Sentiment Trumps Fundamentals: Machine-Learning Evidence on the Drivers of Cryptocurrency Returns

Min Zhang[1], Liangbo Zheng[2,*], Jing Chen[1]

[1] Faculty of Business, Nanjing Normal University Taizhou College, Taizhou, China 225300

[2] Faculty of Business, Information & Human Sciences (FBIHS), Kuala Lumpur University of Science & Technology (KLUST), Kajang, Malaysia 43000

* Email: zoelfank@gmail.com (Corresponding Author)

Received 20 March 2026 · Revised 4 May 2026 · Accepted 18 June 2026 · Published 30 September 2026

Abstract

Whether investor sentiment is more informative than network fundamentals for cryptocurrency returns remains an empirical question. This study evaluates that proposition using a transparent, time-ordered forecasting design for Bitcoin and Ether. Daily prices and on-chain activity are drawn from Coin Metrics community data, and the Bitcoin-oriented Crypto Fear & Greed Index is obtained from Alternative.me. The index is a composite market-mood proxy rather than a pure measure of emotion because its documented components include volatility and market momentum. Daily log returns are forecast using market-history features alone or augmented with sentiment, blockchain-activity proxies, or both. Ridge regression and histogram-based gradient boosting are specified with fixed hyperparameters. Models are developed on data from June 2018 to December 2021, assessed on 2022, and refitted through 2022 before testing on 2023–2024. After common complete-case filtering, each asset has 723 test observations. In the test period, the gradient-boosting market-only root mean squared errors are 2.546 percentage points for Bitcoin and 2.967 for Ether, compared with 2.558 and 2.967 when the sentiment block is added. The sentiment block does not consistently improve forecasting relative to the market-only or network-augmented models, and all gradient-boosting variants have negative out-of-sample R^2 relative to a zero-return forecast. Paired block-bootstrap intervals show substantial uncertainty in the incremental comparisons. The title's proposition is therefore not supported as a general conclusion by this restricted test. The results illustrate the importance of release timing, overlapping information, simple benchmarks, and reproducible validation when interpreting claims about the drivers of cryptocurrency returns.

Keywords: digital assets; blockchain activity; investor attention; return forecasting; gradient boosting; time-series validation; feature-block comparison

Sentiment Trumps Fundamentals: Machine-Learning Evidence on the Drivers of Cryptocurrency Returns

1. Introduction

Cryptocurrency prices respond to a mixture of market trading, expectations, network activity, news, and changes in the wider financial environment. Understanding these relationships is difficult because digital assets differ in design and have short, evolving histories. Bitcoin and Ether are traded continuously, while their underlying networks process transactions under distinct rules and economic mechanisms. A change in network activity may reflect genuine use, speculative trading, technical protocol changes, or several of these at once. Similarly, a measure labeled investor sentiment can partly reflect past returns rather than an independent assessment of market beliefs. These characteristics make it risky to interpret any single predictor as the fundamental cause of a return movement (Kondor et al., 2021; Koutmos, 2018; Liu & Tsyvinski, 2021; Meister & Price, 2024).

The distinction between sentiment and fundamentals is especially contested in markets without a common valuation model. Traditional equity valuation often refers to claims on future cash flows, but cryptocurrency networks do not all provide the same form of economic claim. For some research questions, active addresses, transactions, and network fees are useful indicators of adoption or usage; they do not automatically measure an asset's intrinsic value. The broader literature investigates cryptocurrency risk, cross-sectional returns, and relationships between blockchain activity and prices, but the mechanisms and statistical designs vary (Borri & Shakhnov, 2022; Koutmos, 2018; Liu & Tsyvinski, 2021; Liu et al., 2025). A careful comparison must therefore define what its "fundamental" variables actually measure.

Investor sentiment is also multidimensional. Social-media tone, search intensity, speculative attention, and fear–greed indicators capture different phenomena. Some indicators combine price changes and volatility with social data, creating overlap with market-history variables. A forecasting study that treats such an index as an independent sentiment signal can overstate the distinction between behavioral and market information. Prior research has studied social-media indicators, attention, fear and greed, and their relationships with cryptocurrency returns and volatility, with results dependent on the asset, sample, horizon, and measurement approach (Cavalheiro et al., 2024; Dias et al., 2022; Gaies et al., 2023; Guégan & Renault, 2021).

Machine learning can represent nonlinear relationships and interactions that are not imposed in a simple linear regression. Its flexibility also creates opportunities for overfitting. Cryptocurrency studies have evaluated machine-learning methods using different data, targets, and validation protocols (Berger & Koubová, 2024; Chen et al., 2021; Murray et al., 2023; Ranjan et al., 2023). A model that fits price levels well may exploit persistence without accurately predicting returns. A model selected after repeated inspection of a single test period can appear useful by chance. Data leakage occurs when information unavailable at the forecast origin enters features, transformations, or model selection. All of these problems complicate claims that one class of information dominates another (Arian et al., 2024; Bailey et al., 2016; Yae & Luo, 2023).

This article tests the proposition expressed in its title: do sentiment-oriented features forecast cryptocurrency returns more effectively than observable network-activity features once past market information is accounted for? The analysis is explicitly predictive and does not identify causal effects. It uses publicly available daily observations for Bitcoin and Ether, a Bitcoin-oriented market-mood index from Alternative.me, and Coin Metrics data for price, active addresses, transaction counts, and native-unit transaction fees. We compare a market-history baseline with models that add the sentiment proxy, on-chain proxies, or both. A linear ridge model provides a regularized benchmark; a histogram-based gradient-boosting model represents a nonlinear machine-learning approach.

The research design fixes the asset universe, time windows, feature blocks, two model families, and core hyperparameters. Models are trained on June 2018–December 2021, examined on the following year, and

refitted using information through 2022 before the 2023–2024 test. All predictors use observations dated at least two days before the target return date. A paired moving-block bootstrap summarizes uncertainty in test-period loss differences, and a three-day-lag sensitivity is evaluated on a common set of dates. Because the public data were downloaded retrospectively and the precise historical release and revision times are not archived here, the lag reduces but does not eliminate point-in-time uncertainty.

The test-period evidence does not support a general statement that sentiment trumps fundamentals. In the primary gradient-boosting comparison, adding the sentiment block worsens the Bitcoin test root mean squared error relative to the market baseline and makes almost no difference for Ether. Adding network features changes the results differently across assets. All gradient-boosting specifications have negative out-of-sample R^2 relative to a zero-return forecast. Some validation-period results favor the sentiment-augmented model, illustrating why a single favorable period is not enough to establish stable superiority. These outcomes are reported rather than adjusted to fit the wording of the title.

The study offers four contributions. First, it turns a broad dominance claim into explicit comparisons against market-only, network-augmented, and zero-return benchmarks. Second, it documents the composition and timing limitations of a widely used fear–greed proxy. Third, it supplies two separate asset analyses with a common test window, fixed feature engineering, and reproducible code and source snapshots. Fourth, it reports uncertainty and sensitivity results alongside point estimates, without equating feature importance or predictive association with causal influence. The paper proceeds through related literature, data and methods, empirical results, discussion, and conclusion.

2. Literature Review

2.1 Cryptocurrency returns and economic information

Cryptocurrency returns have been studied through time-series and cross-sectional asset-pricing frameworks. Research on risks and returns of cryptocurrency examines the role of asset-specific predictors and attention relative to traditional financial factors (Liu & Tsyvinski, 2021). Other studies investigate consumption-based pricing, alternative empirical factor models, and the cross-section of digital-asset returns (Borri & Shakhnov, 2022; Dunbar & Owusu-Amoako, 2022; Hwang & Kwon, 2024). Such approaches address different estimands: a cross-sectional strategy that ranks many assets is not the same as predicting the next daily return of one asset. Their findings cannot be assumed to apply unchanged to a two-asset, daily forecasting exercise.

Returns can also contain large jumps and changes in volatility. Studies of cryptocurrency jump risks, the return–volatility relationship, and return spillovers highlight the need to consider tail behavior and market states (Bouri et al., 2017; Jia et al., 2023; Leong & Kwok, 2023). These features matter for statistical evaluation because a small number of extreme days can dominate squared-error metrics. A model can have a lower average loss during one period and fail during another, even if its feature definitions remain unchanged. Mean absolute error and block-resampling intervals complement root mean squared error but do not remove the difficulty of evaluating a volatile, changing market.

Network information has been proposed as a way to characterize use and security in digital-asset systems. Transaction activity, active addresses, and fee patterns can capture activity on a ledger, but their interpretation depends on the protocol and the period (Kondor et al., 2021; Koutmos, 2018; Meister & Price, 2024). One entity can control many addresses, a single address can serve many users, and exchange aggregation can change the relationship between addresses and people. Changes to Ethereum's protocol and fee mechanism further complicate comparisons across time. These observations caution against treating on-chain proxies as a complete or invariant measure of value.

Our "network fundamentals" block is deliberately narrow. It contains transformations of three observable network-activity series and does not include protocol development, miner or validator economics, asset

supply, staking yields, transaction value, macroeconomic conditions, or regulatory developments. A model using these three series may perform poorly even if other fundamental information is important. The expression "network proxies" is therefore used throughout to avoid equating the operational block with all possible fundamentals.

2.2 Sentiment, attention, and market mood

Behavioral research has investigated how investor sentiment, social-media activity, and media coverage relate to cryptocurrency returns. Some studies analyze whether social-media measures contain information for Bitcoin, while others examine asymmetric responses, pandemic-period sentiment, and price jumps (Bouteska et al., 2022; Guégan & Renault, 2021; He & Wang, 2023; Naifar & Altamimi, 2023). Cross-sectional studies consider whether sentiment is related to differences among digital assets (Han, 2025). These papers establish that sentiment is a plausible research topic, not that one proxy must outperform every alternative in every forecasting window.

Attention and sentiment are related but distinct. A search for Bitcoin may indicate curiosity, concern, or investment intent; the count alone does not specify which. Research on attention, quantile dependence, and liquidity uses search activity and social-media measures to investigate market relationships (Choi, 2021; Su et al., 2021; Subramaniam & Chakraborty, 2020; Zhu et al., 2021). Text sentiment models add linguistic classification but introduce their own problems, including sarcasm, bots, changing platform composition, non-English content, and uncertain historical access. Forecasting performance may depend on how those problems are handled and whether information is available before the prediction is made (Amirshahi & Lahmiri, 2025; Koltun & Yamshchikov, 2023; Li & Ma, 2024).

Fear–greed indices combine several inputs into an interpretable score. Research examines the relation between these scores and returns, return distributions, and volatility (Cavalheiro et al., 2024; Dias et al., 2022; Gaies et al., 2023; Gessie, 2026; Szeberényi & Kovács, 2026). The Alternative.me index used in this study is explicitly Bitcoin-oriented. Its provider describes a composite including Bitcoin price volatility and momentum or volume, along with social and other indicators. The provider also notes that the survey component is paused. These details refer to the provider's current documentation; this study does not have a complete archive of historical weights or revisions. Consequently, the index should be described as a Bitcoin-centered market-mood proxy rather than an independent survey of sentiment or a universal sentiment measure for Ether.

The direction of association is not guaranteed. Returns can influence a fear–greed index because price behavior contributes to the index, and the index may be associated with later returns under some conditions. These are different statements. A predictive relationship in one direction does not establish a causal mechanism, and a sentiment–return association can change across horizons or market states (Biswas & Sharma, 2025; Gaies et al., 2023; Gessie, 2026). The present design uses dated lags to limit contemporaneous overlap with the target, but it does not resolve endogeneity or identify the effect of an exogenous change in beliefs.

2.3 Network activity and market structure

Bitcoin and Ethereum differ in consensus, transaction design, fee structures, and applications. Comparing the same named metric across the two systems does not ensure that it has an identical economic meaning. Ethereum transaction fees, for example, are affected by gas consumption and protocol rules, whereas Bitcoin fees are linked to a different transaction and block-space mechanism (Meister & Price, 2024). Network-link patterns can also vary across blockchains (Kondor et al., 2021). We therefore estimate separate models by asset rather than pooling the raw series under a common coefficient.

Network activity can rise during periods of price appreciation or heightened volatility, making a causal interpretation particularly difficult. Transaction volume may reflect user adoption, exchange transfers,

internal wallet movements, or speculative repositioning. Active-address counts can be influenced by address reuse and custodial arrangements. Research has explored the relationship between Bitcoin transaction activity and returns and the behavior of active-address measures in particular market conditions (Koutmos, 2018; Liu et al., 2025). These relationships motivate a forecasting block, but they do not turn each variable into a direct measure of fundamental value.

Cryptocurrency markets can also transmit information and risk across assets. Cross-asset price relationships, high-frequency spillovers, sentiment spillovers, and cyber events are studied in separate literatures (Caporale et al., 2021; Chen & Hu, 2026; Jia et al., 2023; Leong & Kwok, 2025). A Bitcoin-oriented mood index may therefore contain information relevant to Ether, but it can also simply proxy for a common Bitcoin market factor. The present design does not include a separate Ether-specific sentiment index or a complete set of cross-asset returns. That limitation is central to interpreting any comparison between the two assets.

2.4 Forecasting methods and validation

Machine-learning studies use a wide range of targets, features, and algorithms. Some forecast price levels; others forecast returns, volatility, or directional movements. A model can perform well on a highly persistent price series without providing an economically meaningful forecast of a return. Research comparing cryptocurrency forecasting approaches underscores the need to evaluate the target actually relevant to the claim (Berger & Koubová, 2024; Chen et al., 2020; Chen et al., 2021; Murray et al., 2023; Ranjan et al., 2023). In this study, the target is a daily log return expressed in percentage points, and the main performance criterion is held-out squared prediction error.

Ridge regression is included as a regularized linear comparison. Histogram-based gradient boosting can model nonlinearities and feature interactions, but it has more flexibility and can fit noise. Neither model is presumed superior before testing. Research in asset pricing and machine learning demonstrates the potential of flexible predictors while emphasizing the importance of out-of-sample evaluation and the appropriate economic interpretation of the target (Borup, 2019; Byun et al., 2024; Gu et al., 2020; Hansen, 2022). A forecast that reduces a statistical loss by a small amount may still be unhelpful after execution costs or under a different decision objective.

Temporal validation differs from random cross-validation because observations and features are ordered in time. Randomly mixing future and past observations can leak information through training, scaling, model selection, or overlapping windows. Research on time-series cross-validation and out-of-sample evaluation offers a basis for preserving chronology (Deng, 2023; Panagiotidis, 2010). We use a fixed chronological validation and test design, train transformations only on the relevant training sample, and apply the same test dates across feature blocks. This does not guarantee generalization beyond the chosen period.

Repeated testing can produce favorable results by chance. Work on data snooping, multiple comparisons, and backtest overfitting cautions against interpreting the best result from a large search as if it were prespecified (Arian et al., 2024; Bailey et al., 2016; Neuhierl & Schlusche, 2011; Romano & Wolf, 2005; Suhonen et al., 2017; White, 2000). The present study restricts the model set and reports every specified block rather than only the best. Nevertheless, the analysis is exploratory and was not publicly preregistered. Its intervals and rankings should be interpreted in light of the number of comparisons and the possibility that the chosen feature design is one of many plausible alternatives.

2.5 Incremental value and interpretation

The predictive contribution of a block depends on the other information already in the model. A sentiment score correlated with past volatility may appear important in isolation but add little after volatility is included. Conversely, a network proxy may contribute only when combined with another variable. Feature-importance measures can be unstable in the presence of correlated predictors, and an importance score does not by itself identify a causal effect (Altmann et al., 2010; Conn et al., 2019). We therefore focus on performance

differences between explicitly defined feature blocks, with the same market-history baseline for both sentiment and network comparisons.

Forecast-comparison methods evaluate losses rather than stories about individual coefficients. Classic and subsequent work on predictive accuracy and multiple forecast comparisons provides a useful background for testing differences between models (Diebold & Mariano, 1995; Diebold, 2015; Mariano & Preve, 2012). Our confidence intervals use a fixed-length moving-block bootstrap rather than a Diebold–Mariano test, and they are reported as exploratory intervals for the observed test sample. Alternative resampling methods, including the stationary bootstrap, use different block constructions (Kunsch, 1989; Politis & Romano, 1994). We do not claim to implement all of these methods or to obtain causal confidence intervals from forecast errors.

3. Data and Methods

3.1 Research question and scope

The empirical question is whether a sentiment-oriented block or a network-activity block reduces out-of-sample forecasting error after a common set of market-history features is included. The unit of analysis is an asset–date observation. We study Bitcoin and Ether separately because their network metrics and economic mechanisms differ. The dependent variable is the daily log return for the asset's Coin Metrics USD closing-price series, expressed in percentage points. The empirical asset universe is limited to two large cryptocurrencies and cannot represent the full cross-section of digital assets.

The analysis compares four feature sets. The first uses only lagged returns and realized volatility. The second adds the fear–greed block. The third adds the network block. The fourth adds both blocks. These are incremental prediction exercises; neither block is a complete operationalization of its theoretical category. In particular, the sentiment block contains market-derived information and the network block contains only three on-chain series. "Sentiment trumps fundamentals" is evaluated as a hypothesis about these specific representations, not as a universal comparison between all behavioral and economic drivers.

The data cover 1 June 2018 through 31 December 2024 after the required historical lags are constructed. This period contains different market conditions, protocol developments, and trading environments. It is not a random sample of cryptocurrency history, and historical performance can differ materially in other windows. Bitcoin and Ether were chosen as commonly observed assets, not because the results should be generalized to assets with different liquidity, consensus rules, data coverage, or exchange access.

3.2 Sources, provenance, and outcome definition

Price and network data are taken from the public Coin Metrics community repository, with one asset file for Bitcoin and one for Ether: https://github.com/coinmetrics/data. The fields used are PriceUSD, AdrActCnt, TxCnt, and FeeTotNtv. Coin Metrics describes PriceUSD as the USD-denominated daily fixing at 00:00 UTC on the following day. The address field counts active addresses according to the provider's definition, not distinct human users. The transaction field counts on-chain transactions, and the fee field is denominated in the asset's native units. These definitions are provider-specific and can change with protocol and data-processing practices.

The mood proxy is the Crypto Fear & Greed Index supplied by Alternative.me, downloaded from https://api.alternative.me/fng/?limit=0&format=json. The provider's description and attribution rules are at https://alternative.me/crypto/fear-and-greed-index/. Alternative.me is acknowledged here and next to the relevant tables and figure. The index is Bitcoin-oriented and ranges from 0 to 100. Its published methodology includes components based on market volatility and momentum as well as social and other information. The methodology may change; this study does not have a point-in-time archive of its historical component weights.

The source files were downloaded on 27 October 2026 and retained as dated analysis snapshots. Historical observations may have been revised before that date. The analysis uses data only through 31 December 2024, so later observations in the downloaded files do not enter feature construction or fitting for the reported period. A SHA-256 manifest, the original downloaded files, the program, and machine-readable results are supplied in the accompanying reproducibility archive. This allows the reported calculations to be rerun on the same snapshot even if a provider later changes the historical data.

For asset a and date t, let $P_{a,t}$ be the Coin Metrics PriceUSD observation. The target $r_{a,t}$ is defined by Equation 1. Using log returns focuses the question on changes rather than the persistence of price levels. Multiplication by 100 expresses results in percentage points; for example, a root mean squared error of 2.5 means approximately 2.5 percentage points of daily log-return forecast error on the squared-error scale. This quantity is not a percentage return earned by a trading strategy.

$$r_{a,t} = 100\, ln(P_{a,t} / P_{a,t-1}) \quad (1)$$

3.3 Sample construction and missing observations

The daily sources are aligned by their reported date in UTC. Duplicate dates are checked, the index is reindexed to a daily calendar, and features are constructed before the analysis window is selected. No price, sentiment, or network value is forward-filled. Observations are retained only when the outcome and every feature in the most comprehensive feature set are available. The resulting sample is therefore identical across the four feature blocks within an asset, preventing a change in the evaluated dates from being mistaken for an improvement from a feature block.

The downloaded Alternative.me history lacks a value dated 26 October 2024. Because the sentiment features include a two-day lag, a seven-day change, and a trailing seven-day average, that missing value removes eight target dates, 28 October–4 November 2024, from the main test sample. These dates are omitted for both assets and all feature sets. The remaining test observations number 723 per asset rather than the 731 calendar days in 2023–2024. This is complete-case filtering, not an estimate or interpolation of the absent index value. Other missing-data patterns in a different download or asset universe could produce a different sample.

The training interval contains 1,310 observations for each asset from 1 June 2018 through 31 December 2021. The validation interval contains 365 observations in 2022. After the models and feature definitions are fixed, training and validation observations through 31 December 2022 are combined to fit the models used for the held-out 2023–2024 test. The dates and counts are shown in Table 1. The test period is a single historical holdout; its existence does not make the overall study immune to exploratory choices or guarantee that a later period will behave similarly.

Table 1. Chronological Sample and Forecasting Windows

Window	Target dates	Bitcoin n	Ether n
Training	2018-06-01 to 2021-12-31	1,310	1,310
Validation	2022-01-01 to 2022-12-31	365	365
Test	2023-01-01 to 2024-12-31	723	723

Note. The test models are refitted on the training and validation windows. Eight test dates are omitted by complete-case filtering following the missing 2024-10-26 sentiment observation. Sentiment index: Alternative.me; price and network data: Coin Metrics.

The common-sample rule improves the comparability of block-level losses but restricts inference to dates for which the complete set of variables exists. If missingness is related to market stress or service outages, the retained sample may not be representative of all days. We do not model the missingness mechanism. A

production system would need a documented plan for delayed or unavailable data, including whether to fall back to a market-only forecast and how to evaluate that fallback without using information from the future.

3.4 Feature engineering and information timing

All feature construction is based on observations dated at least two days before the target date. For the market-history block, the inputs are returns dated t−2, t−3, and t−7, together with the standard deviation of returns over the trailing seven and 30 observations ending at t−2. These features provide a simple representation of prior market movement and volatility. They do not contain the target return $r_{a,t}$ or the immediately preceding date's return $r_{a,t-1}$. The deliberate gap is conservative with respect to daily date labels, although it cannot prove that every historical provider value was available by a particular decision time.

The sentiment block contains the fear–greed value dated t−2, its change between t−9 and t−2, and its trailing seven-day average ending at t−2. The same Bitcoin-oriented index is used for both Bitcoin and Ether; it is not presented as an Ether-specific mood series. The three transformations may be correlated with one another and with market-history features. They are therefore evaluated as a block rather than interpreted as three independent measures of emotional sentiment.

The network block starts with active addresses, transaction counts, and total native-unit fees. For each series, the program applies $\log(1 + x)$, takes the value dated t−2, and constructs two features: the change relative to t−9 and the deviation from its trailing 30-day average ending at t−2. Six network features result. The log transformation reduces scale differences and permits zero values but does not make the series economically equivalent across assets. Native fees are not converted to USD in order to avoid automatically introducing a contemporaneous price conversion into the network block.

Table 2 summarizes the variables and the timing rules. The output is a prediction of the date-t close-to-close return using features dated no later than t−2. This is a daily return forecast with a lagged information set; it should not be described as a contemporaneous explanation or as a one-day-ahead forecast made at the close of t−1. The common cutoff also means that some information potentially available closer to the target is deliberately excluded.

Table 2. Feature Blocks for the Two-Day Information Lag

Feature set	Inputs and construction	Number
Market only	Three lagged returns and two trailing volatility measures; latest input dated t−2.	5
Market + sentiment	Market features plus the Alternative.me index, its seven-day change, and seven-day mean.	8
Market + network	Market features plus seven-day changes and 30-day deviations in log-transformed active addresses, transaction counts, and native-unit fees.	11
All blocks	Market, sentiment, and network features jointly.	14

Note. All inputs use date labels no later than t−2. The sentiment index includes market-derived information. Sentiment index: Alternative.me; price and network data: Coin Metrics.

A lagged observation is not necessarily a verified point-in-time observation. Coin Metrics and Alternative.me can revise historical data, and the supplied snapshots do not establish when every index value or blockchain metric first became available. The Alternative.me API's date label is not a complete historical record of publication latency. A two-day lag and a three-day sensitivity provide useful timing checks, but they do not remove this uncertainty. A deployable live strategy or a definitive point-in-time backtest would need archived release timestamps, archived versions, and executable market prices.

3.5 Models and training procedure

The first model is ridge regression with a penalty parameter $\alpha = 100$. Its features are standardized using the mean and standard deviation estimated within the training data for that fitted model. For the validation forecast, scaling and coefficients are estimated on the 2018–2021 training interval. For the test forecast, the entire pipeline is refitted on the combined 2018–2022 interval. Standardization is never fitted on the validation or test outcomes before their respective evaluation. Ridge regularization provides a stable linear benchmark when features within and across blocks are correlated.

The second model is histogram-based gradient boosting regression using squared loss, learning rate 0.03, 100 iterations, a maximum of seven leaf nodes, at least 80 training observations per leaf, L2 regularization of 1.0, and random seed 19. Early stopping is disabled. The model is fitted separately for each asset and feature block. No test-period hyperparameter search is performed, and the same settings are used for both assets. The nonlinear model can represent interactions and thresholds, but its flexibility does not guarantee a lower out-of-sample error than a simple benchmark.

The specifications are fixed for this demonstration rather than selected from a large published model competition. Validation results are reported as a separate diagnostic; after those results are observed, the fixed specifications are refitted through 2022 and evaluated on 2023–2024. The work was not publicly preregistered, and the selection of two algorithms, feature transformations, and periods remains a research design choice. Any future claim of robust superiority should be tested on a genuinely new period or independently specified dataset.

Three baselines are useful. The zero-return forecast predicts 0 for every test day. The training-mean benchmark predicts the mean target return in the 2018–2022 fitting sample. Finally, the market-only model uses lagged return and volatility features but excludes the sentiment and network blocks. The zero and training-mean forecasts are not claimed to be optimal, but they provide simple reference points that flexible models should not be allowed to ignore. The market-only comparison assesses incremental information conditional on the chosen market variables.

3.6 Evaluation and uncertainty

Root mean squared error and mean absolute error are calculated on the 723 common test observations for each asset. The main out-of-sample R^2 compares a model's squared error with the zero-return benchmark, not with a mean calculated from the test outcomes. Equation 2 defines RMSE and Equation 3 defines R^2 against zero. A negative value in Equation 3 means that the model's average squared error is higher than the zero-return forecast's error on the same test dates. It is not an in-sample coefficient of determination.

$$RMSE = \sqrt{[(1/n)\, \Sigma_i\, (r_i - \hat{y}_i)^2]} \tag{2}$$

$$R^2_0 = 1 - \Sigma_i\, (r_i - \hat{y}_i)^2 \,/\, \Sigma_i\, r_i^2 \tag{3}$$

For a feature block B, Equation 4 defines relative test mean-squared-error improvement over the market-only model of the same algorithm and asset. Positive values indicate lower squared error for B; negative values indicate higher squared error. The comparison is paired by date, so both models face the same observed returns. This is a conditional comparison of particular feature transformations and a fitted model, not an estimate of the causal effect of sentiment or network use.

$$\Delta_B = 100\, (MSE_{market} - MSE_B) \,/\, MSE_{market} \tag{4}$$

To describe sampling variability in the observed sequence of paired losses, we apply a moving-block bootstrap with 2,000 resamples and a fixed block length of seven calendar days. Candidate blocks are sampled from consecutive observed dates; blocks are not allowed to cross the eight-day missing-data gap. Each resample concatenates selected blocks and is truncated to the original test-sample length. The 2.5th and 97.5th percentiles of the relative loss difference form the displayed interval. The models are not refitted in

the bootstrap, so the intervals condition on the fitted predictions and do not capture all model-estimation, model-selection, or data-vintage uncertainty.

The intervals are exploratory, are not adjusted for multiple comparisons, and should not be read as causal confidence intervals. We do not infer a reliable economic trading advantage from a point estimate or from whether a single interval includes zero. Squared loss places considerable weight on extreme returns, while the financial meaning of an improvement depends on trading rules, costs, execution, and risk constraints that are outside this study (Hansen, 2022; Svogun & Bazán-Palomino, 2022). The interval procedure is intended to reveal the uncertainty around the reported forecast-loss comparison, not to certify a strategy.

3.7 Sensitivity design and reproducibility

A timing sensitivity shifts the latest permissible predictor date from t−2 to t−3. For this exercise, the lagged return features become t−3, t−4, and t−8, and all rolling windows end at t−3. The sentiment and network blocks are shifted consistently. Because the missing Alternative.me date affects a different set of target dates under the longer lag, the two lag specifications are re-evaluated on their intersection of 722 test dates per asset. Models are refitted with the corresponding lagged training data, and the gradient-boosting RMSE values are compared on those common dates.

A change in lag alters both the amount and age of information available. The sensitivity therefore tests a different prediction problem rather than providing an independent replication. Additional assets, alternative sentiment measures, different network proxies, expanding-window refits, and additional non-overlapping test periods would be needed to investigate generality. The present analysis does not report rolling model retraining, trading performance, causal treatment effects, or text-level sentiment classification.

The analysis program uses Python 3.12.14, pandas 2.2.3, NumPy 2.3.5, and scikit-learn 1.8.0. The archive contains the source snapshots, attribution and provenance information, the program, test predictions, results, and the common-date lag comparison. The plotting script uses Matplotlib 3.10.8. Reproducibility of the calculations on the archived data should be distinguished from independent validation of the data providers' historical measurement and revision practices.

4. Results

4.1 Sample characteristics and baseline forecasts

The 2018–2022 fitting sample has a mean daily log return of approximately 0.047 percentage points for Bitcoin and 0.044 for Ether. In the retained 2023–2024 test observations, the corresponding means are approximately 0.240 and 0.148 percentage points, and standard deviations are 2.519 and 2.935 percentage points. These are descriptive statistics for the retained dates, not forecasts or estimates of an expected return for a future period. They also illustrate that the level and variability of returns differ between the fitting and test windows.

The average lagged Alternative.me mood score is approximately 42.37 in the 2018–2022 fitting sample and 58.76 on the retained test dates. Because the same Bitcoin-oriented index is supplied to both models, these averages are the same for the two assets. The change in the distribution is a reminder that a model trained on earlier data is applied to a different market environment. The shift should not be interpreted as a causal explanation for the test-period returns or as evidence that all investor groups were more optimistic.

For Bitcoin, the zero-return benchmark has a test RMSE of 2.5292 percentage points; the training-mean benchmark has an RMSE of 2.5251. For Ether, the corresponding values are 2.9369 and 2.9350. The training-mean benchmark uses the mean computed through 2022, not the mean of the test observations. These simple values provide a scale for assessing the model errors. They do not include transaction costs or imply that any forecast is suitable for a particular trading decision.

The history of cryptocurrency market efficiency is itself contested, with studies examining adaptive behavior, herding, and changing predictability (Hachicha et al., 2023; Karasiński, 2023; Yi et al., 2023). A close result against a simple zero benchmark is therefore informative even when it is not a strong positive finding. A complex model should be evaluated relative to what could have been forecast without the claimed additional information, not solely by whether its predictions vary over time.

4.2 Gradient-boosting comparison

Table 3 reports gradient-boosting validation and test results for the four blocks. The 2022 validation models are trained through 2021. For the 2023–2024 test, the same fixed specifications are refitted through 2022. Bitcoin's market-only test RMSE is 2.5461, while adding the sentiment block gives 2.5580. The network-augmented model gives 2.5389, and the combined model gives 2.5555. On this test window, the sentiment-augmented model is not the lowest-error Bitcoin specification, and all four gradient-boosting models are less accurate in squared error than predicting zero.

Table 3. Gradient-Boosting Validation and Test Performance

Feature set	BTC val. RMSE	BTC test RMSE	BTC test R^2_0	ETH val. RMSE	ETH test RMSE	ETH test R^2_0
Market only	3.4038	2.5461	-0.0135	4.5805	2.9666	-0.0203
Market + sentiment	3.3869	2.5580	-0.0230	4.5497	2.9668	-0.0205
Market + network	3.4209	2.5389	-0.0077	4.6594	3.0068	-0.0482
All blocks	3.4100	2.5555	-0.0210	4.6223	3.0047	-0.0467

Note. RMSE is in percentage points; lower is better. R^2_0 compares test squared error against a zero-return forecast. Validation n = 365 and test n = 723 per asset. Sentiment index: Alternative.me; price and network data: Coin Metrics.

For Ether, the market-only and sentiment-augmented test RMSE values are 2.9666 and 2.9668, respectively. The network-augmented and combined models have RMSE values of 3.0068 and 3.0047. The difference between the two former models is very small, while both are below the two models containing the specified network block. All four remain above the zero-return benchmark's 2.9369 RMSE. The results do not demonstrate that the Bitcoin-oriented index measures Ether-specific sentiment, nor do they establish that network fundamentals are irrelevant to Ether's price formation.

The validation results differ from the test results. In 2022, the sentiment-augmented gradient-boosting model has a Bitcoin RMSE of 3.3869 compared with 3.4038 for the market-only model, and an Ether RMSE of 4.5497 compared with 4.5805. Those validation differences are descriptive and were followed by a refit. In the subsequent holdout, the same addition does not yield a comparable improvement. The contrast illustrates temporal instability and the danger of drawing a broad conclusion from one favorable interval, but it does not identify which market mechanism changed.

The difference between RMSE and MAE rankings is also relevant. For Bitcoin's test period, the network-augmented gradient-boosting model has a slightly lower RMSE than the market-only model but a higher MAE: approximately 1.8026 versus 1.7794. Squared error gives more weight to large forecast misses, whereas absolute error assigns them linearly. A lower RMSE in one comparison therefore does not imply an improvement for every loss function. The program reports both metrics so that readers can inspect the trade-off rather than rely on a single number.

4.3 Incremental comparisons and uncertainty

Figure 1 and Table 4 compare each gradient-boosting block with the market-only model using relative test mean-squared-error improvement. A positive value denotes an improvement; a negative value denotes a

deterioration. For Bitcoin, the sentiment block has an estimated change of −0.94%, with an exploratory 95% moving-block-bootstrap interval of −2.90% to 1.07%. The network block has an estimated improvement of 0.57%, with an interval of −2.05% to 2.71%. The combined block has an estimated change of −0.74%, with an interval of −3.05% to 1.49%.

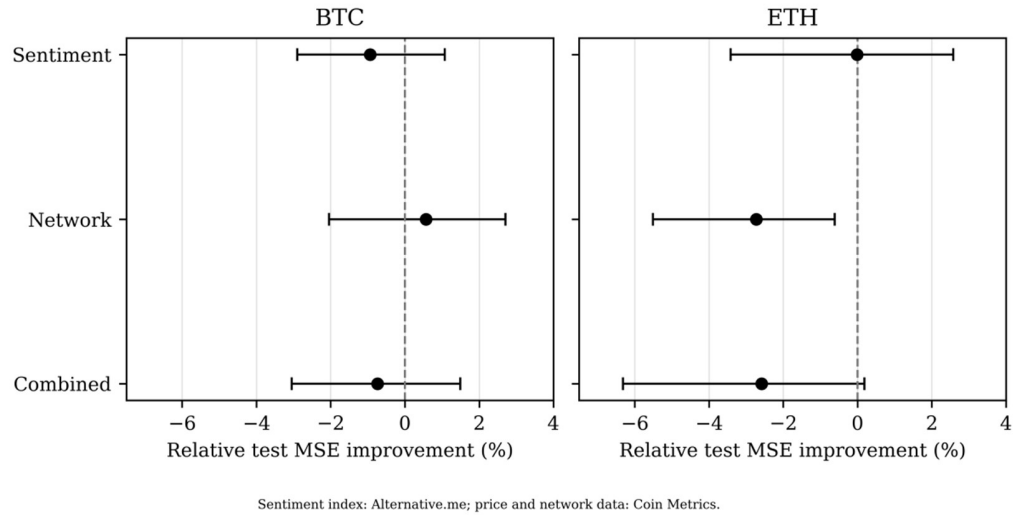


Figure 1. Incremental Forecasting Performance of Sentiment and Network Features

Note. Test-period relative mean-squared-error changes from adding the specified sentiment, network, or both blocks to the market-only gradient-boosting model for Bitcoin (BTC) and Ether (ETH). Points are estimates; bars are exploratory 95% moving-block bootstrap intervals (2,000 resamples, seven-day blocks). Positive values indicate improvement. Sentiment index: Alternative.me; price and network data: Coin Metrics.

Table 4. Incremental Gradient-Boosting Test Results

Asset	Compared block	MSE improvement (%)	95% interval (%)
BTC	Market + sentiment	-0.94	[-2.90, 1.07]
BTC	Market + network	+0.57	[-2.05, 2.71]
BTC	All blocks	-0.74	[-3.05, 1.49]
ETH	Market + sentiment	-0.01	[-3.42, 2.58]
ETH	Market + network	-2.73	[-5.51, -0.61]
ETH	All blocks	-2.59	[-6.32, 0.18]

Note. Each block is compared with the market-only model for the same asset. Positive values indicate improvement. Intervals are exploratory and unadjusted percentile intervals from 2,000 paired seven-day moving-block resamples; models are not refitted. Sentiment index: Alternative.me; price and network data: Coin Metrics.

For Ether, the estimated change from adding the sentiment block is −0.01%, with an interval of −3.42% to 2.58%. The network block has an estimated change of −2.73%, with an interval of −5.51% to −0.61%. Adding both blocks produces an estimated change of −2.59%, with an interval of −6.32% to 0.18%. These intervals are based on a single fitted prediction sequence and are not adjusted for the multiple models and comparisons reported in the paper. The network result is a loss comparison for one specified model and period, not a causal estimate of the value of blockchain activity.

The archive also contains nested comparisons that add one block conditional on the other. In the Bitcoin gradient-boosting model, adding sentiment to the network-augmented model changes test MSE by approximately −1.32%, with an interval spanning zero. Adding network features to the sentiment-augmented model changes test MSE by approximately 0.20%, also with an interval spanning zero. For Ether, the analogous changes are approximately 0.14% and −2.58%. These results show that the question "which block matters?" depends on the baseline and that the two blocks may share information with market history and with each other.

The block intervals condition on the fitted forecasts and preserve short runs of consecutive daily loss observations. They do not include the uncertainty of re-estimating models on a different training sample, of choosing a different block length, or of revising historical provider data. The missing October 2024 sentiment observation also prevents resampled blocks from crossing the resulting gap. The intervals should be treated as a descriptive uncertainty check for these paired losses, not as a complete treatment of nonstationarity or model-selection risk.

4.4 Linear benchmark

Table 5 reports the ridge results. For Bitcoin, the market-only model has a test RMSE of 2.5294, close to the zero-return benchmark's 2.5292. Adding sentiment increases RMSE to 2.5465; adding the network block gives 2.5544; and adding both gives 2.5746. For Ether, the market-only model has a test RMSE of 2.9305, slightly below the zero-return benchmark's 2.9369. The sentiment, network, and combined variants have RMSE values of 2.9475, 2.9352, and 2.9536, respectively.

Table 5. Ridge Regression Validation and Test Performance

Feature set	BTC val. RMSE	BTC test RMSE	BTC test R^2_o	ETH val. RMSE	ETH test RMSE	ETH test R^2_o
Market only	3.3974	2.5294	-0.0002	4.6125	2.9305	0.0044
Market + sentiment	3.3896	2.5465	-0.0138	4.6035	2.9475	-0.0072
Market + network	3.4000	2.5544	-0.0201	4.6261	2.9352	0.0012
All blocks	3.3931	2.5746	-0.0363	4.6270	2.9536	-0.0114

Note. RMSE is in percentage points; lower is better. R^2_o compares test squared error against a zero-return forecast. Validation n = 365 and test n = 723 per asset. Sentiment index: Alternative.me; price and network data: Coin Metrics.

The ridge comparison does not establish that the market-history variables will remain useful in a new period. The positive zero-benchmark R^2 for the Ether market-only model is approximately 0.0044, a small reduction in squared error for this holdout. Bitcoin's corresponding R^2 is approximately −0.0002. Adding more variables can raise or lower test error because estimated coefficients must generalize to new observations; additional information does not guarantee a better forecast when estimation error and distribution changes are present.

The ridge and gradient-boosting results also show that an observed block ranking is model-dependent. For Bitcoin's gradient-boosting model, the network-augmented specification has the lowest RMSE among the four fitted variants, whereas the ridge market-only model has the lowest RMSE among its variants. For Ether, market-only ridge and market-only gradient boosting are each at least as accurate by RMSE as their sentiment-augmented counterparts in the test period. The consistent result is narrower than the title's proposition: the specified sentiment block does not demonstrate stable incremental superiority in this particular comparison.

4.5 Longer information lag and common dates

The three-day-lag sensitivity is evaluated on 722 dates common to both lag designs. For Bitcoin, the gradient-boosting market-only RMSE is 2.5449 with the two-day lag and 2.5847 with the three-day lag. The corresponding sentiment-augmented RMSE values are 2.5570 and 2.5698. Under the longer lag, sentiment modestly improves on the market-only model for Bitcoin, but the network-augmented model's RMSE is lower at 2.5561. Both the information set and the fitted model change, and the comparison remains within the same historical test period.

For Ether, the corresponding two-day market-only and sentiment-augmented RMSE values are 2.9677 and 2.9687. With a three-day lag, they are 2.9648 and 2.9624. The longer-lag sentiment difference is small and the network-augmented model remains less accurate by RMSE. These numbers are evaluated on the same 722 dates within the sensitivity exercise, avoiding a comparison driven solely by the different missing-data windows. They do not prove that a three-day lag captures the true publication time of every historical source value.

The timing sensitivity changes the rank of the sentiment-augmented model relative to the market-only model in both assets, but not the broader interpretation. The differences remain small relative to daily return volatility, and the gradient-boosting forecasts have negative zero-benchmark R^2 on the common-date sensitivity sample. Lag choice, feature overlap, and estimation noise can all affect the ranking. An independent evaluation with archived releases and additional periods is needed before treating any small difference as stable.

5. Discussion

5.1 Interpreting the title's proposition

The study's central empirical result is that the proposed sentiment block does not consistently outperform the selected network block or the market-only benchmark. The headline proposition is therefore not established by the test. In the validation year, sentiment augmentation reduces gradient-boosting RMSE for both assets; in the later test period, that pattern weakens or reverses. The same exercise also shows that network augmentation can fail to improve forecasting. These findings are consistent with a setting in which daily returns are difficult to predict, the chosen signals are noisy, or relationships differ across periods, but the analysis cannot distinguish these explanations causally.

The finding should not be inverted into a claim that fundamentals always trump sentiment. Our network block contains active addresses, transactions, and native fees, and none is a complete valuation model. It omits supply dynamics, protocol risks, adoption quality, decentralized-finance activity, revenue allocation, and macroeconomic conditions. A different representation could yield different results. Likewise, the Alternative.me composite may not capture news tone, institutional positioning, or investor beliefs that could be useful in another design (Chen et al., 2021; Guégan & Renault, 2021; Han, 2025).

This distinction also matters for the word "drivers." The analysis predicts returns using lagged observations and compares forecast errors; it does not estimate structural causal effects. A variable can help predict returns because it proxies for another process or reflects earlier market movements. Conversely, a genuine economic factor can fail to improve a short-horizon forecast if its information is already incorporated into prices. The observed model rankings are evidence about predictive performance under a declared setup, not proof of the mechanisms producing price changes.

5.2 The limits of a composite sentiment proxy

The fear–greed index is attractive because it is public, daily, and straightforward to reproduce. Its composition is also a central limitation. The provider currently describes volatility and momentum or volume components, so part of the score is generated from market behavior that resembles the baseline features. It

also includes social and other signals whose historical definitions and weights may not be constant. An incremental comparison against lagged returns and volatility therefore asks whether the composite adds useful information beyond those particular market features, not whether pure emotional sentiment adds value.

Using the same Bitcoin-oriented score for Ether creates a second interpretive issue. It may capture broad cryptocurrency market conditions, a Bitcoin risk factor, or behavioral attention that spills over to Ether. It does not directly measure the sentiment of Ether users, developers, validators, or investors. Research on cross-asset and sentiment spillovers provides reasons to investigate these channels separately (Chen & Hu, 2026; Leong & Kwok, 2025), but this study does not estimate a spillover model. A richer design could compare asset-specific text, search, survey, and option-implied measures using consistent point-in-time timestamps.

Attention should not be equated with optimism. A spike in searches or social-media posts can occur during a price crash, a security incident, a regulatory announcement, or a rally. The relation between sentiment and prices can also be bidirectional, and market-derived components make reverse dependence especially plausible (Gaies et al., 2023; Gessie, 2026; Zhu et al., 2021). A rigorous causal design would need a credible source of exogenous variation or a carefully designed event study. Merely lagging the sentiment index, while necessary for a forecasting exercise, is not sufficient for causal identification.

5.3 Network metrics and the meaning of fundamentals

The selected network features describe recent changes and deviations in activity. They may reflect transaction demand, congestion, exchange behavior, speculation, or protocol-specific events. Active addresses are not unique users, transaction counts are not necessarily purchases of goods or services, and native fees are not a direct measure of net economic value. In Ethereum, changes in fee rules and network use can alter the meaning of the same series over time (Kondor et al., 2021; Meister & Price, 2024). The use of identical transformations across assets improves procedural comparability while leaving substantive comparability unresolved.

A weak daily prediction from these proxies should not be read as evidence that blockchain activity has no long-run economic relevance. Different horizons can answer different questions: a network adoption measure might be related to long-run use or valuation while contributing little to tomorrow's price change. Likewise, certain network variables may be informative only during congestion or security events, or after separating exchange-related from non-exchange activity. A more complete study could incorporate protocol-specific variables and clearly define whether it targets returns, volatility, usage, or welfare (Caporale et al., 2021; Koutmos, 2018; Liu et al., 2025).

Network and sentiment data can also be affected by the same underlying market event. A price increase may attract users and transactions while simultaneously raising momentum-based sentiment components. A model may distribute predictive weight across correlated variables in different ways without changing its forecast. This is one reason we use block-level loss differences rather than interpreting individual coefficient magnitudes or tree importance as a ranking of true causes (Altmann et al., 2010; Conn et al., 2019). Conditional effects, causal mechanisms, and the economic value of a signal require additional analysis.

5.4 Model performance and the value of simple baselines

The zero-return forecast performs well relative to the gradient-boosting models in the reported test window. For Bitcoin, every gradient-boosting specification has higher squared error than zero; the same is true for Ether. Ridge market-only forecasting yields a small improvement relative to zero for Ether but is close to zero for Bitcoin. These findings do not prove that cryptocurrency markets are perfectly efficient. They show that the specified models do not reliably exploit the chosen information to beat a simple reference on the selected holdout.

Price-level prediction can obscure this distinction. Because a price at one date is often close to its previous level relative to its long-run range, a model that predicts the last observed price may achieve a high level-based fit while having little ability to forecast the return. Evaluating returns and reporting a naive benchmark prevents that persistence from being mistaken for a strong directional signal (Chen et al., 2020; Murray et al., 2023; Ranjan et al., 2023). Other forecasting goals may require different loss functions: volatility prediction, for example, is not evaluated by the return RMSE in this article.

The comparison also demonstrates the risk of treating the most favorable validation result as a final result. Sentiment helps the fixed gradient-boosting model in 2022, but its advantage does not persist in the main 2023–2024 comparison. A researcher who reports only the validation period could reach a more optimistic conclusion. Chronological separation is essential, but a single test period can still be influenced by market regimes, and repeated attempts across many datasets can reintroduce selection bias (Bailey et al., 2016; Romano & Wolf, 2005; White, 2000; Yae & Luo, 2023). Independent replication and disclosure of the full model search are therefore important.

5.5 Implications for business analytics

For a business analytics team, the results support maintaining a clear distinction among data availability, predictive accuracy, and business value. A public daily index can be easy to access yet add little incremental value to an existing market model. A network signal may be economically interesting but require a different horizon or measurement process. Before building a dashboard, teams should specify the decision, forecast origin, expected latency, benchmark, error metric, and acceptable failure modes. These choices determine whether a predictor is useful for a particular task.

A practical deployment would require point-in-time data feeds and records of revisions. For every feature, the system should record the event time, publication time, ingestion time, transformation version, and model version. A lag based on a calendar date is not a substitute for an observed publication timestamp. A production pipeline also needs a rule for missing sentiment values, provider outages, protocol changes, and invalid or revised observations. Those operational controls are not evaluated by the current retrospective exercise.

The results also argue for monitoring performance by asset and period rather than presenting a single aggregate score. Bitcoin and Ether share a mood proxy but have different on-chain mechanics, and their feature-block rankings are not identical. An organization should evaluate whether its model remains calibrated when the distribution of inputs or returns changes. Retraining frequency, a fixed benchmark, drift alerts, and a documented decision to retire an underperforming model can be more valuable than adding another complex predictor without independent evidence.

A business case based on trading would require additional steps that this paper does not undertake: specifying entry and exit times, feasible order types, exchange coverage, fees, spreads, market impact, slippage, leverage, and position constraints. The test uses daily reference prices rather than executable fills. A small decrease in forecast RMSE cannot be translated automatically into profit, and directional accuracy alone is insufficient to establish economic value. These distinctions are consistent with research on data snooping, technical rules, and the economic evaluation of forecasts (Hansen, 2022; Neuhierl & Schlusche, 2011; Svogun & Bazán-Palomino, 2022).

5.6 Limitations and future research

The first limitation is the restricted asset universe. Bitcoin and Ether are large, liquid assets with relatively long data histories; results may differ for smaller tokens, stablecoins, exchange tokens, or assets created later. A broader panel should address survivorship bias, delistings, changing token identifiers, and differences in data coverage. It should also distinguish time-series prediction of each asset from cross-sectional rankings or portfolio-level factors (Borri & Shakhnov, 2022; Leong & Kwok, 2023).

Second, the sentiment index is a Bitcoin-oriented composite and is not a clean behavioral experiment. Its documented price-based components overlap with market predictors, and historical revisions or changes in methodology may not be observable in the downloaded file. Future work could use archived text and publication timestamps, independently measure language and engagement, and compare multiple proxies on the same set of prediction dates. A stronger design would report which sources are available at each forecast origin and how bots, duplicate messages, and platform changes are handled (Amirshahi & Lahmiri, 2025; Koltun & Yamshchikov, 2023; Li & Ma, 2024).

Third, the network block is incomplete and potentially endogenous. It excludes important protocol-specific variables and does not separate exchange transfers, smart-contract activity, and economically distinct users. The choice of seven-day change and 30-day deviation is one plausible transformation among many. Future work should motivate additional variables before testing, assess structural breaks, and examine whether activity measures have different meaning around major geopolitical or cyber events (Caporale et al., 2021; Liu et al., 2025).

Fourth, the time split is fixed and the reported test is a single historical interval. A 2022 validation result and a 2023–2024 test result do not reveal the full distribution of model performance across bull, bear, and transition periods. An expanding-window or rolling-origin design could estimate how rankings evolve, provided that hyperparameter selection is nested within each training window and no future data enter preprocessing. Such an extension should report all windows rather than selecting the most favorable one (Arian et al., 2024; Deng, 2023).

Fifth, the uncertainty intervals are conditional and exploratory. The seven-day block length is a design choice, and the test sample includes a missing-data gap. The bootstrap does not refit the models or correct for every comparison. More comprehensive work could examine alternative block lengths, explicitly model volatility, use forecast-comparison procedures appropriate to nested models, and validate results in additional non-overlapping periods (Diebold & Mariano, 1995; Diebold, 2015; Kunsch, 1989; Mariano & Preve, 2012; Politis & Romano, 1994). The current intervals should not be treated as a definitive ranking of all possible predictors.

Sixth, the historical source files are retrospective snapshots. A two- or three-day lag reduces obvious timing overlap but cannot recreate values as originally published. Revisions, delayed index publication, changes to exchange coverage, and protocol upgrades can affect the apparent history. A live or archived-vintage replication is required to establish a deployable point-in-time forecasting process. The reproducibility archive supports checking the reported numbers on the same snapshot; it does not resolve the historical publication and revision problem.

Finally, sentiment and fundamentals can interact rather than compete as independent causes. Behavioral pressure may influence network use, network events may change investor beliefs, and both may respond to macroeconomic news. Bubbles, volatility, and cross-asset contagion complicate simple comparisons (Berger & Turtle, 2015; Bouri et al., 2017; Leong & Kwok, 2023). Future research should use theory to distinguish measurement, prediction, and causation and should avoid treating a model's selected features as a direct map of market mechanisms.

6. Conclusion

This study evaluates a sentiment-versus-network forecasting proposition using public daily data for Bitcoin and Ether, time-ordered training and test periods, fixed feature blocks, a regularized linear model, and gradient boosting. In the 2023–2024 holdout, the specified Bitcoin-oriented sentiment proxy does not consistently improve return forecasts over market-history or network-augmented alternatives. All gradient-boosting variants have higher squared error than the zero-return benchmark, while the linear market-only model shows a small improvement for Ether.

The evidence therefore does not support a general conclusion that sentiment trumps fundamentals. It supports a narrower methodological conclusion: claims about cryptocurrency return drivers depend on the proxy, information timing, benchmark, horizon, asset, and test period. The public data snapshot and code enable the calculations to be checked, while the limitations of retrospective data, overlapping market information, and non-causal interpretation remain explicit.

Acknowledgement

We acknowledge the researchers whose published work informed this study. Any remaining errors are our responsibility.

Author Contributions

Conceptualization, M.Z. and L.B.Z.; Methodology, M.Z. and L.B.Z.; Formal Analysis, M.Z. and L.B.Z.; Investigation, M.Z. and J.C.; Data Curation, M.Z. and J.C.; Software, M.Z. and J.C.; Validation, M.Z. and L.B.Z. and J.C.; Resources, J.C.; Visualization, M.Z. and J.C.; Writing – Original Draft, M.Z.; Writing – Review & Editing, M.Z. and L.B.Z. and J.C.; Supervision, L.B.Z.; Project Administration, M.Z. and L.B.Z. We have read and approved the final manuscript and take responsibility for the work.

Funding

This research received no external funding.

Conflicts of Interest

We declare no conflicts of interest.

Ethical Approval and Consent

The study analyzes public, aggregate market and blockchain data and does not use individual-level personal data or involve interventions. We take responsibility for compliance with applicable institutional requirements.

Data Availability Statement

The analysis program, downloaded source snapshots, provenance and hashes, test predictions, results, and common-date lag sensitivity are supplied in the reproducibility archive. Sentiment index: Alternative.me (https://alternative.me/crypto/fear-and-greed-index/); price and network data: Coin Metrics Community Data (https://coinmetrics.io/community-network-data/). The snapshots do not establish the historical publication or revision time of each observation.

Declaration of Generative AI Use

We used OpenAI ChatGPT (GPT-6) to assist with literature discovery, drafting, analytical implementation, and document preparation. We take full responsibility for the accuracy of the references, calculations, and final manuscript.

References

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. Bioinformatics, 26(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Amirshahi, B., & Lahmiri, S. (2025). Investigating the effectiveness of Twitter sentiment in cryptocurrency close price prediction by using deep learning. Expert Systems, 42(1), e13428. https://doi.org/10.1111/exsy.13428

Arian, H., Norouzi Mobarekeh, D., & Seco, L. (2024). Backtest overfitting in the machine learning era: A comparison of out-of-sample testing methods in a synthetic controlled environment. Knowledge-Based Systems, 305, 112477. https://doi.org/10.1016/j.knosys.2024.112477

Bailey, D., Borwein, J., López de Prado, M., & Zhu, Q. J. (2016). The probability of backtest overfitting. The Journal of Computational Finance. https://doi.org/10.21314/jcf.2016.322

Berger, D., & Turtle, H. J. (2015). Sentiment bubbles. Journal of Financial Markets, 23, 59–74. https://doi.org/10.1016/j.finmar.2015.01.002

Berger, T., & Koubová, J. (2024). Forecasting Bitcoin returns: Econometric time series analysis vs. machine learning. Journal of Forecasting, 43(7), 2904–2916. https://doi.org/10.1002/for.3165

Biswas, P., & Sharma, C. (2025). Does fear sentiment drive cryptocurrency volatility? Evidence from Google trends data. Finance Research Letters, 86, 108510. https://doi.org/10.1016/j.frl.2025.108510

Borri, N., & Shakhnov, K. (2022). The cross-section of cryptocurrency returns. The Review of Asset Pricing Studies, 12(3), 667–705. https://doi.org/10.1093/rapstu/raac007

Borup, D. (2019). Asset pricing model uncertainty. Journal of Empirical Finance, 54, 166–189. https://doi.org/10.1016/j.jempfin.2019.07.005

Bouri, E., Azzi, G., & Dyhrberg, A. H. (2017). On the return-volatility relationship in the Bitcoin market around the price crash of 2013. Economics, 11(1), 2. https://doi.org/10.5018/economics-ejournal.ja.2017-2

Bouteska, A., Mefteh-Wali, S., & Dang, T. (2022). Predictive power of investor sentiment for Bitcoin returns: Evidence from COVID-19 pandemic. Technological Forecasting and Social Change, 184, 121999. https://doi.org/10.1016/j.techfore.2022.121999

Byun, S.-J., Cho, S., & Kim, D.-H. (2024). Can a machine learn from behavioral biases? Evidence from stock return predictability of deep learning models. Journal of Behavioral and Experimental Finance, 41, 100881. https://doi.org/10.1016/j.jbef.2023.100881

Caporale, G. M., Kang, W.-Y., Spagnolo, F., & Spagnolo, N. (2021). Cyber-attacks, spillovers and contagion in the cryptocurrency markets. Journal of International Financial Markets, Institutions and Money, 74, 101298. https://doi.org/10.1016/j.intfin.2021.101298

Cavalheiro, E. A., Vieira, K. M., & Thue, P. S. (2024). The impact of investor greed and fear on cryptocurrency returns: A granger causality analysis of Bitcoin and Ethereum. Review of Behavioral Finance, 16(5), 819–835. https://doi.org/10.1108/rbf-08-2023-0224

Chen, W., Xu, H., Jia, L., & Gao, Y. (2021). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. International Journal of Forecasting, 37(1), 28–43. https://doi.org/10.1016/j.ijforecast.2020.02.008

Chen, Y.-L., & Hu, M.-C. (2026). Sentiment spillovers from news and social media in cryptocurrency markets. Journal of International Financial Markets, Institutions and Money, 110, 102357. https://doi.org/10.1016/j.intfin.2026.102357

Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. Journal of Computational and Applied Mathematics, 365, 112395. https://doi.org/10.1016/j.cam.2019.112395

Choi, H. (2021). Investor attention and Bitcoin liquidity: Evidence from Bitcoin tweets. Finance Research Letters, 39, 101555. https://doi.org/10.1016/j.frl.2020.101555

Conn, D., Ngun, T., Li, G., & Ramirez, C. M. (2019). Fuzzy forests: Extending random forest feature selection for correlated, high-dimensional data. Journal of Statistical Software, 91(9). https://doi.org/10.18637/jss.v091.i09

Deng, A. (2023). Time series cross validation: A theoretical result and finite sample performance. Economics Letters, 233, 111369. https://doi.org/10.1016/j.econlet.2023.111369

Dias, I. K., Fernando, J. M. R., & Fernando, P. N. D. (2022). Does investor sentiment predict Bitcoin return and volatility? A quantile regression approach. International Review of Financial Analysis, 84, 102383. https://doi.org/10.1016/j.irfa.2022.102383

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. Journal of Business & Economic Statistics, 33(1), 1–1. https://doi.org/10.1080/07350015.2014.983236

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Dunbar, K., & Owusu-Amoako, J. (2022). Cryptocurrency returns under empirical asset pricing. International Review of Financial Analysis, 82, 102216. https://doi.org/10.1016/j.irfa.2022.102216

Gaies, B., Nakhli, M. S., Sahut, J.-M., & Schweizer, D. (2023). Interactions between investors' fear and greed sentiment and Bitcoin prices. The North American Journal of Economics and Finance, 67, 101924. https://doi.org/10.1016/j.najef.2023.101924

Gessie, L. (2026). Do Bitcoin returns move sentiment? Evidence from the crypto fear & greed index. Finance Research Open, 2(1), 100094. https://doi.org/10.1016/j.finr.2026.100094

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. The Review of Financial Studies, 33(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Guégan, D., & Renault, T. (2021). Does investor sentiment on social media provide robust information for Bitcoin returns predictability? Finance Research Letters, 38, 101494. https://doi.org/10.1016/j.frl.2020.101494

Hachicha, F., Masmoudi, A., Abid, I., & Obeid, H. (2023). Herding behavior in exploring the predictability of price clustering in cryptocurrency market. Finance Research Letters, 57, 104178. https://doi.org/10.1016/j.frl.2023.104178

Han, S. (2025). Investor sentiment and cross-section of cryptocurrency returns. Journal of Behavioral and Experimental Finance, 46, 101043. https://doi.org/10.1016/j.jbef.2025.101043

Hansen, E. (2022). Economic evaluation of asset pricing models under predictability. Journal of Empirical Finance, 68, 50–66. https://doi.org/10.1016/j.jempfin.2022.06.001

He, C.-W., & Wang, Y.-J. (2023). Bitcoin price jumps and investor sentiment indicators. Applied Economics Letters, 30(18), 2626–2630. https://doi.org/10.1080/13504851.2022.2102122

Hwang, I., & Kwon, J. H. (2024). Cryptocurrency returns and consumption-based asset pricing. Applied Economics, 56(55), 7393–7408. https://doi.org/10.1080/00036846.2024.2358214

Jia, Y., Wu, Y., Yan, S., & Liu, Y. (2023). A seesaw effect in the cryptocurrency market: Understanding the return cross predictability of cryptocurrencies. Journal of Empirical Finance, 74, 101428. https://doi.org/10.1016/j.jempfin.2023.101428

Karasiński, J. (2023). The adaptive market hypothesis and the return predictability in the cryptocurrency markets. Economics and Business Review, 9(1), 94–118. https://doi.org/10.18559/ebr.2023.1.4

Koltun, V., & Yamshchikov, I. P. (2023). Pump it: Twitter sentiment analysis for cryptocurrency price prediction. Risks, 11(9), 159. https://doi.org/10.3390/risks11090159

Kondor, D., Bulatovic, N., Stéger, J., Csabai, I., & Vattay, G. (2021). The rich still get richer: Empirical comparison of preferential attachment via linking statistics in Bitcoin and Ethereum. Frontiers in Blockchain, 4, 668510. https://doi.org/10.3389/fbloc.2021.668510

Koutmos, D. (2018). Bitcoin returns and transaction activity. Economics Letters, 167, 81–85. https://doi.org/10.1016/j.econlet.2018.03.021

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. The Annals of Statistics, 17(3), 1217–1241. https://doi.org/10.1214/aos/1176347265

Leong, M., & Kwok, S. (2023). The pricing of jump and diffusive risks in the cross-section of cryptocurrency returns. Journal of Empirical Finance, 74, 101420. https://doi.org/10.1016/j.jempfin.2023.101420

Leong, M., & Kwok, S. (2025). Bitcoin spillovers: A high-frequency cross-asset analysis. Financial Review, 60(2), 453–479. https://doi.org/10.1111/fire.12418

Li, S., & Ma, J. (2024). The impact of sentiment and engagement of Twitter posts on cryptocurrency price movement. Finance Research Letters, 65, 105598. https://doi.org/10.1016/j.frl.2024.105598

Liu, K., Manahov, V., & Stafylas, D. (2025). An investigation into the relationship between cryptocurrency active addresses and prices during major geopolitical conflicts. The European Journal of Finance, 31(18), 2323–2351. https://doi.org/10.1080/1351847x.2025.2585957

Liu, Y., & Tsyvinski, A. (2021). Risks and returns of cryptocurrency. The Review of Financial Studies, 34(6), 2689–2727. https://doi.org/10.1093/rfs/hhaa113

Mariano, R. S., & Preve, D. (2012). Statistical tests for multiple forecast comparison. Journal of Econometrics, 169(1), 123–130. https://doi.org/10.1016/j.jeconom.2012.01.014

Meister, B. K., & Price, H. C. W. (2024). Gas fees on the Ethereum blockchain: From foundations to derivative valuations. Frontiers in Blockchain, 7, 1462666. https://doi.org/10.3389/fbloc.2024.1462666

Murray, K., Rossi, A., Carraro, D., & Visentin, A. (2023). On forecasting cryptocurrency prices: A comparison of machine learning, deep learning, and ensembles. Forecasting, 5(1), 196–209. https://doi.org/10.3390/forecast5010010

Naifar, N., & Altamimi, S. (2023). Asymmetric impact of investor sentiment and media coverage news on Bitcoin returns. Managerial Finance, 49(8), 1342–1354. https://doi.org/10.1108/mf-08-2022-0400

Neuhierl, A., & Schlusche, B. (2011). Data snooping and market-timing rule performance. Journal of Financial Econometrics, 9(3), 550–587. https://doi.org/10.1093/jjfinec/nbq032

Panagiotidis, T. (2010). An out-of-sample test for nonlinearity in financial time series: An empirical application. Computational Economics, 36(2), 121–132. https://doi.org/10.1007/s10614-010-9225-z

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. Journal of the American Statistical Association, 89(428), 1303–1313. https://doi.org/10.1080/01621459.1994.10476870

Ranjan, S., Kayal, P., & Saraf, M. (2023). Bitcoin price prediction: A machine learning sample dimension approach. Computational Economics, 61(4), 1617–1636. https://doi.org/10.1007/s10614-022-10262-6

Romano, J. P., & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. Econometrica, 73(4), 1237–1282. https://doi.org/10.1111/j.1468-0262.2005.00615.x

Su, X., Zhan, W., & Li, Y. (2021). Quantile dependence between investor attention and cryptocurrency returns: Evidence from time and frequency domain analyses. Applied Economics, 53(55), 6439–6471. https://doi.org/10.1080/00036846.2021.1940826

Subramaniam, S., & Chakraborty, M. (2020). Investor attention and cryptocurrency returns: Evidence from quantile causality approach. Journal of Behavioral Finance, 21(1), 103–115. https://doi.org/10.1080/15427560.2019.1629587

Suhonen, A., Lennkh, M., & Perez, F. (2017). Quantifying backtest overfitting in alternative beta strategies. The Journal of Portfolio Management, 43(2), 90–104. https://doi.org/10.3905/jpm.2017.43.2.090

Svogun, D., & Bazán-Palomino, W. (2022). Technical analysis in cryptocurrency markets: Do transaction costs and bubbles matter? Journal of International Financial Markets, Institutions and Money, 79, 101601. https://doi.org/10.1016/j.intfin.2022.101601

Szeberényi, A., & Kovács, M. I. (2026). Fear, greed and Bitcoin return distributions: Evidence from digital asset market states. Digital Finance, 8(3), 50. https://doi.org/10.1007/s42521-026-00218-y

White, H. (2000). A reality check for data snooping. Econometrica, 68(5), 1097–1126. https://doi.org/10.1111/1468-0262.00152

Yae, J., & Luo, Y. (2023). Robust monitoring machine: A machine learning solution for out-of-sample R^2-hacking in return predictability monitoring. Financial Innovation, 9(1), 94. https://doi.org/10.1186/s40854-023-00497-z

Yi, E., Yang, B., Jeong, M., Sohn, S., & Ahn, K. (2023). Market efficiency of cryptocurrency: Evidence from the Bitcoin market. Scientific Reports, 13(1), 4789. https://doi.org/10.1038/s41598-023-31618-4

Zhu, P., Zhang, X., Wu, Y., Zheng, H., & Zhang, Y. (2021). Investor attention and cryptocurrency: Evidence from the Bitcoin market. PLOS ONE, 16(2), e0246331. https://doi.org/10.1371/journal.pone.0246331