Disentangling Investor Behavior in Bitcoin Bubbles: Sentiment, Price Expectations, and Nonlinear Dynamics

Enrico C. Mira*[1], Leandro S. Maciel[2], Jéfferson Colombo[3]

[1]Department of Economics, School of Economics, Business, Accounting and Actuary, University of Sao Paulo, Av. Prof. Luciano Gualberto 908, 05508-010, Sao Paulo, Brazil

[2]Department of Business Administration, School of Economics, Business, Accounting and Actuary, University of Sao Paulo, Av. Prof. Luciano Gualberto 908, 05508-010, Sao Paulo, Brazil

[3]Sao Paulo School of Economics, Fundação Getulio Vargas (FGV-EESP), Rua Doutor Plínio Barreto 365, 01313-020, Sao Paulo, Brazil

Abstract

The decoupling of cryptocurrency valuations from firm-specific and macroeconomic fundamentals renders digital assets particularly prone to speculative bubbles driven by investor behavior. This paper examines the behavioral determinants of Bitcoin bubbles by distinguishing general investor sentiment from directional price expectations that capture beliefs about future price dynamics. We construct a large textual dataset covering 2017–2025 from social media and professional news using natural language processing and large language models. Our two-stage econometric framework first identifies bubble episodes through recursive GSADF testing and then evaluates the predictive role of behavioral factors using Logit models and nonlinear machine learning classifiers under time-ordered validation. We document three key results: directional beliefs dominate broad sentiment in explaining bubbles; news signals perform better in linear models, while social media excels in nonlinear settings; and investor behavior exhibits a fundamentally nonlinear relationship with price explosivity.

Keywords: Bitcoin, Speculative Bubbles, Investor Sentiment, Textual Analysis, LLMs.
JEL Code: G12, G41, C53, C58.

*Corresponding author. E-mail address: enrico.c.mira@usp.br.

1 Introduction

The valuation of digital assets constitutes one of the most complex challenges in modern financial economics (Aimable, 2026). Unlike equities or fixed-income assets, where pricing models are anchored in discounted cash flows, dividend streams, or firm/macroeconomic fundamentals, cryptocurrencies lack a clearly defined intrinsic value (Cheah and Fry, 2015). For instance, García-Monleón et al. (2021) propose a theoretical framework in which intrinsic value may arise from the functional role of cryptocurrencies and the value generated by their underlying networks. Nevertheless, the lack of consensus on how to operationalize such intrinsic value decouples cryptocurrency prices from traditional valuation metrics, rendering the market uniquely susceptible to speculative behavior and recurrent episodes of price exuberance (Cheung et al., 2015, Chowdhury and Damianov, 2024, Han and Wang, 2025). In the absence of fundamental anchors, market capitalization becomes heavily dependent on user adoption, network effects and, predominately, on the beliefs and expectations of market participants (Ahn and Kim, 2023, Liu and Tsyvinski, 2021, Wu and Li, 2026). Consequently, understanding the formation of bubbles in this environment requires a rigorous examination of how information—specifically investor sentiment and price expectations—is processed and transmitted into prices (Manahov and Urquhart, 2021, Osman et al., 2024, van Eyden et al., 2023).

Theoretical finance literature distinguishes between rational and irrational bubbles (Boehl and Hommes, 2025, Krichevskiy and Qirjo, 2017, Mesly, 2023). Rational bubbles can emerge in models with rational agents where the asset price contains a bubble component that grows at the required rate of return, satisfying the no-arbitrage condition (Blanchard and Watson, 1982). However, the cryptocurrency market is often characterized by episodes of "irrational exuberance", in which price explosions are driven by behavioral biases, herding, and the activity of noise traders who rely on sentiment rather than fundamental information (De Long et al., 1990, Shiller, 2000). In this context, prices deviate explosively from any proxy of fundamental value due to self-fulfilling prophecies. This is consistent with the framework proposed by Wu et al. (2025), suggesting that the interaction between fundamentalist investors—anchoring expectations to intrinsic, network-based fundamentals—and speculative traders—extrapolating past returns—can give rise to market dynamics in which speculative forces dominate, amplifying persistent and sometimes explosive deviations from fundamental value. To empirically identify price deviations, the literature has relied on recursive testing procedures developed by Phillips et al. (2011) and generalized by Phillips et al. (2015a,b). The Generalized Supremum Augmented Dickey-Fuller (GSADF) test, proposed by Phillips et al. (2015b), provides a robust framework for detecting multiple periodically collapsing bubbles by testing for a transition from a random walk to an explosive autoregressive process.

The extensive application of the GSADF framework in digital asset markets has established a robust empirical foundation for identifying explosive price behavior.[1] Cheung et al. (2015) were among the first to document multiple bubble episodes in the Bitcoin market, attributing these deviations to internal market frictions and major exchange hacks. Moreover, Geuder et al. (2019) and Chaim and Laurini (2019) expanded the scope of this testing framework within the crypto-sphere, utilizing it to disentangle genuine explosive bubble dynamics from periods of mere high volatility, thereby offering a more precise chronology of market exuberance. Subsequent studies such as Bouri et al. (2019) and Enoksen et al. (2020) have corroborated these findings across a broader basket of altcoins, validating the GSADF test's efficacy in capturing the frequent and intense boom-and-bust cycles characteristic of the crypto-asset ecosystem. A comprehensive survey of the cryptocurrency literature documents recurrent pricing bubbles across major digital assets, particularly Bitcoin, Ethereum, and Litecoin, with pronounced bubble phases around 2013 and 2017 (Kyriazis et al., 2020).

While the financial econometric methods for bubble detection are well-established, the behavioral drivers underpinning the episodes of price exuberance remain a subject of intense debate (Li et al., 2025, Steiger and Pelster, 2020, Zhang et al., 2024). In particular, relatively few studies explicitly investigate these behavioral mechanisms in asset markets, particularly for cryptocurrencies where pricing dynamics are arguably more prone to sentiment-driven distortions (Han, 2025, Lin et al., 2023). For instance, Pan (2020) investigate stock market bubbles using the GSADF framework and proxies investor sentiment

[1] Beyond digital assets, the GSADF methodology has been extensively applied to identify speculative bubbles in a wide range of asset classes, such as real estate (Mira et al., 2025, Pavlidis et al., 2016), equity markets (Acharya, 2024, Phillips et al., 2015a), commodities (Ozgur et al., 2021), and foreign exchange (Hu and Oxley, 2017).

through consumer confidence indices, capturing aggregate optimism in traditional equity markets. More recently, Guo et al. (2025) extend the GSADF approach to non-fungible token markets and analyze whether bubble dynamics are correlated with cryptocurrency prices and market sentiment indicators such as the VIX (CBOE volatility index), global economic policy uncertainty, and Google Trends–based attention measures. While these studies provide evidence that sentiment-related variables are statistically associated with bubble formation and persistence, the notion of sentiment they employ remains broad and largely undifferentiated. In particular, these measures conflate distinct behavioral channels—such as affective mood, uncertainty, attention, and expectations—without explicitly identifying which dimension of sentiment is being captured, nor how different behavioral components may operate across digital asset markets.

A critical, yet often overlooked, distinction in this literature is the difference between investor sentiment and price expectations (Baker and Wurgler, 2006, Shiller, 2017). Sentiment broadly refers to the subjective and emotional state of the market—the prevailing atmosphere of bullish or bearish tendencies regarding the asset class, often driven by exogenous news or social trends (Baker and Wurgler, 2006). In contrast, price expectations represent the cognitive component of investor behavior—explicit beliefs regarding the future trajectory of asset prices (e.g., beliefs that prices will rise or fall, or beliefs about momentum) (Greenwood and Shleifer, 2014). While related, these constructs are distinct; an investor may hold a generally positive sentiment toward the underlying technology or the broader market ecosystem yet believe the asset is currently overvalued—negative price expectations (Gennaioli and Shleifer, 2018, Shiller, 2000). Disentangling these effects is crucial for understanding whether bubbles are driven by blind emotional exuberance or by coordinated, albeit potentially irrational, beliefs about price trends (De Long et al., 1990, Shiller, 2017).

Addressing this gap, this paper advances the understanding of speculative dynamics in the Bitcoin market by empirically disentangling investor sentiment from price expectations. We exploit a rich and granular dataset from the Thomson Reuters MarketPsych database, constructed using large-scale natural language processing techniques and state-of-the-art Large Language Models (LLMs), which enables a clear distinction between affective market sentiment and explicit beliefs about future price dynamics. Unlike prior studies that rely on coarse proxies such as trading volume or aggregate search intensity (Goodle Trends) (Biswas and Sharma, 2025, Guo et al., 2025, Neto, 2021, Pan, 2020, Urquhart, 2018), our measures are constructed from large-scale Social media and News media data, which enable the textual classification of narratives into sentiment- and expectation-based components. This approach provides a novel lens through which to assess whether Bitcoin price exuberance is driven primarily by affective sentiment or by coordinated, expectation-driven speculative behavior.

The dataset employed in this study enables the construction of a rich set of complementary text-based metrics that capture distinct dimensions of investor behavior. Specifically, we develop four behavioral classes to assess the role of investor behavior in driving episodes of price exuberance. The first class, "Sentiment", reflects the overall polarity of market mood by contrasting the intensity of positive and negative references. The second class, "Optimism", captures the broader psychological outlook of market participants by comparing the prevalence of optimistic versus pessimistic narratives surrounding the asset. The third class, "Direction" (or price expectations), directly measures belief-based components of behavior by tracking textual references to anticipated price increases relative to expected declines. Finally, "Momentum" is defined as the net balance between references to strengthening and weakening price trends. This momentum-based metric is particularly relevant for bubble analysis, as speculative bubbles are inherently characterized by persistent, self-reinforcing price dynamics and nonlinear deviations from the random walk hypothesis (Basher and Sadorsky, 2022, Hong and Stein, 1999).

Building on the behavioral metrics derived from textual data, our empirical strategy integrates two complementary components. First, we apply the GSADF test (Phillips et al., 2015b) to Bitcoin prices in order to identify and date-stamp periods of explosive behavior consistent with speculative bubbles. Second, conditional on these identified regimes, we estimate a set of explanatory models to assess the role of investor behavior in characterizing bubble dynamics. Specifically, we evaluate the predictive and classificatory power of the Sentiment, Optimism, Direction, and Momentum classes using both linear Logit and nonlinear Random Forest (RF) classifiers. This two-stage framework allows us to examine whether behavioral factors contribute to market instability through complex and potentially nonlinear transmission mechanisms (Basher and Sadorsky, 2022), and to assess their effectiveness in discriminating bubble from non-bubble periods.

This paper contributes to the econometric modeling of speculative dynamics by integrating large-scale textual belief measures with nonlinear predictive methods. First, we show that explicit directional expectations outperform generalized sentiment indicators in forecasting bubble regimes, suggesting that speculative exuberance is driven by coordinated forward-looking beliefs rather than diffuse optimism (Greenwood and Shleifer, 2014). Second, a disaggregated sentiment analysis reveals asymmetric effects: positive sentiment amplifies the likelihood of price explosiveness, whereas negative sentiment acts as a countervailing force, dampening bubble formation. Third, the superior out-of-sample performance of nonlinear machine learning models relative to linear specifications indicates threshold and interaction effects consistent with an inherently nonlinear relationship between investor beliefs and explosive price dynamics. Fourth, we document that social and news media provide complementary predictive information, as models combining both sources outperform single-channel specifications, implying that peer-to-peer communication interacts with structured news narratives to propagate speculative contagion in retail-driven markets (Chen et al., 2014, Da et al., 2011, Garcia et al., 2014). Taken together, these findings demonstrate how machine learning methods, embedded within a disciplined econometric framework, enhance real-time bubble risk monitoring in cryptocurrency markets.

The remainder of the paper is structured as follows. Section 2 describes the methodology, detailing the GSADF procedure and the construction of the textual sentiment and expectation models for bubble detection. Section 3 presents the empirical results, describing the data, identifying bubble periods, and assessing whether different dimensions of investor behavior influence bubble formation through linear or nonlinear channels. Section 4 concludes and suggests topics for further investigation.

2 Methodology

The empirical approach follows a two-stage framework to investigate both the identification and the determinants of speculative bubbles in the cryptocurrency market, with a specific focus on Bitcoin. In the first stage, the GSADF test is employed to detect periods of explosive price behavior, which are subsequently translated into binary indicators that precisely classify bubble and non-bubble regimes over time. In the second stage, these binary bubble indicators serve as the dependent variable in a set of explanatory and predictive models—namely Logit regressions and Random Forest classifiers—designed to evaluate the role of investor sentiment and price expectation metrics in shaping the likelihood of bubble occurrence, while controlling for market microstructure factors such as volatility and trading volume.

2.1 GSADF test for exuberance detection

To detect periods of exuberance in the Bitcoin price series, we adopt the GSADF test developed by Phillips et al. (2015a,b). The GSADF test is based on a recursive implementation of the right-tailed Augmented Dickey-Fuller (ADF) (Dickey and Fuller, 1981) procedure. It evaluates multiple subsamples of the data instead of relying solely on the full sample; hence, it is necessary to establish precise notation for the normalized time scale and the construction of the estimation windows.

The complete sample is normalized to the interval $[0, 1]$, such that each point in time corresponds to a fraction of the total number of observations T. The parameters r_1 and r_2 represent the fractional positions that mark the beginning and the end of a given subsample, with $0 \leq r_1 < r_2 \leq 1$. The regression window length is defined as $r_w = r_2 - r_1$. The parameter r_0 denotes the fixed minimum initial window length required by the econometrician, ensuring that any subsample ending at r_2 satisfies $r_2 \in [r_0, 1]$.

The empirical model employed for the test corresponds to the ADF auxiliary regression equation:

$$\Delta y_t = a_{r_1,r_2} + \gamma_{r_1,r_2} y_{t-1} + \sum_{j=1}^{k} \psi_{j,r_1,r_2} \Delta y_{t-j} + \varepsilon_t, \tag{1}$$

where $\Delta y_t = y_t - y_{t-1}$, y_t denotes the price of the cryptocurrency at instant t, for $t = 1, \ldots, T$; T is the sample size; a_{r_1,r_2}, γ_{r_1,r_2}, ψ_{j,r_1,r_2} are model parameters, for $j = 1, \ldots, k$, and k is the number of lagged augmented terms; and $\varepsilon_t \sim \mathcal{N}(0, \sigma_{r_1,r_2})$ is a normally-distributed error term.

Following Phillips et al. (2015a,b), we set the minimum window size to $r_0 = 0.01 + 1.8/\sqrt{T}$. The rolling window is defined over the fractional sample $[r_1, r_2]$, where r_2 moves from r_0 to 1, and for each r_2, r_1 varies from 0 up to $r_2 - r_0$, ensuring a minimum window length of r_0. This procedure generates a sequence of expanding and shifting windows that scan the entire sample to identify periods of explosivity. The lag order is fixed at $k = 1$ to eliminate additional lagged terms and concentrate exclusively on contemporaneous dynamics, given that Vasilopoulos et al. (2022) report that lag selection based on information criteria can cause substantial size distortions.

The null hypothesis of a unit root is given by $H_0 : \gamma_{r_1,r_2} = 0$, against the alternative of explosive behavior: $H_1 : \gamma_{r_1,r_2} > 0$. To test these hypotheses, the Backward Supremum ADF (BSADF) statistic is calculated as follows:

$$BSADF_{r_2} = \sup_{r_1 \in [0, r_2 - r_0]} \frac{\hat{\gamma}_{r_1,r_2}}{\hat{\text{se}}\left(\hat{\gamma}_{r_1,r_2}\right)}, \tag{2}$$

where $\hat{\text{se}}(\hat{\gamma}_{r_1,r_2})$ denotes the standard error of the estimated coefficient $\hat{\gamma}_{r_1,r_2}$.

To detect episodes of explosive behavior within the full sample, the GSADF statistic is defined as the supremum of the BSADF sequence:

$$GSADF(r_0) = \sup_{r_2 \in [r_0, 1]} BSADF_{r_2}. \tag{3}$$

To identify the timing of exuberance episodes, we apply a date-stamping algorithm based on the trajectory of the $BSADF_{r_2}$ statistic. The start ($\hat{r}_e$) and end ($\hat{r}_f$) of an explosive episode are defined as:

$$\hat{r}_e = \inf_{r_2 \in [r_0, 1]} \left\{ r_2 : BSADF_{r_2} > cv_{r_2}^{\alpha} \right\}, \tag{4}$$

$$\hat{r}_f = \inf_{r_2 \in [\hat{r}_e, 1]} \left\{ r_2 : BSADF_{r_2} < cv_{r_2}^{\alpha} \right\}, \tag{5}$$

where $cv_{r_2}^{\alpha}$ denotes the right-tailed critical value at significance level α.

Given the non-standard distribution of the GSADF statistic and the potential for non-stationary volatility in cryptocurrency data, we adopt the wild bootstrap resampling scheme proposed by Phillips and Shi (2018) to compute the critical values.

Once the periods of explosive bubbles have been identified for the cryptocurrency, the next step is to examine the interrelationships and potential spillover effects during these episodes of price exuberance.

2.2 Investor sentiment and price expectation on bubble dynamics

To assess whether investor sentiment and price expectations predict speculative bubbles in the Bitcoin market, we model the relationship between behavioral indicators and the exuberance periods identified in the previous stage. Let B_t denote a binary variable indicating whether the asset is in a bubble regime at time t, where $B_t = 1$ corresponds to periods of explosive price behavior detected by the GSADF test and $B_t = 0$ otherwise. Having established this binary classification of bubble regimes in the first stage using the GSADF test, the objective of this second stage is to formally characterize the determinants of bubble occurrence by linking B_t to measures of investor behavior and market conditions.

Formally, we assume that the probability of observing a bubble regime at time t is a function of a vector of observable covariates $\mathbf{Z}_t$, represented by the following general probabilistic representation:

$$P(B_t = 1 \mid \mathbf{Z}_t) = f\Big(\beta_0 + \mathbf{\Phi}_t^\top \boldsymbol{\beta} + \mathbf{C}_t^\top \boldsymbol{\gamma}\Big), \tag{6}$$

where $f(\cdot)$ denotes a function mapping the bubbles determinant relationship, which may take different functional forms depending on the estimation approach (e.g., a logistic function in the case of Logit models); $\mathbf{Z}_t = \{\mathbf{\Phi}_t, \mathbf{C}_t\}$, where the vector $\mathbf{\Phi}_t$ represents a set of textual-based metrics capturing distinct dimensions of investor behavior—such as sentiment, optimism, price expectations, and momentum—,while $\mathbf{C}_t$ is a vector of control variables related to market microstructure, including volatility and trading volume. This specification provides the baseline framework for the explanatory and predictive models estimated in this stage, serving as the foundation for both linear and nonlinear classification approaches used to identify the behavioral drivers of speculative bubble regimes.

The vector $\mathbf{\Phi}_t$ is flexibly specified to incorporate different classes of investor behavior indicators, depending on the behavioral dimension under investigation. Rather than estimating a single model, we consider alternative specifications in which $\mathbf{\Phi}_t$ is populated by regressors associated with a specific behavioral channel. In particular, we investigate four distinct classes of models: sentiment, optimism, price direction, and momentum.

In the sentiment model (Sentiment), $\mathbf{\Phi}_t$ includes measures capturing the overall emotional polarity of market discourse:

$$\mathbf{\Phi}_t^{\text{Sentiment}} = [Positive_t, Negative_t]^\top,$$

where $Positive_t$ is the number of positive references and $Negative_t$ is the number of overall negative references.

The optimism model (Optimism) focuses on forward-looking psychological outlook by distinguishing between optimistic and pessimistic narratives expressed in future-oriented language:

$$\mathbf{\Phi}_t^{\text{Optimism}} = [Optimism_t, Pessimism_t]^\top,$$

where $Optimism_t$ ($Pssimism_t$) is the number of references connoting optimism (pessimism), future-tense positive (negative).

The price direction/expectation model (Direction) captures explicit beliefs regarding the future trajectory of Bitcoin prices by contrasting upward versus downward price expectations:

$$\mathbf{\Phi}_t^{\text{Direction}} = [PriceUp_t, PriceDown_t]^\top,$$

where $PriceUp_t$ ($PriceDown_t$) represents the number of references to price increases (decreases).

Finally, the momentum model (Momentum) isolates trend-following behavior by summarizing the balance between narratives emphasizing price trend strength and those highlighting trend weakness:

$$\mathbf{\Phi}_t^{\text{Momentum}} = Momentum_t,$$

where $Momentum_t$ is the difference between number of references to price trend strength and price trend weakness.

These alternative specifications allow us to assess how different dimensions of investor behavior—ranging from subjective sentiment to explicit price expectations and trend-following beliefs—contribute to the probability of observing speculative bubble regimes.

The investor sentiment and expectation measures employed in this study are obtained from the MarketPsych database. This dataset relies on large language models (LLMs) to process and classify large volumes of unstructured textual data, enabling the construction of high-frequency behavioral indicators from natural language.[2] Importantly, the metrics are computed separately using different information sources, namely: (i) Social media, (ii) News media, and (iii) a combined Social + News corpus. By estimating the models across these alternative data sources, the analysis allows us to assess which channel of information transmission is more strongly associated with speculative bubble regimes in Bitcoin markets. This multi-source framework enriches the empirical investigation by explicitly evaluating whether bubbles are more closely linked to narratives originating from professional news outlets, social discourse, or the interaction between both.

To control for market conditions that are known to be intrinsically linked to speculative dynamics, we explicitly account for the stylized facts of cryptocurrency returns, most notably volatility clustering. Hence, we estimate the conditional variance of Bitcoin returns, denoted by σ_t^2, using a standard GARCH(1,1) specification with a zero-mean return equation:

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{7}$$

where y_t denotes log-returns, ε_t is an i.i.d. innovation term, and ω, α, and β are parameters satisfying the usual positivity and stationarity conditions.

The resulting estimated conditional volatility, $\hat{\sigma}_t^2$, is incorporated as a key control variable in the vector $\mathbf{C}_t$. Controlling for volatility is essential, as periods of heightened uncertainty and risk are often

[2]For more details see: `https://www.marketpsych.com/home`.

intertwined with speculative behavior and explosive price dynamics in financial markets. In parallel, we include trading volume ($Volm_t$) as an additional control, reflecting the intensity of market participation. Both variables are motivated by theoretical and empirical models linking speculative bubbles to elevated trading activity and increased return variance, as rational and noise traders interact under heterogeneous beliefs (Mei et al., 2005, Scheinkman and Xiong, 2003). By conditioning on these market microstructure factors, $\mathbf{C}_t = [\hat{\sigma}_t^2, Volm_t]^\top$, the empirical framework isolates the incremental role of investor sentiment and price expectations in explaining the likelihood of bubble regimes.[3]

We estimate four distinct model specifications, each corresponding to a specific dimension of investor behavior embedded in the vector $\mathbf{\Phi}_t$. In all cases, the probability of observing a bubble regime is modeled through a generic link function $f(\cdot)$, allowing for both linear and nonlinear classification methods.

Model I (Sentiment). In the sentiment specification, the vector $\mathbf{\Phi}_t^{\text{Sentiment}}$ captures the overall emotional polarity of market discourse:

$$\mathbf{\Phi}_t^{\text{Sentiment}} = (Positive_t,\ Negative_t)\,, \tag{8}$$

$$P(B_t = 1) = f\Big(\alpha + \beta_1 Positive_t + \beta_2 Negative_t + \mathbf{C}_t^\top \boldsymbol{\gamma}\Big)\,. \tag{9}$$

Model II (Optimism). The optimism specification focuses on forward-looking psychological narratives expressed in future-oriented language:

$$\mathbf{\Phi}_t^{\text{Optimism}} = (Optimism_t,\ Pessimism_t)\,, \tag{10}$$

$$P(B_t = 1) = f\Big(\alpha + \beta_1 Optimism_t + \beta_2 Pessimism_t + \mathbf{C}_t^\top \boldsymbol{\gamma}\Big)\,. \tag{11}$$

Model III (Direction). This specification captures explicit beliefs regarding the future direction of Bitcoin prices:

$$\mathbf{\Phi}_t^{\text{Direction}} = (PriceUp_t,\ PriceDown_t)\,, \tag{12}$$

$$P(B_t = 1) = f\Big(\alpha + \beta_1 PriceUp_t + \beta_2 PriceDown_t + \mathbf{C}_t^\top \boldsymbol{\gamma}\Big)\,. \tag{13}$$

Model IV (Momentum). Finally, the momentum specification isolates trend-following behavior:

$$\mathbf{\Phi}_t^{\text{Momentum}} = (Momentum_t)\,, \tag{14}$$

$$P(B_t = 1) = f\Big(\alpha + \beta_1 Momentum_t + \mathbf{C}_t^\top \boldsymbol{\gamma}\Big)\,. \tag{15}$$

In all models, $\mathbf{C}_t$ denotes the vector of control variables capturing market microstructure conditions, including conditional volatility and trading volume. Models I and II primarily reflect affective psychological dimensions of investor behavior (general market sentiment), whereas Models III and IV capture explicit cognitive beliefs about future price movements and trend persistence. $f(\cdot)$ is implemented using both a logistic regression and Random Forest classifiers. By estimating $f(\cdot)$ through these alternative approaches, we are able to explicitly test whether the transmission of investor behavior to bubble formation operates through linear mechanisms or through more complex, nonlinear relationships characterized by threshold effects and interaction patterns.

Finally, it is worth noting that the raw investor behavior metrics considered in this study, *Positive*, *Negative*, *PriceUp*, *PriceDown*, *Optimism*, *Pessimism*, and *Momentum*, are originally provided by the MarketPsych database as relative frequencies, expressed as percentages of total textual references. To capture the aggregate magnitude of these beliefs, we rescale each metric by multiplying it by the Buzz index, also provided by MarketPsych, which measures the overall volume of asset-specific references. The Buzz index is normalized to lie between zero and one, so that this transformation converts relative frequency measures into volume-weighted indicators, effectively reflecting the absolute intensity of each

[3] Additional control variables, such as economic growth and interest rates, could be included; however, since these measures are specific to individual economies, they may introduce distortions when analyzing the cryptocurrency market, which is inherently global.

narrative driving market behavior.

3 Empirical analyses

This section reports the empirical analyses investigating the dynamics of speculative bubbles in Bitcoin prices. First, we describe the data related to Bitcoin prices and the different investor behavior metrics considered. Second, periods of price exuberance are identified using the GSADF test. Third, we examine the role and predictive power of investor sentiment and price expectations in driving bubble episodes using both linear and nonlinear model specifications.

3.1 Data

Data are comprised of daily closing prices and trading volume in USD for Bitcoin (BTC) for the period from January 1, 2017, to August 31, 2025 (a total of 2,803 observations).[4] Bitcoin was selected due to its position as the leading cryptocurrency in terms of market capitalization and liquidity, serving as a proxy for the broader digital asset market. The sample period spans from the beginning of 2017 to the end of 2025. The 2017 start date reflects the onset of sustained increases in trading activity and market participation, ensuring sufficient liquidity and price discovery for reliable analysis. The period concludes at the end of 2025, aligning with the data collection cutoff for our empirical experiments. This period allows for a robust assessment of the presence and dynamics of speculative bubbles, covering multiple boom and bust cycles over nearly eight years.

Panel A of Table I presents descriptive statistics for the daily Bitcoin price, trading volume, and volatility over the sample period. The results highlight the pronounced variability of Bitcoin prices, with a high standard deviation, reflecting the well-documented dynamics of the cryptocurrency market. Trading volume exhibits substantial dispersion, strong right skewness, and excess kurtosis. Volatility, estimated as the conditional variance from a GARCH(1,1) model as described in Subsection 2.2—Eq. (7), displays significant skewness and kurtosis.

[4] Data were extracted from Yahoo Finance: `https://finance.yahoo.com`. Accessed on August 2, 2025.

Table I. **Descriptive Statistics of the Database**.

Variable	**Mean**	**Maximum**	**Minimum**	**Std. Dev.**	**Skewness**	**Kurtosis**
Panel A: Volume and volatility						
Volume	27,124,650,027.36	350,967,941,479.00	60,851,700.00	22,546,998,856.54	2.13	18.20
Volatility	0.03	0.12	0.02	0.01	2.04	9.63
Panel B: Behavioral metrics extracted from Social media.						
Positive	0.05	0.23	0.00	0.03	1.18	6.81
Negative	0.04	0.29	0.00	0.03	1.35	10.45
Optimism	0.01	0.05	0.00	0.01	0.33	3.73
Pessimism	0.01	0.06	0.00	0.01	0.76	6.75
PriceUp	0.01	0.05	0.00	0.01	1.81	10.56
PriceDown	0.01	0.04	0.00	0.00	2.10	13.88
Momentum	0.00	0.00	-0.00	0.00	1.87	13.28
Panel C: Behavioral metrics extracted from News media.						
Positive	0.04	0.28	0.00	0.03	1.66	9.85
Negative	0.04	0.25	0.00	0.02	1.93	11.37
Optimism	0.01	0.08	0.00	0.01	1.62	10.17
Pessimism	0.01	0.04	0.00	0.00	1.18	6.59
PriceUp	0.01	0.08	0.00	0.01	2.47	13.96
PriceDown	0.01	0.05	0.00	0.00	2.85	17.27
Momentum	0.00	0.01	-0.00	0.00	2.62	17.75
Panel D: Behavioral metrics extracted from News Headline media.						
Positive	0.03	0.24	0.00	0.02	2.30	14.64
Negative	0.04	0.46	-0.00	0.03	3.52	25.81
Optimism	0.01	0.05	-0.00	0.01	1.60	7.67
Pessimism	0.01	0.04	-0.00	0.01	1.81	8.18
PriceUp	0.01	0.09	-0.00	0.01	3.96	27.18
PriceDown	0.01	0.11	-0.00	0.01	4.48	36.83
Momentum	0.00	0.01	-0.01	0.00	1.32	20.22
Panel E: Behavioral metrics extracted from Social & News corpus & News Headline.						
Positive	0.05	0.24	0.00	0.03	1.14	6.85
Negative	0.05	0.28	0.00	0.03	1.46	10.77
Optimism	0.01	0.05	0.00	0.01	0.32	3.96
Pessimism	0.01	0.06	0.00	0.01	0.76	6.61
PriceUp	0.01	0.06	0.00	0.01	1.83	10.59
PriceDown	0.01	0.05	0.00	0.00	2.33	15.12
Momentum	0.00	0.00	-0.00	0.00	1.93	12.83

Notes: This table provides the descriptive statistics of Bitcoin market (price, volume and volatility) data and investor behavior variables for Bitcoin over the period from January 1, 2017, to August 31, 2025.

To capture the behavioral drivers of market dynamics, this study employs the Thomson Reuters MarketPsych Indices (TRMI), a comprehensive set of textual-based behavioral indicators specifically designed to capture investor sentiment and related psychological dimensions in financial markets. The TRMI dataset provides daily, seven-days-a-week measures, originally available at the country level and, more recently, extended to cryptocurrency-specific sentiment indicators, including Bitcoin. The index family encompasses a wide range of behavioral constructs—such as Sentiment, Optimism, Buzz, Trust, Fear, and Joy—which are derived from large-scale textual analysis of both traditional news outlets and social media platforms.

MarketPsych's data are generated using advanced natural-language processing techniques, relying on structured grammatical templates specifically tailored to extract financially relevant meanings from unstructured text. These templates are applied to millions of documents drawn from professional news media and social media, allowing the construction of synthetic quantitative indicators that reflect emotional tone, beliefs, and expectations embedded in market narratives. As documented in prior studies, the TRMI framework has been shown to add significant value in financial applications by elucidating how information from heterogeneous media sources is incorporated into asset prices, return predictability, herding behavior, investment decisions, and crash risk (Eierle et al., 2022, Filip and Pochea, 2023, Gan et al., 2020, Gathergood et al., 2023, Kalyvas et al., 2020, Sun et al., 2016).

This paper is the first to employ MarketPsych sentiment and expectation measures to explicitly analyze the dynamics of speculative bubbles. By linking these textual indicators to periods of explosive price behavior identified through econometric bubble tests, the study extends the literature providing new evidence on how investor narratives and expectations contribute to bubble formation and persistence. Moreover, the availability of sentiment measures constructed separately from news media, social media, and their combination allows the analysis to assess which information channel plays a more prominent role in driving bubble dynamics, thereby offering a richer and more nuanced understanding of belief formation in cryptocurrency markets.

The set of behavioral metrics selected from the TRMI database corresponds exactly to those defined in the Subsection 2.2 and is used to operationalize the four empirical models of investor behavior—Sentiment, Optimism, Direction, and Momentum. TRMI dataset covers the same sample period: January 1, 2017, to August 31, 2025. Panels B, C, and D of Table I report descriptive statistics for these metrics computed from different information sources, namely Social media, News media, and their combined Social & News corpus. In general, the statistics reveal substantial variability, pronounced right skewness, and excess kurtosis across most behavioral indicators, indicating the presence of episodic surges in attention and belief formation. Social media–based measures exhibit markedly higher means and dispersion relative to news-based metrics, reflecting the larger volume and higher intensity of discourse in social platforms. By contrast, news-based indicators tend to display sharper asymmetries and heavier tails, consistent with clustered information releases. The combined Social & News measures amplify these patterns.

3.2 Exuberance periods in the Bitcoin market

Periods of price exuberance for Bitcoin are identified with the GSADF test. The test is applied to the price series to detect episodes of explosive behavior indicative of speculative bubbles. Table II presents the results of the GSADF test. Critical values were generated using 500 bootstrap replications.[5] The testing procedure followed the implementation provided by the "`exuber`" R-package, developed by Vasilopoulos et al. (2022).

The findings in Table II indicate that Bitcoin rejects the null hypothesis of a unit root, with the GSADF statistic significant at the 1% level. This finding provides strong statistical evidence of explosive price behavior—a key characteristic of speculative bubbles—within the sample. As detailed in Table II, Bitcoin experienced a total of 623 days classified as bubble episodes.

Figure 1 provides a visual representation of the historical price dynamics for Bitcoin, highlighting the exuberance periods identified by the GSADF test (shaded grey areas) and distinct clusters of explosive behavior. The test detects intermittent bubble episodes throughout 2019, followed by a prolonged period of exuberance starting in mid-2020. Notably, the period spanning from mid-2020 through early 2022 represents a significant phase of market explosivity. This timeframe coincides with a sharp increase in cryptocurrency valuation, likely driven by the entry of traditional market participants and institutional investors. A prominent example during this phase was the high-profile investment in Bitcoin by Tesla, the electric vehicle manufacturer led by Elon Musk, which signaled growing institutional acceptance and fueled further price appreciation.

[5] Bootstrap critical values are based on 500 replications, a commonly adopted choice in GSADF applications that ensures stable inference without excessive computational burden, according to Phillips and Shi (2018) and Vasilopoulos et al. (2022).

Table II. **Results from the GSADF test for Bitcoin**.

GSADF Statistic	**# of bubble days**	**Bubble Dating (start–end)**
9.29***	623	05/10/17–06/26/17; 06/27/17–06/30/17; 07/04/17–07/05/17 08/05/17–08/06/17; 08/07/17–08/09/17; 08/10/17–09/13/17 10/09/17–10/10/17; 10/11/17–11/12/17; 11/13/17–12/30/17 12/31/17–01/01/18; 01/02/18–01/09/18; 01/10/18–01/11/18 11/24/18–11/28/18; 05/09/19–06/04/19; 06/14/19–07/01/19 07/02/19–07/05/19; 07/06/19–07/11/19; 07/12/19–07/13/19 11/05/20–11/07/20; 11/08/20–11/09/20; 11/11/20–11/26/20 11/28/20–12/08/20; 12/09/20–12/10/20; 12/12/20–05/15/21 10/15/21–10/24/21; 10/25/21–10/26/21; 10/29/21–11/05/21 11/06/21–11/16/21; 11/10/23–11/12/23; 12/04/23–12/11/23 12/13/23–12/15/23; 12/21/23–12/23/23; 01/08/24–01/09/24 01/10/24–01/12/24; 02/12/24–04/17/24; 04/19/24–04/26/24 05/20/24–05/23/24; 05/25/24–05/26/24; 06/04/24–06/07/24 06/09/24–06/11/24; 11/11/24–01/09/25; 01/10/25–01/11/25 01/12/25–02/02/25; 02/03/25–02/05/25

Notes: The table reports the GSADF statistic, the number of days identified as bubble periods, and the corresponding start and end dates (bubble dating) for each detected bubble episode. The total sample comprises 2,803 days. (***) denotes statistical significance at the 1% level.

Furthermore, a substantial and sustained bubble episode beginning in late 2023 and extending through mid-2025 (see in Table II). This recent wave of exuberance reflects a period of pronounced price appreciation in late 2024 and early 2025, which can be associated with the approval of spot Bitcoin Exchange Traded Funds (ETFs) and a more favorable regulatory outlook.[6]

With the bubble periods identified, we proceed to examine the extent to which investor behavior contributes to their formation, considering the cryptocurrency market's heightened susceptibility to market sentiment (Han, 2025, Li and Ma, 2024).

3.3 The role of investor sentiment and price expectations in cryptocurrency market bubbles

To investigate the determinants of the identified exuberance periods, we estimate four linear Logit specifications in which the binary bubble indicator is modeled as a function of investor behavior metrics capturing Sentiment (Model I), Optimism (Model II), price Direction (Model III), and price Momentum (Model IV), as detailed in Subsection 2.2.

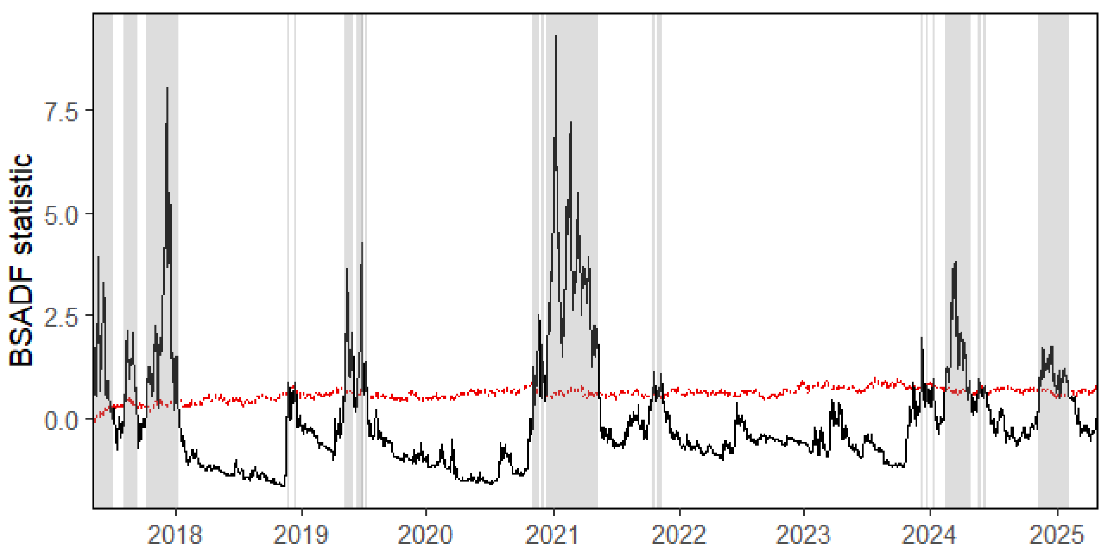
Table III reports the estimation results for these models using three alternative data sources: Social media, News media, and the combined Social & News corpus. The results provide strong evidence of the predictive power of the selected parameters. Notably, the behavioral variables are statistically significant at the 1% level across almost all model specifications.[7] This result holds both in the baseline specification that includes only behavioral measures ("Sent") and in the augmented specification that jointly incorporates behavioral variables and standard market controls—namely volatility and trading volume ("Sent + Ctrl"). These complementary specifications are estimated to assess whether sentiment-based measures provide additional explanatory power beyond that captured by conventional control variables alone. This consistency confirms the robustness of the results, indicating that the explanatory power of investor sentiment and price expectations persists even when controlling for market microstructure factors. Furthermore, the control variables themselves—conditional volatility and trading volume—are

[6]Source: `https://www.bankrate.com/investing/bitcoin-price-history`. Accessed on 19 January 2026.

[7]The stationarity properties of all behavioral and control metrics were assessed using the Augmented Dickey–Fuller (ADF) test of Dickey and Fuller (1981), and the results provided in Table A.1, A showed that the variables are stationary in levels under all test specifications, thereby mitigating concerns regarding spurious inference in the logistic regression models.

significant across all panels, reaffirming that heightened market activity and risk are intrinsic to speculative episodes (Mei et al., 2005, Scheinkman and Xiong, 2003).

[a]Bitcoin prices (in USD) and detected bubbles.



[b]GSADF statistic, bootstrap critical value (red line) and bubbles

Figure 1. Temporal evolution of Bitcoin prices (in USD) and identified exuberance periods (grey shaded areas) based on the GSADF test at the 95% significance level. The BSADF statistic and the bootstrap critical value (dashed red line) are also reported.

A comparative analysis of the information criteria reveals an important distinction between the data sources. As shown in Table III, a comparison of the Akaike Information Criterion (AIC) across information sources indicates that the combined Social & News (Panel D) corpus yields the most informative specifications, consistently achieving the lowest AIC values. Models based exclusively on News media (Panel B) generally rank second, while those relying solely on Social media display comparatively higher AICs. This pattern suggests that social and news-based narratives convey complementary information, and that integrating both sources improves the empirical characterization of price exuberance.

Table III. **Logit Model Parameter Estimates for Bitcoin Bubble Determinants**.

	Model I (Sentiment)		**Model II** (Optimism)		**Model III** (Direction)		**Model IV** (Momentum)		**Bench** (Ctrl)
Variables	**Raw**	**Ctrl**	**Raw**	**Ctrl**	**Raw**	**Ctrl**	**Raw**	**Ctrl**	**Ctrl**
Panel A: Social media.									
Constant	-1.6***	-3.5***	-1.5***	-3.1***	-1.8***	-3.6***	-1.8***	-4.0***	-3.2***
Positive	35.4***	49.2***							
Negative	-26.3***	-76.6***							
Optimism			46.2***	127.5***					
Pessimism			-16.0	-260.7***					
PriceUp					208.6***	172.1***			
PriceDown					-175.6***	-341.1***			
Momentum							2872.2***	2511.1***	
Volatility		42.8***		36.9***		40.4***		38.3***	25.4***
Volume		6.4E-11***		6.9E-11***		5.2E-11***		3.0E-11***	4.1E-11***
AIC	3400.8	2841.1	3542.9	2930.8	3256.0	2857.2	3154.9	2878.9	3140.1
Adj. R^2	0.048	0.205	0.009	0.180	0.089	0.200	0.117	0.194	0.121
MAE	0.354	0.287	0.373	0.298	0.335	0.287	0.324	0.289	0.321
RMSE	0.420	0.379	0.432	0.387	0.408	0.378	0.403	0.379	0.400
Panel B: News media.									
Constant	-1.9***	-3.9***	-1.7***	-3.7***	-2.0***	-3.5***	-1.9***	-3.7***	-3.2***
Positive	46.2***	42.1***							
Negative	-33.2***	-53.4***							
Optimism			130.2***	125.1***					
Pessimism			-111.0***	-216.1***					
PriceUp					135.7***	91.9***			
PriceDown					-106.5***	-184.6***			
Momentum							1449.0***	1205.1***	
Volatility		50.1***		45.4***		33.8***		31.2***	25.4***
Volume		4.4E-11***		4.4E-11***		4.5E-11***		2.9E-11***	4.1E-11***
AIC	3184.0	2813.8	3328.5	2931.3	3163.4	2866.4	3083.2	2869.2	3141.9
Adj. R^2	0.110	0.214	0.070	0.181	0.116	0.199	0.138	0.198	0.122
MAE	0.327	0.282	0.345	0.295	0.324	0.286	0.315	0.288	0.320
RMSE	0.404	0.375	0.415	0.383	0.402	0.377	0.398	0.380	0.399
Panel C: News Headline.									
Constant	-1.7***	-3.6***	-1.3***	-3.2***	-1.7***	-3.5***	-1.2***	-3.4***	-3.3***
Positive	29.0***	13.6***							
Negative	-6.6***	-21.4***							
Optimism			54.6***	20.8**					
Pessimism			-22.5**	-74.7***					
PriceUp					82.8***	35.0***			
PriceDown					-11.2*	-75.2***			
Momentum							543.0***	469.2***	
Volatility		39.2***		31.3***		32.6***		27.0***	26.5***
Volume		5.1E-11***		4.6E-11***		4.8E-11***		4.1E-11***	4.2E-11***
AIC	2765.2	2413.1	2888.5	2507.1	2729.6	2427.5	2788.5	2466.3	2558.1
Adj. R^2	0.057	0.177	0.015	0.145	0.070	0.173	0.049	0.159	0.128
MAE	0.365	0.308	0.385	0.323	0.358	0.309	0.368	0.317	0.330
RMSE	0.427	0.391	0.439	0.401	0.423	0.392	0.428	0.398	0.405
Panel D: Social & News corpus.									
Constant	-1.8***	-3.7***	-1.7***	-3.3***	-2.0***	-3.6***	-2.0***	-4.0***	-3.2***
Positive	44.1***	50.1***							
Negative	-33.7***	-75.8***							
Optimism			96.1***	151.6***					
Pessimism			-63.8***	-295.2***					
PriceUp					205.3***	151.4***			
PriceDown					-161.0***	-290.1***			
Momentum							2566.2***	2242.0***	
Volatility		48.2***		43.0***		39.4***		37.5***	25.4***
Volume		6.1E-11***		6.8E-11***		5.0E-11***		2.8E-11***	4.1E-11***
AIC	3310.3	2796.7	3486.8	2887.5	3161.8	2835.2	3054.5	2827.3	3141.9
Adj. R^2	0.075	0.218	0.026	0.193	0.116	0.208	0.146	0.210	0.122
MAE	0.341	0.281	0.365	0.292	0.322	0.283	0.311	0.283	0.320
RMSE	0.412	0.374	0.427	0.383	0.401	0.376	0.395	0.376	0.399

Notes: This table reports parameter estimates from Logit models examining the determinants of Bitcoin bubble periods under four alternative specifications (Models I–IV), each capturing a distinct dimension of investor behavior. Model I focuses on Sentiment, Model II on Optimism, Model III on price Direction, and Model IV on price Momentum. For each specification, results are presented both without control variables (sentiment-based models only, "Sent") and with controls included ("Sent + Ctrl"). The benchmark specification corresponds to a Logit model estimated solely with control variables, serving as a reference for identifying bubble periods.
Standard errors are omitted for brevity. (***), (**), (*) indicate significance at the 1%, 5%, and 10% levels, respectively. AIC corresponds to the value of the Akaike information criteria of each model.

The parameter estimates for the sentiment variables in Table III align strictly with economic intuition. Across all models, positive sentiment indicators—specifically *Positive*, *Optimism*, and *PriceUp*—exhibit positive and significant coefficients. This confirms that bullish expectations directly fuel the formation of price bubbles. Conversely, negative sentiment metrics—*Negative*, *Pessimism*, and *PriceDown*—display negative coefficients, acting as a countervailing force (Tetlock, 2007).

Collectively, these findings offer a nuanced perspective on the behavioral origins of cryptocurrency bubbles. The relatively stronger performance of News media, and its further improvement when combined with social media, indicates that professionally curated information and peer-to-peer communication jointly shape investor beliefs in cryptocurrency markets (Tetlock, 2007). Furthermore, according to Table III, the fact that variables explicitly capturing price expectations (Model III, Direction) and trend continuity (Model IV, Momentum) provide a better fit than general sentiment metrics indicates that speculative episodes are driven more by the extrapolation of past returns than by unstructured emotional optimism (Garcia et al., 2014, Greenwood and Shleifer, 2014). This evidence supports theoretical frameworks where positive feedback loops, amplified by media narratives, create significant deviations from fundamental value (Hong and Stein, 1999).

3.4 Comparing linear and nonlinear approaches to bubble detection

After establishing the relevance of investor behavior metrics as determinants of Bitcoin bubble episodes, we proceed to assess whether their influence operates through linear or nonlinear channels. To this end, we compare the predictive performance of linear Logit models with nonlinear Random Forest (RF) classifiers.[8] Table IV reports Accuracy and Area Under the Curve (AUC)[9] metrics across three information panels: Social media, News media, and the combined Social & News corpus dataset.

The results from Table IV reveal a substantial performance gap between linear and nonlinear specifications, providing strong evidence that investor behavior impacts market bubbles in a nonlinear manner. Across all panels and model specifications, the Random Forest models consistently outperform their linear (Logit) counterparts. For instance, focusing on the Social media data source (Panel A) and the specification that includes sentiment and control variables, the Random Forest model attains an AUC of 0.931, substantially outperforming the Logit model, which achieves an AUC of 0.790 in Model I based on sentiment-related variables (Positive and Negative). This pattern persists across all sentiment indicators, suggesting that simple linear models fail to capture the complex, likely threshold-dependent dynamics through which investor behavior influences market explosiveness, a finding consistent with recent asset pricing literature demonstrating the superiority of machine learning over standard econometric methods (Gu et al., 2020).

Regarding the sources of information, the results reinforce the findings from the previous analysis: Social media data provides superior predictive signals compared to News media data in the nonlinear model. In the Model I (Sentiment), considering sentiment variables and controls (last column of Table IV), the RF model using Social media data yields an AUC of 0.931 (Panel A), surpassing the AUC of 0.910 obtained using News media data (Panel B). Even when combining both media sources (Panel C), the AUC remains at 0.930, indicating that the marginal contribution of traditional news-based data to the predictive power of social media is negligible. This confirms that retail-driven sentiment on social platforms is the primary informational channel for bubble detection in the Bitcoin market, supporting the view that retail attention is a distinct driver of price pressure (Chen et al., 2014, Da et al., 2011).

[8]The Random Forest models were estimated using 500 trees, with the square root of the available predictors randomly selected at each split. To address class imbalance, we implemented an asymmetric downsampling strategy via stratified bootstrap resampling with replacement, restricting the majority class (non-bubble periods) to a maximum of twice the size of the minority class (bubble periods) in each tree. Additionally, to prevent overfitting, a minimum terminal node size of 5 was imposed. All models and performance metrics were implemented in R using the `randomForest` and `pROC` packages.

[9]Accuracy measures the proportion of correctly classified observations over the total sample, providing an overall indicator of classification performance. The Area Under the Receiver Operating Characteristic Curve (AUC) captures the model's ability to discriminate between bubble and non-bubble periods across all possible classification thresholds, with higher values indicating stronger discriminatory power independent of any specific cutoff.

Table IV. **Performance Comparison Between Logit and Random Forest (RF) Models for Bitcoin Bubble Detection**.

Model specification	**Method**	**Sent** Accuracy	**Sent** AUC	**Sent + Ctrl** Accuracy	**Sent + Ctrl** AUC
Panel A: Social media.					
Model I: Sentiment	Logit	0.768	0.607	0.803	0.800
	RF	0.959	0.992	0.981	0.999
Model II: Optimism	Logit	0.747	0.572	0.799	0.780
	RF	0.963	0.993	0.981	0.999
Model III: Direction	Logit	0.776	0.675	0.811	0.788
	RF	0.946	0.990	0.983	0.999
Model IV: Momentum	Logit	0.772	0.727	0.804	0.780
	RF	0.903	0.968	0.973	0.997
Bench. I - Controls-only	Logit	—	—	0.781	0.724
	RF	—	—	0.958	0.992
Bench. II - Majority classifier	—	0.747	0.500	—	—
Bench. III - Random classifier	—	0.624	0.511	—	—
Panel B: News media.					
Model I: Sentiment	Logit	0.770	0.727	0.810	0.807
	RF	0.946	0.992	0.980	0.998
Model II: Optimism	Logit	0.755	0.682	0.800	0.783
	RF	0.957	0.994	0.981	0.998
Model III: Direction	Logit	0.771	0.745	0.806	0.793
	RF	0.949	0.991	0.980	0.999
Model IV: Momentum	Logit	0.777	0.768	0.796	0.794
	RF	0.909	0.972	0.966	0.996
Bench. I - Controls-only	Logit	—	—	0.782	0.725
	RF	—	—	0.954	0.993
Bench. II - Majority classifier	—	0.747	0.500	—	—
Bench. III - Random classifier	—	0.621	0.503	—	—
Panel C: News Headline.					
Model I: Sentiment	Logit	0.744	0.667	0.789	0.779
	RF	0.957	0.994	0.986	0.999
Model II: Optimism	Logit	0.730	0.593	0.778	0.752
	RF	0.959	0.997	0.983	0.999
Model III: Direction	Logit	0.745	0.701	0.793	0.771
	RF	0.954	0.994	0.981	0.999
Model IV: Momentum	Logit	0.743	0.650	0.773	0.761
	RF	0.875	0.910	0.972	0.996
Bench. I - Controls-only	Logit	—	—	0.770	0.736
	RF	—	—	0.953	0.993
Bench. II - Majority classifier	—	0.731	0.500	—	—
Bench. III - Random classifier	—	0.608	0.494	—	—
Panel D: Social & News corpus.					
Model I: Sentiment	Logit	0.773	0.654	0.813	0.810
	RF	0.951	0.992	0.980	0.999
Model II: Optimism	Logit	0.749	0.601	0.799	0.791
	RF	0.951	0.992	0.982	0.999
Model III: Direction	Logit	0.782	0.713	0.811	0.796
	RF	0.950	0.991	0.976	0.998
Model IV: Momentum	Logit	0.786	0.758	0.804	0.796
	RF	0.911	0.975	0.962	0.996
Bench. I - Controls-only	Logit	—	—	0.782	0.725
	RF	—	—	0.954	0.993
Bench. II - Majority classifier	—	0.747	0.500	—	—
Bench. III - Random classifier	—	0.621	0.503	—	—

Notes: This table reports classification performance metrics—Accuracy and AUC—for Logit and Random Forest (RF) models examining the determinants of Bitcoin bubble periods under four alternative specifications (Models I–IV), each capturing a distinct dimension of investor behavior. Model I focuses on Sentiment, Model II on Optimism, Model III on price Direction, and Model IV on price Momentum. For each specification, results are presented both without control variables (sentiment-based models only, "Sent") and with controls included ("Sent + Ctrl". Three benchmark specifications are also reported for comparison. Bench. I corresponds to a Logit model estimated solely with control variables. Bench. II is the majority-class classifier, which always predicts the dominant class (non-bubble periods). Bench. III is a random classifier that assigns bubble states according to a binomial distribution calibrated to the empirical class proportions observed in the sample.

Furthermore, the findings in Table IV highlight the critical role of investors' price expectations. Among the different sentiment classifications (models I-IV), models incorporating variables explicitly related to price beliefs—such as Sentiment (I) and Direction (III) models (which inherently captures bullish/bearish divergence)—exhibit the highest predictive accuracy. In Panel A of Table IV, considering the RF models with sentiment and controls, the Sentiment model (Model I) reaches the highest AUC (0.931), followed closely by Direction model (Model III) with an AUC of 0.929. In contrast, models based solely on optimism/pessimism metrics (Model II) or based on price Momentum (Model IV) yield slightly lower performance metrics. This suggests that specific expectations regarding future price movements are more informative of impending speculative episodes than generalized market mood (Greenwood and Shleifer, 2014).

In addition, the inclusion of control variables (volatility and volume) significantly enhances model performance across all specifications (see Table IV). The benchmark model based on control variables (Bench. I) provides a strong baseline, with the RF specification achieving an AUC of roughly 0.80 across panels. However, the addition of attention and sentiment variables improves the AUC to levels exceeding 0.90 in the nonlinear framework. This demonstrates that while market microstructure variables are essential (Scheinkman and Xiong, 2003), investor sentiment provides distinct, non-redundant information that is crucial for accurately identifying periods of exuberance (Baker and Wurgler, 2006).

Finally, these results not only corroborate the emerging consensus on the superiority of machine learning in asset pricing (Gu et al., 2020), but significantly advance the literature by isolating the specific nature of the sentiment-return relationship in cryptocurrency markets. While prior studies identify retail attention as a driver of price pressure (Da et al., 2011), our analysis reveals that this predictive advantage is fundamentally dependent on modeling nonlinearity. Moreover, by establishing that explicit directional beliefs outperform general optimism, we refine the operational definition of sentiment for bubble detection. This contributes a methodological precision to the field, suggesting that the mechanism of exuberance is driven less by broad emotional states and more by the nonlinear coordination of specific price beliefs among retail investors.

3.5 Sentiment Dynamics in the Genesis and Termination of Speculative Bubbles

To further understand the lifecycle of cryptocurrency bubbles, we decouple our analysis to examine the specific sentiment dynamics underlying the genesis and termination of these speculative episodes (Shiller, 2000). The theoretical premise dictates that bubbles emerge when market participants aggressively seek capital gains, driven by the strong expectation of future price appreciation (Barberis et al., 2018, Greenwood and Shleifer, 2014). Consequently, in this article, we provide evidence that variables capturing positive sentiment, optimism, momentum, and upward price expectations—extracted from both Social and News text data—help explain this phenomenon in Bitcoin prices (Phillips and Shi, 2018). Specifically, the growth in the volume of positive textual narratives serves as a robust indicator for identifying the onset of a bubble (Tetlock, 2007).

Table V reports the Logit model estimates strictly for the determinants of the beginning of Bitcoin bubbles, focusing on positive behavioral metrics. The empirical results strongly corroborate our theoretical intuition. Variables capturing positive market dynamics—specifically *Positive* sentiment, *Optimism*, *PriceUp*, and *Momentum*—exhibit positive and statistically significant coefficients in nearly all specifications.[10] Economically, the significance of *Momentum* and *PriceUp* highlights the presence of extrapolative price expectations, where past positive returns are projected into the future, fueling the positive feedback loops characteristic of financial bubble dynamics (Barberis et al., 2018, Shiller, 2000). Furthermore, the strong loading on the *Positive* and *Optimism* parameter suggests that subjective sentiment and investor enthusiasm are closely correlated with the emergence of Bitcoin overvaluation (Baker and Wurgler, 2006, Cheah and Fry, 2015).

[10] There is modest but interesting heterogeneity in the statistical significance of the coefficients across panels. For example, the "Positive" indicator is more relevant for News Media and News Headlines than for Social Media alone. Conversely, the "Momentum" indicator is more relevant for Social Media than for News Media or News Headlines.

Table V. **Logit Model Parameter Estimates for the Determinants of Bitcoin *Bubble Onset*.**

	Model I (Positive)		**Model II** (Optimism)		**Model III** (PriceUp)		**Model IV** (Momentum)		**Bench** (Ctrl)
Variables	**Raw**	**Ctrl**	**Raw**	**Ctrl**	**Raw**	**Ctrl**	**Raw**	**Ctrl**	**Ctrl**
Panel A: Social media.									
Constant	-4.5***	-4.8***	-4.4***	-4.6***	-4.6***	-4.9***	-4.5***	-4.7***	-4.4***
Positive	7.2	9.1							
Optimism			18.0	16.7					
PriceUp					49.5**	71.6**			
Momentum							942.7***	1100.3***	
Volatility		7.4		5.0		8.8		8.7	3.4
Volume		-3.1E-12		1.7E-12		-8.0E-12		-4.4E-12	4.1E-12
AIC	490.5	494.0	492.0	495.8	488.7	491.5	485.6	488.8	494.2
Adj. R^2	-0.004	-0.011	-0.007	-0.014	0.000	-0.006	0.006	-0.000	-0.011
MAE	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029
RMSE	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
Panel B: News media.									
Constant	-4.7***	-4.8***	-4.7***	-4.8***	-4.6***	-4.6***	-4.4***	-4.5***	-4.4***
Positive	10.9***	14.7***							
Optimism			40.3***	48.3***					
PriceUp					32.6**	43.1**			
Momentum							300.3**	328.6*	
Volatility		3.9		5.0		1.6		3.6	3.3
Volume		-8.4E-12		-5.6E-12		-7.5E-12		-2.2E-12	4.3E-12
AIC	487.4	490.3	487.4	490.7	488.0	491.2	489.2	493.0	494.4
Adj. R^2	0.003	-0.003	0.003	-0.003	0.002	-0.005	-0.000	-0.008	-0.011
MAE	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029
RMSE	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
Panel C: News Headline.									
Constant	-4.5***	-4.6***	-4.4***	-4.7***	-4.4***	-4.4***	-4.2***	-4.4***	-4.4***
Positive	11.2**	15.6**							
Optimism			37.7*	42.7*					
PriceUp					33.0***	42.6***			
Momentum							158.9	160.0	
Volatility		6.6		9.2		4.9		6.9	7.9
Volume		-1.0E-11		-4.0E-12		-1.1E-11		-1.6E-12	4.5E-13
AIC	419.2	421.5	420.6	423.8	415.7	417.5	421.3	424.9	424.6
Adj. R^2	-0.000	-0.006	-0.004	-0.011	0.008	0.004	-0.006	-0.014	-0.013
MAE	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032
RMSE	0.126	0.126	0.126	0.127	0.126	0.126	0.126	0.126	0.127
Panel D: Social & News corpus.									
Constant	-4.7***	-5.0***	-4.6***	-4.9***	-4.7***	-4.9***	-4.5***	-4.6***	-4.4***
Positive	9.0**	12.9**							
Optimism			29.9	34.3					
PriceUp					49.5**	79.7***			
Momentum							706.3***	861.2***	
Volatility		8.4		6.5		7.9		6.8	3.3
Volume		-7.0E-12		-1.5E-12		-1.2E-11		-5.7E-12	4.3E-12
AIC	489.5	492.6	491.0	494.7	487.7	489.7	486.4	489.6	494.4
Adj. R^2	-0.001	-0.007	-0.004	-0.012	0.003	-0.001	0.005	-0.001	-0.011
MAE	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029
RMSE	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121

Notes: This table reports parameter estimates from Logit models examining the determinants of the beginning of Bitcoin bubble periods under four alternative specifications (Models I, II, III, and IV), each capturing a distinct positive dimension of investor behavior. Model I focuses on Positive sentiment, Model II on Optimism, Model II on upward price expectations (PriceUp), and Model IV on price Momentum. For each specification, results are presented both without control variables ("Raw") and with market controls included ("Ctrl"). The benchmark specification corresponds to a Logit model estimated solely with control variables, serving as a reference. Standard errors are omitted for brevity. (***), (**), (*) indicate significance at the 1%, 5%, and 10% levels, respectively. AIC corresponds to the Akaike Information Criterion, and Adj. R^2 refers to McFadden's adjusted R-squared.

Evaluating the logit model fitting performance across these specifications provides further insights into the optimal information set for bubble detection. The models utilizing the aggregated News corpora, as well as the isolated News Headline models, exhibit superior explanatory metrics. These specifications yield the highest McFadden's Pseudo-R^2 and lowest Akaike Information Criterion (AIC) values. With few exceptions – namely, the Momentum specification with controls in the Headline panel and the Optimism specification with controls in Panels A and D – the models outperform the benchmark. This reinforces the significant explanatory gain provided by sentiment variables in explaining the onset of a bubble. Ultimately, this superior goodness-of-fit indicates that while professional news media provides the most informative channel, the consistent significance of sentiment and price expectation variables across all specifications demonstrates their robust capacity to capture and explain the genesis of cryptocurrency bubbles.

Extending this logic to the termination—or bursting—phase of the bubble, one might intuitively expect a sudden surge in negative sentiment or pessimistic narratives to act as the primary catalyst for a crash. However, our empirical framework yields a different result. As presented in Table VI, models incorporating *Negative* sentiment, *Pessimism*, and *PriceDown* expectations completely fail to achieve statistical significance across all data panels when attempting to predict the end of the exuberance period.

Table VI. **Logit Model Parameter Estimates for the Determinants of Bitcoin *Bubble Termination.***

	Model I (Negative)		Model II (Pessimism)		Model III (PriceDown)		Bench (Ctrl)
Variables	**Raw**	**Ctrl**	**Raw**	**Ctrl**	**Raw**	**Ctrl**	**Ctrl**
Panel A: Social media.							
Constant	-4.4***	-4.8***	-4.4***	-4.8***	-4.3***	-4.7***	-4.6***
Negative	5.0	6.1					
Pessimism			17.2	19.4			
PriceDown					22.0	18.4	
Volatility		10.9		10.9		10.6	10.9
Volume		-2.6E-12		-1.2E-12		4.4E-13	2.2E-12
AIC	491.9	494.9	492.2	495.3	492.4	495.6	493.7
Adj. R^2	-0.006	-0.013	-0.007	-0.013	-0.008	-0.014	-0.010
MAE	0.029	0.029	0.029	0.029	0.029	0.029	0.029
RMSE	0.121	0.121	0.121	0.121	0.121	0.121	0.121
Panel B: News media.							
Constant	-4.5***	-4.7***	-4.5***	-4.8***	-4.3***	-4.7***	-4.6***
Negative	6.8	6.3					
Pessimism			39.9	38.1			
PriceDown					14.2	8.2	
Volatility		8.0		8.0		10.4	10.9
Volume		-1.2E-12		-1.4E-12		1.5E-12	2.4E-12
AIC	491.5	494.9	491.1	494.6	492.7	495.9	493.9
Adj. R^2	-0.005	-0.012	-0.004	-0.011	-0.008	-0.014	-0.010
MAE	0.029	0.029	0.029	0.029	0.029	0.029	0.029
RMSE	0.121	0.121	0.121	0.121	0.121	0.121	0.121
Panel C: News Headline.							
Constant	-4.4***	-4.9***	-4.4***	-4.9***	-4.3***	-4.9***	-4.9***
Negative	3.7	2.9					
Pessimism			16.3	11.9			
PriceDown					5.5	0.9	
Volatility		15.9		17.0		17.5	17.6
Volume		-1.7E-12		-6.5E-13		2.6E-14	1.4E-13
AIC	372.1	374.4	372.5	374.6	372.8	374.8	372.8
Adj. R^2	-0.009	-0.015	-0.010	-0.016	-0.011	-0.016	-0.011
MAE	0.027	0.027	0.027	0.027	0.027	0.027	0.027
RMSE	0.117	0.117	0.117	0.117	0.117	0.117	0.117
Panel D: Social & News corpus.							
Constant	-4.5***	-4.8***	-4.5***	-4.8***	-4.3***	-4.7***	-4.6***
Negative	5.8	7.1					
Pessimism			23.2	27.3			
PriceDown					21.0	16.5	
Volatility		10.1		10.4		10.5	10.9
Volume		-3.3E-12		-2.3E-12		5.9E-13	2.4E-12
AIC	491.8	494.9	492.1	495.2	492.6	495.8	493.9
Adj. R^2	-0.006	-0.012	-0.006	-0.013	-0.007	-0.014	-0.010
MAE	0.029	0.029	0.029	0.029	0.029	0.029	0.029
RMSE	0.121	0.121	0.121	0.121	0.121	0.121	0.121

Notes: This table reports parameter estimates from Logit models examining the determinants of the termination of Bitcoin bubble periods under three alternative specifications (Models I, II, and III), each capturing a distinct negative dimension of investor behavior. Model I focuses on Negative sentiment, Model II on Pessimism, and Model III on downward price expectations (PriceDown). For each specification, results are presented both without control variables ("Raw") and with market controls included ("Ctrl"). The benchmark specification corresponds to a Logit model estimated solely with control variables, serving as a reference. Standard errors are omitted for brevity. (***), (**), (*) indicate significance at the 1%, 5%, and 10% levels, respectively. AIC corresponds to the Akaike Information Criterion, and Adj. R^2 refers to McFadden's adjusted R-squared.

This lack of significance demonstrates a clear asymmetry in the behavioral drivers of market exuberance and reveals a critical underlying mechanism: the termination of a bubble represents the exhaustion of a prolonged period of euphoria and aggressive price escalation. Consequently, the end of this phenomenon is primarily characterized by the weakening and collapse of the specific positive trends that previously sustained it. In other words, a bubble bursts not because negative textual narratives suddenly overpower the market, but because the extrapolative expectations and optimism—the structural fuel of the speculative episode—gradually dissipate as the bubble matures (Barberis et al., 2018, Hong and Stein, 1999).

4 Conclusion

The valuation of digital assets remains a central puzzle in financial economics. Absent the fundamental anchors that stabilize traditional asset classes, cryptocurrencies are particularly exposed to investor-driven dynamics and speculative feedback loops (Müser et al., 2024, Shen and Wu, 2025). Although a growing literature documents the presence of speculative bubbles in cryptocurrency markets (Cheung et al., 2015, Enoksen et al., 2020), comparatively fewer studies examine how measurable belief structures contribute to the emergence and persistence of these episodes. Existing empirical approaches have typically relied on indirect proxies of investor attention—such as trading volume or search intensity—rather than on structured textual measures that disentangle affective sentiment from directional price expectations (Guo et al., 2025, Pan, 2020). This measurement limitation constrains both the econometric modeling and real-time prediction of bubble risk in belief-driven markets.

To address this gap, this study investigates both the occurrence of speculative bubbles in the Bitcoin market and the role of investor sentiment in their formation. We adopt a two-stage empirical framework in which bubble periods are first identified using the GSADF test and subsequently linked to investor behavior metrics derived from a rich textual dataset. Specifically, we exploit sentiment and expectation indicators from the Thomson Reuters MarketPsych database, constructed using large-scale natural language processing and state-of-the-art large language models. To the best of our knowledge, this is the first study to employ MarketPsych-based behavioral measures to analyze speculative bubbles in cryptocurrency markets. Importantly, the framework allows us to distinguish among multiple dimensions of investor behavior, captured through four complementary classes: sentiment, optimism, price direction (expectations), and momentum. Moreover, by estimating both linear (Logit) and nonlinear (Random Forest) models, the analysis explicitly assesses whether the transmission of investor behavior to bubble formation operates through linear or nonlinear mechanisms.

The empirical results provide strong evidence of recurrent speculative bubbles in the Bitcoin market, consistent with previous findings in the literature. Beyond bubble identification, we show that investor sentiment and price expectations play a statistically and economically significant role in explaining the likelihood of bubble episodes. Sentiment-related variables are highly significant across most specifications, display economically intuitive signs, and add explanatory power beyond standard market controls such as volatility and trading volume. A comparison across information sources further reveals that narratives extracted from both news media and social media are informative. This finding suggests that professionally curated information and peer-to-peer communication convey complementary signals in the formation of speculative dynamics.

When allowing for nonlinearities, the results indicate that Random Forest models substantially outperforming linear benchmarks across all behavioral dimensions. These findings highlight the importance of nonlinear transmission mechanisms in cryptocurrency markets, where retail-driven narratives and coordinated belief formation may amplify speculative behavior in a threshold-dependent manner. Notably, models capturing explicit price expectations and trend-following beliefs outperform those based solely on generalized emotional sentiment, indicating that bubbles are driven more by coordinated beliefs about future price movements than by unstructured optimism alone. In addition, the nonlinear models benefit from aggregating social and news-based information, as reflected in higher accuracy values.

Overall, this study contributes to the literature on speculative bubbles and investor sentiment in cryptocurrency markets by offering a more granular characterization of investor behavior and its role in driving price exuberance. From a practical perspective, the results suggest that monitoring sentiment and expectation-based narratives—particularly when extracted from both news and social media—can provide valuable signals regarding the emergence of speculative bubbles. Such information may complement traditional market indicators and support investment decision-making and risk management in cryptocurrency markets.

Future research shall extend this framework to other cryptocurrencies and digital asset classes, including decentralized finance tokens and non-fungible assets. Additionally, further work could incorporate behavioral metrics related to investors' perceptions of technological features, security risks, and protocol integrity. Given the central role of technological trust, cybersecurity, and fraud-related concerns in digital asset valuation, integrating sentiment measures linked to technological and governance dimensions represents a promising avenue for advancing the understanding of price formation in these markets.

References

Acharya, D. (2024). Comparative analysis of stock bubble in S&P 500 individual stocks: A study using sadf and gsadf models. *Journal of Risk and Financial Management*, 17(2).

Ahn, Y. and Kim, D. (2023). Visceral emotions and Bitcoin trading. *Finance Research Letters*, 51:103458.

Aimable, W. (2026). Disentangling market and uncertainty effects in crypto valuation: A portfolio-based analysis. *Pacific-Basin Finance Journal*, 96:103032.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4):1645–1680.

Barberis, N., Shiller, R. J., and Vishny, A. (2018). *A Crisis of Beliefs: Investor Psychology and Financial Fragility*. Princeton University Press, Princeton.

Basher, S. A. and Sadorsky, P. (2022). Forecasting bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility? *Machine Learning with Applications*, 9:100355.

Biswas, P. and Sharma, C. (2025). Does fear sentiment drive cryptocurrency volatility? Evidence from Google trends data. *Finance Research Letters*, 86:108510.

Blanchard, O. J. and Watson, M. W. (1982). Bubbles, rational expectations and financial markets. NBER Working Papers 0945, National Bureau of Economic Research, Inc.

Boehl, G. and Hommes, C. (2025). Rational vs. irrational beliefs in a complex world. *Journal of Economic Behavior & Organization*, 232:106898.

Bouri, E., Shahzad, S. J. H., and Roubaud, D. (2019). Co-explosivity in the cryptocurrency market. *Finance Research Letters*, 29:178–183.

Chaim, P. and Laurini, M. P. (2019). Is Bitcoin a bubble? *Physica A: Statistical Mechanics and its Applications*, 517:222–232.

Cheah, E.-T. and Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, 130:32–36.

Chen, H., De, P., Hu, Y. J., and Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, 27(5):1367–1403.

Cheung, A., Roca, E., and Su, J.-J. (2015). Crypto-currency bubbles: An application of the Phillips-Shi-Yu methodology to Mt. Gox Bitcoin prices. *Applied Economics*, 47(23):2348–2358.

Chowdhury, M. S. R. and Damianov, D. S. (2024). Uncertainty and bubbles in cryptocurrencies: Evidence from newly developed uncertainty indices. *International Review of Financial Analysis*, 91:102949.

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *The Journal of Finance*, 66(5):1461–1499.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, 98(4):703–738.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4):1057–1072.

Eierle, B., Klamer, S., and Muck, M. (2022). Does it really pay off for investors to consider information from social media? *International Review of Financial Analysis*, 81:102074.

Enoksen, F., Landsnes, C., Lučivjanská, K., and Molnár, P. (2020). Understanding risk of bubbles in cryptocurrencies. *Journal of Economic Behavior & Organization*, 176:129–144.

Filip, A. M. and Pochea, M. M. (2023). Intentional and spurious herding behavior: A sentiment driven analysis. *Journal of Behavioral and Experimental Finance*, 38:100810.

Gan, B., Alexeev, V., Bird, R., and Yeung, D. (2020). Sensitivity to sentiment: News vs social media. *International Review of Financial Analysis*, 67:101390.

Garcia, D., Tessone, C. J., Mavrodiev, P., and Perony, N. (2014). The digital traces of bubbles: Feedback cycles between Bitcoin price and social media. *Journal of the Royal Society Interface*, 11(99):20140623.

García-Monleón, F., del Valle, I. D., and Lara, F. J. (2021). Intrinsic value in crypto currencies. *Technological Forecasting and Social Change*, 162:120393.

Gathergood, J., Hirshleifer, D., Leake, D., Sakaguchi, H., and Stewart, N. (2023). Naïve buying diversification and narrow framing by individual investors. *The Journal of Finance*, 78(3):1705–1741.

Gennaioli, N. and Shleifer, A. (2018). *A Crisis of Beliefs: Investor Psychology and Financial Fragility*. Princeton University Press.

Geuder, J., Kinateder, H., and Wagner, N. (2019). Cryptocurrencies as financial bubbles: The case of Bitcoin. *Finance Research Letters*, 31:179–184.

Greenwood, R. and Shleifer, A. (2014). Expectations of returns and expected returns. *Review of Financial Studies*, 27(3):714–746.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Guo, M., Wang, S., and Wei, Y. (2025). Bubbles in NFT markets: Correlated with cryptocurrencies or sentiment indexes? *Applied Economics Letters*, 32(4):491–497.

Han, J. and Wang, Y. (2025). All that glitters: A theory of multiple bubbles with implications for cryptocurrencies. *Journal of Monetary Economics*, 152:103764.

Han, S. (2025). Investor sentiment and cross-section of cryptocurrency returns. *Journal of Behavioral and Experimental Finance*, 46:101043.

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

Hu, Y. and Oxley, L. (2017). Are there bubbles in exchange rates? Some new evidence from G10 and emerging market economies. *Economic Modelling*, 64:419–442.

Kalyvas, A., Papakyriakou, P., Sakkas, A., and Urquhart, A. (2020). What drives Bitcoin's price crash risk? *Economics Letters*, 191:108777.

Krichevskiy, D. and Qirjo, D. (2017). Chapter 7 - Rational Agents and Irrational Bubbles. In Economou, F., Gavriilidis, K., Gregoriou, G. N., and Kallinterakis, V., editors, *Handbook of Investors' Behavior During Financial Crises*, pages 113–127. Academic Press.

Kyriazis, N., Papadamou, S., and Corbet, S. (2020). A systematic review of the bubble dynamics of cryptocurrency prices. *Research in International Business and Finance*, 54:101254.

Li, B., Chen, Y., Wu, H., and Mao, X. (2025). Macroeconomics, geopolitical risk, and resource commodity price bubbles. *Resources Policy*, 101:105478.

Li, S. and Ma, J. (2024). The impact of sentiment and engagement of twitter posts on cryptocurrency price movement. *Finance Research Letters*, 65:105598.

Lin, X., Meng, Y., and Zhu, H. (2023). How connected is the crypto market risk to investor sentiment? *Finance Research Letters*, 56:104177.

Liu, Y. and Tsyvinski, A. (2021). Risks and returns of cryptocurrency. *The Review of Financial Studies*, 34(6):2689–2727.

Manahov, V. and Urquhart, A. (2021). The efficiency of Bitcoin: A strongly typed genetic programming approach to smart electronic Bitcoin markets. *International Review of Financial Analysis*, 73:101629.

Mei, J., Scheinkman, J., and Xiong, W. (2005). Speculative trading and stock prices: Evidence from chinese a-b share premia. Working Paper 11362, National Bureau of Economic Research.

Mesly, O. (2023). Irrational exuberance and deception — Why markets spin out of control. *Journal of Behavioral and Experimental Finance*, 37:100707.

Mira, E. C., Maldonado, W. L., and Tourinho, O. A. F. (2025). Testing for bubbles in the Brazilian commercial real estate market. *Economics Bulletin*, 45(3):1308–1325.

Müser, S., Hemmerich, M., and Schmitz, F. (2024). Individual differences in Bitcoin investment: The role of personality, attitudes, and knowledge. *Journal of Behavioral and Experimental Economics*, 113:102289.

Neto, D. (2021). Are Google searches making the Bitcoin market run amok? A tail event analysis. *The North American Journal of Economics and Finance*, 57:101454.

Osman, M. B., Galariotis, E., Guesmi, K., Hamdi, H., and Naoui, K. (2024). Are markets sentiment driving the price bubbles in the virtual? *International Review of Economics and Finance*, 89:272–285.

Ozgur, O., Yilanci, V., and Ozbugday, F. C. (2021). Detecting speculative bubbles in metal prices: Evidence from GSADF test and machine learning approaches. *Resources Policy*, 74:102306.

Pan, W.-F. (2020). Does investor sentiment drive stock market bubbles? Beware of excessive optimism! *Journal of Behavioral Finance*, 21(1):27–41.

Pavlidis, E., Yusupova, A., Paya, I., Peel, D., Martínez-García, E., Mack, A., and Grossman, V. (2016). Episodes of exuberance in housing markets: In search of the smoking gun. *The Journal of Real Estate Finance and Economics*, 53:419–449.

Phillips, P. C. and Shi, S. (2018). Real time monitoring of asset markets: Bubbles and crises. Cowles Foundation Discussion Papers 2152, Cowles Foundation for Research in Economics, Yale University.

Phillips, P. C. B., Shi, S.-P., and Yu, J. (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, 56(4):1043–1078.

Phillips, P. C. B., Shi, S.-P., and Yu, J. (2015b). Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review*, 56(4):1079–1134.

Phillips, P. C. B., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: Why did exuberance escalte asset values? *International Economic Review*, 52(1):201–226.

Scheinkman, J. A. and Xiong, W. (2003). Overconfidence and speculative bubbles. *Journal of Political Economy*, 111(6):1183–1219.

Shen, D. and Wu, Y. (2025). The role of Guru investor in Bitcoin: Evidence from Kolmogorov-Arnold networks. *Research in International Business and Finance*, 75:102789.

Shiller, R. J. (2000). *Irrational Exuberance*. Princeton University Press.

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, 107(4):967–1004.

Steiger, S. and Pelster, M. (2020). Social interactions and asset pricing bubbles. *Journal of Economic Behavior & Organization*, 179:503–522.

Sun, L., Najand, M., and Shen, J. (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking & Finance*, 73:147–164.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Urquhart, A. (2018). What causes the attention of Bitcoin? *Economics Letters*, 166:40–44.

van Eyden, R., Gupta, R., Nielsen, J., and Bouri, E. (2023). Investor sentiment and multi-scale positive and negative stock market bubbles in a panel of G7 countries. *Journal of Behavioral and Experimental Finance*, 38:100804.

Vasilopoulos, K., Pavlidis, E., and Martínez-García, E. (2022). exuber: Recursive right-tailed unit root testing with R. *Journal of Statistical Software*, 103(10):1–26.

Wu, N. and Li, M. (2026). Digital finance empowerment, social media sentiment, and asset price bubbles. *International Review of Financial Analysis*, 109:104806.

Wu, Q., Guo, G., Li, X., Singh, R., and Zhang, T. (2025). Bitcoin's fundamental value and speculative behavior: A new framework for price dynamics. *The North American Journal of Economics and Finance*, 80:102509.

Zhang, X., Zhu, M., Tian, Y., and Zedda, S. (2024). Detecting house price bubbles in G7 countries: New evidence and heterogeneous determinants. *Finance Research Letters*, 69:106107.

A Unit Root Tests

Table A.1. **ADF Test Statistics for Investor Bahavior and Control Variables**.

	ADF (none)		ADF (constant)		ADF (constant and trend)	
Variable	Stat	Lags	Stat	Lags	Stat	Lags
Panel A: Social media.						
Positive	−4.209***	1	−11.778***	1	−12.266***	1
Negative	−6.053***	1	−15.678***	1	−16.056***	1
Optimism	−3.902***	1	−11.706***	1	−11.814***	1
Pessimism	−5.308***	1	−15.069***	1	−15.379***	1
PriceUp	−5.466***	1	−13.011***	1	−14.083***	1
PriceDown	−8.143***	1	−17.691***	1	−18.122***	1
Momentum	−11.388***	1	−14.799***	1	−15.041***	1
Panel B: News media.						
Positive	−10.708***	1	−27.204***	1	−29.216***	1
Negative	−11.449***	1	−29.268***	1	−29.316***	1
Optimism	−10.014***	1	−24.416***	1	−26.556***	1
Pessimism	−10.376***	1	−28.825***	1	−28.935***	1
PriceUp	−12.741***	1	−25.722***	1	−27.621***	1
PriceDown	−13.377***	1	−25.466***	1	−26.245***	1
Momentum	−15.981***	1	−19.718***	1	−20.353***	1
Panel C: News & Social corpus.						
Positive	−5.183***	1	−15.505***	1	−16.452***	1
Negative	−6.691***	1	−18.392***	1	−18.759***	1
Optimism	−4.675***	1	−15.154***	1	−15.585***	1
Pessimism	−5.676***	1	−17.330***	1	−17.678***	1
PriceUp	−7.356***	1	−17.993***	1	−19.813***	1
PriceDown	−9.068***	1	−19.814***	1	−20.400***	1
Momentum	−12.492***	1	−16.264***	1	−16.681***	1
Panel D: Control variables.						
Volume	−7.614***	1	−13.776***	1	−15.032***	1
Volatility	−3.077***	1	−9.498***	1	−10.175***	1

Notes: This table reports the Augmented Dickey–Fuller (ADF) test statistics for the investor behavior measures and control variables (trading volume and volatility) employed in the bubble identification models. The ADF tests are conducted under specifications with no constant and trend (none), with constant and no trend (constant) and including both constant and trend (constant and trend). Lag lengths are selected according to the Akaike Information Criterion (AIC).
(***), (**), (*) denote statistical significance at the 1%, 5%, and 10% levels, respectively.