

Co-movement Without Leadership: Bitcoin's Regime-Level Influence on Altcoins, and Why Its Predictability Is Not Profitability

Abdullah¹

¹ COMSATS University Islamabad, Lahore Campus

Correspondence: sp23-bcs-003@cuilahore.edu.pk

JEL codes: G14 (Information and Market Efficiency), G17 (Financial Forecasting), C58 (Financial Econometrics), C45 (Neural Networks)

Keywords: Bitcoin; altcoins; trend regimes; volatility spillover; lead-lag; machine learning; Kolmogorov–Arnold networks; transaction costs; market efficiency

Abstract

Bitcoin–altcoin correlation is strong, but Bitcoin's short-horizon predictive lead is close to zero. Using two years (2024–2026) of hourly and 4-hour Binance data for Bitcoin and nine major altcoins, we separate these two claims — usually conflated — and locate where leadership actually lives. Contemporaneous return correlation averages 0.772, while the one-hour-ahead correlation is approximately 0.006, and Holm-corrected Granger tests leave no robust one-way hourly lead. Leadership exists instead at the trend-regime level: after a Bitcoin regime flip, altcoins that currently disagree with Bitcoin's regime converge to it within 24 hours at a rate 15.1 percentage points above a matched, disagreement-conditioned control (out-of-sample, $p < 0.001$). The effect is positive in five of five walk-forward folds, in fifteen of fifteen parameter specifications, in every yearly subsample, under non-overlapping events, and replicates on a second exchange in a second quote currency — yet trading it returns -0.343% per event after realistic costs. A four-architecture ensemble (LSTM, XGBoost, Transformer, KAN) trained on the same regime target reproduces the conclusion independently: 81% regime classification F1 decomposes into 52% directional accuracy and cost-negative signals. We further document a measured Goodhart effect in continuously trained systems: any metric used to accept model updates becomes unusable for ensemble calibration. Predictability and profitability are distinct properties, and classification metrics without economic decomposition systematically overstate cryptocurrency forecastability.

1. Introduction

Practitioner commentary on cryptocurrency markets rests on a folk theorem: Bitcoin leads, altcoins follow. The claim is quoted daily, drives the popular “altcoin season” rotation framework, and — despite a decade of academic work on cryptocurrency interdependence — has, per our searches, no direct peer-reviewed treatment. The academic literature establishes strong and time-varying co-movement (Aslanidis et al., 2019; Katsiampa, 2019), documents volatility spillovers with Bitcoin as a major source (Koutmos, 2018; Yi et al., 2018), and reports asymmetric responses to Bitcoin’s declines (Demir et al., 2021; Šíla et al., 2024). But co-movement, spillover and asymmetry are all statements about *contemporaneous or realized* relationships. The folk theorem is a statement about *prediction* — that observing Bitcoin tells you what altcoins will do next — and prediction is a different property requiring a different test.

This paper separates the two claims on 2024–2026 data and finds they have opposite answers. Bitcoin and nine major altcoins co-move strongly: the average contemporaneous hourly correlation is 0.772. Bitcoin’s predictive lead at the same horizon is essentially zero: the average one-hour-ahead correlation is 0.006, two orders of magnitude smaller, and after Holm–Bonferroni correction a battery of eighteen directional Granger tests retains two significant results pointing in opposite directions. At hourly resolution, the relationship is co-movement, not leadership.

Leadership does exist — one level of abstraction higher. We define a trend regime as the volatility-normalized sign of the EMA(12)–EMA(26) spread and ask an event-conditional question: after Bitcoin’s regime flips, does an altcoin whose regime currently *disagrees* with Bitcoin’s converge to the new regime within 24 hours? Against a matched control — the identical convergence question evaluated at moments when Bitcoin did not flip, under the identical disagreement restriction — the answer is yes: out-of-sample, the pooled follow rate is 48.8% against a control of 33.7%, a lift of +15.1 percentage points ($p < 0.001$, $n = 486$), positive for all nine coins. Regime transitions propagate; hour-to-hour returns do not. This reconciles the practitioner intuition with its absence from high-frequency return data: capital reallocation across assets takes hours — the median convergence lag is roughly half a day — so leadership is visible only in persistent state transitions, not in next-hour returns.

The propagation effect is unusually robust for this literature. It is positive in five of five expanding-window walk-forward folds, positive and significant in all fifteen cells of a parameter-specification grid, present in every yearly subsample, essentially unchanged when events are thinned to eliminate overlapping evaluation windows, and — replicated end-to-end on Coinbase USD pairs, a different exchange, matching engine, user base and quote currency — if anything slightly stronger (+16.5pp). And yet it is not tradeable: executing every propagation event in the implied direction and holding 24 hours returns −0.343% per event after a 0.30% round-trip cost, with only two of nine coins net-positive. The informative move is largely complete before a position taken on the signal can capture it.

A second, independent strand of evidence reproduces this conclusion from a different direction. A production forecasting system — a heterogeneous ensemble of LSTM, XGBoost, Transformer and Kolmogorov–Arnold network (KAN) architectures trained on the same regime target under a leak-free evaluation protocol — achieves 80.99% macro-F1 on held-out data, comfortably above its 71.06% persistence baseline. Decomposing those same signals economically: directional accuracy of issued BUY/SELL signals is 52.26%, barely above a coin flip, and the mean BUY signal is cost-negative. The classification metric and the economic value of the same predictions differ by an amount large enough to invert a deployment decision.

The paper makes four contributions. First, the co-movement/leadership separation itself, with the regime-propagation construct — an event-conditional test against a matched, disagreement-conditioned control, validated under a pre-registered out-of-sample gate — which per our searches has no precedent in the regime-switching or spillover literatures. Second, a metric decomposition of the same signals into classification performance and economic value, quantifying how much a reported F1 overstates tradeability; prior work has shown that predictability need not imply profitability for single-asset Bitcoin strategies (Sebastião & Godinho, 2021; Bysik & Ślepaczuk, 2026; Arain & Snudden, 2026), while cross-asset predictability has recently been reported as tradable (Kurihara & Matsumoto, 2026; Guo et al., 2024) — we extend the negative result to cross-asset regime propagation and to the classification-metric decomposition, engaging that positive prior art directly on horizon, turnover and cost assumptions. Third, era replications with the original literature’s own machinery: on 2024–2026 data, NARDL finds the celebrated crash-asymmetry effect (Demir et al., 2021) largely absent, and a rolling Diebold–Yilmaz exercise shows Cholesky-identified spillover “dominance” to be nearly unidentified under contemporaneous correlation of 0.77 — the answer is decided by variable ordering, not economics. Fourth, a measured Goodhart effect in continuously trained deployment systems: when model updates are accepted by a validation-split gate, the gated metric inflates while true out-of-time performance stands still, and ensemble weights calibrated on that metric invert against true model quality. We have not found this stated for financial machine-learning ensembles.

Throughout, negative economic results are reported as findings rather than suppressed. Claims of profitable cryptocurrency strategies are common and weakly evidenced; a robustness-checked effect reported *with* its economic failure is, we argue, the more credible contribution.

2. Related literature

The literature on cryptocurrency interdependence divides into four strands, each answering a different question about Bitcoin’s role. All sources below were verified against the Crossref DOI registry, with findings taken from the papers’ own abstracts and text rather than secondary summaries.

Interdependence and correlation. Aslanidis et al. (2019) apply a generalized DCC model and find correlations among cryptocurrencies positive but substantially time-varying, while correlations with traditional assets are negligible — crypto moves as a self-referential bloc. Katsiampa (2019) documents BTC–ETH conditional volatility interdependence responsive to news using a diagonal BEKK model, noting Ether can act as a hedge against Bitcoin. Ji et al. (2019) and Bouri et al. (2021) extend the connectedness picture dynamically and across quantiles, finding integration that strengthens in the tails.

Price leadership. Ciaian et al. (2018) study seventeen currencies over 2013–2016 and find the Bitcoin–altcoin relationship significantly stronger in the short run than the long run, with macro-financial indicators dominating altcoin price formation at longer horizons. Sifat et al. (2019) apply VECM, Granger causality, ARDL and wavelet coherence to hourly and daily BTC–ETH data and report *bi-directional* causality — a caution against assuming one-way Bitcoin leadership.

Volatility spillover. The methodological foundation is the Diebold & Yilmaz (2012) variance-decomposition spillover index. Koutmos (2018) applies it to eighteen cryptocurrencies and concludes Bitcoin is the dominant contributor of return and volatility spillovers. Yi et al. (2018) qualify dominance with a LASSO-VAR network over fifty-two coins: mega-caps propagate the most shocks, but several low-profile coins are also significant net transmitters — Bitcoin is a major, not exclusive, source. Balcilar & Ozdemir (2023) extend to 2022 and find mixed evidence for behavioural spillover patterns, with stablecoins behaving distinctly.

Asymmetry. Demir et al. (2021), applying the NARDL framework of Shin et al. (2014) to ETH, XRP and LTC over 2015–2019, find that decreases in Bitcoin’s price affect altcoins more than increases, intensifying after the 2017 crash. Šíla et al. (2024) decompose volatility into good and bad components under a TVP-VAR and find downturn information spills over substantially faster than comparable surges. Bouri et al. (2019) document significant, time-varying herding that intensifies with uncertainty — a candidate behavioural mechanism.

Machine-learning forecastability and economic value. Sebastião & Godinho (2021) find machine-learning crypto strategies’ profitability fragile under changing market conditions; Bysik & Ślepaczuk (2026) show walk-forward Bitcoin ML strategies that beat statistical benchmarks fail to survive transaction costs. The older economic-value tradition (Fleming et al., 2001) and the efficiency literature (Fama, 1970; Urquhart, 2016) frame the question: statistical forecastability is neither necessary nor sufficient for economic value. On architectures, Shen & Wu (2025) apply Kolmogorov–Arnold networks to Bitcoin prediction; heterogeneous regime-classification ensembles containing a KAN, evaluated leak-free against a persistence baseline, are absent from the literature per our searches.

Cross-asset predictability and its tradability. Most directly related to this paper are three recent results on whether cross-cryptocurrency predictability is *tradable*. Kurihara and Matsumoto (2026) document lagged price transmission from Bitcoin to altcoins at high frequency and report trading strategies that exploit it profitably — a positive-on-tradability

finding our cost-adjusted, event-conditional result stands in tension with, and which we engage directly on horizon, event definition, turnover and cost assumptions. Guo et al. (2024) find that lagged returns of other cryptocurrencies predict a focal coin and build a cross-sectional long-short portfolio that remains profitable after transaction costs; their estimand — continuous lagged-return predictors with cross-sectional turnover — differs from our discrete, event-conditioned propagation of trend regimes. Arain and Snudden (2026) settle the single-asset case: across twelve Bitcoin indices, statistically significant forecast gains yield significant excess profits for only two, and only during large-swing periods. Together these papers bound the debate to which our cross-asset, cost-adjusted result contributes.

The gap. A systematic search found no peer-reviewed treatment of the “altcoin season” / Bitcoin-dominance rotation framework, despite its ubiquity in practitioner commentary; available material is confined to exchange marketing and trading blogs, which fail the verification standard adopted here. Two further gaps motivate this work: the dominant findings above were estimated on 2013–2019 data, predating spot-ETF approval and institutional participation; and existing studies rarely separate contemporaneous co-movement from predictive lead — distinct claims with very different practical implications.

3. Data and methodology

3.1 Data

Binance spot OHLCV for ten USDT-quoted pairs — BTC plus ETH, BNB, SOL, XRP, ADA, DOGE, AVAX, DOT, LINK — from 2024-07-30 to 2026-07-30: 4,378 aligned 4-hour candles and 8,758 1-hour candles. Series are inner-joined on timestamp so all cross-sectional comparisons use jointly observed candles. Cross-exchange replication uses Coinbase USD pairs for the same window (BNB excluded; it does not trade on Coinbase).

3.2 Causality discipline

Every rolling statistic used for classification is shifted by one period, so no observation contributes to classifying itself. Volatility regimes, shock thresholds and trend regimes are all computed from strictly trailing windows.

3.3 Trend-regime definition

The regime definition is shared with the production forecasting system of Section 5, so the research and the deployed models measure the same object:

```
score_t = (EMA12(close)_t - EMA26(close)_t) / close_t
vol_t   = σ(1-period log returns, 180 periods, lagged) × √6
z_t     = score_t / vol_t
regime  = uptrend if z > 0.20; downtrend if z < -0.20; else neutral
```

3.4 Six analyses

- (1) Rolling 30/90-day Pearson correlations; (2) OLS beta to Bitcoin, partitioned into calm and stress regimes at the 25th/75th percentiles of trailing Bitcoin volatility; (3) cross-correlation at 1–24 hour leads in both directions, plus pairwise Granger causality tested both ways; (4) an event study on Bitcoin 4-hour returns exceeding ± 2 trailing-30-day standard deviations, with responses partitioned by shock sign; (5) the regime-propagation analysis of Section 3.5; and (6) a VAR(4) forecast-error variance decomposition (Diebold & Yilmaz, 2012) on standardized squared returns.

Inference. 95% confidence intervals by percentile bootstrap (1,000 resamples), paired resampling where statistics are computed on joint series; proportions compared by two-proportion z-tests.

3.5 The propagation construct and its control

The propagation analysis asks: *after Bitcoin's regime flips, does a currently-disagreeing altcoin converge to Bitcoin's new regime within 24 hours?* The treatment rate is conditional on current disagreement, since agreement offers no convergence to observe.

The control design deserves emphasis, because it determined the substantive conclusion. A first implementation compared the treatment rate against the *unconditional* probability that the two regimes agree. That comparison is invalid: the unconditional rate is dominated by already-agreeing observations, which the treatment measure excludes by construction. It produced a mean lift of -0.222 with zero of nine coins positive — the exact opposite of the corrected result. The corrected design uses a **matched control**: the identical convergence question, evaluated on observations where Bitcoin did *not* just flip, under the identical disagreement restriction. This yields $+0.133$ mean lift with nine of nine coins positive (in-sample). Persistence affects both arms equally and is differenced out; what remains is the marginal effect of the Bitcoin flip itself. We foreground the inversion deliberately: the conclusion was determined entirely by control specification, not by the data, and both specifications are preserved in the repository's history.

3.6 Validation protocol

Because in-sample effects can reflect overfitting, the propagation result was subjected to a temporal hold-out: rates estimated on the first 60% of aligned observations, evaluated on the final 40%, with the matched control recomputed within each split so no information crosses the boundary. A pre-registered gate required (i) pooled out-of-sample lift significant at $p < 0.05$ and (ii) more than half of coins showing positive out-of-sample lift. Results failing the gate were not to be reported as effects (nor deployed to the companion system); the gate specifies in advance what would falsify the claim.

4. Results

4.1 Correlation structure

Statistic	Value
Mean full-sample BTC–altcoin correlation	0.732
Correlations positive	9 / 9
Mean 90-day rolling range (max – min)	0.320
Highest / lowest coupling	ETH 0.819 / XRP 0.655

Table 1. Correlation structure, 4h returns, 2024–2026. All bootstrap CIs exclude zero.

Consistent with Aslanidis et al. (2019): positive, large, and time-varying — the 0.32 average swing in the 90-day rolling correlation means a static number would misrepresent the relationship.

4.2 Lead–lag: the central distinction

Statistic	Value
Mean contemporaneous correlation	0.772
Mean 1-hour-ahead correlation (BTC leading)	≈ 0.006
Granger BTC → altcoin significant (uncorrected)	6 / 9
Granger altcoin → BTC significant (uncorrected)	4 / 9
Bi-directional	4 / 9

Table 2. Lead–lag structure, 1h data.

The two-order-of-magnitude gap between contemporaneous and one-hour-ahead correlation is the principal descriptive result: **the relationship is co-movement, not lead.** Bi-directionality partially replicates Sifat et al. (2019). The Granger battery is treated properly in Section 4.7 — correction for multiple testing nearly eliminates it.

4.3 Volatility spillover

Coin	BTC share of variance (%)
SOL	48.4
ETH	48.1
DOGE	33.5
LINK	31.8
BNB	24.1
AVAX	23.1
XRP	17.2
ADA	15.2

Coin	BTC share of variance (%)
DOT	14.0
Mean	28.4

Table 3. VAR(4) FEVD, standardized squared 4h returns, horizon 10.

Bitcoin’s own variance is 92.4% self-driven. On this sample Bitcoin explains a minority of altcoin volatility variance — meaningful, clearly the largest single external source, but not dominant. This contradicts Koutmos (2018) and supports the more qualified reading of Yi et al. (2018). Section 4.8 shows the contradiction is largely an identification artifact interacting with the sample era.

4.4 Event study: the asymmetry that wasn’t

Bitcoin 4-hour shocks beyond $\pm 2\sigma$: 109 up-shocks, 143 down-shocks. Mean altcoin return over the following 24 hours:

Coin	After BTC up (%)	After BTC down (%)	Asymmetry (pp)
XRP	+0.219	+0.886	+0.667
ETH	+0.023	−0.163	+0.139
BNB	−0.151	+0.060	−0.092
DOGE	−0.409	+0.126	−0.283
SOL	−0.587	−0.273	−0.314
LINK	−0.461	−0.074	−0.388
ADA	−0.637	+0.182	−0.455
AVAX	−0.589	−0.132	−0.457
DOT	−1.066	−0.255	−0.811
Mean			−0.222

Table 4. Post-shock 24h altcoin returns by Bitcoin shock sign.

Only two of nine coins show the larger downside response the literature predicts (Demir et al., 2021; Šila et al., 2024). The dominant pattern is mean reversion: altcoins tend to drift *down* after Bitcoin surges and partially rebound after Bitcoin crashes. Section 4.8 replicates this test with the original NARDL machinery and reaches the same conclusion — the asymmetry belongs to the earlier era.

4.5 Beta by volatility regime

Statistic	Value
Mean beta to BTC	1.274
Mean R^2	0.529
Mean (stress beta – calm beta)	−0.303

Statistic	Value
Coins with higher beta under stress	0 / 9

Table 5. OLS beta to Bitcoin under calm vs stress regimes.

Altcoins amplify Bitcoin's moves ($\beta > 1$), but the loading *falls* under stress for all nine coins — the direction opposite to a naïve herding prediction (Bouri et al., 2019), though beta and CSAD-based herding measure different quantities. A plausible reading: in turbulent periods altcoins develop more idiosyncratic variance (their own liquidations, narratives, unlocks), lowering fitted beta even while correlation stays high.

4.6 Regime propagation, out-of-sample

Coin	Follow rate	Control	Lift	p	n
LINK	0.660	0.404	+0.256	0.0006	50
ETH	0.625	0.467	+0.158	0.043	48
XRP	0.518	0.366	+0.153	0.027	54
ADA	0.418	0.274	+0.144	0.025	55
BNB	0.500	0.361	+0.139	0.039	58
SOL	0.480	0.343	+0.137	0.055	50
AVAX	0.434	0.300	+0.134	0.046	53
DOGE	0.455	0.328	+0.127	0.059	55
DOT	0.349	0.284	+0.065	0.276	63
Pooled	0.488	0.337	+0.151	<0.001	486

Table 6. Out-of-sample propagation: follow rate within 24h of a BTC regime flip vs matched control (final 40% of sample). Pooled $z = 6.61$.

The pre-registered gate is passed decisively: all nine coins positive out-of-sample, pooled $p < 0.001$. Notably, the out-of-sample mean lift (+0.146) slightly exceeds the in-sample estimate (+0.123) — the opposite of an overfitting signature. Convergence, when it happens, is fast: pooled across 550 convergence events, the lag quartiles are 4, 12 and 16 hours, with 47.5% of convergences complete within 8 hours.

4.7 Robustness

Walk-forward. Five expanding-window folds (does the single 60/40 split matter?):

Fold	Follow	Control	Lift	p	n
1	0.316	0.297	+0.020	0.659	117
2	0.629	0.336	+0.294	<0.001	89
3	0.532	0.331	+0.200	<0.001	126
4	0.421	0.349	+0.072	0.079	152

Fold	Follow	Control	Lift	p	n
5	0.446	0.324	+0.121	0.010	110

Table 7. Expanding-window walk-forward validation.

Five of five folds positive, three individually significant, mean +0.141. The effect varies in magnitude but never reverses sign.

Specification grid. Fifteen combinations ($\tau \in \{0.10 \dots 0.30\} \times \text{EMA} \in \{8/21, 12/26, 20/50\}$): 15/15 positive and 15/15 significant; lift range +0.086 to +0.178, mean +0.140. The production specification ($\tau=0.20$, EMA 12/26) sits mid-range — the headline is not a favourable parameter selection. Two features of the grid bear directly on the provenance concern of Section 7: EMA(12, 26) is the conventional MACD pairing, a specification that predates and is independent of this sample rather than a value tuned on it; and the production $\tau=0.20$ does not maximize the lift within the grid. Selection on the overlapping product data therefore cannot account for the effect — it is a property of the regime construct across the whole specification family, not of the particular parameters carried over from the deployed system.

Yearly subsamples. 2024 +0.142, 2025 +0.138, 2026 +0.118, all $p < 0.001$.

Independence and cross-exchange replication. Two further attacks: events thinned so no 24-hour evaluation window overlaps another (removing the dependence objection to the pooled z-test), and the entire 60/40 out-of-sample design replicated on Coinbase USD pairs — a different exchange, matching engine, user base and quote currency:

Sample	Events	Lift	p
Binance USDT — all events	486	+0.150	<0.001
Binance USDT — non-overlapping	375	+0.142	7×10^{-6}
Coinbase USD — all events (8/8 coins positive)	—	+0.165	<0.001
Coinbase USD — non-overlapping	—	+0.151	9×10^{-6}

Table 8. Independence thinning and second-exchange replication.

Multiple-testing correction of the Granger battery. Raw Granger p-values in Table 2 are minima over lags 1–6, uncorrected. Holm–Bonferroni across the eighteen-test battery collapses 10/18 significant results to **2/18** — and the two survivors point in opposite directions (BTC→DOT; XRP→BTC). The correction *strengthens* the paper’s central claim: there is no robust one-way hourly lead from Bitcoin.

Dominance-conditioned propagation. As a direct test of the practitioner “altseason” framework, we condition propagation on Bitcoin’s relative strength versus an equal-weighted altcoin index: lift is +0.152 when Bitcoin is gaining on altcoins and +0.108 when losing, both $p < 0.001$. Regime leadership strengthens modestly when capital rotates toward Bitcoin but does not disappear otherwise — to our knowledge the first quantitative content behind the heuristic.

4.8 Replications with the original literature’s methods

A referee’s natural objection is that Sections 4.3–4.4 contradict published findings using simpler machinery than the originals. We therefore re-ran the originals’ methods on our sample.

NARDL (the method of Demir et al., 2021), daily data 2024–2026, HAC errors. Short-run asymmetry is significant for **one of nine** coins (BNB, $p = 0.014$); point estimates lean toward decrease-dominance for six of nine. Using the original machinery on the modern era, the crash-asymmetry effect is largely absent — the contradiction with Demir et al. is attributable to the **sample era**, not to methodological simplification.

Rolling Diebold–Yilmaz (the framework of Koutmos, 2018), 180-day windows, Cholesky identification. With Bitcoin ordered first in the factorization, Bitcoin’s to-others spillover share is 54.8% and it is the top transmitter in 100% of windows; ordered last, 1.0% and 0%. With contemporaneous correlation near 0.77, Cholesky-identified “dominance” is close to unidentified: whoever is ordered first receives credit for the common shock. Our order-invariant full-sample estimate (28.4%, Table 3) lies inside these bounds. Claims about *which* coin dominates volatility transmission are highly sensitive to identification; generalized (order-invariant) decompositions should be treated as the minimum standard.

4.9 Economic significance: the decisive negative result

Trading each propagation event in the implied direction (long on a flip to uptrend, short on a flip to downtrend), holding 24 hours, at 0.15% per side:

Coin	n	Gross (%)	Net (%)	Win rate
ETH	82	+0.757	+0.457	0.476
LINK	71	+0.603	+0.303	0.451
BNB	76	+0.166	−0.134	0.539
DOGE	78	+0.062	−0.238	0.449
DOT	73	+0.039	−0.261	0.479
XRP	76	−0.137	−0.437	0.447
AVAX	64	−0.601	−0.901	0.438
SOL	66	−0.630	−0.930	0.379
ADA	70	−0.646	−0.946	0.457

Coin	n	Gross (%)	Net (%)	Win rate
Mean			-0.343	

Table 9. Per-event returns trading the propagation signal, 0.30% round-trip cost.

Only two of nine coins are profitable after costs and the mean is negative. The mechanism is visible in the lag distribution of Section 4.6: with a median convergence lag of roughly half a day, the informative move is largely complete before a position taken on the signal can capture it. The effect is real, robust — and too small and too slow to survive frictions.

5. The forecasting-system strand

A production forecasting system built on the same regime target provides an independent evidence strand. Four architectures — LSTM (48×1h sequences), XGBoost (~90 features), Transformer (96×4h sequences) and a Kolmogorov–Arnold network (19 features) — are trained under a leak-free protocol: scale-free features only, scalars fit on the training split alone, per-symbol temporal 70/15/15 splits with purge gaps, volatility-adaptive labels, and macro series lagged one day. Four system-level findings are relevant to the paper’s argument.

(a) The target, not the model, determines attainable performance. The same four architectures, features and pipeline, evaluated on two targets: raw 24-hour direction yields macro-F1 of 0.31–0.36 against a 0.33 chance floor — unpredictable for every architecture. The 24-hour trend-regime target yields 0.72–0.78. The forty-point swing came entirely from re-specifying *what is predicted*. Reported crypto “accuracy” is uninterpretable without the target definition.

(b) A heterogeneous ensemble beats its best member. On held-out test data: LSTM 0.7819, KAN 0.7663, XGBoost 0.7534, Transformer 0.7215, blended ensemble **0.8019** macro-F1. Because the regime target is ~71% persistent, every result is reported against the persistence baseline (“current regime continues”, F1 0.7106): the ensemble’s genuine skill margin is +9.1 points.

(c) Metric decomposition: 81% F1 is 52% direction. On 5,940 held-out predictions, the same signals score 80.99% regime macro-F1 and **52.26%** directional accuracy on issued BUY/SELL calls (coin flip = 50%). Mean forward 24h return of BUY signals after a 0.30% round trip: **-0.050%**. The F1 scores classification of a lagging, persistent state; direction is what a trade needs. Papers reporting only the former invite a ~28-point misreading. This mirrors Section 4.9 exactly: two independent analyses, one conclusion — predictability and profitability are distinct properties.

(d) A Goodhart effect in gated continuous learning. The system fine-tunes continuously, accepting an update only if it does not regress on a fixed out-of-time holdout. Three measured observations: (i) scored on a *rolling* window, a stable model appeared to decay (0.71 → 0.63) while its fixed-holdout score was unchanged — the yardstick moved, not the model. (ii) Because the acceptance gate selects on the validation split, accepted updates

inflated validation scores (KAN 0.756→0.865; Transformer 0.758→0.851) while holdout scores stood still; ensemble weights calibrated on the validation split then *inverted* against true quality — the best model by holdout (LSTM, 0.782) received 0.2% weight and the worst (Transformer, 0.722) received 33.5%. Re-calibrating on the holdout restored sane weights. In a continuously trained system, **any metric used for acceptance becomes unusable for calibration**. (iii) With the fixed-holdout gate in place, continuous learning genuinely improved the model out-of-time (LSTM 0.7066 → 0.7819). We have not found observation (ii) stated for financial ML ensembles.

6. Discussion

6.1 Reconciling the contradictions with the literature

Three findings contradict published work: Bitcoin’s spillover share is a minority, not dominant (vs Koutmos, 2018); crash asymmetry is largely absent (vs Demir et al., 2021; Šíla et al., 2024); and beta falls rather than rises under stress (vs the herding prediction of Bouri et al., 2019). Section 4.8 shows the first two contradictions survive replication with the original papers’ own machinery, which rules out methodological simplification as the explanation. The most plausible reading is structural change: the 2024–2026 market — post spot-ETF approval, with institutional participation, deeper derivatives markets and materially higher altcoin liquidity — differs from the 2013–2019 samples underpinning the canonical results. Under those conditions altcoins plausibly carry more idiosyncratic variance, which would simultaneously reduce Bitcoin’s spillover share, lower stress-beta, and permit the mean-reverting post-shock pattern of Table 4. Where our replication disagrees with the older papers, the disagreement is a finding about the era, not an error in either study.

6.2 Statistical significance without economic significance

The robustness evidence pulls in two directions, and the tension is the paper’s most defensible contribution. The propagation effect is unusually robust for this literature — positive in every fold, specification, subsample, and on a second exchange. Yet trading it loses 0.343% per event after costs. Both statements are true simultaneously, and reporting only one of them would mislead in either direction. The decomposition of Section 5(c) shows the same structure inside a single model’s outputs: a classification metric can be excellent while the tradeable content of the same signals is nil. This extends the single-asset cost results of Sebastião & Godinho (2021) and Bysik & Ślepaczuk (2026) to cross-asset regime propagation, and adds the metric-decomposition mechanism that explains *how* the overstatement arises. In the tradition of Fleming et al. (2001), forecast evaluation should report economic value; in crypto ML, we argue it must, because the gap between the two is systematically large. The finding is also consistent with an efficiency reading (Fama, 1970; cf. Urquhart, 2016): a predictable regime pattern whose exploitation cost exceeds its return can persist indefinitely without implying exploitable inefficiency.

6.3 What the co-movement/leadership separation buys

Practitioner reasoning of the form “Bitcoin just moved, so this altcoin will follow” is not supported: the hourly predictive correlation is ~ 0.006 and the corrected Granger battery is empty. But the intuition is not baseless — it is mis-located. Regime transitions propagate (+15.1pp out-of-sample, median lag about half a day); hour-to-hour returns do not. The reconciliation is economically sensible: capital reallocation across assets takes hours, so leadership appears only at the persistence horizon. This also explains why the effect resists monetization — by the time convergence is observable, most of the move has happened.

6.4 Honest deployment as a design property

Two candidate features were tested and *rejected* for the companion deployment: hourly BTC→altcoin prediction (predictive correlation ~ 0.006) and a crash-asymmetry signal (contradicted by our own data, Table 4). Both are published on the live system alongside their rejection evidence. Every published probability appears beside its matched control rate, since “49%” alone invites reading as a promise, whereas “49% against a 34% baseline” exposes the information content. Directional accuracy and after-cost expectancy are displayed beside the headline F1. We suggest honest presentation is an engineering property that can be specified and tested — refusing to serve untrained weights, surfacing baselines, publishing rejections — rather than a disclaimer paragraph.

7. Limitations

1. **Sample period.** Two years; findings may not generalize across market regimes.
2. **Stablecoin denomination.** All primary pairs are USDT-quoted; Balcilar & Ozdemir (2023) show stablecoins behave distinctly, so USDT dynamics are embedded in every series. The Coinbase USD replication partially addresses this.
3. **Granger causality is predictive, not structural.**
4. **Multiple testing.** Addressed for the Granger battery via Holm–Bonferroni; other marginal per-coin results warrant caution.
5. **Hold-out design.** The gate uses one 60/40 temporal split; the five-fold walk-forward of Table 7 addresses multiple origins, but folds are not independent samples.
6. **Regime-definition dependence.** Results are conditional on the EMA/ τ family; the 15-cell grid bounds sensitivity within that family, not outside it.
7. **Survivorship in universe selection.** The ten-coin universe comprises 2026’s major pairs studied backwards; all were top-30 at sample start, which bounds but does not eliminate the bias. Claims concern persistent major altcoins, not the full cross-section.
8. **Parameter provenance.** $\tau=0.20$ and EMA(12,26) were chosen for the forecasting system before the research phase on overlapping data, so the production specification is not independently pre-registered. Two facts bound the concern (Section 4.7): EMA(12,26) is the standard MACD pairing — a convention

independent of this sample, not a tuned value — and within the 15-cell grid the production specification sits mid-range rather than at the maximum, with all fifteen specifications positive and significant. The result is thus a property of the regime construct across the specification family, not of the specific parameters inherited from the product; but a specification chosen strictly out-of-sample would still be the stronger design.

9. **Event dependence.** Bitcoin flips can fall within one another's evaluation windows; the non-overlapping re-test (Table 8) addresses this directly.

8. Conclusion

On 2024–2026 data, Bitcoin and major altcoins co-move strongly and Bitcoin predicts almost nothing about the next hour. Bitcoin's leadership is real but lives at the trend-regime level, where it is robust to every falsification attempt we could construct — walk-forward, specification grid, yearly subsamples, independence thinning, a second exchange — and still is not worth trading after costs. A production-grade four-architecture ensemble on the same target reproduces the conclusion from inside the machine-learning frame: excellent classification metrics, coin-flip directional value. The general lesson we draw for cryptocurrency forecasting research is methodological: report the economic decomposition of every classification result, report baselines that absorb persistence, correct multiple tests, and treat any metric used for model acceptance as spent for calibration purposes. Predictability and profitability are distinct properties; conflating them flatters both models and markets.

Declarations

Data and code availability. All analyses are reproducible from the project repository (see Appendix A); Binance and Coinbase public market data, accessed 2026-07-30. [Repository link / Zenodo DOI to be added at submission.]

Ethics. Public market data; no human subjects.

AI-assistance disclosure. The authors used Claude (Anthropic) for code development, data analysis, literature search and drafting assistance; all analyses, claims and citations were verified by the authors, who take full responsibility for the content.

Conflicts of interest. The authors operate the non-commercial deployment described in Section 6.4. [Adjust if monetization status changes before submission.]

Appendix A. Reproduction

```
.venv/Scripts/python.exe research/btc_effects.py          -> logs/btc_research
h.json              (Tables 1–5, in-sample propagation)
```

```
.venv/Scripts/python.exe research/validate_propagation.py -> logs/propagation_validation.json (Table 6)
.venv/Scripts/python.exe research/robustness.py -> logs/robustness.json (Tables 7, 9; grid; subsamples)
.venv/Scripts/python.exe research/original_methods.py -> logs/original_methods.json (Section 4.8; Holm; lag distribution)
.venv/Scripts/python.exe research/bulletproof.py -> logs/bulletproof.json (Table 8)
python diagnose_tradeability.py -> logs/tradeability.json (Section 5c)
python _stage_eval_ensemble.py -> logs/ensemble_eval.json (Section 5b)
```

References

All 44 references verified against the Crossref DOI registry (verification dates and citation counts in research/bibliography.json). Every entry follows the same protocol — no unverified citation. The 2026-09 expansion (20 → 41) added the economic-value, ML-in-finance, concept-drift, regime-switching, market-efficiency and multiple-testing strands; of those 21, twelve carry Semantic Scholar-verified abstracts and nine are DOI/metadata-verified only (abstracts publisher-elided in the indices). Three further entries — Kurihara & Matsumoto (2026), Guo et al. (2024) and Arain & Snudden (2026) — were added from the 2026-09 novelty check as the closest prior art on cross-asset predictability and its tradability, each Crossref-verified.

- Akyildirim, E., Goncu, A., Sensoy, A. (2020). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-020-03575-y>
- Antonakakis, N., Chatziantoniou, I., Gabauer, D. (2019). Cryptocurrency market contagion: Market uncertainty, market complexity, and dynamic portfolios. *Journal of International Financial Markets, Institutions and Money*. <https://doi.org/10.1016/j.intfin.2019.02.003>
- Arain, Snudden (2026). When are statistical forecast gains economically relevant? Evidence from Bitcoin returns. *Journal of Forecasting*, 45(3), 1245–1260. <https://doi.org/10.1002/for.70077>
- Ardia, D., Bluteau, K., Rüede, M. (2019). Regime changes in Bitcoin GARCH volatility dynamics. *Finance Research Letters*. <https://doi.org/10.1016/j.frl.2018.08.009>
- Aslanidis, N., Bariviera, A.F., Martínez-Ibañez, O. (2019). An analysis of cryptocurrencies conditional cross correlations. *Finance Research Letters*. <https://doi.org/10.1016/j.frl.2019.04.019>
- Balcilar, M., Ozdemir, H. (2023). On the risk spillover from Bitcoin to altcoins: The fear of missing out and pump-and-dump scheme effects. *Journal of Risk and Financial Management*, 16(1), 41. <https://doi.org/10.3390/jrfm16010041>
- Bariviera, A.F. (2017). The inefficiency of Bitcoin revisited: A dynamic approach. *Economics Letters*. <https://doi.org/10.1016/j.econlet.2017.09.013>

- Bouri, E., Gupta, R., Roubaud, D. (2019). Herding behaviour in cryptocurrencies. *Finance Research Letters*. <https://doi.org/10.1016/j.frl.2018.07.008>
- Bouri, E., Saeed, T., Vo, X.V., Roubaud, D. (2021). Quantile connectedness in the cryptocurrency market. *Journal of International Financial Markets, Institutions and Money*. <https://doi.org/10.1016/j.intfin.2021.101302>
- Bysik, M., Ślepaczuk, R. (2026). Machine learning-based Bitcoin trading under transaction costs: Evidence from walk-forward forecasting. arXiv preprint (q-fin), submitted 2026-05-19.
- Campbell, J.Y., Thompson, S.B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4), 1509–1531. <https://doi.org/10.1093/rfs/hhm055>
- Caporale, G.M., Zekokh, T. (2019). Modelling volatility of cryptocurrencies using Markov-Switching GARCH models. *Research in International Business and Finance*. <https://doi.org/10.1016/j.ribaf.2018.12.009>
- Ciaian, P., Rajcaniova, M., Kancs, d'A. (2018). Virtual relationships: Short- and long-run evidence from BitCoin and altcoin markets. *Journal of International Financial Markets, Institutions and Money*. <https://doi.org/10.1016/j.intfin.2017.11.001>
- Corbet, S., Meegan, A., Larkin, C., Lucey, B., Yarovaya, L. (2018). Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*. <https://doi.org/10.1016/j.econlet.2018.01.004>
- Della Corte, P., Sarno, L., Tsiakas, I. (2009). An economic evaluation of empirical exchange rate models. *The Review of Financial Studies*, 22(9), 3491–3530. <https://doi.org/10.1093/rfs/hhn058>
- Demir, E., Simonyan, S., García-Gómez, C.D., Lau, C.K.M. (2021). The asymmetric effect of bitcoin on altcoins: Evidence from the nonlinear autoregressive distributed lag (NARDL) model. *Finance Research Letters*. <https://doi.org/10.1016/j.frl.2020.101754>
- Diebold, F.X., Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2011.02.006>
- Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417. <https://doi.org/10.2307/2325486>
- Fischer, T., Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*. <https://doi.org/10.1016/j.ejor.2017.11.054>
- Fleming, J., Kirby, C., Ost diek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, 56(1), 329–352. <https://doi.org/10.1111/0022-1082.00327>
- Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 44. <https://doi.org/10.1145/2523813>

- Gu, S., Kelly, B., Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273. <https://doi.org/10.1093/rfs/hhaa009>
- Harvey, C.R., Liu, Y., Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1), 5–68. <https://doi.org/10.1093/rfs/hhv059>
- Jaquart, P., Dann, D., Weinhardt, C. (2021). Short-term bitcoin market prediction via machine learning. *The Journal of Finance and Data Science*, 7, 45–66. <https://doi.org/10.1016/j.jfds.2021.03.001>
- Guo, Sang, Tu, Wang (2024). Cross-cryptocurrency return predictability. *Journal of Economic Dynamics and Control*, 163, 104863. <https://doi.org/10.1016/j.jedc.2024.104863>
- Ji, Q., Bouri, E., Lau, C.K.M., Roubaud, D. (2019). Dynamic connectedness and integration in cryptocurrency markets. *International Review of Financial Analysis*. <https://doi.org/10.1016/j.irfa.2018.12.002>
- Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters*. <https://doi.org/10.1016/j.econlet.2017.06.023>
- Katsiampa, P. (2019). Volatility co-movement between Bitcoin and Ether. *Finance Research Letters*. <https://doi.org/10.1016/j.frl.2018.10.005>
- Koutmos, D. (2018). Return and volatility spillovers among cryptocurrencies. *Economics Letters*. <https://doi.org/10.1016/j.econlet.2018.10.004>
- Krauss, C., Do, X.A., Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*. <https://doi.org/10.1016/j.ejor.2016.10.031>
- Kristoufek, L. (2018). On Bitcoin markets (in)efficiency and its evolution. *Physica A: Statistical Mechanics and its Applications*. <https://doi.org/10.1016/j.physa.2018.02.161>
- Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., Zhang, G. (2018). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12), 2346–2363. <https://doi.org/10.1109/TKDE.2018.2876857>
- Kurihara, Matsumoto (2026). Price transmission from Bitcoin to altcoins: High-frequency evidence and implications for trading strategy. *Asia-Pacific Financial Markets*. <https://doi.org/10.1007/s10690-026-09589-z>
- Nadarajah, S., Chu, J. (2017). On the inefficiency of Bitcoin. *Economics Letters*. <https://doi.org/10.1016/j.econlet.2016.10.033>
- Sebastião, H., Godinho, P. (2021). Forecasting and trading cryptocurrencies with machine learning under changing market conditions. *Financial Innovation*, 7, 3. <https://doi.org/10.1186/s40854-020-00217-x>
- Sezer, O.B., Gudelek, M.U., Ozbayoglu, A.M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90, 106181. <https://doi.org/10.1016/j.asoc.2020.106181>

- Shen, Y., Wu, C. (2025). The role of Guru investor in Bitcoin: Evidence from Kolmogorov-Arnold Networks. *Research in International Business and Finance*. <https://doi.org/10.1016/j.ribaf.2025.102789>
- Shin, Y., Yu, B., Greenwood-Nimmo, M. (2014). Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework. In *Festschrift in Honor of Peter Schmidt* (pp. 281–314). Springer. https://doi.org/10.1007/978-1-4899-8008-3_9
- Sifat, I.M., Mohamad, A., Mohamed Shariff, M.S.B. (2019). Lead-lag relationship between Bitcoin and Ethereum: Evidence from hourly and daily data. *Research in International Business and Finance*. <https://doi.org/10.1016/j.ribaf.2019.06.012>
- Šíla, J., Kočenda, E., Kristoufek, L., Kukačka, J. (2024). Good vs. bad volatility in major cryptocurrencies: The dichotomy and drivers of connectedness. *Journal of International Financial Markets, Institutions and Money*. <https://doi.org/10.1016/j.intfin.2024.102062>
- Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters*. <https://doi.org/10.1016/j.econlet.2016.09.019>
- Welch, I., Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4), 1455–1508. <https://doi.org/10.1093/rfs/hhm014>
- White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5), 1097–1126. <https://doi.org/10.1111/1468-0262.00152>
- Yi, S., Xu, Z., Wang, G.-J. (2018). Volatility connectedness in the cryptocurrency market: Is Bitcoin a dominant cryptocurrency? *International Review of Financial Analysis*. <https://doi.org/10.1016/j.irfa.2018.08.012>