

Article

Hybrid Regime-Switching Models for Cryptocurrency Prices: An Asset-Dependent Performance Analysis Using Markov Chains and Random Forests

Steve Karam ^{1,*} , Joseph El Maalouf ² and Nadine Dirani ¹

¹ College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait; nadine.dirani@aum.edu.kw

² School of Science—Dubai Branch, Saint Joseph University of Beirut, Beirut 345008, Lebanon; joseph.maalouf2@usj.edu.lb

* Correspondence: steve.karam@aum.edu.kw

Abstract

This study develops a leakage-free hybrid Markov–Random Forest framework for cryptocurrency price forecasting and evaluates it on Bitcoin and Ethereum. Daily OHLCV features are lagged by one trading day to prevent look-ahead bias, while regime labels are assigned from observed price changes using a two-state Markov chain with increasing and decreasing states. Regime-specific Random Forest models are then tuned independently via time-series cross-validation, allowing the predictive structure to adapt to regime-specific market conditions. The empirical results exhibit clear asset dependence. For Ethereum, the hybrid model outperforms the standalone Random Forest on magnitude-based metrics, attaining lower MAE and RMSE while also delivering a modest improvement in directional accuracy. Regime-specific tuning further identifies distinct optimal hyperparameter configurations across the increasing and decreasing states, suggesting that Ethereum’s upward and downward dynamics are structurally heterogeneous and can be better captured through regime-aware learning. By contrast, for Bitcoin, the standalone Random Forest delivers superior magnitude forecasting performance, while the regime-specific models differ only in tree depth and share the remaining tuning parameters, indicating that regime conditioning adds limited incremental value in a more persistent market. Statistical tests reinforce these findings. For Ethereum, Diebold–Mariano tests show that the hybrid significantly outperforms the standalone Random Forest under squared loss, while the absolute-loss comparison is only marginal. Across both assets, directional accuracy remains close to random chance, confirming the limited predictability of next-day price direction from lagged OHLCV features. Overall, the hybrid framework is most valuable when regime-specific dynamics are sufficiently distinct, offering improved forecasting performance and greater interpretability than a single global model.



Academic Editors: Jin-Ting Zhang and Tianming Zhu

Received: 14 May 2026

Revised: 24 June 2026

Accepted: 26 June 2026

Published: 30 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: Bitcoin; Ethereum; Markov regime switching; Random Forest; hybrid forecasting; cryptocurrency volatility

1. Introduction

Bitcoin and other cryptocurrencies have brought pronounced volatility and structural shifts to financial markets, with abrupt transitions between bullish and bearish regimes [1,2]. These assets display nonstationarity, fat-tailed distributions, and regime changes that challenge traditional time-series models, such as ARIMA and GARCH, that assume parameter

stability. Although deep learning methods often achieve strong predictive results [3], their opacity limits interpretability in financial applications.

Bitcoin, as a digital store of value, exhibits more persistent trading regimes than Ethereum, whose platform-driven dynamics show a more balanced split between increasing and decreasing states. This contrast motivates the question of whether regime-specific modelling adds value in one asset more than another, rather than assuming that a single forecasting strategy should work uniformly across cryptocurrencies.

Markov switching frameworks [4] provide a natural way to model regime transitions, while Random Forests [5] capture nonlinear relationships in lagged OHLCV features but do not explicitly encode regime dependence. A hybrid design that separates regime detection from prediction can therefore combine interpretability with nonlinear forecasting capacity.

To investigate this question, this paper develops a leakage-free hybrid Markov–Random Forest framework in which a two-state Markov chain identifies market regimes from daily price changes and regime-specific Random Forest models are tuned via time-series cross-validation. The empirical analysis compares Bitcoin and Ethereum to determine whether regime conditioning adds value beyond a single global nonlinear model and whether any benefits appear in magnitude forecasting, directional prediction, or both.

Our contributions include: separating regime detection from prediction in a leakage-free framework; identifying when regime conditioning improves a global nonlinear model; comparing Bitcoin and Ethereum to reveal asset-dependent behaviour; and evaluating magnitude and directional performance with statistical tests and regime diagnostics.

The paper proceeds with the Literature Review, Methodology, Empirical Results for Bitcoin and Ethereum, Discussion, and Conclusion.

2. Literature Review

2.1. Regime-Switching Models in Finance

Markov switching models form a cornerstone of financial econometrics for capturing structural breaks and regime changes [4]. Hamilton’s framework posits unobserved state transitions via Markov processes, characterising markets by distinct regimes. This framework has gained traction: Caporale et al. [6] identified high/low-volatility regimes in Bitcoin using MS-GARCH, while Bouri et al. [7] showed regime-switching factors outperform single-regime models across cryptos. Recent hybrids integrate ML, such as Agakishiev’s [8] HMM with reinforcement learning, improving transition forecasts. Yet regimes often remain latent within black-box architectures, obscuring their predictive role. Bitcoin shows persistent regimes due to maturity, whereas Ethereum’s distinct platform dynamics may yield different regime characteristics, an underexplored cross-asset gap.

2.2. Machine Learning for Cryptocurrency Forecasting

ML outperforms parametric models in crypto prediction by handling nonlinearities. Tree-based and neural methods demonstrate strong performance: Basher et al. [9] found RFs beat logits with 75–85% directional accuracy over multi-day horizons. Hamayel et al. [10] reported strong GRU point forecasts, sans regimes. Karam [11] showed RF/SVR topping LSTMs via simple OHLCV features, calling into question the complexity needs. Recent hybrid approaches have emerged that combine hidden Markov models with machine learning; for example, Tripathy et al.’s [12] HMM-LSTM and Karam’s [13] Monte Carlo RF for volatility. Still, interpretability is lagging behind, and Bitcoin dominates the cryptocurrency literature, while Ethereum’s platform-specific dynamics, including consensus mechanism changes and decentralised finance (DeFi) activity, have received comparatively less attention.

2.3. Research Gap and Contribution

Existing hybrids embed regimes opaquely [8,14], conflate point vs. directional metrics, and assume that regime switching universally benefits cryptocurrencies without testing asset-dependent effects or the conditions under which global nonlinear models might already suffice. We separate regime detection (Markov) from prediction (RF), test BTC/ETH performance to reveal asset-dependent effects, and prioritise directional accuracy, demonstrating when regime conditioning adds value beyond global models, advancing clarity, generalizability, and practical applicability.

3. Methodology

3.1. Data and Preprocessing

The empirical analysis is conducted using daily cryptocurrency market data for Bitcoin and Ethereum, two digital assets with distinct market structures and trading histories. For both assets, the raw dataset contains daily Open, High, Low, Close, and Volume series obtained from the *Yahoo Finance* API using the USD-quoted tickers BTC–USD (Bitcoin) and ETH–USD (Ethereum). All data were downloaded on 18 February 2026; no forward-filled or imputed values were used. The term “Price” in the empirical analysis refers to the daily Close, which serves as the one-step-ahead prediction target p_t . The Bitcoin dataset spans the period from 17 September 2014 to 30 November 2023, yielding a total of 3362 daily observations. This extended time horizon covers multiple market cycles, including early post-adoption volatility, the 2017 bull run, the subsequent 2018 correction, the 2020–2021 expansion, and the 2022–2023 contraction phase. Over the full sample, Bitcoin prices range from \$178 to \$67,567, with a mean of \$14,325 and a standard deviation of \$16,038, reflecting the pronounced volatility and regime variation characteristic of mature cryptocurrency markets.

The Ethereum dataset covers a shorter observation window reflecting a later stage of Ethereum’s market development, from 3 February 2021 to 3 February 2026, and consists of 1826 daily observations. This sample captures Ethereum’s post-maturation phase, including the decentralised finance expansion, the 2022 market downturn, and the subsequent recovery. Ethereum prices over this period range from \$993.64 to \$4831.35, with a mean of \$2567.41 and a standard deviation of \$896.40. Compared to Bitcoin, Ethereum exhibits a narrower price range, consistent with an asset whose market dynamics may still be maturing.

To eliminate look-ahead bias from intra-day features, the High, Low, Open, and Volume variables are lagged by one trading day prior to model training and evaluation. For each observation at time t , the model receives High_{t-1} , Low_{t-1} , Open_{t-1} , and Volume_{t-1} as inputs, reflecting values that are fully observed before the closing price p_t is realised. Calendar features (Day, Month, Year) are contemporaneous by nature and do not introduce leakage. This lagging procedure results in the loss of one observation per dataset, leaving 3361 observations for Bitcoin and 1825 for Ethereum after preprocessing.

To preserve the temporal structure of the data, an 80 percent training and 20 percent testing split is applied chronologically for each asset, without shuffling, ensuring that no future observations influence model training. For Bitcoin, this corresponds to 2689 observations used for model training over the period from September 2014 to January 2022, and 673 observations reserved for out-of-sample evaluation from January 2022 to November 2023. During the test period, Bitcoin prices range between \$15,787 and \$47,466, representing a wide spectrum of volatility conditions. An analogous chronological split is applied to the Ethereum dataset: 1460 observations for training from February 2021 to February 2025, and 366 observations for testing from February 2025 to February 2026, covering test prices from \$1473 to \$4831.

Feature construction combines lagged market information with simple temporal indicators. The predictive feature set includes the lagged Open, High, Low, and Volume variables (Open_{t-1} , High_{t-1} , Low_{t-1} , Volume_{t-1}), along with calendar-based features derived from the timestamp, namely, the day of the week, month, and year. Lagged Close (or “price”) values are not included as predictors, ensuring that the models do not exploit trivial autoregressive dependence from p_{t-1} to p_t and that the reported performance is obtained under a strictly leakage-free feature construction.

3.2. Markov Chain Regime Detection

The market regimes are identified directly from observed daily price movements using a first-order Markov chain. For each cryptocurrency, let p_t denote the observed price at day t . Regime classification is based on the sign of the one-period price change between consecutive days.

Specifically, the regime at time t is defined as

$$X_t = \begin{cases} I, & \text{if } p_t > p_{t-1}, \\ D, & \text{if } p_t < p_{t-1}, \end{cases}$$

where I and D denote increasing and decreasing market states, respectively. This classification rule relies solely on observable price changes and avoids the use of latent variables or additional distributional assumptions.

The resulting sequence of regimes $\{X_t\}$ is modelled as a first-order Markov chain, under the assumption that the probability of transitioning to a new regime depends only on the current regime. For each asset, the transition matrix P is estimated empirically from the training data by counting observed regime transitions, following the standard procedure in Hamilton [4,15]. The entry P_{ij} represents the probability of moving from regime i at time t to regime j at time $t + 1$, and is computed as

$$\hat{P}_{ij} = \frac{n_{ij}}{\sum_k n_{ik}},$$

where n_{ij} denotes the number of observed transitions from state i to state j .

The stationary distribution π is obtained by solving

$$\pi P = \pi, \quad \sum_i \pi_i = 1,$$

which yields the long run proportion of time spent in each regime. In addition, average regime returns are computed as the mean percentage price change conditional on the prevailing regime, allowing for an economic interpretation of each state.

3.3. Model Architectures

3.3.1. Pure Markov Chain Baseline

The pure Markov chain model serves as a structural baseline that isolates the predictive information contained in regime dynamics alone. This approach does not rely on explanatory variables or learning algorithms, but instead assumes that price evolution is governed by the long-run behaviour of the estimated Markov chain. Under the assumption of ergodicity, the stationary distribution $\pi = (\pi_I, \pi_D)$ represents the long-run proportion of time the market spends in each regime.

Let the one-period simple return be defined as

$$R_t = \frac{p_t - p_{t-1}}{p_{t-1}}.$$

For each regime $i \in \{I, D\}$, the regime-specific average return r_i is computed as the empirical conditional mean:

$$r_i = \mathbb{E}[R_t \mid X_t = i] \approx \frac{1}{N_i} \sum_{t: X_t=i} \frac{p_t - p_{t-1}}{p_{t-1}},$$

where N_i denotes the number of observations classified in regime i .

Using the stationary distribution and regime-specific returns, the expected multiplicative price change implied by the Markov chain is given by the stationary weighted average growth factor:

$$\sum_{i=1}^2 \pi_i (1 + r_i).$$

Price predictions are then generated by anchoring this expected growth factor to the first observed price in the test period, denoted p_{t_0} . The resulting forecast is defined as

$$\hat{p}_t = p_{t_0} \sum_{i=1}^2 \pi_i (1 + r_i) = p_{t_0} \left(1 + \sum_{i=1}^2 \pi_i r_i \right), \quad \forall t.$$

This formulation yields a constant price prediction over the test horizon and is explicitly designated as a structural lower bound rather than an operational forecasting model. Crucially, because this baseline generates a static trajectory, any implied directional tracking metrics do not reflect predictive capability or directional market-timing skill; instead, they mathematically represent the unconditional proportion of positive-return days within the out-of-sample test partition. Consequently, this structural bound is excluded from directional accuracy comparisons to prevent any misleading interpretation of forecasting competency. The primary purpose of this baseline is to isolate and quantify the explanatory power of unconditional regime parameters under stationarity, establishing a minimal baseline for magnitude-based tracking errors against which the machine learning and hybrid architectures can be strictly benchmarked.

As an additional naïve benchmark, a persistence baseline forecasts the next-day closing price as the most recently observed closing price, i.e., letting p_t denote the closing price at day t , the persistence forecast is defined as $\hat{p}_t = p_{t-1}$, so that the forecast at time t equals the observed closing price at time $t - 1$.

3.3.2. Random Forest Benchmark

A Random Forest regressor is employed as the machine learning benchmark, following the ensemble learning framework introduced by [5]. This model captures nonlinear relationships and interaction effects among market features without incorporating regime information, thereby serving as a purely data-driven comparison to the Markov-based approaches.

For each cryptocurrency, the Random Forest model is trained on the full training dataset using lagged Open, High, Low, and Volume (Open_{t-1} , High_{t-1} , Low_{t-1} , Volume_{t-1}) together with contemporaneous calendar features (Day, Month, Year) as inputs. The model hyperparameters are selected using `GridSearchCV` combined with five-fold time-series cross-validation implemented through `TimeSeriesSplit` in the `scikit-learn` library [16]. This procedure preserves temporal ordering by sequentially training on observations from periods one through k and validating on period $k + 1$, thereby avoiding look-ahead bias, which is particularly critical in financial forecasting. All the

Random Forest models and the associated `GridSearchCV` procedures are initialised with a fixed random seed (`random_state = 42`) to ensure deterministic and reproducible results across runs.

The hyperparameter search space includes the number of trees, maximum tree depth, and minimum sample requirements for node splitting. Specifically, the grid explores the number of trees in the range from 50 to 200; maximum depth values of 5, 10, 20; and unrestricted depth, minimum samples per leaf between 1 and 5, and minimum samples required to split an internal node between 2 and 5. All continuous features are standardised prior to training to ensure numerical stability.

The resulting optimised Random Forest model provides a strong nonlinear benchmark that reflects the predictive performance achievable through machine learning alone, without explicit modelling of regime dynamics. Feature importance measures are extracted after training based on the mean decrease in impurity across trees, allowing for an interpretable assessment of the relative contribution of each input variable. However, mean-decrease-in-impurity importance is known to exhibit upward bias toward continuous or high-cardinality predictors [17], so the microstructural conclusions drawn from these rankings should be interpreted as indicative rather than as unbiased estimates of each feature's true causal influence.

Table 1 reports the final hyperparameters selected by `GridSearchCV` for the globally tuned Random Forest benchmark models for Bitcoin and Ethereum.

Table 1. Globally tuned Random Forest hyperparameters selected via time-series cross-validation.

Asset	n_estimators	max_depth	min_samples_split	min_samples_leaf
Bitcoin	100	20	5	1
Ethereum	200	10	5	2

3.3.3. Hybrid Regime-Switching Model

The hybrid regime-switching framework can be summarised through the following procedural steps.

1. Input data. Collect daily price series $\{p_t\}$ and leakage-free feature vectors $\{x_t\}$ for the training and test periods, constructed as described in Section 3.1.
2. Regime identification. For each time index $t \geq 2$, assign a regime label $X_t \in \{I, D\}$ based solely on the price change $p_t - p_{t-1}$:

$$X_t = \begin{cases} I & \text{if } p_t > p_{t-1} \quad (\text{Increasing}), \\ D & \text{if } p_t \leq p_{t-1} \quad (\text{Decreasing}), \end{cases}$$

where exact ties ($p_t = p_{t-1}$) are conservatively assigned to D . The first observation ($t = 1$) is assigned to D as a neutral fallback. This yields a two-state regime sequence $\{X_t\}$ that depends only on past and current prices, introducing no future information.

3. Markov chain estimation. Using the labelled training sequence $\{X_t\}$, compute regime transition counts and estimate the 2×2 transition probability matrix $\mathbf{P}$:

$$\mathbf{P} = \begin{pmatrix} P_{II} & P_{ID} \\ P_{DI} & P_{DD} \end{pmatrix}, \quad P_{ij} = \Pr(X_{t+1} = j \mid X_t = i), \quad i, j \in \{I, D\}.$$

The two-state chain is ergodic by construction: both off-diagonal transition probabilities P_{ID} and P_{DI} are strictly positive for both assets, guaranteeing the existence and uniqueness of the stationary distribution π . Derive $\pi = (\pi_I, \pi_D)$ by solving $\pi\mathbf{P} = \pi$ subject to $\pi_I + \pi_D = 1$.

4. Regime-specific model training. For each regime $i \in \{I, D\}$, construct a training subset containing all observations at index $t \geq 3$ such that the *previous-day* regime satisfies $X_{t-1} = i$. A Random Forest regressor RF_i is trained on the corresponding feature vectors $\{x_t : X_{t-1} = i\}$ and closing price targets $\{p_t : X_{t-1} = i\}$. Hyperparameters for each RF_i are selected independently via `GridSearchCV` with time-series cross-validation on the regime-specific training subset, using the same parameter grid as the global Random Forest benchmark. This ensures that the only structural difference between the global Random Forest and each regime-specific model is the training data subset on which it is fitted, not the model class or tuning procedure. The resulting best hyperparameters (together with regime-specific CV RMSE values and the number of training observations per regime) are reported in Table 2.

Table 2. Regime-specific Random Forest hyperparameters.

Asset	Regime	max_depth	min_leaf	min_split	n_est	CV RMSE	Train n
Bitcoin	I (Increasing)	20	1	2	200	\$12,701.92	1455
	D (Decreasing)	None	1	2	200	\$13,260.53	1233
Ethereum	I (Increasing)	None	1	2	100	\$189.25	743
	D (Decreasing)	10	1	5	100	\$155.95	716

For Bitcoin, the increasing and decreasing regimes select different optimal parameters, with the increasing regime using `max_depth = 20` and the decreasing regime using `max_depth = None`; however, both regimes share the same `min_leaf = 1`, `min_split = 2`, and `n_est = 200`, confirming that the regime-specific models are structurally similar except for tree depth. For Ethereum, the two regimes also select distinct parameter configurations: the increasing regime favours unrestricted tree depth (`max_depth = None`), with a CV RMSE of 189.25; while the decreasing regime favours greater regularisation (`max_depth = 10`), with a CV RMSE of 155.95. This asymmetry confirms that upward and downward price movements in Ethereum exhibit genuinely different underlying dynamics that regime-specific tuning can exploit, and it documents the complete hybrid-model specifications required for reproducibility.

Algorithms 1 and 2 provide a formal procedural summary of the training and prediction phases of the hybrid framework described above.

Algorithm 1 Hybrid Markov–Random Forest: Training Phase

```

1: Input: Daily closing prices  $\{p_t\}_{t=1}^T$ ; daily raw features
      ( $\text{Open}_t, \text{High}_t, \text{Low}_t, \text{Volume}_t, \text{Day}_t, \text{Month}_t, \text{Year}_t$ );
      chronological training index set  $\mathcal{T}_{\text{train}}$ .

2: Step 1: Construct lagged features and targets without leakage.
3: for  $t = 2, \dots, T$  do
4:    $x_t \leftarrow (\text{Open}_{t-1}, \text{High}_{t-1}, \text{Low}_{t-1}, \text{Volume}_{t-1}, \text{Day}_t, \text{Month}_t, \text{Year}_t)$ 
5:    $y_t \leftarrow p_t$ 
6: end for
7: Restrict  $\{x_t, y_t\}$  to  $t \in \mathcal{T}_{\text{train}}$ .

8: Step 2: Assign regimes from realized prices using training data only.
9:  $X_1 \leftarrow D$ 
10: for  $t = 2, \dots, \max(\mathcal{T}_{\text{train}})$  do
11:   if  $p_t > p_{t-1}$  then
12:      $X_t \leftarrow I$  ▷ Increasing
13:   else
14:      $X_t \leftarrow D$  ▷ Decreasing or flat
15:   end if
16: end for
17: Regimes  $X_t$  depend only on  $\{p_1, \dots, p_t\}$  and are computed once from historical data.

```

Algorithm 1 *Cont.*

18: **Step 3: Estimate the Markov chain on training regimes.**

19: Compute transition counts

$$N_{ij} = \#\{t \in \mathcal{T}_{\text{train}} : X_t = i, X_{t+1} = j\}, \quad i, j \in \{I, D\}.$$

20: Set

$$P_{ij} = \frac{N_{ij}}{\sum_{k \in \{I, D\}} N_{ik}}.$$

21: Solve for the stationary distribution $\pi = (\pi_I, \pi_D)$ such that

$$\pi \mathbf{P} = \pi, \quad \pi_I + \pi_D = 1.$$

22: **Step 4: Train regime-specific Random Forest models.**

23: **for** each regime $i \in \{I, D\}$ **do**

24: $\mathcal{T}_{\text{train}}^{(i)} \leftarrow \{t \in \mathcal{T}_{\text{train}} : t \geq 3, X_{t-1} = i\}$

25: $X^{(i)} \leftarrow \{x_t : t \in \mathcal{T}_{\text{train}}^{(i)}\}$

26: $y^{(i)} \leftarrow \{y_t : t \in \mathcal{T}_{\text{train}}^{(i)}\}$

27: Fit RF_i^* on $(X^{(i)}, y^{(i)})$ using time-series GridSearchCV over the common hyperparameter grid.

28: **end for**

29: **Output:** Transition matrix $\mathbf{P}$, stationary distribution π , and fitted models RF_I^* and RF_D^* .

Algorithm 2 Hybrid Markov–Random Forest: Prediction Phase

1: **Input:** Fitted models RF_I^* and RF_D^* ; historical regimes $\{X_t\}_{t \in \mathcal{T}_{\text{train}}}$; daily prices and features for the test period; chronological test index set $\mathcal{T}_{\text{test}}$.

2: **Step 1: Extend regime assignment to the test period using only past prices.**

3: **for** $t = \max(\mathcal{T}_{\text{train}}) + 1, \dots, \max(\mathcal{T}_{\text{test}})$ **do**

4: **if** $p_t > p_{t-1}$ **then**

5: $X_t \leftarrow I$

6: **else**

7: $X_t \leftarrow D$

8: **end if**

9: **end for**

10: **Step 2: Generate leakage-free hybrid predictions.**

11: **for** each $t \in \mathcal{T}_{\text{test}}$ **do**

12: Use X_{t-1} , which is computed from information available up to time $t - 1$.

13: $x_t \leftarrow (\text{Open}_{t-1}, \text{High}_{t-1}, \text{Low}_{t-1}, \text{Volume}_{t-1}, \text{Day}_t, \text{Month}_t, \text{Year}_t)$

14: **if** $X_{t-1} = I$ **then**

15: $\hat{p}_t^{(\text{Hybrid})} \leftarrow \text{RF}_I^*(x_t)$

16: **else**

17: $\hat{p}_t^{(\text{Hybrid})} \leftarrow \text{RF}_D^*(x_t)$

18: **end if**

19: **end for**

20: **Output:** Hybrid predictions $\{\hat{p}_t^{(\text{Hybrid})}\}_{t \in \mathcal{T}_{\text{test}}}$.

3.4. Evaluation Metrics

Model performance is evaluated using a set of complementary metrics designed to capture both magnitude accuracy and directional effectiveness. This multi-metric approach reflects the distinction between traditional point forecasting objectives and the requirements of trading-oriented applications.

Magnitude prediction accuracy is assessed using the mean absolute error and root mean squared error. Mean absolute error is defined as

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|,$$

and provides a robust measure of average absolute deviation. Root mean squared error is given by

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2},$$

and penalises large prediction errors more heavily. In addition, the mean absolute percentage error is used to assess relative accuracy in a scale-independent manner, and is defined as

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{|y_t|}.$$

The coefficient of determination R^2 is also reported to quantify the proportion of variance in observed prices explained by each model.

Directional performance is evaluated using directional accuracy, which measures the proportion of periods for which the predicted price movement correctly matches the realised price movement [18]. Formally, directional accuracy is defined as

$$\text{DA} = \frac{1}{n-1} \sum_{t=2}^n \mathbb{I}(\text{sign}(y_t - y_{t-1}) = \text{sign}(\hat{y}_t - y_{t-1})),$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. This metric is particularly relevant for trading and decision-making contexts, where correctly predicting the direction of price changes is often more important than minimising absolute forecast errors.

To assess whether the observed performance differences between models are statistically meaningful, the Diebold–Mariano (DM) test [19] is applied to forecast error differentials under both squared loss (RMSE-based) and absolute loss (MAE-based), using a Newey–West HAC variance estimator [20] with bandwidth $h = \lfloor 4(n/100)^{2/9} \rfloor$ and a Bartlett kernel. The DM test statistic is

$$\text{DM} = \frac{\bar{d}}{\sqrt{\hat{V}_{\text{HAC}}(\bar{d})}},$$

where $\bar{d}$ is the sample mean of the loss differential $d_t = L(e_{A,t}) - L(e_{B,t})$ and $\hat{V}_{\text{HAC}}$ is the Newey–West variance of $\bar{d}$. Under H_0 of equal predictive accuracy, $\text{DM} \xrightarrow{d} \mathcal{N}(0, 1)$. For directional accuracy comparisons, a circular block bootstrap proportion test is used [21] with block size $b = \lfloor n^{1/3} \rfloor$ and $B = 5000$ resamples to preserve serial dependence. All test statistics and p -values are reported uniformly across all comparisons.

4. Empirical Results

4.1. Bitcoin

4.1.1. Regime Characteristics

The estimated Markov transition matrix for Bitcoin indicates a clear degree of regime persistence,

$$\hat{\mathbf{P}}_{\text{BTC}} = \begin{pmatrix} 0.5162 & 0.4838 \\ 0.5714 & 0.4286 \end{pmatrix}, \quad \hat{\boldsymbol{\pi}}_{\text{BTC}} = [0.5415, 0.4585],$$

These results show that increasing market conditions persist with a probability of 51.62 percent, while decreasing regimes remain active with a probability of 42.86 percent. Such persistence supports the relevance of modelling Bitcoin price dynamics through regime-dependent structures rather than assuming homogeneous behaviour over time.

The regime-specific average returns reveal a clear asymmetry between market states. During increasing regimes, the average daily return is $r_I = +2.59\%$, while decreasing regimes are associated with an average return of $r_D = -2.54\%$. This near symmetry in magnitude but opposite direction highlights the strong directional structure present in Bitcoin price movements and provides further justification for conditioning predictive models on the prevailing regime. The distribution of regimes in the training data, with 54 percent classified as increasing and 46 percent as decreasing, confirms that both regimes are well represented, ensuring reliable estimation of regime-specific dynamics.

4.1.2. Model Performance Comparison

Table 3 reports the out-of-sample performance of the models on the Bitcoin test set under the fully leakage-free two-state protocol. Among the learned models, the standalone Random Forest achieves the best magnitude forecasting performance, with an RMSE of \$1755.28 and an R^2 of 0.95, while the hybrid regime-switching model attains an RMSE of \$1956.34 and an R^2 of 0.93. The hybrid model also delivers a slightly higher directional accuracy than the standalone Random Forest, although the difference is marginal. The persistence baseline remains the strongest magnitude benchmark overall, while the Markov chain serves as a regime-only structural lower bound.

Table 3. Bitcoin test set performance comparison. Bold indicates the best value in each metric across the learned models.

Model	MAE (\$)	RMSE (\$)	MAPE (%)	R^2	Dir. Acc. (%)
Random Forest	1489.14	1755.28	5.84	0.95	49.55
Hybrid regime-switching	1634.30	1956.34	6.45	0.93	49.70
Persistence baseline	523.52	829.40	1.89	0.99	48.07

4.1.3. Statistical Significance

Table 4 reports the Diebold–Mariano test results for Bitcoin. Under squared loss, the Random Forest significantly outperforms the hybrid ($DM = 3.3774$, $p < 0.001$), and this pattern persists under absolute loss as well ($DM = 2.7692$, $p < 0.001$), indicating that the RF advantage is robust across loss functions and not driven by autocorrelated or heteroskedastic forecast errors. The block bootstrap tests for directional accuracy are not statistically significant ($p = 0.9458$ for hybrid vs. RF and $p = 0.5492$ for hybrid vs. persistence baseline), suggesting that neither model captures meaningful day-ahead directional signal for Bitcoin.

Table 4. Diebold–Mariano and block bootstrap test results for Bitcoin. The DM statistic uses Newey–West HAC variance (Bartlett kernel). The directional accuracy (DA) test uses a circular block bootstrap ($B = 5000$, $b = 8$). *** $p < 0.001$, where pp denotes percentage points, and n.s. denotes not significant.

Comparison	Loss/Test Type	Stat./Diff.	p -Value	Sig.
Hybrid vs. RF	Squared (RMSE)	3.3774	< 0.001	***
Hybrid vs. RF	Absolute (MAE)	2.7692	< 0.001	***
Hybrid vs. persistence baseline	Squared (RMSE)	11.1449	< 0.001	***
Hybrid vs. RF (DA)	Block bootstrap	0.1488 pp	0.9458	n.s.
Hybrid vs. persistence baseline (DA)	Block bootstrap	1.6369 pp	0.5492	n.s.

4.1.4. Visualisation and Diagnostics

Figure 1 compares point forecasts generated by the Random Forest and hybrid models over the Bitcoin test period. The standalone Random Forest closely tracks observed price

movements, reflecting its strong performance in magnitude prediction. The hybrid model produces smoother forecasts while remaining responsive to regime-dependent dynamics.

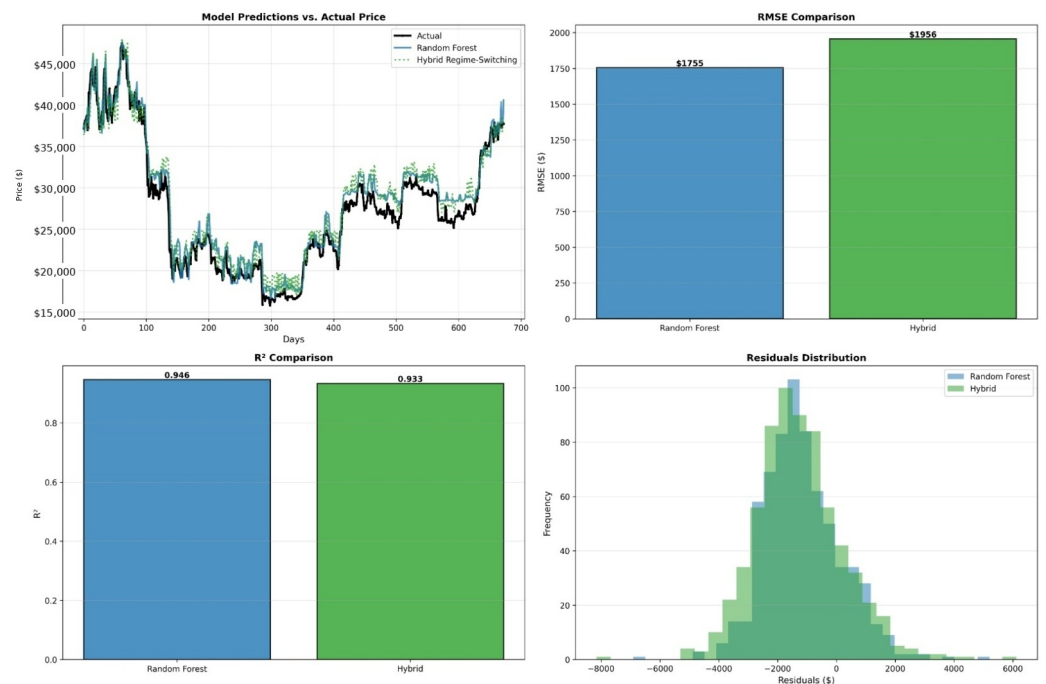


Figure 1. Point forecasts across models (Random Forest and hybrid).

Figure 2 provides a more detailed view of the hybrid model by decomposing its predictions according to the detected market regime. Distinct predictive behaviours are observed across increasing and decreasing states, illustrating the model's explicit regime awareness. This visualisation highlights how the hybrid framework adapts its forecasts based on prevailing market conditions rather than applying a single global prediction rule.

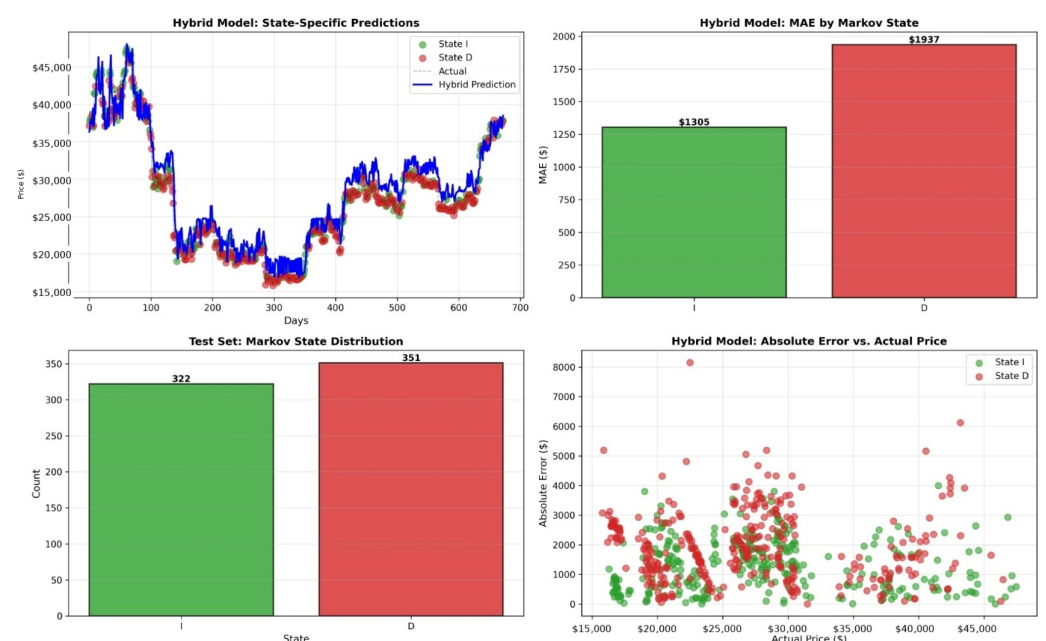


Figure 2. Hybrid regime-specific performance analysis.

Regime-specific error diagnostics further support these observations. A feature importance analysis reinforces the earlier findings, with the Low, High, and Open prices accounting for 35.848 percent, 48.471 percent, and 15.412 percent of total importance, respectively. Temporal features contribute negligibly, suggesting that price-level information dominates the forecasting process in Bitcoin markets.

4.2. Ethereum

4.2.1. Ethereum Regime Characteristics

The estimated Markov transition matrix for Ethereum is

$$\hat{\mathbf{P}}_{\text{ETH}} = \begin{pmatrix} 0.4738 & 0.5262 \\ 0.5469 & 0.4531 \end{pmatrix}, \quad \hat{\boldsymbol{\pi}}_{\text{ETH}} = [0.5096, 0.4904],$$

with self-persistence probabilities of 47.38% (I) and 45.31% (D), compared to Bitcoin's 51.6% (I) and 42.9% (D).

These estimates indicate that Ethereum exhibits moderate regime persistence overall, with a more symmetric balance between increasing and decreasing states than Bitcoin. The stationary distribution shows that Ethereum spends almost an equal amount of time in both regimes (approximately 51% vs. 49%).

4.2.2. Ethereum Model Performance

Table 5 reports the out-of-sample performance for Ethereum. The hybrid regime-switching model achieves the best overall magnitude performance across all metrics, with an MAE of \$99.17, RMSE of \$137.37, MAPE of 3.29%, and R^2 of 0.97, outperforming the standalone Random Forest (MAE: \$106.91, RMSE: \$150.11, MAPE: 3.57%, R^2 : 0.97) on all magnitude metrics. Directional accuracy for the hybrid (51.23%) modestly exceeds that of the Random Forest (50.96%), though both remain close to random chance, consistent with the general difficulty of day-ahead directional prediction across cryptocurrency markets.

Table 5. Ethereum test set performance comparison. Bold indicates the best value in each metric across the learned models.

Model	MAE (\$)	RMSE (\$)	MAPE (%)	R^2	Dir. Acc. (%)
Random Forest	106.91	150.11	3.57	0.97	50.96
Hybrid regime-switching	99.17	137.37	3.29	0.97	51.23
Persistence baseline	81.45	119.02	2.72	0.98	49.32

4.2.3. Ethereum Statistical Significance

Table 6 reports the statistical significance tests for Ethereum, evaluating both predictive error distributions and directional accuracy. Under squared loss (RMSE), the proposed hybrid framework significantly outperforms the standalone global Random Forest ($DM = -2.6061$, $p = 0.0221$), while the absolute-loss comparison is only marginally significant ($DM = -2.2481$, $p = 0.052$). These results suggest that the hybrid's magnitude forecasting gains are more robust under squared loss than under absolute loss.

In contrast, the hybrid does not outperform the persistence baseline under squared loss ($DM = 3.6509$, $p = 0.0002$), indicating that the persistence rule remains stronger on this criterion. Taken together, these comparisons indicate that the hybrid model is not uniformly superior across all benchmarks in the Ethereum experiment.

When directional forecasting is evaluated using the circular block bootstrap test with block size $b = 7$, the results are consistent with the Bitcoin findings: the observed directional gains are not statistically significant. Although the hybrid achieves a modest improvement in directional accuracy, this edge is statistically indistinguishable from both

the standalone Random Forest (+0.2740 percentage points, bootstrap $p = 0.9202$) and the persistence baseline (+2.19178 percentage points, bootstrap $p = 0.3534$). This suggests that, for Ethereum, regime conditioning improves magnitude prediction more reliably than it improves day-ahead directional timing.

Table 6. Diebold–Mariano and block bootstrap test results for Ethereum. The DM statistic uses Newey–West HAC variance (Bartlett kernel). The directional accuracy (DA) test uses a circular block bootstrap ($B = 5000$, $b = 7$). *** $p < 0.001$, * $p < 0.05$, where pp denotes percentage points, and n.s. denotes not significant.

Comparison	Loss/Test Type	Stat./Diff.	p -Value	Sig.
Hybrid vs. RF	Squared (RMSE)	−2.6061	0.0221	*
Hybrid vs. RF	Absolute (MAE)	−2.2481	0.052	n.s.
Hybrid vs. persistence baseline	Squared (RMSE)	3.6509	0.0002	***
Hybrid vs. RF (DA)	Block Bootstrap	0.2740 pp	0.9202	n.s.
Hybrid vs. persistence baseline (DA)	Block bootstrap	2.19178 pp	0.3534	n.s.

Overall, the Ethereum results show a clear separation between magnitude and direction: the hybrid model delivers statistically significant gains over the Random Forest in squared-loss error reduction, but not in directional accuracy, and it does not improve upon the persistence baseline under squared loss. This pattern is consistent with a market in which regime-specific structure can be exploited for better point forecasts, while the short-horizon direction remains difficult to predict from lagged OHLCV features alone.

4.2.4. Ethereum Visualisation and Diagnostics

Figure 3 compares the point forecasts generated by the Random Forest and hybrid models over the Ethereum test period. The Random Forest closely follows observed price movements, reflecting its strong magnitude-prediction performance. The hybrid model exhibits smoother dynamics that partially adapt to detected regime changes.



Figure 3. Random Forest and hybrid point forecasts on the ETH test period.

Figure 4 presents a four-panel diagnostic summary for Ethereum, including the actual-versus-predicted trajectories, RMSE and R^2 comparisons, and residual distributions for the benchmark and hybrid models. Overall, the figure shows that the hybrid model tracks the test series closely, while the error distributions remain comparable across models.



Figure 4. Analysis of hybrid's regime-specific performance.

4.3. Cross-Asset Comparative Analysis

Regime Persistence Comparison. Table 7 compares the estimated Markov persistence properties and forecasting performance across Bitcoin and Ethereum under the leakage-free evaluation protocol.

Table 7. Cross-asset comparison of Markov chain properties and model performance under the leakage-free evaluation protocol. Bold indicates the best RMSE value per asset.

Asset	P_{II}	P_{DD}	π_I	π_D	RF RMSE	Hybrid RMSE
Bitcoin	0.5162	0.4286	0.5415	0.4585	\$1755.28	\$1956.34
Ethereum	0.4738	0.4531	0.5096	0.4904	\$150.11	\$137.37

Interpretation: Bitcoin and Ethereum exhibit broadly similar but not identical regime persistence patterns. Bitcoin's average self-persistence, computed as the mean of P_{II} and P_{DD} , is 47.24 percent, while Ethereum's is 46.35 percent. Bitcoin exhibits stronger self-persistence in the increasing state, whereas Ethereum has a higher decreasing-state self-transition probability. Despite these similarities, the models behave differently across assets: the global Random Forest performs better for Bitcoin, while the hybrid achieves lower RMSE for Ethereum. This suggests that regime persistence alone is not sufficient to determine hybrid usefulness; the asymmetry of transitions and the extent to which a global nonlinear model already captures the signal also matter.

5. Discussion

The cross-asset evidence suggests that the usefulness of regime conditioning depends less on the mere presence of regime persistence and more on whether the predictive mapping from lagged features to prices changes materially across regimes. In other words, a hybrid architecture is most justified when the market exhibits regime-dependent nonlinear

structure that a single global model cannot represent efficiently. This perspective helps explain why the proposed framework does not deliver uniform gains across assets.

For Bitcoin, the results imply that regime labels derived from daily price changes add limited incremental information once lagged OHLCV inputs are already available to a global Random Forest. The similarity of the regime-specific model configurations indicates that the conditional relationship between predictors and next-day price is broadly stable across the two detected states. In such a setting, splitting the sample by regime may reduce effective training information without creating enough structural specialisation to offset that loss.

An important benchmark-design detail is that the persistence baseline is information advantaged relative to the learned models. By construction, it forecasts $\hat{p}_t = p_{t-1}$ and therefore has direct access to the lagged closing price, whereas the Random Forest and hybrid architectures are deliberately denied p_{t-1} in their feature sets to prevent trivial autoregressive dependence. Consequently, the strong magnitude performance of the persistence rule should be interpreted as a stronger-than-naïve reference rather than as evidence that the learned models are underperforming.

For Bitcoin, this asymmetry is particularly visible: the persistence baseline attains an RMSE of \$829.40, compared with \$1755.28 for the Random Forest benchmark (Table 3). This large gap primarily reflects the informational advantage of using the lagged closing price itself, not a failure of the leakage-free learned models, so the fact that the Random Forest and hybrid architectures remain competitive against such a favourable baseline understates rather than overstates their practical value.

Ethereum presents a different forecasting environment. The distinct regime-specific hyperparameters indicate that the increasing and decreasing states are associated with different levels of nonlinear complexity and regularisation needs, which is consistent with the idea that upward and downward market phases are generated by different local dynamics. For this asset, the hybrid model appears to function not simply as a regime classifier followed by two regressors, but as a mechanism for partitioning the forecasting problem into more homogeneous subproblems.

A second implication concerns the difference between magnitude forecasting and directional forecasting. Across both assets, the strongest improvements appear in error-based metrics such as MAE and RMSE rather than in directional accuracy, which remains close to chance even when magnitude performance improves. This suggests that the proposed framework is better understood as a tool for refining conditional expectation estimates than as a reliable mechanism for short-horizon market timing.

From a practical perspective, the results suggest that hybrid regime-switching models should not be adopted as default forecasting architectures for cryptocurrency markets. Their value appears to be conditional on observable evidence of regime-dependent predictive structure, such as materially different regime-specific validation outcomes or clearly differentiated model complexity across states. A useful modelling workflow is therefore to treat regime conditioning as a testable design choice rather than as an assumption, and to verify its value empirically on an asset-by-asset basis before deployment.

6. Limitations and Future Work

Several limitations should be acknowledged. First, regime identification is based on simple daily price changes, which ensures transparency but may overlook more nuanced intra-day dynamics or volatility driven regimes. Second, the analysis relies on a limited set of OHLCV and temporal features, which, while effective, may not fully capture macroeconomic effects, cross-asset spillovers, or order-flow information. Third, regime persistence is estimated over fixed training periods and assumed stable during testing, an assumption

that may not hold in rapidly evolving cryptocurrency markets. Future research could extend this framework by incorporating richer feature sets such as technical indicators, cross-asset signals, and on-chain metrics, as well as alternative regime definitions based on volatility or volume dynamics. Additionally, adaptive or rolling regime estimation and probabilistic regime assignment could improve robustness in environments with frequent structural change. Finally, expanding the analysis to a broader universe of cryptocurrencies would further clarify when regime-aware modelling provides consistent advantages and when simpler global models are preferable.

7. Conclusions

This study proposes a hybrid regime-switching framework that integrates Markov chain regime detection with asset-specific Random Forest regressors to enhance cryptocurrency price forecasting. Empirical analysis of Bitcoin and Ethereum demonstrates pronounced asset-dependent performance, with the hybrid model achieving superior magnitude forecasting for Ethereum under squared loss, while for Bitcoin the global Random Forest outperforms the hybrid, and the directional accuracy remains close to random chance for both models under the leakage-free protocol.

Principal Contributions. The findings delineate three core insights: first, the efficacy of regime conditioning derives principally from regime asymmetry rather than persistence magnitude alone; second, Random Forests excel in magnitude-based metrics for Bitcoin, whereas the hybrid yields superior magnitude forecasting for Ethereum when regime-specific dynamics are structurally distinct; third, feature importance underscores microstructural divergences, with Bitcoin exhibiting equilibrated OHLC contributions and Ethereum prioritising intra-day extremes.

Implications and Extensions. This framework equips practitioners with diagnostics for regime-utility assessment and informs researchers on reconciling directional versus point forecasting objectives. Prospective extensions encompass on-chain metric integration and dynamic regime adaptation amid market maturation.

Author Contributions: Conceptualization, S.K. and J.E.M.; methodology, S.K. and J.E.M.; software, S.K.; validation, S.K.; formal analysis, S.K.; investigation, S.K.; resources, S.K. and J.E.M.; data curation, S.K.; writing—original draft, S.K.; writing—review and editing, S.K., J.E.M. and N.D.; visualization, S.K.; supervision, S.K.; funding acquisition, S.K. and N.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: The data used in this study are publicly available from Yahoo Finance (<https://finance.yahoo.com>). Daily price data for Bitcoin (BTC-USD) and Ethereum (ETH-USD) were retrieved via the `yfinance` Python library and are freely accessible to the public. No new data were created in this study.

Conflicts of Interest: The authors declare no conflicts of interest

References

1. Baur, D.G.; Dimpfl, T. Asymmetric volatility in cryptocurrencies. *Econ. Lett.* **2018**, *173*, 148–151. [[CrossRef](#)]
2. Katsiampa, P. Volatility estimation for Bitcoin: A comparison of GARCH models. *Econ. Lett.* **2017**, *158*, 3–6. [[CrossRef](#)]
3. Liu, M.; Li, G.; Li, J.; Zhu, X.; Yao, Y. Forecasting the price of Bitcoin using deep learning. *Financ. Res. Lett.* **2021**, *40*, 101755. [[CrossRef](#)]
4. Hamilton, J.D. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* **1989**, *57*, 357–384. [[CrossRef](#)]
5. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]

6. Caporale, G.M.; Zekokh, T. Modelling volatility of cryptocurrencies using markov-switching garch models. *Res. Int. Bus. Financ.* **2019**, *48*, 143–155. [[CrossRef](#)]
7. Bouri, E.; Christou, C.; Gupta, R. Forecasting returns of major cryptocurrencies: Evidence from regime-switching factor models. *Financ. Res. Lett.* **2022**, *49*, 103193. [[CrossRef](#)]
8. Agakishiev, I.; Härdle, W.K.; Becker, D.; Zuo, X. Regime switching forecasting for cryptocurrencies. *Digit. Financ.* **2025**, *7*, 107–131. [[CrossRef](#)]
9. Basher, S.A.; Sadorsky, P. Forecasting bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility? *Mach. Learn. Appl.* **2022**, *9*, 100355. [[CrossRef](#)]
10. Hamayel, M.J.; Owda, A.Y. A novel cryptocurrency price prediction model using GRU, LSTM and bi-LSTM machine learning algorithms. *AI* **2021**, *2*, 477–496. [[CrossRef](#)]
11. Karam, S. On Integrating Technical Analysis with Machine Learning for Cryptocurrency Price Forecasting. *Comput. Math.* **2025**, *6*, 1168–1178. [[CrossRef](#)]
12. Tripathy, N.; Hota, S.; Mishra, D.; Satapathy, P.; Nayak, S.K. Empirical Forecasting Analysis of Bitcoin Prices: A Comparison of Machine learning, Deep learning, and Ensemble learning Models. *Int. J. Electr. Comput. Eng. Syst.* **2024**, *15*, 21–29. [[CrossRef](#)]
13. Karam, S. A Novel Hybrid Monte Carlo-Random Forest Framework for Improved Financial Predictions. *Comput. Math.* **2025**, *6*, 2570–2581. [[CrossRef](#)]
14. Figà-Talamanca, G.; Focardi, S.; Patacca, M. Regime switches and commonalities of the cryptocurrencies asset class. *N. Am. J. Econ. Financ.* **2021**, *57*, 101425. [[CrossRef](#)]
15. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 1994; Chapter 22.
16. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
17. Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [[CrossRef](#)] [[PubMed](#)]
18. Pesaran, M.H.; Timmermann, A. A simple nonparametric test of predictive performance. *J. Bus. Econ. Stat.* **1992**, *10*, 461–465. [[CrossRef](#)]
19. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [[CrossRef](#)]
20. Newey, W.K.; West, K.D. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* **1987**, *55*, 703–708. [[CrossRef](#)]
21. Politis, D.N.; Romano, J.P. The Stationary Bootstrap. *J. Am. Stat. Assoc.* **1994**, *89*, 1303–1313. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.