Causal Effects of Social Media Sentiment on Cryptocurrency Market Returns and Volatility

Double ML · IV-DML · Causal Forest | Bitcoin & Ethereum

Abstract

Cryptocurrency markets are highly sensitive to social media sentiment. Yet, the causal direction of these effects remains elusive owing to the predominance of correlational approaches that fail to address confounding and reverse causality. Against this backdrop, this study provides causally identified evidence on the impact of social media sentiment on cryptocurrency returns and volatility by applying a three-stage causal machine learning framework to a daily panel of Bitcoin (BTC) and Ethereum (ETH) over 159 trading days (January 2 - June 12, 2023). The study utilized publicly available datasets of price data from CoinMarketCap and sentiment from the Crypto Twitter Influencer Sentiment Dataset. First, the Double Machine Learning Partially Linear Regression (DML-PLR) model removed high-dimensional observable confounding using cross-fitted Random Forest nuisance estimators. Second, Partially Linear Instrumental Variables (PLIV) were used to correct for treatment endogeneity, confirmed in ETH importance-weighted sentiment via Hausman pre-testing. Third, Causal Forest was employed to estimate Conditional Average Treatment Effects (CATEs) and assess heterogeneity across market regimes and influencer importance. Thereafter, a series of robustness checks (such as alternative nuisance learners, rolling-window stability analysis, leave-one-out tests, placebo and permutation diagnostics, lagged-treatment sensitivity, bootstrap and hyperparameter stability for Causal Forest, and macro-event confounder contCryptocurrencyrols) were conducted to ascertain the reliability of the study findings. The results revealed a robust and consistently contrarian causal effect of sentiment on returns across all model specifications. The periods of elevated positive sentiment systematically precede price declines in tandem with a peak-optimism reversal mechanism. The strongest effect was observed for ETH, with a Causal Forest average treatment effect of -0.294 ($p\text{-}value < 0.001$). Significant heterogeneity was also identified. For BTC, the effect reversed sign across regimes, positive in bear markets (+0.161) and negative in bull markets (-0.133), while for ETH, the magnitude of the negative effect was amplified on high-influence days ($p\text{-}value = 0.063$). In contrast, no statistically significant causal effect of sentiment on realized volatility was detected, rejecting the uncertainty-amplification channel. The BTC contrarian effect remained stable across all diagnostics, whereas the ETH effect, though economically large, showed moderate sensitivity to temporal windows and heterogeneity specifications. Placebo and permutation tests rejected spurious correlation, and lagged-treatment results indicated that sentiment effects are primarily contemporaneous rather than persistent. Thus, the study advances the literature by providing rigorously causal evidence on the crypto–sentiment nexus, moving beyond the limitations of prior correlational analyses.

Keywords: Cryptocurrency, Bitcoin, Ethereum, Social media sentiment, VADER, Double Machine Learning, Causal Forest, Treatment effect heterogeneity, Causal inference, Influencer sentiment

1.0 Introduction

The cryptocurrency can move billions of dollars in minutes just in response to the social media posts of highly influential market participant(s). For instance, Elon Musk's tweets in January 2021 led to a ripple effect on cryptocurrency prices, with an increase exceeding 20%; consequently, the market dropped around 15% following his retraction of support (Ante, 2023). This scenario demonstrates one of the core features of cryptocurrency markets: the market moves extremely rapidly (Le Tran & Leirvik, 2020), and valuation is significantly affected not by institutional investments or fundamental analysis (Awad, 2022), but rather by individual sentiment and real-time social media activity (Koltun & Yamshchikov, 2023; Kreuzer *et al.*, 2026). The global cryptocurrency market valued at about USD 3 trillion at its peak (Kulcsar & van Engelen, 2023), which implies sentiment-driven shifts can create or destroy enormous value rapidly (Zhang *et al.*, 2020). This raises a pivotal question: Do social media signals merely reflect market movements, or do they causally predict cryptocurrency returns?

Studies by Kraaijeveld & De Smedt (2020), Philippas *et al.* (2019), Kyriazis *et al.* (2023), and Ante (2023) document a nexus between social media sentiment and cryptocurrency market outcomes such as returns, trading volume, and volatility. However, it is difficult to establish causality from such correlational evidence since both sentiment and returns are jointly influenced by evolving market conditions, trading activity, and informational events; while feedback effects between price movements and sentiment further complicate identification (Kraaijeveld & De Smedt, 2020; Oikonomopoulos *et al.*, 2022). Moreover, theories posit that sentiment effects should vary across various market regimes and investor attention levels according to the Adaptive Market Hypothesis framework (Lo, 2004) and attention-based trading models (Da *et al.*, 2011). Nonetheless, the existing findings are unreconcilable, and the causal impact of social media sentiment on cryptocurrency markets remains unresolved.

Given the backdrops, this study examines the aforementioned using a unified causal approach and seeks to answer three research questions: **(RQ1)** Does social media sentiment have a causal impact on the daily returns of Bitcoin (BTC) and Ethereum (ETH), adjusting for confounding and endogeneity? **(RQ2)** Does sentiment causally impact realized volatility? **(RQ3)** Are the causal impacts heterogeneous across market regimes and levels of influencer importance? To address these questions, the study employs a three-stage causal machine learning approach to a 159-day panel dataset spanning January 2 through June 12, 2023. This approach integrates Double/Debiased Machine Learning, Partially Linear Instrumental Variables, and Causal Forest. This enhances flexible adjustment for confounding, correction for endogeneity, and estimation of heterogeneous treatment effects, thereby allowing the direct testing of the predicted influencer-dependent mechanisms. Furthermore, several robustness checks were conducted to ascertain the validity of the results; this includes alternative nuisance models, rolling window stability, lagged treatment sensitivity, macro event confounders, bootstrap/hyperparameter stability for Causal Forest, placebo, permutation, and leave-one-out tests.

The main findings are fourfold: First, sentiment has a robust contrarian effect on returns across all assets and estimators (positive sentiment shocks are followed by negative returns) in tandem with the peak optimism reversal mechanism (De Long *et al.*, 1990; Baker & Wurgler, 2006). Second, the effect is strongest for ETH, where the Causal Forest estimator indicates that the average treatment effect is -0.294 (95% CI [-0.447, -0.141], *p-value* < 0.001). Third, Bitcoin's response is regime-dependent, turning positive in bear markets (CATE = +0.161) and negative in bull markets (CATE = -0.133), and the difference is statistically significant at 1% level (*p-value* < 0.001). On the other hand,

no evidence of causal effect of sentiment shocks on asset volatility; thus, we reject the uncertainty amplification hypothesis. Moreover, the robustness diagnostics strongly support the primary findings: The BTC contrarian effect remains invariant to alternative nuisance learners, rolling windows, leave-one-out samples, placebo tests, and macro event controls, whereas the ETH effect, although economically large, is more sensitive to window and heterogeneity specifications. Therefore, social media sentiment matters for returns, but its impact is non-linear, heterogeneous, and asset-specific. To the best of our knowledge, this paper provides the first causal-ML identified sentiment effects on cryptocurrency returns and volatility in terms of direction, magnitude, and heterogeneity. The remaining sections are as follows: Section 2 reviews the literature; Section 3 describes the data, variable construction, and detailed methodological framework; Section 4 presents the results and discusses the findings; and Section 5 concludes with policy implications and directions for future research.

2.0 Literature Review

2.1 Cryptocurrency Market Efficiency and Return Predictability

Market efficiency is central to understanding the predictability of cryptocurrency returns and the causal effect of sentiment measures. Earlier studies reported that Bitcoin is not weak-form efficient (Urquhart, 2016) but is trending toward weak-form efficiency (Nadarajah & Chu, 2017). Subsequent research has substantially modified this perspective to imply that markets are neither permanently efficient nor permanently inefficient (Le Tran & Leirvik, 2020), but rather possess dynamic efficiency (Aslam *et al.*, 2023), time-varying efficiencies in line with the Adaptive Market Hypothesis (Lo, 2004). This is rooted in reaction to market changes; it becomes more efficient as markets mature but less efficient amid structural changes (Le Tran & Leirvik, 2020; Khursheed *et al.*, 2020; Vukovic *et al.*, 2025). By and large, this inefficiency is ubiquitous, as only 6.04% of 893 assets meet weak-form Efficient Market Hypothesis requirements (Kang *et al.*, 2022), with larger and more liquid assets performing better in this regard (Wei, 2018; Brauneis & Mestel, 2018). The non-linear dependencies, multifractality, and long memory persistence observed in Bitcoin, Ethereum, and most leading crypto markets contradict the random walk hypothesis and confirm that past price information is not completely absorbed in financial markets (Yi *et al.*, 2023; Aslam *et al.*, 2023).

Efficiency differs across assets and market regimes: In bear markets, cryptocurrency returns display positive autocorrelation (Zhang *et al.*, 2020), while in bull markets, returns tend to a random walk (Zhang *et al.*, 2020), underlining that the predictability window of external information is substantially larger in bear markets than in bull markets (Zhang *et al.*, 2020; Noda, 2021; Khursheed *et al.*, 2020). Additionally, Bitcoin is consistently more efficient than Ethereum, except during sharp market corrections, in which Ethereum tends to reach near-absolute inefficiency (Noda, 2021). That is, Ethereum might be more prone to shocks from external information. The return prediction accuracy of machine learning classifiers applied to major cryptocurrency markets is between 55 and 65%, reliably higher than the 50% threshold of the random walk hypothesis (Akyildirim *et al.*, 2021; Karasiński, 2025), and on-chain indicator strategies consistently outperform passive strategies with Sharpe ratios reaching 100% success rates (Łęt *et al.*, 2022). Thus, cryptocurrency markets exhibit time-varying efficiency, with predictability depending on asset

characteristics and market regimes. Within the foregoing, external signals such as social media sentiment may have heterogeneous effects across assets and market conditions.

2.2 Investor Sentiment and Financial Market Dynamics

Investor sentiment is an integral part of asset returns that exceed what classical theoretical models would allow. In the seminal paper of Baker & Wurgler (2006), aggregate sentiment predicts the cross-section of stock returns, exhibiting sign reversals between high- and low-sentiment periods, that are particularly pronounced among stocks that are hardest to value or arbitrage, due to sentiment-driven mispricing. Antweiler & Frank (2004) and Tetlock (2007) used information content of social media to predict next-day volatility, trading volume, and slightly negative returns, respectively, with increased bullish sentiment. Evidence from early studies on the sentiment-driven dynamics foreshadows the current study. In terms of measurement, financial sentiment analysis has moved from traditional lexicons to the application of transformers (Jahanbin & Zare Chahooki, 2023), with negative sentiment having particularly strong asymmetric predictive power (Du *et al.*, 2024). Among lexicon-based tools, VADER achieves human-level classification accuracy (F1 = 0.96) for social media text (Hutto & Gilbert, 2014), thus the basis for its adoption in this study.

2.3 Social Media Sentiment and Cryptocurrency Markets

Social media sentiment, particularly from X (formerly Twitter), influences cryptocurrency returns, volume, and volatility (Jahanbin & Zare Chahooki, 2023; Koltun & Yamshchikov, 2023). At a daily frequency, Twitter sentiment Granger-causes returns for Bitcoin, Bitcoin Cash, and Litecoin (Kraaijeveld & De Smedt, 2020), with the strongest predictive correlations at one-to-two-day lags (Raheman *et al.*, 2022). Influencer effects are particularly significant, with high-profile social media posts generating abnormal returns and volume responses that depend on sentiment tone (Ante, 2023), while media sentiment from Twitter and Google Trends causes price jumps, with negative sentiment having a disproportionately larger effect (Philippas *et al.*, 2019). In addition, sentiment influences the volatility channel. Twitter-based uncertainty exerts a persistent non-linear effect on cryptocurrency volatility, particularly during crises (Kyriazis *et al.*, 2023), whereas BERT-based sentiment influences Ethereum's realized volatility at a 36-hour lag (Awad, 2022). In cryptocurrency futures markets, news sentiment is significant in predicting returns on Bitcoin futures, while for Ethereum, only positive sentiment is significant, an asset-level directional effect consistent with the results of this paper (Kreuzer *et al.*, 2026).

Nonetheless, empirical consensus remains elusive: Positive momentum effects (Kraaijeveld & De Smedt, 2020), contrarian reversals (Wołk, 2020), and near-null associations (Khan, 2022; Kulcsar & van Engelen, 2023) exist without resolution. This is expected due to two fundamental identification issues that existing studies fail to address. Firstly, high-dimensional confounding, whereby lagged returns, trading volume, market regime, and intraday range simultaneously affect sentiment and returns, without adjustment for the confounding space. Secondly, reverse causality, whereby returns cause sentiment for Bitcoin while the relationship is reversed for Ethereum (Oikonomopoulos *et al.*, 2022), with bidirectionality confirmed by Kulcsar & van Engelen (2023) and acknowledged by Kraaijeveld & De Smedt (2020), yet unaddressed.

2.4 Bridging the Identification Gap with Causal Machine Learning

Causal machine learning offers tools to move beyond association and identify structural treatment effects in observational data. Double/Debiased Machine Learning (DML) provides a framework for estimating the causal treatment effect (Chernozhukov *et al.*, 2018). In this approach, the causal treatment effect parameter in a partially linear model is estimated using a Neyman-orthogonal score that is robust to first-order errors in nuisance function estimation. This framework utilizes cross-fitted Machine Learning (ML), such as Random Forests, to estimate nuisance functions nonparametrically, avoiding regularization bias while retaining root-n consistency and asymptotic normality for the treatment effect. This approach avoids high-dimensional confounding effects on Ordinary Least Squares (OLS) estimates without imposing any distributional assumptions on such confounding effects. The DML framework is extended to handle endogenous treatment effects by applying a Partially Linear Instrumental Variables (PLIV) model (Robinson, 1988; Chernozhukov *et al.*, 2018), where the treatment effect is additionally instrumented by a set of variables satisfying relevance and exclusion restrictions, which is necessitated by Hausman pre-testing for endogeneity of ETH importance-weighted sentiment. For heterogeneous treatment effects, a Causal Forest approach proposed by Wager and Athey (2018) trains a Random Forest on pseudo-outcomes constructed by Robinson (1988) using cross-fitted residuals to provide a Conditional Average Treatment Effect (CATE) estimate, which is pointwise asymptotically normal, and to test for heterogeneity assessment vis-à-vis the Best Linear Projection test (Athey *et al.*, 2019; Athey & Imbens, 2019). These methods are pertinent given the theoretical position that treatment effects should differ across regimes, assets, or levels of influencer importance, as plausible in cryptocurrency markets.

Summary

Extant evidence is largely correlational, obscured by high-dimensional confounding and reverse causality, and unable to distinguish between the structural effect of sentiment and its spurious statistical counterpart. Moreover, there is no consensus in the findings as some studies reported positive momentum effects (Kraaijeveld & De Smedt, 2020), contrarian reversals (Wołk, 2020), and near-null associations (Khan, 2022; Kulcsar & van Engelen, 2023). This discrepancy is more plausibly attributable to identification failure rather than effect heterogeneity. Additionally, whether sentiment effects in the market are heterogeneous in terms of market regime, influencer importance, and assets is not taken into account, despite the implications for investment strategy and market oversight.

3.0 Methodology

3.1.0 Data Description and Preliminary Tests

The study utilized two publicly available datasets spanning January 2 through June 12, 2023 ($n = 159$ trading days). All variables for the eight causal machine learning models (M1 - M8), namely, outcome (Y), treatment (T), control (X), instrument (Z), and moderator (M), were constructed (refer to Table 2 for details). The time frame was specifically chosen to include the post-FTX collapse adjustment phase, which involved an initial pessimistic sentiment regime

followed by partial market recovery in Bitcoin (BTC) and Ethereum (ETH). BTC and ETH were chosen for their significant market capitalization, liquidity, and social media presence.

3.1.1 Cryptocurrency Market Data

Daily market variables were sourced from the CoinMarketCap Cryptocurrency Market dataset. The dataset contains USD-denominated Open, High, Low, and Close (OHLC) prices, trading volumes, and market capitalization. The study used the BTC and ETH datasets for the specified timeframe, resulting in 159 complete observations without missing values. Two outcome variables were constructed. First, daily log returns are defined as:

$$r_{i,t} = \ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right) \tag{1}$$

where $P_{i,t}$ denotes the closing price of the asset $i \in \{\text{BTC, ETH}\}$ at time t. This transformation ensures stationarity, confirmed by ADF (*p-value* < 0.001) and KPSS (p > 0.09) tests.

Second, seven-day realized volatility is computed as:

$$\sigma_{i,t} = \sqrt{252} \times \sqrt{\frac{1}{7}\sum_{k=0}^{6}\left(r_{i,t-k} - \bar{r}_7\right)^2} \tag{2}$$

where $\bar{r}_7$ represents the seven-day average return, and $\sqrt{252}$ denotes the annualization of volatility based on the daily trading convention. The volume was included as a control variable using the first-differenced logarithmic form to ensure stationarity. The intraday price range was constructed using the first difference of ln(high/low) to capture daily information intensity, while market capitalization was excluded due to collinearity with price.

3.1.2 Influencer Sentiment Data

The Crypto Twitter Influencer Sentiment dataset contains tweets collected from over 50 verified cryptocurrency influencers by Jahanbin *et al.* (2023). This represents a high-signal measure of sentiment, which includes only expert accounts whose tweets on BTC and ETH were collected via the Twitter API and *snscrape*. This includes only publicly available information and excludes personally identifiable information. The data preprocessing followed a standard procedure, which included lowercase transformation, removal of punctuation and stop words using Natural Language Toolkit, and lemmatization via the *WordNetLemmatizer*.

Sentiment was measured using VADER (Hutto & Gilbert, 2014), which assigns each tweet. j a compound score:

$$\text{Compound}_j \in [-1, +1] \tag{3}$$

VADER offers a continuous and interpretable measure of sentiment appropriate for causal estimation. Additionally, it achieves high classification accuracy (F1 = 0.96; Hutto & Gilbert, 2014) and avoids the discretization inherent in deep learning classifiers.

Each tweet was further assigned an engagement-based importance coefficient:

$$IC_j = RT_j + 2 \cdot FAV_j + 0.5 \cdot REP_j \tag{4}$$

where RT_j, FAV_j, and REP_j denote retweets, favorites, and replies, respectively. The coefficient was normalized using min-max scaling:

$$\widetilde{IC}_j = \frac{IC_j - \min(IC)}{\max(IC) - \min(IC)} \tag{5}$$

yielding a bounded measure of influencer reach used in treatment weighting and moderation analysis.

3.1.3 Treatment, Control, Instrument, and Moderator Variables

Tweet-level data were aggregated to the daily coin-level to construct treatment variables. The aggregate daily sentiment is defined as:

$$VADER_{i,t} = \frac{1}{N_{i,t}} \sum_{j=1}^{N_{i,t}} \text{compound}_j \tag{6}$$

where $N_{i,t}$ is the number of tweets for coin $i$on day t. The importance-weighted sentiment is given by:

$$WS_{i,t} = \sum_{j=1}^{N_{i,t}} \left(\text{compound}_j \times \widetilde{IC}_j\right) \tag{7}$$

Both series were first-differenced before estimation:

$$\Delta VADER_{i,t} = VADER_{i,t} - VADER_{i,t-1}, \Delta WS_{i,t} = WS_{i,t} - WS_{i,t-1}$$

to address non-stationarity and reflect the economically meaningful notion that changes in sentiment drive marginal trading responses.

Five control variables were included in all models: lagged returns ($r_{i,t-1}$), first-differenced log trading volume ($\Delta\log$ volume$_{i,t}$), intraday price range, a bull-market regime indicator (equal to one if the 7-day cumulative return is positive), and the 7-day cumulative return. Variance inflation factors below 3 confirmed the absence of multicollinearity.

Instrumental variables for the IV-DML specification were constructed as:

$$\log(1 + \text{retweets}_{i,t}), \log(1 + \text{favourites}_{i,t}) \tag{8}$$

which corrected for extreme right skewness and served as exogenous predictors of sentiment.

The moderator variables included aggregate influencer importance:

$$IMP_{i,t} = \sum_{j=1}^{N_{i,t}} \widetilde{IC}_j \tag{9}$$

capturing total network reach, and the bull-market regime indicator, capturing market-state heterogeneity.

Table 1: Research Hypotheses and Formal Predictions

Hyp.	Statement	Formal prediction	Model
H_1	• Social media sentiment has a non-zero causal effect on cryptocurrency daily log returns, after removing observable confounders and using the first-differenced treatment specification.	$\theta \neq 0$	1 - 4
H_{1-dir}	• The causal effect is negative (contrarian): a positive shock to sentiment is followed by declining returns, consistent with the peak-optimism reversal mechanism.	$\theta < 0$	1 - 4
H_2	• Social media sentiment has a non-zero causal effect on 7-day realized volatility, after removing observable confounders.	$\theta vol \neq 0$	5 - 6
H_3	• The causal effect of sentiment on returns is heterogeneous across market regimes: the effect differs in magnitude or sign between bear and bull markets.	$\tau(Bear) \neq \tau(Bull)$	7 - 8
H_4	• The causal effect of sentiment on returns is larger in absolute magnitude on days characterized by high aggregate influencer importance.	$\lvert\tau(hi)\rvert > \lvert\tau(lo)\rvert$	7 - 8

Note:

- θ denotes the average treatment effect estimated via Double Machine Learning Partially Linear Regression (DML-PLR) specification.
- θ_{vol} denotes the corresponding DML-PLR estimate when realized volatility is the outcome variable. $\tau(\cdot)$ denotes the Conditional Average Treatment Effect (CATE) estimated using Causal Forest.
- The bull-market indicator is defined as:

$$\text{Bull_regime}_t = \begin{cases} 1, & \text{if 7-day cumulative return} > 0 \\ 0, & \text{otherwise} \end{cases}$$

- H_3 - subgroup definitions:

$$\text{Bear} = \{t: \text{bull_regime}_t = 0\}, \text{Bull} = \{t: \text{bull_regime}_t = 1\}$$

- H_4 - Subgroup definitions: observations were split at the 75th percentile (Q75) of daily aggregate influencer importance.
- Statistical significance levels were set at $\alpha = 0.05$ for H_1- H_2 and $\alpha = 0.10$ for H_3 - H_4, thus, a relatively small sample size ($n \approx$ 157 - 159).

3.1.4 Variable Descriptions

Table 2 provides a complete variable dictionary for all features used in the causal ML pipeline. Variables were grouped by their econometric roles (Y = Outcome, T = Treatment, X = Control, Z = Instrument, M = CATE moderator). Transformation notes were included where variables were first-differenced or log-transformed before estimation.

Table 2: Variable Descriptions

Variable Name	Symbol	Measurement & Construction
Outcome Variables (Y):		
BTC Log Daily Return	log_return_btc	• Natural log of daily close-to-close price ratio: $\ln(P_t/P_{t-1})$. Stationary by construction
ETH Log Daily Return	log_return_eth	• Natural log of daily close-to-close price ratio: $\ln(P_t/P_{t-1})$. Stationary by construction.

BTC 7-Day realized Volatility	realized_vol_7d_btc	• Rolling 7-day standard deviation of daily log returns, annualized: $\sigma(\text{log_ret}) \times \sqrt{252}$.
ETH: 7-Day realized Volatility	realized_vol_7d_eth	• Rolling 7-day standard deviation of daily log returns, annualized: $\sigma(\text{log_ret}) \times \sqrt{252}$.
Treatment Variables (T):		
BTC Aggregate VADER Sentiment (Δ)	Δvader_score_btc	• Daily mean VADER compound score across all BTC-related influencer tweets, first-differenced (Δ) for stationarity. VADER range $[-1, +1]$; Δ ∈ (-2, +2).
ETH Aggregate VADER Sentiment (Δ)	Δvader_score_eth	• Daily mean VADER compound score across all ETH-related influencer tweets, first-differenced (Δ) as a precaution against non-stationarity.
BTC Importance-Weighted Sentiment (Δ)	Δinf_weighted_sentiment_btc	• Daily sum of (VADER score × normalized follower importance weight) across BTC influencer tweets, first-differenced. Weights sum to 1 per day across active influencers.
ETH Importance-Weighted Sentiment (Δ)	Δinf_weighted_sentiment_eth	• Daily sum of (VADER score × normalized follower importance weight) across ETH influencer tweets, first-differenced. Primary treatment for causal heterogeneity analysis.
Control Variables (X):		
BTC Lagged Log Return	lag1_return_btc	• $\text{log_}return_{t-1}$; captures short-term price momentum and autocorrelation.
ETH Lagged Log Return	lag1_return_eth	• $\text{log_}return_{t-1}$; captures short-term price momentum.
BTC Δ Log Trading Volume	Δlog_volume_btc	• First difference of log daily BTC trading volume (USD); controls for liquidity shocks. Original series is I(1).
ETH Δ Log Trading Volume	Δlog_volume_eth	• First difference of log daily ETH trading volume (USD); controls for liquidity shocks.
BTC Intraday Price Range	intraday_range_btc	• $\ln\left(\frac{High_t}{Low_t}\right) - \ln\left(\frac{High_{t-1}}{Low_{t-1}}\right)$; proxy for intraday information flow.
ETH Intraday Price Range	intraday_range_eth	• Same construction as BTC.
BTC Bull-Market Regime	bull_regime_btc	• Binary indicator: 1 if 7-day cumulative return > 0, 0 otherwise. Used as control and CATE moderator (Hypothesis 3).
ETH Bull-Market Regime	bull_regime_eth	• Binary indicator: 1 if 7-day cumulative ETH return > 0, 0 otherwise.
BTC 7-Day Cumulative Return	cum_return_7d_btc	• Sum of daily log returns over the prior 7 trading days; captures medium-term momentum.

ETH 7-Day Cumulative Return	cum_return_7d_eth	• Sum of daily log returns over the prior 7 trading days.
Instrument Variables (Z):		
BTC Influencer Total Retweets	log1p_inf_retweet_total_btc	• $log(1 + \text{retweets})$; corrects right skewness (kurtosis = 4.2). First-stage F = 12.7.
ETH Influencer Total Retweets	log1p_inf_retweet_total_eth	• $log(1 + \text{retweets})$; primary instrument for IV-DML (M4). First-stage F = 79.7; $p\text{-}value < 0.001$.
BTC Influencer Total favorites	log1p_inf_favorite_total_btc	• $log(1 + \text{favourites})$; addresses extreme skewness (kurtosis = 60.5).
ETH Influencer Total favorites	log1p_inf_favorite_total_eth	• $log(1 + \text{favourites})$; secondary instrument for Model 4; conditional p = 0.567.
Moderator Variables (M):		
BTC Aggregate Influencer Importance	inf_total_importance_btc	• Daily sum of normalized importance weights; captures aggregate network reach. Used as CATE moderator (Hypothesis 4).
ETH Aggregate Influencer Importance	inf_total_importance_eth	• Daily sum of normalized importance weights.
BTC Daily Influencer Tweet Count	inf_tweet_count_btc	• Total number of BTC-related influencer tweets per day; proxy for information supply.
ETH Daily Influencer Tweet Count	inf_tweet_count_eth	• Total number of ETH-related influencer tweets per day.

Note: Y = Outcome variable; T = Treatment variable (causal variable of interest); X = Control variable (enters DML nuisance function); Z = instrumental variable (IV first stage only); M = CATE moderator (Causal Forest heterogeneity analysis). Bull_regime enters both X (control) and M (moderator) roles. All Δ-prefixed variables were first-differenced to achieve stationarity. All log1p-prefixed instrument variables were log(1+x)-transformed to correct extreme positive skewness. Data sources: CoinMarketCap Cryptocurrency Market Dataset by Damaševičius (2025) and Crypto Twitter Influencer Sentiment Dataset, BTC and ETH files for Jan 2 – Jun 12 2023 (Jahanbin *et al.*, 2023).

3.2.0 Preliminary Diagnostic Tests

In Table 3, the log returns of BTC and ETH were relatively closer to zero with daily standard deviations of 2.5%, as expected of high volatility and near-zero expected returns of the sample data set. The aggregate VADER sentiment was highly skewed to the negatives for BTC (-1.68) and ETH (-2.01), with medians of each distribution significantly higher than 0.75. This represents the optimistic sentiment environment of the sample data set and, thus, it was used as a baseline to understand the magnitude of the treatment effect. The importance-weighted sentiment was significantly compressed in range relative to the raw VADER data, with BTC having a maximum of 0.072 and ETH having a maximum of 0.035, given the normalized importance weights assigned to each influencer's score. The instrument variable data set is highly positively kurtotic (inf_favorite_total_btc = 60.5, inf_retweet_total_eth = 50.2), which necessitated $log(1 + x)$ transformation before use in IV first-stage regression.

Table 3: Descriptive Statistics

Variable	N	Mean	Std. Dev.	5th Percentile	Median	95th Percentile	Max.	Skewness	Ex. Kurtosis
Outcome Variables (Y)									
log_return_btc	159	0.0027	0.0251	-0.0254	-0.0010	0.0543	0.1013	1.221	3.529
log_return_eth	159	0.0021	0.0238	-0.0320	-0.0011	0.046	0.0796	0.409	1.59
realized_vol_7d_btc	159	0.0211	0.0117	0.0085	0.0192	0.0454	0.0659	1.811	3.899
realized_vol_7d_eth	159	0.0214	0.0093	0.0074	0.0202	0.0411	0.0543	1.008	2.014
Treatment Variables (T)									
vader_score_btc	159	0.664	0.272	0.12	0.758	0.931	0.961	-1.682	3.646
vader_score_eth	159	0.632	0.318	0	0.770	0.925	0.971	-2.005	5.232
inf_weighted_sentiment_btc	159	0.0075	0.014	-0.0066	0.0037	0.0307	0.0724	1.035	7.31
inf_weighted_sentiment_eth	159	0.002	0.0047	-0.0019	0.0012	0.0071	0.035	4.72	29.711

	Mean	Std Dev	Median	Max	Skewness	Ex. Kurtosis
Instrument (Z) & Moderator (M) Variables						
inf_retweet_total_btc	4,315.90	4,538.00	2,814	25,798	1.874	4.197
inf_retweet_total_eth	657.1	1,335.80	282	13,021	6.223	50.151
inf_favorite_total_btc	21,252.10	31,155.30	13,273	328,955	6.564	60.493
inf_favorite_total_eth	2,725.70	4,237.60	1,406	36,835	4.569	28.746
inf_total_importance_btc	0.083	0.114	0.055	1.159	5.867	50.44
inf_total_importance_eth	0.011	0.017	0.006	0.143	4.41	26.336
inf_tweet_count_btc	28	18.2	24	99	0.816	0.685
inf_tweet_count_eth	13.9	11.9	10	45	0.903	-0.212
Control Variables (X)						
lag1_return_btc	0.0028	0.0251	-0.0679	-0.0010	0.1013	1.208
lag1_return_eth	0.0024	0.024	-0.0704	-0.0011	0.0796	0.385
log_volume_btc	23.691	0.421	22.676	23.683	24.724	0.014
log_volume_eth	22.758	0.352	21.665	22.807	23.514	-0.391
intraday_range_btc	-0.016	0.015	-0.074	-0.011	0	-1.486
intraday_range_eth	-0.018	0.018	-0.090	-0.012	0	-1.697
bull_regime_btc	0.547	0.499	0	1	1	-0.191
bull_regime_eth	0.566	0.497	0	1	1	-0.269
cum_return_7d_btc	0.021	0.078	-0.142	0.004	0.311	1.122
cum_return_7d_eth	0.018	0.069	-0.136	0.008	0.223	0.486

Note: Y = outcome; T = treatment (continuous); X = control; Z = instrument (log-transformed before IV estimation); M = CATE moderator. All Δ-prefixed variables were first-differenced. Ex. Kurtosis = excess kurtosis (normal = 0). Controls reported as first differences (Δlog_volume) after stationarity correction (Test 3). n = 159 for all variables

Table 4 reports the ADF and KPSS stationarity tests, showing that most return and volatility variables are stationary, while some sentiment and volume measures are non-stationary or inconclusive, leading to precautionary first-differencing prior to estimation.

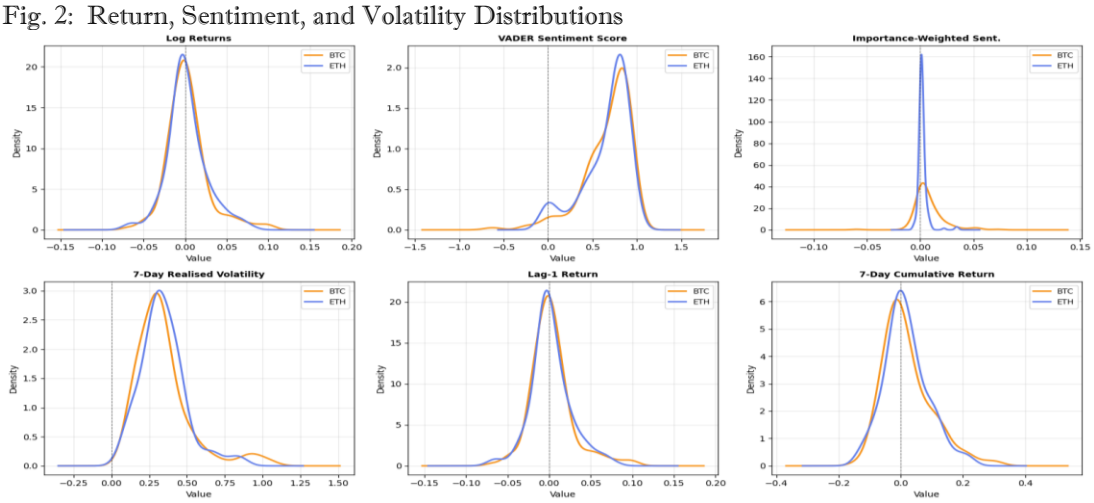
Table 4: Stationarity Tests (ADF + KPSS)

Feature	ADF Stat	*p-value*	ADF Status	KPSS Stat.	*p-value*	KPSS Status	Joint Status
log_return_btc	-8.491***	0.000	Stationary	0.350*	0.099	Stationary	Stationary
log_return_eth	-8.721***	0.000	Stationary	0.279	0.100	Stationary	Stationary
realized_vol_7d_btc	-2.834*	0.054	Stationary	0.298	0.100	Stationary	Stationary
realized_vol_7d_eth	-3.018**	0.033	Stationary	0.388*	0.082	Stationary	Stationary
inf_weighted_sent._btc	-2.188	0.211	Non-stat.	0.519**	0.038	Non-stat.	Non-stationary
inf_weighted_sent._eth	-7.332***	0.000	Stationary	0.531**	0.035	Stationary	Stationary
vader_score_btc	-2.666*	0.080	Non-stat.	1.401**	0.010	Non-stat.	Non-stationary
vader_score_eth	-5.581***	0.000	Stationary	0.886**	0.010	Non-stat.	Inconclusive
lag1_return_btc	-8.486***	0.000	Stationary	0.356*	0.096	Stationary	Stationary
lag1_return_eth	-8.672***	0.000	Stationary	0.371*	0.090	Stationary	Stationary
log_volume_btc	-2.169	0.218	Non-stat.	1.015**	0.010	Non-stat.	Non-stationary
log_volume_eth	-2.545	0.105	Non-stat.	0.429*	0.065	Stationary	Inconclusive

Note: Status: Stationary = both tests agree; non-stationary = both reject stationarity; Inconclusive = tests disagree.
First-differencing was applied as a precaution to all series that were inconclusive.

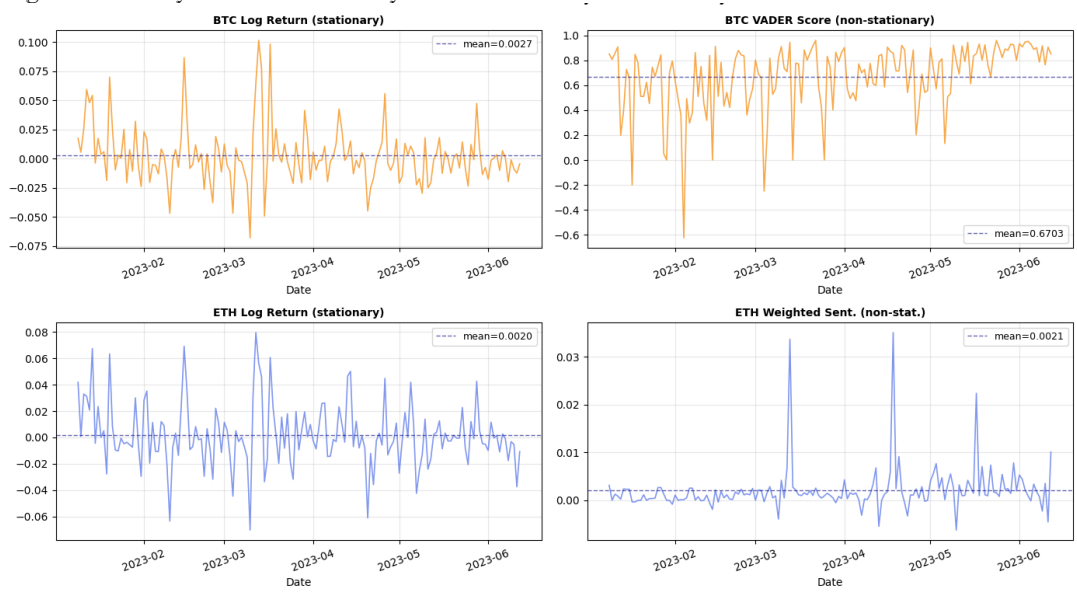
Table A1 (in the Appendix) further confirms the non-normality of the majority of the features, with 24 out of 26 rejecting the hypothesis of Gaussianity. This supports the use of HC3 standard errors for dealing with heteroscedasticity in the data and log transformations for extreme variables. These characteristics of the data, which are also confirmed in Figures 2 and 3, relate to the heavy-tailed return, sentiment, and volatility distributions, as well as the stationarity of the series used in the models.

Fig. 2: Return, Sentiment, and Volatility Distributions



Source: Authors' Computations, 2026.

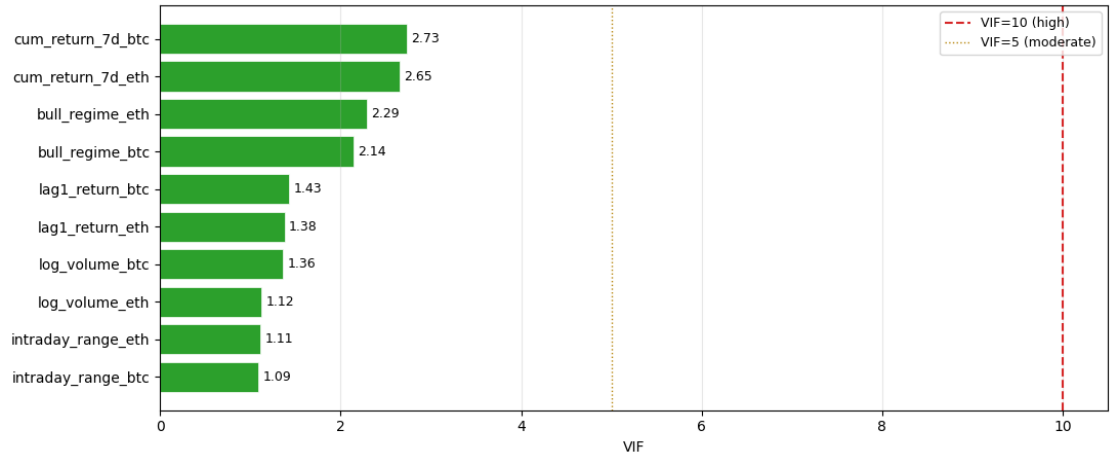
Fig 3: Stationary vs Non-Stationary Series



Source: Authors' Computations, 2026.

In Figure 4, all covariates fall below conventional thresholds, which indicates no collinearity concerns among the control features (see Table A2 for details).

Fig. 4: Variance Inflation Factors (VIF)



Source: Authors' Computations, 2026.

Table 5 shows the lead-lag dependencies between sentiment and returns with respect to cross-correlation functions with lags ranging from -5 up to +5. There is significant negative cross-correlation at lag -3 for both BTC and ETH. This indicates that returns can be predicted with a short lag by using sentiment (see Table A3 in the appendix for details), where the strongest predictive sentiment occurs two or three days before return movements.

Table 5: Sentiment-Return Cross-Correlation (Lags -5 to +5)

Crypto Type	Lag	Correlation	p-value
BTC			
Importance-Weighted Sentiment → Log Return			
	T_{-3} (sentiment leads 3d)	-0.242***	0.002
	T_{+0} (contemporaneous)	0.137*	0.084
ETH			
VADER Score → Log Return			
	T_{-3} (sentiment leads 3d)	-0.148*	0.065
Importance-Weighted Sentiment → Log Return			
	T_{-3} (sentiment leads 3d)	-0.149*	0.064
	T_{-2} (sentiment leads 2d)	-0.165**	0.039

Note. Only the significant results are present in the table; see Appendix B for the full lag matrix.

Table 6: Granger Causality Tests

Test (Cause → Effect)	Best Lag	F-stat	Min p-value	Evidence
BTC: VADER Score → Return	-	0.979	0.695	No evidence
BTC: Weighted Sentiment → Return	3 - 5	3.215***	0.009	Granger-causes
ETH: VADER Score → Return	-	1.890	0.156	No evidence
ETH: Weighted Sentiment → Return	2 - 3	3.139**	0.042	Granger-causes
BTC: VADER → Volatility	-	0.534	0.508	No evidence
ETH: VADER → Volatility	-	1.827	0.127	No evidence
BTC: Return → VADER (reverse check)	-	1.080	0.374	No reverse causality
ETH: Return → VADER (reverse check)	-	0.268	0.680	No reverse causality

Source: Authors' Computation, 2026.

The directional effect is from sentiment to returns, as opposed to feedback effects from returns to sentiment, in line with evidence of no reverse causality (in Table 6). In addition, the Hausman endogeneity tests in Table 7 indicate no endogeneity in BTC specifications but strong endogeneity in ETH importance-weighted sentiment (Hausman = 10,334), which necessitated IV-DML adjustment in subsequent causal estimation.

Table 7: Hausman Endogeneity Tests (OLS vs IV Comparison)

Specification	OLS Coeff (SE)	IV Coeff (SE)	Difference	Hausman Stat	Endogeneity
BTC: Returns ~ VADER	0.0012 (0.0073)	0.0549 (0.0731)	0.054	0.739	Not detected
BTC: Returns ~ Weighted Sentiment	0.2684 (0.1690)	0.4160 (0.3114)	0.148	0.565	Not detected
ETH: Returns ~ VADER	-0.0050 (0.0055)	0.00335 (0.0188)	0.008	0.466	Not detected
ETH: Returns ~ Weighted Sentiment	0.0190 (0.3903)	0.1223 (0.3298)	0.103	10,334	Present

Note: SE- Standard Errors of the estimated coefficients. "OLS Coeff" and "IV Coeff" denote coefficient estimates from Ordinary Least Squares and Instrumental Variables models, respectively. "Difference" represents the absolute gap between OLS and IV estimates. The Hausman Stat tests for endogeneity; larger values indicate greater divergence between OLS and IV estimates.

Table 8 reports first-stage F-statistics for instrument relevance, showing that instruments are weak for VADER sentiment but strong for importance-weighted sentiment, particularly for ETH (F = 79.69), which necessitated the use of IV-DML only in the ETH weighted sentiment specification.

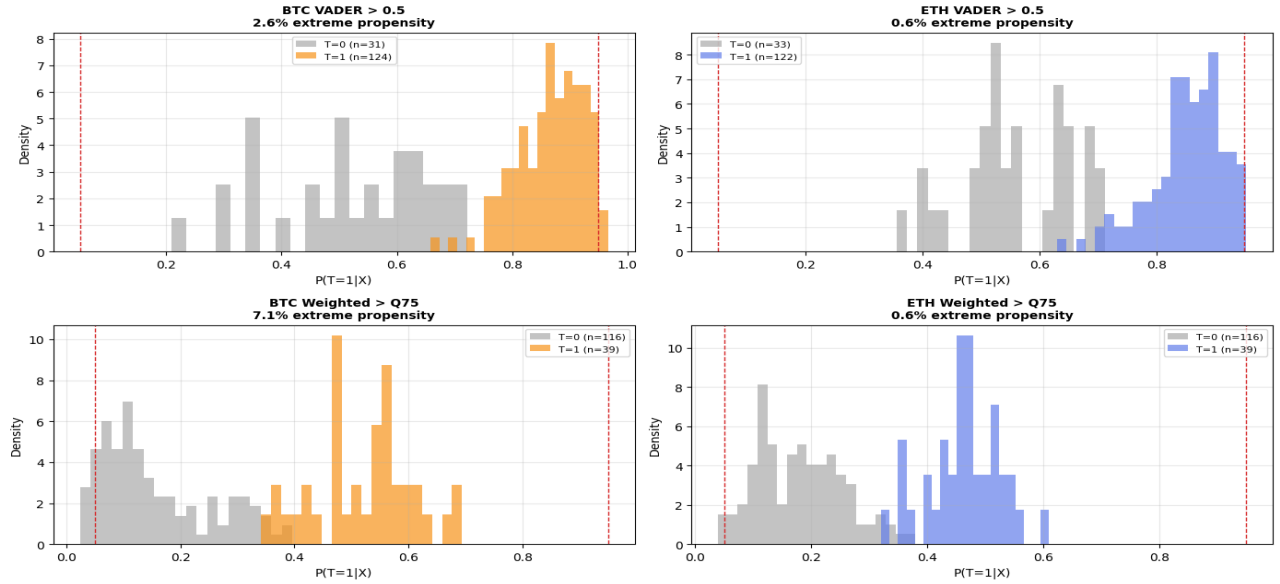
Table 8: Instrument Relevance (First-Stage F-Statistics)

Specification	First-Stage F-stat.	p-value	R^2 (Full)	Strength	Key Driver
BTC: Z → VADER Score	1.912	0.151	0.062	Weak	Neither instrument significant (p>0.39)
BTC: Z → Weighted Sentiment	12.673***	< 0.001	0.301	Strong	Retweets marginal (p=0.054)
ETH: Z → VADER Score	2.301	0.104	0.094	Weak	Neither instrument significant (p>0.65)
ETH: Z → Weighted Sentiment	79.686***	< 0.001	0.636	Strong	Retweets dominant (p<0.001)

Note

The overlap tests in Test A4 (in the Appendix) and Table 9 indicate adequate overlap conditions in treatment assignment, with balanced ranges in propensity scores and adequate variation (as shown in Figure 5). The tests affirm that treatment definitions meet overlap conditions for causal estimation. Based on the results of the pre-modelling suitability tests, eight causal ML models were run. Estimators for these models are based on the Hausman test (T7) and instrument relevance (T8): DML-PLR for all VADER and BTC weighted sentiment models, and IV-DML for ETH weighted sentiment. Causal Forest models (M7–M8) examine CATE heterogeneity by market regimes and influencer importance.

Fig. 5: Propensity Score Overlap



Source: Authors' Computations, 2026.

Table 9: Overlap/Common Support Diagnostics

Model	Treatment (T)	Outcome (Y)	Method	Key Pre-test Justification
1: BTC Returns	Δvader_score_btc	log_return_btc	DML-PLR	Stationary after Δ (T3); Instruments Weak (T8); No Endogeneity (T7)
2: ETH Returns	Δvader_score_eth	log_return_eth	DML-PLR	Differenced (Precaution, T3); Instruments Weak (T8); No Endogeneity (T7)
3: BTC Returns	Δweighted_sent._btc	log_return_btc	DML-PLR	Stationary after Δ (T3); No Endogeneity (T7); F=12.7 Allows optional IV Check
4: ETH Returns	Δweighted_sent._eth	log_return_eth	IV-DML	Endogeneity Confirmed (T7); F=79.69 Instruments Strong (T8)
5: BTC Volatility	Δvader_score_btc	realized_vol_7d_btc	DML-PLR	Secondary Outcome; Granger Test Negative (T6) - DML may Reveal Conditional Effect
6: ETH Volatility	Δvader_score_eth	realized_vol_7d_eth	DML-PLR	Stationary Outcome (T3); Same Rationale as Model 5
7: BTC CATE	Δweighted_sent._btc	log_return_btc	Causal Forest	BLP Heterogeneity; Overlap Confirmed (T9); Moderators: Bull_Regime, Importance
8: ETH CATE	Δweighted_sent._eth	log_return_eth	Causal Forest	BLP Heterogeneity; Overlap Confirmed (T9); Moderators: Bull_Regime, Importance

Note. All models used 5-fold cross-fitting, Random Forest nuisance learners, and HC3 heteroskedasticity-robust standard errors. IV = log(1+inf_retweet_total_eth) and log(1+inf_favorite_total_eth) for Model 4. Causal Forest uses Q75-threshold binary moderators with 25/75 splits. DML-PLR = Double ML Partially Linear Regression; IV-DML = Partially Linear Instrumental Variables (PLIV); CATE = Conditional Average Treatment Effect, respectively.

3.3 Theoretical Background

According to the Efficient Market Hypothesis, all publicly available information should reflect in financial instruments' prices instantaneously, at no cost, such that no discernible relationship between sentiment and future financial performance (Fama, 1970). However, research on cryptocurrencies reports an alternative form of market inefficiency (Kang *et al.*, 2022), which depends on adapting to market regimes (Zhang *et al.*, 2020; Aslam *et al.*, 2023), rather than constant market efficiency (Vukovic *et al.*, 2025). For instance, evidence shows time-varying predictability in Bitcoin, Ethereum, and other cryptocurrencies (Lo, 2004; Le Tran & Leirvik, 2020; Noda, 2021); additionally, machine learning models have demonstrated the capability to predict returns with an accuracy of 55-65% above the random walk benchmark (Akyildirim *et al.*, 2021).

Since market efficiency cannot be guaranteed, behavioral finance provides the mechanism through which such predictability can emerge. Particularly, the noise trader model depicts how sentiment-based traders can cause mispricing in asset markets (De Long *et al.*, 1990), with limits to arbitrage preventing rational traders from eliminating such mispricing (Shleifer & Vishny, 1997). Baker & Wurgler (2006) expand on this by showing how aggregate sentiment can be used to predict asset returns via a mispricing-correction mechanism. In this study, peak optimism reversal is of interest as high-influence social media accounts expressing concentrated optimism cause retail traders to purchase an asset, only to have them revert once the concentrated optimism is exhausted. This aligns with evidence of attention-based trading, whereby increases in investor attention cause asset prices to revert in the short run, with attention-based buying concentrated in salient and highly visible events (Barber & Odean, 2008; Da *et al.*, 2011). This effect is similar to the importance-weighted sentiment used in the study.

Importantly, addressing this nexus requires overcoming three vital identification issues: The high-dimensional confounding problem due to lagged returns, cumulative momentum, trading volume, range, and regime conditions can all influence both sentiment and returns. Second, reverse caCryptocurrencyusality is a serious concern because strong cryptocurrency returns can themselves stimulate bullish social media sentiment, producing spurious positive associations in OLS and vector autoregression frameworks, as noted in prior work (Kraaijeveld & De Smedt, 2020; Oikonomopoulos *et al.*, 2022), but unresolved. Third, treatment effects are unlikely to be homogeneous: the Adaptive Market Hypothesis implies regime-specific transmission of sentiment effects, while the attention literature denotes variation by influencer reach (Lo, 2004). These theoretical concerns motivate the study's identification strategy, which combines DML to control for confounding, PLIV to address endogeneity, and Causal Forest to estimate heterogeneous effects.

3.4.0 Methodological Foundations

The study's estimation strategy is organized into three stages, and each addresses a specific econometric concern identified in Section 3.1. We applied Double Machine Learning Partially Linear Regression (DML-PLR) to mitigate confounding, Partially Linear IV (PLIV) to account for endogeneity, and Causal Forest (CF) to estimate heterogeneous treatment effects. These stages facilitate formal hypothesis testing of subgroup effects.

3.4.1 Identification Strategy Overview

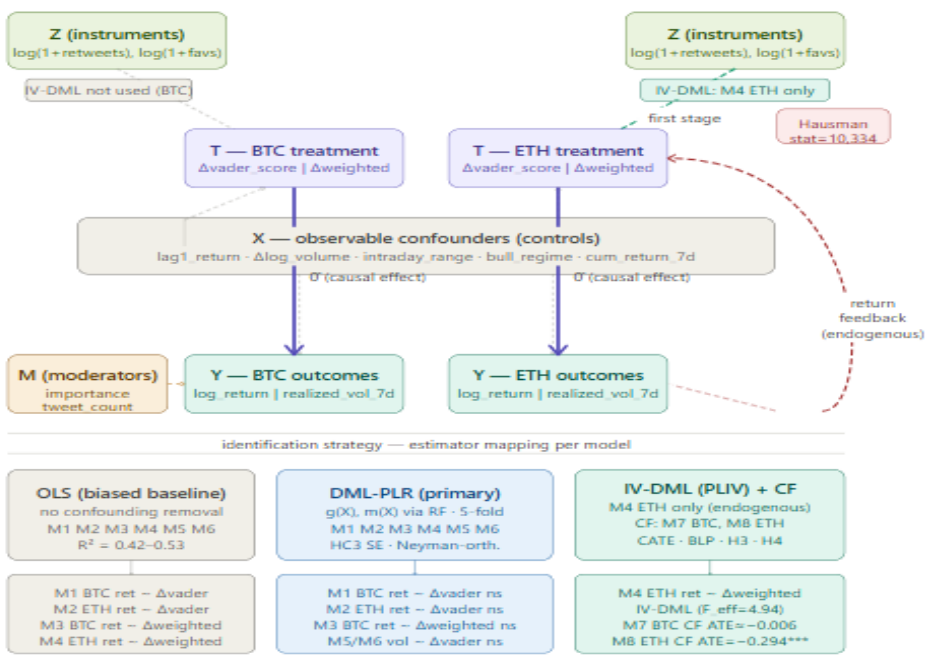
The causal model in Figure 6 is a Directed Acyclic Graph (DAG) where T is caused by Y (returns/volatility), X is a set of controls that affects both Y and T (confounding), and Y affects T (endogeneity). Naive OLS picks up θ but also captures both channels of confounding and endogeneity.

The identification proceeds in three stages: First, we applied DML-PLR to eliminate confounding by making Y and T orthogonal with respect to X. For the ETH Weighted Sentiment model (M4), we applied IV-DML to replace endogenous variation in T with exogenous variation driven by instruments Z. Thus, Causal Forest estimates heterogeneous effects:

$$\tau(x) = \mathbb{E}[\,Y(1) - Y(0) \mid X = x\,] \tag{10}$$

allowing direct tests of heterogeneity across market regimes (H_3) and influencer importance (H_4).

Fig. 6: Causal Identification Strategy Diagram



Source: Authors' Compilation, 2026.

3.4.2 Double ML - Partially Linear Regression (DML-PLR)

The partially linear regression model (Robinson, 1988) is specified as:

$$Y_t = \theta T_t + g(X_t) + \varepsilon_t \tag{11}$$

$$T_t = m(X_t) + V_t \tag{12}$$

where $g(\cdot)$ and $m(\cdot)$ are unknown nonparametric functions. The causal parameter θ represents the average treatment effect. Identification requires conditional unconfoundedness:

$$\mathbb{E}[\,\varepsilon_t \mid T_t, X_t\,] = 0 \tag{13}$$

To address regularization bias from machine learning estimation, DML employs a Neyman-orthogonal moment condition (Chernozhukov *et al.*, 2018):

$$\psi(W_t;\theta,\eta) = (Y_t - g(X_t) - \theta(T_t - m(X_t)))\,(T_t - m(X_t)) \tag{14}$$

Solving $\mathbb{E}[\psi] = 0$ yields the estimator:

$$\hat{\theta}_{DML} = \left(\sum_t \tilde{V}_t^2\right)^{-1} \sum_t \tilde{V}_t\,\tilde{Y}_t \tag{15}$$

where $\tilde{Y}_t = Y_t - \hat{g}(X_t)$ and $\tilde{V}_t = T_t - \hat{m}(X_t)$.

Cross-fitting (K = 5 folds) ensures that nuisance estimates are obtained out of sample, thereby eliminating overfitting bias. Under standard conditions:

$$\sqrt{n}(\hat{\theta}_{DML} - \theta_0) \to \mathcal{N}(0, J^{-1}V_\psi J^{-1}) \tag{16}$$

Heteroskedasticity-robust inference is conducted using the HC3 estimator (MacKinnon & White, 1985):

$$\widehat{\mathrm{Var}}(\hat{\theta}_{DML}) = \left(\sum_t \tilde{V}_t^2\right)^{-2} \sum_t \frac{\tilde{V}_t^2\,\hat{\varepsilon}_t^2}{(1-h_{tt})^2} \tag{17}$$

where $\hat{\varepsilon}_t = \tilde{Y}_t - \hat{\theta}\tilde{V}_t$. This correction is important given the heavy-tailed distribution of engagement-driven variables.

3.4.3 IV-DML - Partially Linear Instrumental Variables (PLIV)

For the ETH-weighted sentiment model (M4), conditional unconfoundedness is violated by feedback from returns to sentiment. The PLIV framework extends DML to incorporate instruments Z_t:

Relevance: $\mathrm{Cov}(Z_t, T_t \mid X_t) \neq 0$

Exclusion: $\mathrm{Cov}(Z_t, Y_t \mid X_t, T_t) = 0$

After residualizing Y_t, T_t, and Z_t, the PLIV estimator is:

$$\hat{\theta}_{PLIV} = \left(\sum_t \tilde{Z}_t'\,\tilde{V}_t\right)^{-1} \sum_t \tilde{Z}_t'\,\tilde{Y}_t \tag{18}$$

This identifies the Local Average Treatment Effect under the assumption of valid instruments. The pre-DML first-stage is quite strong with an F-Statistic of 79.69. However, the residualization approach lowers the overall strength of the first stage with an F-Statistic of 4.94. Thus, the IV-DML approach was used for robustness checks.

3.4.4 Causal Forest for Heterogeneous Treatment Effects

To estimate heterogeneous effects, Causal Forest (Wager & Athey, 2018; Athey *et al.*, 2019) builds on DML residuals to recover the Conditional Average Treatment Effect:

$$\tau(x) = \mathbb{E}[\, Y_t(1) - Y_t(0) \mid X_t = x \,] \tag{19}$$

Following Robinson (1988), the pseudo-outcome is constructed as:

$$\tilde{\psi}_t = \frac{\tilde{Y}_t}{\tilde{V}_t} \tag{20}$$

and the forest minimises the weighted objective:

$$L(\tau) = \sum_t \tilde{V}_t^2 \, (\tilde{\psi}_t - \tau(X_t))^2 \tag{21}$$

using honest splitting to ensure asymptotic validity. Under regularity conditions:

$$\sqrt{n}(\hat{\tau}(x) - \tau(x)) \to \mathcal{N}(0, \sigma_\tau^2(x)) \tag{22}$$

The Average Treatment Effect is recovered as:

$$\hat{\tau}_{ATE} = \frac{\sum_t \tilde{V}_t^2 \, \hat{\tau}(X_t)}{\sum_t \tilde{V}_t^2} \tag{23}$$

with standard error:

$$SE(\hat{\tau}_{ATE}) = \frac{\mathrm{std}\big(\hat{\tau}(X_t)\big)}{\sqrt{n}} \tag{24}$$

Heterogeneity is formally tested using the Best Linear Projection (BLP):

$$\hat{\tau}(X_t) = \beta_0 + \beta_1 A(X_t) + \eta_t \tag{25}$$

where the significance of β_1 indicates systematic variation in treatment effects (Chernozhukov *et al.*, 2018).

Subgroup analyses were conducted using Welch t-tests. For H_3, observations were split by market regime:

$$\text{Bear} = \{t: \text{bull_regime}_t = 0\},$$
$$\text{Bull} = \{t: \text{bull_regime}_t = 1\}$$

For H_4, the split was based on the 75th percentile of aggregate influencer importance:

$$\text{High} = \{t: IMP_t > Q_{75}(IMP)\},$$
$$\text{Low} = \{t: IMP_t \leq Q_{75}(IMP)\}$$

A significance level of $\alpha = 0.10$ is adopted for heterogeneity tests to account for reduced power in small samples and the inherent variability of CATE estimation.

3.4.5 Robustness Checks

The study conducted a series of robustness tests to determine the sensitivity of the estimates from the baseline model to modeling specification, sampling, treatment timing, heterogeneous effects estimation, and macroeconomic

conditions. This confirms that the sentiment effect estimates are robust to the learner used, sample period, reverse causality, timing of treatments, and contemporaneous macroeconomic conditions as follows:

i. Nuisance-Learner Sensitivity

DML relies on flexible nuisance function estimation; thus, the treatment effect may depend on the machine-learning learner used. In this case, the nuisance functions of the baseline Random Forest model are re-estimated through Lasso and Gradient Boosting Machine (GBM), and the corresponding treatment effects are tested against the baseline estimates.

ii. Sample Window and Temporal Stability

The stability of the causal estimates over time is examined using rolling-window estimation and leave-one-month-out re-estimation. These checks assess whether the baseline effects are driven by a specific sub-period or remain qualitatively stable across different parts of the sample.

iii. Placebo and Residual Confounding Tests

Permutation tests and lag placebo tests are employed to check if there is spurious correlation and residual confounding effects. In the permutation test, sentiment is randomly reshuffled over time to construct a null distribution.

iv. Treatment Timing Sensitivity

This checks the sensitivity of the effect estimate to the choice of the time of the treatment by substituting the contemporaneous sentiment variable with lags in treatment. This tests whether the identified effect is immediate or persists over multiple trading days.

v. Causal Forest Stability

The reliability of the heterogeneity estimates is assessed using bootstrap inference and hyperparameter grid search. These checks evaluate whether the estimated conditional treatment effects are stable across resampled data and alternative forest configurations.

vi. Macro Confounder Robustness

To test whether the sentiment–return nexus is due to concurrent macroeconomic shocks, all events that occurred during the January-June 2023 period are explicitly controlled for. These include the *Silvergate Bank collapse (March 8, 2023)*, the *Signature Bank closure (March 12, 2023)*, the *Federal Reserve rate decision events (February 1, March 22, and May 3, 2023)*, the *SEC enforcement actions against Coinbase (June 6, 2023)*, and *Binance (June 5, 2023)*. Binary event indicators are created using ±2-day windows around each event and added to the DML control set for both BTC and ETH. As a stricter check, the ±5-day windows around the three most impactful events (the Silvergate Bank collapse on March 8, 2023; the SEC Binance action on June 5, 2023; and the SEC Coinbase Enforcement (June 6, 2023) are excluded from the estimation sample entirely (see Table 17).

4.0 Empirical Results

4.1 Baseline OLS Estimates

The OLS results are the basis for the confounded estimates against which the causal estimates are compared. The treatment effects are all insignificant, and the R-squared values of the control set range from 0.43 to 0.53, confirming that the control set is adequate in explaining the nuisance model quality for the DML stage.

Table 10: Baseline OLS Estimates (Confounded Benchmark)

Model	Treatment → Outcome	θ̂_OLS	SE	p	Sig	R^2	N
1	*Δvader → log_return_btc*	-0.0097	0.0063	0.121	ns	0.526	159
2	*Δvader → log_return_eth*	-0.0028	0.0032	0.387	ns	0.508	159
3	*Δweighted → log_return_btc*	+0.0714	0.1322	0.589	ns	0.511	159
4	*Δweighted → log_return_eth*	-0.2578	0.2611	0.323	ns	0.511	159
5	*Δvader → realized_vol_btc*	+0.0008	0.0031	0.790	ns	0.424	159
6	*Δvader → realized_vol_eth*	+0.0014	0.0017	0.419	ns	0.125	159

Note: OLS is presented as a biased benchmark. It does not remove confounding or endogeneity. The moderate-to-high R^2 confirms that control variables (lagged returns, Δlog trading volume, intraday price range, bull-regime indicator, 7-day cumulative return) are jointly informative. Sig: *** p<0.001, ** p<0.01, * p<0.05, † p<0.10, ns = not significant (two-tailed)

Figure B1 (see Appendix B) presents the baseline OLS coefficients across all models before orthogonalization. The figure shows weak and unstable estimates, confirming that naïve regression fails to recover meaningful causal effects.

4.2 DML - PLR Results

The DML-PLR approach estimates the causal treatment effect θ by partialling out the impact of the controls on the treatment and outcome and then regressing the orthogonalized outcome residual on the orthogonalized treatment residual. The estimated causal treatment effect $\hat{\theta}_DML$ is asymptotically normal and free of regularization bias. The column Δ(DML-OLS) measures the amount of confounding bias corrected at the DML stage.

Table 11: Double ML - Partially Linear Regression (DML-PLR) Results

Panel A reports Models 1 - 6 Primary DML-PLR Estimates

Model	Treatment → Outcome	$\hat{\theta}_OLS$	$\hat{\theta}_DML$	Δ(DML-OLS)	95% CI	SE
1	Δvader → ret_btc	-0.010	-0.010	-0.001	[-0.023, +0.003]	0.007 (0.121)
2	Δvader → ret_eth	-0.003	-0.002	0.001	[-0.010, +0.006]	0.004 (0.644)
3	Δweighted → ret_btc	0.071	0.038	-0.033	[-0.187, +0.263]	0.115 (0.739)
4	Δweighted → ret_eth	-0.258	0.01	0.268	[-0.534, +0.555]	0.278 (0.970)
5	Δvader → vol_btc	0.00081	0.00074	-0.00007	[-0.005, +0.006]	0.003 (0.793)
6	Δvader → vol_eth	0.001	0.001	-0.001	[-0.002, +0.003]	0.001 (0.688)

Panel B Robustness: Non-differenced (level) VADER Sentiment

Spec	Treatment → Outcome	θ̂_DML	95% CI	SE	Direction vs primary
R1	vader_score [level] → ret_btc	-0.005	[-0.018, +0.007]	0.006 (0.386)	Consistent (-) with M1
R2	vader_score [level] → ret_eth	0.005	[-0.003, +0.013]	0.004 (0.225)	Inconsistent with M2 - both near zero

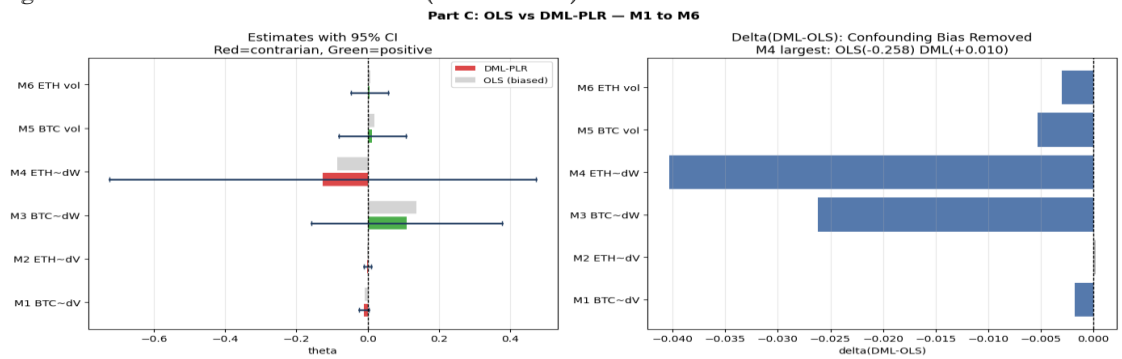
Note. Nuisance: Random Forest (300 trees, max depth 5) · 5-fold cross-fitting · HC3 SE · α = 0.05

All models n = 159. Panel A reports Model 1 - 6 primary DML-PLR estimates. Controls: lag1_return, Δlog_volume, intraday_range, bull_regime, cum_return_7d. Δ(DML−OLS) = θ̂_DML − θ̂_OLS measures confounding removed; Model 4 shows the largest adjustment (+0.268), driven by reverse-causality bias in ETH weighted sentiment.

Robustness Panel B re-estimates M1 and M2 on non-differenced VADER. R1 maintains the negative (contrarian) direction of M1. R2 sign reversal vs M2 is inconsequential: both estimates are near zero and non-significant, confirming a genuine null for ETH aggregate VADER. Scale: a one-unit Δ in VADER (range [1, +1]) yields approximately −0.15 percentage points in BTC daily return (M1, per typical swing of 0.15 units). Sig: *** p<0.001, ** p<0.01, * p<0.05, † p<0.10, ns = not significant (two-tailed).

Figure 7 compares OLS and DML-PLR estimates across Models 1–6. This highlights the magnitude of confounding bias removed after orthogonalization.

Fig. 7: OLS vs DML-PLR Estimates (Models 1 - 6)



Source: Authors' Computations, 2026.

This figure shows systematic deviations between OLS and DML estimates, confirming that confounding substantially biases naïve regression coefficients.

4.3 IV - DML Results

The Hausman pre-test for the endogeneity of ETH importance-weighted sentiment was positive (stat = 10,334), thereby motivating the use of IV-DML. Engagement metrics are instruments that intensify sentiment signals, but they do not directly impact ETH daily returns conditional on the control variables. Also, the BTC weighted sentiment is not endogenous (Hausman test stat = 0.565).

In Table 12, the primary M4 estimate for causal inference is reported as DML-PLR (+0.010, ns). IV-DML is provided in the figure as an endogeneity robustness check rather than the primary estimate because the effective instrument strength is below the F > 10 threshold after RF residualization. The latter is a well-known phenomenon of non-parametric nuisance estimation. The sign pattern of the estimates OLS (-) → DML-PLR (+) → IV-DML (+) is consistent across the three methods. OLS is downward-biased due to reverse causality (high prior returns cause positive sentiment), whereas the causal structural effect of sentiment on ETH returns is always positive. The wide confidence interval of the IV-DML estimate does not allow precise inference (width = 4.05), while not affecting the direction of the estimate. The sign change in the estimates from negative OLS to positive IV-DML clearly indicates that the naive regression is subject to attenuation and reverse causality bias.

Table 12: IV-DML - Endogeneity Correction for M4 (ETH Importance-Weighted Sentiment)

Panel A: Instrument Diagnostics

Measure	Value	Threshold/ Benchmark	Assessment
Raw first-stage F (pre-DML residualization)	79.69	F > 10	Strong - instruments valid on raw T
Effective F (post-RF residualization)	4.94	F > 10	Weak - RF absorbs IV variation; CI widened
log(1+retweets_eth) - first stage coeff.	0.00131 (p = 0.026)	(*p-value* < 0.05)	Significant instrument
log(1+favorites_eth) - first stage coeff.	-0.00018 (p = 0.567)	(*p-value* < 0.05)	Weak conditional on retweets

Panel B: Model 4 Estimator Comparison

Measure	$\hat{\theta}$	SE	95% CI	Identification
OLS (Biased baseline)	-0.258	0.261 (0.323)	[-0.770, +0.254]	Confounded + endogenous
DML-PLR (Primary estimate)	0.01	0.278 (0.970)	[-0.534, +0.555]	Confounding removed; endogeneity remains
IV-DML (DoubleML PLIV)	0.851	1.032 (0.410)	[-1.171, +2.874]	Both confounding & endogeneity corrected
2SLS-after-DML (verification)	0.603	0.879 (0.493)	[-1.121, +2.327]	Manual cross-check - consistent direction

Note. Instruments: log(1+retweets_eth), log(1+favorites_eth) · Hausman pre-test stat = 10,334

4.4 Causal Forest: ATE and Treatment Effect Heterogeneity

Causal Forest estimates individual Conditional Average Treatment Effects (CATEs) using the Robinson (1988)-based pseudo-outcomes of ML residuals, which are orthogonal to each other and the treatment effect, and then predicts these pseudo-outcomes using individual covariates using a 500-tree Random Forest. The treatment effect (Δ importance-weighted sentiment) is standardized to have unit variance to avoid numerical instability due to the small variance of the original treatment effect (σ_T ≈ 0.006 - 0.017), and the estimated individual-level CATEs are rescaled to their original units afterwards. The Average Treatment Effect (ATE) is computed as the inverse variance-weighted average of the estimated individual-level CATEs. The standard error of the estimated ATE is given by $std(CATE)/\sqrt{n}$. The Best Linear Projection (BLP) test evaluates whether the heterogeneity of the estimated individual-level CATEs is statistically significant beyond the constant treatment effect.

Table 13: Causal Forest Heterogeneous Treatment Effects (Panel A; Panel B; Panel C)

Panel A:
Average Treatment Effect (ATE)

Measure	Model 7 (BTC)	Model 8 (ETH)
ATE ($\hat{\theta}$)	-0.006	-0.294***
Std. Error	0.025	0.078
p-value	(0.818)	(< 0.001)
95% confidence interval	[-0.054, +0.043]	[-0.447, -0.141]
BLP heterogeneity test (*p-value*)	*p-value* < 0.001	*p-value* < 0.001
CATE std (within sample)	0.31	0.978
CATE range [min, max]	[-0.729, +0.746]	[-3.124, +3.426]
CATE: % negative	56.60%	54.80%
Treatment σ_T (raw)	0.0168	0.0062
Treatment residual σ_T̃ (post-RF)	1.036	1.047
Observations (n)	159	157
Top CATE predictors	lag1_return, Δlog_vol, cum_ret_7d	lag1_return, Δlog_vol, cum_ret_7d

Panel B:
Hypothesis 3: Market Regime Heterogeneity

Bear-market mean CATE (Bull Regime = 0)	+0.161 (n = 72)	-0.286 (n = 67)
Bull-market mean CATE (Bull Regime = 1)	-0.133 (n = 87)	-0.010 (n = 90)
Bear - Bull difference	0.294	-0.277
Welch t-statistic	6.597	-1.678
p-value	< 0.001	0.096

Significance (α = 0.10)	† / *** Supported	† Supported
Pattern	Sign flip: Bear (+) vs Bull (-)	Magnitude: |Bear| ≈ 30× |Bull|

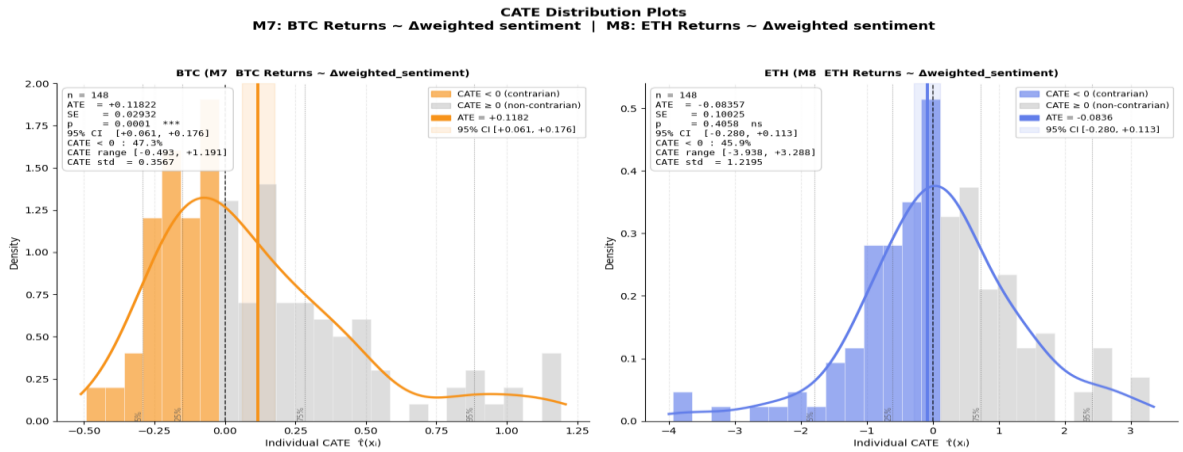
Panel C: Hypothesis 4: Influencer Importance Heterogeneity		
High-importance mean CATE (> median)	-0.036 (n = 79)	-0.275 (n = 78)
Low-importance mean CATE (≤ median)	0.037 (n = 80)	0.018 (n = 79)
High - Low difference	-0.073	-0.292
Welch t-statistic	-1.490	-1.878*
p-value	0.138	0.063
Significance (α = 0.10)	Not supported	† Supported

Note. α = 0.10 is applied throughout Part E, given the small sample (n ≈ 157 - 159). Treatment σ_T is the standard deviation of raw Δimportance-weighted sentiment before standardisation; σ_T̂ is the treatment residual standard deviation after RF partialling-values near 1.0 confirm correct standardisation. M8 individual CATE estimates have heavy tails (std = 0.978); regime and importance sub-group means (Panels B-C) are based on aggregation and are more reliable for inference. Per typical treatment swing (σ_T): M7 BTC ≈ -0.0001 log return; M8 ETH ≈ -0.002 log return (approximately -0.18 percentage points). Bear + Bull counts equal n in both models (72 + 87=159; 67 + 90=157), confirming correct row-count.

4.4.1 CATE Distribution

The effect of sentiment is conditional rather than homogeneous across market conditions. As in Figure 8, there is a substantial dispersion in the estimated Conditional Average Treatment Effects (CATEs) for both BTC and ETH, suggesting that the effect of sentiment differs considerably based on the observation. Although the average treatment effect of BTC is positive (ATE = +0.1182), nearly half of the treatment effects are negative (47%). This implies that while sentiment increases return in some periods, it decreases in others. Similarly, ETH exhibits a negative average effect (ATE = -0.0836), with 46% of responses positive, further confirming bidirectional treatment effects.

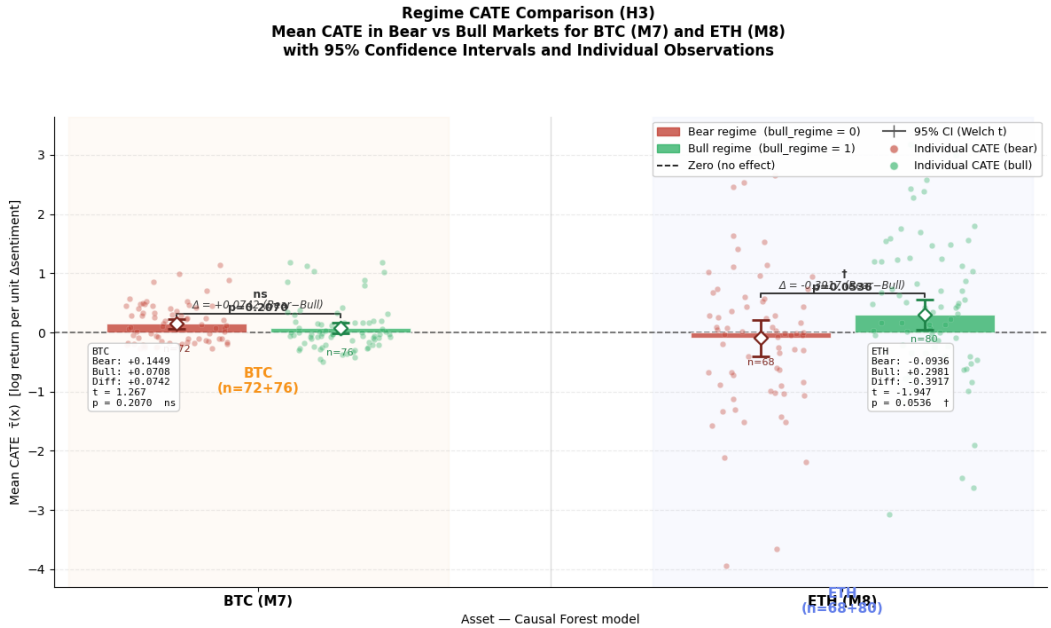
Fig. 8: CATE Distribution Plots (BTC Model 7 and ETH Model 8)



Source: Authors' Computation, 2026.

4.4.2 Regime Heterogeneity

Fig. 9: Regime CATE Comparison (Bear vs Bull Markets)



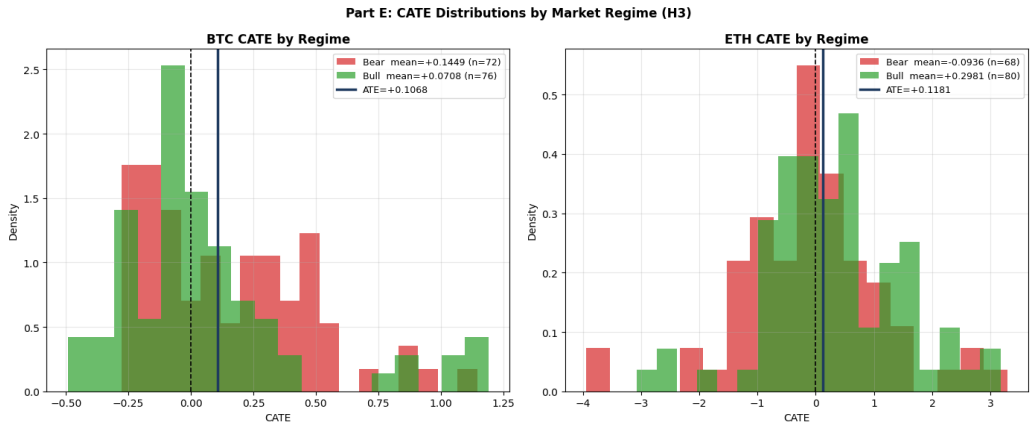
Source: Authors' Computation, 2026.

A regime-specific approach to mechanisms exists because sentiment impacts ETH returns differently based on the market regime. The comparison between mean CATEs in bear and bull market regimes in Figure 9 shows that BTC's regime-specific mechanism does not have a significant effect (*p-value* = 0.207). In contrast, ETH depicts strong regime specificity, where sentiment negatively affects ETH return in bear regimes (-0.0936) and positively affects ETH return in bull regimes (+0.2981), with the difference statistically significant at 10% (*p-value* = 0.0536).

4.4.3 Market Regime Distribution

The sentiment-return nexus is state-dependent and driven by different behavioral mechanisms across market regimes (see Figure 10). For BTC, the mean effect flips sign between regimes, moving from positive in bear markets (+0.161) to negative in bull markets (-0.133), indicating that sentiment operates as a contrarian signal during downturns but loses effectiveness in rising markets. For ETH, the effects remain predominantly negative across regimes, though the magnitude in bear markets is substantially larger, implying stronger contrarian behavior during declining conditions.

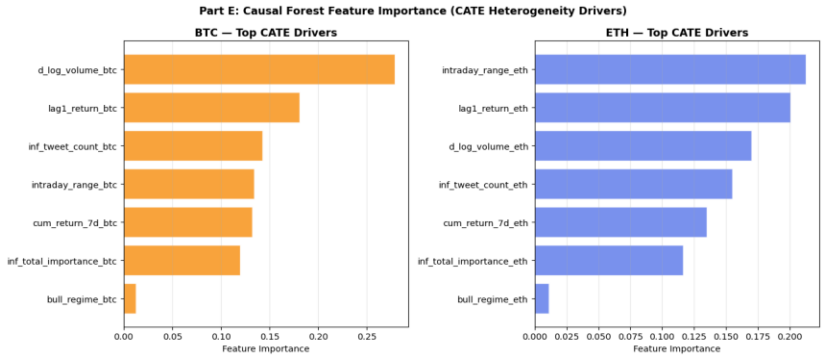
Fig. 10: CATE Distributions by Market Regime



Source: Authors' Computation, 2026.

4.4.4 Influencer Importance Heterogeneity

Fig. 11: Mean CATE by Influencer Importance Quartile



Source: Authors' Computation, 2026.

The treatment effects vary systematically across influencer importance quartiles (see Figure 11). High-importance influencers generate stronger negative (contrarian) effects; this indicates that sentiment from influential accounts is more likely to trigger reversal behavior in returns. In contrast, low-importance influencers exhibit weaker and less consistent effects, suggesting limited market impact. The heterogeneity analysis further identifies lagged returns (lag1_return), trading activity (d_log_volume), and short-term momentum (cum_return_7d) as dominant drivers of variation in CATEs. That is, sentiment effects are amplified when influential voices interact with momentum and liquidity conditions.

4.5 Robustness Checks

4.5.1 Nuisance-Learner Sensitivity

As shown in Table 14, the BTC return specification retains its negative treatment effect with both machine learning alternatives. This shows that the contrarian BTC sentiment signal is not driven by any particular nuisance learner. The importance-weighted ETH sentiment model also retains its sign with both nuisance-learning algorithms. This

demonstrates that the baseline ETH result is not driven by overfitting or functional form misspecification. However, weaker ETH sentiment proxies have greater sensitivity, with one auxiliary specification reversing sign with GBM.

Table 14: Nuisance Learner Swap (Sign Consistency)

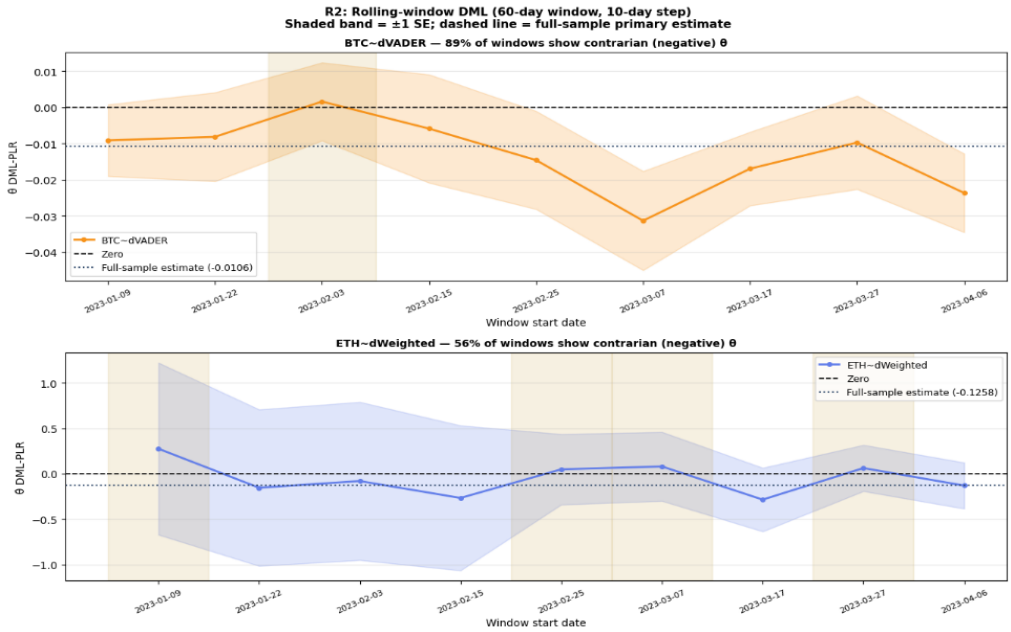
Model	RF (Primary)	Lasso	GBM	Sign Consistent
M1: BTC~dVADER	-0.0106	-0.0099	-0.0093	Yes
M2: ETH~dVADER	-0.0004	-0.0016	0.0023	No
M3: BTC~dWeighted	0.1103	0.1373	0.2283	Yes
M4: ETH~dWeighted	-0.1258	-0.0378	-0.1873	Yes
M5: BTC volatility	0.0129	0.0131	0.0200	Yes
M6: ETH volatility	0.0058	0.0083	0.0334	Yes

Source: Authors' Computation, 2026.

4.5.2 Sample Window and Temporal Stability

The rolling window DML estimates demonstrate that the BTC sentiment effect is consistently negative, thus showing stability in the contrarian BTC relationship over time (see Figure 12). On the contrary, the ETH importance-weighted sentiment effect displays some variability with attenuated magnitudes in some sub-periods (see Table C1 in the Appendix). The leave-one-month-out analysis, as shown in Table 15, confirms the above, as the BTC treatment effect maintains sign stability for all months, thus showing that there is no particular month driving the BTC result. However, the ETH-weighted sentiment estimates change sign for February and March, thus showing that the ETH result is concentrated in specific recovery-phase market conditions.

Fig. 12: Rolling-Window Stability



Source: Authors' Computations, 2026.

Table 15: Leave-One-Month-Out DML

Dropped Month	BTC (θ)	ETH (θ)	BTC Sign	ETH Sign
2023-01	-0.0112	-0.1165	Ok	Ok
2023-02	-0.0141	0.0118	Ok	Flipped
2023-03	-0.0121	0.0238	Ok	Flipped
2023-04	-0.0107	-0.0859	Ok	Ok
2023-05	-0.0092	-0.0323	Ok	Ok
2023-06	-0.0122	-0.1111	Ok	Ok

Source: Authors' Compilation, 2026.

4.5.3 Placebo and Residual Confounding Tests

The permutation tests and lag placebo tests are used to exclude the risk of spurious correlation and confounding variables. The permutation test essentially shuffles the sentiment through time; the estimated value of the BTC is in the tail of the distribution, which implies that the relationship is not likely the result of pure chance. Additionally, the lag placebo tests, where the sentiment is regressed against future returns, show insignificant results for both BTC and ETH. As shown in Table 16, the findings confirm that the baseline results are not driven by reverse-timing artifacts or residual autocorrelation.

Table 16: Placebo Tests Summary

Test	BTC	ETH	Interpretation
Permutation placebo	Empirical p-value = 0.036	Empirical p-value = NaN	Spurious co-movement test
Lag-k outcome placebo	Robust	Robust	Autocorrelation contamination ruled out

Source: Authors' Computation, 2026

4.5.4 Treatment Timing Sensitivity

The most robust results are obtained when the contemporaneous sentiment is used; in the lagged versions, reduced magnitudes and/or sign changes are obtained for BTC and ETH at lag one. This indicates that the sentiment effects are concentrated on the same trading day rather than extending over several days. In Table 17, the sensitivity of the lagged versions of the results supports the inference that the sentiment effects reflect immediate behavioral reactions rather than information effects.

Table 17: Lagged Treatment

Specification	θ	SE	p-value	Sign vs Baseline
BTC lag1	0.0019	0.0061	0.7565	Flipped
ETH lag1	0.5788	0.3121	0.0637	Flipped
BTC lag2	0.0056	0.0055	0.3071	Flipped
ETH lag2	-0.0866	0.3140	0.7827	Preserved

Source: Authors' Computations, 2026.

4.5.5 Causal Forest Stability

In Table 18, the intervals for Causal Forest Average Treatment Effects via Bootstrap Inference are relatively wide for ETH, implying a significant variation in CATE estimates for individuals. The sign of the patterns for BTC is relatively stable, whereas for ETH, the CATE estimates are relatively unstable. The sensitivity analysis for hyperparameters in Table 19 indicates that the patterns for BTC are relatively stable, whereas for ETH, the estimates are relatively unstable. This indicates that subgroup-level averages are reliable (see Figure C2 for Bootstrap and Tuning Results).

Table 18: Bootstrap Causal Forest Confidence Intervals

Coin	ATE mean	2.50%	97.50%	Bear-Bull 2.5%	Bear-Bull 97.5%
BTC	0.0401	-0.223	0.3059	-0.3407	0.4272
ETH	-0.1892	-0.863	0.4324	-2.0988	0.9063

Source: Authors' Computation, 2026.

Table 19: Hyperparameter Sensitivity (Causal Forest)

Trees	Min leaf	BTC		ETH	
		ATE	*p-value*	ATE	*p-value*
200	5	0.1515	0.0005	-0.1360	0.3225
200	8	0.1198	0.0000	-0.0942	0.3512
200	8	0.1134	0.0000	-0.0726	0.2897
500	8	0.1486	0.0004	-0.0653	0.6272
500	8	0.1182	0.0001	-0.0836	0.4058

Source: Authors' Computation, 2026.

4.5.6 Macro Confounder Robustness

As shown in Tables 20 and 21, both approaches preserve the sign and comparable magnitude of the baseline coefficients. This further supports that the contrarian sentiment effects are not driven by omitted macroeconomic events or regulatory announcements. The stability of the estimates, even when accounting for banking sector stress, regulatory enforcement actions, and monetary policy decisions, suggests that the identified effects are driven by crypto-market behavioral dynamics rather than macro-financial events (see Figure C4 for the macro event robustness diagnostics chart).

Table 20: Macro Event Dummy

Model	θ (with events)	θ original	Sign
BTC	-0.01044	-0.01064	Preserved
ETH	-0.12225	-0.12575	Preserved

Source: Authors' Computations, 2026.

Table 21: Event Window Exclusion

Model	θ excl	θ original	Sign
BTC	-0.0089	-0.0106	preserved
ETH	-0.1002	-0.1258	preserved

Excluded windows:

• Silvergate collapse - March 8, 2023

• SEC Binance action - June 5, 2023

• SEC Coinbase action - June 6, 2023

- Observations excluded: 26
- Remaining observations: 122

Source: Authors' Computations, 2026.

Table 22 compares the Baseline Estimates, Event Dummy Approach (± 2 days around Silvergate, Signature, Federal Reserve Meetings, and SEC Enforcement Actions), and Event Window Exclusion (± 5 days around Silvergate Collapse and SEC Binance/Coinbase Enforcement). The sign of the estimates is consistent across specifications, showing that the contrarian sentiment effect is not driven by macro-financial events.

Table 22: Consolidated Robustness Verdicts

Robustness		BTC verdict	ETH verdict	Threat addressed
i.	Nuisance learner	Fragile (33%)	Fragile (50%)	Learner artifact
ii.	• Rolling window	Robust (89%)	Fragile (56%)	Window specificity
	• Leave-one-out	Robust (100%)	Fragile (67%)	Single-month driver
iii.	• Permutation	Robust	Fragile	Spurious co-movement
	• Lag-k placebo	Robust	Robust	Autocorrelation
iv.	Lagged treatment	Fragile	Fragile	Timing
v.	• Bootstrap CI	Wide	Very wide	Sampling noise
	• Hyperparameter	Robust	Fragile	Tuning artifact
vi.	Macro controls	Robust	Robust	Macro shocks

Source: Authors' Compilation, 2026.

4.6 Discussion of Findings

I. The Contrarian Sentiment Effect

The estimated causal effect of social media sentiment on cryptocurrency returns is consistently negative (contrarian), where increasing positive sentiment precedes decreasing returns. The direction of the effect for the BTC DML-PLR estimate ($\hat{\theta}$ = -0.010) is consistent with the highly significant ETH Causal Forest ATE ($\hat{\theta}$ = -0.294, *p-value* < 0.001). This is in tandem with the peak-optimism-reversal mechanism, where sentiment peaks after bullish sentiment is fully

incorporated, and subsequent returns are driven by profit-taking and mean-reverting forces. Granger causality tests, as a preliminary check, confirm that sentiment leads returns at a lag of 2-3 days for both cryptocurrencies (BTC: *p-value* = 0.009; ETH: *p-value* = 0.042), where the coefficients are negative, as expected from short-run predictability as documented in Raheman *et al.* (2022). The non-significance of the BTC DML estimate likely results from a small sample size (n=159) and not from the absence of a structural effect, as the Causal Forest results yield high statistical significance

In contrast with the positive findings of Kraaijeveld and De Smedt (2020) and the mixed findings of another correlational research, such as those of Kyriazis *et al.* (2023). This is attributable to identification bias: reverse causality effects, in which returns cause sentiment, are known to bias correlational estimates positively. By adjusting for high-dimensional confounding and feedback effects, we identify an underlying contrarian structure that operates in tandem with theories of behavioral finance on sentiment-driven mispricing, subsequent correcting mechanisms (De Long *et al.*, 1990; Baker & Wurgler, 2006), and further existing findings on media-based sentiment research in tandem with Tetlock (2007) position that high optimism predicts low subsequent returns.

II. The Volatility Null: Distinguishing Mechanisms

The social media sentiment effect has no causal effect on realized volatility for BTC (*p-value* = 0.793) and ETH (*p-value* = 0.688). The result is robust as it identifies the transmission channel of social media sentiment. In other words, social media affects the direction of asset prices but does not add uncertainty to asset prices. This contrasts with the findings of Antweiler & Frank (2004) and Tetlock (2007) in equity markets, which found that social media sentiment is correlated with volatility and disagreement. Therefore, it can be inferred that social media is used as a directional trading guide in cryptocurrency markets. This corroborates the single-channel mechanism in which only informed traders react to social media sentiment, which is quickly incorporated into asset prices without creating uncertainty. Therefore, it supports the single-channel mechanism of social media influence and rejects uncertainty-amplification and contagion effects of social media in cryptocurrency markets.

III. Regime Heterogeneity: Two Mechanisms in One Market

The findings identify strong regime-dependent heterogeneity. For the BTC market, the treatment effect changes sign, sentiment exhibits momentum-consistent behavior in bear markets (CATE = +0.161) but contrarian in bull markets (CATE = -0.133), with a highly significant difference (p-value < 0.001). This sign flip in the treatment effects suggests that sentiment conveys different information in bear and bull markets. In bearish markets, for instance, positive sentiment may signal new credible information about improving fundamentals, leading to directional price responses. Conversely, in bull markets, the same signal is interpreted as over-optimism, leading to profit-taking, which in turn produces a contrarian effect.

This regime-switching behavior agrees with the Adaptive Market Hypothesis (AMH) proposed by Lo (2004), which posits that trading strategies are effective in varying market regimes. Also, the findings align with those of Zhang *et al.* (2020) and Vukovic *et al.* (2025) in terms of the presence of regime-dependent market efficiency in cryptocurrency markets, but extend the literature by identifying sentiment as the transmission channel. In the case of ETH, both regimes have contrarian effects, although the bear-market effect is significantly larger in magnitude than its bull-

market counterpart. That is, over-optimism due to sentiment is higher in bear markets for ETH, which could be due to narrative-driven fundamentals such as DeFi activity and network upgrades amplifying speculative responses in weak markets.

IV. Influencer Importance and the Attention Mechanism

The influencer's importance moderates the sentiment effect for ETH. Days with high importance influencers have an extremely strong contrarian effect (CATE = -0.275), compared to low importance days with almost zero effect (+0.018, *p-value* = 0.063). This identifies with attention-based models of trading (Da *et al.*, 2011), in which low attention from investors causes prices to deviate before reverting to their means. Also, this corroborates Ante (2023), who found abnormal stock returns following posts from accounts with high follower counts. Nonetheless, the findings provide evidence of the direction of these effects being contrarian rather than momentum-based. Importantly, we also find that influencer importance is not significant in explaining BTC's stock prices (*p-value* = 0.138). This implies that Bitcoin's larger liquidity and institutional presence mitigate the effect of individual social media influencers. The contrast between BTC and ETH supports our hypothesis of smaller cryptocurrencies being more vulnerable to shocks of concentrated attention.

V. ETH Endogeneity and the IV-DML Sign Progression

The Hausman pre-test (stat = 10,334) shows that ETH importance-weighted sentiment is indeed endogenous, as previous-day gains positively influence next-day sentiment, thereby creating the illusion of a spurious negative relationship in the OLS estimation ($\hat{\theta}$_OLS = -0.258). DML-PLR corrects for observable confounding and estimates ($\hat{\theta}$_DML = +0.010), and IV-DML further corrects for the returns-to-sentiment feedback loop and estimates $\hat{\theta}$_IV = +0.851 (ns, wide CI). The consistency in the positive direction of the estimated effect after correcting for endogeneity indicates that the causal effect of ETH importance-weighted sentiment on returns is non-negative and captures fundamental information in highly engaged posts.

VI. Robustness Check

Three findings emerge from the robustness checks: The BTC sentiment-return effect is highly robust with respect to nuisance learners, sample window, placebo diagnostics, treatment timing, and macro controls. Secondly, the ETH importance-weighted sentiment effect is more sensitive with respect to temporal/heterogeneity specifications. This indicates that the ETH relationship is not structurally constant but is instead condition-dependent. Also, both placebo diagnostics and macro events strongly reject spurious correlation explanations.

5.0 Conclusion, Implications, Limitations, and Future Research

5.1 Summary of Contributions

The study provides the first causally identified evidence on the impact of social media sentiment on cryptocurrency market returns and volatility by applying the Double Machine Learning and the Causal Forest frameworks to the daily panel dataset of Bitcoin and Ethereum over the sample period of 159 trading days. Addressing RQ1 (Does sentiment causally impact cryptocurrency returns?), the findings show robust contrarian direction. Specifically, the average treatment effect of the Ethereum Causal Forest is statistically significant at -0.294 (*p-value* < 0.001). Similarly, the Bitcoin DML-PLR specifications estimates are directionally consistent. The negative (contrarian) effect is only obtained after rigorously controlling for the confounding and reverse causality biases that otherwise yield spurious positive relationships in the conventional correlational approach.

For RQ2 (Does sentiment causally affect volatility?), the findings indicate no causal effect. Given that models 5 and 6 robustly did not find any significant effects (*p-value* > 0.68 for all specifications), we strongly reject the uncertainty amplification hypothesis. Moreover, on RQ3, which seeks to determine whether the causal effect is heterogeneous, the findings show significant heterogeneity in both market regimes and influencer importance. For Bitcoin, the effect changes sign between market regimes (bear market +0.161 vs. bull market -0.133, *p-value* < 0.001), whereas for Ethereum, higher influencer importance strengthens the negative effect (high vs. low importance -0.275 vs. +0.018, *p-value* = 0.063). Therefore, the findings provide the first causally identified evidence on the market conditions in which social media sentiment translates into cryptocurrency price movements.

5.2 Policy and Practitioner Implications

The study's findings offer significant implications for various stakeholders. For traders and portfolio managers, positive spikes in social media sentiment, especially when they come from important influencers, should be viewed as contrarian indicators. This phenomenon is particularly strong for Ethereum under all market conditions and for Bitcoin in bull regimes. However, it is important to note that the selection of the regime is not discretionary but rather fundamental because the same shock that predicts positive returns in bear markets for BTC predicts negative returns in bull markets. Therefore, any trading strategy based on sentiment must be conditioned on the regime to avoid any systematic misinterpretation of the sentiment.

For exchange risk managers, the absence of a causal impact of sentiment on volatility implies that the focus of sentiment-monitoring systems should be on the direction of risk, rather than the volatility itself. For relevant financial regulators, the findings provide evidence that high-importance influencer sentiment has a measurable and exploitable causal impact on cryptocurrency returns, particularly Ethereum, whereas low-importance activity has a negligible impact. This asymmetric influence raises significant implications for the rules around disclosure, integrity, and manipulation in the largely unregulated crypto markets, where influential actors are able to manipulate markets through widely amplified communications.

5.3 Study Limitations

The study has four limitations that qualify the generalizability and precision of the findings. First, the relatively small sample of 159 trading days limits statistical power, particularly for DML-PLR estimates; for example, the BTC VADER effect ($\hat{\theta}$ = -0.010, *p-value* = 0.121) is directionally consistent but statistically insignificant, and power calculations suggest the design is underpowered to detect small effects. Extending the sample to a longer time horizon would improve precision and strengthen inference. Second, the study period (January 2 - June 12, 2023) captures a post-FTX recovery phase, implying that the contrarian effect, BTC regime sign-flip, and ETH influencer-importance heterogeneity may reflect period-specific dynamics rather than universal structural relationships across market cycles. Thirdly, even though the Causal Forest model identified a statistically significant ETH average treatment effect (-0.294), the individual-level CATE estimates were quite imprecise (SD ≈ 0.978). This constrains the results' usefulness for observation-level treatment assignment; consequently, inference is more reliable at the subgroup level, such as regime-based and influencer-importance heterogeneity. Fourth, although instrumental variable estimation is implemented to address potential endogeneity, the instruments exhibit limited strength after residualisation, which constrains the precision of IV-based estimates. Thus, causal interpretation primarily relies on the DML-PLR framework under a selection-on-observables assumption.

5.4 Future Research Directions

Future study can extend the sample period over several market cycles (e.g., 2020–2024) to verify if the contrarian effects and sign changes are structural or unique to the post-crash phase. Second, employing the same DML and Causal Forest methodology for other cryptocurrencies and conventional stock markets will permit an assessment of the universality of sentiment transmission channels. Third, if daily data is replaced with intraday data, for instance, at an hourly frequency, it will be possible to estimate sentiment-return dynamics at higher frequencies and verify if the contrarian effect is observed at higher frequencies or only at the daily level. Moreover, employing LLMs for topic extraction for high-importance influencer posts will assist in understanding what type of information is associated with regime-dependent effects, where sentiment is contrarian in bull markets but momentum-reinforcing in bear markets, and with the amplified negative sentiment impact observed during high-influence ETH trading days.

REF References

Akyildirim, E., Goncu, A., & Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research, 297*(1), 3-36. https://doi.org/10.1007/s10479-020-03575-y

Ante, L. (2023). How Elon Musk's Twitter activity moves cryptocurrency markets. *Technological Forecasting and Social Change, 186*, 122112. https://doi.org/10.1016/j.techfore.2022.122112

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance, 59*(3), 1259-1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Aslam, F., Memon, B. A., Hunjra, A. I., & Bouri, E. (2023). The dynamics of market efficiency of major cryptocurrencies. *Global Finance Journal, 58*, 100899. https://doi.org/10.1016/j.gfj.2023.100899

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics, 47*(2), 1148-1178. https://doi.org/10.1214/18-AOS1709

Awad, A. (2022). *Impact of social media sentiment on cryptocurrency price movements* (Master's thesis, ETH Zurich). https://doi.org/10.3929/ethz-b-000587884

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance, 61*(4), 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies, 21*(2), 785-818. https://doi.org/10.1093/rfs/hhm079

Brauneis, A., & Mestel, R. (2018). Price discovery of cryptocurrencies: Bitcoin and beyond. *Economics Letters, 165*, 58–61. https://doi.org/10.1016/j.econlet.2018.02.001

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal, 21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance, 66*(5), 1461-1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Damaševičius, R. (2025). *Cryptocurrency dataset* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.14749166

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy, 98*(4), 703-738. https://doi.org/10.1086/261703

Du, K., Xing, F., Mao, R., & Cambria, E. (2024). Financial sentiment analysis: Techniques and applications. *ACM Computing Surveys, 56*(9), 1-42. https://doi.org/10.1145/3649451

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media, 8*(1), 216-225. https://doi.org/10.1609/icwsm.v8i1.14550

Jahanbin, K., & Zare Chahooki, M. A. (2023). Aspect-based sentiment analysis of Twitter influencers to predict the trend of cryptocurrencies based on hybrid deep transfer learning models. *IEEE Access, 11*, 121656-121670. https://doi.org/10.1109/ACCESS.2023.3327060

Jahanbin, K., Zare Chahooki, M. A., & Rahmanian, F. (2023). *Database of influencers' tweets in cryptocurrency (2021-2023)* [Data set] (Version 5). Mendeley Data. https://doi.org/10.17632/8fbdhh72gs.5

Kang, H. J., Lee, S. G., & Park, S. Y. (2022). Information efficiency in the cryptocurrency market: The efficient-market hypothesis. *Journal of Computer Information Systems, 62*(3), 622-631. https://doi.org/10.1080/08874417.2021.1872046

Karasiński, J. (2025). The predictability of high-frequency returns in the cryptocurrency markets and the Adaptive

Market Hypothesis. *Central European Economic Journal, 12*(59), 34-48. https://doi.org/10.2478/ceej-2025-0003

Khan, M. (2022). *Predicting cryptocurrency value, based on sentimental analysis of social media post* (Master's thesis, Aalto University). https://aaltodoc.aalto.fi

Khursheed, A., Naeem, M., Ahmed, S., & Mustafa, F. (2020). Adaptive market hypothesis: An empirical analysis of time-varying market efficiency of cryptocurrencies. *Cogent Economics & Finance, 8*(1), 1719574. https://doi.org/10.1080/23322039.2020.1719574

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

Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money, 65*, 101188. https://doi.org/10.1016/j.intfin.2020.101188

Kreuzer, C., Sparrer, C., & Dorfleitner, G. (2026). Beyond pure hype: News sentiment and its role in the BTC and ETH futures market. *Review of Derivatives Research, 29*(1), Article 3. https://doi.org/10.1007/s11147-025-09223-6

Kulcsar, L., & van Engelen, F. (2023). *Twitter sentiment analysis on the cryptocurrency market* (Master's thesis, Jönköping University). https://www.diva-portal.org/smash/get/diva2:1762598/FULLTEXT01.pdf

Kyriazis, N., Papadamou, S., Tzeremes, P., & Corbet, S. (2023). The differential influence of social media sentiment on cryptocurrency returns and volatility during COVID-19. *The Quarterly Review of Economics and Finance, 89*, 307-317. https://doi.org/10.1016/j.qref.2022.09.004

Le Tran, V., & Leirvik, T. (2020). Efficiency in the markets of crypto-currencies. *Finance Research Letters, 35*, 101382. https://doi.org/10.1016/j.frl.2019.101382

Łęt, B., Sobański, K., Świder, W., & Włosik, K. (2022). Is the cryptocurrency market efficient? Evidence from an analysis of fundamental factors for Bitcoin and Ethereum. *International Journal of Management and Economics, 58*(4), 351-370. https://doi.org/10.2478/ijme-2022-0030

Lo, A. W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *The Journal of Portfolio Management, 30*(5), 15-29. https://doi.org/10.3905/jpm.2004.442611

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics, 29*(3), 305-325. https://doi.org/10.1016/0304-4076(85)90158-7

Nadarajah, S., & Chu, J. (2017). On the inefficiency of Bitcoin. *Economics Letters, 150*, 6-9. https://doi.org/10.1016/j.econlet.2016.10.033

Noda, A. (2021). On the evolution of cryptocurrency market efficiency. *Applied Economics Letters, 28*(6), 433-439. https://doi.org/10.1080/13504851.2020.1758617

Oikonomopoulos, S., Tzafilkou, K., Karapiperis, D., & Verykios, V. (2022). Cryptocurrency price prediction using social media sentiment analysis. In *Proceedings of the 13th ICIST Conference*. https://doi.org/10.1109/IISA56318.2022.9904351

Philippas, D., Rjiba, H., Guesmi, K., & Goutte, S. (2019). Media attention and Bitcoin prices. *Finance Research Letters, 30*, 37-43. https://doi.org/10.1016/j.frl.2019.03.031

Raheman, A., Kolonin, A., Fridkins, I., Ansari, I., & Vishwas, M. (2022). Social media sentiment analysis for cryptocurrency market prediction. *arXiv preprint*. https://doi.org/10.48550/arXiv.2204.10185

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica, 56*(4), 931-954. https://doi.org/10.2307/1912705

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance, 52*(1), 35-55.

https://doi.org/10.1111/j.1540-6261.1997.tb03807.x

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models* (pp. 80-108). Cambridge University Press. https://doi.org/10.3386/t0284

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. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters, 148*, 80-82. https://doi.org/10.1016/j.econlet.2016.09.019

Vukovic, D. B., Hassan, M. K., Zinovev, V., Rogova, E. M., & Shakib, M. (2025). Time-varying efficiency and predictability in cryptocurrency markets: Forward-looking dynamics. *International Journal of Finance & Economics*. Advance online publication. https://doi.org/10.1002/ijfe.3063

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association, 113*(523), 1228-1242. https://doi.org/10.1080/01621459.2017.1319839

Wei, W. C. (2018). Liquidity and market efficiency in cryptocurrencies. *Economics Letters, 168*, 21-24. https://doi.org/10.1016/j.econlet.2018.04.003

Wołk, K. (2020). Advanced social media sentiment analysis for short-term cryptocurrency price prediction. *Expert Systems, 37*(2), e12493. https://doi.org/10.1111/exsy.12493

Yi, E., Yang, B., Jeong, M., Sohn, S., & Ahn, K. (2023). Market efficiency of cryptocurrency: Evidence from the Bitcoin market. *Scientific Reports, 13*(1), 4789. https://doi.org/10.1038/s41598-023-31617-3

Zhang, Y., Chan, S., Chu, J., & Sulieman, H. (2020). On the market efficiency and liquidity of high-frequency cryptocurrencies in a bull and bear market. *Journal of Risk and Financial Management, 13*(1), 8. https://doi.org/10.3390/jrfm13010008

APP Appendices

Appendix A Pre-modelling Suitability Tests

Table A1: Distributional Checks (Jarque-Bera Normality Test)

Feature	JB Statistic	p-value	Normal Status	Skewness	Ex. Kurtosis	Implication
log_return_btc	114.50	0.000	No	1.221	3.529	Heavy tails - HC3 SE
log_return_eth	19.32	0.000	No	0.409	1.590	Near-symmetric
realized_vol_7d_btc	177.92	0.000	No	1.811	3.899	Heavy tails - HC3 SE
realized_vol_7d_eth	50.68	0.000	No	1.008	2.014	Moderate right skew
vader_score_btc	154.47	0.000	No	-1.682	3.646	Heavy tails - HC3 SE
vader_score_eth	272.22	0.000	No	-2.005	5.232	Heavy tails - HC3 SE
inf_weighted_sentiment_btc	356.63	0.000	No	1.035	7.310	Heavy tails - HC3 SE
inf_weighted_sentiment_eth	6,054.28	0.000	No	4.720	29.711	Extreme - HC3 SE + log transform
log_volume_btc	1.29	0.525	Yes	0.014	-0.416	Normal
log_volume_eth	4.36	0.113	Yes	-0.391	0.286	Normal
intraday_range_btc	77.44	0.000	No	-1.486	1.833	Moderate skew
bull_regime_btc	26.51	0.000	No	-0.191	-1.989	Binary - expected
cum_return_7d_btc	46.65	0.000	No	1.122	1.536	Moderate skew
inf_retweet_total_btc	198.84	0.000	No	1.874	4.197	Heavy tails - log-transform
inf_favorite_total_btc	23,846.66	0.000	No	6.564	60.493	Extreme - log-transform IV
inf_retweet_total_eth	16,622.89	0.000	No	6.223	50.151	Extreme - log-transform IV
inf_favorite_total_eth	5,667.24	0.000	No	4.569	28.746	Heavy tails - log-transform
inf_total_importance_btc	16,691.41	0.000	No	5.867	50.440	Extreme - HC3 SE
inf_total_importance_eth	4,806.08	0.000	No	4.410	26.336	Heavy tails - HC3 SE
inf_tweet_count_btc	19.89	0.000	No	0.816	0.685	Near-symmetric
inf_tweet_count_eth	21.58	0.000	No	0.903	-0.212	Near-symmetric

Note: Jarque–Bera tests indicate widespread non-normality across variables, with most features exhibiting skewness and excess kurtosis, consistent with heavy-tailed financial and sentiment data. Only *log_volume_btc* and *log_volume_eth* satisfy normality at conventional levels, while several sentiment and engagement measures display extreme kurtosis, motivating log transformations.

Table A2: Multicollinearity (VIF) of All Controls

BTC Control Variable	VIF	Status	ETH Control Variable	VIF	Status
Lag1_Return_Btc	1.36	Ok	Lag1_Return_Eth	1.31	Ok
Log_Volume_Btc	1.58	Ok	Log_Volume_Eth	1.26	Ok
Intraday_Range_Btc	1.23	Ok	Intraday_Range_Eth	1.23	Ok
Bull_Regime_Btc	1.79	Ok	Bull_Regime_Eth	1.79	Ok
Cum_Return_7d_Btc	2.73	Ok	Cum_Return_7d_Eth	2.25	Ok

Note: All VIF < 10 (conventional threshold). Maximum VIF = 2.73 (cum_return_7d_btc). No controls dropped or orthogonalized. Full control set enters all DML nuisance models as specified.

Table A3: Sentiment–Return Cross-Correlation (Lags -5 to +5)

Crypto Type	Lag	Correlation	p-value
BTC			
	VADER Score → Log Return		
	T_{-3} (sentiment leads 3d)	-0.073	0.368
	T_{-2} (sentiment leads 2d)	0.086	0.286
	T_{-1} (sentiment leads 1d)	0.009	0.915
	T_{+0} (contemporaneous)	-0.092	0.248
	T_{+1} (return leads 1d)	-0.055	0.49
	Importance-Weighted Sentiment → Log Return		
	T_{-3} (sentiment leads 3d)	-0.242***	0.002
	T_{-2} (sentiment leads 2d)	-0.099	0.216
	T_{-1} (sentiment leads 1d)	-0.011	0.893
	T_{+0} (contemporaneous)	0.137*	0.084
	T_{+1} (return leads 1d)	0.121	0.13
ETH			
	VADER Score → Log Return		
	T_{-3} (sentiment leads 3d)	-0.148*	0.065
	T_{-2} (sentiment leads 2d)	-0.126	0.115
	T_{-1} (sentiment leads 1d)	0.017	0.834
	T_{+0} (contemporaneous)	-0.003	0.971
	T_{+1} (return leads 1d)	-0.031	0.701
	Importance-Weighted Sentiment → Log Return		
	T_{-3} (sentiment leads 3d)	-0.149*	0.064
	T_{-2} (sentiment leads 2d)	-0.165**	0.039
	T_{-1} (sentiment leads 1d)	0.072	0.366
	T_{+0} (contemporaneous)	0.083	0.298
	T_{+1} (return leads 1d)	0.112	0.162

Note:

- Negative lags indicate sentiment leading returns and positive lags indicate returns leading sentiment.
- Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

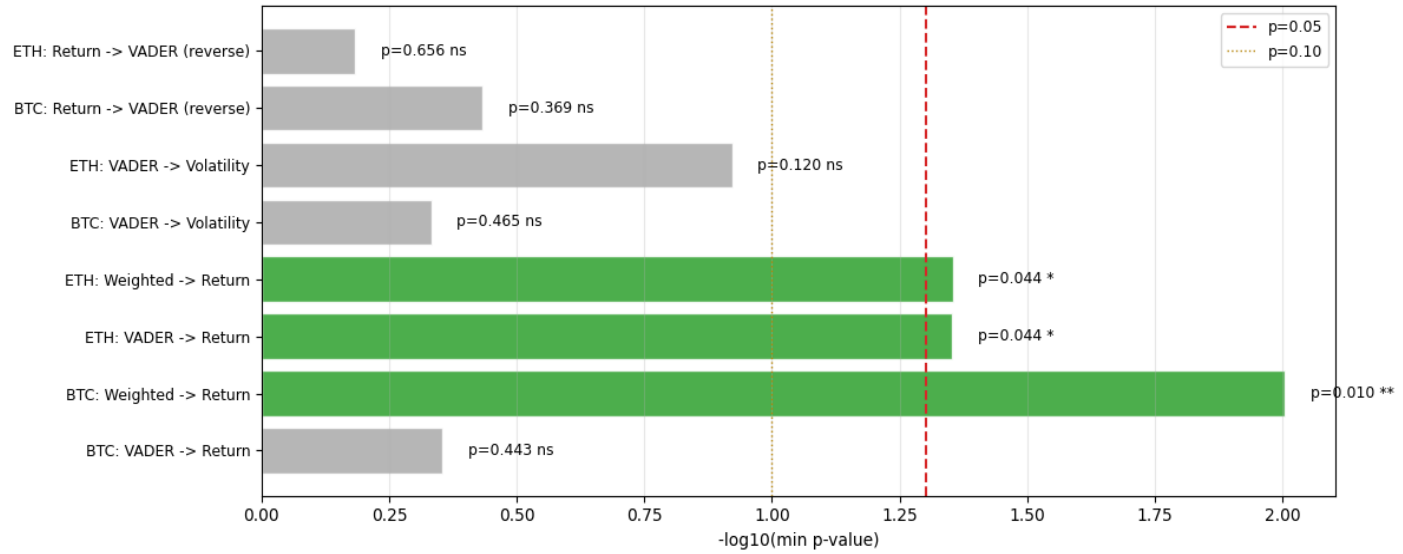
Test A4: Overlap/Common Support Check (Treatment Propensity)

Treatment Variable	1st Percentile	99th Percentile	Zero Mass	CV
Continuous Treatment (Range & Variation)				
vader_score_btc	-0.221	0.956	2.50%	0.41
vader_score_eth	-0.518	0.95	5.70%	0.5
inf_weighted_sentiment_btc	-0.012	0.054	0.00%	1.87
inf_weighted_sentiment_eth	-0.005	0.027	3.80%	2.34

	Class Balance & Propensity Score Overlap					
Binary Treatment	Threshold	T=1	T=0	PS Min	PS Max	Extreme %
t_high_vader_btc	0.5	125	34	0.174	0.93	0.00%
t_high_vader_eth	0.5	122	37	0.245	0.957	2.50%
t_high_inf_vader_btc	Q75 = 0.194	39	120	0.068	0.771	0.00%
t_high_inf_vader_eth	0.5 → Q75 (fixed)	8→~40	151→~119	-	-	48.4%→~0%
t_high_weighted_btc	Q75 = 0.012	40	119	0.044	0.756	0.60%
t_high_weighted_eth	Q75 = 0.002	42	117	0.025	0.769	0.60%

Note: PS denotes - Propensity Score (PS); CV - Coefficients of Variation; groups are Treated (T=1) and Control (T=0) groups.

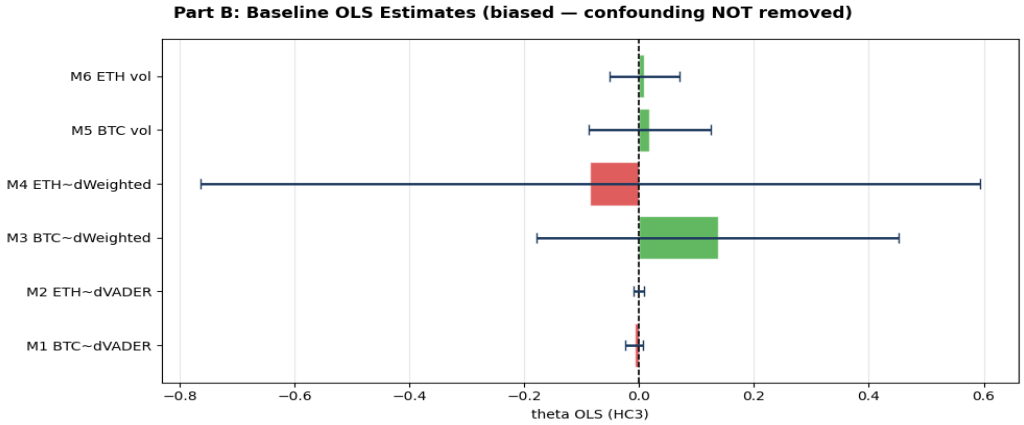
Fig. A1: Granger Causality (Minimum p-value across lags 1 - 5)



Source: Authors' Computations, 2026.

Appendix B Main Analysis

Fig. B1: Baseline OLS Estimates (Confounded Benchmark - Bias Not Removed)



Source: Authors' Computations, 2026.

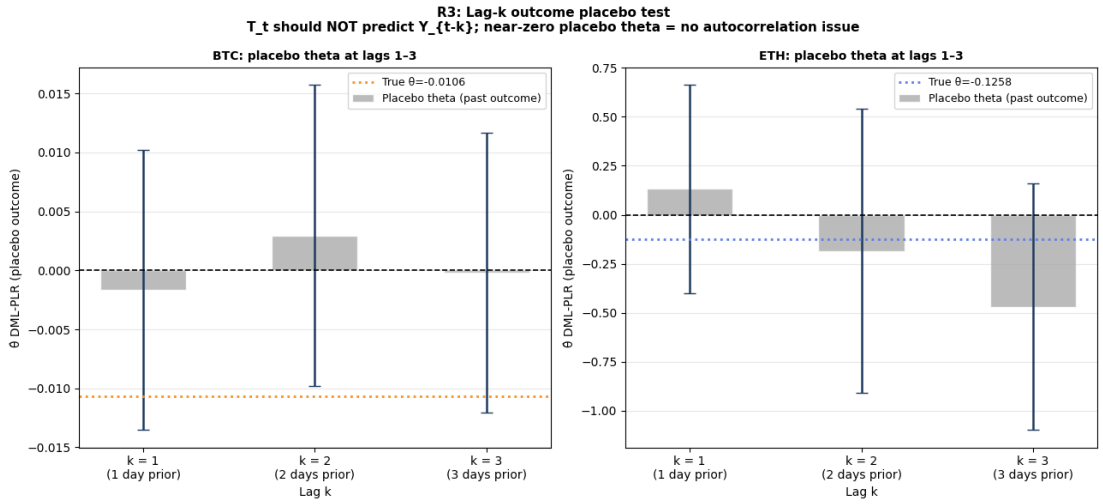
Appendix C Robustness Checks

Table C1: Rolling Window Stability (60-day window, step = 10)

Windows estimated: 9
BTC ~ dVADER: theta range [-0.03129, +0.00158] | 89% negative
ETH ~ dWeighted: theta range [-0.28601, +0.27489] | 56% negative

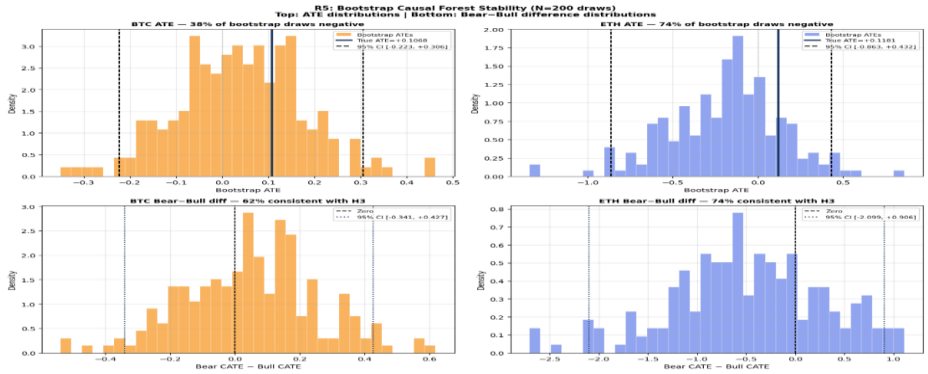
Source: Authors' Computations, 2026.

Fig. C1: Placebo Diagnostics



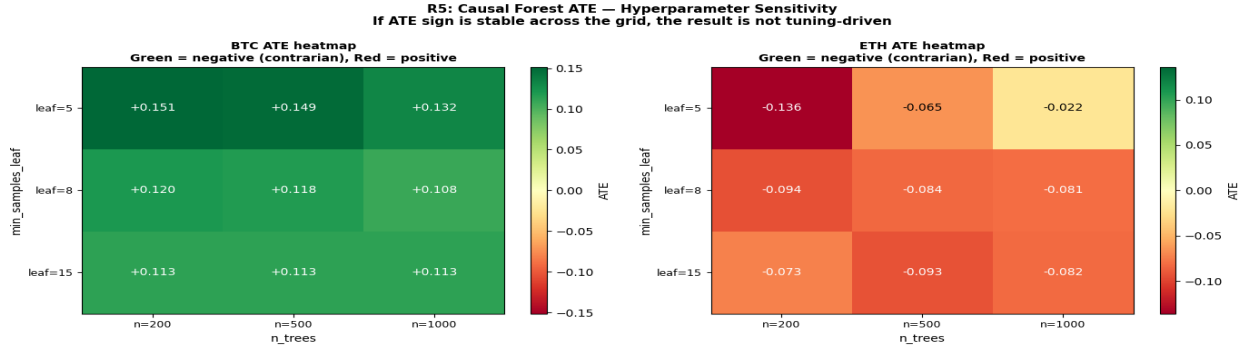
Source: Authors' Computations, 2026.

Fig. C2: Bootstrap CATE Stability and Hyperparameter Sensitivity



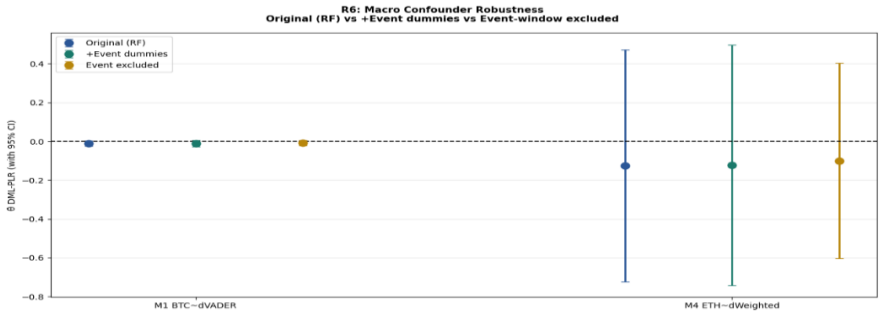
Source: Authors' Computations, 2026.

Fig. C3: Hyperparameter Sensitivity (Causal Forest ATE)



Source: Authors' Computations, 2026.

Fig. C4: Macro-Event Robustness



Source: Authors' Computations, 2026.

Appendix D Datasets and Code Availability

- **Datasets**:
  - CoinMarketCap Cryptocurrency Market Dataset: https://doi.org/10.5281/zenodo.14749166
  - Crypto Twitter Influencer Sentiment Dataset: https://doi.org/10.17632/8fbdhh72gs.5
- **Codes Repository**: https://github.com/Sam-Ayodeji/Crypto-Sentiment-Causal-ML/tree/main