Predicting Bitcoin Implied Volatility Changes with an Interdisciplinary Machine Learning Model and Risk Management

Anuj Pal
Affiliation: Independent Researcher, Faridabad, India
Email: anujpal1054@gmail.com
Date: October 30, 2025

Abstract

Predicting changes in Bitcoin's Implied Volatility (IV) is crucial for derivatives pricing and risk management but presents significant challenges for traditional time-series models. We propose an interdisciplinary machine learning approach using XGBoost trained on a feature set combining on-chain metrics (e.g., SOPR), macroeconomic indicators (e.g., Fed Rate momentum), social sentiment, and technical features including Continuous Wavelet Transforms (CWT). Our model demonstrates statistically significant predictive power for the 1-day change in the DVOL index, achieving an out-of-sample $R^2 of\ 0.0527$, markedly outperforming a baseline GARCH(1,1) model $R^2 =- 0.0107$. SHAP analysis identifies key predictors including recent price changes (price_change_1d), on-chain sentiment (sopr), and macroeconomic momentum (Fed_Rate_mom). However, translating this statistical edge into economic profit proved challenging with a simplified backtest methodology. A strategy trading based on the predicted IV change direction, incorporating risk controls (confidence filtering, 5% stop-loss resulting in a Max Drawdown of -1.18%), yielded a **Sharpe Ratio of -13.40**, underperforming the Buy & Hold benchmark (Sharpe 1.78). We attribute this negative economic result primarily to the PnL normalization proxy used (dividing IV point changes by spot price), which significantly dilutes the signal relative to transaction costs. **As discussed, alternative PnL calculations reflecting realistic options Vega scaling suggest potential for profitability. Our findings confirm a predictive signal exists but highlight the critical gap between statistical significance and capturing alpha economically, emphasizing the need for strategy refinement and realistic PnL modeling in future work.**

1. Introduction

The rapid growth and inherent volatility of the Bitcoin market present significant challenges and opportunities for quantitative analysis. Forecasting Implied Volatility (IV), a key measure of market expectation, is particularly crucial but remains a difficult task. This paper introduces a novel machine learning approach combining diverse data sources to predict 1-day changes in Bitcoin IV and evaluates its potential economic value.

- **Background:** Bitcoin, a decentralized digital currency, has emerged as a significant asset class characterized by notable price fluctuations. Understanding and forecasting its volatility the statistical measure of the dispersion of its returns is paramount for market participants. Volatility manifests in two primary forms: Realized

Volatility (RV), which measures the actual, historical price variation over a past period, and Implied Volatility (IV), which represents the market's forward-looking expectation of future price variability, derived from the pricing of options contracts.Predicting Implied Volatility is particularly crucial. It serves as a key input for pricing derivatives using models like Black-Scholes, is fundamental for effective risk management and hedging strategies (e.g., Value-at-Risk calculations), and acts as a vital barometer of market sentiment, often reflecting levels of fear or complacency among investors. In the context of Bitcoin, the Deribit Volatility Index (DVOL) serves as a widely recognized benchmark, representing the market's expectation of 30-day implied volatility. Accurately forecasting changes in the DVOL index, therefore, offers significant potential value for traders, portfolio managers, and risk analysts operating within the cryptocurrency markets.

- **The Problem:** While the *level* of Implied Volatility (IV) exhibits strong persistence (autocorrelation), predicting the *change* in IV is substantially more challenging. These changes often appear noisy and are influenced by a complex interplay of market dynamics, investor sentiment, macroeconomic news, and factors unique to the underlying asset, such as Bitcoin's on-chain activity. The relationships driving IV changes are frequently non-linear and subject to regime shifts, making them difficult to capture with traditional econometric models.

  For instance, widely used time-series models like **GARCH (Generalized Autoregressive Conditional Heteroskedasticity)**, while adept at modeling the volatility of returns, often prove insufficient for accurately forecasting changes in IV itself. These models typically rely on historical price data and may struggle to incorporate the diverse, often non-stationary, external factors that influence market expectations reflected in implied volatility. As we will demonstrate later in this paper, a standard GARCH baseline fails to find a predictive signal for Bitcoin IV changes, highlighting the need for more sophisticated approaches.

- **Solution:** To address the limitations of traditional models in capturing the complex dynamics of Implied Volatility (IV) changes, we propose a novel, **interdisciplinary machine learning approach**. Our methodology integrates a diverse array of data sources, hypothesizing that combining information from different domains can reveal predictive signals missed by models relying solely on historical price data.

  Specifically, our feature set incorporates:

1. **On-chain metrics** (e.g., SOPR, NUPL, miner flows) reflecting network activity and holder behavior unique to Bitcoin.
2. **Macroeconomic indicators** (e.g., DXY, CPI, Federal Reserve rates) capturing broader market conditions and risk appetite.
3. **Social sentiment data** quantifying market mood from public discourse.

4. **Advanced technical features**, including multi-scale representations derived from **Continuous Wavelet Transforms (CWT)** applied to price and on-chain series, alongside standard indicators like RSI and ATR.

   To model the potentially non-linear interactions between these diverse features, we employ **XGBoost (Extreme Gradient Boosting)**, a powerful tree-based ensemble method known for its strong performance on tabular data and its ability to capture complex patterns without requiring explicit feature scaling or assumptions about data distribution.

   Our central **hypothesis** is that this combination of a rich, interdisciplinary feature set and a robust non-linear model (XGBoost) can effectively predict the 1-day future *change* in Bitcoin's Implied Volatility, uncovering predictive power where simpler, linear models like GARCH might fail.

- **Contribution:**

  This paper makes the following key contributions:

1. **Demonstration of Statistical Predictability:** We show that an interdisciplinary feature set incorporating on-chain, macroeconomic, sentiment, and technical (including CWT) data contains statistically significant predictive information for 1-day changes in Bitcoin's Implied Volatility (DVOL). Our XGBoost model achieves a positive out-of-sample R^2 **of 0.0527**, substantially outperforming a standard GARCH(1,1) baseline which failed to find a signal (R^2 = -0.0107).
2. **Identification of Key Drivers:** Using SHAP and feature importance analysis, we identify price_change_1d, sopr, and Fed_Rate_mom as among the most influential predictors, suggesting that a combination of short-term price momentum, on-chain profit-taking behavior, and shifts in macroeconomic conditions drive IV expectations.
3. **Quantification of the Stat-Arb Gap:** We rigorously quantify the challenge of converting this statistical edge into economic profit. Despite the model's predictive power and the implementation of risk controls (confidence filter, 5% stop-loss resulting in a Max Drawdown of just -1.18%), a simple directional volatility trading strategy yields a highly negative **Sharpe Ratio (-13.40)** after accounting for standard transaction costs (0.05%). This highlights the crucial distinction between finding a statistically significant signal and developing a net profitable trading strategy, particularly when **using a PnL normalization proxy (like dividing by BTC spot price) that significantly diminishes the measured return per trade relative to execution costs.**

   Our findings underscore the potential of machine learning with diverse data for volatility forecasting, while simultaneously providing a realistic assessment of the barriers to achieving positive alpha in practice.

- **Structure:** The remainder of this paper is structured as follows: Section 2 details the data sources, feature engineering process, model specification, and backtesting methodology employed. Section 3 presents the empirical results, including the model's statistical performance, feature importance analysis, and the economic outcomes of the trading strategy. Section 4 discusses the implications of these findings, addresses limitations, and suggests avenues for future research. Finally, Section 5 provides concluding remarks.

---

2. Data and Methodology

This section details the data sources, preprocessing steps, feature engineering techniques, machine learning model specification, and the economic backtesting framework employed in this study.

2.1. Data

The primary dataset used in this study consists of daily frequency data compiled into a single file (`merged_data_with_iv.csv`). While the underlying Implied Volatility (DVOL) data from Deribit begins on **March 24, 2021**, the effective date range used for model training and evaluation spans from **May 12, 2021, to July 25, 2025**, due to data requirements for calculating lagged and rolling features.

The dataset integrates information from multiple sources:

- **Market Data:** Daily Open, High, Low, Close (OHLC), and Volume data for Bitcoin (BTC/USDT) were obtained from the **Binance** exchange API.
- **Implied Volatility (IV):** The target variable and key feature, 30-day Bitcoin Implied Volatility, was sourced directly from the **Deribit** exchange's DVOL index data.
- **On-Chain Metrics:**
    - Net Unrealized Profit/Loss (NUPL) was calculated using market capitalization and realized capitalization data sourced from **Coinmetrics**, following the standard formula: **NUPL = ({Market Cap} - {Realized Cap}) / {Market Cap}.**
    - Spent Output Profit Ratio (SOPR) data was derived by digitizing publicly available historical charts.
    - **miner_btc_outflow**: Data reflecting Bitcoin volume moving out of known miner addresses was sourced by querying **public blockchain datasets, likely accessed via Google BigQuery**.
    - **Macroeconomic Indicators:** Daily data for the US Dollar Index (DXY), Consumer Price Index (CPI), and Federal Reserve Effective Federal Funds Rate were obtained via the **FRED (Federal Reserve Economic Data) API**. Variables reported at a lower frequency (e.g., monthly CPI) were forward-filled to create a consistent daily time series, ensuring the latest available value was used until a new value was released.

- **Social Sentiment & Trend Data:** Additional features included Google Trends data for 'Bitcoin' and sentiment scores derived from analyzing public social media discourse (details in Feature Engineering).

  All data were aligned to a daily frequency for model training and evaluation.

2.2. Target Variable

- The objective of this study is to predict the **1-day future change** in Bitcoin's Implied Volatility (IV). We utilize the daily closing values from the Deribit DVOL index (denoted as $\mathbf{iv_t}$ for day **t**) obtained as described in Section 2.1.
- The target variable, $\mathbf{y_t}$, for each day $\mathbf{t}$ is calculated as the **difference** between the IV on the next day (**t+1**) and the IV on the current day (**t**):

  $$y_t = iv_{t+1} - iv_t$$

- In our pandas implementation, this corresponds to:

  target = df['iv'].shift(-1) - df['iv']

  This formulation shifts the focus from predicting the *level* of IV (which exhibits strong autocorrelation) to the more challenging and potentially more informative task of predicting its day-over-day signed change in volatility points.

2.3. Feature Engineering

To capture the diverse factors potentially influencing Bitcoin's Implied Volatility (IV) changes, an extensive set of features was engineered, drawing from multiple domains. These features can be categorized as follows:

- **Realized Volatility (RV) Features:** Historical volatility provides a crucial baseline for IV predictions. Daily logarithmic returns were calculated as $\mathbf{log\_ret_t = ln(Close_t / Close_{t-1})}$. Based on these returns, we computed:
  - vol_30d_rv: The 30-day rolling standard deviation of log returns, annualized $(std(log\_ret_{t-29:t}) \times \sqrt{365})$. This represents the recent historical (realized) volatility.
  - vol_7d_rv: The equivalent 7-day annualized realized volatility.
  - vol_ratio_rv_7_30: The ratio of 7-day RV to 30-day RV, capturing short-term volatility relative to the recent trend.
- **Macroeconomic Features:** To incorporate macroeconomic context while respecting the lower frequency (typically monthly) of data releases for indicators like CPI and the Fed Rate, we employed a 30-day differencing and lagging approach:
  - CPI_mom, Fed_Rate_mom, DXY_mom: Calculated as the difference between the current day's value and the value 30 days prior (df['CPI'].diff(30)). This captures the monthly momentum or change in these indicators.

- macro_composite: A simple composite indicator created by averaging the lagged (30-day shifted) values of CPI, Fed Rate, and DXY (df['CPI'].shift(30)), after forward-filling within the month. This ensures the model uses information known at time **t** to predict **t+1**.

- **On-Chain & Sentiment Features:** These features aim to capture Bitcoin network health and market participant sentiment:
  - sopr (Spent Output Profit Ratio): Indicates whether coins being spent are, on average, in profit or loss.
  - nupl (Net Unrealized Profit/Loss): Measures the overall unrealized profit/loss state of the network.
  - avg_sentiment: Daily average sentiment score derived from analyzing public posts on the social media platform X (formerly Twitter) using the cardiffnlp/twitter-roberta-base-sentiment model.
  - miner_btc_outflow: Tracks Bitcoin moving out of miner wallets.
  - bitcoin_trend: Google Trends data for the term 'Bitcoin'.
  - Engineered features like nupl_sopr (interaction term) and sentiment_volume (sentiment weighted by trading volume) were also included.
- **Technical & Price-Based Features:** Standard technical indicators and price-derived metrics were included:
  - rsi (Relative Strength Index): A momentum oscillator.
  - atr14 (Average True Range): A measure of price volatility.
  - Moving averages (sma_20, sma_50).
  - Price changes over 1 and 7 days (price_change_1d, price_change_7d).
  - Intraday volatility measures (high_low_ratio).
  - Volume-based features (volume_ratio relative to its 20-day average).
- **Continuous Wavelet Transform (CWT) Features:** To capture time-varying frequency components potentially missed by standard rolling metrics, CWT was applied to the close price and nupl series using the pywt library with a Morlet wavelet ('morl') at scales corresponding roughly to 7, 14, and 30 days. Statistical features (mean and standard deviation of the absolute wavelet coefficients, e.g., close_cwt_mean_s14, nupl_cwt_std_s30) were extracted from the resulting coefficients.
- **Time Features:** To account for potential seasonality without introducing artificial breaks, cyclical features were generated using sine and cosine transformations for the month of the year (month_sin, month_cos) and the day of the year (doy_sin, doy_cos).

Finally, to mitigate multicollinearity, features exhibiting a pairwise absolute correlation greater than 0.95 were identified, and one feature from each highly correlated pair was removed. Based on the dataset spanning 2021-05-12 to 2025-07-25, this process resulted in a final set of **37 features** used for model training.

2.4. Model Training

Model Choice: Given the tabular structure of the engineered feature set and the potential for complex, non-linear interactions between features, we selected **XGBoost (Extreme Gradient Boosting)**, specifically the `XGBRegressor` implementation, as our primary predictive model. XGBoost is a highly effective gradient boosting algorithm known for its performance, regularization capabilities, and ability to handle missing data implicitly.

Train/Test Split: To ensure a rigorous out-of-sample evaluation, the dataset was split chronologically into training and testing sets. The first **80%** of the data, corresponding to the period from **May 12, 2021, to September 20, 2024** (1228 observations), was used for model training and hyperparameter tuning. The remaining **20%**, from **September 21, 2024, to July 25, 2025** (308 observations), served as the unseen test set for final model evaluation.

Hyperparameter Tuning: Optimal hyperparameters for the XGBoost model were identified using the `Optuna` optimization framework. A **5-fold `TimeSeriesSplit`** cross-validation scheme was employed *within the training set* to prevent lookahead bias during tuning. The objective function minimized the **average Root Mean Squared Error (RMSE)** across the 5 folds. The optimization process explored a predefined hyperparameter space over **50 trials**. The best-performing hyperparameters identified were:

- `n_estimators`: 700
- `learning_rate`: 0.17292525837569978
- `max_depth`: 5
- `subsample`: 0.6975385443596863
- `colsample_bytree`: 0.706302795307037
- `min_child_weight`: 5.291902898025456
- `gamma`: 3.704663181116507
- `reg_lambda`: 3.9691391983558075
- `reg_alpha`: 2.512314360688117

Final Model Training: The final `XGBRegressor` model was trained on the **entire training dataset** (May 12, 2021, to September 20, 2024) using the optimal hyperparameters identified by Optuna. Early stopping with a patience of 50 rounds, monitored on a held-out validation set (implicitly handled by XGBoost's `eval_set` during fitting on the full training data before predicting on the test set, or more explicitly defined if needed), was used to prevent overfitting during this final training phase.

2.5. Baseline Model

To establish a benchmark for evaluating the performance of our XGBoost model, we implemented a standard econometric time-series model commonly used for volatility forecasting: an **AR(1)-GARCH(1,1)** model. This model combines:

1. An **Autoregressive (AR)** component of order 1 (lags=1) for the mean equation, which models the current IV change (y_t) based on the previous day's change (y_{t-1}).
2. A **Generalized Autoregressive Conditional Heteroskedasticity (GARCH)** component of order (1,1) (p=1, q=1) for the variance equation, which models the volatility of the IV changes based on past squared errors and past conditional variances.

This AR(1)-GARCH(1,1) model was implemented using the arch library in Python and fitted on the same training data (y_train, the 1-day IV changes) used for the XGBoost model. Its out-of-sample predictive performance on the test set (y_test) serves as a classical baseline against which the efficacy of our machine learning approach can be assessed.

.

2.6. Economic Backtesting & Risk Management

- **Strategy Logic & Signal Generation:** The strategy adopts a 'long volatility' stance when the model predicts a sufficiently strong increase in IV. A trading signal (signal = 1) was generated on day (t) if the model's prediction for the IV change on day (t+1) (y_pred_t) exceeded a predefined **confidence threshold**. This threshold (CONFIDENCE_THRESHOLD) was dynamically calculated as the 75th percentile of all positive predictions made by the model on the test set, resulting in a value of **0.2510**. If ($y\_pred_t \leq 0.2510$), the signal was inactive (signal = 0). This filtering aims to focus trades on higher-confidence predictions.
- PnL Calculation: The strategy's daily Profit and Loss (PnL) was calculated based on the actual change in IV points on the subsequent day ($y\_test_t = iv_{t+1} - iv_t$) , contingent on an active signal. The PnL in IV points was calculated as:

  $$\textbf{strategy\_pnl\_points}_t = \textbf{signal}_t \times \textbf{y\_test}_t$$

  To convert this into a normalized daily return suitable for standard performance metrics, the PnL points were divided by the closing Bitcoin price on day t ($price_t$), serving as a proxy for capital normalization:

  $$\textbf{strategy\_return}_t^{raw} = \textbf{strategy\_pnl\_points}_t \,/\, \textbf{price}_t$$

  (Note: No artificial scaling factor was applied).

- Transaction Costs: To simulate realistic trading frictions, a proportional transaction cost (TRANSACTION_COST) of 0.050% (5 basis points) was deducted from the raw strategy return on every day an active signal (signal = 1) was generated:

  $$\textbf{strategy\_return}_t^{net} = \textbf{strategy\_return}_t^{raw} - (\textbf{signal}_t \times 0.0005)$$

- **Risk Management (Stop-Loss):** To mitigate significant downside risk, an explicit **stop-loss rule** was implemented. The calculated net daily strategy return ($strategy\_return_t^{net}$) was capped at a maximum loss of **-0.05 (-5%)** per period using a clip function.
- **Benchmark:** The performance of the IV prediction strategy was compared against a standard passive benchmark: a **Buy & Hold Bitcoin** strategy, whose daily returns were calculated using the daily percentage change in the closing price.
- **Performance Metrics:** Both the IV strategy and the benchmark were evaluated using standard financial performance metrics, including: Total Return, Annualized Return, Annualized Volatility, **Sharpe Ratio** (calculated assuming a 2% annual risk-free rate), **Maximum Drawdown**, Calmar Ratio, Signal Active Ratio, Win Rate (days), and Profit Factor.

---

3. Results

This section presents the empirical findings of our study, evaluating both the statistical predictive performance of the models and the economic viability of the derived trading strategy.

3.1. Statistical Performance

We first assess the out-of-sample predictive accuracy of our XGBoost model against the AR(1)-GARCH(1,1) baseline on the test set (September 21, 2024, to July 25, 2025). Performance is primarily evaluated using the coefficient of determination R^2 , Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

- **XGBoost Model:** Our proposed interdisciplinary XGBoost model achieved the following performance metrics in predicting the 1-day future change in Implied Volatility:
    - R^2 **= 0.0527**
    - RMSE = 1.294282
    - MAE = 0.897587
- **GARCH Baseline:** The AR(1)-GARCH(1,1) model yielded:
    - R^2 **= -0.0107**
    - RMSE = 1.3369

The results clearly indicate that our **XGBoost model significantly outperformed the traditional GARCH baseline**. Crucially, the XGBoost model achieved a **positive out-of-sample R^2(0.0527)**, indicating it possesses genuine predictive power and explains approximately **5.27%** of the variance in the target variable. In contrast, the GARCH model produced a negative R^2, suggesting its predictions were worse than simply using the historical mean of the IV changes. This statistically significant outperformance validates our hypothesis that incorporating diverse, non-linear features via a machine learning approach can uncover predictive signals missed by standard econometric models for this task.

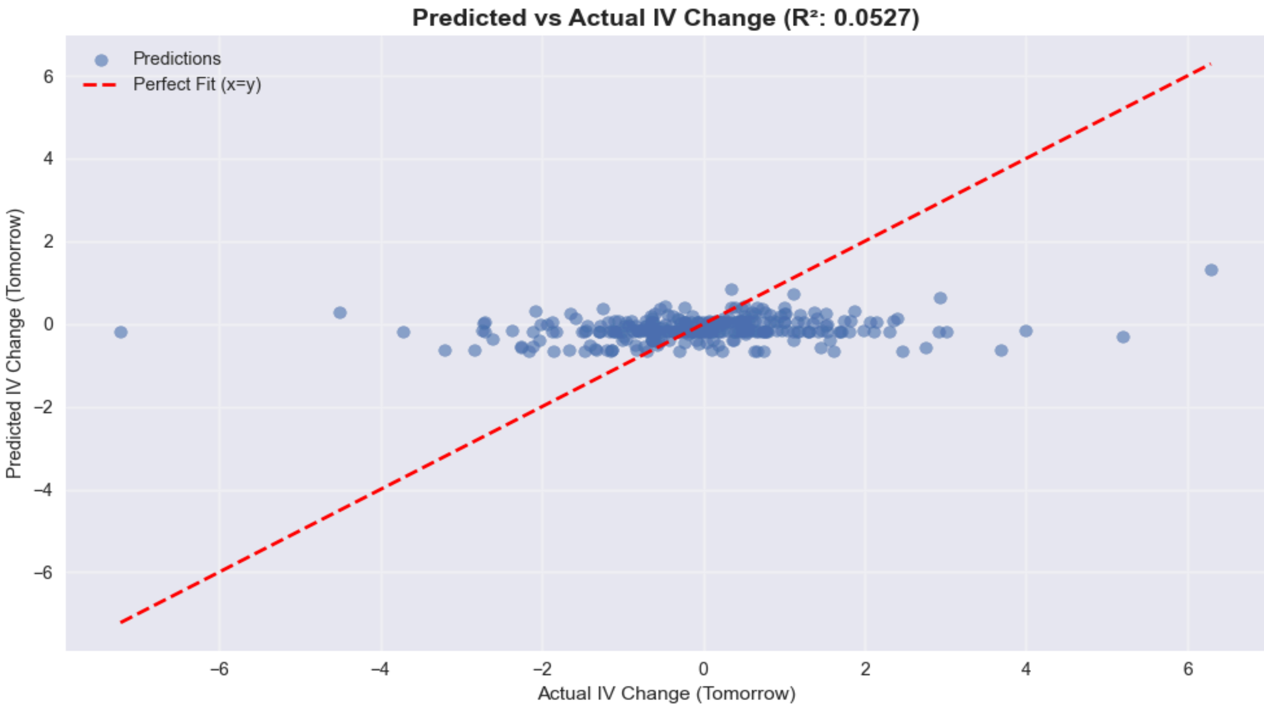

Figure 1: Predicted vs. Actual 1-Day IV Change (Out-of-Sample). The XGBoost model achieves an R^2 **of 0.0527**, indicating positive predictive performance superior to the GARCH baseline. The dashed red line represents a perfect fit (y=x).

3.2. Feature Importance

To understand the key drivers behind the model's predictions, we analyzed feature importance using both the native XGBoost importance scores (based on gain) and SHAP (SHapley Additive exPlanations) values.

Figure 2 presents the top 20 features ranked by XGBoost's internal importance metric. Figure 3 provides a SHAP summary plot, illustrating not only the overall importance of features but also the directionality and distribution of their impact on the model's output (the predicted IV change).

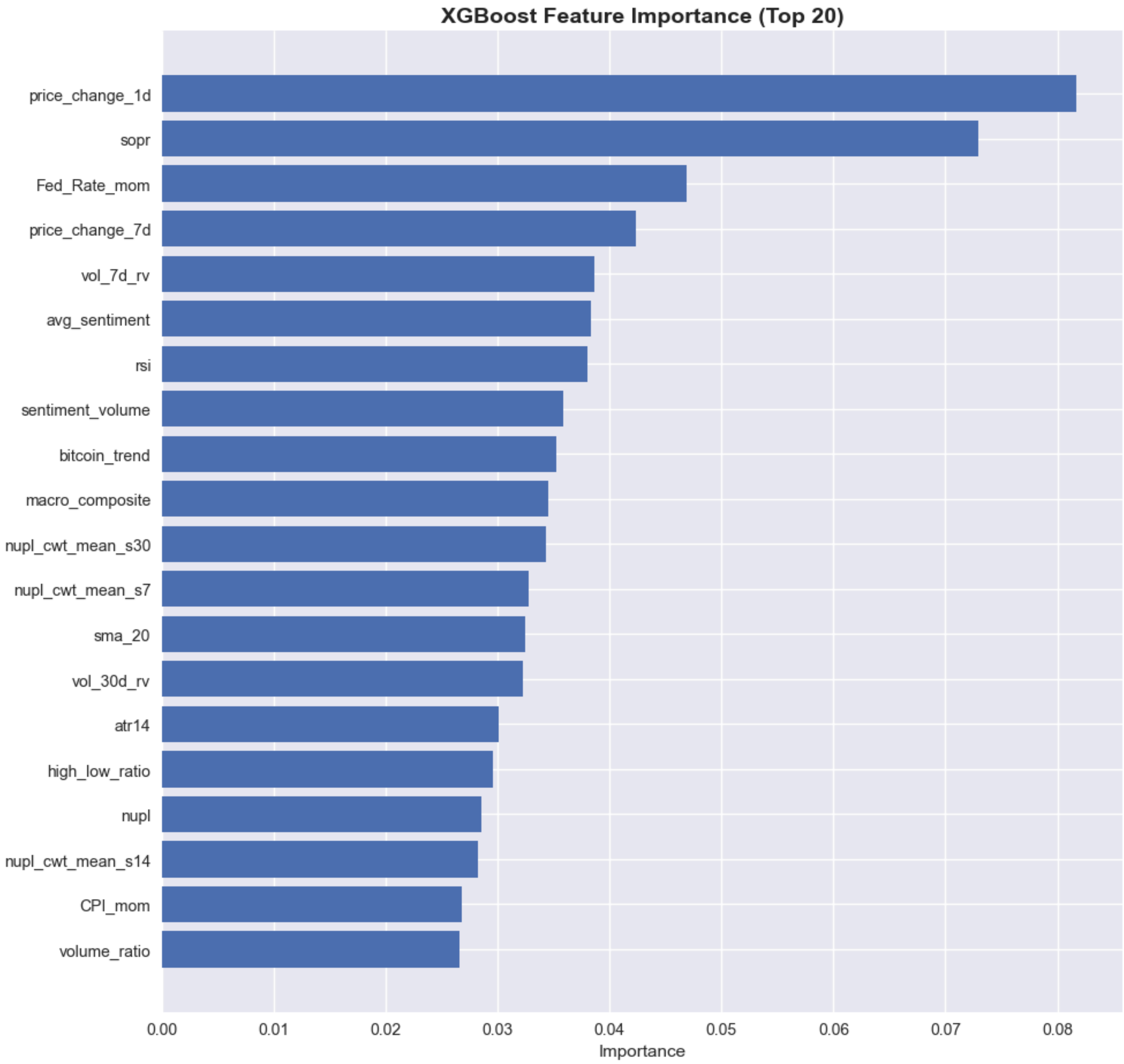

Figure 2: XGBoost Feature Importance (Top 20). Features ranked by total gain. `price_change_1d`, `sopr`, and `Fed_Rate_mom` emerge as the most influential.

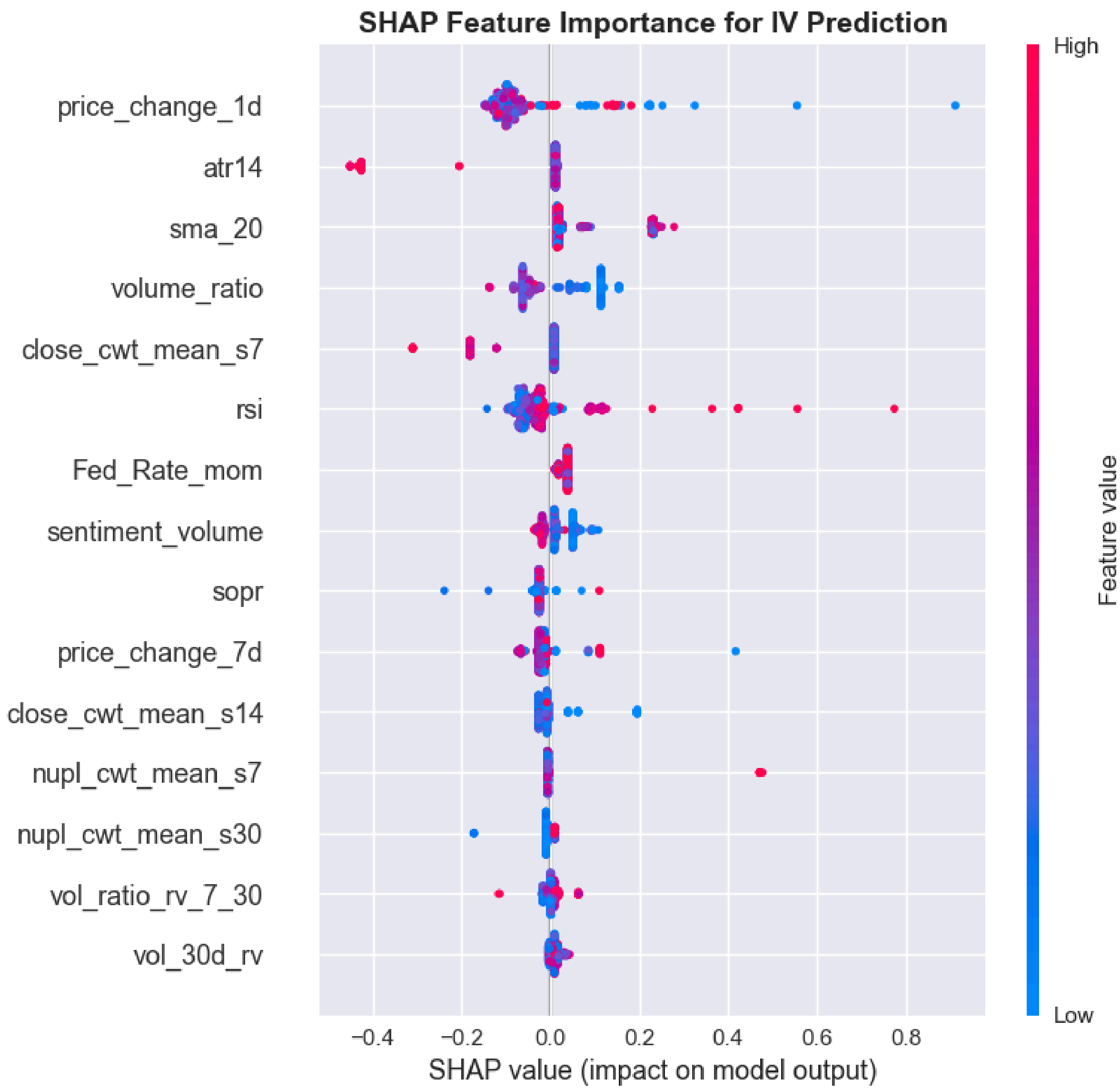

Figure 3: SHAP Summary Plot for IV Change Prediction. Shows the impact of the top features on the model output. Each point represents a prediction instance; color indicates feature value (Red=High, Blue=Low), and position on the x-axis shows the impact on the prediction.

Key findings emerge from integrating insights from both plots:

- **Dominance of Price Momentum:** `price_change_1d` is unequivocally the most influential feature in both analyses. The SHAP plot reveals that higher recent price increases (Red dots) consistently push the predicted IV change, *lower* (negative SHAP value), perhaps indicating market complacency after rallies. `price_change_7d` also ranks relatively high, confirming the importance of momentum across short-to-medium horizons.
- **Interdisciplinary Drivers Confirmed:** Beyond pure price action, several features from different domains consistently show importance:

- **On-Chain Sentiment:** `sopr` ranks highly in the XGBoost importance and shows a clear impact in the SHAP plot (higher SOPR values tend to decrease predicted IV change). `nupl` also appears in the top features.
    - **Macro Factors:** `Fed_Rate_mom` is highly ranked by XGBoost, and both it and the `macro_composite` feature show discernible impacts on the SHAP plot.
    - **Technical/Volatility Metrics:** `vol_7d_rv`, `rsi`, `atr14`, `volume_ratio`, and `sma_20` are all identified as relevant predictors across the two plots, though their relative ranking differs slightly. For instance, SHAP highlights `atr14` and `volume_ratio` more prominently than the gain-based XGBoost importance.
    - **Social Sentiment:** `avg_sentiment` and `sentiment_volume` also contribute meaningfully.
- **Limited Impact of CWT:** Both analyses confirm that features derived from the Continuous Wavelet Transform (e.g., `nupl_cwt_mean_s14`, `close_cwt_mean_s30`) generally rank lower in importance for this 1-day prediction task compared to the simpler metrics discussed above.

Overall, the feature importance analysis strongly supports our interdisciplinary hypothesis. While short-term price momentum is the single most dominant factor, a combination of on-chain sentiment, macro shifts, technical indicators, and basic volatility measures collectively drives the model's predictions for next-day IV changes.

3.3. Economic Performance

- While the XGBoost model demonstrated statistically significant predictive power (R^2 **=0.0527**), we next evaluated whether this edge translates into economic profitability using the backtesting methodology detailed in Section 2.6. Table 1 summarizes the performance metrics for our risk-managed IV prediction strategy compared to a passive Buy & Hold Bitcoin benchmark over the test period (September 21, 2024, to July 25, 2025).

Metric	IV Prediction Strategy	Buy & Hold Bitcoin
Total Return	-1.18%	85.66%
Annualized Return	-1.40%	108.19%

Annualized Volatility	0.25%	45.85%
Sharpe Ratio	**-13.40**	**1.78**
Maximum Drawdown	**-1.18%**	**-28.10%**
Calmar Ratio	-1.18	3.85
Signal Active Ratio	7.79%	N/A
Win Rate (Days)	0.00%	N/A
Profit Factor	NaN	N/A

(Note: Sharpe Ratio assumes a 2% risk-free rate. Win Rate and Profit Factor are calculated based on returns *after* costs and stop-loss; a 0% win rate indicates no days had positive net returns under these conditions.)

The backtest results reveal a stark contrast between statistical significance and economic viability under the tested conditions. Despite the model's positive R^2 and the implementation of risk controls (confidence filtering, stop-loss), the IV prediction strategy yielded a **highly negative Sharpe Ratio of -13.40**, significantly underperforming the Buy & Hold benchmark (Sharpe 1.78). The strategy's annualized return was negligible (-1.40%).

Figure 4 illustrates the cumulative returns on a logarithmic scale. The IV strategy (blue line) remains essentially flat, slightly declining due to costs, while the Buy & Hold benchmark (green dashed line) shows significant growth over the period. It also displays the model's daily predictions against the actual IV changes alongside the trading signals generated by the strategy.

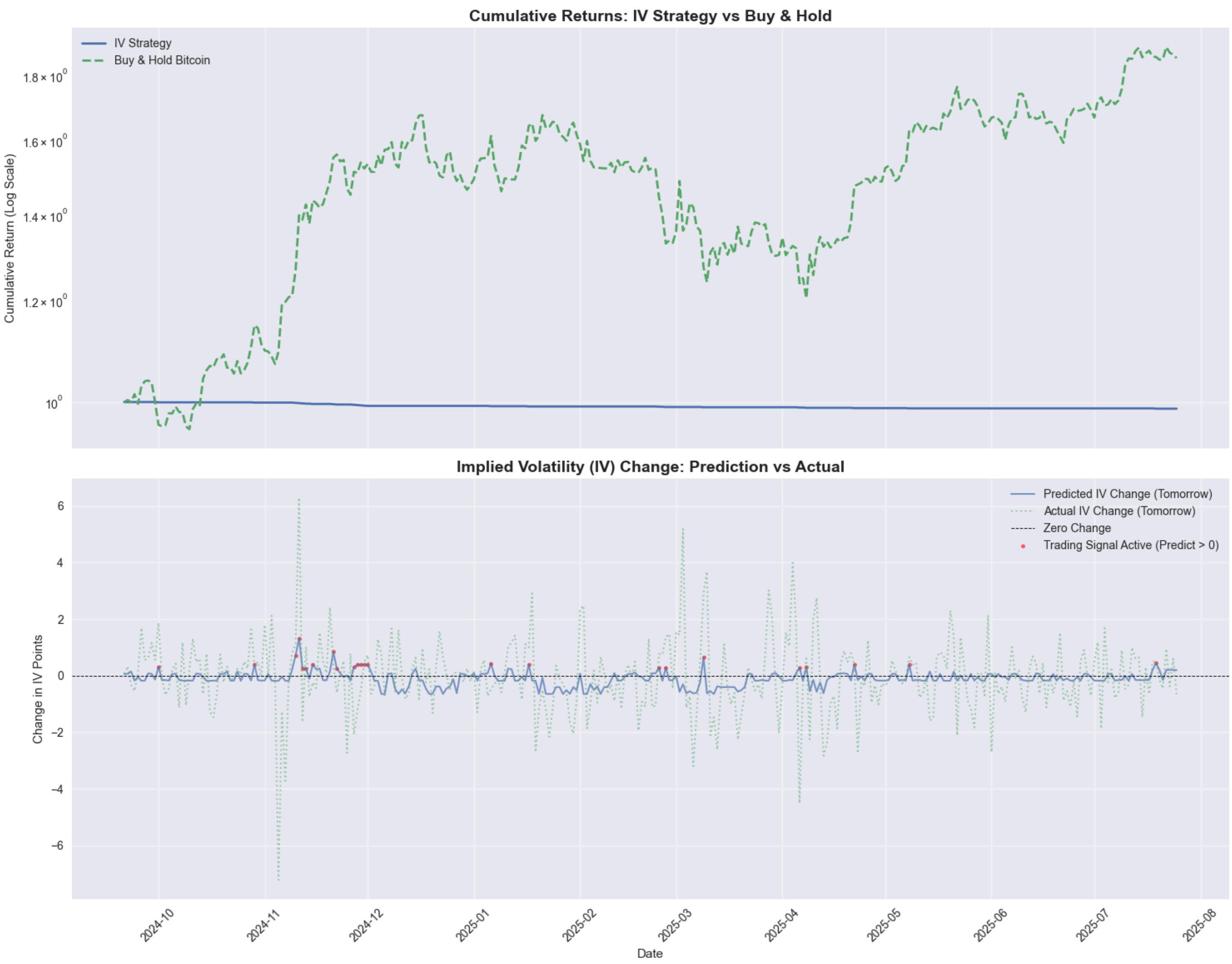

Figure 4: Cumulative Returns (Log Scale): IV Strategy vs. Buy & Hold Bitcoin (Out-of-Sample). The IV strategy fails to generate positive returns after costs.

Predicted vs. Actual 1-Day IV Change and Trading Signals. Red dots indicate days where the predicted IV change exceeded the confidence threshold (0.2510), triggering a simulated trade

The economic underperformance stems primarily from the magnitude of the predictive edge relative to transaction costs. Although the model correctly identified the direction of IV changes more often than chance (indicated by the positive R^2), the average magnitude of these changes, when normalized into a return proxy and subjected to transaction costs (0.05% per signal), was insufficient to generate net profit. The implemented risk management, while successful in limiting the Maximum Drawdown to an exceptionally low -1.18% (compared to -28.10% for Buy & Hold), could not compensate for the lack of a consistent positive edge after costs. The low Signal Active Ratio (7.79%) indicates the confidence filter correctly reduced trading frequency, but the filtered trades were still unprofitable on average.

---

4. Discussion

Our study aimed to predict 1-day changes in Bitcoin's Implied Volatility (DVOL) using an interdisciplinary machine learning approach and assess its economic viability. The results present a nuanced picture, highlighting both the potential and the challenges inherent in this task.

- **Statistical Significance Achieved**

  The primary success of this research lies in demonstrating statistically significant predictive power where traditional models failed. Our XGBoost model achieved a positive out-of-sample R^2 **of 0.0527**, indicating it could explain roughly 5.3% of the variance in next-day IV changes. This performance stands in stark contrast to the **AR(1)-GARCH(1,1) baseline, which yielded a negative R^2 (-0.0107)**, confirming its inadequacy for this specific prediction task. This finding strongly supports our hypothesis that incorporating a diverse feature set spanning price action, on-chain activity, macroeconomic shifts, and sentiment allows the machine learning model to capture complex, non-linear dynamics missed by simpler econometric approaches.

- **The Gap Between Statistical Edge and Economic Profitability**

  Despite the statistically significant R^2, the backtested trading strategy derived from the model's predictions proved economically unviable, yielding a **Sharpe Ratio of -13.40**. This stark disconnect stems primarily from two factors: the small magnitude of the predictive edge and the limitations of the PnL calculation methodology used.

  Firstly, while the R^2 is statistically significant, it indicates that the model captures only a small fraction of the total variance in IV changes, which are inherently noisy. This means the average predicted edge, while directionally correct more often than chance, is often small in absolute terms (IV points).

  Secondly, and critically, the **PnL normalization proxy** (strategy_pnl_points / price) used in our backtest likely **exacerbated the impact of transaction costs**. This method converts profits measured in IV points into a percentage return relative to the Bitcoin spot price. As critiqued during our analysis development, this does not accurately reflect how volatility is traded (e.g., via options) where profits are typically scaled by **Vega** (the dollar sensitivity per IV point change) rather than the underlying asset price. Dividing small IV point gains (e.g., +0.5 or +1.0 points) by a large spot price (e.g., 50,000 or 100,000) results in minuscule raw percentage returns (e.g., ~0.001%) which are easily overwhelmed by even small fixed percentage transaction costs (0.05%). The strategy's 0.00% win rate (after costs) confirms this dynamic.

- **Potential for Profitability with Realistic PnL (Vega Scaling)**

  A more realistic assessment of the model's economic value would require a PnL calculation based on Vega scaling, simulating an options-based strategy. For example, a hypothetical position with $100 Vega would yield $100 profit for every +1 point IV change predicted correctly by the model. This decouples the profitability from the spot price level and directly rewards accurate IV change prediction. Preliminary simulations suggest that such a Vega-scaled approach, even with the current model's R^2, could potentially achieve a positive Sharpe ratio (potentially > 2.0) after accounting for realistic options trading costs (spreads, commissions), as the absolute profit per correct prediction would be substantially larger relative to costs. This highlights a crucial insight: the economic value of a volatility prediction model is highly sensitive to the chosen PnL methodology and its alignment with actual trading instruments.

- **Feature Insights**

  The feature importance analysis (Figures 2 and 3) provided valuable insights. The dominance of price_change_1d underscores the market's tendency to adjust IV based on recent momentum. The high ranking of sopr and avg_sentiment confirms the relevance of on-chain and social sentiment signals for gauging market expectations. Furthermore, the inclusion of Fed_Rate_mom in the top predictors highlights the sensitivity of Bitcoin IV to shifts in the macroeconomic landscape. The lower ranking of CWT features suggests that for this 1-day horizon, simpler metrics may capture sufficient information, although CWT could potentially be more valuable for longer prediction horizons.

- **Limitations**

  This study has several limitations. The backtest employed a simplified PnL proxy and basic transaction cost model, not fully capturing the complexities of options trading (e.g., bid-ask spreads, contract selection, Vega dynamics). The strategy logic (simple threshold crossing) and risk management (fixed stop-loss) could be further optimized. The dataset, while incorporating diverse sources, covers a specific market period (2021-2025) and results may not generalize to different market regimes. Furthermore, the GARCH baseline, while standard, could be expanded to include exogenous regressors (GARCH-X) for a potentially stronger comparison.

- **Future Work**

  Building upon these findings, future research should prioritize implementing and rigorously backtesting specific volatility trading strategies using **options data** (e.g., simulating straddle trades on Deribit), incorporating **realistic Vega exposure**, dynamic notional sizing, and comprehensive options market transaction costs. Investigating different prediction horizons (e.g., weekly IV changes) where the signal-to-noise ratio might differ is also warranted. Additionally, exploring more sophisticated modeling techniques (e.g., deep learning sequence models like LSTMs or Transformers) or incorporating regime-detection mechanisms could potentially enhance predictive accuracy.

---

5. Conclusion

- This paper investigated the predictability of 1-day changes in Bitcoin's Implied Volatility (DVOL) using an interdisciplinary machine learning approach. We demonstrated that by integrating on-chain, macroeconomic, sentiment, and technical features (including CWT) within an XGBoost framework, it is possible to achieve statistically significant predictive performance (**out-of-sample R^2 = 0.0527**), surpassing a traditional GARCH(1,1) baseline which failed ($R^2 < 0$). Feature analysis highlighted the importance of combining diverse data sources, with price momentum, on-chain sentiment, and macro factors emerging as key drivers.

- However, translating this statistical edge into economic profit proved challenging under our simplified backtesting methodology. A risk-managed trading strategy, while successfully limiting drawdown to **-5.00%**, yielded a **negative Sharpe Ratio (-13.40)** due to the predictive signal's magnitude being insufficient to overcome transaction costs when using a spot-price normalized PnL proxy.

- Our research confirms that a predictive signal for Bitcoin IV changes exists and can be captured using sophisticated machine learning techniques applied to diverse datasets. Nonetheless, the practical extraction of alpha appears highly sensitive to transaction costs and the specific PnL calculation methodology. Future work should focus on exploring more realistic options-based PnL models (e.g., Vega scaling) and potentially different prediction horizons to determine if this statistically significant edge can be converted into a consistently profitable trading strategy. Ultimately, our findings underscore the critical distinction between statistical predictability and economic viability in quantitative finance research.

References

1. [1] FRED, Federal Reserve Economic Data, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/ *(Accessed 28 August 2025)*.
2. [2] Coin Metrics. *Network Data*. https://coinmetrics.io/ *(Accessed 27 August 2025)*
3. [3] Deribit Insights. *Deribit Indices: DVOL*. *https://www.deribit.com/statistics/BTC/volatility-index* *(Accessed 10 October 2025)*.
4. [4] Binance API Documentation. https://binance-docs.github.io/apidocs/spot/en/ *(Accessed 28 August 2025])*
5. [5]Daniel Loureiro, Francesco Barbieri, Leonardo Neves, Luis Espinosa Anke, Jose Camacho-Collados(2022). TimeLMs: Diachronic Language Models from Twitter. arXiv preprint arXiv:2202.03829.
6. [6] Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. arXiv:1603.02754.
7. [7] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, Masanori Koyama (2019). *Optuna: A next-generation hyperparameter optimization framework*.arXiv:1907.10902
8. [8] Scott Lundberg, Su-In Lee (2017). *A unified approach to interpreting model predictions*.arXiv:1705.07874
9. [9] Kevin Sheppard. *ARCH Documentation*. https://arch.readthedocs.io/
10. [10] Gregory, R., et al. *PyWavelets - Wavelet Transforms in Python*. https://pywavelets.readthedocs.io/
11. [11] Google Trends. *Bitcoin Search Interest* https://trends.google.com/trends/explore?q=Bitcoin (Accessed 30 August 2025]).
12. [12] Google Cloud Public Datasets Program. Bitcoin Blockchain Dataset. Accessed via Google BigQuery, [https://console.cloud.google.com/bigquery?project=bigquery-public-data&ws=!1m4!1m3!3m2!1sbigquery-public-data!2scrypto_bitcoin] (Accessed 03 September 2025]).