Fixed Event-Day Dummies for US Macro Releases Did Not Improve Daily Crypto Volatility Forecasts: A Registered Out-of-Sample Test on Six Coins

Fernando Varani Espoladore
Independent researcher
This version: 24 November 2026

Abstract

We ask whether fixed day dummies for scheduled US macro events help forecast the daily realised volatility of large crypto-assets. The events are FOMC rate decisions and the first releases of nonfarm payrolls (NFP) and CPI. The data are 5-minute Binance prices for six USDT-margined perpetual futures (BTC, ETH, XRP, BNB, DOGE, SOL) from January 2020 to September 2026. The baseline is a per-coin HAR model of log daily realised volatility with a weekend dummy. The test model adds three event-day dummies, with separate coefficients for each coin. Both models were fitted once on 2020–2023 and then applied unchanged to January 2024 – September 2026. They were judged by squared log error on 85 event dates (510 coin-days). The test and its five pass marks were registered before the single look at outcomes. The registration records two tries: a counting step and that look. The evaluation window was not used for fitting, but earlier related studies had already used it. The registration is timestamped only in the author's private records. The registered test killed the claim, and all five marks failed. The pooled loss ratio (test model over baseline) was 1.069, against a pass mark of 0.95 or less. The Newey–West (Diebold–Mariano-type) t of the per-date loss difference was +1.75, against a pass mark of −2 or less (two-sided $p \approx 0.08$). So the dummies did not improve forecasts; the point estimate was a deterioration, but it is not significant at the 5% level. Only one coin of six improved, and the loss rose in both halves of the window. The claim also needs the baseline to under-forecast event days, and this failed too: its errors on event dates were not significantly larger than on matched non-event days one week earlier ($t = +1.42$, against a required +1.65). The result concerns one specification, fixed per-coin dummies estimated on 2020–2023. It does not show that US releases leave crypto volatility untouched.

Keywords: realised volatility; cryptocurrencies; macroeconomic announcements; FOMC; CPI; nonfarm payrolls; HAR model; forecast

evaluation; pre-registration; negative result

Introduction

Scheduled US macroeconomic announcements move prices and raise volatility in currency, bond and equity markets (Andersen, Bollerslev, Diebold and Vega 2003). FOMC days are also unusual for equity returns (Lucca and Moench 2015). Several studies report that crypto-asset prices respond to US monetary policy announcements as well (Corbet et al. 2020; Karau 2023). Crypto trades around the clock, and a standard daily volatility model such as the HAR (Corsi 2009) has no calendar of these events. If event days are systematically more volatile than such a model expects, the simplest fix is one dummy per event type.

We registered this claim as a forecasting test with fixed pass marks before looking at any outcome. The registration also made numeric predictions. Table 1 sets each one beside its outcome. The predictions about counts and spread were close. The predictions about the claim itself mostly failed:

- The fitted FOMC coefficient was in range on average but not coin by coin.
- The CPI coefficient fell below its predicted range on average.
- The payrolls coefficient had the wrong sign for every coin.
- Every prediction about forecast accuracy failed.

The registration also gave the test about a 0.35 chance of passing. (Another field of the same record rounds this to 0.4.) That is a probability, not a prediction that can hold or fail. A single failed test is consistent with it.

Table 1. Registered predictions and outcomes.

Registered prediction	Outcome	Held?
Fit-period event dates about 120 (FOMC ~31, NFP ~47, CPI ~47)	123 (30 / 47 / 47)	Yes
Evaluation event dates about 85 (FOMC ~22; ~40 / ~45 per half); about 500 coin-days	85 (22; 41 / 44); 510	Yes
SD of the baseline's date-mean residual 0.22–0.28	0.267	Yes
Smallest effect the t mark can detect, 0.12–0.16 log	0.165	Just above the range
FOMC coefficient +0.06 to +0.15	Mean +0.084, in range; 3 of 6 coins outside (BTC above; BNB, DOGE below)	Mean yes; per coin, half
CPI coefficient +0.08 to +0.20	Mean +0.072, below the floor; 2 of 6 coins below (XRP, DOGE)	No (mean); 4 of 6 coins in range
NFP coefficient +0.04 to +0.10	Mean −0.058; negative for all 6 coins	No
Pooled loss ratio 0.85–0.95	1.069	No
Newey–West t −1.8 to −3.5	+1.75	No
Loss ratio below 1 in both halves	1.072 and 1.064	No
5–6 coins below 1; BTC and ETH strongest, DOGE weakest	1 coin (ETH); BTC the worst (1.138)	No
Control t (event dates vs. a week earlier) +2 to +4	+1.42	No

This is a forecasting study only. No trading rule is built or evaluated, and transaction costs do not enter.

Data

Prices. We use 5-minute candles for the Binance USDT-margined perpetual futures BTCUSDT, ETHUSDT, XRPUSDT, BNBUSDT, DOGEUSDT and SOLUSDT, from 1 January 2020 to 30 September 2026. These come from Binance's public kline archive. Timestamps are bar starts in UTC. For each coin and UTC calendar day, realised volatility (Andersen and Bollerslev 1998; Andersen, Bollerslev, Diebold and Labys 2003) is RV = sqrt(sum of squared 5-minute log close-to-close returns within the day). A day is valid if it has at least 280 of the 288 possible bars and RV > 0. The target is y = log RV. Every coin had a median of 288 bars per day.

Missing or invalid days. The HAR features are built on each coin's sequence of valid days:

- lag1 is the previous valid day;
- the "7-day" and "30-day" means are the means of the previous 7 and 30 valid days.

When a day is invalid, it is skipped, and the window reaches one calendar day further back. BTC and ETH have no invalid days. The other four coins have 1 to 6 invalid days

each in almost seven years (Table 2), so the effect is negligible. The first 30 valid days of each coin are a warm-up and do not enter the fit.

Table 2. Coverage per coin. Valid days are reported by the script. The fit-day and event-day counts are derived from each coin's first data day, its valid-day total and the 30-day warm-up; the script did not output them directly. Where the invalid days fall is not known, so ranges are given. Fit days include weekends.

Coin	First data day	Valid days, 2020-01 to 2026-09	Fit days to 2023-12-31	Fit-period event days FOMC / NFP / CPI (approx.)
BTC	1 Jan 2020	2465 (all)	1431	30 / 47 / 47
ETH	1 Jan 2020	2465 (all)	1431	30 / 47 / 47
XRP	6 Jan 2020	2454	1420–1426	30 / 47 / 47
BNB	10 Feb 2020	2424	1390–1391	30 / 45 / 46
DOGE	10 Jul 2020	2273	1239–1240	27 / 40 / 41
SOL	14 Sep 2020	2202	1168–1174	26 / 38 / 38

Each per-coin event coefficient therefore rests on 26 to 47 fit-period event days.

Event calendar. We use three kinds of event date:

- the Federal Reserve's scheduled FOMC decision days;
- first-release days of US nonfarm payrolls;
- first-release days of US CPI.

The payrolls and CPI dates come from first-release vintages in ALFRED. The 08:30 New York releases and the 14:00 New York FOMC statement both fall inside the UTC day with the same date, so each flag is set on that UTC day. Before using the calendar, the script checks that five known event dates are present. We note one apparent calendar error, found while writing this paper: the April 2020 payrolls report is dated 11 May 2020, but it was released on 8 May 2020. That date is in the fit period only. In late 2025 the calendar follows the delayed releases of that period: September CPI on 24 October, and payrolls on 20 November and 16 December. It has no separate October CPI.

Sample counts. A date can carry two flags. In the fit period, 124 event flags fall on 123 dates because 10 June 2020 was both a CPI day and an FOMC day. In the evaluation window, 86 flags fall on 85 dates because 12 June 2024 was likewise both a CPI day and an FOMC day. The counts of dates use only weekdays on which at least three coins have a residual. This is why the January 2020 events (still inside every coin's warm-up) do not appear. All six coins are present on every one of the 85 evaluation dates ($6 \times 85 = 510$ coin-days).

Table 3. Sample sizes.

Item	Value
Fit window	target days up to 2023-12-31
Fit-period event dates (FOMC / NFP / CPI flags)	123 dates (30 / 47 / 47 = 124 flags)
Evaluation window	weekdays, 2024-01-01 to 2026-09-30
Evaluation event dates (FOMC / NFP / CPI flags)	85 dates (22 / 32 / 32 = 86 flags)
First half (2024-01 to 2025-04) / second half (2025-05 to 2026-09)	41 / 44 dates
Evaluation event coin-days	510
Control dates (one week before an event date)	73

Method

Baseline. For each coin we fit y(t) = a + b1·y(t−1) + b7·mean y(t−7..t−1) + b30·mean y(t−30..t−1) + c·weekend(t) + e(t) by OLS. This is a HAR model in logs (Corsi 2009) with a Saturday/Sunday dummy. It is fitted once on target days up to 31 December 2023 and then applied unchanged.

Test model. The test model is the same regression plus three dummies for the target day: FOMC(t), NFP(t) and CPI(t). It is fitted per coin on the same window and also applied unchanged. Each coin gets its own three event coefficients. The HAR coefficients are re-estimated jointly with the dummies.

Evaluation set and loss. The evaluation set is the UTC weekdays from 2024-01-01 to 2026-09-30 that carry at least one flag and have residuals for at least three coins. The loss is squared log error, $(y - \text{forecast})^2$, which is the squared error of log RV.

Pass marks. All five were fixed at registration, and the claim survives only if all five hold:

1. Pooled loss ratio ≤ 0.95. This is the sum of the test model's squared errors over the baseline's, across all event coin-days.
2. Newey–West $t \leq -2$. For each date we average the loss difference (test minus baseline) over that date's coins. The t is that of the mean of these per-date differences, using a Bartlett-kernel long-run variance with 3 lags and the dates in time order (Newey and West 1987). Averaging within a date first absorbs the strong correlation across coins. This is a Diebold–Mariano-type comparison (Diebold and Mariano 1995).
3. The loss ratio is below 1 in each half of the window.
4. At least five of the six coins have their own loss ratio below 1.
5. A control comparison with Welch $t \geq 1.65$ (see below).

Interpreting the t statistic. Both models' parameters are estimated once on a fixed window, and we compare the forecasts that this produces. The t statistic is therefore best read as Giacomini and White (2006) read it: as a test of equal predictive ability of the two forecasting methods, estimation noise included. Their framework allows a fixed estimation window and nested models. It is not a test of whether the population event coefficients

are zero. The test model nests the baseline, so when the true coefficients are small, estimation noise alone tends to raise its out-of-sample loss (Clark and West 2007). That is the relevant question here: whether adding the dummies, as estimated, improves the forecasts.

Lag choice. The Bartlett lags run over the sequence of event dates, not over calendar days. Event dates are on average about 12 calendar days apart (85 dates in 1,004 days), so lags of 1–3 events cover neighbouring events up to roughly five weeks apart. The lags allow for neighbouring events that share a volatility regime the HAR does not fully absorb. Three lags is the practice's standard setting, used in many earlier studies. For $T = 85$, the common rule of thumb floor(4(T/100)^(2/9)) also gives 3.

The control comparison (mark 5). The registration calls this a "placebo control", but it does not work like a usual placebo, where a non-significant result is good news. It tests a premise of the claim: that the baseline under-forecasts event days more than it under-forecasts ordinary days. It compares the baseline's date-mean residual (log RV minus forecast, averaged over coins) on event dates with the same quantity on control dates. A control date is an unflagged weekday exactly seven days before an event date. The mark requires event-date residuals to be larger, with Welch $t \geq 1.65$. The comparison guards against a spurious gain: if the baseline under-forecasts all days after its fit window, any upward shift would look helpful. A result below 1.65 is a failure for the claim.

Why 73 control dates against 85 event dates. Twelve event dates have no usable date one week earlier:

- In 7 cases the earlier date was itself an event date: the CPI days of 11 September 2024, 11 December 2024, 12 March 2025, 11 June 2025, 11 March 2026 and 10 June 2026 (each a week before an FOMC day), and the payrolls day of 4 September 2026 (a week before a CPI day).
- In 3 cases the earlier date was a US market holiday: Good Friday on 29 March 2024 and 3 April 2026, and Independence Day on 4 July 2024.
- In 2 cases the earlier date fell in the 24 December – 2 January exclusion: 29 December 2023 (also before the window) and 2 January 2026.

The exclusion of the unscheduled market closure of 9 January 2025 removed no date.

Why Welch rather than a paired test. The registration specified a Welch two-sample t. This followed an earlier study on the same window, in which control dates lay both before and after each event date and so did not pair one-to-one. Here the dates do pair one-to-one, but only 73 of the 85 event dates have a partner. A paired test would drop the other 12. A paired t would also remove any persistent level that a pair shares. If pair members are positively correlated, a paired t would have a smaller standard error than the Welch t. We did not compute the paired t: it was not registered, and adding it after the look would be a further, unregistered try.

Power (registered rule, derived). Power was computed in the counting step from the spread of the residuals only, before any event coefficient was fitted or any outcome mean was seen. Let $\bar{r}_t$ be the baseline's date-mean residual on event date t, let m be its sample

mean over the N event dates, and let σ be its standard deviation. Suppose the test model raises every event coin-day forecast by the same amount $s > 0$ and changes nothing else. Then each coin's residual becomes $r - s$. Its loss difference is $(r - s)^2 - r^2 = s^2 - 2s{\cdot}r$, and the per-date loss difference is $d_t = s^2 - 2s{\cdot}\bar{r}_t$. The mean of d_t is $s^2 - 2s{\cdot}m$, and its standard deviation is $2s{\cdot}\sigma$. Ignoring autocorrelation:

$t_{dm} = \sqrt{N}\,(s^2 - 2s{\cdot}m) \,/\, (2s{\cdot}\sigma) = -\sqrt{N}\,(m - s/2) \,/\, \sigma$.

So the t mark ($t_{dm} \le -2$) needs $m \ge s/2 + 2\sigma/\sqrt{N}$. For 80% power (adding 0.84 standard errors), the true mean event-day residual μ must satisfy $\mu - s/2 \ge 2.84{\cdot}\sigma/\sqrt{N}$. With $\sigma = 0.267$ (the SD of the date-mean residual over all evaluation weekdays) and $N = 85$, this gives $\mu - s/2 \ge 0.082$.

The registered rule assumes the fitted shift equals the true one ($s = \mu$). Then t_{dm} has half the expected size of the plain residual t, $\sqrt{N}{\cdot}\mu/\sigma$, and the requirement becomes $\mu \ge 2 \times 0.082 = 0.165$ log. The fitted per-coin coefficients were not used, because they did not yet exist at the counting step. For a different fixed shift the bar moves. For example, with $s = 0.10$, near the middle of the predicted coefficients, the requirement would be $\mu \ge 0.132$. In the actual study the shift differs by coin and event type (and is negative for payrolls), and the HAR coefficients are re-estimated, so the formula is a guide rather than an exact power calculation. It also uses the plain standard deviation, while the test uses a Newey–West variance.

Joint power. The joint power of all five marks was not computed before the look. At the design effect ($\mu = s = 0.165$), it is at most the 80% of the t mark. Rough calculations done for this paper use the same registered spread and no outcome data. They put the halves mark at about 95%: in each half, the expected z of $m - s/2$ is about 2.0. They put the control mark at about 99%: the expected Welch t is about 3.9. The per-coin mark and the 0.95 ratio mark depend on coin-level residual spreads that the counting step did not record. The joint power at the design effect is therefore below 80%, by an amount we did not estimate.

Design choices inherited from earlier studies. The registration records two tries, but several choices were fixed by earlier work of the practice on the same data and window. These choices bound the researcher degrees of freedom that remain in this study (Table 4).

Table 4. Origin of design choices.

Choice	Origin
UTC-day RV from 5-minute log returns; valid day ≥ 280 bars	The baseline model, an earlier study of the practice (registered October 2026)
HAR terms (1, 7, 30 days) on log RV; weekend dummy; per-coin OLS	Baseline model
Fit once to 31 Dec 2023, apply unchanged; evaluation Jan 2024 – Sep 2026	Baseline model (its single evaluation window)
Squared log error	Baseline model
≥ 3 coins per date; one value per date (mean over coins)	Earlier US-holiday study on the same window
Control dates 7 days apart; exclusions (NYSE holidays, 24 Dec – 2 Jan, 9 Jan 2025); Welch t	US-holiday study (which used dates both 7 days before and after)
Newey–West, Bartlett kernel, 3 lags	Practice's standard setting across earlier studies
Judging both halves and every coin	A lesson from the practice's earlier negative test of event-day effects in US equity index futures volatility (gains concentrated in some years)
New in this study: the event set (FOMC, NFP, CPI) and UTC-day timing; per-coin dummies; the half split at 1 May 2025; the five thresholds; control dates before the event only	This registration

Registration and procedure. The design and the analysis script were produced by an AI assistant working without a person at the keyboard. A second, independent AI review approved them before registration. Before the look, the script's measuring functions were run on a synthetic six-coin panel built on the real event calendar and the real coin start dates. According to the registration, pure noise failed at least one mark, and planted event effects of +0.25 to +0.35 log passed all five.

The registration recorded the claim, the reasoning, the numeric predictions, the baseline, the pass marks and a disclosure of earlier use of the window. The practice's records give these machine-written times (CET, 24 November 2026):

- registration, 10:23:08;
- counting step, 10:23:43;
- the single look, 10:24:45;
- verdict, 10:25:02.

The registration's list of amendments is empty. The analysis script was identified at both runs by the same 16-hex-digit content fingerprint (d0e084e145622e11), and the later replication found it unchanged.

These records are private. The registration was not lodged with a public registry, and no hash or public timestamp was published before the look. Readers therefore cannot independently confirm that the pass marks were fixed before the look and never

amended. The claim rests on the author's records, and the registration text and script are not public at the time of writing. We list this as a limitation.

Prior use of the evaluation window (disclosed). The 2024–2026 period was not used for fitting, but it is not fresh data. It was the baseline's single window for judging overall accuracy. It was also used by a related test of crypto volatility on US holidays and by a study of 5-minute crypto volatility around the same releases. According to the registration, none of those studies looked at daily errors on event days.

Results

Table 5. Registered marks and outcomes (event dates, 2024-01 to 2026-09).

Measure	Pass mark	Result	Pass?
Pooled loss ratio, test / baseline (510 coin-days)	≤ 0.95	1.069	No
Newey–West t of per-date loss difference (85 dates)	≤ -2	+1.75	No
Worse half's loss ratio (first 1.072, second 1.064)	< 1	1.072	No
Coins with own loss ratio < 1	≥ 5 of 6	1 of 6	No
Control comparison: event-date minus control-date baseline residual, Welch t	≥ 1.65	+1.42	No

All five marks failed, and the registered test killed the claim. On event dates, the dummies raised the mean per-date squared log error by 0.0054. That is a deterioration as a point estimate. The registration records the t statistic only. Using a normal approximation, $t = +1.75$ gives a one-sided p of about 0.96 against the hypothesis that the dummies improve forecasts, and a two-sided p of about 0.08. So the dummies did not improve forecasts, and the deterioration is not significant at the 5% level. Within each half, the t was +1.33 (first half) and +1.22 (second half).

In the control comparison, the baseline's mean residual on event dates exceeded that on control dates by 0.056 log ($t = +1.42$, one-sided $p \approx 0.08$ by the normal approximation). This is in the expected direction but short of the registered 1.65.

Table 6. Results by coin: event-day loss ratio and the event coefficients fitted on 2020–2023 (log RV). Predicted ranges: FOMC +0.06 to +0.15, NFP +0.04 to +0.10, CPI +0.08 to +0.20.

Coin	Loss ratio	FOMC coef.	NFP coef.	CPI coef.
BTC	1.138	+0.184 (above)	−0.024 (below)	+0.134
ETH	0.953	+0.098	−0.020 (below)	+0.092
XRP	1.043	+0.091	−0.081 (below)	+0.031 (below)
BNB	1.094	+0.019 (below)	−0.034 (below)	+0.086
DOGE	1.089	−0.008 (below)	−0.143 (below)	−0.007 (below)
SOL	1.132	+0.118	−0.045 (below)	+0.095
Mean	—	+0.084 (in range)	−0.058 (below)	+0.072 (below)

The coefficient predictions, one at a time:

- FOMC. The mean (+0.084) is in the predicted range, but three of the six coins are outside it: BTC above, BNB and DOGE below.
- CPI. The mean (+0.072) is below the predicted floor of +0.08. XRP and DOGE are below the range, and the other four coins are in it.
- Payrolls. The coefficient is negative for all six coins, against a predicted positive range. In 2020–2023, payroll days were slightly less volatile than the baseline expected.

Table 7. Results by event type (reported, not pass marks).

Event type	Dates	Loss ratio, test / baseline	Baseline mean residual on these dates	Mean fitted coefficient
FOMC	22	0.877	+0.090	+0.084
NFP	32	1.068	+0.029	−0.058
CPI	32	1.181	−0.026	+0.072
All event dates	85	1.069	+0.022	—

In 2024–2026 the baseline's average error on event dates was small (+0.022 log), far below the 0.165 log that the t mark needed for 80% power. The pattern by event type matches the formula in the Method section. A fixed shift s lowers the loss only if the mean residual m exceeds s/2 (for $s > 0$):

- On FOMC days, $m = +0.090$ exceeded $s/2 \approx 0.042$, and the dummies helped (ratio 0.877).
- On CPI days, the baseline's mean residual was slightly negative, so the positive CPI shift learned on 2020–2023 hurt.

- On payroll days, the shift was negative while the residual was positive, so it also hurt.

This split by event type was reported, not tested. The FOMC result comes from 22 dates and was seen after the look, so it is not evidence of an FOMC effect.

Other reported numbers:

- Control dates with the shift applied. Applying each source event's coefficients to its control date gave a loss ratio of 1.132. This is what one expects when a shift is applied to ordinary days.
- All weekdays. Over all evaluation weekdays, event and non-event, the test model's loss ratio was 1.008. Most or all of this excess comes from the event days themselves. The script did not report a ratio for non-event days alone, but the reported numbers bound it. The baseline's summed squared error on the 510 event coin-days is $510 \times 0.0054 / 0.069 \approx 40.0$, and the excess loss there is about 2.77. There are about 3,800 non-event weekday coin-days (718 weekdays $\times$ 6 coins $-$ 510). If the baseline's mean squared error on them lies between 0.06 and 0.10, the non-event-day loss ratio is between about 0.997 and 1.002. An earlier draft explained the 1.008 by the re-estimated HAR coefficients; that is not supported, and the coefficient changes were not output.

What this does and does not show

It shows that, in this registered design, fixed per-coin FOMC, payrolls and CPI dummies did not improve out-of-sample daily volatility forecasts for six large coins. The dummies were estimated on 2020–2023 and judged on January 2024 – September 2026. The registered test killed the claim, and every mark failed:

- The pooled loss ratio was 1.069, against a mark of ≤ 0.95.
- The t was $+1.75$, against a mark of ≤ -2. The point estimate was a deterioration, though not significant at 5%.
- The loss rose in both halves and for five of six coins.
- The baseline's event-day errors were not significantly larger than its errors a week earlier.

Because the claim was killed, the planned later test on data collected after registration will not be run.

It does not show that US macro releases have no effect on crypto volatility. The limits are these:

- Specification. Only fixed per-coin dummies were tested. Each per-coin coefficient rests on 26 to 47 fit-period event days (Table 2). Estimation noise alone can therefore make the larger model's forecasts worse (Clark and West 2007). A pooled specification, with one coefficient per event type shared across coins, is a natural alternative that we did not test. Effects that change over time, or that depend on the size of the surprise in the release, were not tested either.
- Loss function. Squared log error is one loss among several. QLIKE, or squared error on the level of RV, can rank volatility forecasts differently (Patton 2011). Neither was

registered, and we do not report them.

- Power. At 80% power, the t mark could detect an average event-day under-forecast of about 0.165 log when the fitted shift equals the true one. Smaller effects would usually be missed. The joint power of all five marks was lower and was not computed.
- Horizon. The study measures volatility over the whole UTC day only. It says nothing about the minutes around a release, where volatility can rise even when the full day does not stand out.
- Data reuse. The evaluation window was held out from fitting but already existed, and earlier related studies had used it. Many design choices were inherited from them (Table 4). The result is an out-of-sample test in time, not a test on data that did not exist when the test was registered.
- Verifiability of the registration. The order "registration, then counting, then look" rests on the practice's private, machine-written timestamps, which fall within about two minutes of one another on the day this paper was written. No public timestamp or hash was published.
- Calendar. One payrolls date in the fit period appears to be misdated (see Data). Its effect on the result is likely negligible but was not checked.
- FOMC days. The lower loss on FOMC days (Table 7) is a subgroup finding from 22 dates. It would need its own registered test.

Nothing here is investment advice or a recommendation to trade any instrument.

Reproducibility

The study is a single analysis script with three modes:

- count: counts and power only;
- look: the single look at outcomes;
- check: the synthetic panel.

The script reads stored 5-minute candles and the event calendar and downloads nothing. Only the synthetic check uses random numbers, and it is seeded. After the look, the study was re-run end to end with the script unchanged. All 37 numbers shared between the re-run and the registered record matched exactly. The prices are public Binance market data (USDT-margined perpetual futures klines). The event dates come from public Federal Reserve and ALFRED sources. The registration text and the script are not public at the time of writing.

Parts of the analysis and the writing were done with an AI assistant.

References

Andersen, T. G., and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review, 39(4), 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. Econometrica, 71(2), 579–625.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Vega, C. (2003). Micro effects of macro announcements: Real-time price discovery in foreign exchange. American Economic Review, 93(1), 38–62.

Clark, T. E., and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics, 138(1), 291–311.

Corbet, S., Larkin, C., Lucey, B., Meegan, A., and Yarovaya, L. (2020). Cryptocurrency reaction to FOMC announcements: Evidence of heterogeneity based on blockchain stack position. Journal of Financial Stability, 46, 100706.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics, 7(2), 174–196.

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263.

Giacomini, R., and White, H. (2006). Tests of conditional predictive ability. Econometrica, 74(6), 1545–1578.

Karau, S. (2023). Monetary policy and Bitcoin. Journal of International Money and Finance, 137, 102880.

Lucca, D. O., and Moench, E. (2015). The pre-FOMC announcement drift. Journal of Finance, 70(1), 329–371.

Newey, W. K., and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3), 703–708.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics, 160(1), 246–256.