKT), size factor (SMB), momentum factor (MOM), volatility factor (VOL), and liquidity factor ($ILLIQ$).[9] The estimation is based on a 60-week rolling window. The time-varying slope coefficient obtained from this regression is $\beta_{i,t}^{FSI}$. Specifically, we estimate the model below:

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

[9]Description of risk factors can be found in Table A2 of the Internet Appendix.

where $rx_{i,t}$ is the cryptocurrency return at time t, FSI_t represents the fundamental sentiment index at time t. We include controls in the regression to account for other determinants of cryptocurrency returns.

Interpretation of the betas. We provide an economic interpretation of the betas. Specifically, we provide in Figures A1 and A2 of the Internet Appendix the time-series plots of each currency beta, and the rolling beta of the low and high currency portfolio. We observe that larger cryptocurrencies such as BTC are more stable in comparison to smaller and more volatile cryptocurrencies such as ADA. Figure 5 shows the time-series average of the betas with respect to the fundamental sentiment factor. The cryptocurrencies are sorted based on their beta. Thus, those cryptocurrencies that appear to the left of the figure have the most negative exposure to the fundamental sentiment index and the cryptocurrencies that appear at the right of the figure demonstrate the most positive betas.

[FIGURE 5 ABOUT HERE]

Based on the token classification in Table A3, we find there are cross-sectional differences in the token types from low to high beta currencies. Currencies that have a high beta with respect to the fundamental sentiment index are typically platform or general payment tokens. Token returns co-move positively with our measure of fundamental sentiment. An increase in congestion on the blockchain or an increase in the transaction benefits can impact the utility of these tokens, in accordance with the stylized model presented in Section 2. Therefore these currencies are riskier as they are used predominantly as a medium of exchange or by having transfer value on the blockchain platform. Alternatively, currencies with a low (negative) beta are typically governance tokens. These are different to platform and general payment tokens as they are not used to make payments but are rather staked in protocols to vote on governance proposals. The news captured in fundamental sentiment, such as blockchain characteristics like the hash rate and other mining related costs, have less effect on the valuation of these tokens. Therefore governance tokens provide a hedge against our measure of fundamental sentiment.

Fundamental Sentiment Betas and Value Factors. We now turn to examine characteristics that can explain the cross-sectional variation in fundamental sentiment betas. We run Fama and MacBeth (1973) cross-sectional regressions for Fundamental Sentiment Index betas β^{FSI} on different measures of value (See Appendix Table A2 for more details). Thus, our model takes the form:

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

where $\hat{\beta}^{FSI}_{i,t}$ denotes the 60-week rolling betas with the FSI index. $Value$ represents different measures of cryptocurrency value (Cong et al., 2021). These value measures record the difference in returns in portfolios sorted by transaction-to-market cap ratio, user-to-market cap ratio, and address-to-market cap ratio. Value captures broadly the network effects of a currency that has higher transactions, users, and addresses. These currencies are precisely ones that are used more for payments on a blockchain platform or as a general payment currency. In contrast, governance tokens are staked by users to vote on proposals and are used less in the secondary market to trade and for payment services or administering fees. Therefore these tokens have low value, as measured by the transactions, users, and address to market cap ratio.

Table 2 displays the average coefficients of contemporaneous cross-sectional regressions that are estimated on a weekly basis. We find that the value factors are strong positive predictors of the cross-section of the fundamental sentiment betas. This finding indicates the strong connection of the value factors with the fundamental sentiment index and highlights the ability of the BERT model to extract meaningful topics. The average cross-sectional R-squares range from 9% to 12%.

[TABLE 2 ABOUT HERE]

4.2 Descriptive Statistics

In this section, we sort cryptocurrencies into portfolios based on their exposure to the fundamental sentiment index.

Fundamental Sentiment Portfolio Construction. At time t, we sort cryptocurrencies into quartiles based on their previous week (i.e. $t-1$) betas with FSI. We rebalance our portfolios weekly. The first portfolio (P_1) includes currencies with the lowest betas, while the fourth portfolio (P_4) covers currencies with the highest betas. We then construct a zero-cost portfolio (HML_{FSI}), which goes short the first portfolio (P_1) and long the high beta portfolio (P_4).

Summary Statistics. If the fundamental sentiment index is a pricing factor for the cross-section of cryptocurrency returns, there should be a significant dispersion in excess returns between low-beta and high-beta portfolios. Therefore the corresponding spread portfolio HML_{FSI} should generate statistically significant excess returns. Table 3 reports summary statistics of portfolios sorted on the exposure to the fundamental sentiment index (β_{FSI}).

Table 3 shows that going long in cryptocurrencies with the highest exposure to FSI (β_{FSI}) while short-selling cryptocurrencies with the most negative exposure to FSI yields average positive excess returns. The average portfolio returns increase monotonically with the FSI beta. The HML_{FSI} portfolio yields an annualized average excess return of 65% with a Sharpe ratio of 1.24 per annum. The fundamental sentiment strategy exhibits positive skewness and excess kurtosis.

[TABLE 3 ABOUT HERE]

Figure 6 displays cumulative returns of the fundamental sentiment strategy (HML_{FSI}) and the market factor. We observe that the FSI strategy is profitable and outperforms the market portfolio. In addition, the strategy is less affected by crashes in the cryptocurrency market. In contrast, we observe that the market portfolio exhibits poor performance in the period between 2019 and 2020 as well as during the cryptocurrency crash in November 2021. The market portfolio demonstrates a positive performance in the period between the end of 2020 and April 2021 which was arguably the period that the bitcoin exhibited significant performance with other cryptocurrencies to follow a similar pattern.

[FIGURE 6 ABOUT HERE]

Figure 7 shows the portfolio turnover of the FSI strategy. Specifically, we report the frequency with which each currency appears in the low and high fundamental sentiment portfolios. Panel A and Panel B of Figure 7 show the results for low and high FSI portfolios, respectively. We find the FSI strategy is driven more by BAT, GNO, and NEO in the low FSI portfolio and ADA, DOGE, LINK, and XLM in the high FSI portfolio. For example, GNO appears in the low beta portfolio almost 70% of the weeks that we re-balance our portfolio, and NEO appears 60% of the total number of holding periods. Similarly, XLM appears in almost 70% of the weeks in the high fundamental sentiment portfolio. DOGE, ADA and LINK are typically in the high beta portfolio in 60% of our sample.

[FIGURE 7 ABOUT HERE]

4.3 Fundamental Sentiment Portfolios and Other Investment Strategies

In this section, we test whether our sentiment factor offers significant alphas after controlling for market, size, momentum, liquidity, and volatility risk factors. The first column of Table 4 offers results for a contemporaneous regression of the spread portfolio of the fundamental strategy (HML_{FSI}) on the market factor. The coefficient of the market portfolio is positive but statistically insignificant, whereas the alpha is 67.6% annually and statistically significant with a Newey and West (1987) t-statistic of 2.66. Table 4 also shows the link between HML_{FSI} and other conventional investment strategies. Specifically, we consider a two-factor model that comprises a market and a size factor. We find that the FSI strategy offers an alpha of 78% which is statistically significant. In a similar fashion, we augment the previous model with a momentum factor and the FSI strategy offers an alpha of 72.8%. Finally, we include liquidity and volatility factors by adding one factor each time and find that these models also offer an alpha of 72.8% that is statistically significant at 1% significance level. Our results suggest that the HML_{FSI} strategy can generate a positive and statistically significant alpha even after considering conventional asset pricing models.

[TABLE 4 ABOUT HERE]

We then examine the link between HML_{FSI} and other fundamental related risk factors in the cryptocurrency literature. In particular, we test if our factors are explained by fundamental risk factors constructed in Cong et al. (2021). We regress HML_{FSI} contemporaneously on three value factors to see if these value factors can explain the returns generated by FSI. Table 5 displays results with three independent variables: the value factors are constructed based on the transaction-to-market ratio (T/M), user-to-market ratio (U/M), and address-to-market ratio (A/M) respectively.[10] The coefficients for all three value factors

[10] Description of risk factors can be found in Table A2 of the Internet Appendix.

are positive and strongly significant, with t-statistics of 3.90, 3.76, and 3.63, respectively. This is intuitive and supports our hypothesis that our measure of fundamental sentiment is interpreted as an over-under valuation of a currency. The alphas of the regressions also remain positive and statistically significant in all regressions. Therefore, the results highlight an important finding: while value factors are positively correlated with HML_{FSI}, they cannot fully explain HML_{FSI}. It means that HML_{FSI} captures a different dimension of fundamental cryptocurrency characteristics beyond the three value factors.

[TABLE 5 ABOUT HERE]

4.4 Asset Pricing Tests

4.4.1 Framework

We start with a framework to conduct our asset pricing tests. Under general conditions, there exists a stochastic discount factor (SDF) M_t, which can price the excess returns of any asset i, $rx_{i,t}$.

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

Following Bhambhwani et al. (2021), we assume the SDF is a linear function of observable factors F_t, where μ_F where f_t are factors centered around their means and b is a vector of parameters.

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

Using the equation for the SDF, we can write returns as a linear function of factor betas:

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

where β_i measures the exposure of returns to factor i, and λ is a measure of the risk price associated with factor i.[11] We will use this standard linear-beta representation of the SDF in cross-sectional regressions of expected returns on factor-beta. In our first method in section 4.4.2, we will use individual currencies as test assets and estimate the risk price of our fundamental sentiment factor and whether it can explain the cross-section of cryptocurrency returns. We will then conduct a two-step Fama Macbeth regression in section 4.4.3, where we use portfolios sorted on lagged sentiment measures to estimate the risk price.

4.4.2 Estimating the Price of Risk

Test assets. Our test assets are individual currencies rather than portfolios. Ang, Liu, and Schwarz (2018) suggest that grouping stocks into portfolios shrinks the betas' cross-sectional dispersion, which leads to a less efficient estimate of factor risk premia. Bali, Brown, and Tang (2017) estimate the risk price of economic uncertainty using individual stocks. In the context of currencies, Barroso, Kho, Rouxelin, and Yang (2018) test the risk price of global imbalances using individual currencies.

Cross-sectional Regressions. We now investigate the risk price of β^{FSI}. Having estimated β^{FSI} from equation (5), we investigate the cross-sectional relation between these betas and expected excess returns at the cryptocurrency level (Bali et al., 2017). In particular, we run weekly cross-sectional regressions at each time t:

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

[11] $\beta_i = \mathbb{E}[(f_t - \mu_F)(f_t - \mu_F)']^{-1}\mathbb{E}[(f_t - \mu_F)'R_{i,t}]$ 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.

where $X_{i,t}$ is a set of control variables, including β^{MKT}, β^{Size}, $\beta^{Momentum}$, and $\beta^{Volatility}$ estimated from Equation (5). We then take the time-series average of slope coefficients $\lambda_{1,t}$ and report its Newey and West (1987) t-statistic and average adjusted R^2.

Table 6 summarizes results regarding the estimation of risk prices of β^{FSI} from regressions (2) and (3). The univariate regression results shown in the first column suggest a negatively significant link between the β^{FSI} and the cross-section of future cryptocurrency excess returns. The coefficient of β^{FSI} is 0.004 with a t-statistic of 2.61. This positive coefficient for β^{FSI} is in line with the portfolio results shown in Table 3, meaning that taking a long position in currencies with higher β^{FSI} predicts positive returns in the following period. To examine the economic significance of this result, we compute the difference in average β^{FSI} between P_1 and P_4 from Table 3, which is 2.76 [=1.46 - (-1.30)]. If a currency were to move from P_1 to P_4, its expected return would increase by 1.10% [=2.76 × 0.004] per week. Therefore, the risk price of the β^{FSI} is both statistically and economically significant. The β^{FSI} coefficient is robust to adding factors controlling for market, volatility, and momentum. In the full specification, the risk price of β^{FSI} is 0.005 with a Newey and West (1987) t-statistic of 2.00. This finding further verifies that the fundamental sentiment index is a strong positive predictor of cryptocurrency returns after accounting for other determinants of cryptocurrency risk premia.

[TABLE 6 ABOUT HERE]

Having found evidence of strong predictive power of β^{FSI} for the next week's cryptocurrency returns, we now test whether our sentiment factors have predictive power at a longer horizon. To this end, we regress cryptocurrency excess returns from $t+2$ weeks to $t+12$ weeks ahead on β^{FSI} at time t and the same set of control variables that we considered in the previous section.

In Table 7, we observe that the coefficient of β^{FSI} is positive and strongly significant up to 5 weeks ahead and gradually fades away with the exception of weeks 7 to 10 where the coefficient is positive and significant. Thus, we can argue that the strategy reverses in 5 to 11 weeks.

[TABLE 7 ABOUT HERE]

4.4.3 Fama Macbeth Asset pricing Tests

We apply a Fama and MacBeth (1973) (FMB) two-pass regression. Our portfolios $P1$ to $P4$ are defined in section 4.2, and are constructed by sorting portfolios based on lagged values of the β_{FSI} respectively. We then construct a measure of returns of the high beta portfolio (P4) less the returns of low beta portfolio (P1), HML_{FSI}. For each sentiment measure, in the first stage, we run contemporaneous time-series regressions of currency portfolio excess returns on the risk factors. In our baseline specification, we only include the market factor. We conduct FMB two-pass regressions for a number of alternative specifications in section 4.4.4.

$$rx_{i,t} = \alpha_{i,t} + \beta_i^{FSI} HML_{FSI,t} + \beta_i^{MKT} MKT_t + \epsilon_{i,t}, \qquad i = P1, P2, P3, P4 \tag{10}$$

In the second stage, we perform cross-sectional regressions of average portfolio returns on factor loadings, calculated in the previous step, to obtain 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 report the results for a two-factor model that consists of the market factor (MKT) and the sentiment factors in Panel A of Table 8. We provide estimates for the implied risk factor (λ) and the corresponding Newey and West (1987) t-statistic, the root mean square error (RMSE), and cross-sectional R-squared. We find that the risk price λ^{FSI} is 1.30% per week with a Newey and West (1987) t-statistic of 2.90. RMSE and R^2 of this regression are 0.0018 and 0.86, respectively. Overall, results in this section provide further evidence that

λ^{FSI} are priced in the cross-section of cryptocurrency returns. We also report t-statistics based on Shanken (1992) standard errors, which account for the error-in-variable problem – the fact the regressors of the second pass regression are estimated in the first pass regression. We find that our results remain highly significant at the 5% significance level.

Panel B of Table 8 considers more test assets. Specifically, we consider 24 test assets that include four size, four momentum, four liquidity, four volatility, four value, and four fundamental sentiment portfolios. The purpose of this exercise is to account for lucky factors. Lewellen, Nagel, and Shanken (2010); Harvey et al. (2022) argue that is relatively easy to find factors that can price the cross-section of portfolios with a strong factor structure. Thus, we consider a larger cross-section of 24 test assets to account for this possibility. Panel B of Table 8 shows that the risk price λ^{FSI} is 1% per week, and it is statistically significant.

[TABLE 8 ABOUT HERE]

4.4.4 Can Fundamental Sentiment improve other asset pricing models?

In this section, we explore the role of FSI in improving existing cryptocurrency asset pricing models. We augment a set of existing factor models with the FSI factor and the 24 test assets. Table 9 reports the prices of risk and the corresponding t-statistics as well as the RMSE and the cross-sectional R^2 of the regression. In Panel A, we show results for a two-factor model that comprises a market (MKT) and size factor. We find the market factor is not priced in the cross-section of cryptocurrency returns which is in line with the literature and size if positively priced. We augment the model with the HML_{FSI} factor, and we find that the FSI factor is priced in the cross-section of cryptocurrency returns. The cross-sectional R-square increased from 15% to 36% which highlights an improvement in terms of goodness-of-fit in the model. We perform a similar exercise in Panel B where we consider a three-factor model with a market factor, size, and momentum. The inclusion of the fundamental factor in the model increases the cross-sectional R-square from 17% to 38%, and the price of risk for FSI is positive and statistically significant. Finally, Panel C investigates the performance of a five-factor model that includes the market, size, momentum, liquidity, and volatility models. We augment the model with the FSI factor and the cross-sectional R-square increases from 25% to 39%, and the price of risk for the fundamental sentiment factor is significant at a 5% significance level.

In all cases, adding HML_{FSI} significantly improves the existing asset pricing models for cryptocurrencies in terms of goodness-of-fit and statistical significance. Our results remain significant after we consider t-statistics that are based on Shanken (1992) standard errors.

[TABLE 9 ABOUT HERE]

4.5 Robustness tests

4.5.1 Top 15 cryptocurrencies

To ensure that smaller cryptocurrencies do not drive our results, we replicate the strategy with the top 15 cryptocurrencies ranked by average market capitalization during our sample period.[12] We report the portfolio sorting results in Table 10.[13]

We show results when sorting cryptocurrencies based on β^{FSI}. The average returns increase in a monotonic fashion from portfolio 1 to portfolio 3. The HML_{FSI} portfolio achieves an excess return of 77% annually with a Newey and West (1987) t-statistic of 2.06. The Sharpe ratio of this portfolio is 1.04 per annum. Therefore we provide evidence that our results are robust to the sample choice of cryptocurrencies.

[TABLE 10 ABOUT HERE]

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

[13] We limit the number of portfolios to three due to the limited number of cryptocurrencies available at the beginning of the sample.

4.5.2 Alternative specifications to estimate β^{FSI}

We estimate β^{FSI} based on alternative specifications to equation (5), which uses 5 factors, market, size, momentum, volatility and liquidity in addition to the sentiment factor. In the first alternative specification, we only control for the market (MKT) factor when we estimate the betas. The model takes the following form:

$$rx_{i,t} = \alpha_{i,t} + \beta_{i,t}^{SI} SI + \beta_{i,t}^{MKT} MKT_t + \epsilon_{i,t}, \tag{12}$$

We also consider a specification with the market (MKT), size (SMB), and momentum (MOM) factors. The model takes the form below:

$$rx_{i,t} = \alpha_{i,t} + \beta_{i,t}^{FSI} FSI + \beta_{i,t}^{MKT} MKT_t + \beta_{i,t}^{SMB} SMB_t + \beta_{i,t}^{MOM} MOM_t + \epsilon_{i,t}. \tag{13}$$

We then construct the long-short strategy based on past β^{FSI}. Summary statistics of portfolios for specification (12) and (13) are shown in Table 11. Constructing long-short portfolios always generates positive and statistically significant returns when alternative specifications to estimate β^{FSI} are used. Therefore our fundamental sentiment factor is robust to including alternative factor models as controls.

[TABLE 11 ABOUT HERE]

4.5.3 Alternative proxies for sentiment

We construct alternative measures of sentiment of fundamental trading factors to equation (4). One alternative is to use the number of negative words over the total number of words (equation (14)), or the net negative words over the total number of positive and negative words (equation (15)).

$$Sent = 1 - \frac{\text{Number of negative words}}{\text{Total number of words}} \tag{14}$$

$$Sent = \frac{\text{Number of positive words-Number of negative words}}{\text{Number of positive words+Number of negative words}} \tag{15}$$

We report the results for these two alternative proxies in Table A4 of the Internet Appendix. Specifically, Panel A shows results for the first measure which considers only negative sentiment. Panel B offers results for the second measure. The results suggest that these alternative measures of sentiment generate robust factors in predicting cryptocurrency returns. In other words, long-short portfolios that buy high fundamental sentiment cryptocurrencies and sell low fundamental sentiment cryptocurrencies offer an annualized return of 62% and 60%, respectively. The corresponding Sharpe ratios are 1.21 and 1.16 per annum. Thus, our results are robust to alternative specifications of the sentiment.

4.5.4 Different Number of Portfolios

Our main results focus on quartile portfolios. Table A5 of the Internet Appendix shows that the choice of the number of portfolios does affect our results. Specifically, we show similar results for tercile in Panel A and quintile portfolios in Panel B. The returns of the fundamental sentiment strategy are 55% and 62% per annum.

4.5.5 Diversification Benefits

Table A6 of the Internet Appendix examines the diversification benefits that the fundamental sentiment factor could offer to other factors in the literature. Specifically, we investigate whether the fundamental sentiment factor could improve the Sharpe ratios of the market, size, illiquidity, volatility, and momentum factors. Thus, we form an equally weighted portfolio between each factor and the fundamental factor. Panel A shows summary statistics of each factor. Panel B shows the summary statistics of the factors after blending them with the HML_{FSI} factor. The last row of Panel B shows the weight of each factor in the portfolio. The last column of Panel B considers an equally weighted portfolio of all factors and the fundamental sentiment. For this reason, the weight for the fundamental sentiment is only 16% in this portfolio. We find that the fundamental sentiment factor offers strong diversification benefits to all other strategies that we consider. For example, the annualized Sharpe ratio for the market portfolio increases 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. There is also a significant improvement in the equally-weighted portfolio that considers all factors. The Sharpe ratio increases from 0.91 to 1.41 per annum. Overall, we find that fundamental sentiment provides strong diversification benefits to a large number of factors.

4.5.6 Sentiment of Other Topics

To alleviate the concern that our results are due to data mining, we construct the sentiment index for other topics identified by BERT topic modeling. In this way, we also examine the role of other types of sentiment in the cryptocurrency market. These topics include lending, regulation, payment, derivatives, social media, hedging, and technical trading. Our results are in Table 12. Word clouds for these topics are provided in the Internet Appendix Figures A3 to A8. The sentiment of most of these topics does not predict cross-sectional cryptocurrency returns, and the HML portfolios constructed using these factors are statistically insignificant. The only exception is the sentiment of the technical trading topic which offers negative and statistically significant returns. We find that a strategy that goes long low technical sentiment (TSI) cryptocurrencies and sells high technical sentiment cryptocurrencies offers a return of 71% with a Sharpe ratio is 1.30 per annum. Figures A1 to A7 provide word clouds of the most prominent words in each topic.

[TABLE 12 ABOUT HERE]

Technical Sentiment. The absence of observable fundamentals could lead investors to rely more on price patterns. In imperfect markets, Treynor and Ferguson (1985), Brown and Jennings (1989), Hong and Stein (1999), Cespa and Vives (2012), Edmans, Goldstein, and Jiang (2015), Han, Zhou, and Zhu (2016), Keloharju, Linnainmaa, and Nyberg (2019) show that past stock prices offer important information for future prices. This finding implies that technical indicators, which are based on past prices, could be important trading signals. Brock, Lakonishok, and LeBaron (1992) and Lo, Mamaysky, and Wang (2000) show empirical evidence that technical indicators are profitable signals in the stock market.

We find that discussions in the media about price movements have strong implications for technical trading in the cryptocurrency market. This is particularly important in the absence of a standardized accounting framework that could offer reliable financial measurements. For example, Detzel, Liu, Strauss, Zhou, and Zhu (2021), among others, focus on the bitcoin and find that 1- to 20-week moving averages of daily prices forecast bitcoin returns in-sample and out-of-sample. They show theoretically in an equilibrium model that when there is uncertainty about growth in fundamentals, rational learning by investors with different priors could lead to strong predictability of returns by moving average rules. Therefore, we conjecture that the sentiment of discussions about price movements from experts in the media should offer important information for the cross-sectional of cryptocurrency returns.

We explore the robustness of this finding in Table A7 of the Internet Appendix where we run Fama and MacBeth (1973) regressions of cryptocurrency returns at time $t+1$ on TSI betas and a number of controls at time t. The control variables are size, momentum, liquidity, and volatility. We find that technical sentiment is a strong negative predictor of the cross-section of cryptocurrency returns even after accounting for these factors. In Table A8 we run similar cross-sectional regressions, and we add the FSI betas in the regression so as to assess whether one factor subsumes the predictive ability of the other. We find that both factors are priced in the cross-section of cryptocurrency returns with an opposite sign. Thus, they offer distinct information. This is not surprising given that the two factors exhibit a very low correlation of 0.04.

Overall, we find that technical sentiment is an important factor for the cross-section of cryptocurrency returns. Investors require a risk premium for holding cryptocurrencies with high pessimism about technical trading. We also show that this factor is unrelated to the fundamental factors which contain unique information.

5 Conclusion

This paper investigates the cross-sectional predictive ability of text-based fundamentals in the cryptocurrency market. We develop a novel text-based factor-pricing framework that significantly improves our understanding of the cross-section of cryptocurrency returns. We collect news articles that mention the top 43 cryptocurrencies by market capitalization and implement Bidirectional Encoder Representations from Transformers (BERT) topic modeling to identify the most prominent topics related to fundamentals, technical trading, regulation, lending, payments, derivatives, social media, and hedging. We then identify factors that capture fundamentals and analyze their sentiment using a difference in frequency between positive and negative words.

We find that the most important text-based measures are factors that capture cryptocurrency fundamentals. Fundamental analysis considers factors affecting the demand and supply of a cryptocurrency, including hash rate, mining technology, transaction costs, and institutional demand for liquidity. Classifying cryptocurrencies based on their token types, we show that high-beta currencies are typically platform or general payment tokens. These token returns show a positive correlation with fundamental sentiment, reflecting their sensitivity to changes in blockchain efficiency and transaction benefits. Conversely, low-beta currencies are often governance tokens and are a hedge against our measure of fundamental sentiment.

Constructing a portfolio sort based on the betas with respect to Fundamental-based sentiment (FSI), we find this factor is priced in the cross-section of cryptocurrency returns. Cryptocurrencies that co-move positively with the fundamental sentiment factor are riskier, and investors demand a risk premium as compensation.

We show that the text-based fundamental sentiment factor is priced in the cross-section of cryptocurrency returns, offering information over and above other existing factor models in the literature. In sum, our findings have important implications for investors in the cryptocurrency market, and highlight the importance of considering fundamental sentiment factors in their investment decisions.

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*.

Bali, T. G., S. J. Brown, and Y. Tang (2017). Is economic uncertainty priced in the cross-section of stock returns? *Journal of Financial Economics 126*(3), 471–489.

Barroso, P., F. Kho, F. Rouxelin, and L. Yang (2018). Do external imbalances matter in explaining the cross-section of currency excess returns? *Available at SSRN 3232396*.

Bhambhwani, S., S. Delikouras, and G. M. Korniotis (2021). Blockchain characteristics and the cross-section of cryptocurrency returns. *Available at SSRN 3342842*.

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.

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.

Chava, S., W. Du, and B. Malakar (2021). Do managers walk the talk on environmental and social issues? Technical Report 3900814, Georgia Tech Scheller College of Business Research Paper.

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*.

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

Gorodnichenko, Y., T. Pham, and O. Talavera (2023, February). The voice of monetary policy. *American Economic Review 113*(2), 548–84.

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.

Harris, Z. S. (1954). Distributional structure. *Word 10*(2-3), 146–162.

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.

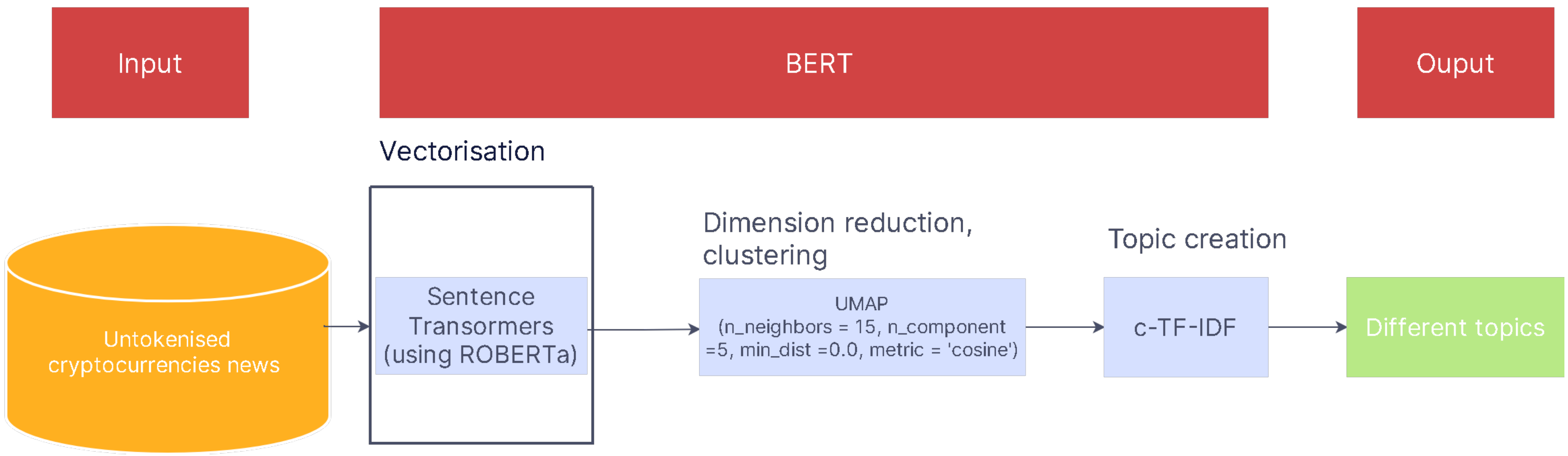
Perone, C. S., R. Silveira, and T. S. Paula (2018). Evaluation of sentence embeddings in downstream and linguistic probing tasks. Technical report, arXiv:1806.06259.

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 modelling

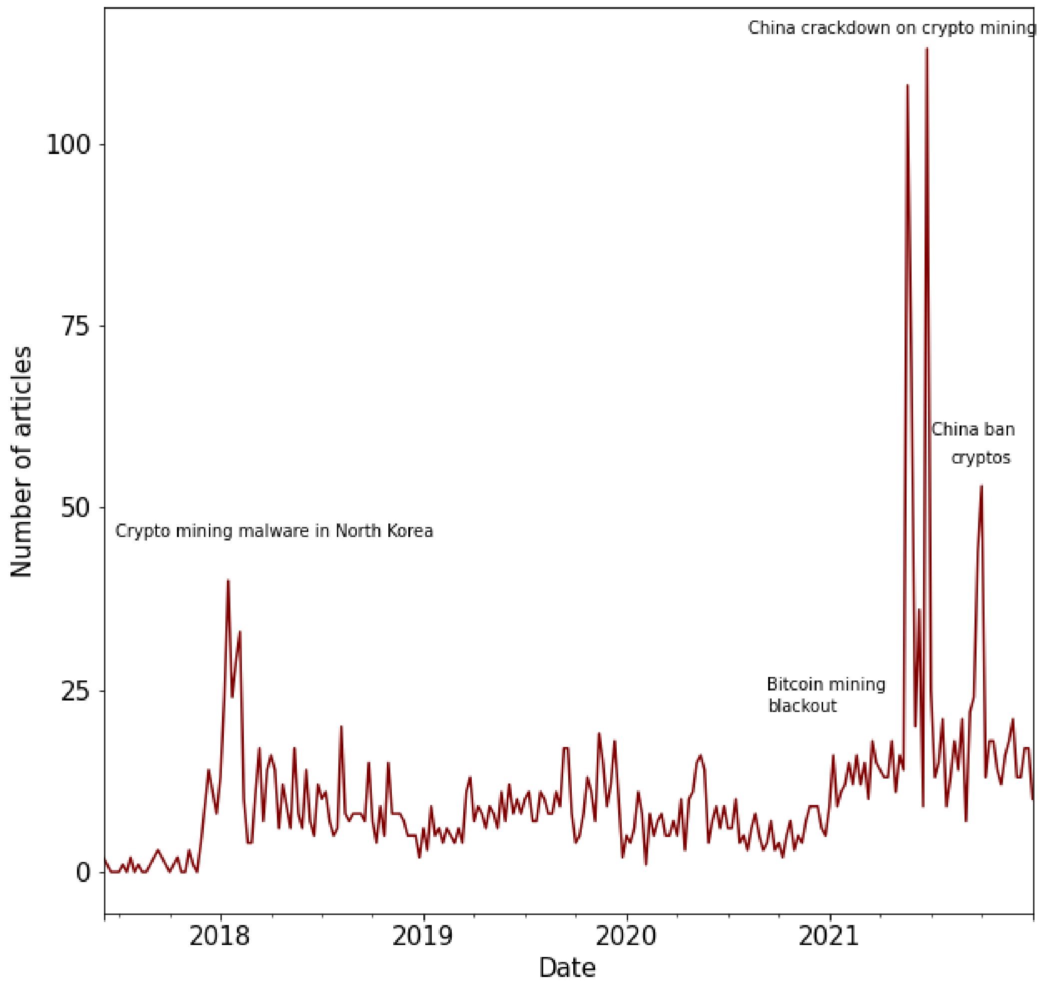


This graph shows a summary of the BERT algorithm.

Figure 2. Fundamental Topic generated from BERT topic modeling

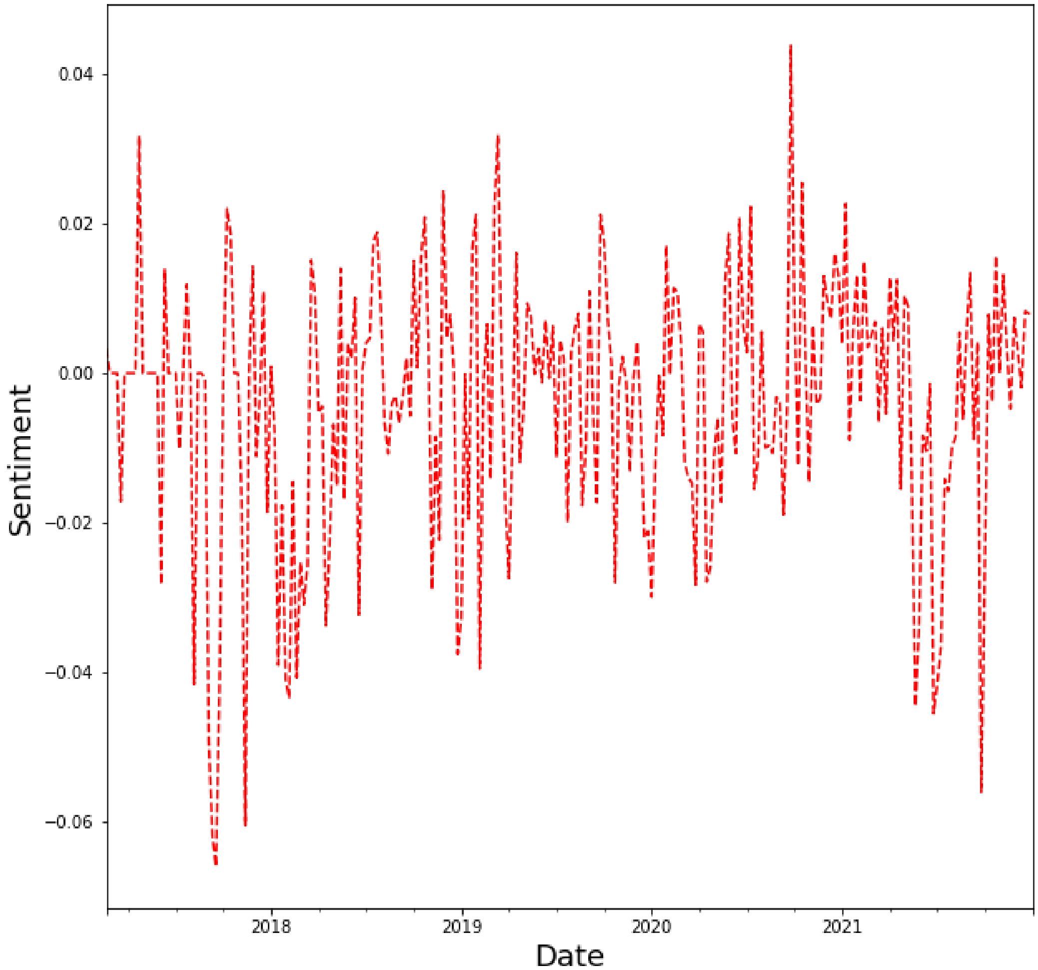
This graph shows keywords for the Fundamental topic. The data is weekly between June 2017 and December 2021.

Figure 3. Fundamental News Articles



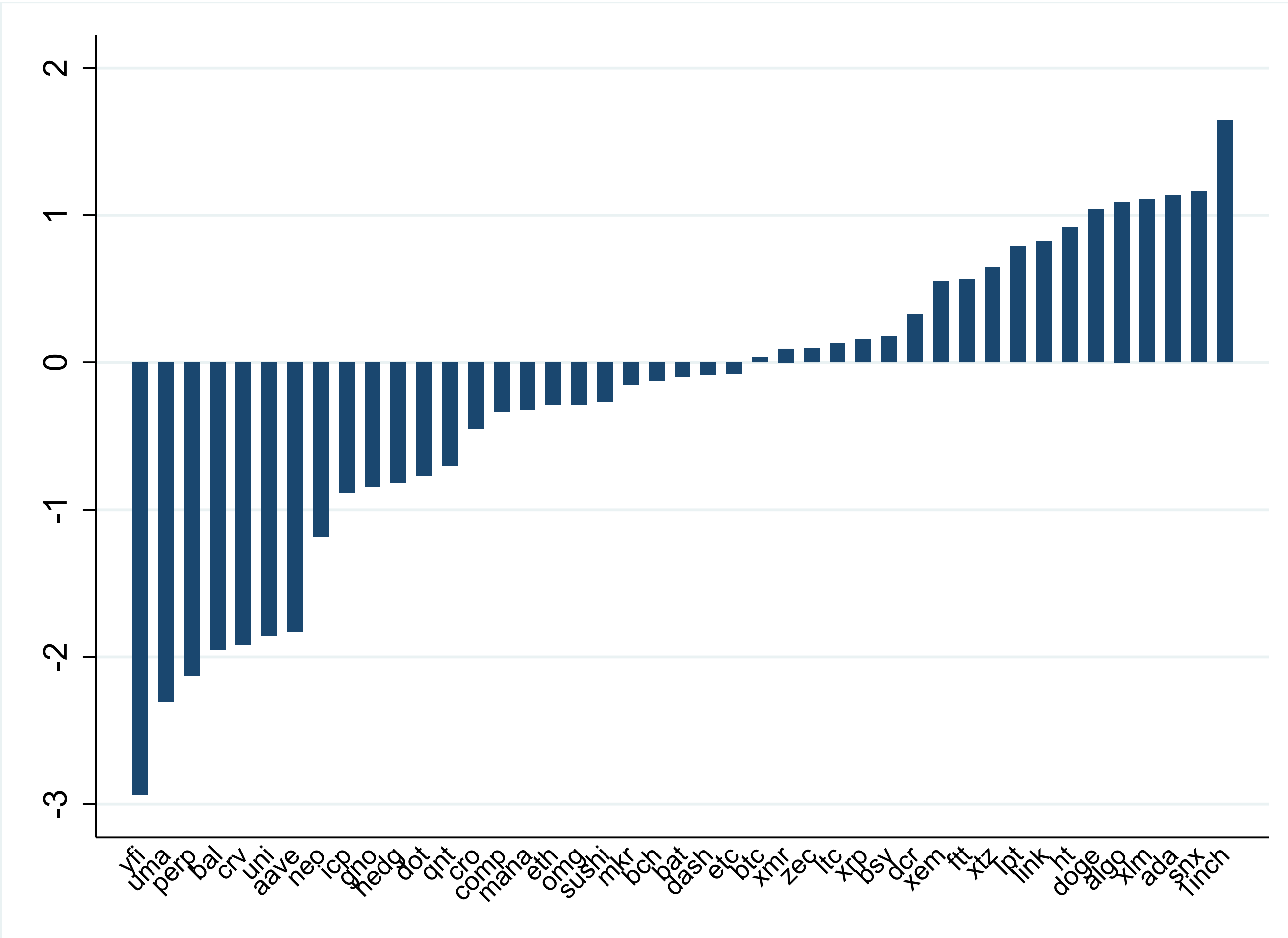
This graph shows the Fundamental News Articles. The data is weekly between June 2017 and December 2021.

Figure 4. Fundamental Sentiment Index



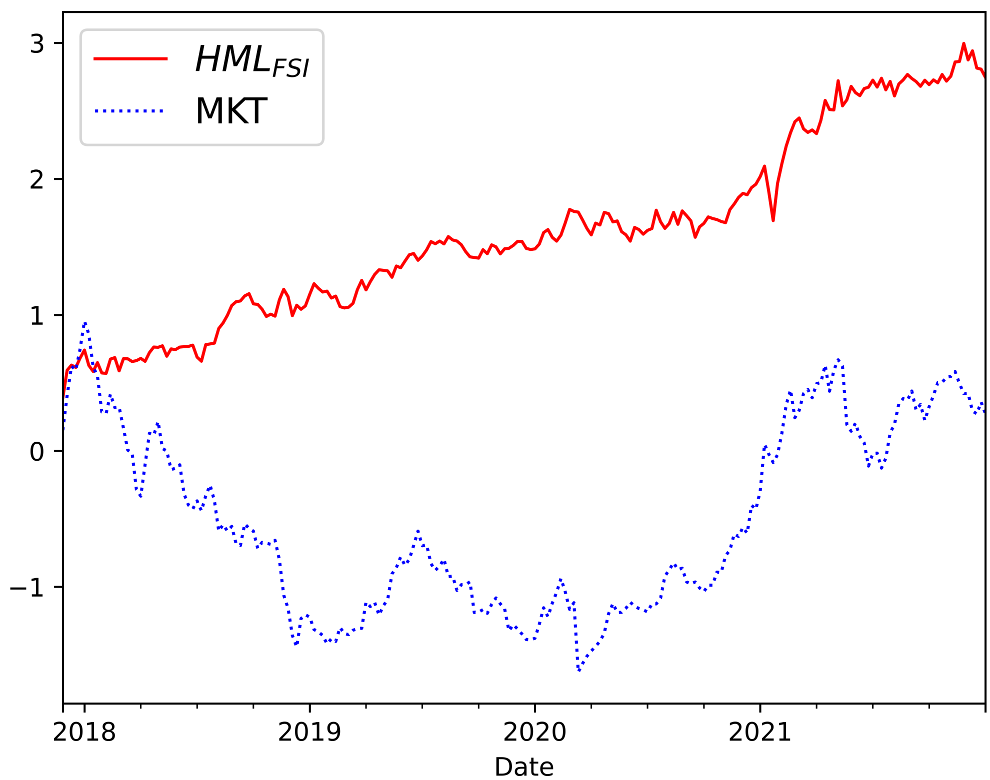
This graph shows the Fundamental Sentiment Index. The data is weekly between June 2017 and December 2021.

Figure 5. Average FSI beta by cryptocurrency



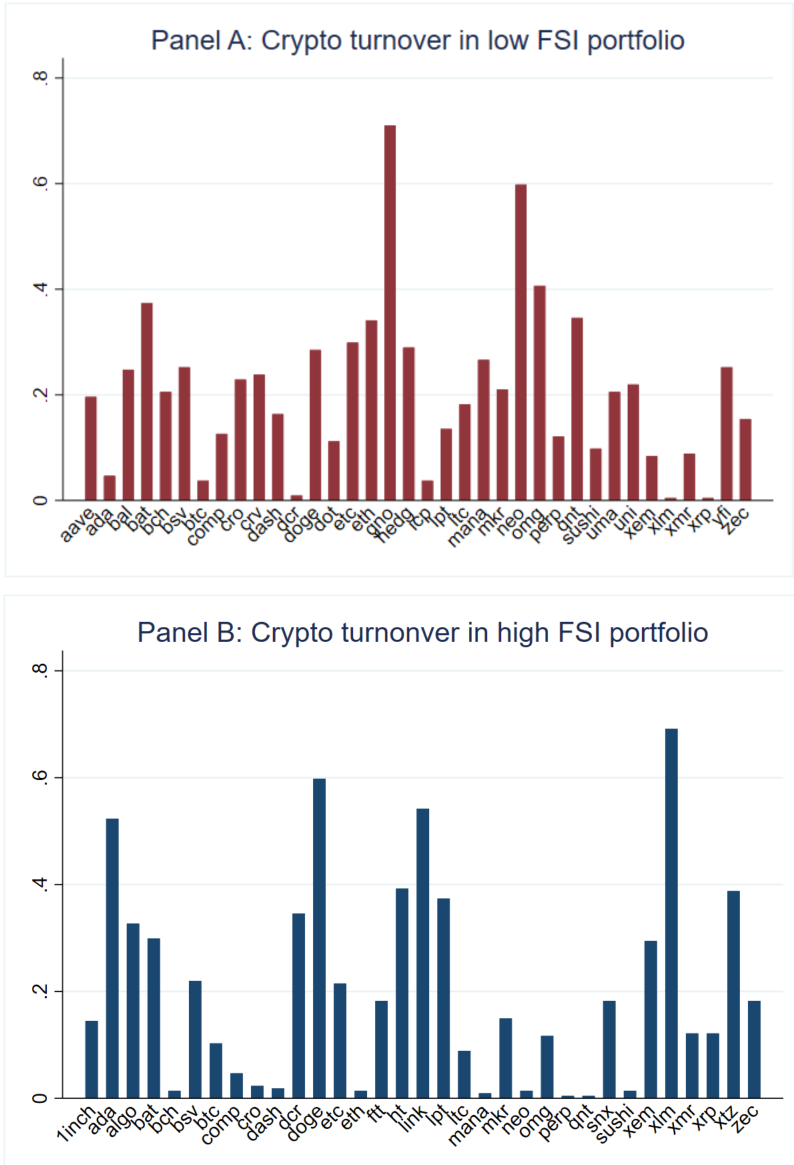
This graph shows the average FSI beta by cryptocurrency. The data is weekly between June 2017 and December 2021.

Figure 6. Cumulative returns of Fundamental Sentiment Index Strategy



This graph shows the cumulative returns of the Fundamental Sentiment Index (HML_{FSI}), and market portfolio (MKT). The data is weekly between June 2017 and December 2021.

Figure 7. FSI Portfolio Turnover



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

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 model below:

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

where $\hat{\beta}^{FSI}_{i,t}$ denotes the 60-week rolling betas with the FSI index. $Value$ represents different measures of cryptocurrency value Cong et al. (2021). 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.

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) n = 1	(2) n = 2	(3) n = 3	(4) n =4	(5) n = 5	(6) n = 6	(7) n = 7	(8) n = 8	(9) n = 9	(10) n = 10	(11) n = 11	(12) n = 12
β^{FSI}_t	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^{MKT}_t	-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. Asset Pricing Tests

This table reports regressions results for the two-factor model, including the MKT and FSI risk factors. Panel A uses as test assets 4 FSI portfolios. Portfolios are rebalanced weekly. Panel B shows results for test assets that include four size portfolios, four momentum portfolios, four liquidity portfolios, four volatility portfolios, four value portfolios, and four FSI portfolios. Newey and West (1987) t-statistics 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: Four FSI portfolios				
	λ_{MKT}	λ_{FSI}	RMSE	R^2
FMB	-0.007	0.013***	0.0017	0.81
(NW)	[-0.12]	[2.90]		
(SH)	[-0.12]	[2.53]		
Panel B: 24 Test Assets				
	λ_{MKT}	λ_{FSI}	RMSE	R^2
FMB	-0.002	0.010*	0.005	0.15
(NW)	[-0.23]	[1.86]		
(SH)	[-0.26]	[1.82]		

Table 9. Adding Fundamental news to existing asset pricing models

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 10. 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 11. 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:

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

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

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

Table 12. 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						Panel B: Regulation Sentiment Portfolio					
	P_1	P_2	P_3	P_4	HML		P_1	P_2	P_3	P_4	HML
Mean	-0.14	0.34	-0.12	0.00	0.14	Mean	-0.19	0.19	0.02	0.09	0.28
					[0.19]						[1.20]
Std	1.00	0.99	0.89	0.97	0.46	Std	0.98	0.99	0.92	1.01	0.56
SR					0.31	SR					0.49
Panel C: Payment Sentiment Portfolio						**Panel D: Derivatives Sentiment Portfolio**					
	P_1	P_2	P_3	P_4	HML		P_1	P_2	P_3	P_4	HML
Mean	-0.09	-0.25	0.41	0.03	0.12	Mean	-0.09	-0.019	0.20	0.18	0.26
					[1.05]						[0.93]
Std	0.95	0.95	1.00	0.96	0.45	Std	0.96	0.99	0.96	0.97	0.49
SR					0.28	SR					0.54
Panel E: Social Media Sentiment Portfolio						**Panel F: Hedging Sentiment Portfolio**					
	P_1	P_2	P_3	P_4	HML		P_1	P_2	P_3	P_4	HML
Mean	-0.22	0.16	-0.02	0.20	0.42	Mean	-0.02	-0.13	0.24	-0.05	0.01
					[1.31]						[0.05]
Std	0.97	0.98	0.92	1.00	0.52	Std	1.00	0.93	0.98	0.96	0.51
SR					0.80	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)

Appendix 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 (16) subject to the budget constraints in Equations (17) and (18).

$$\max_{q_t, s_t, \hat{q}_t} \qquad u(c_t^y) + \beta \mathbb{E}_t[u(c_{t+1}^o)] \tag{16}$$

subject to:

$$c_t^y = e_t - s_t - q_t p_t - \hat{q}_t \hat{p}_t - \psi_t q_t p_t, \tag{17}$$

$$c_{t+1}^o = s_t(1 + r_t) + q_t p_{t+1} + \hat{q}_t \hat{p}_{t+1} + \theta_{t+1} q_t p_{t+1}. \tag{18}$$

First order conditions:

$$-p_t - \psi_t p_t u'(c_t^y) + \beta \mathbb{E}_t u'(c_{t+1}^o)(p_{t+1} + \theta_{t+1} p_{t+1}) = 0 \tag{19}$$

$$-u'(c_t^y) + (1 + r_t)\beta \mathbb{E}_t u'(c_{t+1}^o) = 0 \tag{20}$$

$$-\hat{p}_t u'(c_t) + \beta \mathbb{E}_t u'(c_{t+1})\hat{p}_{t+1} = 0 \tag{21}$$

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.

$$p_t = \beta \mathbb{E}_t \left[\frac{u'(c_{t+1}^o)}{u'(c_t^y)} \frac{1 + \theta_{t+1}}{1 + \psi_t} p_{t+1} \right] \tag{22}$$

$$\beta = \frac{1}{1 + r_t} \frac{u'(c_t^y)}{\mathbb{E}_t u'(c_{t+1}^o)} \tag{23}$$

$$\hat{p}_t = \beta \mathbb{E}_t \left[\frac{u'(c_{t+1}^o)}{u'(c_t^y)} \hat{p}_{t+1} \right] \tag{24}$$

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

Appendix B: Examples of Fundamental Analysis Articles

2.1 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 Ã¢â‚¬Å“the internet of blockchain. A Growing Platform 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 MATICÃ 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."

2.2 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

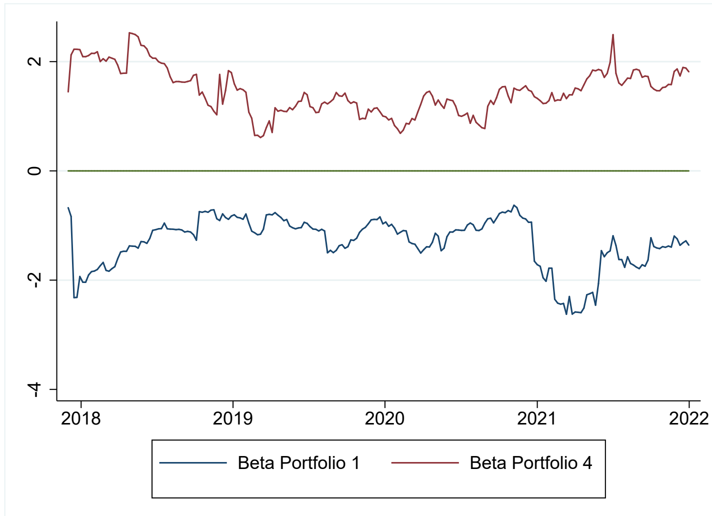
"*ethereum gas fees have exploded in 2021, which has been a hindrance to both inexpensive nfts, and also defi platforms that were designed to deal with small amounts of value.*" (Sentiment Score -0.09)

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)

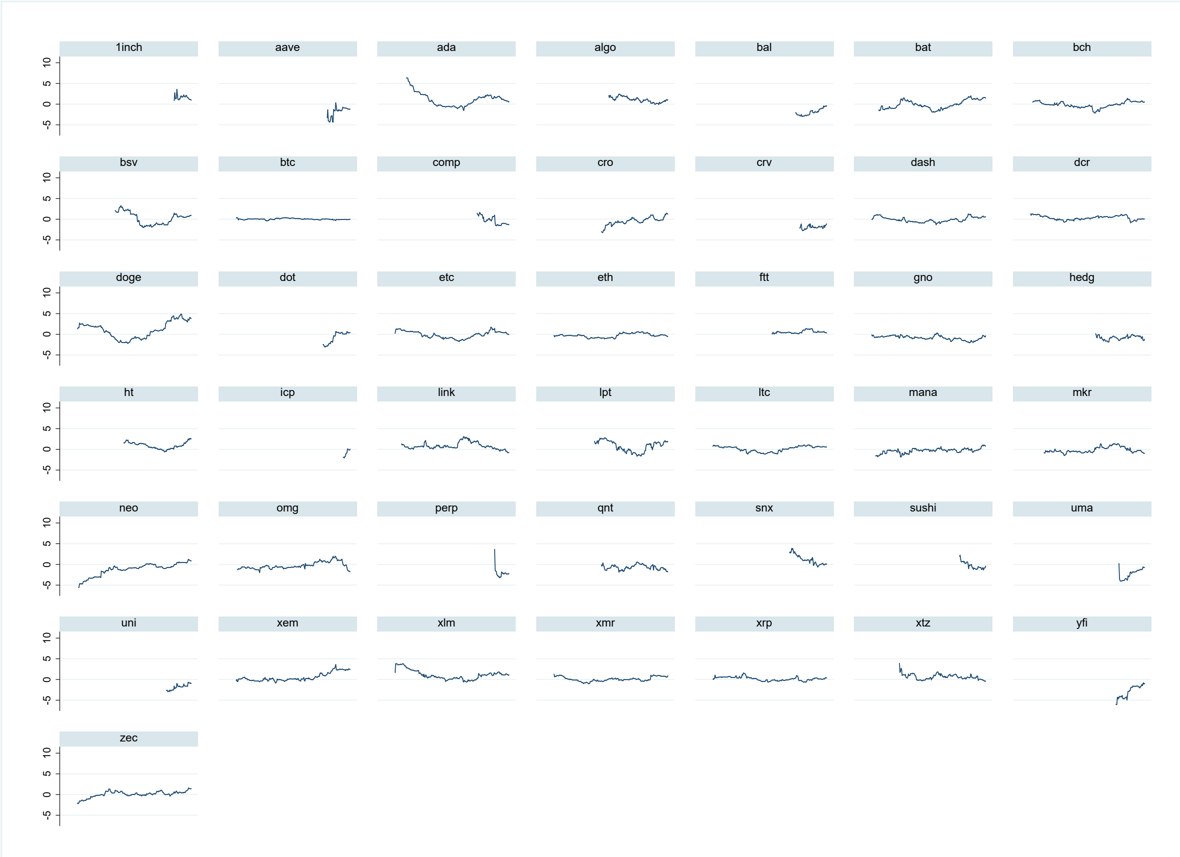
Appendix C: Additional Figures

Figure A1. Rolling beta for Portfolio 1 and Portfolio 4



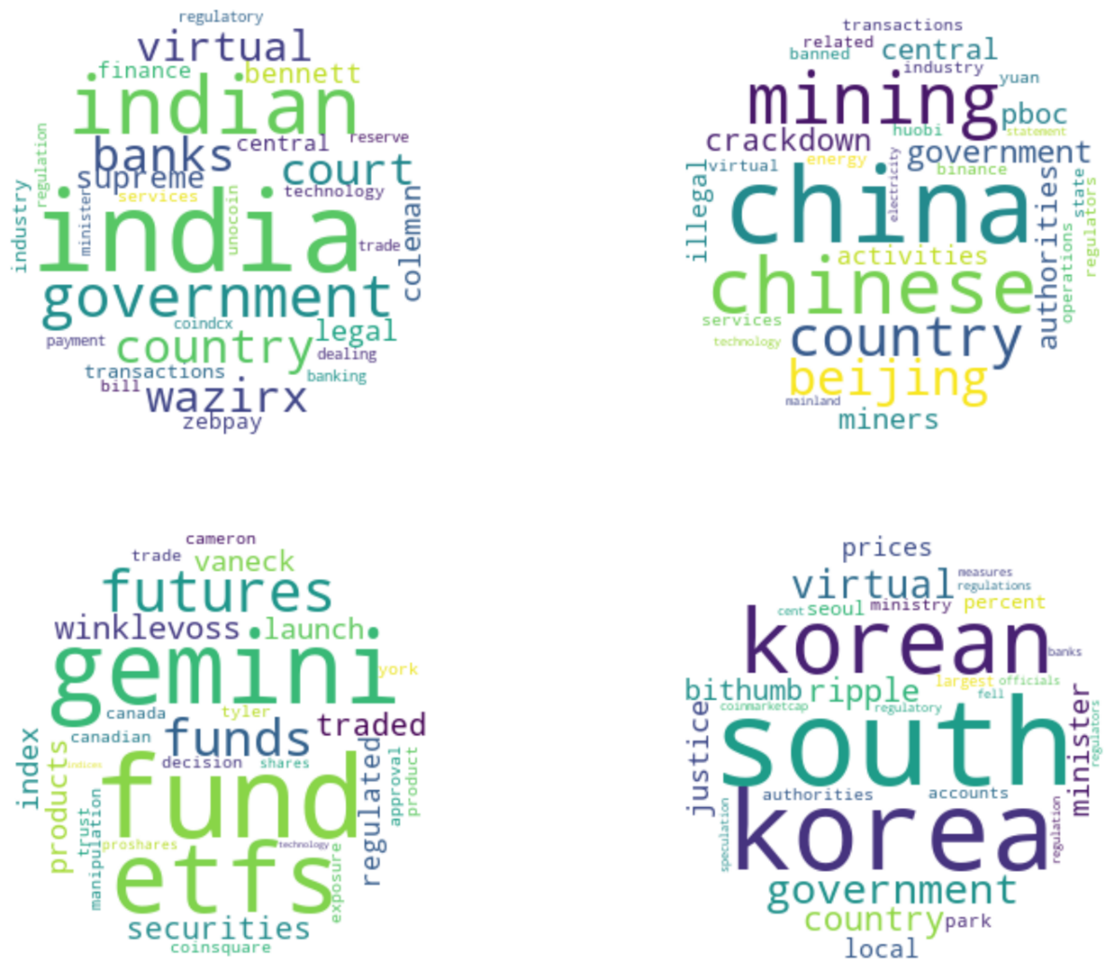
This graph shows rolling betas for Portfolio 1 and Portfolio 4. The data is weekly between June 2017 and December 2021.

Figure A2. Rolling beta by cryptocurrency



This graph shows rolling betas by cryptocurrency. The data is weekly between June 2017 and December 2021.

Figure A3. Regulation Topics generated from BERT topic modelling



The figure shows keywords for 4 Regulation topics. The data are between June 2017 and December 2021.

Figure A4. Lending Topic generated from BERT topic modeling

This graph shows keywords for the Lending topic. The data is weekly between June 2017 and December 2021.

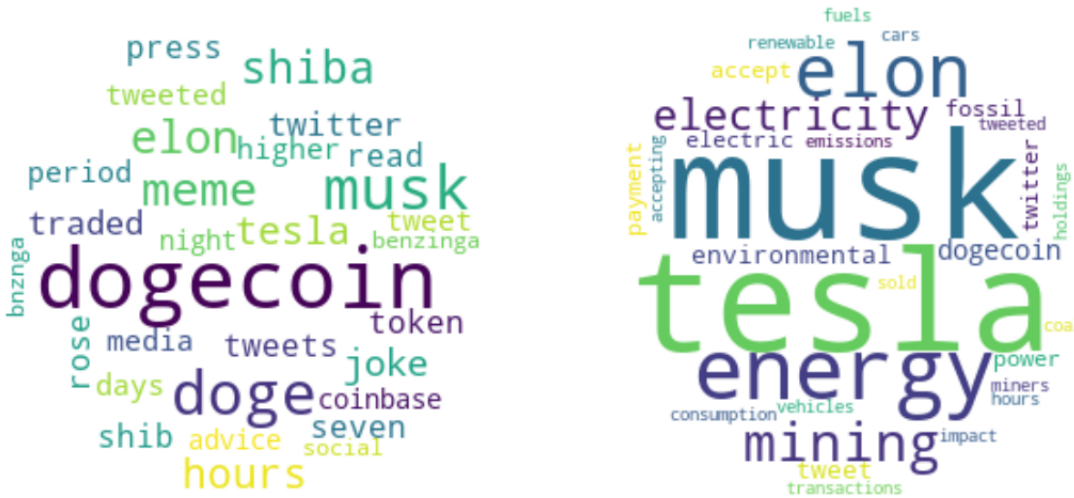
Figure A5. Payment Topic generated from BERT topic modeling

This graph shows keywords for the Payment topic. The data is weekly between June 2017 and December 2021.

Figure A6. Technical Derivatives Topics generated from BERT topic modelling

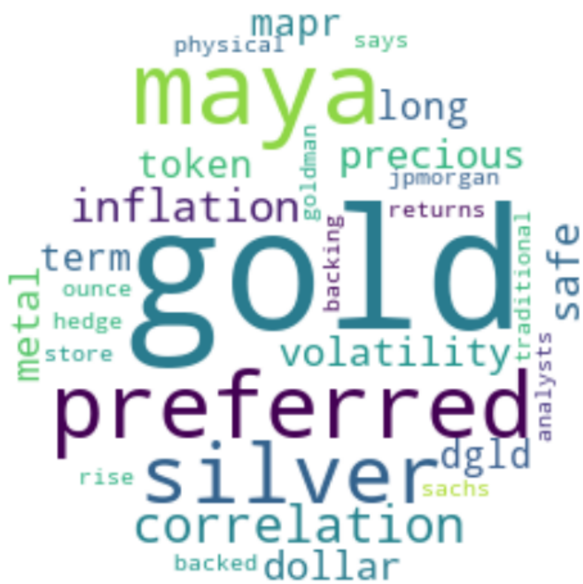
The figure shows keywords for 2 Technical Derivatives topics. The data are between June 2017 and December 2021.

Figure A7. Social Media Topics generated from BERT topic modeling



The figure shows keywords for 2 Social Media topics. The data are between June 2017 and December 2021.

Figure A8. Hedging Topic generated from BERT topic modeling



This graph shows keywords for the Hedging topic. The data is weekly between June 2017 and December 2021.

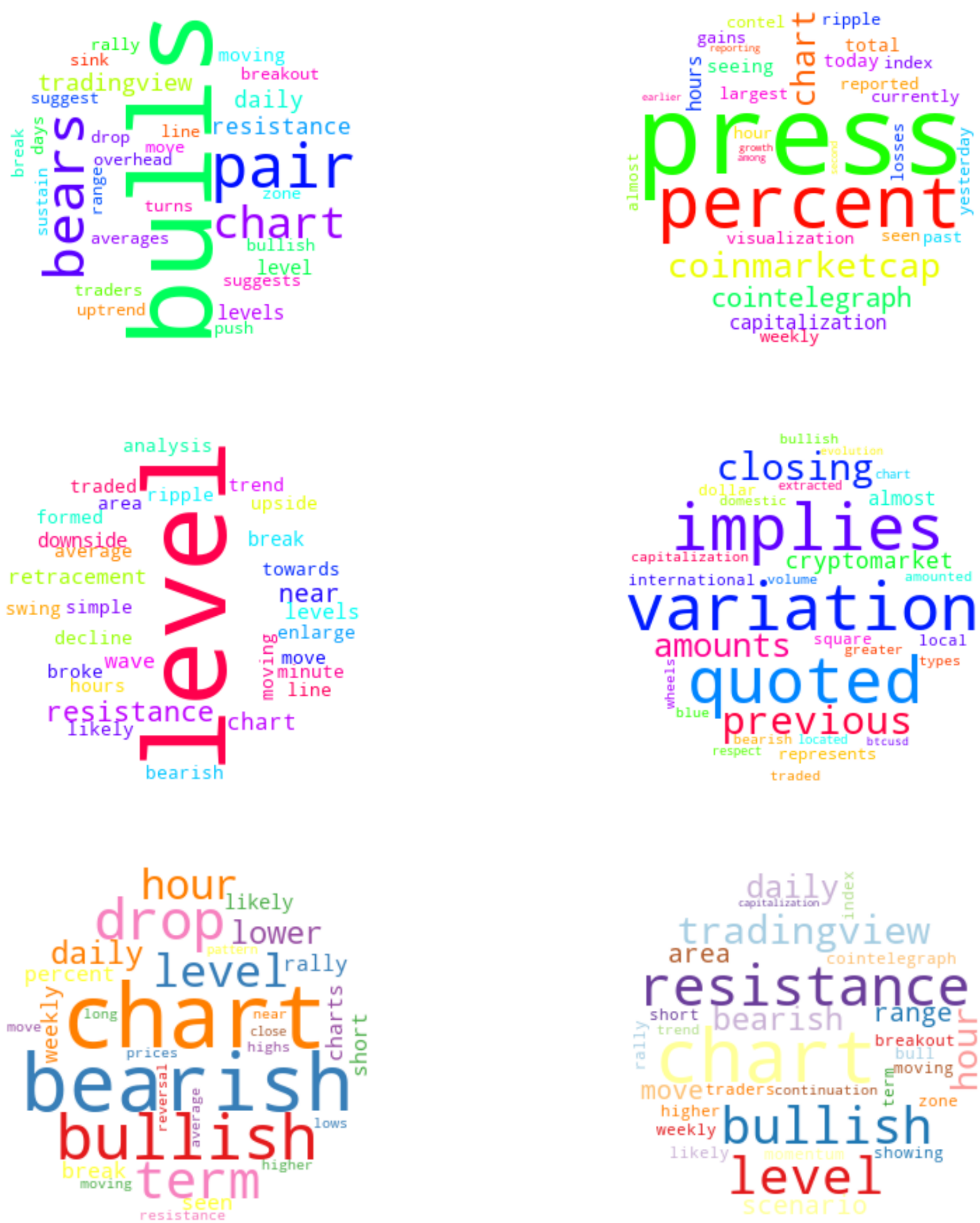
Appendix D: Additional Tables

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.

Full Sample					
Year	Number of coins	Total Market capitalization	Sample to total cryptocurrency market capitalization	volatility	Number of accounts
2017	20	661	0.87	0.91	73957.8
2018	25	145	0.78	0.91	64978.62
2019	30	195	0.83	0.91	58252.67
2020	40	654	0.91	0.91	58126.36
2021	43	1,750	0.82	0.91	71446.09

Figure A9. Technical Topics generated from BERT topic modelling



The figure shows keywords for 6 Technical topics. The topics are generated from BERT topic modelling algorithm based on Factiva news articles about cryptocurrencies between June 2017 and December 2021.

Table A2. Variable descriptions

This table reports descriptions of variables used in the paper.

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

Table A3. Token Classification

Ticker	Token Type	Description
yfi	governance	Yearn Finance is a decentralized finance (DeFi) platform.
uma	governance	UMA is a decentralized financial contracts platform.
perp	governance	Perpetual Protocol provides decentralized perpetual contracts for crypto assets.
bal	governance	Balancer is an automated portfolio manager and liquidity protocol.
crv	governance	Curve is a decentralized exchange optimized for stablecoins.
uni	governance	Uniswap is a decentralized trading protocol and liquidity provider.
aave	governance	Aave is a lending and borrowing protocol in the DeFi space.
neo	governance	NEO is a blockchain platform for the development of digital assets and smart contracts.
icp	governance	Internet Computer is a blockchain-based computing platform.
gno	governance	Gnosis is a platform for prediction markets and decentralized applications.
hed	governance	Hedera Hashgraph is a public distributed ledger technology.
gdot	platform	Polkadot is a multi-chain blockchain platform enabling interoperability.
qnt	platform	Quant Network aims to enable seamless blockchain interoperability.
cro	product	Crypto.com offers various crypto-related products and services.
comp	governance	Compound is a decentralized lending and borrowing protocol.
mana	product	Decentraland is a virtual reality platform powered by the Ethereum blockchain.
eth	platform	Ethereum is a decentralized platform for building applications and smart contracts.
omg	platform	OMG Network aims to facilitate fast and scalable payments.

Table A3. Token Classification (continued)

Ticker	Token Type	Description
sushi	governance	SushiSwap is a decentralized exchange and AMM protocol.
mkr	governance	MakerDAO is a decentralized organization behind the DAI stablecoin.
bch	general payment	Bitcoin Cash is a peer-to-peer electronic cash system.
bat	product	Basic Attention Token is used to reward content creators and users.
dash	platform	Dash focuses on fast and private digital transactions.
etc	platform	Ethereum Classic is a continuation of the original Ethereum blockchain.
btc	general payment	Bitcoin is a decentralized digital currency.
xmr	general payment	Monero focuses on private and untraceable transactions.
zec	platform	Zcash aims to provide enhanced privacy features for transactions.
ltc	general payment	Litecoin is a peer-to-peer cryptocurrency.
xrp	platform	XRP is the native cryptocurrency of the Ripple network.
bsv	general payment	Bitcoin SV aims to scale Bitcoin for large blocks and fast transactions.
dcr	general payment	Decred is a cryptocurrency with a strong focus on governance.
xem	platform	NEM is a blockchain platform for managing assets and data.
ftt	platform	FTX Token is associated with the FTX cryptocurrency exchange.
xtz	platform	Tezos is a self-amending blockchain platform.
lpt	platform	Livepeer is a decentralized video transcoding network.
link	platform	Chainlink provides decentralized oracle services.

Table A3. Token Classification (continued)

Ticker	Token Type	Description
ht	platform	Huobi Token is associated with the Huobi cryptocurrency exchange.
doge	general payment	Dogecoin started as a joke but has become a popular cryptocurrency.
algo	platform	Algorand is a blockchain platform focusing on speed and security.
xlm	platform	Stellar is a platform for cross-border payments and remittances.
ada	platform	Cardano aims to provide a more secure and scalable blockchain platform.
snx	general payment	Synthetix is a platform for creating synthetic assets on Ethereum.
1inc	platform	1inch is a decentralized exchange aggregator and liquidity protocol.

Table A4. 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:

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

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

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 A5. 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 A6. 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 A7. 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 A8. 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