

JUNE 2026

POLICY BRIEF

FROM NLP TO HYPE AND FINANCIAL BUBBLES: INTEGRATING NEWS ATTENTION WITH BUBBLE DETECTION MODELS

HÉLYETTE GEMAN

*In 2017, eight scientists from the Google research team published in the journal *Advances in Neural Info Processing Systems* the remarkable article “Attention is all you need,” which introduced a Transformer neural network architecture. The paper has been cited over 173,000 times and ranks among the top 100 most cited papers of the 21st century. It builds on the attention principle introduced in 2014 by Bahdanau, Cho and Turing Award winner Bengio, who proposed neural machine translation by jointly learning to align and translate. This transformer approach has become the main architecture for a wide variety of AI tasks, including large language models. In machine learning, “attention” refers to a mechanism that allows models to focus on specific parts of the input data during the learning process and to determine the relative importance of each component within a sequence.*

Turning to financial economics, financial news—whether in terms of volume, unusual frequency, or sentiment (positive versus negative tone)—has long attracted the attention of researchers seeking to forecast market dynamics—“buy on rumors, sell on news.” Financial bubbles, however, remain among the most challenging phenomena to model and trade. Traditional models relying solely on price dynamics often fail to detect bubbles in real time, a key objective for stock picking and portfolio selection. Advances in natural language processing (NLP) now enable researchers to quantify market attention and sentiment using financial news and social media activity.

This paper builds on recent research on sentiment in financial markets and integrates these insights into quantitative bubble detection models derived from the Log-Periodic Power Law (LPPL) literature, while incorporating a Hype Index that measures disproportionate news attention at a given moment, in order to obtain a hype-adjusted view of speculative dynamics.

Within this framework, sentiment and news intensity modify bubble scores derived from price dynamics. The resulting Hyped Log-Periodic Power Law (HLPPL) model improves the identification of emerging bubbles and enables the detection of negative bubbles, corresponding to temporarily overvalued assets. The approach further highlights the importance of the choice of numéraire with respect to which prices are expressed (e.g., gold versus the dollar), emphasizing that bubbles must be assessed relative to a chosen reference asset.

Empirical illustrations across equities and cryptocurrencies show how media attention and narrative amplification interact with price dynamics during speculative episodes. Taken together, these results suggest that incorporating information flows and market narratives can significantly improve the early detection and interpretation of financial bubbles.

HÉLYETTE GEMAN

I. INTRODUCTION

Financial bubbles have long attracted the attention of economists, policymakers, and market participants because of their profound impact on financial stability and economic cycles. Historical episodes include the Dutch Tulip Mania of the seventeenth century and the South Sea Bubble in the UK—known as the world’s first financial crash—in which Isaac Newton himself is said to have lost the equivalent of 40 million pounds in today’s currency when South Sea Company stock fell by about 80% from its peak of around £1,000 in August 1720. More recent examples, including the dot-com boom of the late 1990s, the housing market bubble preceding the 2008 global financial crisis, and the rapid rise of technology and artificial intelligence-related equities, highlight both the recurring nature of speculative episodes and the persistent difficulty of recognizing bubbles before they burst.

A central challenge in the study of bubbles is that speculative dynamics often appear indistinguishable from strong fundamental growth while they are unfolding. Traditional approaches to bubble detection, therefore, rely heavily on price dynamics, attempting to identify statistical signatures such as super-exponential growth or accelerating volatility patterns. Among these approaches, the LPPL framework, introduced by Sornette et al. (1996), aims to predict the “critical date” at which the bubble will burst. This framework has been particularly influential in modeling the characteristic price patterns observed during bubble regimes.

However, price trajectories alone may not fully capture the mechanisms that drive speculative dynamics. Financial markets are strongly influenced by information flows, narratives, and investor attention. News coverage, social media discussions, and market commentary can amplify optimistic expectations and reinforce positive feedback loops among investors. As a result, speculative episodes are often accompanied by rapid increases in media attention and narrative intensity.

Recent advances in NLP make it possible to quantify these information flows systematically. By analyzing large corpora of financial news and textual data, researchers can construct measures of market sentiment, narrative intensity, and media attention. These signals provide a new source of information that complements traditional price-based indicators.

The article builds on three strands of recent research that combine natural language processing techniques and investor sentiment with quantitative finance, namely a Hype Index and a quantitative bubble detection model derived from the LPPL framework. Combining these approaches provides a framework for integrating price dynamics with systematic measures of market attention and sentiment extracted from textual data.

The central idea is that news attention can amplify the self-reinforcing feedback loops that characterize financial bubbles; in another asset class, bank runs arise from a similar mechanism, as described, for instance, in the model proposed by the Nobel Prize winners Diamond and Dybvig. When rising prices attract media coverage and investor interest, this attention can in turn reinforce further price increases, creating a narrative-driven amplification mechanism that accelerates speculative dynamics. Hence, incorporating measures of sentiment and hype into bubble diagnostics provides a more comprehensive view of market behavior than price dynamics alone.

In this sense, the analysis connects the literature on speculative bubbles and critical phenomena in financial markets with the recent advances in machine learning and natural language processing applied to financial text data.

II. MARKET SENTIMENT AND VOLATILITY PREDICTION

In the paper titled “A Sentiment Analysis Approach for the Prediction of Market Volatility,” Deveikyte et al. (2022) explore the predictive power of sentiment analysis on financial market behavior, focusing in particular on the FTSE 100. More precisely, the authors investigate whether sentiment extracted from financial news headlines and Twitter posts can forecast next-day market returns and volatility. The authors employ NLP techniques, including sentiment scoring and Latent Dirichlet Allocation (LDA) for topic modeling, when processing textual data. In NLP, LDA is a generative statistical model in which documents are represented as random mixtures of a small number of latent topics, and each topic is characterized by a probability distribution over words. In this case, these features are then used as inputs into a logistic regression classifier to predict the direction of market volatility.

Cao et al. (2025) further underscores the importance of information flows and market narratives in shaping asset price dynamics. They introduce a Hype-Adjusted Probability Measure, which incorporates sentiment extracted from financial news into traditional probabilistic frameworks used in financial modeling, where “hype” captures the effect of excessive news circulating in both general and financial media.

The central idea is that investor beliefs are influenced not only by price movements and fundamental information, but also by the tone and intensity of media coverage surrounding an asset. Using NLP techniques, financial news can be analyzed to quantify positive and negative narratives associated with a company or sector. Incorporating these signals into probabilistic frameworks allows models to capture how narratives and sentiment influence expectations about future price dynamics.

A classical definition of the daily sentiment score was proposed by Gabrovsek et al. (2016) as:

$$Sent_d = \frac{N_d(pos) - N_d(neg)}{N_d(pos) + N_d(Neut) + N_d(neg) + 3}$$

Where N_d represents the number of positive, negative, or neutral news headlines at time d .

While sentiment measures capture the direction or tone of news coverage, they do not fully account for the intensity of media attention. An asset may receive overwhelmingly positive sentiment but relatively little coverage, or it may attract extremely high levels of attention regardless of sentiment. To address this distinction, Cao et al. (2025) introduce the Hype Index, a measure designed to quantify disproportionate media attention.

The Hype Index compares the frequency with which an asset is mentioned in financial news with a baseline reflecting its economic scale, typically measured through market capitalization. When an asset receives significantly more attention than would be expected based on its size, it can be described as “hyped.” Conversely, assets receiving relatively little attention may be considered under-hyped.

This concept provides a practical way to measure the imbalance between media attention and the firm’s fundamentals. Empirical evidence suggests that periods of extreme hype often coincide with rapid price appreciation, elevated volatility, and heightened speculative activity—features commonly associated with the formation of financial bubbles.

Consequently, for bubble detection, the Hype Index provides a complementary signal to price-based diagnostics, allowing practitioners to incorporate information about market attention and narrative amplification alongside traditional price dynamics.

III. FINANCIAL BUBBLES DETECTION

A large body of financial economics literature has investigated bubbles, focusing on their definition, identification, and attempts to estimate their expected burst dates. One strand of this literature analyzes financial bubbles using only price trajectories. Among the most influential approaches is the LPPL framework developed by Didier Sornette (1995) and collaborators. LPPL models capture the accelerating growth and oscillatory behavior often observed during speculative bubble phases, where prices exhibit super-exponential dynamics (as opposed to the geometric Brownian motion underlying standard option pricing models), accompanied by increasingly frequent fluctuations.

The LPPL framework models the price trajectory as approaching a critical time t_c representing the theoretical end of the bubble regime. As the system approaches this critical point, price dynamics accelerate while oscillations become more frequent. The log-price in the LPPL model is expressed as follows:

$$p(t) = A + B(t_c - t)^m + C(t_c - t)^m \cos(\omega \ln(t_c - t) + \phi)$$

where A is the baseline price level, B captures the super-exponential growth rate, C determines the amplitude of log-periodic oscillations, $m \in (0, 1)$ is the critical exponent, ω is the log-periodic frequency, ϕ is the phase parameter, and t_c denotes the critical time corresponding to the theoretical termination of the bubble regime.

In the “volatility-confined” LPPL formulation (Lin, Ren, and Sornette, 2014), the price trajectory follows this super-exponential structure while the residual component remains mean-reverting. This property improves the stability of calibration and allows researchers to estimate bubble dynamics using rolling windows without incorporating future information, which would be undesirable.

Despite their success, price-based models face an important limitation: they rely solely on observed price dynamics and therefore do not incorporate information about media attention and investor sentiment, which often drive speculative behavior. In modern financial markets, information flows and narrative amplification can play a central role in reinforcing positive feedback loops among investors.

As a result, price trajectories alone may be insufficient to detect emerging bubbles in their early stages. Incorporating measures of news attention and sentiment can therefore provide valuable complementary signals, helping to distinguish between price dynamics driven by fundamentals and those amplified by market narratives.

IV. HLPPL MODEL

To integrate price dynamics with information flows derived from textual data, we introduce the HLPPL framework. The objective is to combine the traditional LPPL bubble representation—based on a log power law for the log stock price, growing faster than the geometric Brownian motion underlying standard option pricing models—with signals capturing media attention (hype) and sentiment extracted from financial news using NLP.

The starting point of the approach is the concept of a bubble score, which measures the deviation between the observed market price and the price implied by the LPPL model fit. Formally, the bubble score is defined as:

$$\epsilon(t) = \ln p(t) - \ln \hat{p}(t)$$

where $\ln p(t)$ denotes the observed log market price at time t , $\ln \hat{p}(t)$ represents the LPPL model-fitted price, and $\epsilon(t)$ captures the deviation between the two. This quantity measures the extent to which market prices diverge from the trajectory predicted by the bubble model.

Bubble scores provide a convenient way to quantify the strength of bubble dynamics in asset prices. When the observed price exceeds the LPPL fitted price, the bubble score is positive:

$$\epsilon(t) \geq 0 \text{ if } \ln p(t) \geq \ln \hat{p}(t)$$

Conversely, when the observed price falls below the LPPL fitted trajectory, the bubble score becomes negative:

$$\epsilon(t) < 0 \text{ if } \ln p(t) < \ln \hat{p}(t)$$

While positive scores correspond to classical bubble behavior, negative bubble scores capture situations where asset prices fall significantly below the trajectory implied by the LPPL structure. These cases can be interpreted as “negative bubbles,” corresponding to temporarily undervalued assets or accelerated downward price dynamics.

However, price-based bubble scores alone may still fail to capture the influence of information flows and market narratives. To address this limitation, the HLPPL framework incorporates two additional signals derived from textual data:

H_t : the Hype Index, measuring the intensity of media attention,
 S_t : the news sentiment signal, capturing the tone of news coverage.

The adjusted bubble scores are defined as:

$$\text{Positive Score} = \epsilon(t) + \alpha_{\text{hype}} H_t + \alpha_{\text{sentiment}} S_t$$

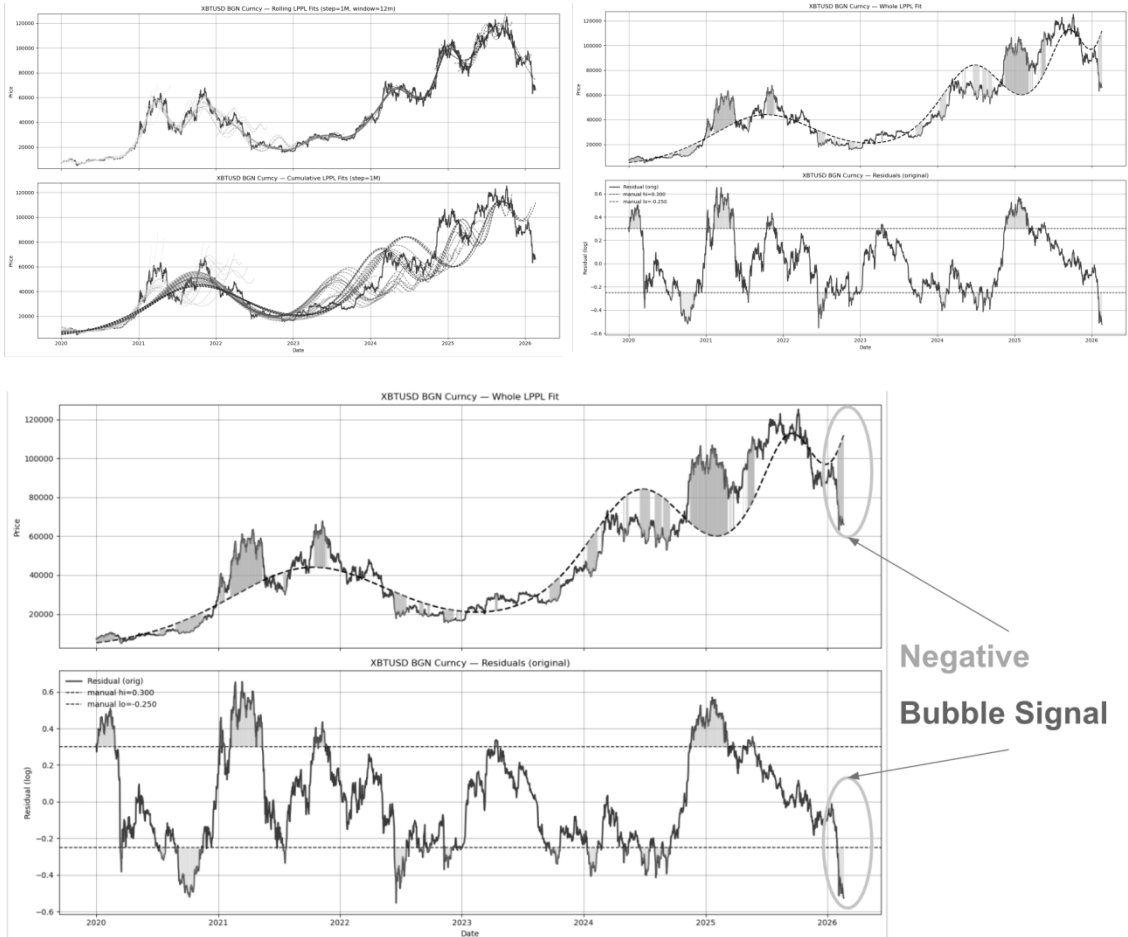
$$\text{Negative Score} = \epsilon(t) - \alpha_{\text{hype}} H_t + \alpha_{\text{sentiment}} S_t$$

where α_{hype} and $\alpha_{\text{sentiment}}$ represent weights controlling the influence of hype and sentiment signals. The Hype Index H_t captures the level of news attention associated with the asset, while the sentiment measure $s_t \in (-1,1)$ reflects the polarity of media coverage.

In this formulation, hype and sentiment act as adjustment terms that modify the bubble score derived from price dynamics alone. When strong news attention and positive sentiment coincide with accelerating price dynamics, the adjusted bubble score increases, reinforcing

the bubble signal. Conversely, when price movements occur without corresponding narrative amplification, the additional signals act as corrective buffers, reducing the likelihood of false bubble detection.

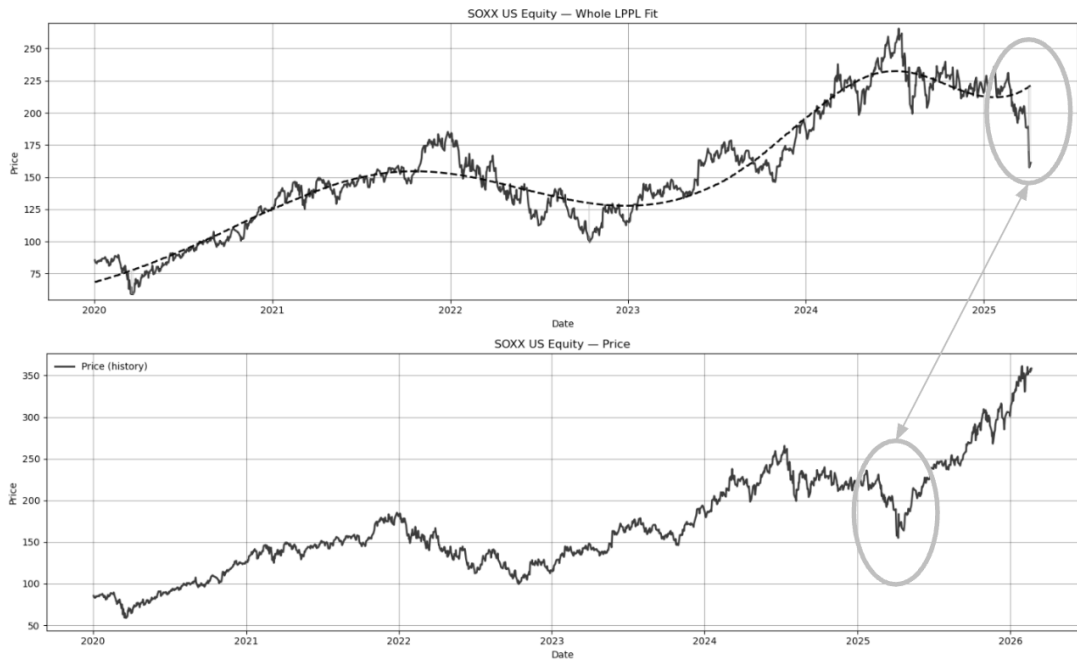
By integrating LPPL-based diagnostics with NLP-derived measures of market attention and sentiment, the HLPPL framework provides a more comprehensive approach to detecting speculative dynamics in financial markets.



V. EMPIRICAL EXAMPLES

Empirical illustrations demonstrate how the integrated framework can identify bubble signals across multiple markets.

Semiconductor equities provide a useful example due to the rapid growth associated with artificial intelligence infrastructure. The SOXX index shows periods of accelerated growth when price dynamics alone suggest potential bubble behavior.



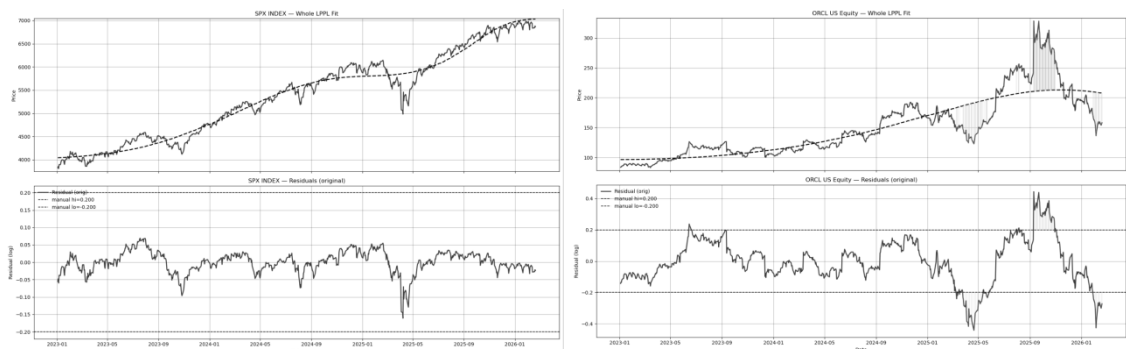
SOXX US Equity

VI. BUBBLE THRESHOLDS

Bubble scores alone do not automatically generate actionable signals. To convert scores into bubble signals, threshold values must be learned from historical data.

These thresholds are not constant. Instead, they depend on the asset, the time period, and the path of market dynamics. Machine learning techniques can be used to estimate optimal thresholds based on historical bubble episodes.

This adaptive approach allows the model to distinguish between normal periods of growth and true bubble dynamics.



Bubble Threshold Comparison: SPX Index vs. ORCL US Equity

VII. NEWS AND SENTIMENT

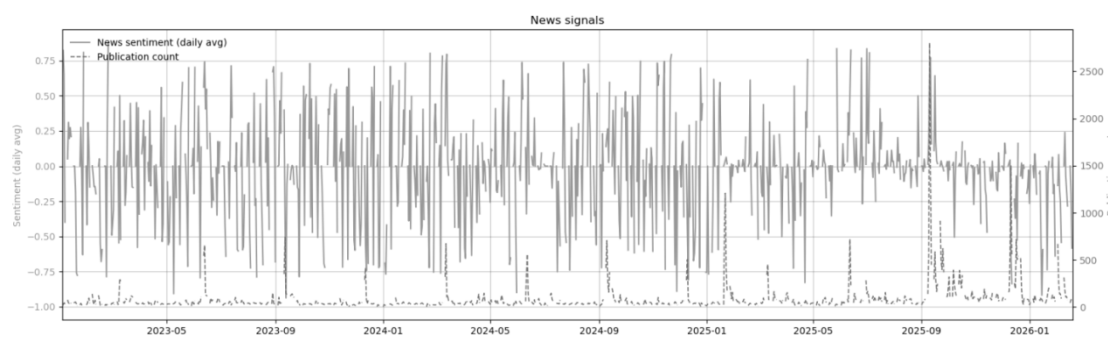
Financial bubbles are not driven solely by price dynamics. Narratives, investor attention, and media coverage often play a critical role in amplifying speculative behavior. The rapid diffusion of information through financial news and social media platforms creates feedback loops between price movements and market narratives. As a result, integrating news signals into quantitative models can significantly improve the detection of bubble dynamics.

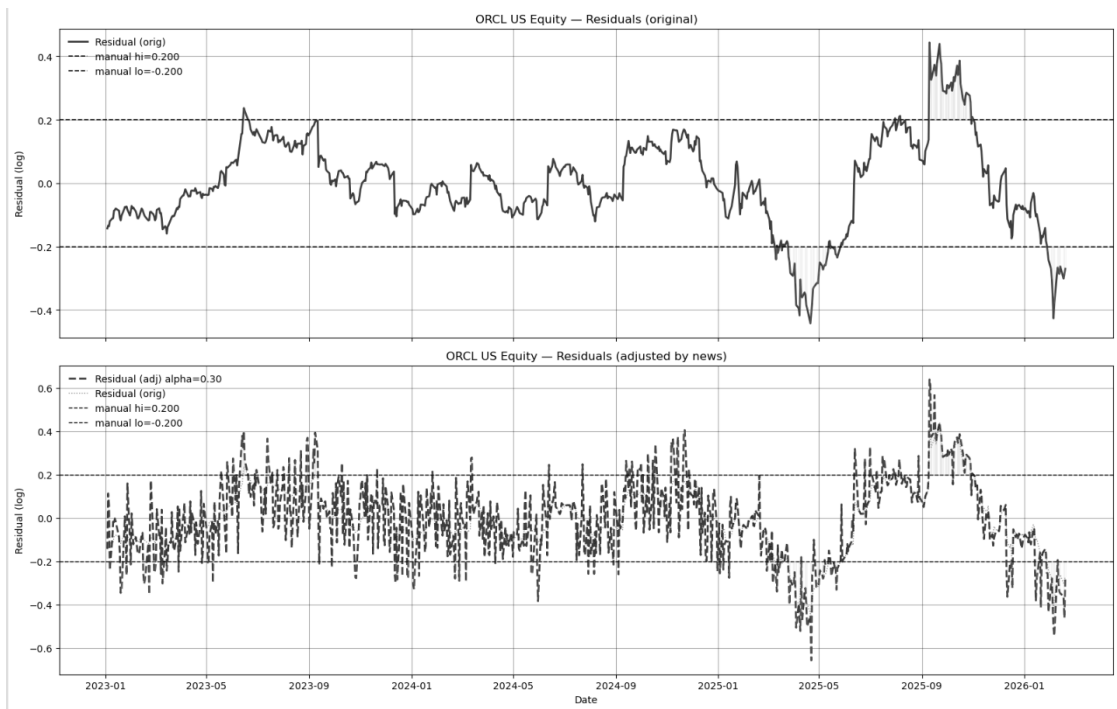
In this framework, we incorporate news intensity and sentiment measures into the bubble detection process. These signals are extracted using natural language processing techniques applied to financial news articles and other textual data sources. Two types of signals are particularly relevant:

- News Attention (Hype) – measured by the intensity with which a company or asset is mentioned in financial news relative to a baseline level of attention.
- News sentiment – measured by the average polarity of news coverage, capturing whether the tone of reporting is positive, neutral, or negative.

The integration of these signals allows the bubble detection model to distinguish between price movements driven by fundamentals and those amplified by excessive market attention.

To illustrate the effect of news and sentiment signals, we examine the case of Oracle Corporation (ORCL). Figure X presents the news signal dynamics alongside the residual structure of the LPPL model both before and after incorporating news and sentiment adjustments.





The top plot shows the evolution of news signals, including publication counts and aggregated sentiment scores. Periods of elevated attention correspond to spikes in news coverage, which often coincide with major corporate announcements or market narratives. The bottom subplots compare two residual structures:

- Top panel: residuals from the LPPL model using price dynamics alone;
- Bottom panel: residuals after adjusting news and sentiment signals.

The inclusion of news and sentiment information introduces corrective buffers to the bubble detection process. When price movements are supported by strong news-driven narratives, the adjusted residuals reflect this amplification. Conversely, when price fluctuations occur without corresponding news support, the model dampens the bubble score.

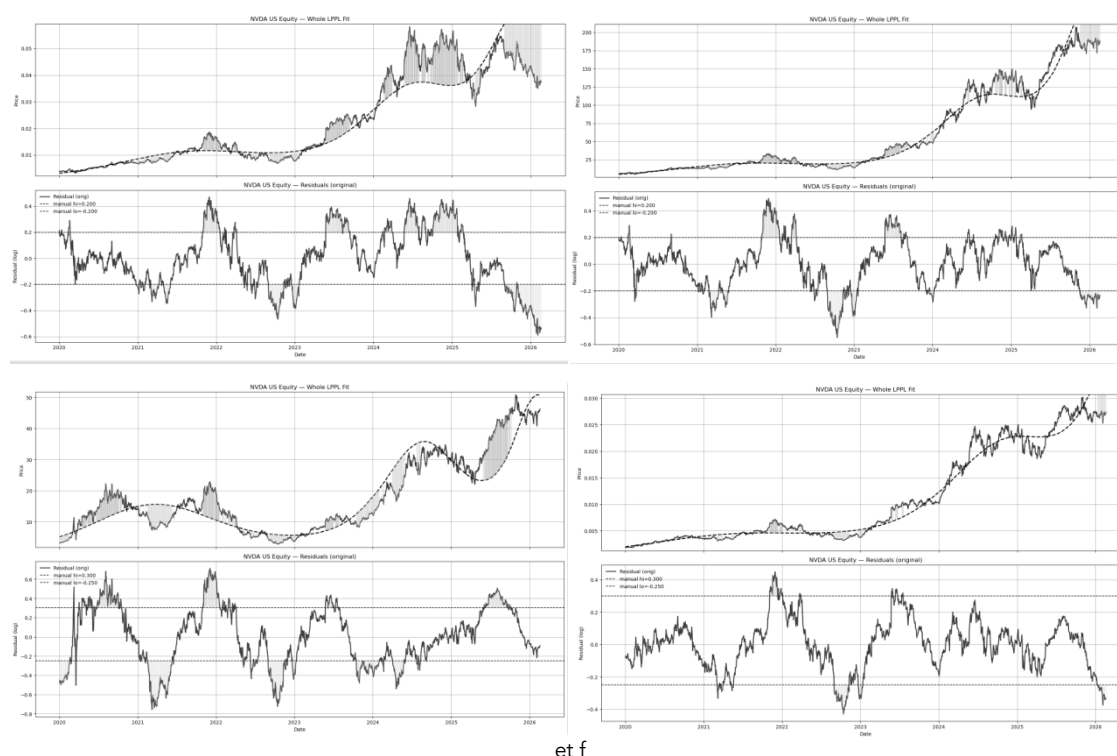
This adjustment improves the robustness of bubble detection by reducing false positives that arise from short-term volatility or purely technical price movements.

More broadly, the integration of news signals highlights the role of information flows as a catalyst for speculative dynamics. Excessive media attention can accelerate the positive feedback loops that characterize financial bubbles, while negative news shocks may contribute to the rapid unwinding of speculative positions.

VIII. CHANGE OF NUMÉRAIRE AND RELATIVE BUBBLES

An important conceptual insight is that bubbles are inherently relative phenomena. Asset prices are always measured relative to a chosen numéraire, such as a currency, a bond, or a market index.

Changing the numéraire can reveal new perspectives on bubble dynamics. For example, analyzing the price of NVIDIA relative to gold or relative to the S&P 500 can produce different interpretations of whether a bubble exists. This perspective highlights that bubble detection should account for relative valuation rather than relying solely on absolute price levels.



et f

CONCLUSION

This paper presents an integrated framework linking price-based bubble diagnostics with measures of market attention derived from textual data. Building on the LPPL framework, we introduce the HLPPL approach, which adjusts traditional bubble scores using signals capturing news intensity and sentiment extracted from financial news.

The central contribution of the framework is to combine quantitative price dynamics with information flows and market narratives. While LPPL-based diagnostics provide a structural description of bubble-like price trajectories, the addition of hype and sentiment signals allows the model to incorporate the role of media attention and narrative amplification in speculative episodes. This integration improves the robustness of bubble detection by strengthening signals when price dynamics and narratives reinforce one another, while reducing false positives when price movements occur without corresponding information flows.

The empirical illustrations presented in this paper demonstrate how the framework can be applied across different markets, including technology equities and cryptocurrencies. In particular, examples involving semiconductor equities, AI-related companies, and digital assets highlight how speculative dynamics often emerge alongside rapid increases in news coverage and narrative intensity.

More broadly, the analysis emphasizes that financial bubbles are not purely price-driven phenomena. They are shaped by feedback loops between prices, investor expectations, and information flows. Advances in natural language processing now make it possible to quantify these information flows systematically, providing new tools for studying speculative dynamics in financial markets.

The paper also highlights an important conceptual insight: bubbles are inherently relative phenomena, as asset prices are always evaluated relative to a chosen numéraire. Examining price dynamics under alternative numéraires can therefore reveal additional perspectives on speculative behavior and valuation dynamics.

Future research may further extend this framework to other asset classes (such as interest rates) and information sources. Potential directions include applications to commodities, macroeconomic indicators, and alternative data such as social media or prediction markets. As advances in artificial intelligence continue to improve the analysis of large textual datasets, integrating NLP-based signals with financial modeling may provide valuable new approaches for understanding market dynamics and identifying bubbles in the making.

References

- Cao, Z., Geman, H., (2025) 'A hype-adjusted probability measure for NLP stock return forecasting'. *Frontiers in Artificial Intelligence*
- Cao, Z., Geman, H., et al. 'Identifying and Quantifying Financial Bubbles with the Hyped Log-Periodic Power Law Model'. Working paper.
- Deveikyte, J., Geman, H., Piccarii, C., and Proveti, A (2022) 'A sentiment analysis approach to the prediction of market volatility', *Frontiers in Artificial Intelligence*
- Lin, L., Ren, R., Sornette, D. (2014). 'The volatility-confined LPPL model', *International Review of Financial Analysis*
- Vaswani, A et al (2017) Attention is all you need', *Adv in Neural Info Processing Systems*

ABOUT THE AUTHOR



HÉLYETTE GEMAN

Hélyette Geman is Senior Fellow at the Policy Center for the New South, Director of the Commodity Finance Center at Birkbeck, University of London, and a Research Professor at the Johns Hopkins University, who focuses on Probability, Commodities and Finance. She was Chairman of the Finance Department at ESSEC Business School from 1988 to 1995, and the Director of Dauphine Master 203 at the University of Paris Dauphine from 1995 to 2005.

Hélyette Geman is a Graduate from Ecole Normale Supérieure in Mathematics and holds a PhD in Probability and a PhD in Finance. She worked as a scientific expert for EDF Trading, Louis Dreyfus, BMP Billiton, Bunge Invivo and financial institutions. Her book "Commodities and Commodity Derivatives" published in 2005 by Wiley Finance has become the reference in the field.

Hélyette Geman has published more than 140 papers in major journals and counts among her former PhD students Nassim Taleb, author of the Black Swan.

ABOUT THE POLICY CENTER FOR THE NEW SOUTH

The Policy Center for the New South (PCNS) is a Moroccan think tank aiming to contribute to the improvement of economic and social public policies that challenge Morocco and the rest of Africa as integral parts of the global South.

The PCNS pleads for an open, accountable and enterprising "new South" that defines its own narratives and mental maps around the Mediterranean and South Atlantic basins, as part of a forward-looking relationship with the rest of the world. Through its analytical endeavours, the think tank aims to support the development of public policies in Africa and to give the floor to experts from the South. This stance is focused on dialogue and partnership, and aims to cultivate African expertise and excellence needed for the accurate analysis of African and global challenges and the suggestion of appropriate solutions.

[Read more](#)

All opinions expressed in this publication are those of the author.

Policy Center for the New South

Rabat Campus of Mohammed VI Polytechnic University,
Rocade Rabat Salé - 11103
Email : contact@policycenter.ma
Phone : +212 (0) 537 54 04 04
Fax : +212 (0) 537 71 31 54



www.policycenter.ma

