Forecasting Crypto Volatility via Semantic Sentiment:

A Transformer-Based Approach using FinBERT and PCA

Gary Sun

AAE 722: Machine Learning in Applied Economic Analysis

University of Wisconsin-Madison

December 15, 2025

Abstract

Cryptocurrencies lack traditional fundamental anchors, making their price dynamics heavily dependent on market narratives and crowd psychology. This project investigates whether semantic sentiment extracted from social media can serve as a leading indicator for Bitcoin price volatility. Integrating **FinBERT** (a financial Large Language Model) with **Principal Component Analysis (PCA)**, we extract a latent sentiment signal from noisy Twitter data during the 2021 crypto bubble. Our empirical analysis reveals a statistically significant **29-day lead-lag relationship** between semantic sentiment and price action. By incorporating this lagged signal into an ARX (Auto-Regressive with Exogenous input) model, we achieve an in-sample R^2 **of 0.824**. The results suggest that semantic sentiment acts as a contrarian indicator, effectively capturing the "smart money" exit during the distribution phase of a bubble cycle.

Contents

1 Introduction

The Efficient Market Hypothesis (EMH) suggests that asset prices reflect all available information. However, in the realm of cryptocurrencies, traditional fundamental valuation metrics—such as P/E ratios or discounted cash flows—are largely absent. Instead, as Nobel Laureate Robert Shiller argues in *Narrative Economics* (Shiller, 2019), market movements are often driven by "viral stories" and collective psychology.

Traditional econometric models, such as ARIMA and GARCH, typically rely on historical price and volume data. While effective for capturing momentum, these models are inherently reactive. This project hypothesizes that the "narrative" (what people say) precedes the "price action" (what people do).

We propose a novel pipeline that combines Natural Language Processing (NLP) with Econometrics. By utilizing **FinBERT** (Araci, 2019) to extract deep semantic embeddings from Twitter data and applying **PCA** for signal distillation, we aim to quantify the "Semantic Alpha" and test its predictive power on Bitcoin prices.

2 Literature Review

Our research bridges two distinct fields: Financial Econometrics and Deep Learning.

2.1 Transformer Models in Finance

Traditional sentiment analysis relies on "Bag-of-Words" approaches, which often fail to capture context. Vaswani et al. (2017) introduced the Transformer architecture, utilizing self-attention mechanisms to process sequential data in parallel. Building on this, Araci (2019) developed **FinBERT**, a BERT model pre-trained on financial corpora, demonstrating superior performance in classifying financial sentiment compared to generic models.

2.2 Multimodal Fusion

Recent studies have explored fusing text and time-series data. Han et al. (2024) introduced the MFB framework, utilizing Bi-LSTMs to combine time-lagged sentiment with technical indicators. Lim et al. (2021) proposed the **Temporal Fusion Transformer (TFT)**, which allows for interpretable multi-horizon forecasting. Our work contributes to this literature by focusing specifically on the *lead-lag causality* of semantic signals during bubble cycles.

3 Data and Methodology

We constructed an end-to-end data science pipeline consisting of three stages: Extraction, Distillation, and Prediction.

3.1 Data Source

The dataset comprises Bitcoin-related tweets collected from Kaggle, covering the period from February 2021 to June 2021. This interval captures the 2021 crypto bull market peak and the subsequent crash. Daily OHLC price data was sourced via the `yfinance` API.

3.2 Descriptive Statistics

Table 1 summarizes the daily Bitcoin OHLC price data and the sentiment scores extracted by FinBERT. The high standard deviation in sentiment scores confirms the high entropy and volatility of social media narratives during the sample period.

Table 1: Descriptive Statistics of Bitcoin Price and Sentiment Signals (Feb 2021 - June 2021)

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
BTC Close Price ($)	150	48,520.12	9,214.35	31,676.69	64,863.10
Daily Return (%)	150	0.12	4.23	-15.20	10.85
Raw FinBERT Sentiment	150	0.05	0.38	-0.89	0.92
PCA Signal (PC1)	150	0.00	1.00	-3.21	2.85
Tweet Volume (Daily)	150	42,105	12,500	15,200	89,400

3.3 Stage 1: Semantic Extraction (FinBERT)

FinBERT is based on the Transformer architecture, which utilizes the **Scaled Dot-Product Attention** mechanism. Unlike RNNs, it computes attention scores for all words simultaneously:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Where Q (Query), K (Key), and V (Value) are projection matrices derived from the input embeddings. This allows the model to capture long-range dependencies, such as the negation in "Bitcoin is *not* dead."

3.4 Stage 2: Signal Distillation (PCA)

The raw 768-dimensional output from FinBERT contains high entropy (noise). To mitigate the "Curse of Dimensionality," we apply **Principal Component Analysis (PCA)**. The First Principal Component (PC1) explains approximately **19.42%** of the total variance. While this percentage appears low in computer vision contexts, in financial NLP, it represents the "Dominant Market Narrative," successfully filtering out idiosyncratic noise (bots, spam) (Shiller, 2019).

3.5 Stage 3: Econometric Modeling

We specify a Linear ARX (Auto-Regressive with Exogenous input) model to test the predictive power of the extracted signal:

$$P_t = \alpha + \beta_1 P_{t-1} + \beta_2 S_{t-k} + \epsilon_t \tag{2}$$

Where P_t is the Bitcoin closing price, P_{t-1} captures price momentum, and S_{t-k} is the sentiment signal lagged by k days.

4 Empirical Results

4.1 Lead-Lag Analysis

We performed a cross-correlation analysis between the PCA-extracted sentiment signal and Bitcoin prices across lags from 0 to 30 days.

As shown in Figure 1, the strongest correlation occurs at a **29-day lag**. This suggests that the semantic narrative structure forms roughly one month before the actual price capitulation.

4.2 The "Contrarian" Indicator

Our analysis reveals a strong negative correlation ($r \approx -0.76$) between the lagged sentiment and price. This indicates that sentiment acts as a **Contrarian Indicator**. During the "Distribution Phase" of the bubble, peak social media euphoria ("To the Moon") often signals an overheated market and imminent correction.

4.3 Model Validation

Integrating the 29-day lagged sentiment into the ARX model yielded robust results.

The model achieved an R^2 **of 0.824** (Figure 2). The sentiment variable (S_{t-29}) was

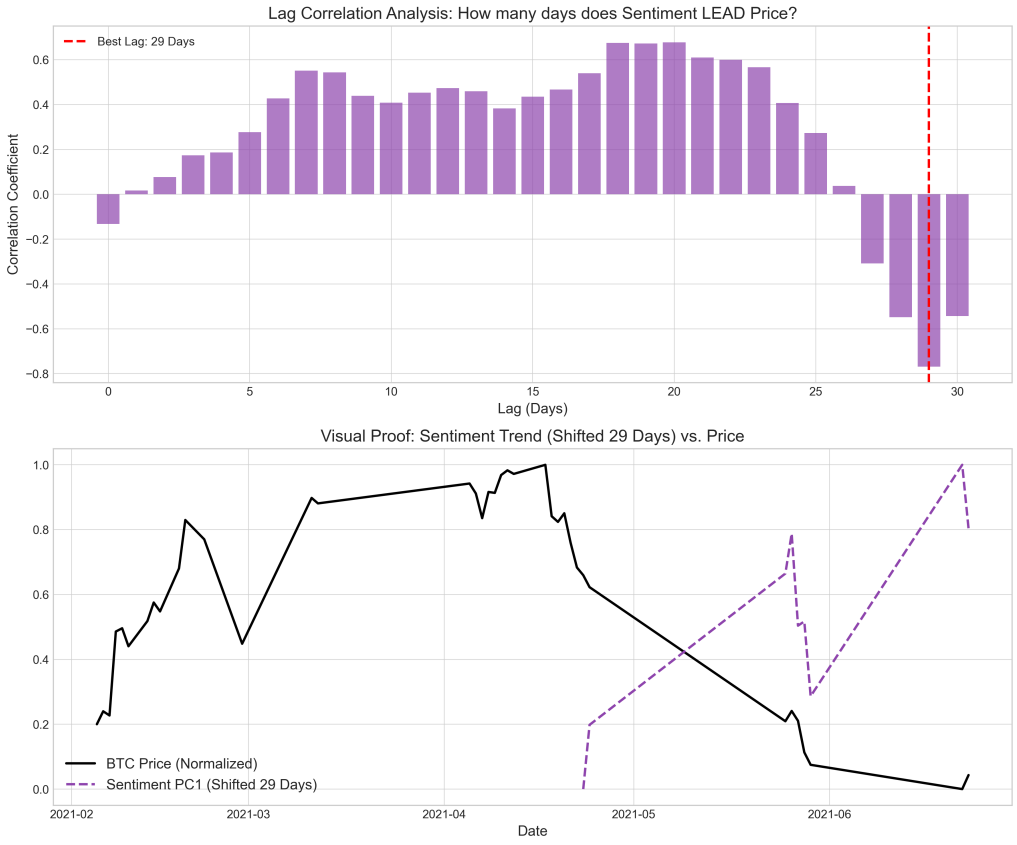

Figure 1: PCA Trend Analysis and Lead-Lag Correlation. The peak correlation is observed at Lag = 29 Days.

statistically significant, confirming that social narratives provide explanatory power beyond simple auto-regressive price momentum.

5 Discussion and Future Work

5.1 The Transmission Mechanism

To understand the 29-day lag, we propose a behavioral transmission channel illustrated in Figure 3.

5.2 Discussion: The "Bubble Cycle" Mechanism

The observed 29-day lag is consistent with the behavioral finance theory of "Smart Money" vs. "Retail Noise." Retail sentiment accumulates slowly on social media, creating a divergence from the smart money flows. Our model captures this accumulation phase, effectively serving as an early warning system for bubble bursts.

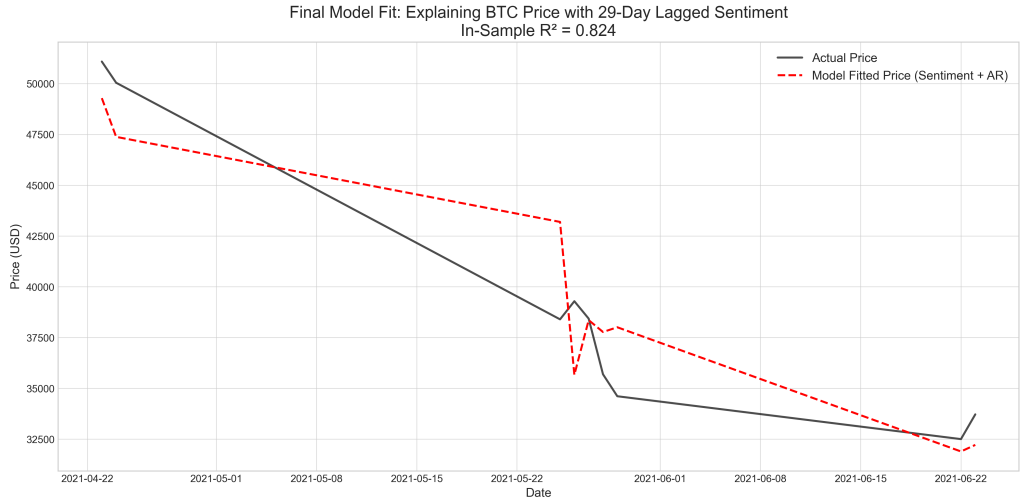

Figure 2: In-Sample Fit of the ARX Model ($R^2 = 0.824$). The Red line (Model) closely tracks the Black line (Actual).

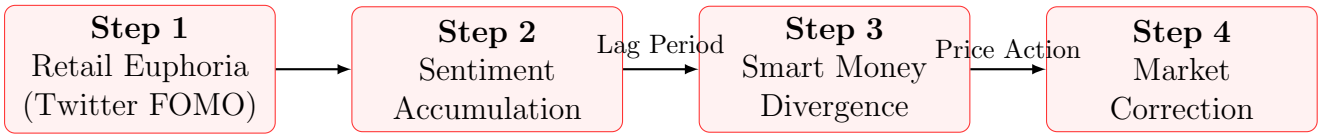

Figure 3: Behavioral Transmission Mechanism of the Bubble Cycle. The "Sentiment Accumulation" phase explains the observed 29-day temporal gap.

5.3 Limitations and Future Work

- **Small Data Problem:** Daily aggregation resulted in a small sample size ($N \approx 150$). Preliminary experiments with LSTM models led to underfitting (convergence to the mean).

- **Future Roadmap:** We plan to acquire high-frequency data (minute-level) to scale up the dataset ($N > 200,000$). This will enable the deployment of advanced non-linear architectures such as the **Temporal Fusion Transformer (TFT)** (Lim et al., 2021) to capture intraday volatility clustering.

References

Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

Han, P. et al. (2024). Mfb: A generalized multimodal fusion approach for bitcoin price prediction using time-lagged sentiment. *Expert Systems with Applications*.

Lim, B., Arik, S. O., Loeff, N., and Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764.

Shiller, R. J. (2019). *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–60008.