

On Chain Activity and Seven Day Volatility Forecasts for Bitcoin and Ethereum

Hanyuan Yin

Research manuscript | 6 October 2026

Abstract

This study tests whether delayed blockchain activity improves forecasts of volatility over the next seven daily returns of Bitcoin and Ethereum beyond recent price information. A fixed Coin Metrics snapshot supplies 2,922 daily observations per asset for 2018 through 2025. Ridge and random forest models are fitted on 2018 through 2021 origins, selected on 2022 through 2023 origins and evaluated on 724 origins per asset in 2024 through 2025. The target is an unannualized daily return RMS over seven future observations. In the main specification, adding activity reduces Bitcoin Ridge MAE by 2.51 percent, with a 28 day moving block bootstrap 95 percent interval of -4.22 to 7.81 percent, while the Bitcoin forest MAE rises by 1.75 percent. Ethereum MAE rises by 20.00 percent for Ridge and 14.07 percent for the forest, with both paired intervals favoring the price models. A longer activity delay retains the adverse Ethereum result. Adding activity to a volume controlled Bitcoin Ridge model reduces MAE by 9.63 percent, although that forecast remains less accurate than the original price only model. Validation selects price only Ridge for both dashboard assets. The results support context dependent information value rather than a general forecasting improvement. The accompanying dashboard presents historical forecasts and validation calibrated error bands. Because the data contain provider revisions and the outcome uses daily rather than intraday returns, the evidence is a retrospective risk monitoring evaluation with limited real time and latent volatility interpretation.

Keywords Digital assets; blockchain activity; volatility forecasting; risk monitoring; temporal validation

1 Introduction

Risk monitoring requires an estimate of how widely returns may vary, rather than a prediction of whether an asset price will rise. For digital assets, public blockchains provide additional measurements of transactions, active addresses and transaction fees. These measurements might reveal changing demand for network use or congestion before that information is fully reflected in a price based volatility forecast. They can also respond to past price movements or unrelated changes in network mechanics. Their availability therefore creates an empirical question, not a reason to assume that adding them improves a risk model.

This study asks whether lagged on chain activity adds predictive information for the next seven daily returns of Bitcoin and Ethereum after controlling for recent return and volatility history. The focus is incremental forecasting value. A chain augmented model is compared with a price history model using the same algorithm, target, training period and validation rule. This pairing separates the contribution of the additional information set from the benefit of switching to a more flexible algorithm. Simple historical volatility and exponentially weighted volatility forecasts provide additional reference points.

The investigation uses a fixed public Coin Metrics snapshot covering 2018 through 2025. Models are estimated on 2018 through 2021 origins, selected on 2022 through 2023 origins and evaluated on 2024 through 2025 origins. Seven origin dates are removed before the validation and test boundaries so that a training or validation target does not extend into the next partition. The target is the root mean square of seven future daily log returns. It is a daily return measure of short horizon variability, rather than a volatility estimate constructed from intraday trades.

Three design choices make the study useful for a small research project. First, the input set is compact and available without a subscription or a proprietary order book. Second, each asset is modeled separately, allowing differences in network architecture and native fee units. Third, evaluation retains both unsuccessful and successful augmented models and checks a longer feature delay and an additional market volume control. The study consequently addresses whether a feasible public data implementation produces a stable improvement, rather than whether a sufficiently extensive search can produce a favorable backtest.

The contribution is an applied comparison with a reproducible research dashboard. It does not establish a causal effect of blockchain activity on prices. It also does not reconstruct the historical versions of the provider's data. The downloaded files contain revised historical measurements, so even correctly lagged features support a retrospective evaluation rather than a verified real time deployment record. This distinction remains visible in the dashboard and constrains the practical interpretation of the results.

2 Related research

Volatility forecasting commonly exploits persistence across several time scales. Corsi (2009) develops the heterogeneous autoregressive model of realized volatility, combining daily, weekly and monthly components. The present price feature set borrows the idea of multiple lookback horizons, but uses daily return squares and a regularized regression. It is therefore not an implementation of the original intraday realized volatility HAR model. Retaining this distinction avoids attributing the precision of a high frequency volatility measure to seven daily observations.

On chain activity has already been investigated as a volatility predictor. Gkillas, Tantoula and Tzagarakis (2021) study transaction activity and Bitcoin realized volatility with HAR and random forest forecasts, reporting improved out of sample accuracy when blockchain transaction counts are included. Their evidence motivates an incremental information test, while its target and data resolution differ from this study. Here the outcome is the RMS of seven future daily log returns, the sample includes Ethereum, and activity features are deliberately delayed. These choices test a narrower and more accessible implementation; they do not constitute a claim to the first use of blockchain activity in forecasting.

Ridge regression shrinks coefficients to reduce sensitivity to correlated predictors (Hoerl and Kennard, 1970). Random forests average regression trees and can represent nonlinear interactions (Breiman, 2001). Neither property guarantees that an additional data source is helpful. A feature set can improve a validation period while failing after a change in network usage or market conditions. Comparing both algorithms with and without the activity block therefore examines whether the measured benefit depends on the modeling choice.

Forecast evaluation also depends on the loss and dependence structure. Patton (2011) studies loss functions for evaluating volatility forecasts using imperfect volatility proxies. This study reports QLIKE on positive variance proxies alongside errors on the volatility scale, without claiming that its seven day proxy satisfies every assumption required for latent volatility forecast rankings. Adjacent targets share six returns. Moving block resampling, following the approach introduced by Künsch (1989), is used to retain local dependence in paired loss comparisons. Its intervals remain conditional on this fixed experiment and are not a substitute for evidence across new assets or future market regimes.

3 Data and information timing

The source is the official coinmetrics/data repository. Raw BTC and ETH CSV files were retrieved on the client date 6 October 2026 and retained with source URLs and SHA256 hashes. The study window is fixed at 1 January 2018 through 31 December 2025, with 2,922 daily records for each asset. In this window, dates are unique and continuous and the six retained core columns have no missing observations. No missing day is filled, no observation is winsorized and no future value is interpolated.

The two assets are selected for the explicitly scoped comparative question; the sample does not represent the full population of digital assets.

PriceUSD supplies the daily USD closing price. AdrActCnt counts active addresses, TxCnt counts blockchain transactions, TxTfrCnt counts transfers and FeeTotNtv measures aggregate fees in the native asset unit. The reported daily USD spot volume is reserved for a sensitivity analysis. Address activity is not a count of independent people. One person can control several addresses, and automated or protocol specific activity can change the count. Transactions and transfers follow different network definitions. Fees are not directly comparable across BTC and ETH units, and total Ethereum fees can include burned fees. The study therefore derives within asset fee per transaction features and fits separate asset models (Coin Metrics, 2026a, 2026b, 2026c).

The marked CSV day must be distinguished from the nominal forecast issue time. Coin Metrics documents PriceUSD for day t as the closing reference rate at 00:00 UTC on day t plus one. Accordingly, an origin labelled t uses prices through P_t and is recorded as a nominal forecast issued at the following midnight. The target uses returns with labels t plus one through t plus seven. Its final price is available at midnight on t plus eight. Chain features use values no later than the marked day t minus two in the main specification. This ordering prevents a contemporaneous price from being treated as known at the beginning of its marked day (Coin Metrics, 2026d).

The timestamp discipline does not recover real time vintages. The provider documents historical recalculations and data cleanses, and the source files contain completion fields consistent with later calculation of historical rows. A two or three day delay can impose a conservative order of observations, but cannot prove which version was visible to a user at the time. Results are described throughout as forecasts simulated from a revised historical snapshot (Coin Metrics, 2026e).

Table 1 Data fields and analysis roles

Source field	Interpretation	Role
PriceUSD	USD closing price	Return history and target
AdrActCnt	Active addresses	Delayed level and change
TxCnt	Blockchain transactions	Delayed level and change
TxTfrCnt	Blockchain transfers	Delayed level and change
FeeTotNtv	Native transaction fees	Fees per transaction
volume_reported_spot_usd_1d	Reported USD spot volume	Symmetric sensitivity control

Source: Coin Metrics Community data and metric documentation. Complete study window: 2,922 daily records per asset. Fields are interpreted within each network.

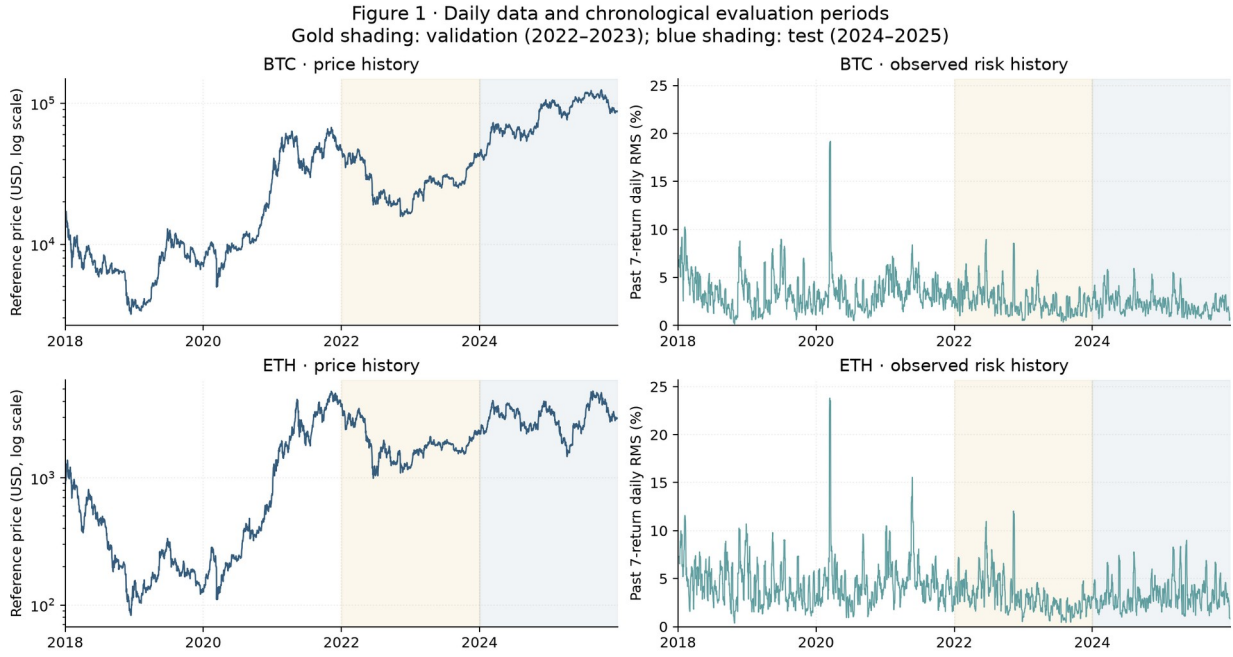


Figure 1 Historical prices and backward seven return RMS. Prices are on logarithmic USD axes with asset specific ranges; risk panels use a common percentage scale. Shading identifies validation and test calendar periods, while exact eligible origins and purges are given in Table 2. Source: Coin Metrics snapshot and archived calculations.

4 Forecast design

4.1 Target and units

Let P_t denote the closing price with CSV label t and define r_t as $\ln(P_t \text{ divided by } P_{(t-1)})$. For each forecast origin, the variance proxy v_t is one seventh of the sum of $r_{(t+h)}$ squared for h equal to one through seven. The forecasting target σ_t is the square root of v_t . It is expressed as an unannualized daily log return RMS over the coming seven observations. Multiplication by 100 gives the percentage units used in the figures and tables. A value of 2 percent does not mean a 2 percent predicted weekly loss, and no annualization factor is applied.

The target uses squared returns around zero rather than subtracting a seven day sample mean. With only seven daily observations it is a noisy measure of subsequent variability and does not capture intraday variation or tail loss probabilities. Models estimate $\ln(\max(\sigma_t, 10^{-8}))$ and positive volatility forecasts are obtained by exponentiation. No test based bias correction or recalibration is applied. This transformation stabilizes training, but exponentiating a log target prediction need not estimate the conditional mean of variance and can understate large volatility episodes.

4.2 Time partitions

The training, validation and test partitions are defined by the marked origin date, not by randomly sampled rows. Each partition contains only origins whose complete future target can be constructed from the fixed sample. Training and validation origin windows end on 24 December before the following boundary. Their final labelled target return is 31 December. The next partition begins with a 1 January origin, so targets do not overlap across partitions. The first training origin follows the thirty return warmup.

Table 2 Time partitions per asset

Partition	Origin dates	Origins	Final target label
Training	2018-02-01 to 2021-12-24	1423	2021-12-31
Validation	2022-01-01 to 2023-12-24	723	2023-12-31
Test	2024-01-01 to 2025-12-24	724	2025-12-31

Forecast issue time is the next midnight UTC after each origin label. Test origins comprise 366 in 2024 and 358 in 2025. The final scored forecast is nominally issued on 25 December 2025.

4.3 Information sets

The price information set contains five features: log RMS over the most recent one, seven and thirty returns, the most recent log return and the sum of the most recent seven log returns. Rolling windows end at the marked origin t . A volatility floor of 10^{-8} prevents taking the logarithm of zero. No price input uses an observation beyond the nominal post close issue time.

The activity block contains eight features. Active addresses, transactions and transfers are each transformed by $\log(1+x)$. Native fees per transaction are transformed by $\log(\max(\text{FeeTotNtv} \text{ divided by } \text{TxCnt}, 10^{-12}))$. Each of these four levels is paired with its difference from seven marked days earlier. All eight features are shifted by two days in the main experiment and by three days in a pre specified delay sensitivity. Using native fees avoids directly multiplying the activity measure by the USD asset price, while retaining the possibility that fees themselves respond to market conditions.

The volume control specification adds the $\log(1+x)$ level and seven day log change of reported daily USD spot volume, both delayed by two marked days. The same two volume features are included on both sides of each price versus activity comparison. This prevents a favorable activity result from being attributed solely to the availability of an additional generic measure of market activity. Reported volume is a provider measurement and is not an estimate of executable liquidity or order book depth.

4.4 Models and selection

Models are trained separately for each asset. The Historical7 benchmark repeats the backward seven return RMS. EWMA applies a fixed decay of 0.94 to past squared returns, using `adjust=False`, and takes the square root. These benchmarks do not receive information from the validation or test targets. The decay is a fixed conventional benchmark choice, not a parameter estimated from the test period.

Ridge and random forest regressions use the log volatility target. Ridge standardization is fitted only on the training features, and its penalty is selected from 0.1, 1, 10 and 100. The forest uses 300 trees, all candidate features per split and a fixed seed of 20261006. Maximum depth is selected from 4 and 8 and minimum leaf size from 10 and 30. Each algorithm and information set receives the same small validation grid and is selected independently by validation MAE on the untransformed volatility scale. This is a comparison of equally tuned model families, rather than identical fitted parameters. A deterministic grid order resolves ties.

All preprocessing and final fitted models remain based on the training partition. Validation selects parameters and supplies descriptive error bands, but is not appended for a final model refit. The test partition is never used to select features, parameters, the dashboard model or an error threshold. A time stamped protocol was saved before model results were generated. It is a local experiment record, not an externally registered pre analysis plan.

4.5 Losses and uncertainty

The primary forecast loss is MAE, the average absolute difference between forecast and observed RMS volatility. RMSE gives greater weight to large errors. Both are shown in percentage points of daily log

return RMS. The secondary QLIKE loss is the mean of v_t divided by predicted variance, minus the logarithm of that ratio, minus one. Predicted variance is the squared positive volatility forecast. QLIKE is dimensionless; its ranking is reported separately because the models are selected on volatility MAE, not on a variance optimal objective.

For each asset and algorithm, incremental improvement is 100 times the price model MAE minus the chain model MAE, divided by the price model MAE. Positive values favor the augmented information set. The primary comparison includes four pairs: Ridge and forest for BTC and for ETH. A 95 percent percentile interval is computed by resampling aligned pairs of test absolute losses in overlapping contiguous blocks of 28 origins, concatenating sampled blocks and truncating to the original length. There are 2,000 replicates with a fixed seed. Block lengths of 14 and 56 origins are sensitivity checks. The procedure is non circular, so blocks do not join the end of the sample to its beginning.

Block resampling acknowledges overlapping seven return targets and longer local persistence, but does not make the loss process stationary. Intervals are conditional on the selected fitted models and this historical sample; they exclude training and selection uncertainty and are not adjusted for four simultaneous comparisons. The delay and volume specifications and annual breakdowns are descriptive stability checks, not additional opportunities to select the best test result.

The dashboard shows an additional descriptive 90 percent error band. For each frozen model, its radius is the validation absolute error order statistic at rank $\text{ceil}(0.9 \text{ times } (n+1))$, capped at n . The lower endpoint is $\max(0, \text{forecast minus radius})$ and the upper endpoint is $\text{forecast plus radius}$. Empirical test coverage and width are reported. This is a historical error band calibrated on overlapping, potentially nonstationary validation observations. It does not provide a formal conformal or future coverage guarantee.

5 Results

Table 3 and Figure 2 report all six main test forecasts for each asset. Mean observed subsequent RMS is 2.270 percent for BTC and 3.325 percent for ETH. Each result uses 724 identical test origins. Price only Ridge has MAE of 0.771 percentage points for BTC and 1.222 for ETH. Both values are lower than the historical seven day and EWMA benchmark errors. These comparisons establish that the price feature baseline is informative before examining the activity block.

The lowest BTC test MAE is 0.751 percentage points for chain augmented Ridge. The same forecast has RMSE of 1.024, compared with 0.995 for price Ridge, and QLIKE of 0.476, compared with 0.355. Thus its small average absolute error advantage comes with worse performance under losses that respond differently to large errors or variance underprediction. Test MAE is an evaluation result and does not replace the validation model selection rule.

Table 3 Main test forecast errors

Asset	Model	MAE pp	RMSE pp	QLIKE
BTC	Historical 7 day	0.949	1.249	0.770
BTC	EWMA	0.808	1.039	0.380
BTC	Ridge price	0.771	0.995	0.355
BTC	Ridge plus chain	0.751	1.024	0.476
BTC	Forest price	0.803	1.037	0.386
BTC	Forest plus chain	0.817	1.035	0.359
ETH	Historical 7 day	1.548	2.113	1.199

Asset	Model	MAE pp	RMSE pp	QLIKE
ETH	EWMA	1.324	1.733	0.472
ETH	Ridge price	1.222	1.601	0.414
ETH	Ridge plus chain	1.467	1.802	0.421
ETH	Forest price	1.238	1.614	0.405
ETH	Forest plus chain	1.413	1.773	0.414

MAE and RMSE are percentage points of unannualized daily return RMS; QLIKE is dimensionless. Lower values are better. Main activity lag is two marked days. All rows have 724 test origins.

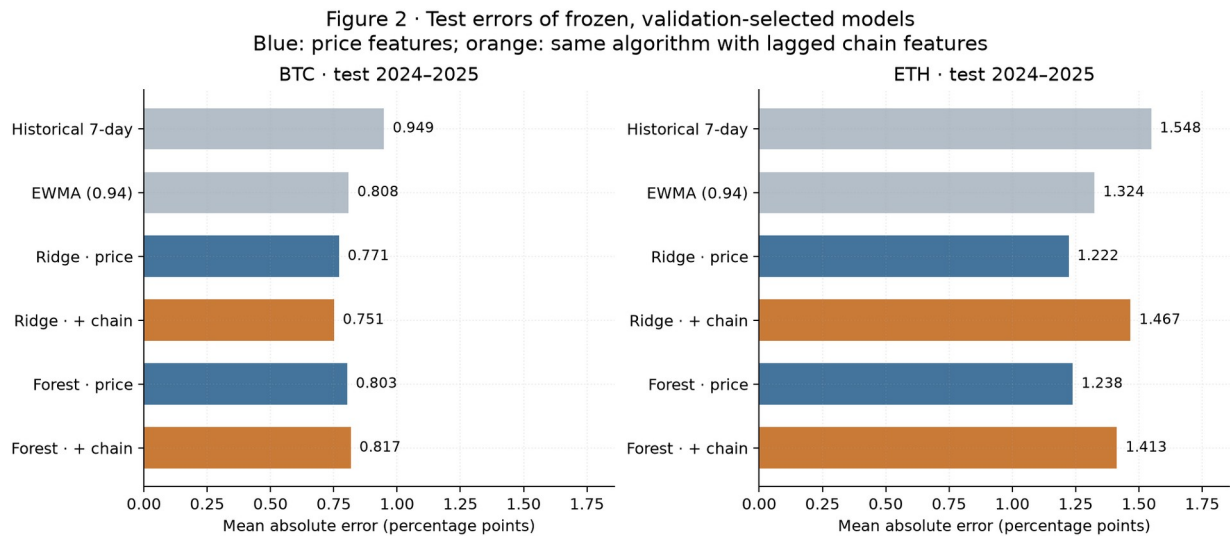


Figure 2 Mean absolute errors for the main specification. Both panels share the same horizontal scale; 724 test origins per asset. Benchmarks and all paired information sets are retained. Exact RMSE and QLIKE values appear in Table 3.

5.1 Incremental activity information

For BTC, Ridge improves MAE by 2.51 percent, but the paired 95 percent interval is [-4.22, 7.81] percent. It includes zero and both a modest deterioration and a modest improvement. The forest improvement is -1.75 percent, with interval [-7.27, 3.38]. Its point estimate favors retaining the price information set. This experiment does not provide clear primary evidence that the activity block improves BTC forecasts.

For ETH, Ridge improvement is -20.00 percent and forest improvement is -14.07 percent. The corresponding intervals are [-27.28, -12.32] and [-19.90, -7.90] percent. Negative improvement means higher forecast error after adding activity. Both intervals lie below zero under the primary block choice. These are conditional paired comparisons, not multiplicity adjusted hypothesis tests, but their direction is consistent across both algorithms.

The level of the ETH forecasts helps describe the deterioration. Mean predicted RMS rises from 3.594 to 4.137 percent for Ridge, compared with an observed mean of 3.325. The forest also predicts a higher average level after activity is added. This establishes an overprediction pattern in the held out period; it does not identify which feature or protocol change caused it. Figure 4 retains the validation selected price forecasts and their observed future outcomes for inspection.

Table 4 Incremental activity information in the main test

Asset	Algorithm	MAE reduction percent	95 percent interval
BTC	Ridge	+2.51	[-4.22, 7.81]
BTC	RF	-1.75	[-7.27, 3.38]
ETH	Ridge	-20.00	[-27.28, -12.32]
ETH	RF	-14.07	[-19.90, -7.90]

Positive reduction favors chain features. Percentile intervals use 2,000 non circular moving block bootstrap replicates with 28 origin blocks. They are conditional on fixed model selection and are not adjusted for four comparisons.

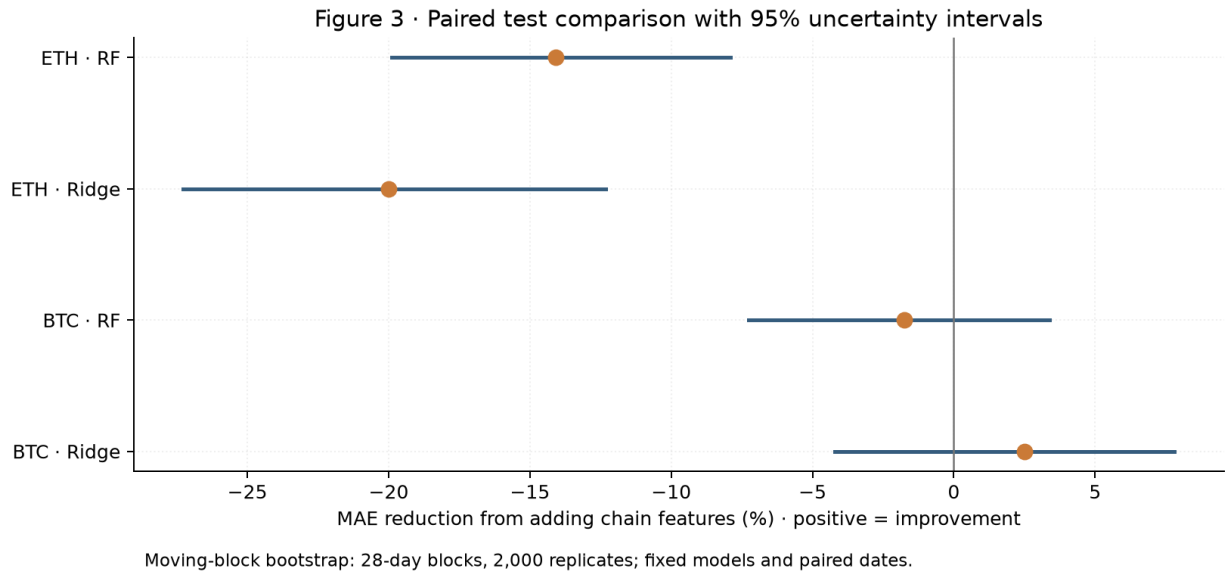


Figure 3 Paired MAE reductions and 95 percent block bootstrap intervals. A zero reduction indicates equal average error. BTC intervals straddle zero; both ETH intervals favor price only models. The displayed intervals use 28 origin blocks.

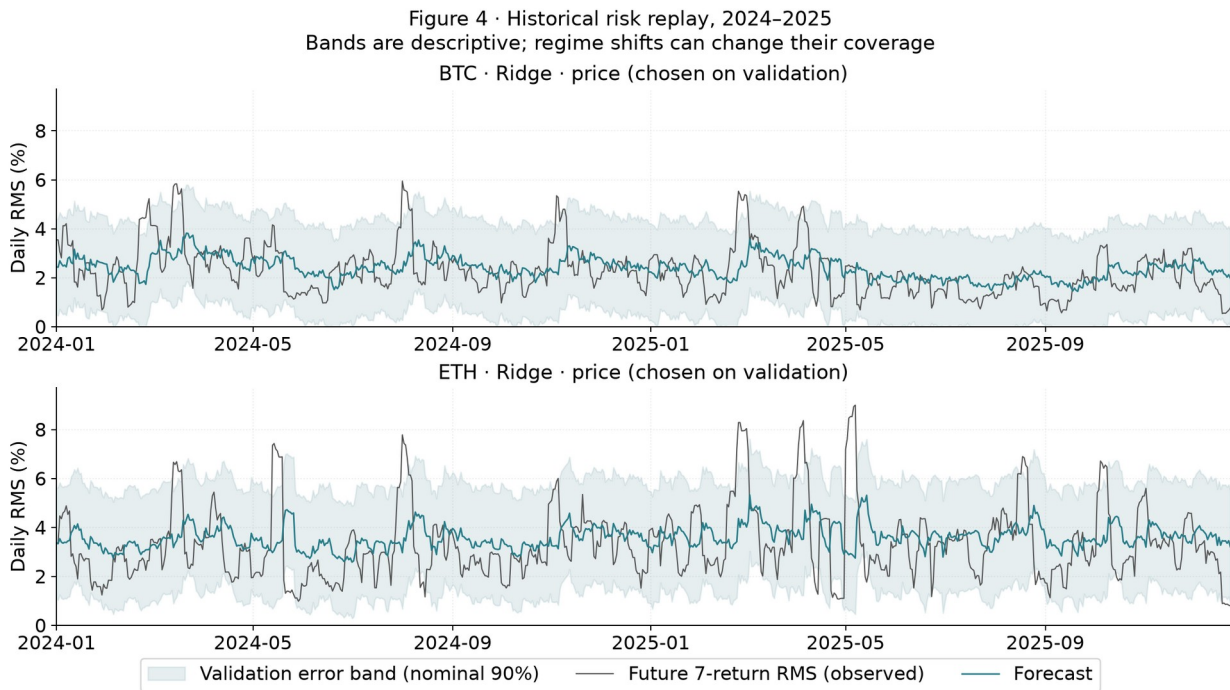


Figure 4 Future observed seven return RMS and forecasts aligned by marked origin date. Price only Ridge is chosen using validation for each asset. Bands use validation absolute errors with nominal 90 percent calibration; they are descriptive rather than guaranteed prediction intervals. Each curve contains 724 test origins and both panels use a common volatility scale.

5.2 Stability checks

The BTC Ridge result changes across years. Its MAE reduction is -1.54 percent in 2024 and 7.20 percent in 2025. The BTC forest changes from 0.41 to -4.25 percent. In contrast, the ETH activity block raises error for both algorithms in both years, with a larger deterioration in 2025. The annual results are descriptive; there is no annual parameter reselection, and the two subsets contain 366 and 358 origins respectively.

Delaying activity by three marked days produces BTC Ridge improvement of 4.62 percent, with primary interval [-1.39, 9.90]. The BTC forest deteriorates by 6.07 percent. ETH deterioration remains 20.64 percent for Ridge and 16.41 percent for the forest. An extra delay therefore does not establish a stable activity benefit across algorithms and assets.

The volume control reveals an important qualification. Adding activity to the BTC Ridge model with symmetric volume controls reduces MAE by 9.63 percent, with interval [3.21, 15.32]. However, the price plus volume baseline has MAE of 0.912 percentage points, and its chain augmented version has MAE of 0.825. Both are worse than the original price only Ridge value of 0.771. The favorable incremental comparison does not justify choosing that specification for the dashboard. The volume controlled forest still worsens for BTC, and both ETH models worsen. Activity information can therefore help a particular restricted comparison without improving the best validation selected implementation.

With 14 and 56 origin blocks, both main BTC intervals continue to include zero and both main ETH intervals remain below zero. Appendix Table A2 retains these ranges. Some sensitivity conclusions depend on the block choice: for lag three BTC Ridge, the 56 origin interval barely clears zero at 0.02 to 8.71 percent, while the 14 and 28 origin intervals include zero. This marginal change is not promoted to the primary finding. All comparisons remain based on a single historical split.

Table 5 MAE reductions by test year

Asset	Algorithm	2024 percent	2025 percent
BTC	Ridge	-1.54	+7.20
BTC	RF	+0.41	-4.25
ETH	Ridge	-14.37	-24.77
ETH	RF	-11.15	-16.67

Main lag two specification. 2024 has 366 origins; 2025 has 358. Positive values favor activity features. Models remain fixed.

Table 6 Pre specified information set sensitivities

Specification	BTC Ridge	BTC RF	ETH Ridge	ETH RF
Main lag two	+2.51	-1.75	-20.00	-14.07
Lag three	+4.62	-6.07	-20.64	-16.41
Volume controls on both sides	+9.63	-2.27	-11.43	-7.95

Entries are paired MAE reductions in percent, with 724 test origins. Every specification is tuned on validation using the same grid. Intervals for all 28 origin comparisons are shown in Figure 6 and retained in bootstrap.csv.

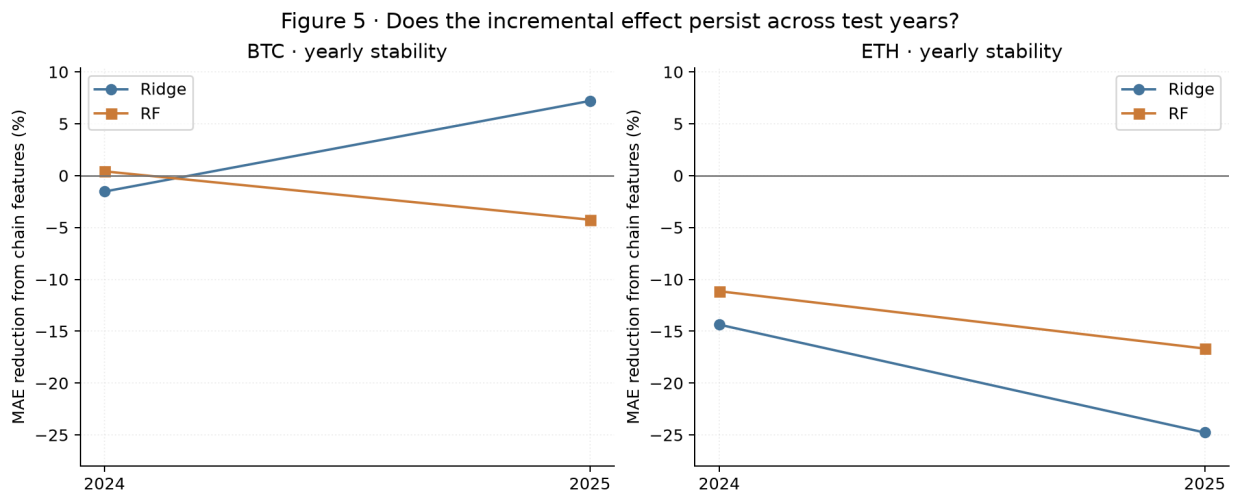


Figure 5 Annual paired MAE reductions using the same fixed main models. Common vertical scales and a zero line retain the direction and size of the comparisons. The connectors link two annual estimates, not an interpolated monthly trend.

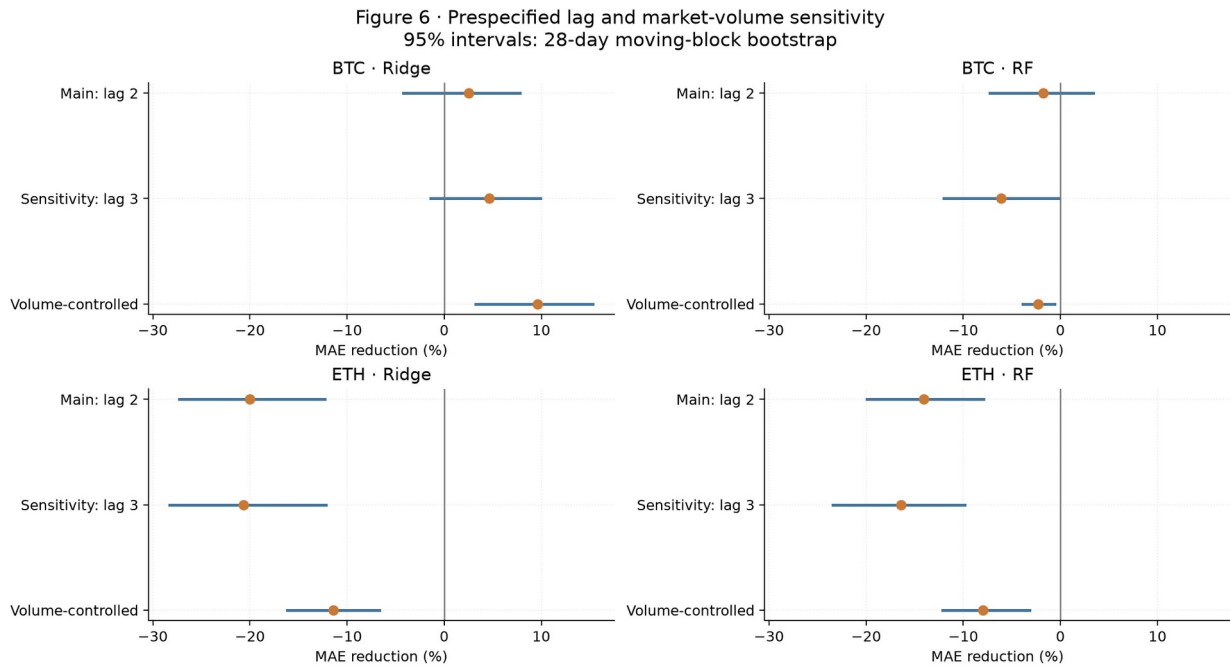


Figure 6 MAE reductions and 95 percent intervals for main lag two, lag three and symmetric volume controls. Common scales permit comparisons across assets and algorithms. Intervals use the primary 28 origin block length; sensitivities do not replace the original specification.

5.3 Error bands and model selection

Validation selects price only Ridge with alpha 0.1 for both assets. Its validation MAE is 1.053 percentage points for BTC and 1.249 for ETH. The dashboard follows this rule even though chain Ridge has slightly lower BTC test MAE. Selecting that test winner would use information unavailable when the model choice was fixed.

For the selected price Ridge models, the empirical test coverage of the nominal 90 percent validation error band is 95.03 percent for BTC and 87.02 percent for ETH. Mean widths are 3.919 and 4.582 percentage points. A wide band and a clipped zero lower endpoint can coexist with a useful point forecast. The ETH coverage shortfall and BTC excess coverage show why a validation calibration label must not be interpreted as a future coverage guarantee. These widths and coverage rates measure uncertainty about the daily RMS target, not a confidence interval for an asset return.

Table 7 Validation selected dashboard forecasts

Asset	Selected model	Test band coverage	Mean band width pp
BTC	Ridge price	95.03%	3.919
ETH	Ridge price	87.02%	4.582

Selection uses validation MAE only. Nominal 90 percent band radius is the 652nd smallest absolute validation error among 723 observations. Coverage is descriptive on 724 test origins.

6 Research dashboard

The accompanying application is a research monitoring prototype that replays the fixed test period. It displays the forecast issue date, the future seven return outcome, the selected model and its historical error band. The forecast comparison view retains the price and activity information sets side by side, so a reader can inspect the incremental test rather than infer value from the presence of a blockchain chart. Source inspection preserves the underlying reviewed rows, model definitions, date conventions and data attribution.

A practical monitoring workflow begins with the validation selected price baseline and treats additional chain indicators as candidates that must earn a place through subsequent evidence. The fixed period can be explored by asset and model without recalculating parameters from test outcomes. Historical activity charts describe network observations; they do not establish a risk signal or explain a price movement causally. The forecast and realized target must remain distinguishable during replay, because the latter is known only after the horizon has elapsed.

The application is served locally and can be preserved as a portable snapshot. Its data end on 31 December 2025, and its last scored origin is 24 December, nominally issued on 25 December. It has no verified live feed, scheduled refresh or point in time data store. Turning it into an operational service would require licensed current inputs, versioned availability timestamps and new prospective validation. A model output is a variability estimate, not a buy or sell recommendation.

7 Discussion and limitations

The central finding is that public blockchain measurements do not provide a stable incremental improvement under this compact forecasting design. BTC offers a small MAE advantage for one algorithm with uncertainty spanning zero, a different sign for the forest and a reversal between test years. ETH activity augmentation raises MAE in both model families and both years. The positive BTC volume controlled Ridge comparison qualifies the conclusion: information value depends on the conditioning set and estimator, and an incremental gain need not improve the most useful available baseline.

The results differ in scope from the favorable Bitcoin evidence of Gkillas et al. (2021). Differences in the asset sample, target horizon, daily rather than intraday measurement, feature delay and evaluation period could all change the comparison. This study does not isolate which difference is responsible and does not refute a result about high frequency realized volatility. It shows that transferring the general idea to a simple two asset public data workflow requires another empirical check.

One plausible explanation for instability is that recorded network use combines economically relevant demand with automation, protocol mechanics and changes in transaction fees. ETH network and fee rules change within the study window, while the four observed series are treated as consistent predictor families. The current documentation does not verify an unchanged fee schema across every historical year. These possibilities motivate further feature audits; they are not mechanisms established by this experiment. Aggregate feature correlations or forest importance would not by themselves identify causality.

The fixed training policy makes deterioration after 2021 easy to observe, but restricts the conclusion to models frozen at that point. An expanding or rolling fit might adapt better, and a GARCH family, other regularization choices or different activity transformations could alter performance. None is evaluated here. Adding new models after viewing this test would create another research iteration and would require a new untouched evaluation period before claiming prospective improvement.

The statistical evidence is limited by two selected assets, one chronological split and dependent overlapping outcomes. Moving blocks retain local dependence, but cannot fully represent structural change or all uncertainty from fitting and tuning. The reported intervals are conditional and unadjusted for multiple comparisons. Daily return RMS is also a noisy risk proxy, with no intraday, order book, tail probability or portfolio loss measurement. Consequently, the results do not establish economic trading gains, a capital requirement or a value at risk estimate.

Finally, the provider snapshot can contain historical revisions and cannot establish the exact information available to an investor at each nominal issue time. Conservative lags address chronological ordering, not revision bias. Data license conditions limit this package to attributed noncommercial research use. These constraints allow a reproducible academic application while preventing the current demonstration from being represented as an unrestricted commercial or live system.

8 Conclusion

For the fixed 2018 through 2025 Coin Metrics sample, lagged activity does not deliver a reliable general improvement in seven future daily return RMS forecasts beyond recent price history. BTC Ridge reduces MAE slightly in the main comparison, but its interval includes zero and other evaluation losses worsen. BTC forests do not show a stable benefit. ETH augmented models incur higher errors, including under an extra activity delay and symmetric volume controls. A favorable BTC Ridge volume controlled comparison remains conditional on a weaker baseline.

The resulting risk monitoring prototype therefore uses price only Ridge chosen on validation for both assets and keeps the activity experiments visible as evidence. The practical contribution is a complete and inspectable forecasting workflow with truthful timing, uncertainty and source limits. A prospective study with versioned data and a fresh evaluation window is the next step needed to establish deployment value.

References

- Fulvio Corsi (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174–196. <https://doi.org/10.1093/jjfinec/nbp001>
- Konstantinos Gkillas; Maria Tantoula; Manolis Tzagarakis (2021). Transaction activity and bitcoin realized volatility. *Operations Research Letters*, 49(5), 715–719. <https://doi.org/10.1016/j.orl.2021.06.016>
- Andrew J. Patton (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256. <https://doi.org/10.1016/j.jeconom.2010.03.034>
- Arthur E. Hoerl; Robert W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
- Leo Breiman (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Hans R. Künsch (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3), 1217–1241. <https://doi.org/10.1214/aos/1176347265>
- Coin Metrics (2026a). Active Addresses. Documentation or repository snapshot accessed 6 October 2026. <https://docs.coinmetrics.io/network-data/network-data-overview/addresses/active-addresses>
- Coin Metrics (2026b). Transactions. Documentation or repository snapshot accessed 6 October 2026. <https://docs.coinmetrics.io/network-data/network-data-overview/transactions/transactions>
- Coin Metrics (2026c). Fees. Documentation or repository snapshot accessed 6 October 2026. <https://docs.coinmetrics.io/network-data/network-data-overview/fees-and-revenue/fees>
- Coin Metrics (2026d). FAQs: PriceUSD and ReferenceRate timestamp conventions. Documentation or repository snapshot accessed 6 October 2026. <https://gitbook-docs.coinmetrics.io/service-and-support/faqs>
- Coin Metrics (2026e). Status Page. Documentation or repository snapshot accessed 6 October 2026. <https://github.com/coinmetrics/docs-website/blob/master/status-page.md>
- Coin Metrics (2026f). Community data archive and LICENSE. Documentation or repository snapshot accessed 6 October 2026. <https://github.com/coinmetrics/data>

Appendix A Reproduction details

The companion research directory includes the two fixed input CSV files, source attribution, a time stamped protocol, executable `run_research.py`, an executed notebook, package versions and machine readable results. The outputs preserve predictions on the validation and test partitions, annual and overall losses, selected parameter grids, aligned paired losses, block bootstrap intervals and a dashboard replay dataset. Figures are exported as both PNG and SVG. The local protocol SHA256 is `fc58ab43d62c4713c751a94823547c14809ef0a3731605e6973c6d67df7e46da`.

The recorded environment uses Python 3.12, NumPy 2.5.3, pandas 3.0.6, scikit-learn 1.9.1 and Matplotlib 3.11.2. Reproduction uses the fixed local data and `requirements.lock.txt`, then runs `python run_research.py` from the research directory. The notebook provides readable retained outputs; rerunning

the script regenerates the experiment. Model seeds and deterministic bootstrap case seeds are archived. Small numerical differences on another platform should be reconciled against the stored predictions rather than interpreted as a new finding.

The minimum training warmup is thirty returns. Eligible origin counts are 1,423 for training, 723 for validation and 724 for test per asset. No test observation enters scaling or parameter selection. In the main information set, price features number five and the activity block adds eight; the volume control adds two features to both information sets. The complete parameter choices for every sensitivity are retained in `selected_parameters.json`. Table A1 reports the main fitted choices.

Table A1 Main model parameter choices

Asset	Model	Selected parameters	Validation MAE pp
BTC	Ridge price	alpha 0.1	1.053
BTC	Ridge plus chain	alpha 100	1.064
BTC	Forest price	depth 4; leaf 30	1.110
BTC	Forest plus chain	depth 4; leaf 30	1.150
ETH	Ridge price	alpha 0.1	1.249
ETH	Ridge plus chain	alpha 0.1	1.341
ETH	Forest price	depth 4; leaf 30	1.344
ETH	Forest plus chain	depth 8; leaf 10	1.303

Ridge uses training fitted standardization. All forests use 300 trees and all available features per split. The minimum validation loss selects each information set separately.

Table A2 Main comparison block length sensitivity

Asset and algorithm	14 origin interval	28 origin interval	56 origin interval
BTC Ridge	[-4.68, 9.77]	[-4.22, 7.81]	[-3.25, 7.04]
BTC RF	[-6.98, 2.56]	[-7.27, 3.38]	[-7.22, 3.24]
ETH Ridge	[-28.00, -12.12]	[-27.28, -12.32]	[-25.87, -12.00]
ETH RF	[-20.76, -7.91]	[-19.90, -7.90]	[-19.49, -7.67]

Each cell is a 95 percent percentile interval for paired MAE reduction in percent. The model estimates and point reductions are unchanged. All intervals use 2,000 resampling replicates.

Appendix B Research preparation disclosure

This manuscript and the accompanying analyses were prepared with AI assistance for coding, source review, drafting and formatting. Model estimates and reported statistics derive from the archived data and executable analysis, rather than simulated illustrative results. The named student author must review the argument, sources, code, contribution and disclosure before any submission, and comply with the selected journal or institution policy. No independent authorship attestation, funding statement or conflict of interest declaration is inferred from the source materials.

Coin Metrics data and derived data displays retain attribution and CC BY-NC 4.0 conditions. The package is prepared for noncommercial research. A broader manuscript license does not replace the license of third party data or authorize a commercial monitoring service.