

When Is News Really News? Evidence from Cryptocurrency Markets*

Ulrich Aiounou[†]

Abstract

This paper examines whether news sentiment provides additional predictive information for cryptocurrency returns and volatility beyond that contained in market variables. I use asset-specific news headlines collected from GDELT to construct two daily sentiment measures based on a financial lexicon and FinBERT. The analysis covers Bitcoin, Ether, Solana, and XRP, with multiple forecast horizons. Seven forecasting models are considered, including linear regressions, tree-based methods, and a neural network. Forecasting performance is evaluated using an expanding-window approach, with RMSE comparisons and the Clark-West test. I consider both a common sample and asset-native samples to examine the sensitivity of the results to the sample design. The analysis also examines whether the predictive contribution of sentiment differs between high- and low-volatility market conditions.

The results indicate that news sentiment provides more favorable improvements for volatility forecasting than for returns. Statistically supported return gains are relatively limited and rarely persist across the two sample designs. In contrast, several volatility forecasts benefit from sentiment information, with the largest improvements generally obtained using Multilayer Perceptron (MLP), particularly for XRP. However, some of these gains weaken or reverse when different sample periods are considered. The results also show that improvements occur under different market volatility conditions, without a consistent advantage for either regime.

Keywords: Cryptocurrency forecasting; Sentiment analysis; Volatility forecasting; Financial news; Machine learning; FinBERT; Clark west test; Market regimes.

*We acknowledge research support by the French National Research Agency Grants ANR-17-EURE-0020 and ANR-21-CE26-0007-01 (project MLEforRisk) and by the Excellence Initiative of Aix-Marseille University - A*MIDEX.

[†]Aix-Marseille University, CNRS, AMSE, Marseille, France, Email: dimitri.aiounou@univ-amu.fr

1 Introduction

Cryptocurrency markets have attracted considerable attention from investors and researchers because of their high volatility and their development as a distinct class of financial assets (Urquhart 2016; Liu and Tsyvinski 2021). Unlike traditional financial markets, cryptocurrency trading continues throughout the week, while information about individual assets circulates through financial news, online media, and social networks. These characteristics raise the question of whether information beyond market variables can improve forecasting performance. Despite the increasing interest in cryptocurrencies, predicting their returns and volatility remains challenging. Price movements can be abrupt, and the relevance of predictive variables may depend on the cryptocurrency, market conditions, and forecast horizon. From an efficient-market perspective, publicly available information may already be reflected in prices (Fama 1970). Behavioral finance, however, suggests that investor beliefs and limits to arbitrage may influence financial markets (Baker and Wurgler 2007). Nevertheless, these perspectives do not directly establish whether news sentiment contains additional predictive information beyond observable market conditions.

Advances in Natural Language Processing (NLP) and machine learning have increased interest in textual information for financial forecasting. Several studies use financial news, social media, and online discussions to investigate their relationship with market outcomes. In equity markets, Tetlock (2007) finds that media pessimism is associated with short-run stock price movements, while Antweiler and Frank (2004) identifies predictive information for volatility in online stock discussions. Similar relationships have been investigated in cryptocurrency markets using social media sentiment and attention indicators (Mai et al. 2018; Kraaijeveld and De Smedt 2020; Shen, Urquhart, and P. Wang 2019). More recently, Brauneis and Sahiner (2026) examine the contribution of news sentiment to cryptocurrency volatility forecasting using different models, while Fiszeder, Orzeszko, and Pietrzyk (2026) find no statistically significant out-of-sample improvements from news-based indicators for Bitcoin returns and volatility in their application. These studies suggest that textual information may contain predictive information, although its contribution depends on the information source, forecast target, and empirical design. Moreover, identifying a relationship between sentiment and market outcomes does not necessarily imply that sentiment improves out-of-sample forecasts beyond the information already contained in market variables

Against this background, this study examines whether asset-specific news sentiment improves out-of-sample forecasts of cryptocurrency returns and volatility beyond the information already contained in market variables. News sentiment refers here to the positive or negative tone expressed in news headlines about a given cryptocurrency, rather than a direct measure of investors' beliefs, expectations, or attention. The analysis covers Bitcoin (BTC), Ether (ETH), Solana (SOL), and XRP,

allowing comparisons across different cryptocurrencies while maintaining comparable market and news information. The selection of these assets and its limitations are discussed in Section 3.

I construct two sentiment indicators, Lexicon and FinBERT, using asset-specific headlines collected from GDELT. For each cryptocurrency, I compare models using only market variables with models including one sentiment measure at a time. The forecast targets are cumulative returns and cumulative squared returns over horizons of one, seven, fourteen, and thirty days. Several forecasting models are considered, including linear regressions, regularized linear models, tree-based methods, and a neural network. Forecasting performance is evaluated using an expanding-window approach, with RMSE comparisons and the Clark–West test to assess the additional predictive contribution of sentiment. I also consider two sample designs: a common sample covering the same period for all cryptocurrencies and asset-native samples using their respective available histories. This allows cross-asset comparisons while examining the sensitivity of the results to the sample design. The forecasting methodology is described in Section 5.

The results indicate that news sentiment provides additional predictive information under some specifications, although its contribution is not systematic. Improvements in return forecasting are relatively limited and sensitive to the evaluation sample. The evidence is more favorable for volatility, with some substantial improvements obtained using MLP, particularly for XRP. However, these gains depend on the market benchmark, forecast horizon, and cryptocurrency. Some improvements remain statistically supported across the two sample designs, while others deteriorate or become negative. The analysis of volatility regimes also shows that sentiment can improve forecasting performance during both high- and low-volatility periods, without a consistent advantage for either regime.

The main contribution of this study is to provide a comparative evaluation of the incremental forecasting value of news sentiment across cryptocurrencies, forecast targets, horizons, and estimators. By comparing two sentiment measures with the same market benchmark within each estimator, the analysis examines whether the additional information contained in media tone improves predictions and whether these improvements remain stable across different sample designs and market conditions. The objective is therefore not to establish a causal relationship between news sentiment and cryptocurrency markets, but to identify the conditions under which sentiment information contributes to forecasting performance.

The remainder of this study is organized as follows. The literature section 2 review presents the main findings on textual information and financial forecasting. Section 3 describes the data, including market variables and sentiment indicators. Section 5 presents the forecasting methodology and evaluation procedures. Section 6 reports the empirical results and discusses their relationship with the literature.

Finally, Section 7 concludes the study.

2 Literature review

The literature on textual information in financial markets has examined several related concepts, including media tone, investor sentiment, and market attention. Although these measures may be related, they capture different aspects of the information environment. This distinction is important when comparing their predictive contribution across studies.

Several studies provide evidence that textual information can contribute to financial forecasting. However, the results depend on the information source, forecast target, market conditions, and estimation method. Moreover, identifying a relationship between sentiment and market outcomes does not necessarily imply that sentiment improves out-of-sample forecasts beyond the information already contained in market variables. The following review presents the main findings from traditional financial markets and cryptocurrencies, with particular attention to these differences.

2.1 Media tone in equity markets and investor-sentiment perspectives

The relationship between information and asset prices has been extensively studied in financial economics. The efficient market hypothesis discussed by Fama (1970) provides a framework in which asset prices reflect available information, limiting the possibility of systematically predicting returns from publicly available information. However, limited predictive performance from a particular sentiment indicator does not necessarily establish market efficiency. Behavioral finance provides another perspective on the role of investor beliefs in financial markets. For example, Baker and Wurgler (2007) argue that investor sentiment has a stronger influence on stocks that are difficult to value or arbitrage. Their findings suggest that the relationship between sentiment and market outcomes may depend on asset characteristics, particularly valuation uncertainty and trading constraints. Nevertheless, this does not necessarily imply that media tone directly measures investor sentiment or that newer assets are more sensitive to sentiment information.

An important contribution to the relationship between financial news and market outcomes is Tetlock (2007). Using a Wall Street Journal column, the author finds that higher media pessimism is associated with downward pressure on stock prices, followed by a subsequent reversal. The study also shows that unusually high pessimism is associated with increased trading volume. These findings suggest that the relationship between news sentiment and market outcomes may depend on the

forecast horizon and the variable considered. Other studies have examined the contribution of textual information to return and volatility prediction. Antweiler and Frank (2004) analyze messages published on Internet stock discussion boards and find predictive information for volatility, while the effects on returns are statistically significant but economically small. Similarly, GARCÍA (2013) find that the predictive relationship between financial news sentiment and stock returns is particularly important during recessions, suggesting that market conditions may influence the contribution of sentiment information.

The literature has also considered social media as a source of information about collective mood. For instance, Bollen, Mao, and Zeng (2011) report improvements in predicting the Dow Jones Industrial Average using specific Twitter mood dimensions. Their findings depend on the dimension considered, and predicting market direction differs from reducing forecast errors for continuous return variables. Overall, these studies show that textual information can be relevant for financial prediction, with a contribution that depends on the forecast target, information source, and market conditions. They also motivate examining whether similar relationships can be observed in cryptocurrency markets.

2.2 Sentiment, attention and cryptocurrency prediction

The increasing availability of online information has encouraged several studies to investigate the relationship between sentiment and cryptocurrency markets. Early contributions mainly focused on social media and online communities, where investors exchange information and express their expectations about market developments. For example, Kim et al. (2016) use comments and replies from online communities to predict price and transaction fluctuations for Bitcoin, Ether, and XRP. Similarly, Mai et al. (2018) examine the contribution of social media content to Bitcoin price prediction. Their results indicate that more bullish forum content is associated with higher future Bitcoin values, with messages from less active users providing particularly useful information. They also identify differences between forums and Twitter and report out-of-sample predictive improvements. These findings suggest that the source and characteristics of online information may influence its contribution to forecasting.

The relationship between sentiment and cryptocurrency returns is also examined by Kraaijeveld and De Smedt (2020). Using nine cryptocurrencies, the authors find that lexicon-based sentiment provides Granger-predictive information for Bitcoin, Bitcoin Cash, and Litecoin, while a bullishness measure is informative for EOS and TRON. Their findings indicate that the contribution of sentiment differs across cryptocurrencies and sentiment measures. However, a significant Granger-predictive relationship does not necessarily imply an improvement in out-of-sample forecasting accuracy relative to a model already including market information.

Another important distinction concerns sentiment and market attention. While sentiment measures the positive or negative direction of expressed information, attention is generally associated with the amount of information or online activity related to an asset. Abraham et al. (2018) examine the relationship between Twitter activity and the price direction of Bitcoin and Ether. Their results indicate that tweet volume provides useful information, while tweet sentiment is less informative in their application. Similarly, Shen, Urquhart, and P. Wang (2019) find that the number of Bitcoin-related tweets contains predictive information for next-day trading volume and realized volatility, but not for next-day returns. These studies suggest that attention may contain useful predictive information even when sentiment provides limited improvements. However, the two measures capture different aspects of online information and should not be interpreted as equivalent.

More recent studies have focused on sentiment extracted directly from cryptocurrency news. Brauneis and Sahiner (2026) examine the contribution of news sentiment to short-interval cryptocurrency volatility forecasting using several forecasting models. Their results indicate that including sentiment can improve or deteriorate forecasting performance depending on the estimator. This suggests that the contribution of sentiment may depend on the forecasting method rather than being systematic across model specifications. Similarly, Fiszeder, Orzeszko, and Pietrzyk (2026) investigate ChatGPT-based news indicators for Bitcoin returns and volatility. Although they identify in-sample relationships, their results do not show statistically significant improvements in out-of-sample forecasting performance. These two studies are particularly relevant because they evaluate the contribution of news-derived information to cryptocurrency forecasting. Nevertheless, their empirical settings differ from the present study. In particular, Brauneis and Sahiner (2026) consider short-interval volatility forecasting, while Fiszeder, Orzeszko, and Pietrzyk (2026) focus on Bitcoin using ChatGPT-based news indicators.

The present analysis considers two measures of sentiment extracted from the same source of asset-specific news headlines. It also evaluates cumulative daily returns and volatility over several forecast horizons for four cryptocurrencies. These differences are important because the contribution of sentiment may depend on the forecast target, information representation, and evaluation period. Taken together, the literature indicates that online information may contain predictive information, but its contribution to out-of-sample forecasting remains dependent on the empirical setting. In particular, the results obtained from social media, online communities, and news headlines cannot be directly compared, since these sources may capture different aspects of the information environment.

2.3 Cryptocurrencies in relation to traditional assets

Cryptocurrencies differ from traditional financial assets in several aspects, including their market structure, risk characteristics, and trading activity. These differences may influence the relationship between information and market outcomes. Liu and Tsyvinski (2021) examine the risk and return characteristics of cryptocurrencies and find that their behavior differs from that of stocks, currencies, and precious metals in their sample. They also identify time-series momentum and investor attention as important variables for cryptocurrency return prediction. These findings motivate considering cryptocurrency-specific information rather than assuming that conventional financial predictors are equally relevant in these markets.

The relationship between cryptocurrencies and traditional assets is also examined by Corbet et al. (2018). Their results indicate relatively limited connections between the cryptocurrencies and traditional financial assets considered, although these relationships vary over time. Similarly, Urquhart (2016) finds that conclusions about Bitcoin’s informational efficiency differ between the full sample and more recent observations. These findings suggest that the choice of the evaluation period may be important when examining predictive relationships in cryptocurrency markets. Another difference concerns the economic meaning of news information. In equity markets, financial news may contain information about firms’ earnings, cash flows, and other characteristics directly related to their valuation. Cryptocurrency news, in contrast, may concern market developments, regulation, adoption, or security events without necessarily being associated with equivalent corporate cash flows.

The trading structure also differs, since cryptocurrency markets operate continuously, including weekends, while equity markets generally follow specific trading sessions. These differences may influence how information is incorporated into prices and how forecast horizons are defined. However, they do not establish that sentiment information is necessarily more useful in cryptocurrency markets. Such a conclusion would require a direct comparison with traditional assets using comparable information sets and evaluation procedures.

2.4 Text representations and forecasting models

The measurement of sentiment is an important aspect of the literature, since different methods may capture different characteristics of textual information. One common approach consists of using dictionaries that classify words according to their positive or negative meaning. LOUGHRAN and MCDONALD (2011) show that general-purpose sentiment dictionaries can incorrectly classify some expressions in financial texts. They propose dictionaries adapted to financial language to address this problem. Their findings motivate the use of domain-specific vocabularies

when measuring sentiment. In this study, I use a dictionary constructed specifically for cryptocurrency news rather than the dictionary proposed by these authors.

Another approach relies on machine learning methods for text classification. For example, Araci (2019) develop FinBERT, a language model adapted to financial texts, and report improvements in sentiment classification performance. Compared with dictionary-based approaches, FinBERT can use the context of words when classifying sentiment. However, better sentiment classification does not necessarily imply better forecasting performance. Even when sentiment is measured more accurately, it may contain little additional information beyond market variables. Moreover, aggregating individual sentiment scores into daily observations may remove some of the information contained in the original headlines.

The choice of forecasting model is another important consideration. In equity return prediction, Gu, Kelly, and Xiu (2020) find strong forecasting performance from tree-based models and neural networks, partly because of their ability to capture nonlinear interactions between predictive variables. This motivates comparing different forecasting models rather than relying on a single estimator. A sentiment variable may provide limited improvements in a linear model while containing information that can be captured through more complex relationships. Nevertheless, the performance of flexible models in equity markets does not necessarily establish their superiority in cryptocurrency forecasting.

Finally, the evaluation of forecasting performance must account for the additional estimation noise introduced by enriched models. In particular, Clark and West (2007) show that standard comparisons of mean squared forecast errors can be affected when comparing nested models. Their approach provides a framework for evaluating whether additional predictors improve forecasting performance relative to a more parsimonious benchmark. The implementation of the forecasting models and statistical tests considered in this study is presented in Section 5.

2.5 Position of the present study

The literature provides evidence that textual information can contribute to financial forecasting. However, the results differ across information sources, forecast targets, and estimation methods. Moreover, identifying a predictive relationship between sentiment and market outcomes does not necessarily imply that sentiment improves out-of-sample forecasts beyond the information already contained in market variables. Recent studies have examined this question using news-based sentiment indicators, although their results remain dependent on the empirical setting. This raises the question of whether the predictive contribution of media tone is consistent across cryptocurrencies, forecast targets, horizons, and forecasting models.

The present study examines this question using asset-specific news headlines for four cryptocurrencies. Two sentiment measures are compared with the same mar-

ket benchmark across several estimators and forecast horizons. The analysis also considers different sample periods and market volatility conditions to evaluate the stability of the results. The contribution is therefore a comparative assessment of the conditions under which news sentiment provides additional predictive information for cryptocurrency returns and volatility. The focus remains on out-of-sample forecasting performance rather than the identification of a causal relationship between media coverage and market outcomes.

3 Data

The dataset used in this study is a combination of two datasets: cryptocurrency market observations and asset-matched online news. This section describes the final dataset, the construction process, the forecast targets, predictive variables, and some descriptive statistics.

3.1 News collection and asset matching

The news database used in this study is collected from the Global Database of Events, Language, and Tone (GDELT) Project, which continuously monitors online news sources worldwide and provides structured access to news metadata and article content through its API. The news is collected using asset-specific keyword queries that include cryptocurrency names and symbols (for example, "BTC" and "BTCUSD" for Bitcoin) and common alternative spellings. Each query is made one UTC day at a time, with at most 250 returned records per asset. GDELT API returns are a mix of general and specialist online news, and all the returns are filtered again with the same set of lists. This way of collecting is designed to maximize the relevance and coverage of the news corpus. Finally, for each cryptocurrency, all articles published on day t are aggregated by mean into a single daily corpus from which sentiment indicators are computed.

The analysis covers Bitcoin (BTC), Ether (ETH), Solana (SOL), and XRP. Two samples are used: a common sample that begins on 18 December 2020 and allows comparisons over the date range for all assets, and a native sample corresponding to the whole available data history for each asset. The native sample begins on 5 September 2017 for BTC and ETH, 30 August 2020 for SOL, and 23 May 2018 for XRP. Table 1 summarizes the coverage of the different assets. Bitcoin receives by far the greatest media attention, with news available on 78% of days and an average of more than 50 articles per day. Ether follows with coverage of about 73% of days and 12 articles per day. Solana has articles available on approximately 64% of days, and XRP reaches nearly 69%.

Table 1: Market and matched-news coverage by asset

Asset	First date	Days	News days	Coverage	Mean/day
BTC	17 Aug 2017	3,059	2,381	77.8%	51.01
ETH	17 Aug 2017	3,059	2,226	72.8%	12.41
SOL	11 Aug 2020	1,969	1,258	63.9%	6.17
XRP	4 May 2018	2,799	1,942	69.4%	7.79

Note. *First date* is the first market observation. *News days* have at least one asset-matched headline. *Coverage* divides news days by all days; mean titles per day includes zero-news days. The panel is capped at 31 December 2025; usable forecast samples are shorter after feature and target alignment.

3.2 Market data and predictors

Cryptocurrency market data are obtained from the Binance platform using the CCXT API for the USDT trading pairs of the four assets. The variables collected are the common daily open, high, low, close, and volume for each asset. For each cryptocurrency, the sample spans the period from January 1, 2017, to December 31, 2025, with daily-frequency collection.

Let P_t denote the daily closing price of a given cryptocurrency on date t . The daily return is defined as:

$$r_t = \log(P_t) - \log(P_{t-1}), \quad (1)$$

The predictive variables used are chosen to cover distinct observable aspects of recent market conditions. They capture short-run return persistence or reversal, recent risk, intraday price dispersion, trading activity, and price trends. These groups of conditions correspond to the momentum, volatility, and liquidity-related signals often used in empirical machine learning applications to asset returns (Gu, Kelly, and Xiu 2020).

Return dynamics and persistence or reversal I include some return lags to capture the persistence of recent price movements at different delays (Liu and Tsyvinski 2021).

$$R_{\ell,t}^{\text{lag}} = r_{t-\ell}, \quad \ell \in \{1, 2, 3, 5, 10\}.$$

Recent risk and price dispersion.

- I include two variables to capture historical volatility measures: the observed standard deviation of return variability over 5 and 10 days. They are established inputs in cryptocurrency return analysis (Liu and Tsyvinski 2021;

Brauneis and Sahiner (2026). They are defined by

$$\text{RV}_{k,t}^{\text{past}} = \left\{ \frac{1}{k-1} \sum_{j=0}^{k-1} (r_{t-j} - \bar{r}_{k,t})^2 \right\}^{1/2}, \quad \bar{r}_{k,t} = \frac{1}{k} \sum_{j=0}^{k-1} r_{t-j}, \quad k \in \{5, 10\}. \quad (2)$$

- Another variable is the within-day price dispersion, which is used in the literature to capture the impact of price extremes (Parkinson 1980). I use the normalized measure:

$$\text{Range}_t = \frac{H_t - L_t}{P_t}.$$

- Finally, in this group, I use the width of the 20-day Bollinger bands to capture longer-window price dispersion relative to the prevailing price level. Using the original construction and its application to daily Bitcoin prices (Bollinger 2001; Gerritsen et al. 2020), the bands are placed two standard deviations above and below the moving average.

$$\text{BBWidth}_{20,t} = \frac{4 \text{sd}(P_t, \dots, P_{t-19})}{\text{SMA}_{20,t}}.$$

Price trend and momentum.

- Moving average returns are used in the literature to measure price trends in cryptocurrency analysis (BROCK, LAKONISHOK, and LeBARON 1992; Gerritsen et al. 2020). I use the 5-day and 10-day simple moving averages to summarize the recent level of the closing price over two horizons. They are computed as follows:

$$\text{SMA}_{k,t} = \frac{1}{k} \sum_{j=0}^{k-1} P_{t-j}, \quad k \in \{5, 10\}.$$

- The 14-day relative strength index compares smoothed positive and negative closing-price changes. This indicator follows the construction introduced by Wilder (1978) and has been applied to daily Bitcoin prices (Gerritsen et al. 2020). It is computed as

$$\text{RSI}_{14,t} = 100 - \frac{100}{1 + G_t/L_t^*},$$

where G_t and L_t^* are the exponentially smoothed positive and negative closing-price changes over 14 days. The notation L_t^* distinguishes the smoothed loss from the daily low price L_t .

All rolling and lagged quantities at date t use current or past data only. Observations made unavailable by lags or rolling windows are excluded from estimation.

3.3 News-based and sentiment variables

To measure news tone sentiment, I construct two complementary measures based on the headlines of the news. The first is built following the approach of Tetlock (2007), which is called lexicon sentiment. It is built on a project-specific vocabulary and a financial-language classifier for cryptocurrency news.

Lexicon sentiment This first sentiment measure is computed using a specific dictionary designed to identify the direction of the cryptocurrency market from news headlines. The dictionary is organized before any estimation or forecast evaluation. It is built around four recurring financial news categories: price direction and trading outlook; adoption and regulatory approval; security and legal events; and financial distress. A positive term means an upward market development or a favorable event, and a negative term means a downward market development or an adverse event. The dictionary was fixed independently of the results, so the outcomes cannot influence the obtained sentiment variable.

Table 2: Dictionary used to construct lexicon sentiment

Positive	surge, soar, soars, rally, rallies, bull, bullish, breakout, gain, gains, upside, record, ath, approval, approved, buy, accumulate, strong, beat, beats, growth, optimism, rebound, recovery, adoption, upgrade, inflows, partnership.
Negative	crash, plunge, dump, dumps, bear, bearish, selloff, sell-off, sell, risk, fear, hack, hacked, exploit, bankrupt, lawsuit, ban, banned, fraud, collapse, panic, liquidation, outflows, scam, investigation, downgrade, default.

Inflectional forms are also in the dictionary because the scoring relies on exact token matches: for example, gain and gains or rally and rallies. Headlines are converted to lower case and tokenized. Hyphenated terms are preserved. Concretely, for title i , let N_i^+ and N_i^- be the numbers of tokens that match entries in the positive and negative lists. Its lexicon score is

$$s_i^{\text{lex}} = \begin{cases} \frac{N_i^+ - N_i^-}{N_i^+ + N_i^-}, & N_i^+ + N_i^- > 0, \\ 0, & N_i^+ + N_i^- = 0, \end{cases} \quad (3)$$

Each positive term then contributes +1 and each negative term -1 before normalization. Every term has the same weight to preserve the lexicon as a transparent benchmark. I divide by the total number of matched terms to prevent a headline from receiving a more extreme score because it contains more polar words and to

yield a measure bounded by $[-1, 1]$. A title containing only positive matches then scores +1, one containing only negative matches scores -1, and mixed terms are placed between both limits according to the balance of matches. Titles without a matched term receive a zero score. Daily lexicon sentiment, S_t^{lex} , is the mean of s_i^{lex} across the titles that are matched to a given asset on day t .

FinBERT sentiment The second measure is computed from a pretrained BERT text classifier on a financial corpus called FinBERT from Araci (2019). The model is applied to each title without any fine-tuning. For each title i , it provides three class logits and converts them by a softmax function into positive, negative, and neutral probabilities. Denote by $z_{i,c}$ the classifier logit for class $c \in \{\text{positive}, \text{negative}, \text{neutral}\}$; the reported class probability is

$$p_{i,c} = \frac{\exp(z_{i,c})}{\exp(z_{i,+}) + \exp(z_{i,-}) + \exp(z_{i,0})}. \quad (4)$$

Denote by p_i^+ , p_i^- , and p_i^0 , respectively, the probabilities of positive, negative, and neutral classification. The sentiment of article i is computed as:

$$s_i^{\text{FB}} = p_i^+ - p_i^-. \quad (5)$$

For either sentiment measure, the daily sentiment is the average tone (sentiment) of the day's matched titles. Then, for either scoring method $m \in \{\text{lex}, \text{FB}\}$, the daily aggregation is

$$S_{a,t}^m = \frac{1}{N_{a,t}} \sum_{i \in \mathcal{A}_{a,t}} s_i^m \quad \text{if } N_{a,t} > 0, \quad S_{a,t}^m = 0 \quad \text{if } N_{a,t} = 0, \quad (6)$$

where $\mathcal{A}_{a,t}$ is the set of matched titles for asset a on day t and $N_{a,t}$ is its size. Days without news publications are set as neutral, meaning the value of the sentiment variables is set to 0.

4 Descriptive statistics

This section describes some descriptive statistics on key variables; the results are in Table 3. Table 3 presents the descriptive statistics of some key variables. There is substantial heterogeneity across the cryptocurrencies, both in their market characteristics and in the behavior of the sentiment variables.

Average daily returns are close to zero for all assets, but the dispersion differs. Bitcoin has the lowest daily return standard deviation (volatility), 0.04, whereas SOL has the highest one, 0.06. The backward-looking volatility and intraday-range

Table 3: Descriptive statistics by cryptocurrency

Asset	Variable	Mean	Std. dev.	Median
BTC	Daily return	0.00	0.04	0.00
	5-day return standard deviation	0.03	0.02	0.03
	Lexicon sentiment	0.07	0.16	0.03
	FinBERT sentiment	0.01	0.21	0.00
	Log volume	10.56	1.05	10.59
	Intraday range	0.05	0.04	0.04
ETH	Daily return	0.00	0.05	0.00
	5-day return standard deviation	0.04	0.03	0.03
	Lexicon sentiment	0.10	0.18	0.00
	FinBERT sentiment	0.07	0.21	0.01
	Log volume	12.80	1.06	12.95
	Intraday range	0.07	0.05	0.05
SOL	Daily return	0.00	0.06	0.00
	5-day return standard deviation	0.05	0.03	0.04
	Lexicon sentiment	0.06	0.16	0.00
	FinBERT sentiment	0.06	0.21	0.00
	Log volume	15.09	0.63	15.08
	Intraday range	0.09	0.06	0.08
XRP	Daily return	0.00	0.05	0.00
	5-day return standard deviation	0.04	0.03	0.03
	Lexicon sentiment	0.07	0.20	0.00
	FinBERT sentiment	0.03	0.22	0.00
	Log volume	19.18	0.99	19.26
	Intraday range	0.07	0.07	0.05

Note. The table uses each asset’s available daily sample through 31 December 2025. Daily return is the log return in Equation (1); the 5-day return standard deviation is the backward-looking measure in Equation (2). Sentiment variables are daily headline averages and follow the zero-news convention in Equation (6). Intraday range is $(H_t - L_t)/P_t$. Values are rounded to two decimal places.

statistics give the same ordering: the mean 5-day return standard deviation rises from 0.03 for BTC to 0.05 for SOL, and the mean intraday range rises from 0.05 to 0.09. ETH and XRP are in intermediate positions.

Trading activity introduces a different kind of variation. Mean log volume varies from 10.56 for BTC to 19.18 for XRP. However, the interpretation of these values as a ranking of market size should be done with caution: the value of an asset traded is recorded in different units, and prices differ. Changes in log volume within an asset remain useful for describing its own trading evolution.

The news variables add a second source of heterogeneity. Both headline tone measures have modestly positive means, from 0.06 to 0.010 for lexicon sentiment and from 0.01 to 0.07 for FinBERT sentiment. ETH has the highest mean lexicon score (0.1) and BTC the lowest one (0.01). These averages hide daily variation: the standard deviations are 0.16–0.20 for lexicon sentiment and 0.21–0.22 for FinBERT sentiment. In particular, the standard deviations of FinBERT sentiment are very close across all the assets, even though its average score differs. We can also remark that there are zero medians across the assets; this can also reflect days with no matched headline, neutral language, or offsetting tones within a day.

Figure 1 shows the evolution of 7-day rolling averages of the lexicon-based and FinBERT sentiment indicators for each cryptocurrency. The 7-day window is chosen just because it is easier to read; other horizons are available in the annex A. We can notice that sentiment measures fluctuate around zero throughout the sample, consistent with the near-zero averages observed in Table 3. The lexicon and FinBERT curves often move in the same direction, but their levels and individual peaks do not always agree. Thus, the two methods capture related but not identical aspects of headline tone.

5 Methodology

This section describe the methodology design, the models used and the forecast variables.

5.1 Research design and information sets

This subsection describes the research design used to evaluate whether sentiment information improves the forecasts compared with a model that uses only market information. The comparison is made separately for each cryptocurrency, forecast target, horizon, and estimator. I use the same set of market predictors, denoted by X_t , in all models and add one sentiment variable at a time. The three sets of

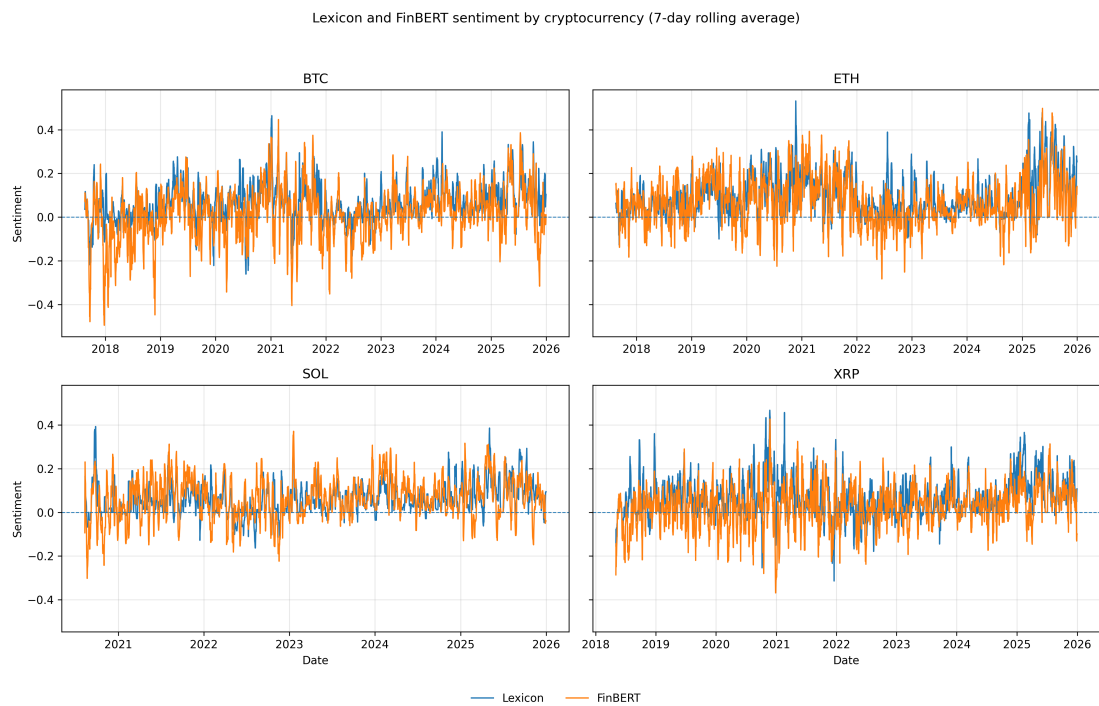


Figure 1: Seven-day rolling averages of Lexicon and FinBERT sentiment by cryptocurrency

predictive variables considered are:

$$\begin{aligned}\text{MARKET} &: X_t, \\ \text{MARKET+LEXICON} &: (X_t, S_t^{\text{lex}}), \\ \text{MARKET+FINBERT} &: (X_t, S_t^{\text{FB}}).\end{aligned}$$

For a given estimator, the model using only market information and the model including sentiment can be represented as follows:

$$\underbrace{Y_{t,h} = f(X_t) + u_{t,h}}_{\text{market benchmark}} \quad \text{versus} \quad \underbrace{Y_{t,h} = g(X_t, S_t) + v_{t,h}}_{\text{sentiment-enriched model}} \quad (7)$$

where S_t represents either S_t^{lex} or S_t^{FB} .

The only difference between the two models is the inclusion of the sentiment variable. The observations, forecast target, horizon, estimator, market predictors, and forecast dates remain the same. All predictive variables at date t are constructed using information available at the end of UTC day t , while the forecast target starts on day $t + 1$. This ensures that future market observations are not used to construct the predictors. However, it does not mean that the sentiment variables are exogenous, since the news published during day t may also be influenced by market conditions observed on the same day.

Endogeneity and causal interpretation. There are several reasons why the sentiment variables may be endogenous. First, changes in prices and volatility during day t can influence the news published on that same day. For example, an important increase in the price of a cryptocurrency may lead to more positive news about the asset. In this case, sentiment may reflect market movements rather than being their cause. Second, both sentiment and market conditions may be influenced by the same events, such as regulatory announcements, macroeconomic news, security incidents, or changes in market attention. Other unobserved factors, such as liquidity conditions, and measurement errors related to news matching or sentiment classification may also affect the relationship between sentiment and future market outcomes.

For these reasons, the empirical design does not impose the following conditional mean assumption:

$$E(v_{t,h} \mid X_t, S_t) = 0. \quad (8)$$

Although these sources of endogeneity cannot be completely eliminated, the methodology is designed to limit some of the problems they may create for forecasting. The forecast target starts on day $t + 1$, meaning that the sentiment variables do not contain information from future returns. The market benchmark includes

several variables that capture recent market conditions, such as returns, volatility, trading activity, price dispersion, and price trends. By comparing the model including sentiment with this benchmark, I evaluate whether sentiment contains additional information beyond what is already available from market variables. Data reprocessing, hyperparameter tuning, and model estimation are performed using only the information available before each forecast. The predictive performance is then evaluated on subsequent observations. This out-of-sample evaluation limits the risk of look-ahead bias and overfitting. However, it cannot eliminate the potential correlation between sentiment and unobserved factors that influence market conditions.

The objective of this study is therefore not to identify the causal effect of media sentiment on cryptocurrency markets, but to evaluate its contribution to forecasting returns and volatility. An improvement in forecasting performance may indicate that sentiment provides new information, reflects the reaction of the market to recent events, or captures other factors that are not directly observed in the market variables. No instrumental variable, natural experiment, or within-day event analysis is used to identify a causal relationship. Similarly, a statistically significant Clark-West test indicates that the inclusion of sentiment improves the predictive performance relative to the market benchmark, conditional on X_t . It does not establish that media sentiment causes subsequent market movements.

5.2 Forecast targets

I consider four forecast horizons, $h \in \{1, 7, 14, 30\}$, corresponding to one day, one week, two weeks, and approximately one month. Since cryptocurrency markets operate seven days a week, these horizons correspond to the same number of daily observations. For each horizon, I construct two forecast targets: cumulative returns and cumulative squared returns.

The cumulative log-return is defined as:

$$Y_{t,h}^r = \sum_{j=1}^h r_{t+j}, \quad (9)$$

The second target is the cumulative squared return, which is used as a measure of volatility over the forecast horizon. It is computed as follows:

$$Y_{t,h}^v = \sum_{j=1}^h r_{t+j}^2. \quad (10)$$

The sum of squared returns provides an observable measure of price variation over the forecast period for all cryptocurrencies. However, a single daily squared

return can be a noisy measure of the underlying volatility, particularly for $h = 1$. Using longer horizons allows the aggregation of several daily observations, which can reduce some of this noise. For both targets, the forecast period starts at $t + 1$ and ends at $t + h$.

For horizons greater than one day, consecutive forecast targets overlap because they include some of the same daily observations. This overlap is taken into account when constructing the estimation samples and performing the statistical tests. At each forecast date, the last $h - 1$ potential training observations are excluded, since their corresponding targets are not yet fully observed. Similarly, the construction of a target requires the next h daily returns, meaning that the last observations of the dataset cannot be used as forecast origins.

For the common sample, which ends on 31 December 2025, the last available forecast dates are 30 December, 24 December, 17 December, and 1 December 2025 for $h = 1, 7, 14$, and 30 , respectively.

5.3 Forecasting models

I use seven forecasting models to evaluate whether the contribution of sentiment information depends on the forecasting method. These models include linear regressions, regularized linear regressions, tree-based methods, and a neural network.

OLS and regularized linear models. The first model is Ordinary Least Squares (OLS), which provides a linear benchmark without regularization. I also consider three regularized linear models: Ridge, Lasso, and Elastic Net. These methods estimate a linear relationship between the predictive variables and the forecast target, but differ in the penalty applied to the coefficients.

Ridge uses an L_2 penalty, while Lasso uses an L_1 penalty. Elastic Net combines both penalties and is defined by the following optimization problem:

$$\min_{\beta_0, \boldsymbol{\beta}} \frac{1}{2n} \sum_{t=1}^n (Y_{t,h} - \beta_0 - Z'_t \boldsymbol{\beta})^2 + \alpha \left[\rho \|\boldsymbol{\beta}\|_1 + \frac{1 - \rho}{2} \|\boldsymbol{\beta}\|_2^2 \right], \quad (11)$$

where Z_t represents the vector of predictive variables, which can contain only market variables or both market and sentiment variables. The parameter α controls the regularization strength, while ρ determines the relative contribution of the L_1 and L_2 penalties. Lasso and Ridge are particular cases of this penalty structure.

The use of regularization is relevant in this application because some of the market variables may be correlated, particularly return lags, moving averages, and technical indicators. Regularization can reduce the variability of the estimated coefficients and improve forecasting performance. Moreover, Lasso and Elastic Net can reduce the influence of variables that provide little predictive information (Hastie, Tibshirani, and Friedman 2009).

Tree ensembles. I consider two tree-based methods: Random Forest and XGBoost. Random Forest, introduced by Breiman (2001), combines several regression trees constructed from bootstrap samples and random subsets of predictive variables. The final forecast is obtained by averaging the predictions of the individual trees.

For a given B trees, the Random Forest forecast is computed as:

$$\hat{Y}_{t,h}^{\text{RF}} = \frac{1}{B} \sum_{b=1}^B T_b(Z_t),$$

where $T_b(Z_t)$ represents the prediction of tree b . To limit overfitting, I impose a minimum number of observations in the terminal leaves. The values considered are 10, 20, and 40 observations, depending on the hyperparameter configuration selected during validation.

The second method is XGBoost, proposed by Chen and Guestrin (2016). Unlike Random Forest, which averages the predictions of several trees, XGBoost constructs trees sequentially. Each additional tree is used to improve the predictions of the previous ensemble by correcting its residual errors. The method also uses regularization, a learning rate, and subsampling of observations and predictive variables to control the complexity of the model.

The XGBoost forecast can be represented as:

$$\hat{Y}_{t,h}^{\text{XGB}} = \mu + \sum_{b=1}^{B^*} \eta T_b(Z_t),$$

where μ is the initial prediction, η is the learning rate, and B is the number of trees retained. The learning rate is selected during hyperparameter tuning, while the number of trees is determined using early stopping rather than being fixed at the maximum number allowed.

Both Random Forest and XGBoost can capture nonlinear relationships and interactions between predictive variables without requiring these relationships to be specified in advance.

Feed-forward neural network. The last model is a Multilayer Perceptron (MLP), which is a feed-forward neural network composed of an input layer, one or several hidden layers, and an output layer. I use the Rectified Linear Unit (ReLU) as the activation function in the hidden layers and a linear function in the output layer.

For each hidden layer ℓ , the transformation is defined as:

$$a^{(\ell)} = \max\{0, W^{(\ell)} a^{(\ell-1)} + b^{(\ell)}\},$$

where $a^{(0)} = Z_t$, $W^{(\ell)}$ is the matrix of weights, and $b^{(\ell)}$ is the vector of biases. The output of the last hidden layer is then used to produce the forecast through a linear transformation. The architectures considered are relatively small, with a decreasing number of neurons across successive hidden layers. This choice is motivated by the size of the available daily estimation samples. Using a large neural network with a limited number of observations may increase the risk of overfitting. The model is trained using the Adam optimizer, mini-batches, L_2 regularization of the weights, and early stopping. These procedures are used to control the training process and limit overfitting.

All seven models use the same predictive variables within a given information specification. The linear models capture linear relationships between the predictors and the target, while Random Forest, XGBoost, and MLP can also capture nonlinear relationships and interactions. This distinction is relevant because the predictive contribution of a variable may depend on how it interacts with other variables. In financial forecasting, the ability of machine learning models to capture nonlinear interactions has been identified as one of the reasons for their predictive performance (Gu, Kelly, and Xiu 2020). By considering these different models, I can evaluate whether sentiment information improves forecasts only under some model specifications or whether its contribution is also observed across different forecasting methods.

5.4 Preprocessing and hyperparameter selection

Before estimating the models, some variables are standardized to account for differences in their scales. The predictive variables are standardized for OLS, Ridge, Lasso, Elastic Net, and MLP. For Lasso, Elastic Net, and MLP, the forecast targets are also standardized. The predictions obtained from these models are then transformed back to their original scale before evaluating their forecasting performance. For Random Forest and XGBoost, both the predictive variables and the targets are used in their original units. To avoid using future information, the standardization parameters are estimated using only the corresponding training observations. During hyperparameter tuning, the scalars are fitted separately for each training fold. They are then re-estimated using the available expanding training sample during the final forecasting exercise. This ensures that observations from the validation or test periods do not influence the preprocessing of the training data.

The hyperparameters are selected using the initial 80% of the sample, while the remaining 20% is reserved for the final out-of-sample evaluation. Hyperparameter tuning is performed separately for each combination of sample, cryptocurrency, forecast target, horizon, information specification, and estimator. I use a five-fold expanding-window cross-validation procedure to select the hyperparameters. The observations are divided chronologically, without shuffling, to preserve the temporal

ordering of the data. At each fold, the model is estimated using past observations and evaluated on a subsequent validation period. Since the forecast targets overlap for horizons greater than one day, I introduce a gap of $h - 1$ observations between the training and validation periods. This prevents the training targets from containing returns that belong to the validation period. For each combination of hyperparameters, the model is estimated and evaluated across the five validation folds. The selection criterion is the average Root Mean Squared Error (RMSE) obtained over these folds. The configuration with the lowest average validation RMSE is selected.

This procedure follows the principle of evaluating model complexity using observations that were not used for estimation, as discussed by Gu, Kelly, and Xiu (2020). It allows the selection of hyperparameters based on forecasting performance while keeping the final test sample independent of the selection process.

Let $\mathcal{G}$ denote the set of candidate hyperparameter configurations. The selected configuration is defined as:

$$\hat{\theta} = \arg \min_{\theta \in \mathcal{G}} \frac{1}{5} \sum_{k=1}^5 \left\{ \frac{1}{|\mathcal{V}_k|} \sum_{t \in \mathcal{V}_k} (Y_{t,h} - \hat{Y}_{t,h}^{(\theta,k)})^2 \right\}^{1/2}, \quad (12)$$

where $\mathcal{V}_k$ is the validation sample of fold k , and $\hat{Y}_{t,h}^{(\theta,k)}$ is the forecast obtained from a model estimated using only the corresponding training observations. Once the hyperparameters are selected, they remain fixed throughout the final 20% out-of-sample period. However, the models are re-estimated as new observations become available. This means that the regression coefficients, trees, and neural network weights are updated using an expanding training window, while the selected hyperparameters remain unchanged. The hyperparameters are selected separately for the market benchmark and the models including sentiment. However, for a given estimator, the same candidate hyperparameter grid and validation folds are used. This ensures that the comparison between the models is based on the same validation procedure.

For XGBoost and MLP, I also use early stopping to limit overfitting during model estimation. The last 10% of the current training sample is reserved for this purpose. These observations are not used to fit the model parameters and are separated from the model-fitting observations by an additional gap of $h - 1$ observations. For XGBoost, the training process stops when the validation performance does not improve for 30 consecutive rounds. For MLP, the maximum number of training epochs is fixed at 200, with early stopping after 10 epochs without improvement. I use a fixed random seed of zero for Random Forest, XGBoost, and MLP to make the estimation results reproducible. Table 4 presents the candidate values and fixed parameters used for each model. For the regularized linear models, the penalty parameters cover several orders of magnitude to allow different levels of regularization. For the tree-based models, I consider different tree depths and minimum leaf

sizes to control model complexity and reduce the risk of overfitting. For MLP, the candidate architectures are limited to relatively small networks, given the number of daily observations available for estimation. The objective is to consider different levels of model complexity without making the hyperparameter search unnecessarily large. The final selection is based on validation performance rather than on the fit obtained from the training sample, following the machine learning forecasting approach discussed by Gu, Kelly, and Xiu (2020).

Table 4: Hyperparameter grids and fixed model settings

Estimator	Tuned parameters	Fixed settings
OLS	None	Intercept included
Ridge	$\alpha \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100, 1000\}$	Intercept included
Lasso	$\alpha \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$	Intercept; cyclic updates; 10,000-iteration cap
Elastic Net	$\alpha \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$; $\rho \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$	Intercept; cyclic updates; 10,000-iteration cap
Random Forest	maximum depth $\in \{3, 5, 8\}$; minimum leaf size $\in \{10, 20, 40\}$; predictors per split $\in \{\sqrt{p}, 0.5p\}$	300 trees; bootstrap sampling; seed 0
XGBoost	maximum depth $\in \{2, 3, 5\}$; learning rate $\in \{0.01, 0.05, 0.10\}$; minimum child weight $\in \{1, 5, 10\}$	At most 2,000 trees; row and column subsampling 0.8; $L_1 = 0$; $L_2 = 1$; 30-round early stopping; seed 0
MLP	hidden layers $\in \{(32), (32, 16), (32, 16, 8)\}$; weight decay $\in \{10^{-5}, 10^{-4}, 10^{-3}\}$; learning rate $\in \{10^{-4}, 10^{-3}\}$	ReLU; Adam; batch size 64; 200-epoch cap; patience 10; seed 0

Note. p is the number of predictors. The XGBoost tree count is a maximum; early stopping determines the effective number of trees in each fit.

For each sample design, the hyperparameter selection procedure is performed 96 times for each estimator. This corresponds to the combination of four cryptocurrencies, two forecast targets, four horizons, and three information sets ($4 \times 2 \times 4 \times 3 = 96$). Table 5 summarizes the hyperparameters selected during the final forecasting exercise. For each estimator, the table reports the most frequently selected combination of hyperparameters and the number of times this combination was selected. The last column presents all the values that were selected at least once. The results show that the selected hyperparameters vary across cryptocurrencies, forecast targets, and horizons. However, some configurations are selected more frequently than

others. For example, Ridge most often selects the largest regularization parameter considered in the grid. Similarly, Random Forest frequently selects relatively shallow trees with a large minimum number of observations in the terminal leaves. These results indicate that regularization and the control of model complexity play an important role in the forecasting exercise.

5.5 Out-of-sample forecasting design

The forecasting exercise is conducted using an expanding-window approach. After excluding observations that are unavailable because of the construction of the predictive variables or forecast targets, I divide each sample into two parts. The first 80% of the observations are used as the initial estimation sample, while the remaining 20% are reserved for the out-of-sample evaluation. The division is chronological, without any random selection of observations.

The forecasts are generated in consecutive blocks of 14 observations. I use blocks of 14 observations to reduce the computational cost associated with repeated model estimation, particularly for the machine learning models. This choice allows the estimation sample to be updated regularly without requiring the models to be re-estimated at every forecast date. Before each block, the model is re-estimated using an expanding window that includes all the observations whose forecast targets are already available at that date. The procedure can be represented as follows:

$$\underbrace{1, \dots, R}_{\text{initial estimation}} \longrightarrow \underbrace{R + 1, \dots, R + 14}_{\text{forecast block}} \longrightarrow \text{expand, refit, and repeat.}$$

After each block of 14 forecasts, the estimation window is expanded to include the new observations for which the targets have become available. The model is then re-estimated before generating the next block of forecasts. This procedure is repeated until the end of the test sample. The last block may contain fewer than 14 observations. The hyperparameters selected during the cross-validation procedure remain fixed throughout the out-of-sample evaluation. However, the parameters estimated by the models, such as regression coefficients, regression trees, or neural network weights, are updated at each re-estimation. This allows the models to incorporate new information without repeating the hyperparameter selection procedure.

To ensure that the comparison between the market benchmark and the models including sentiment is valid, all forecasts within a given comparison must correspond to the same dates and the same realized targets. The forecast results are therefore matched by date, and the comparison is considered invalid if some dates are missing, duplicated, or associated with different observed target values. No random division or shuffling of observations is used during the estimation, hyperparameter selection, or out-of-sample forecasting procedures.

Table 5: Hyperparameters selected in the definitive runs

Estimator	Sample	Most frequent joint setting (count out of 96)	Values selected at least once
Ridge	Common	$\alpha = 1000$ (78)	$\alpha \in \{10, 100, 1000\}$
Ridge	Native	$\alpha = 1000$ (87)	$\alpha \in \{0.1, 1, 10, 100, 1000\}$
Lasso	Common	$\alpha = 1$ (48)	$\alpha \in \{0.01, 0.1, 1\}$
Lasso	Native	$\alpha = 1$ (51)	$\alpha \in \{0.01, 0.1, 1\}$
Elastic Net	Common	$(\alpha, \rho) = (1, 0.50)$ (24)	$\alpha \in \{0.01, 0.1, 1\}$; all five candidate ρ values
Elastic Net	Native	$(\alpha, \rho) = (1, 0.50)$ (24)	$\alpha \in \{0.1, 1\}$; all five candidate ρ values
Random Forest	Common	depth 3, leaf 40, $\sqrt{p}$ (51)	all depths; all leaf sizes; both feature-sampling rules
Random Forest	Native	depth 3, leaf 40, $\sqrt{p}$ (47)	all depths; all leaf sizes; both feature-sampling rules
XGBoost	Common	depth 5, rate 0.01, child weight 10 (9)	every candidate value of all three tuned parameters
XGBoost	Native	depth 2, rate 0.01, child weight 5 (14)	every candidate value of all three tuned parameters
MLP	Common	(32, 16, 8), decay 10^{-3} , rate 10^{-4} (32)	all architectures, decay values, and learning rates
MLP	Native	(32, 16, 8), decay 10^{-3} , rate 10^{-4} (14)	all architectures, decay values, and learning rates

Note. Common denotes the common 2020-2025 calendar sample; Native denotes the asset-specific longer samples. OLS has no selected hyperparameters. The exact setting and inner-validation RMSE for every asset-target-horizon-information-set combination are archived with the forecast results.

5.6 Forecast evaluation and inference

The forecasting performance of the models is evaluated using the Root Mean Squared Error (RMSE). For a given forecast $\hat{Y}_{t,h}$, the forecast error is defined as:

$$e_{t,h} = Y_{t,h} - \hat{Y}_{t,h}.$$

The RMSE is then computed over the P out-of-sample observations as follows:

$$\text{RMSE}_h = \left(\frac{1}{P} \sum_{t=1}^P e_{t,h}^2 \right)^{1/2}. \quad (13)$$

To evaluate the contribution of sentiment information, I compare the RMSE of each sentiment-enriched model with that of the corresponding market benchmark. The percentage improvement in RMSE is defined as:

$$\Delta \text{RMSE}(\%) = 100 \frac{\text{RMSE}_M - \text{RMSE}_E}{\text{RMSE}_M}, \quad (14)$$

where M represents the model using only market variables and E the model including sentiment information. A positive value indicates that the inclusion of sentiment reduces the forecasting error compared with the market benchmark. A negative value means that the market benchmark provides more accurate forecasts. However, a reduction in RMSE does not necessarily mean that the improvement is statistically significant. For this reason, I also use the Clark-West test to evaluate whether sentiment information provides additional predictive content.

5.6.1 Clark-West inference for nested forecasts

The comparison between the market benchmark and the sentiment-enriched model corresponds to a comparison between nested models. The market benchmark is nested within the enriched model because the latter includes the same market variables and one additional sentiment variable. When the predictive contribution of sentiment is zero, the enriched model reduces to the market benchmark.

This characteristic is important when testing differences in forecasting performance. Under the null hypothesis that sentiment does not provide additional predictive information, the enriched model estimates additional parameters whose population values are zero. The estimation of these parameters introduces additional estimation noise, which can increase the mean squared forecast error of the enriched model. Consequently, a direct comparison of the mean squared errors may favor the market benchmark even when both models have the same predictive accuracy in population. To account for this problem, I use the test proposed by Clark and West (2007), which adjusts the difference in squared forecast errors.

Let $\hat{Y}_{M,t}$ and $\hat{Y}_{E,t}$ denote the forecasts obtained from the market benchmark and the sentiment-enriched model, respectively. Their corresponding forecast errors are defined as:

$$e_{M,t} = Y_{t,h} - \hat{Y}_{M,t}, \quad e_{E,t} = Y_{t,h} - \hat{Y}_{E,t}.$$

The adjusted difference in squared forecast errors is computed as follows:

$$f_t = e_{M,t}^2 - e_{E,t}^2 + (\hat{Y}_{M,t} - \hat{Y}_{E,t})^2. \quad (15)$$

The first two terms represent the difference between the squared forecast errors of the market benchmark and the sentiment-enriched model. The last term is the adjustment introduced by Clark and West to account for the additional estimation noise associated with the enriched model. A positive average adjusted difference therefore indicates that the model including sentiment performs better than the market benchmark after taking into account this adjustment.

For P out-of-sample forecasts, the average adjusted difference is defined as:

$$\bar{f} = \frac{1}{P} \sum_{t=1}^P f_t.$$

To conduct the statistical test, it is also necessary to estimate the variance of this adjusted difference. However, for forecast horizons greater than one day, the targets overlap because they contain some of the same daily returns. This overlap may introduce serial correlation in the forecast errors and consequently in the adjusted loss differential.

To account for this serial correlation, I use a Heteroskedasticity and Autocorrelation Consistent (HAC) variance estimator with Bartlett weights. First, the sample autocovariance at lag ℓ is computed as:

$$\hat{\gamma}_\ell = \frac{1}{P} \sum_{t=\ell+1}^P (f_t - \bar{f})(f_{t-\ell} - \bar{f}). \quad (16)$$

The HAC long-run variance is then estimated as follows:

$$\hat{\Omega}_f = \hat{\gamma}_0 + 2 \sum_{\ell=1}^L \left(1 - \frac{\ell}{L+1}\right) \hat{\gamma}_\ell, \quad L = \min\{h-1, P-1\}. \quad (17)$$

The number of lags depends on the forecast horizon. For the one-day horizon, $L = 0$, and the estimator reduces to the variance of the adjusted difference. For longer horizons, additional autocovariances are included to account for the overlap between consecutive forecast targets. I also apply a finite-sample correction by

multiplying the estimated HAC variance by $P/(P-1)$. The Clark-West test statistic is then computed as:

$$CW = \frac{\sqrt{P-1} \bar{f}}{\sqrt{\hat{\Omega}_f}}. \quad (18)$$

The objective is to test whether the inclusion of sentiment information improves forecasting performance relative to the market benchmark. I therefore consider the following one-sided hypotheses:

$$H_0 : E(f_t) \leq 0 \quad \text{against} \quad H_1 : E(f_t) > 0.$$

The corresponding p -value is obtained from the upper tail of the standard normal distribution:

$$p = 1 - \Phi(CW),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

This procedure is equivalent to testing the intercept in a regression of f_t on a constant using the same Bartlett HAC covariance estimator. A large positive value of the Clark-West statistic provides evidence in favor of the model including sentiment information. In contrast, a negative value does not support an improvement relative to the market benchmark. For each test, I use forecasts obtained on identical dates and with the same observed target values. Since the analysis considers several cryptocurrencies, forecast targets, horizons, estimators, and sentiment measures, a large number of comparisons are performed. The reported p -values are not adjusted for multiple testing. I do not interpret statistical significance in isolation, but also consider the magnitude of the RMSE improvements and the consistency of the results across the different model specifications.

Volatility-regime evaluation. In addition to the full-sample evaluation, I examine whether the contribution of sentiment information depends on market volatility conditions. The objective is to determine whether sentiment provides more predictive information during periods of relatively high or low volatility. In this analysis, a volatility regime refers to a market condition defined using the recent observed volatility of a cryptocurrency. It does not correspond to an unobserved regime estimated by a statistical model or to a structural change in the market. To identify these regimes, I use the trailing 10-day standard deviation of daily returns, $RV_{10,t}^{\text{past}}$, previously defined in Equation (2). This variable measures recent return variability using information available at the forecast date. Compared with a single daily squared return, the use of a 10-day window makes the measure less sensitive to an isolated daily observation. At the same time, it allows recent changes in market volatility to be captured.

The choice of this variable is also motivated by the fact that it is already included among the market predictors. Therefore, the identification of volatility regimes does not require the construction of an additional variable using future market observations. The 10-day window is fixed before the forecast evaluation and provides a relatively short-term measure of market volatility. I define three volatility regimes using the 25th and 75th percentiles of the historical distribution of this variable. For each cryptocurrency a and sample design, let $q_{a,0.25}$ and $q_{a,0.75}$ denote these two percentiles, respectively.

The volatility regime at date t is defined as:

$$\mathcal{R}_{a,t} = \begin{cases} \text{low volatility,} & \text{RV}_{10,t}^{\text{past}} \leq q_{a,0.25}, \\ \text{central volatility,} & q_{a,0.25} < \text{RV}_{10,t}^{\text{past}} < q_{a,0.75}, \\ \text{high volatility,} & \text{RV}_{10,t}^{\text{past}} \geq q_{a,0.75}. \end{cases} \quad (19)$$

Observations below the 25th percentile are classified as low-volatility periods, while those above the 75th percentile are classified as high-volatility periods. The remaining observations correspond to the central volatility regime. The analysis focuses on the low- and high-volatility periods, while the observations in the central regime remain included in the full-sample forecast evaluation.

The use of quartiles allows the identification of relatively distinct volatility conditions while retaining a sufficient number of observations in each group. In the calibration sample, approximately 25% of the observations are assigned to each extreme regime. Using narrower parts of the distribution would reduce the number of observations available for the statistical tests. The percentiles are also computed separately for each cryptocurrency. This is important because the level of volatility differs substantially across assets, as shown in the descriptive statistics. A common numerical threshold could therefore represent different market conditions depending on the cryptocurrency.

By using asset-specific percentiles, high and low volatility are defined relative to the historical volatility distribution of each asset. The use of volatility quantiles to distinguish market conditions is also found in the literature. For example, Clarida, Davis, and Pedersen (2009) examine relationships in currency markets under different volatility conditions and identify high-volatility periods using the upper quartile of the volatility distribution. To avoid using information from the test sample when defining the regimes, the percentiles are estimated using only the initial calibration period. For each cryptocurrency and sample design, this period starts at the beginning of the available sample and ends on the day preceding the first out-of-sample forecast. Once estimated, the percentile thresholds remain fixed throughout the forecasting exercise. The same thresholds are used across forecast targets, horizons, and estimators for a given cryptocurrency and sample design. This ensures that future observations do not influence the classification thresholds. Moreover, the

regime assigned to a forecast date t depends only on volatility information available at that date.

After classifying the forecast observations, I evaluate the models separately for the low- and high-volatility regimes. For each regime, the RMSE and Clark-West statistic are recomputed using only the corresponding out-of-sample observations. As in the full-sample analysis, the market benchmark and the sentiment-enriched model are compared using the same dates and realized forecast targets. However, the regime-specific evaluation uses fewer observations than the full-sample analysis. In addition, the observations within a given regime are not necessarily consecutive. For these reasons, the estimated improvements and the corresponding unadjusted p -values are interpreted with more caution. A reduction in RMSE indicates that sentiment improves forecasting accuracy under the corresponding volatility conditions. When this improvement is also supported by a positive and statistically significant Clark-West statistic, it provides evidence of additional predictive content relative to the market benchmark.

The results are interpreted separately for each cryptocurrency, forecast target, horizon, and estimator. They therefore describe the predictive contribution of sentiment under the observed news coverage and the selected measure of news tone. Finally, comparing the results obtained with Lexicon and FinBERT sentiment allows me to examine whether the contribution of sentiment information also depends on the method used to extract sentiment from news headlines.

6 Empirical Results

This section presents the results of the forecasting exercise and evaluates whether news sentiment improves the prediction of cryptocurrency returns and volatility. As previously described, news sentiment refers to the tone of the headlines associated with each cryptocurrency. The sentiment variables therefore measure media tone rather than investor sentiment directly.

I first present the results obtained from the common sample, which allows comparisons across the four cryptocurrencies over the same period. The asset-native samples are then considered to examine whether the results depend on the sample period. In this analysis, an improvement is considered statistically supported when the model including sentiment has a lower RMSE than the market benchmark and the Clark-West test is significant at the 5% level. This criterion combines the improvement in forecasting accuracy defined in Equation (14) with the statistical test presented in Equation (18).

The common-sample evaluation contains 368 forecast observations at the one-day horizon and 362 at the thirty-day horizon, covering the period from late 2024 through 2025. A total of 224 comparisons are performed for each forecast target. In

return forecasting, the inclusion of sentiment reduces the RMSE in 54 comparisons, of which 13 are statistically supported. The results are more favorable for volatility, with 67 RMSE improvements and 38 statistically supported gains. The Clark-West test is significant in 42 volatility comparisons, including four cases where the RMSE increases.

These results suggest that sentiment information contributes more to volatility forecasting than to return forecasting. However, the majority of the specifications do not produce statistically supported improvements. Moreover, these counts summarize the individual comparisons rather than representing a joint statistical test.

Table 6 summarizes the results of all comparisons. The detailed tables report only the estimator-sentiment combinations for which at least one Clark-West p -value is below 10%, while the summary table includes all the comparisons performed.

Table 6: Overview of the incremental forecasting contribution of sentiment in the common sample.

Target	Pairs	Lower RMSE	Unchanged RMSE	CW $p < 0.05$	Lower RMSE and CW $p < 0.05$
Cumulative return	224	54	64	13	13
Cumulative squared return	224	67	52	42	38

Notes: Counts cover four assets, four horizons, seven estimators and two sentiment indicators per target. Lower RMSE denotes a percentage improvement greater than 10^{-9} ; unchanged RMSE denotes an absolute percentage difference no greater than 10^{-9} . CW denotes the one-sided Clark-West test with Bartlett HAC standard errors and $h - 1$ lags. Its adjusted statistic can be significant even when raw RMSE increases. Undefined tests for zero-variance differentials do not count as rejections. P-values are unadjusted, and these counts are descriptive rather than a joint test.

6.1 Return predictability

Table 7 presents the return forecasting results obtained from the common sample.

Overall, the results show that news sentiment provides limited improvements in return forecasting. None of the one-day comparisons produces a statistically supported gain in the common sample, while the most important improvements are generally observed at longer forecast horizons.

The largest improvements for Bitcoin are obtained with MLP at the thirty-day horizon, with RMSE reductions of 7.67% and 7.24% using FinBERT and lexicon sentiment, respectively. Both improvements are statistically significant, with Clark-West p -values of 0.0084 and 0.0117. Random Forest also provides statistically supported gains at the same horizon, although their magnitude is considerably smaller, at approximately 0.6%. With XGBoost, the improvements are around 0.45-0.50%

Table 7: Out-of-sample RMSE improvements (%) of sentiment-augmented models relative to market-only benchmarks for return prediction (Clark-West test) using the common 2020-2025 sample.

Model	$h = 1$	$h = 7$	$h = 14$	$h = 30$
BTC				
MLP FinBERT	-0.35	-0.81	-2.35	7.67***
MLP Lexicon	-0.95	-0.53	-2.30	7.24**
RF FinBERT	0.04	-0.10	-0.14	0.63**
RF Lexicon	-0.03	-0.15	-0.22	0.62**
XGB FinBERT	0.08	-0.15	0.45**	0.63*
XGB Lexicon	0.10	0.92**	0.49**	0.72
ETH				
XGB Lexicon	-0.32	2.10***	-1.56	-2.22
SOL				
MLP FinBERT	-0.48	-1.29	3.56**	8.96***
MLP Lexicon	-0.56	-1.86	2.88**	11.02***
XRP				
MLP FinBERT	0.03	0.68*	-7.29	7.17
OLS FinBERT	0.05	-0.06	0.32*	0.58**
Ridge FinBERT	0.03	-0.09	0.04	0.21*

Notes: Positive values indicate lower RMSE for the sentiment-augmented model. Stars report one-sided upper-tail Clark-West significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Only rows with at least one significant horizon are shown. A significant Clark-West result can accompany a negative raw RMSE change because the test adjusts for parameter-estimation noise.

at fourteen days and 0.92% at seven days when lexicon sentiment is included. Thus, sentiment information can improve Bitcoin return forecasts, although its contribution depends on the estimator and forecast horizon.

Solana exhibits the largest improvements among the four cryptocurrencies in the common sample. At the thirty-day horizon, MLP reduces the RMSE by 8.96% with FinBERT and 11.02% with lexicon sentiment, with corresponding Clark-West p -values of 0.0022 and 0.0020. Improvements are also observed at fourteen days, reaching 3.56% and 2.88%, respectively.

The results are less favorable for Ether and XRP. Only one statistically supported improvement is observed for Ether, with XGBoost and lexicon sentiment reducing the seven-day RMSE by 2.10%. Similarly, XRP has only one supported gain, ob-

tained with OLS and FinBERT at the thirty-day horizon, with an improvement of 0.58%. Although MLP produces larger RMSE reductions of approximately 7% for XRP at the same horizon, these improvements are not statistically significant at the 5% level.

The regularized linear models provide different results. In all common-sample return comparisons, Lasso and Elastic Net produce forecasts identical to those of their respective market benchmarks. This means that including sentiment does not change their predictions or realized forecast errors. However, this does not necessarily indicate that sentiment contains no predictive information. It only shows that its inclusion does not affect the forecasts obtained under the selected model configurations, without providing information about which coefficients were selected.

To examine the sensitivity of these results to the sample period, I also consider the asset-native samples. The number of statistically supported return improvements is relatively similar, with eleven gains compared with thirteen in the common sample. However, their distribution across cryptocurrencies changes substantially. Bitcoin and Solana have no supported improvements in their native samples, while Ether and XRP have seven and four, respectively. Only the XGBoost model with lexicon sentiment for Ether at the seven-day horizon provides a statistically supported improvement in both sample designs. The corresponding RMSE reductions are 2.10% in the common sample and 1.43% in the native sample.

The relationship between forecast horizon and the contribution of sentiment also varies across samples. In the native sample of Ether, the improvements obtained with MLP and FinBERT decrease as the forecast horizon increases. The RMSE reductions are 1.32% and 1.05% at one and seven days, respectively, with both results statistically supported. The improvement then decreases to 0.36% at fourteen days before becoming negative at thirty days, with a value of -1.15% . Neither of these last two results is statistically supported. A similar pattern is observed with lexicon sentiment, where the RMSE improvements decrease from 1.46% at one day to 1.35% at seven days and 0.45% at fourteen days, before becoming negative at thirty days (-0.33%). These results suggest that the contribution of sentiment to Ether return forecasts is more favorable at short horizons in the native sample. In contrast, Bitcoin exhibits a different pattern in the common sample. With MLP and FinBERT, including sentiment increases the RMSE at horizons up to fourteen days, while an improvement of 7.67% is obtained at thirty days. Therefore, the results do not suggest a common relationship between forecast horizon and the predictive contribution of sentiment. Instead, this contribution depends on the cryptocurrency, estimator, and sample period.

Overall, the results indicate that news sentiment can provide additional predictive information for cryptocurrency returns under some model specifications. The largest improvements in the common sample are generally observed at longer hori-

zons, particularly for Bitcoin and Solana with MLP. However, these gains are not consistently observed in the asset-native samples, suggesting that the forecasting contribution of sentiment is sensitive to the sample period. Finally, these findings concern the incremental predictive content of the sentiment measures considered in this study. They do not establish the informational efficiency of cryptocurrency markets or identify how rapidly news information is incorporated into prices.

6.2 Volatility predictability

Table 8 presents the volatility forecasting results obtained from the common sample. Overall, the results are more favorable for volatility forecasting than for returns, with the largest improvements generally obtained with MLP. In the common sample, 21 of the 32 MLP comparisons produce lower RMSE and statistically significant Clark-West tests at the 5% level. These represent 21 of the 38 statistically supported volatility improvements, while the remaining gains are distributed across the other estimators. This suggests that model flexibility may play an important role in capturing the predictive information contained in sentiment. The largest improvements are observed for XRP, particularly with MLP. The inclusion of FinBERT sentiment reduces the RMSE by 22.44%, 30.22%, 29.33%, and 56.38% at horizons of one, seven, fourteen, and thirty days, respectively. Similar improvements are obtained with lexicon sentiment, reaching 23.38%, 30.14%, 29.58%, and 56.33% over the same horizons. All eight comparisons are statistically significant, with Clark-West p -values below 1%. Random Forest also provides statistically supported improvements at seven, fourteen, and thirty days, although their magnitude is considerably smaller, ranging from approximately 1.2% to 2.3%.

However, the magnitude of the improvements obtained with MLP requires some caution. At the thirty-day horizon, the market-only MLP has an RMSE of 0.12338, compared with 0.05382 when FinBERT sentiment is included. Although this represents a substantial improvement, the market-only Lasso achieves an even lower RMSE of 0.05146. Therefore, the inclusion of sentiment considerably improves the performance of MLP relative to its market benchmark, but does not necessarily make it the best forecasting model. In particular, a large improvement relative to a less accurate benchmark should not be interpreted as evidence of overall model superiority.

Solana also benefits from sentiment information, although the improvements are generally smaller than those observed for XRP. The largest MLP gains are obtained at the fourteen-day horizon, where FinBERT and lexicon sentiment reduce the RMSE by 10.62% and 10.84%, respectively. At seven days, the corresponding improvements are 2.31% and 1.02%, while at thirty days they reach 2.19% and 2.37%. The one-day results are less favorable, with a slight deterioration when FinBERT is included and an improvement of only 0.30% with lexicon sentiment.

Table 8: Out-of-sample RMSE improvements (%) of sentiment-augmented models relative to market-only benchmarks for volatility prediction (Clark-West test) using the common 2020-2025 sample.

Model	$h = 1$	$h = 7$	$h = 14$	$h = 30$
BTC				
ElasticNet FinBERT	0.20***	0.00	0.00	-0.01
MLP FinBERT	-0.98***	-12.71**	0.82**	15.33***
MLP Lexicon	-0.96***	-12.80**	0.84**	15.39***
OLS FinBERT	0.04	0.17**	0.14**	-0.03
RF FinBERT	0.43*	-0.92	-0.53	-0.43
RF Lexicon	0.60*	-1.50	-0.63	-0.59
Ridge FinBERT	0.10	0.22**	0.15*	-0.01
XGB Lexicon	0.81*	-1.22	-2.88	-2.67
ETH				
Lasso FinBERT	1.27***	0.00	0.00	0.00
MLP FinBERT	0.13*	-2.63	-1.23	12.42***
MLP Lexicon	-0.14	-2.55	-1.28	11.32***
SOL				
Lasso Lexicon	3.83***	-0.37	0.00	0.00
MLP FinBERT	-0.11*	2.31***	10.62***	2.19**
MLP Lexicon	0.30**	1.02***	10.84***	2.37**
OLS FinBERT	-0.81	-0.17	1.63***	0.59**
Ridge FinBERT	-0.03	-0.11	1.75***	0.97**
XGB FinBERT	-3.76	-0.53	1.85***	-1.72
XRP				
MLP FinBERT	22.44***	30.22***	29.33***	56.38***
MLP Lexicon	23.38***	30.14***	29.58***	56.33***
RF FinBERT	-0.75	2.31***	1.28***	2.27**
RF Lexicon	-0.73	2.10***	1.23**	2.01**
Ridge FinBERT	0.00	0.09*	0.07*	-0.02

Notes: Positive values indicate lower RMSE for the sentiment-augmented model. Stars report one-sided upper-tail Clark-West significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Only rows with at least one significant horizon are shown. A significant Clark-West result can accompany a negative raw RMSE change because the test adjusts for parameter-estimation noise.

Improvements are also observed with other estimators. At the one-day horizon, Lasso with lexicon sentiment reduces the RMSE by 3.83%. Similarly, the inclusion of FinBERT provides statistically supported improvements of 1.63%, 1.75%, and

1.85% at fourteen days with OLS, Ridge, and XGBoost, respectively.

Bitcoin and Ether also exhibit important improvements at the thirty-day horizon. With MLP, FinBERT and lexicon sentiment reduce Bitcoin’s RMSE by 15.33% and 15.39%, respectively. The corresponding reductions for Ether are 12.42% and 11.32%. All four comparisons have Clark-West p -values below 1%. Nevertheless, these improvements are not observed at all horizons. Bitcoin’s MLP forecasts become less accurate when sentiment is included at one and seven days, while Ether exhibits similar deteriorations at seven and fourteen days. These differences suggest that the contribution of sentiment depends on the forecast horizon and cryptocurrency.

The results obtained from the asset-native samples provide additional information about the stability of these improvements. Compared with return forecasting, a larger number of volatility improvements remain statistically supported across both sample designs. In total, fifteen comparisons involving the same cryptocurrency, horizon, estimator, and sentiment measure are statistically supported in both samples, compared with only one for return forecasting. Some of the largest improvements remain substantial. For example, Ether’s thirty-day MLP gains are still approximately 8.5-8.8%, while Solana retains improvements close to 6% at fourteen days. Similarly, XRP continues to exhibit important thirty-day gains of approximately 31%. However, other results change considerably. Bitcoin’s thirty-day MLP improvements observed in the common sample become RMSE deteriorations of approximately 20-28% in the native sample. A similar reversal is observed for XRP at the one-day horizon. At seven days, the native-sample MLP with FinBERT still provides a small improvement for XRP, while lexicon sentiment increases the RMSE by 34.97%. These differences indicate that the contribution of sentiment to volatility forecasting is also sensitive to the sample period, even though more improvements remain statistically supported across both designs than in return forecasting.

The relationship between forecast horizon and predictive performance also differs across cryptocurrencies. Solana provides an interesting example, with MLP improvements concentrated around intermediate horizons in the common sample. Although the gains at seven, fourteen, and thirty days are statistically supported, the largest reductions in RMSE are observed at fourteen days. The positive improvements at thirty days nevertheless indicate that the predictive contribution of sentiment does not disappear at longer horizons. This pattern is not entirely preserved in the native sample. With FinBERT, Solana’s MLP improvement increases from 5.92% at fourteen days to 11.31% at thirty days. XRP exhibits a different pattern in the common sample, with slightly smaller improvements at fourteen days followed by substantially larger gains at thirty days. Together with the results obtained for Bitcoin, these differences do not suggest a common horizon beyond which sentiment information becomes less useful for volatility forecasting. Table 9 summarizes some

of these differences by presenting illustrative return and volatility forecasting results obtained with the same MLP-FinBERT specification.

Table 9: Selected MLP-FinBERT RMSE gains across forecast horizons

Sample	Asset	Target	$h = 1$	$h = 7$	$h = 14$	$h = 30$
Common	SOL	Volatility	-0.11	2.31*	10.62*	2.19*
Native	SOL	Volatility	1.36*	-0.08	5.92*	11.31*
Native	ETH	Return	1.32*	1.05*	0.36	-1.15
Common	BTC	Return	-0.35	-0.81	-2.35	7.67*
Common	BTC	Volatility	-0.98	-12.71	0.82*	15.33*
Common	XRP	Volatility	22.44*	30.22*	29.33*	56.38*
Native	XRP	Volatility	-6.61	0.72*	-10.67	31.05*

Notes: Entries are percentage RMSE improvements relative to the same estimator’s Market baseline. Negative values indicate worse forecasts. An asterisk denotes a positive gain with a one-sided Clark–West p -value below 5%. The profiles illustrate different horizon patterns and are not a test of differences between horizons. Targets accumulate returns or squared returns through h ; horizons are in calendar days.

Overall, the volatility forecasting results provide more favorable evidence for the contribution of news sentiment than the return forecasting results. The largest improvements are generally obtained with MLP, and several remain statistically supported across the two sample designs. However, their magnitude and stability depend on the cryptocurrency, forecast horizon, estimator, and evaluation sample. Therefore, sentiment information can improve volatility forecasts under some specifications, but its contribution is not systematic. In particular, the large improvements observed with certain models should be interpreted relative to their corresponding market benchmarks rather than as evidence that sentiment-enriched models always provide the most accurate forecasts.

6.3 Volatility regimes

This subsection examines whether the contribution of sentiment information to volatility forecasting depends on market volatility conditions. Tables 10 and 11 present the results obtained during high- and low-volatility periods, respectively. The regimes are defined using the historical volatility measure described in Equation (19).

Overall, the results do not suggest that sentiment information is systematically more useful during high- or low-volatility periods. In the common sample, 39 comparisons produce statistically supported RMSE improvements during high-volatility

periods, compared with 38 during low-volatility periods. Although these numbers are relatively similar, the results differ substantially across cryptocurrencies and forecast horizons. Moreover, these counts do not establish whether the differences in predictive performance between regimes are statistically significant.

Bitcoin exhibits important differences between the two regimes, particularly with MLP at longer forecast horizons. At thirty days, the inclusion of FinBERT sentiment reduces the RMSE by 51.96% during high-volatility periods, while it increases the RMSE by 13.52% during low-volatility periods. Similar results are obtained with lexicon sentiment. At fourteen days, the MLP improvements are approximately 38% in the high-volatility regime, compared with RMSE deteriorations of around 5% in the low-volatility regime. These results suggest that sentiment information may be more useful for Bitcoin volatility forecasting during periods of high market volatility, particularly at longer horizons. However, this interpretation requires caution because the high-volatility sample contains only nineteen forecast observations at each horizon.

Solana presents a different pattern. At the fourteen-day horizon, MLP with FinBERT reduces the RMSE by 10.19% during high-volatility periods and 12.56% during low-volatility periods. The corresponding improvements with lexicon sentiment are 10.84% and 13.03%, respectively. At thirty days, both sentiment measures increase the RMSE by approximately 4-5% in the high-volatility regime, while improvements of around 9% are observed in the low-volatility regime. Unlike Bitcoin, these results indicate that sentiment can provide larger improvements during relatively calm market conditions.

XRP also exhibits important improvements with MLP in both volatility regimes, although their magnitude depends on the forecast horizon. At seven days, FinBERT reduces the RMSE by 34.23% in the high-volatility regime and 56.49% in the low-volatility regime. The corresponding improvements at fourteen days are 29.75% and 51.34%. However, the relationship reverses at thirty days, with an improvement of 72.95% during high-volatility periods compared with 55.67% during low-volatility periods. Similar patterns are observed with lexicon sentiment. These results suggest that the contribution of sentiment depends not only on market volatility conditions but also on the forecast horizon.

An important limitation of this analysis is the number of observations available in each regime. The common-sample high-volatility subsets contain only nineteen forecast observations for Bitcoin, sixty-one for Ether, thirty-seven for Solana, and between fifty-one and sixty-five for XRP. The large improvements observed for Bitcoin during high-volatility periods are therefore based on a relatively small number of observations.

Moreover, the differences between regimes should be interpreted with caution. A statistically supported improvement in one regime and the absence of such an

improvement in another do not necessarily imply that the difference between the two regimes is statistically significant. The results presented here describe how the realized forecasting performance varies across different market conditions.

Overall, the analysis shows that the contribution of sentiment information to volatility forecasting varies across market conditions. There is no consistent evidence that sentiment is more useful during high-volatility periods. The results instead suggest that its predictive contribution depends on the cryptocurrency, forecast horizon, and estimator, with improvements observed during both high- and low-volatility periods.

6.4 Interpretation and robustness

Tables 12 and 13 present the return and volatility forecasting results obtained from the asset-native samples, respectively.

The differences observed across forecast horizons require some caution in their interpretation. First, the horizons considered in this study do not exceed thirty days. Therefore, the results concern the predictive contribution of sentiment over relatively short periods rather than its long-run persistence. Moreover, the forecast targets are cumulative, meaning that each horizon includes all daily returns or squared returns between $t + 1$ and $t + h$, rather than only the outcome observed on the last day. The estimated models, benchmark forecast errors, and available forecast observations also differ across horizons. In addition, the Clark-West tests evaluate each sentiment-enriched model against its corresponding market benchmark, but do not directly test whether the improvements differ significantly between horizons. Then, a change in the sign of the RMSE improvement should be interpreted as a change in forecasting performance, not as a reversal of the effect of media sentiment on prices. Identifying how long the market responds to news sentiment would require a separate analysis of returns at individual forecast leads. Establishing a causal effect would also require an appropriate identification strategy.

Three main findings emerge from the analysis. First, sentiment information provides more favorable improvements for cumulative squared returns than for cumulative returns, although these improvements are not observed across all model specifications. Second, the largest gains are generally obtained with MLP, while the other estimators produce smaller improvements, unchanged forecasts, or deteriorations. Finally, some results remain statistically supported across the two sample designs, whereas others change considerably or even become negative.

The differences between the common and asset-native samples also indicate that the results are sensitive to the estimation and evaluation periods. The native-sample out-of-sample evaluation begins in spring 2024 for Bitcoin and Ether, compared with late 2024 in the common sample. XRP also has an earlier evaluation period, while the difference is smaller for Solana. These changes may affect forecasting perfor-

mance through several channels, including the length of the estimation history, the selected hyperparameters, and the market conditions observed during the evaluation period. However, the present analysis does not allow us to identify the individual contribution of each of these factors. Another possible explanation for the differences across cryptocurrencies is their level of market development. Cryptocurrencies with longer trading histories may have more established trading practices and information processing mechanisms, potentially reducing the additional information contained in daily news sentiment. Nevertheless, the results do not provide consistent evidence in favor of this explanation. For example, XRP was launched in June 2012, before the launch of Ether in July 2015 (XRP Ledger 2026; Tual 2015). It would therefore be difficult to explain the results by simply distinguishing Bitcoin and Ether from Solana and XRP based on market age. Moreover, XRP exhibits some of the largest volatility forecasting improvements, while Bitcoin and Ether also provide statistically supported gains at longer horizons. Several of these results change across sample designs, although the relative ages of the cryptocurrencies remain unchanged. Market development may therefore play a role, but the age of a cryptocurrency alone does not explain the observed differences.

The comparison between FinBERT and lexicon sentiment also provides interesting results. In the common sample, FinBERT produces 23 statistically supported volatility improvements, compared with 15 for lexicon sentiment. However, some of the largest MLP improvements are very similar across the two sentiment measures. These findings indicate that both approaches can provide useful predictive information. Nevertheless, the forecasting exercise compares each sentiment measure with the market benchmark rather than directly testing whether FinBERT outperforms lexicon sentiment. Therefore, the results do not establish the superiority of one method over the other or explain why their forecasting performances are sometimes similar.

6.5 Comparison with the literature

The results obtained in this study can be compared with several studies examining the predictive contribution of textual information in financial markets, particularly cryptocurrencies. Concerning return forecasting, the limited and sample-dependent improvements differ from some of the more favorable findings in the literature. For example, Kraaijeveld and De Smedt (2020) identify Granger-predictive relationships between sentiment and cryptocurrency returns for selected assets, while Mai et al. (2018) find predictive information in Bitcoin community content. These differences do not necessarily contradict their findings. The present study uses news headlines rather than community content, considers different sample periods and cumulative forecast targets, and evaluates the additional contribution of sentiment relative to a market benchmark for each estimator. The results therefore indicate that the

predictive relationships documented in other settings do not necessarily translate into consistent improvements in out-of-sample RMSE when daily news sentiment is included.

The findings are more favorable for volatility forecasting, which is also consistent with some previous studies. For instance, Shen, Urquhart, and P. Wang (2019) find that the number of Bitcoin-related tweets contains predictive information for realized volatility, but not for next-day returns. Although the result is similar in terms of the difference between return and volatility predictability, the variables used are not the same. Their study considers the volume of Twitter activity, while this study focuses on the sentiment expressed in news headlines. Moreover, volatility is measured here using cumulative squared returns.

A similar distinction between returns and volatility has also been documented in traditional financial markets. Antweiler and Frank (2004) find that stock message board activity helps predict volatility, while its effects on returns are economically small. The stronger contribution of sentiment to volatility forecasting observed in this study is therefore not necessarily specific to cryptocurrency markets. It is consistent with a broader finding in the literature that the predictive usefulness of textual information may depend on the forecast target.

More recent studies focusing on cryptocurrency news also provide relevant comparisons. Brauneis and Sahiner (2026) examine the contribution of news sentiment to cryptocurrency volatility forecasting using different machine learning models. Their results indicate that including sentiment can improve or deteriorate forecasting performance depending on the model specification. This finding is consistent with the heterogeneity observed in the present study, particularly the substantial improvements obtained with MLP compared with the more limited results of other estimators. Nevertheless, the forecast targets and evaluation periods differ, which prevents a direct comparison of the magnitude of the improvements.

Similarly, Fiszeder, Orzeszko, and Pietrzyk (2026) do not find statistically significant out-of-sample improvements from ChatGPT-based news indicators when forecasting Bitcoin returns and volatility in their empirical setting. In contrast, the present analysis identifies some statistically supported improvements for Bitcoin, particularly in volatility forecasting. However, these improvements depend on the forecast horizon, estimator, and sample period. The differences between the two studies therefore do not establish a general contradiction, but illustrate how the contribution of news sentiment can depend on the empirical design.

The variation in forecasting performance across horizons also raises questions about how long sentiment information remains useful. In traditional financial markets, Tetlock (2007) find that pessimistic news is associated with downward pressure on stock prices, followed by a subsequent reversal. Some of the horizon patterns observed in this study, particularly the larger improvements at intermediate horizons

followed by weaker performance, could suggest that the predictive usefulness of sentiment changes over time. The results do not establish a similar price reversal mechanism. The RMSE measures forecasting accuracy rather than the direction of the relationship between sentiment and returns. Also, the cumulative forecast targets combine observations from several days, making it impossible to identify the effect at a specific forecast lead. The differences across cryptocurrencies and sample designs also prevent a general conclusion about the duration of sentiment information. For example, Solana exhibits relatively large improvements at intermediate horizons in the common sample, while Bitcoin provides some of its largest gains at thirty days. Therefore, the results do not indicate a common horizon beyond which news sentiment loses its predictive usefulness.

The importance of the forecasting method can also be related to the machine learning literature. Gu, Kelly, and Xiu (2020) show that flexible machine learning models can achieve strong forecasting performance in equity markets, partly because of their ability to capture nonlinear interactions between predictive variables. In the present study, the largest improvements are also frequently obtained with MLP. This suggests that the contribution of sentiment may depend on the ability of the estimator to capture more complex relationships between sentiment and market variables. The results do not directly establish that nonlinear interactions explain these improvements. In some cases, the sentiment-enriched MLP still produces less accurate forecasts than another estimator using only market information. Therefore, greater model flexibility does not necessarily guarantee better forecasting performance or a larger contribution from sentiment information.

The volatility-regime analysis provides another point of comparison with the literature. GARCÍA (2013) find that the predictive contribution of financial news sentiment to equity returns is particularly important during recessions. The results obtained here also show that the contribution of sentiment can vary across market conditions. However, improvements are observed during both high- and low-volatility periods, without a consistent advantage for one regime. It is also important to distinguish the market conditions considered in the two studies. Recessions are not equivalent to periods of high cryptocurrency volatility, and the forecast targets are different. The results therefore support the idea that the predictive contribution of sentiment may depend on market conditions, but do not establish that sentiment is systematically more useful during periods of market stress.

Finally, the cryptocurrency market has some characteristics that distinguish it from traditional financial markets, including continuous trading and substantial differences in risk and market attention across assets. Liu and Tsyvinski (2021) document some of these characteristics, which provide additional motivation for studying sentiment information in cryptocurrency forecasting. However, market attention and news sentiment are different variables, and the present analysis does not iden-

tify the contribution of attention itself. The results also do not establish that news sentiment has a stronger predictive contribution in cryptocurrency markets than in equity markets. Such a comparison would require a common empirical design with comparable information sets, forecast targets, and sample periods. Overall, the results extend the existing literature by examining the additional forecasting contribution of asset-specific news sentiment across several cryptocurrencies, forecast horizons, and estimators. The main finding is that sentiment information can improve volatility forecasts, although its contribution is not systematic and depends on the cryptocurrency, forecasting model, horizon, and sample period. These improvements concern the predictive information contained in media tone and should not be interpreted as evidence of a causal effect on market outcomes or as a direct measure of investor sentiment.

7 Conclusion

This study examined whether news sentiment provides additional predictive information for cryptocurrency returns and volatility beyond that contained in market variables. The results show that the contribution of sentiment depends on the forecast target, cryptocurrency, horizon, and estimator. The improvements are relatively limited for return forecasting and sensitive to the evaluation sample. Some important gains are observed at longer horizons, particularly for Bitcoin and Solana with MLP in the common sample. However, most of these improvements are not preserved in the asset-native samples. Therefore, the results do not suggest that sentiment is systematically useful for return forecasting, although it can improve predictions under some specifications.

The evidence is more favorable for volatility forecasting, where several models including sentiment provide statistically supported improvements. The largest gains are generally obtained with MLP, particularly for XRP, while Bitcoin and Ether also benefit at longer horizons in the common sample. Some improvements remain substantial across the two sample designs, notably for Ether, Solana, and XRP, while others change considerably or become negative, as observed for Bitcoin at the thirty-day horizon. These differences also do not suggest a consistent relationship between the age of a cryptocurrency and the predictive contribution of sentiment.

The volatility-regime analysis further shows that the contribution of sentiment varies across market conditions. Although some important improvements are observed during high-volatility periods, others are larger during relatively calm periods. Consequently, the results do not establish that sentiment information is systematically more useful when market volatility is high. Moreover, an improvement relative to a particular market benchmark does not necessarily mean that the sentiment-enriched model provides the most accurate forecasts among all estimators.

Future research could examine whether these improvements remain stable across different neural network initializations and additional evaluation periods. Using identical forecast dates across sample designs could also help distinguish the influence of the estimation history from that of the evaluation period. Direct comparisons between FinBERT and lexicon sentiment, as well as formal tests of differences across volatility regimes, could provide additional evidence about their respective contributions. Finally, considering more detailed information about news topics and events could help identify which aspects of media coverage contribute to forecasting.

References

- Abraham, Jethin et al. (2018). “Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis”. In: *SMU Data Science Review* 1.3. URL: <https://api.semanticscholar.org/CorpusID:52950647>.
- Akerlof, George A. and Robert J. Shiller (2009). *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton University Press. ISBN: 9780691145921. URL: <http://www.jstor.org/stable/j.ctv36mk90z> (visited on 05/19/2026).
- Antweiler, Werner and Murray Z. Frank (2004). “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards”. In: *The Journal of Finance* 59.3, pp. 1259–1294. DOI: <https://doi.org/10.1111/j.1540-6261.2004.00662.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2004.00662.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2004.00662.x>.
- Araci, Dogu (2019). “FinBERT: Financial Sentiment Analysis with Pre-trained Language Models”. In: *CoRR* abs/1908.10063. arXiv: 1908.10063. URL: <http://arxiv.org/abs/1908.10063>.
- Baker, Malcolm and Jeffrey Wurgler (June 2007). “Investor Sentiment in the Stock Market”. In: *Journal of Economic Perspectives* 21.2, pp. 129–152. DOI: 10.1257/jep.21.2.129. URL: <https://www.aeaweb.org/articles?id=10.1257/jep.21.2.129>.
- Balcilar, Mehmet et al. (2017). “Can Volume Predict Bitcoin Returns and Volatility? A Quantiles-Based Approach”. In: *Economic Modelling* 64, pp. 74–81. DOI: 10.1016/j.econmod.2017.03.019. URL: <https://doi.org/10.1016/j.econmod.2017.03.019>.
- Bollen, Johan, Huina Mao, and Xiaojun Zeng (2011). “Twitter mood predicts the stock market”. In: *Journal of Computational Science* 2.1, pp. 1–8. ISSN: 1877-7503. DOI: <https://doi.org/10.1016/j.jocs.2010.12.007>. URL: <https://www.sciencedirect.com/science/article/pii/S187775031100007X>.

- Bollinger, John (2001). *Bollinger on Bollinger Bands*. New York: McGraw-Hill. ISBN: 9780071373685. URL: <https://www.mheducation.com/highered/mhp/product/bollinger-bollinger-bands>.
- Brauneis, Alexander and Mehmet Sahiner (Mar. 1, 2026). “Crypto Volatility Forecasting: Mounting a HAR, Sentiment, and Machine Learning Horserace”. In: *Asia-Pacific Financial Markets* 33.1, pp. 379–411. ISSN: 1573-6946. DOI: 10.1007/s10690-024-09510-6. URL: <https://doi.org/10.1007/s10690-024-09510-6>.
- Breiman, Leo (2001). “Random Forests”. In: *Machine Learning* 45.1, pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. URL: <https://doi.org/10.1023/A:1010933404324>.
- BROCK, WILLIAM, JOSEF LAKONISHOK, and BLAKE LeBARON (1992). “Simple Technical Trading Rules and the Stochastic Properties of Stock Returns”. In: *The Journal of Finance* 47.5, pp. 1731–1764. DOI: <https://doi.org/10.1111/j.1540-6261.1992.tb04681.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1992.tb04681.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1992.tb04681.x>.
- Chen, Tianqi and Carlos Guestrin (Aug. 2016). “XGBoost: A Scalable Tree Boosting System”. In: pp. 785–794. DOI: 10.1145/2939672.2939785.
- Chu, Jeffrey, Saralees Nadarajah, and Stephen Chan (July 2015). “Statistical Analysis of the Exchange Rate of Bitcoin”. In: *PloS one* 10, e0133678. DOI: 10.1371/journal.pone.0133678.
- Chuen, David, Li Guo, and Yu Wang (Dec. 2017). “Cryptocurrency: A New Investment Opportunity?” In: *The Journal of Alternative Investments* 20, pp. 16–40. DOI: 10.3905/jai.2018.20.3.016.
- Clarida, Richard, Josh Davis, and Niels Pedersen (Nov. 2009). *Currency Carry Trade Regimes: Beyond the Fama Regression*. Working Paper 15523. National Bureau of Economic Research. DOI: 10.3386/w15523. URL: <http://www.nber.org/papers/w15523>.
- Clark, Todd E. and Kenneth D. West (2007). “Approximately normal tests for equal predictive accuracy in nested models”. In: *Journal of Econometrics* 138.1. 50th Anniversary Econometric Institute, pp. 291–311. ISSN: 0304-4076. DOI: <https://doi.org/10.1016/j.jeconom.2006.05.023>. URL: <https://www.sciencedirect.com/science/article/pii/S0304407606000960>.
- Corbet, Shaen et al. (2018). “Exploring the dynamic relationships between cryptocurrencies and other financial assets”. In: *Economics Letters* 165, pp. 28–34. ISSN: 0165-1765. DOI: <https://doi.org/10.1016/j.econlet.2018.01.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0165176518300041>.

- Diebold, Francis X. and Roberto S. Mariano (1995). “Comparing Predictive Accuracy”. In: *Journal of Business & Economic Statistics* 13.3, pp. 253–263. DOI: 10.1080/07350015.1995.10524599. eprint: <https://www.tandfonline.com/doi/pdf/10.1080/07350015.1995.10524599>. URL: <https://www.tandfonline.com/doi/abs/10.1080/07350015.1995.10524599>.
- Fama, Eugene F. (1970). “Efficient Capital Markets: A Review of Theory and Empirical Work”. In: *The Journal of Finance* 25.2, pp. 383–417. ISSN: 00221082, 15406261. URL: <http://www.jstor.org/stable/2325486> (visited on 05/20/2026).
- Fiszeder, Piotr, Witold Orzeszko, and Radosław Pietrzyk (Apr. 30, 2026). “News sentiment analysis using ChatGPT for Bitcoin price dynamics”. In: *Journal of Big Data* 13.1, 106. ISSN: 2196-1115. DOI: 10.1186/s40537-026-01392-x. URL: <https://doi.org/10.1186/s40537-026-01392-x>.
- GARCÍA, DIEGO (2013). “Sentiment during Recessions”. In: *The Journal of Finance* 68.3, pp. 1267–1300. DOI: <https://doi.org/10.1111/jofi.12027>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12027>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12027>.
- Gensler, Gary (Jan. 10, 2024). *Statement on the Approval of Spot Bitcoin Exchange-Traded Products*. U.S. Securities and Exchange Commission. URL: <https://www.sec.gov/newsroom/speeches-statements/gensler-statement-spot-bitcoin-011023> (visited on 09/28/2026).
- Gerritsen, {Dirk F.} et al. (May 2020). “The profitability of technical trading rules in the Bitcoin market”. English. In: *Finance Research Letters* 34. ISSN: 1544-6123. DOI: 10.1016/j.frl.2019.08.011.
- Gu, Shihao, Bryan Kelly, and Dacheng Xiu (May 2020). “Empirical Asset Pricing via Machine Learning”. In: *The Review of Financial Studies* 33.5, pp. 2223–2273. ISSN: 0893-9454. DOI: 10.1093/rfs/hhaa009. eprint: <https://academic.oup.com/rfs/article-pdf/33/5/2223/33209812/hhaa009.pdf>. URL: <https://doi.org/10.1093/rfs/hhaa009>.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (Feb. 2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*. ISBN: 0387848576.
- Kim, Young Bin et al. (2016). “Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies”. In: *PLoS ONE* 11. URL: <https://api.semanticscholar.org/CorpusID:15944052>.
- Kraaijeveld, Olivier and Johannes De Smedt (2020). “The predictive power of public Twitter sentiment for forecasting cryptocurrency prices”. In: *Journal of International Financial Markets, Institutions and Money* 65.C. DOI: 10.1016/j.intfin.2020.101188. URL: <https://ideas.repec.org/a/eee/intfin/v65y2020ics104244312030072x.html>.

- Liu, Yukun and Aleh Tsyvinski (June 2021). “Risks and Returns of Cryptocurrency”. In: *The Review of Financial Studies* 34.6, pp. 2689–2727. ISSN: 0893-9454. DOI: 10.1093/rfs/hhaa113. eprint: <https://academic.oup.com/rfs/article-pdf/34/6/2689/38095668/hhaa113.pdf>. URL: <https://doi.org/10.1093/rfs/hhaa113>.
- LOUGHRAN, TIM and BILL MCDONALD (2011). “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks”. In: *The Journal of Finance* 66.1, pp. 35–65. DOI: <https://doi.org/10.1111/j.1540-6261.2010.01625.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2010.01625.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2010.01625.x>.
- Mai, Feng et al. (2018). “How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis”. In: *Journal of Management Information Systems* 35, pp. 19–52. URL: <https://api.semanticscholar.org/CorpusID:13665291>.
- Parkinson, Michael (1980). “The Extreme Value Method for Estimating the Variance of the Rate of Return”. In: *The Journal of Business* 53.1, pp. 61–65. ISSN: 00219398, 15375374. URL: <http://www.jstor.org/stable/2352357> (visited on 10/05/2026).
- Phillips, Peter C. B., Yangru Wu, and Jun Yu (Feb. 2011). “EXPLOSIVE BEHAVIOR IN THE 1990s NASDAQ: WHEN DID EXUBERANCE ESCALATE ASSET VALUES?” In: *International Economic Review* 52.1, pp. 201–226. DOI: None. URL: <https://ideas.repec.org/a/ier/iecrev/v52y2011i1p201-226.html>.
- Shen, Dehua, Andrew Urquhart, and Pengfei Wang (2019). “Does twitter predict Bitcoin?” In: *Economics Letters* 174, pp. 118–122. ISSN: 0165-1765. DOI: <https://doi.org/10.1016/j.econlet.2018.11.007>. URL: <https://www.sciencedirect.com/science/article/pii/S0165176518304634>.
- Tetlock, Paul (Feb. 2007). “Giving Content to Investor Sentiment: The Role of Media in the Stock Market”. In: *Journal of Finance* 62, pp. 1139–1168. DOI: 10.2139/ssrn.685145.
- Tual, Stephan (July 30, 2015). *Ethereum Launches*. Ethereum Foundation. URL: <https://blog.ethereum.org/2015/07/30/ethereum-launches> (visited on 09/28/2026).
- U.S. Securities and Exchange Commission (May 23, 2024). *Order Granting Accelerated Approval of Proposed Rule Changes, as Modified by Amendments Thereto, to List and Trade Shares of Ether-Based Exchange-Traded Products*. URL: <https://www.sec.gov/rule-release/34-100224> (visited on 09/28/2026).
- Urquhart, Andrew (2016). “The inefficiency of Bitcoin”. In: *Economics Letters* 148, pp. 80–82. ISSN: 0165-1765. DOI: <https://doi.org/10.1016/j.econlet.2016>.

- 09.019. URL: <https://www.sciencedirect.com/science/article/pii/S0165176516303640>.
- Wilder Jr., J. Welles (1978). *New Concepts in Technical Trading Systems*. Greensboro, NC: Trend Research. ISBN: 9780894590276. URL: <https://windsorpublishing.com/product/new-concepts-in-technical-trading-systems/>.
- XRP Ledger (2026). *History*. URL: <https://xrpl.org/about/history> (visited on 09/28/2026).

Table 10: Out-of-sample RMSE improvements (%) of sentiment-augmented models relative to market-only benchmarks for volatility prediction (Clark–West test) using the common 2020–2025 sample in the high-volatility regime.

Model	$h = 1$	$h = 7$	$h = 14$	$h = 30$
BTC				
MLP FinBERT	-5.38	-8.01	38.16***	51.96***
MLP Lexicon	-5.66	-8.06	38.39***	52.12***
OLS FinBERT	-0.15	0.22	0.20	0.07*
OLS Lexicon	0.40**	0.03***	-0.31	0.07
RF FinBERT	1.04	-0.40	1.63***	4.99***
RF Lexicon	0.72	0.40	1.73***	4.64***
Ridge Lexicon	0.03*	-0.07	-0.40	0.08
ETH				
MLP FinBERT	0.98*	-5.10	8.46***	17.02***
MLP Lexicon	0.58	-4.75	7.11***	12.41***
RF Lexicon	-0.38	-0.85	-3.25	2.04*
XGB Lexicon	-0.33	0.21	-11.43	1.19**
SOL				
MLP FinBERT	-1.71	6.73**	10.19***	-4.93*
MLP Lexicon	-1.53	7.64***	10.84***	-4.13*
OLS FinBERT	-1.10	-0.22	8.20***	7.72***
OLS Lexicon	-0.13	0.19	0.25	1.16*
RF FinBERT	-0.00	0.42	-1.32	1.35*
Ridge FinBERT	-1.04	-0.11	4.93***	3.85***
Ridge Lexicon	-0.13	0.21	0.23	0.64*
XGB FinBERT	-9.12	-3.32	1.21	4.81*
XGB Lexicon	-3.80	-1.10	3.88	4.89*
XRP				
MLP FinBERT	31.43***	34.23***	29.75***	72.95***
MLP Lexicon	32.04***	34.88***	29.39***	72.83***
RF FinBERT	-1.73	3.34***	2.42*	3.83***
RF Lexicon	-1.26	3.07***	2.80*	3.50**
XGB FinBERT	-0.92	6.64***	-4.03	4.83**
XGB Lexicon	-3.61	6.43***	1.02	4.83**

Notes: Positive values indicate lower RMSE for the sentiment-augmented model. Stars report one-sided upper-tail Clark–West significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Only rows with at least one significant horizon are shown. A significant Clark–West result can accompany a negative raw RMSE change because the test adjusts for parameter-estimation noise.

Table 11: Out-of-sample RMSE improvements (%) of sentiment-augmented models relative to market-only benchmarks for volatility prediction (Clark-West test) using the common 2020-2025 sample in the low-volatility regime.

Model	$h = 1$	$h = 7$	$h = 14$	$h = 30$
BTC				
MLP FinBERT	-3.13***	-17.66**	-4.94	-13.52***
MLP Lexicon	-3.06***	-17.70*	-4.88	-13.44***
ETH				
Lasso FinBERT	1.22**	0.00	0.00	0.00
MLP FinBERT	-0.74	-5.16	-4.22	1.31**
MLP Lexicon	-0.94	-5.16	-4.37	1.49**
OLS FinBERT	0.03*	-0.00	-0.22	1.64***
OLS Lexicon	-0.14	0.61*	1.75*	2.97***
RF FinBERT	-0.03	0.27	-3.35	2.52***
RF Lexicon	0.01	0.37*	-2.51	3.22**
Ridge FinBERT	-0.00	-0.09	-0.25	1.54**
Ridge Lexicon	-0.15	0.34*	1.15*	2.29***
XGB Lexicon	0.24*	-0.28	-0.34	0.11
SOL				
Lasso Lexicon	6.36***	-0.99	0.00	0.00
MLP FinBERT	2.58***	2.76***	12.56***	8.62**
MLP Lexicon	1.99***	0.56***	13.03***	8.83**
RF Lexicon	-2.21	0.40	0.75**	-1.00
Ridge FinBERT	0.85***	0.12	1.35*	1.50*
XGB FinBERT	3.74***	-0.10	2.59**	-3.26
XGB Lexicon	3.47***	-0.91	-1.11	0.05
XRP				
MLP FinBERT	39.41***	56.49***	51.34***	55.67***
MLP Lexicon	34.44***	56.67***	51.48***	55.57***
OLS FinBERT	1.44	0.56*	0.26	-0.25
RF FinBERT	-2.04	2.10***	1.18**	2.82**
RF Lexicon	-0.74	1.80***	0.72*	2.60**
Ridge FinBERT	0.62	0.42*	0.24	-0.13
XGB FinBERT	0.73	3.72	-2.33	3.48***
XGB Lexicon	-17.52	3.20	1.86	3.48***

Notes: Positive values indicate lower RMSE for the sentiment-augmented model. Stars report one-sided upper-tail Clark-West significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Only rows with at least one significant horizon are shown. A significant Clark-West result can accompany a negative raw RMSE change because the test adjusts for parameter-estimation noise.

Table 12: Out-of-sample RMSE improvements (%) of sentiment-augmented models relative to market-only benchmarks for return prediction (Clark-West test) using asset-native samples (BTC and ETH: 2017-2025; SOL: 2020-2025; XRP: 2018-2025).

Model	$h = 1$	$h = 7$	$h = 14$	$h = 30$
BTC				
RF FinBERT	-0.09	-0.07	0.19	0.72*
XGB FinBERT	0.00	0.41*	-1.20	4.86
ETH				
MLP FinBERT	1.32**	1.05**	0.36*	-1.15
MLP Lexicon	1.46***	1.35**	0.45**	-0.33
XGB FinBERT	-0.09	1.41**	-2.80	1.77
XGB Lexicon	0.15*	1.43**	-2.28	1.70
SOL				
XGB Lexicon	-0.30	-0.92	0.80*	-1.44
XRP				
MLP FinBERT	0.17	-7.94	-3.35	2.70***
MLP Lexicon	0.53**	-7.96	-3.65	3.08**
OLS FinBERT	-0.06	0.04	0.10**	0.09*
RF FinBERT	0.10*	0.13*	0.13	0.43*
RF Lexicon	0.09*	0.15	0.19	0.45*

Notes: Positive values indicate lower RMSE for the sentiment-augmented model. Stars report one-sided upper-tail Clark-West significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Only rows with at least one significant horizon are shown. A significant Clark-West result can accompany a negative raw RMSE change because the test adjusts for parameter-estimation noise.

Table 13: Out-of-sample RMSE improvements (%) of sentiment-augmented models relative to market-only benchmarks for volatility prediction (Clark–West test) using asset-native samples (BTC and ETH: 2017–2025; SOL: 2020–2025; XRP: 2018–2025).

Model	$h = 1$	$h = 7$	$h = 14$	$h = 30$
BTC				
MLP FinBERT	-1.20**	-13.84**	-7.74**	-27.99
MLP Lexicon	-0.83**	-9.97**	-8.40**	-20.18
RF FinBERT	-0.78	-1.62	-2.08	2.09***
RF Lexicon	-0.62	-1.53	-1.91	1.72**
XGB FinBERT	4.05**	0.30**	0.02	-2.05
XGB Lexicon	-27.30	-0.34*	-2.01	-0.37
ETH				
MLP FinBERT	-1.39*	7.67***	8.60***	8.53***
MLP Lexicon	-1.76	7.71***	8.86***	8.78***
RF FinBERT	0.46***	0.22	-0.78	0.21
RF Lexicon	0.34*	0.52*	-0.53	-0.10
SOL				
MLP FinBERT	1.36**	-0.08***	5.92***	11.31***
MLP Lexicon	2.19**	0.27***	6.12***	9.31***
OLS FinBERT	-1.26	0.24**	1.49***	0.42**
Ridge FinBERT	-1.25	0.32***	1.83***	0.80*
XGB FinBERT	-2.21	-6.71	1.51***	-4.12
XRP				
MLP FinBERT	-6.61	0.72**	-10.67	31.05***
MLP Lexicon	-5.76	-34.97	-11.30	31.15***
RF FinBERT	0.50*	0.43*	0.03	0.43
RF Lexicon	0.56**	0.35*	0.26	0.34
XGB FinBERT	0.21	5.69**	1.01	1.54**
XGB Lexicon	0.20	2.54*	0.93	0.55

Notes: Positive values indicate lower RMSE for the sentiment-augmented model. Stars report one-sided upper-tail Clark–West significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Only rows with at least one significant horizon are shown. A significant Clark–West result can accompany a negative raw RMSE change because the test adjusts for parameter-estimation noise.

A Appendix

A.1 Complementary plots

Here is plot of lexicon-based sentiment and FinBERT sentiment by cryptocurrency for different horizons.

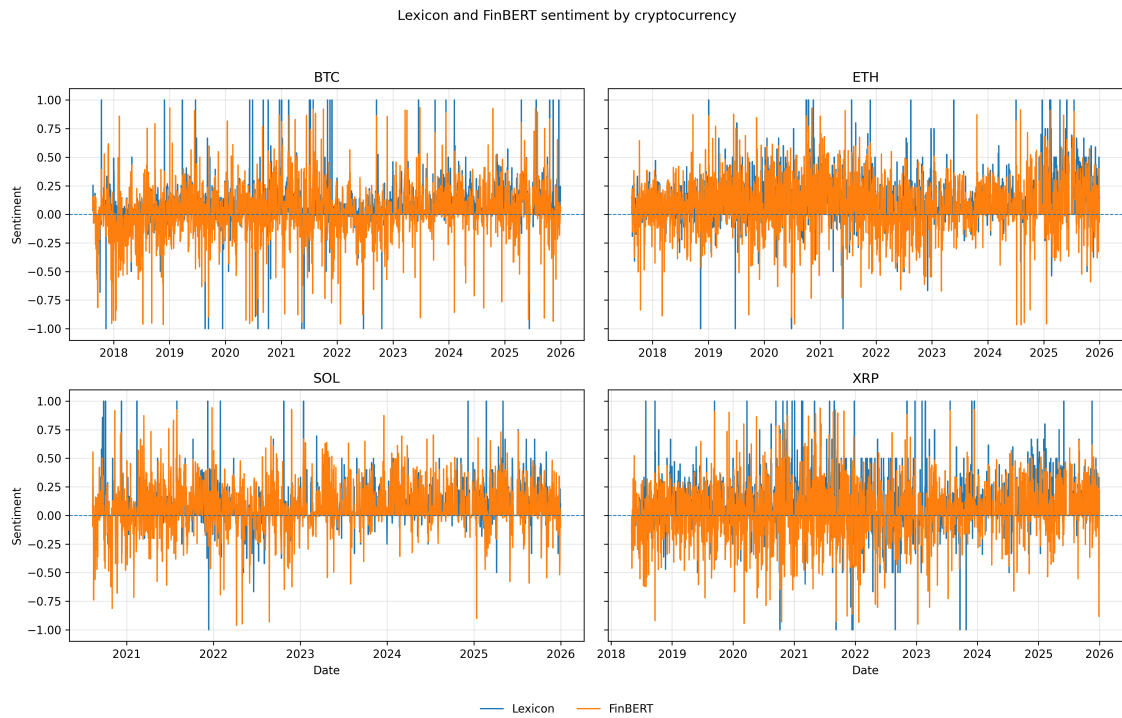


Figure 2: 1-day rolling averages of Lexicon and FinBERT sentiment indicators by cryptocurrency.

References

- Abraham, Jethin et al. (2018). “Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis”. In: *SMU Data Science Review* 1.3. URL: <https://api.semanticscholar.org/CorpusID:52950647>.
- Akerlof, George A. and Robert J. Shiller (2009). *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton University Press. ISBN: 9780691145921. URL: <http://www.jstor.org/stable/j.ctv36mk90z> (visited on 05/19/2026).

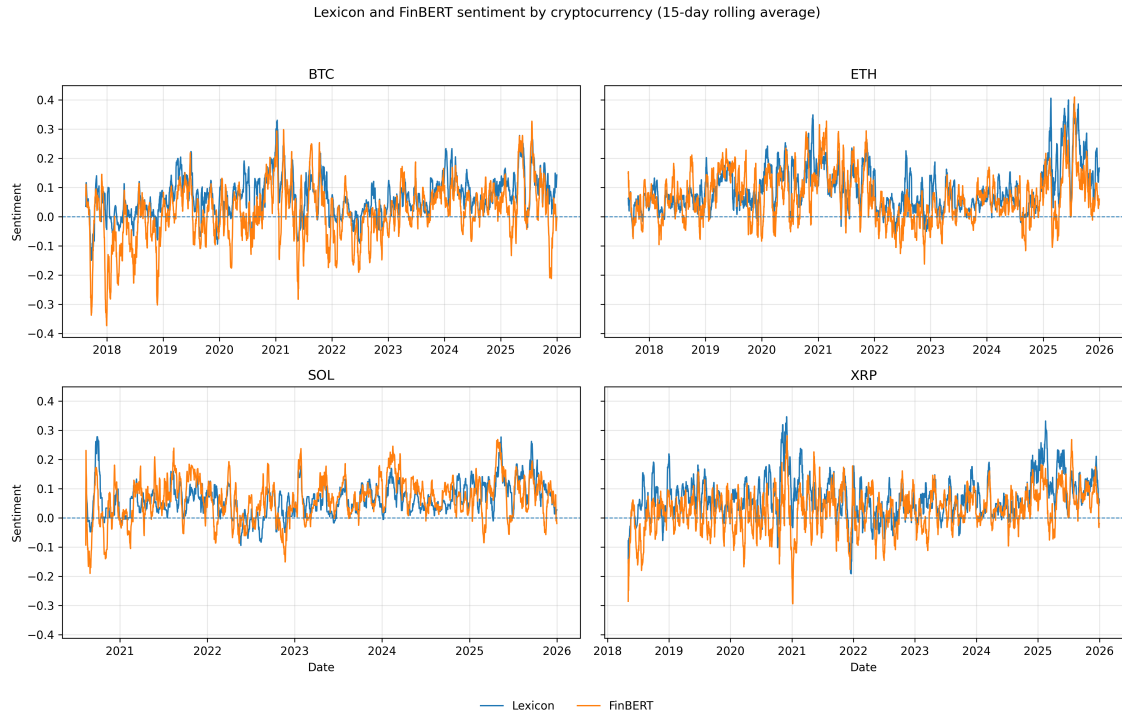


Figure 3: 15-day rolling averages of Lexicon and FinBERT sentiment indicators by cryptocurrency.

- Antweiler, Werner and Murray Z. Frank (2004). “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards”. In: *The Journal of Finance* 59.3, pp. 1259–1294. DOI: <https://doi.org/10.1111/j.1540-6261.2004.00662.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2004.00662.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2004.00662.x>.
- Araci, Dogu (2019). “FinBERT: Financial Sentiment Analysis with Pre-trained Language Models”. In: *CoRR* abs/1908.10063. arXiv: 1908.10063. URL: <http://arxiv.org/abs/1908.10063>.
- Baker, Malcolm and Jeffrey Wurgler (June 2007). “Investor Sentiment in the Stock Market”. In: *Journal of Economic Perspectives* 21.2, pp. 129–152. DOI: 10.1257/jep.21.2.129. URL: <https://www.aeaweb.org/articles?id=10.1257/jep.21.2.129>.
- Balcilar, Mehmet et al. (2017). “Can Volume Predict Bitcoin Returns and Volatility? A Quantiles-Based Approach”. In: *Economic Modelling* 64, pp. 74–81. DOI: 10.1016/j.econmod.2017.03.019. URL: <https://doi.org/10.1016/j.econmod.2017.03.019>.

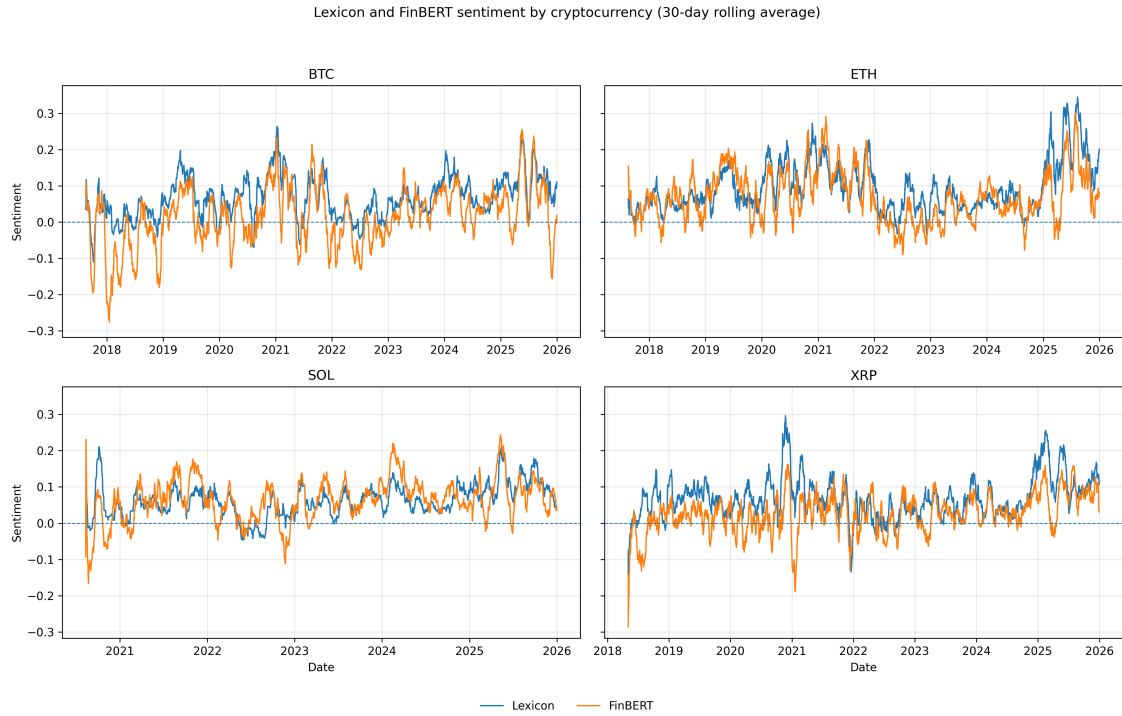


Figure 4: 1-month rolling averages of Lexicon and FinBERT sentiment indicators by cryptocurrency.

- Bollen, Johan, Huina Mao, and Xiaojun Zeng (2011). “Twitter mood predicts the stock market”. In: *Journal of Computational Science* 2.1, pp. 1–8. ISSN: 1877-7503. DOI: <https://doi.org/10.1016/j.jocs.2010.12.007>. URL: <https://www.sciencedirect.com/science/article/pii/S187775031100007X>.
- Bollinger, John (2001). *Bollinger on Bollinger Bands*. New York: McGraw-Hill. ISBN: 9780071373685. URL: <https://www.mheducation.com/highered/mhp/product/bollinger-bollinger-bands>.
- Brauneis, Alexander and Mehmet Sahiner (Mar. 1, 2026). “Crypto Volatility Forecasting: Mounting a HAR, Sentiment, and Machine Learning Horserace”. In: *Asia-Pacific Financial Markets* 33.1, pp. 379–411. ISSN: 1573-6946. DOI: 10.1007/s10690-024-09510-6. URL: <https://doi.org/10.1007/s10690-024-09510-6>.
- Breiman, Leo (2001). “Random Forests”. In: *Machine Learning* 45.1, pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. URL: <https://doi.org/10.1023/A:1010933404324>.
- BROCK, WILLIAM, JOSEF LAKONISHOK, and BLAKE LeBARON (1992). “Simple Technical Trading Rules and the Stochastic Properties of Stock Returns”.

- In: *The Journal of Finance* 47.5, pp. 1731–1764. DOI: <https://doi.org/10.1111/j.1540-6261.1992.tb04681.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1992.tb04681.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1992.tb04681.x>.
- Chen, Tianqi and Carlos Guestrin (Aug. 2016). “XGBoost: A Scalable Tree Boosting System”. In: pp. 785–794. DOI: 10.1145/2939672.2939785.
- Chu, Jeffrey, Saralees Nadarajah, and Stephen Chan (July 2015). “Statistical Analysis of the Exchange Rate of Bitcoin”. In: *PloS one* 10, e0133678. DOI: 10.1371/journal.pone.0133678.
- Chuen, David, Li Guo, and Yu Wang (Dec. 2017). “Cryptocurrency: A New Investment Opportunity?” In: *The Journal of Alternative Investments* 20, pp. 16–40. DOI: 10.3905/jai.2018.20.3.016.
- Clarida, Richard, Josh Davis, and Niels Pedersen (Nov. 2009). *Currency Carry Trade Regimes: Beyond the Fama Regression*. Working Paper 15523. National Bureau of Economic Research. DOI: 10.3386/w15523. URL: <http://www.nber.org/papers/w15523>.
- Clark, Todd E. and Kenneth D. West (2007). “Approximately normal tests for equal predictive accuracy in nested models”. In: *Journal of Econometrics* 138.1. 50th Anniversary Econometric Institute, pp. 291–311. ISSN: 0304-4076. DOI: <https://doi.org/10.1016/j.jeconom.2006.05.023>. URL: <https://www.sciencedirect.com/science/article/pii/S0304407606000960>.
- Corbet, Shaen et al. (2018). “Exploring the dynamic relationships between cryptocurrencies and other financial assets”. In: *Economics Letters* 165, pp. 28–34. ISSN: 0165-1765. DOI: <https://doi.org/10.1016/j.econlet.2018.01.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0165176518300041>.
- Diebold, Francis X. and Roberto S. Mariano (1995). “Comparing Predictive Accuracy”. In: *Journal of Business & Economic Statistics* 13.3, pp. 253–263. DOI: 10.1080/07350015.1995.10524599. eprint: <https://www.tandfonline.com/doi/pdf/10.1080/07350015.1995.10524599>. URL: <https://www.tandfonline.com/doi/abs/10.1080/07350015.1995.10524599>.
- Fama, Eugene F. (1970). “Efficient Capital Markets: A Review of Theory and Empirical Work”. In: *The Journal of Finance* 25.2, pp. 383–417. ISSN: 00221082, 15406261. URL: <http://www.jstor.org/stable/2325486> (visited on 05/20/2026).
- Fiszeder, Piotr, Witold Orzeszko, and Radosław Pietrzyk (Apr. 30, 2026). “News sentiment analysis using ChatGPT for Bitcoin price dynamics”. In: *Journal of Big Data* 13.1, 106. ISSN: 2196-1115. DOI: 10.1186/s40537-026-01392-x. URL: <https://doi.org/10.1186/s40537-026-01392-x>.

- GARCÍA, DIEGO (2013). “Sentiment during Recessions”. In: *The Journal of Finance* 68.3, pp. 1267–1300. DOI: <https://doi.org/10.1111/jofi.12027>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12027>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12027>.
- Gensler, Gary (Jan. 10, 2024). *Statement on the Approval of Spot Bitcoin Exchange-Traded Products*. U.S. Securities and Exchange Commission. URL: <https://www.sec.gov/newsroom/speeches-statements/gensler-statement-spot-bitcoin-011023> (visited on 09/28/2026).
- Gerritsen, {Dirk F.} et al. (May 2020). “The profitability of technical trading rules in the Bitcoin market”. English. In: *Finance Research Letters* 34. ISSN: 1544-6123. DOI: 10.1016/j.frl.2019.08.011.
- Gu, Shihao, Bryan Kelly, and Dacheng Xiu (May 2020). “Empirical Asset Pricing via Machine Learning”. In: *The Review of Financial Studies* 33.5, pp. 2223–2273. ISSN: 0893-9454. DOI: 10.1093/rfs/hhaa009. eprint: <https://academic.oup.com/rfs/article-pdf/33/5/2223/33209812/hhaa009.pdf>. URL: <https://doi.org/10.1093/rfs/hhaa009>.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (Feb. 2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*. ISBN: 0387848576.
- Kim, Young Bin et al. (2016). “Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies”. In: *PLoS ONE* 11. URL: <https://api.semanticscholar.org/CorpusID:15944052>.
- Kraaijeveld, Olivier and Johannes De Smedt (2020). “The predictive power of public Twitter sentiment for forecasting cryptocurrency prices”. In: *Journal of International Financial Markets, Institutions and Money* 65.C. DOI: 10.1016/j.intfin.2020.101188. URL: <https://ideas.repec.org/a/eee/intfin/v65y2020ics104244312030072x.html>.
- Liu, Yukun and Aleh Tsyvinski (June 2021). “Risks and Returns of Cryptocurrency”. In: *The Review of Financial Studies* 34.6, pp. 2689–2727. ISSN: 0893-9454. DOI: 10.1093/rfs/hhaa113. eprint: <https://academic.oup.com/rfs/article-pdf/34/6/2689/38095668/hhaa113.pdf>. URL: <https://doi.org/10.1093/rfs/hhaa113>.
- LOUGHRAN, TIM and BILL MCDONALD (2011). “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks”. In: *The Journal of Finance* 66.1, pp. 35–65. DOI: <https://doi.org/10.1111/j.1540-6261.2010.01625.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2010.01625.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2010.01625.x>.

- Mai, Feng et al. (2018). “How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis”. In: *Journal of Management Information Systems* 35, pp. 19–52. URL: <https://api.semanticscholar.org/CorpusID:13665291>.
- Parkinson, Michael (1980). “The Extreme Value Method for Estimating the Variance of the Rate of Return”. In: *The Journal of Business* 53.1, pp. 61–65. ISSN: 00219398, 15375374. URL: <http://www.jstor.org/stable/2352357> (visited on 10/05/2026).
- Phillips, Peter C. B., Yangru Wu, and Jun Yu (Feb. 2011). “EXPLOSIVE BEHAVIOR IN THE 1990s NASDAQ: WHEN DID EXUBERANCE ESCALATE ASSET VALUES?” In: *International Economic Review* 52.1, pp. 201–226. DOI: None. URL: <https://ideas.repec.org/a/ier/iecrev/v52y2011i1p201-226.html>.
- Shen, Dehua, Andrew Urquhart, and Pengfei Wang (2019). “Does twitter predict Bitcoin?” In: *Economics Letters* 174, pp. 118–122. ISSN: 0165-1765. DOI: <https://doi.org/10.1016/j.econlet.2018.11.007>. URL: <https://www.sciencedirect.com/science/article/pii/S0165176518304634>.
- Tetlock, Paul (Feb. 2007). “Giving Content to Investor Sentiment: The Role of Media in the Stock Market”. In: *Journal of Finance* 62, pp. 1139–1168. DOI: 10.2139/ssrn.685145.
- Tual, Stephan (July 30, 2015). *Ethereum Launches*. Ethereum Foundation. URL: <https://blog.ethereum.org/2015/07/30/ethereum-launches> (visited on 09/28/2026).
- U.S. Securities and Exchange Commission (May 23, 2024). *Order Granting Accelerated Approval of Proposed Rule Changes, as Modified by Amendments Thereto, to List and Trade Shares of Ether-Based Exchange-Traded Products*. URL: <https://www.sec.gov/rule-release/34-100224> (visited on 09/28/2026).
- Urquhart, Andrew (2016). “The inefficiency of Bitcoin”. In: *Economics Letters* 148, pp. 80–82. ISSN: 0165-1765. DOI: <https://doi.org/10.1016/j.econlet.2016.09.019>. URL: <https://www.sciencedirect.com/science/article/pii/S0165176516303640>.
- Wilder Jr., J. Welles (1978). *New Concepts in Technical Trading Systems*. Greensboro, NC: Trend Research. ISBN: 9780894590276. URL: <https://windsorpublishing.com/product/new-concepts-in-technical-trading-systems/>.
- XRP Ledger (2026). *History*. URL: <https://xrpl.org/about/history> (visited on 09/28/2026).