Online appendix to "Is the day's range already made? A random-walk calculation with nothing fitted, tested on index, currency and crypto futures"

Fernando Varani Espoladore
Independent researcher
This version: 22 October 2026

This appendix holds the robustness checks, the power analysis, the full results on the other markets, a one-hour version and the full description of the simulated sessions. Tables and sections are numbered S1, S2, ...; Tables 1 and 2 and Appendices A and B refer to the paper.

S1. Results by period

Table S1 splits the sample by period. The plug-in gap moves within its intervals except for WDO in 2025, when the range held 4.0 points more often than stated (0.5 to 6.7); with the true variance that gap is +0.2. The calibration slope is lowest in 2025 on WIN (0.38, 0.22 to 0.63), a year in which the volatility forecast was most biased.

Table S1. By period (plug-in and true-variance versions)

Market	Period	Readings (sessions)	Held	Plug-in gap (95% CI)	True-variance gap (95% CI)	Plug-in slope	Same-time AUC
WIN	2021-11 to 2024	10,486 (749)	52.0%	−1.4 (−3.9 to +0.8)	−2.1 (−4.3 to −0.2)	0.67	0.701
WIN	2025	3,381 (246)	60.7%	+2.3 (−2.7 to +6.3)	−0.7 (−4.0 to +2.8)	0.38	0.650
WIN	2026 to September	2,590 (185)	62.7%	−1.3 (−7.3 to +2.4)	−2.8 (−6.9 to −0.3)	0.48	0.680
WDO	2021-11 to 2024	10,486 (749)	59.1%	−0.8 (−2.8 to +1.1)	−1.2 (−2.8 to +0.4)	0.58	0.680
WDO	2025	3,472 (248)	58.6%	+4.0 (+0.5 to +6.7)	+0.2 (−2.3 to +3.0)	0.66	0.681
WDO	2026 to September	2,534 (181)	59.2%	+2.8 (−2.0 to +6.5)	+1.0 (−2.7 to +3.7)	0.56	0.681

WIN and WDO both have 749 sessions in 2021–2024 by coincidence: each has two sessions the other lacks (WIN alone on 14 December 2023 and 27 September 2024; WDO alone on 22 December 2021 and 8 March 2024).

Dependence. Readings within a session overlap, sessions depend on each other through volatility, and WIN and WDO share hours and news. Table S1b repeats the main intervals with moving blocks of 10 to 120 sessions and with whole calendar months resampled as clusters (`revision/dependence.py`, 400 resamples each). The conclusions do not depend on the choice. The plug-in gap's interval includes zero on both markets, and the plug-in slope's lies well below one. The conditional slope's lies below the simulated 0.93 in every scheme except 120-session blocks on WIN, where its upper end is 0.935. The WIN true-variance gap's lies below zero in every scheme, and WDO's includes zero in every scheme.

Table S1b. 95% intervals of the main results under different dependence assumptions (five-minute variance; allowance a fifth of the range)

Resampling	WIN plug-in gap	WIN plug-in slope	WIN conditional slope	WIN true-variance gap	WDO plug-in gap	WDO plug-in slope	WDO conditional slope	WDO true-variance gap
Estimate	−0.6	0.58	0.77	−1.9	+0.7	0.58	0.78	−0.6
Blocks of 10 sessions	−2.5 to +1.3	0.47 to 0.70	0.64 to 0.92	−3.5 to −0.3	−0.9 to +2.6	0.49 to 0.66	0.68 to 0.90	−2.0 to +0.8
Blocks of 20 (the paper)	−2.6 to +1.6	0.46 to 0.71	0.63 to 0.92	−3.5 to −0.4	−0.9 to +2.3	0.47 to 0.66	0.66 to 0.89	−1.8 to +0.8
Blocks of 40	−2.5 to +1.6	0.44 to 0.69	0.61 to 0.90	−3.5 to −0.3	−0.8 to +2.3	0.48 to 0.65	0.68 to 0.85	−1.7 to +0.7
Blocks of 60	−2.4 to +1.5	0.46 to 0.70	0.64 to 0.90	−3.5 to −0.4	−0.7 to +2.2	0.49 to 0.64	0.68 to 0.86	−1.6 to +0.3
Blocks of 120	−2.6 to +1.4	0.45 to 0.71	0.61 to 0.93	−3.4 to −0.3	−0.9 to +2.3	0.51 to 0.63	0.69 to 0.85	−1.5 to +0.3
Calendar months (clusters)	−2.5 to +1.5	0.44 to 0.72	0.64 to 0.92	−3.5 to −0.3	−0.9 to +2.4	0.49 to 0.68	0.66 to 0.91	−1.9 to +0.8

Gaps in points. Estimates as in Table 1 (16,457 WIN and 16,492 WDO readings).

S2. The size of extensions, and the search for a cause

The size of extensions. Under the Brownian law with the true variance, the extension of each reading has a known distribution with an atom at zero (no extension). Its randomised probability integral transform (uniform on the atom's interval when the range was not extended) is uniform under the law; on the reference simulation 10.0% of readings fall in its top tenth (9.7 to 10.3), as they should. The conditional transforms are compared with the same shares on the simulations, which are 9.2 to 10.1% in each tail. On WIN, given that the range was extended at all, 13.0% of the transforms fall in the top tenth of the conditional law (11.3 to 15.0; simulated 10.0) and 8.1% in the bottom (7.2 to 9.4; simulated 9.8); given an extension beyond a fifth of the range, 12.7% (10.7 to 15.2; simulated 9.2 to 9.6) and 9.0% (8.0 to 10.0). On WDO the top shares are 9.9% and 10.1%, as under the law, but small extensions are rarer than the law implies (6.7% in the bottom tenth given an extension, 5.9 to 7.7, against 9.8 to 10.1% simulated). On WIN, when the range is extended it tends to be extended somewhat further than a Brownian path with the same variance would extend it; on WDO it does not. Table S3 adds a related pattern on WIN: relative to the simulation, the range is touched no more often (the true-variance gap at an allowance of zero is +0.9 points above the simulated value) but extended by a fifth or three tenths more often (−1.6 and −1.2). Fewer small touches with more large extensions would fit fat-tailed moves, or extremes that hold for a while and then break; both are conjectures we did not test (the sticky alternative of Table S4, which only resists,

produces more holding, not larger breaks). WDO's fewer small extensions and its whole-session variance ratio of 0.97 (0.95 to 0.99, an interval excluding one) suggest mild mean reversion there, also untested.

The search for a cause. Autocorrelation of short-horizon moves over the whole session is not it: the ratio of the variance of 30-minute moves to six times that of 5-minute moves is 0.99 on WIN (0.97 to 1.02), where momentum strong enough to produce a gap of this size would push it to about 1.17 (Section S5). From 15:00 to the close the ratio is 0.93 on WIN (0.88 to 0.99) and 0.96 on WDO (0.91 to 1.01), against 1.00 in the simulation: late-session moves partly reverse, which should make the range hold more, not less, so late-session momentum of the kind documented as intraday momentum (Gao, Han, Li & Zhou, 2018; Baltussen, Da, Lammers & Martens, 2021) does not explain the WIN gap either. Fat tails of moves within five-minute intervals, which realised variance from five-minute closes does not register as range, and price behaviour around B3's closing call remain candidates. We leave them untested.

Volatility that responds to the price. The simulations draw volatility independently of the price, so they contain no feedback from the path to its own volatility, under which the formula is not exact. Two such feedbacks were simulated (`revision/simfb.py`). In the first, volatility rises after down moves (leverage); in the second, it rises after the price sets a new high or low for the session (breakout). Each multiplies the reference simulation's volatility by $\exp(f)$, where f decays by 3% a minute and is pushed by each minute's standardised move or by each new extreme. The strength of each was set to reproduce a statistic of WIN's volatility alone, never whether the range held: the correlation between a five-minute move and the change in variance over the next half hour, and the change in variance in the half hour after a new extreme relative to all minutes. The forecast error's spread was held at the data's 0.66. Table S2d gives the results on 10,000 sessions each. In WIN's data both statistics are close to zero, and slightly on the side of calm after new extremes. Matched to them, the feedbacks are almost absent and the pipeline is the reference's. Even far stronger feedbacks, well outside the data's statistics, raise the conditional slope rather than lower it, and leave the share of large extensions at the law's 10%; strong breakout feedback makes the range hold more often with the true variance (+1.0 point), the opposite of WIN's −1.9, although it does give a true-variance slope above one (1.19, against the data's 1.11). Feedback of these two kinds does not explain the calibration residual or the WIN departure.

Table S2d. Simulations with volatility that responds to the price, against WIN's data

Simulation	Leverage statistic	Breakout statistic	Forecast error sd	Plug-in slope	Conditional slope	True-variance gap	True-variance slope	Top tenth given an extension
WIN data	−0.008	−0.019	0.66	0.58	0.77	−1.9	1.11	13.0%
Reference (volatility independent of the price)	+0.001	+0.016	0.66	0.67	0.93	+0.1	1.00	9.3%
Leverage, matched to the data ($\lambda = 0.002$)	−0.011	−0.002	0.65	0.67	0.93	+0.1	0.99	9.3%
Breakout, matched to the data ($\beta = -0.0025$)	−0.004	−0.017	0.65	0.67	0.93	0.0	0.98	9.3%
Leverage, strong ($\lambda = 0.06$)	−0.206	−0.042	0.70	0.70	0.97	+0.2	0.94	9.3%
Breakout, strong ($\beta = 0.04$)	−0.002	+0.265	0.64	0.77	0.97	+1.0	1.19	9.1%

Statistics of volatility (columns 2–3) measured on one-minute candles, 10:00–17:00; the pipeline columns as in Table 1 and Section S2 (allowance a fifth, five-minute variance). Simulated sessions are 565 minutes long (WIN's long session), so their levels differ slightly from Table 1's reference.

S3. The volatility estimator

Table S2 repeats the level and slope at five sampling intervals and with a subsampled five-minute estimator (the average of the five offset grids; Zhang, Mykland & Aït-Sahalia, 2005), each beside the simulated random walk at the same interval. The level depends on the estimator by several points, in the data and in the simulation alike. Read against the simulation at the same interval, WIN's plug-in level is 0.4 to 2.1 points above it at every interval, and WDO's 1.2 to 2.4 points from five minutes upwards. At one and two minutes WDO is 6.2 and 3.8 points above, and its true-variance version 4.0 and 1.9: the known inflation of realised variance by market-microstructure noise at fine sampling (the volatility signature of Andersen, Bollerslev, Diebold & Labys, 2000; Hansen & Lunde, 2006; Aït-Sahalia, Mykland & Zhang, 2005; Bandi & Russell, 2008), which makes the remaining session look more volatile than it is. This is the episode of the first version, in which a one-minute estimator produced an apparent departure on WDO that a registered test then rejected (the paper's Appendix A); the five-minute estimator of the main results was chosen after it. From five minutes upwards the slopes fall with the interval in data and simulation alike, because a coarser estimate of today's activity makes the forecast noisier.

Table S2. Level gap (held minus says, points) and calibration slope by sampling interval, data and simulated random walk at the same interval

Interval	WIN plug-in gap (sim.)	WIN true-variance gap (sim.)	WIN slope (sim.)	WDO plug-in gap	WDO true-variance gap	WDO slope
1 min	+1.4 (−0.7)	+0.3 (−0.1)	0.67 (0.78)	+5.5	+3.9	0.66
2 min	+0.9 (−0.8)	−0.5 (−0.1)	0.66 (0.75)	+3.0	+1.8	0.66
5 min	−0.5 (−1.5)	−1.8 (−0.5)	0.58 (0.70)	+0.9	−0.5	0.58
10 min	−1.7 (−2.6)	−2.5 (−1.0)	0.50 (0.59)	−0.6	−1.6	0.46
15 min	−3.0 (−3.4)	−3.1 (−1.2)	0.42 (0.46)	−2.2	−2.4	0.36
5 min, subsampled	+0.1	−1.3	0.64	+1.2	−0.2	0.63

Data on all 1,200 (WIN) and 1,198 (WDO) sessions; simulation 6,000 sessions at each interval (the simulated values apply to both markets). The true-variance slope is 1.11–1.16 at every interval in the data and 0.98–1.01 in the simulation.

Forecasting the variance and forecasting the event. A better forecast of the variance need not give a better probability, because an error in the variance matters more for some readings (near a level) than for others. Table S2c compares, for each sampling interval, the variance forecast's accuracy against one common target, the subsampled five-minute variance of the rest of the session, with the plug-in probability's accuracy (`revision/round5.py`). The spread of the log forecast error orders the intervals as the probability's ranking skill and calibration slope do, on both markets (one tie): from 1 and 2 minutes (best) to 15 minutes (worst). The Brier score follows it on WIN but not on WDO. There the one-minute forecast tracks the variance best but has the worst Brier score of the three finest intervals, because microstructure noise inflates its scale and biases the level of the probability (Table S2) while leaving its ranking intact. Tracking the variance buys ranking and slope; getting its scale right buys level.

Table S2c. Variance forecast accuracy (against a common target) and plug-in probability accuracy, by sampling interval

Interval	WIN log error sd	WIN QLIKE	WIN Brier	WIN slope	WIN AUC	WDO log error sd	WDO QLIKE	WDO Brier	WDO slope	WDO AUC
1 min	0.589	0.229	0.1618	0.67	0.717	0.617	0.234	0.1631	0.66	0.703
2 min	0.598	0.241	0.1618	0.66	0.715	0.621	0.254	0.1606	0.66	0.698
5 min	0.645	0.291	0.1656	0.58	0.697	0.659	0.290	0.1635	0.58	0.680
10 min	0.713	0.390	0.1712	0.50	0.675	0.723	0.384	0.1691	0.46	0.655
15 min	0.786	0.518	0.1781	0.42	0.650	0.786	0.494	0.1764	0.36	0.633
5 min, subsampled	0.624	0.265	0.1632	0.64	0.706	0.640	0.270	0.1619	0.63	0.685

Readings shared by every interval (16,737 WIN, 16,772 WDO). Log error: log of the common target over the forecast; its sd removes any constant bias; QLIKE = mean of exp(e) − e − 1, which also penalises the bias.

The forecast error's distribution. The conditional predictive version describes the forecast's log error as normal, with a mean that depends on the morning's scaling ratio. Two richer descriptions were tried (`revision/round4.py`), each estimated at every reading time from the previous 250 sessions only. The first keeps the mean and replaces the normal by the error's own past residuals (25 quantiles). The second lets both the mean and the spread depend on the scaling ratio and on the volatility regime, measured as the previous 20 sessions' average remaining variance relative to its median over the window, and uses the past standardised residuals' own distribution. Table S2b gives the results. On the simulation both are, as they should be, no better than the normal version. In the data neither closes the gap to the simulation: the slopes rise by at most 0.03 on WIN and not at all on WDO, and the Brier scores move by less than 0.001. Neither of these alternatives explains the residual miscalibration; they do not rule out every other shape or regime-dependent mechanism.

Table S2b. Calibration slope (95% CI), level gap (points) and Brier score of the predictive versions (five-minute variance; allowance a fifth of the range)

Version	WIN slope	WIN gap	WIN Brier	WDO slope	WDO gap	WDO Brier	Simulation slope
Plug-in	0.58 (0.46 to 0.71)	−0.6	0.1661	0.58 (0.48 to 0.65)	+0.7	0.1636	0.67 (0.64 to 0.69)
Conditional, normal error	0.77 (0.61 to 0.92)	−1.6	0.1619	0.78 (0.67 to 0.88)	+0.1	0.1578	0.93 (0.90 to 0.96)
Conditional, past residuals	0.77 (0.61 to 0.91)	−1.9	0.1620	0.78 (0.68 to 0.88)	−0.3	0.1576	0.93 (0.90 to 0.96)
Conditional on the regime, past residuals	0.80 (0.68 to 0.94)	−2.4	0.1612	0.78 (0.67 to 0.87)	+0.2	0.1580	0.94 (0.91 to 0.97)

All readings (16,457 WIN, 16,492 WDO; simulation 279,160). CIs: moving-block bootstrap over sessions, blocks of 20 (400 resamples; simulation 200). The normal version here re-estimates the paper's conditional version on all readings with at least 20 past sessions; its slopes agree with Table 1 within 0.01.

S4. The allowance, the trading hours, the change of data source and the feeds compared

The allowance. Table S3 gives the level at allowances of zero (the plain double no-touch event), a tenth and three tenths of the range. Against the simulation, the plug-in level is 0.8 to 2.9 points above on both markets at every allowance, consistent with the forecast bias

described in the paper; the true-variance gaps are within about one point of the simulation on WDO at every allowance, and on WIN 1.2 to 1.6 points beyond it at allowances of a fifth and three tenths, but not at zero or a tenth.

Table S3. Level by allowance (points)

Allowance	WIN held	WIN plug-in gap (sim.)	WIN true-variance gap (95% CI) (sim.)	WDO held	WDO plug-in gap	WDO true-variance gap (95% CI)
0	33.4%	+3.1 (+0.2)	+0.8 (−0.6 to +2.1) (−0.1)	35.0%	+2.9	+0.5 (−0.8 to +1.9)
0.1	46.1%	+1.3 (−0.6)	−0.6 (−2.3 to +0.8) (−0.3)	48.1%	+1.4	−0.7 (−2.1 to +0.8)
0.2	55.5%	−0.6 (−1.4)	−1.9 (−3.5 to −0.3) (−0.3)	59.0%	+0.7	−0.6 (−1.8 to +0.8)
0.3	63.9%	−0.9 (−2.1)	−1.6 (−2.7 to −0.4) (−0.4)	67.8%	+0.8	+0.2 (−0.9 to +1.5)

Rows 0, 0.1 and 0.3 on all sessions; row 0.2 on the sessions of the paper's Table 1. Simulated values in brackets (20,000 sessions).

Trading hours. Restricting the readings to sessions whose 20-session memory has a single closing time (1,056 WIN sessions, 960 WDO) changes the plug-in gap to −0.7 (−2.7 to +1.3) and +0.8 (−1.0 to +2.6), and the true-variance gap to −2.0 (−3.8 to −0.5) and −0.5 (−1.9 to +0.9).

The change of data source. WIN's 2026 sessions without their first 20 (whose volatility memory straddles the change from the broker's to the platform's series) give a plug-in gap of −1.6 (−7.3 to +3.0) and a true-variance gap of −3.0 (−7.3 to +0.3), against −1.3 and −2.8 for all of 2026.

The feeds compared. B3's own trade files for 2021–2026 were not available to us, so we compared the feeds we hold where they overlap (`revision/datacheck.py`). On WIN, the broker's and the platform's series share 184 sessions (30 December 2025 to 24 September 2026). Once the platform's back-adjustment is divided out (one constant per contract, which steps between sessions on the four roll days and never inside a session), high, low and close are identical in all but four of 102,672 matched minutes. Computed on each feed, the 2,520 readings of those sessions give the same outcome in all but one, and plug-in probabilities that differ by 0.0016 on average. One-minute bars built from the platform's trade-by-trade prints (24 to 27 sessions from 21 August 2026) agree with both in high and low on 99.95% of minutes and in the outcome of every reading; one session is incomplete in the prints. On WDO, the two series share 90 sessions (19 May to 23 September 2026) and are identical in every matched minute and every reading. The bars built from WDO's prints differ from both by one tick (half a point) in about half of the minutes, and in 2 of 308 outcomes; the paper does not use them, but the forward test of the paper's Appendix A will, for lack of another daily WDO feed.

Rolls and isolated prints. A contract switched inside a session would show as a single outsized one-minute move. One-minute moves of the close larger than 20 times the session's median occur in 66 WIN and 48 WDO sessions of the broker's series; 7.6% and 12.5% of them fall within a session of an expiry, against 7.2% and 14.6% of all sessions, and most fall at 09:30 or 10:30, when US and Brazilian figures are released. One isolated print was found: WDO on 26 June 2025 at 18:00, 80 points below the minutes on either side (5,411.5 against about 5,490), which sets that session's low and counts all its 14 readings as extended. Leaving the session out changes WDO's held share by 0.05 points and its true-variance gap by 0.06; it is kept. WIN's largest one-minute move (2,060 points at 11:30 on 26 January 2026) is in both feeds, and the session later traded higher still.

S5. What the tests could have detected

The main finding is a non-rejection in level for the plug-in version, so its power matters. Table S4 runs three alternatives through the same generator and pipeline (6,000 sessions each; Section S8): momentum and mean reversion of short-horizon moves, strong enough to move the variance ratio of 30-minute to 5-minute moves to 1.17 and 0.88, and "sticky" extremes, a small drift away from the running high or low whenever the price comes within one minute's standard deviation of it, the support-and-resistance behaviour of Osler (2003).

Table S4. Effect of each alternative against the simulated random walk, and the power of a two-sided 5% test at the data's standard errors (WIN; WDO in brackets)

Alternative	Variance ratio 30/5 min	Plug-in gap shift (power)	True-variance gap shift (power)	True-variance slope shift (power)	AUC
None (random walk)	1.00	–	–	–	0.690
Momentum	1.17	−2.9 points, 0.84 (0.91)	−2.8 points, 0.95 (0.99)	+0.04, 0.09 (0.10)	0.704
Mean reversion	0.88	+2.9 points, 0.83 (0.90)	+2.9 points, 0.95 (0.99)	+0.01, 0.06 (0.06)	0.682
Sticky extremes	0.80	+12.6 points, 1.00 (1.00)	+11.9 points, 1.00 (1.00)	−0.09, 0.29 (0.31)	0.650

Standard errors of the data (WIN / WDO): plug-in gap 1.0 / 0.9 points, true-variance gap 0.8 / 0.7 points, true-variance slope 0.06 / 0.06.

The level tests can detect momentum or mean reversion of this size with probability 0.83 to 0.99, and sticky extremes with certainty; the calibration slope has little power against any of them. At 80% power the smallest detectable shift of the true-variance gap is 2.2 points on WIN and 1.9 on WDO (2.8 and 2.5 for the plug-in gap). Interpolating linearly between the random walk and the simulated alternatives, that corresponds to momentum with a variance ratio of about 1.13, or mean reversion with about 0.91. The whole-session variance ratio in the data (0.99 on WIN, 0.97 on WDO) itself rules out effects of this strength over the whole session.

Weaker effects, and effects confined to part of the session, can escape these tests, and the WIN true-variance gap of −1.9 points is of the size these tests detect only part of the time.

S6. Other markets in full

Bitcoin and Ether. The calculation ran unchanged on the two coins, 2020 to 2023, with the exact probability (the registered test used the 400-step routine; its marks were met for ranking and level and missed for the surest third, the paper's Appendix A). Three routines give the level, and each figure below names its own. (i) This paper's analyses (`revision/other_markets.py`, 12,609 readings on 1,401 days, the first 20 of the 1,421 days kept as the predictive version's memory): the plug-in probability says 32.5% for BTC and 34.0% for ETH, and the range held in 34.3% and 34.4% of readings, gaps of +1.7 (+0.3 to +3.1) and +0.3 (−1.1 to +1.9). (ii) The registered script with the exact probability (12,789 readings on 1,421 days): 32.5% and 34.0% said, 34.4% and 34.3% held, gaps +1.9 and +0.3. (iii) The registered script with its 400-step routine, the registered outcome: 34.0% and 35.4% said, gaps +0.4 and −1.1 (the paper's Appendix A, "largest gap 1.1 points", is ETH's). The probabilities are far more extreme than on WIN (routine ii): in the third of readings where the calculation is least sure it says 1.3% and 1.8% and the range held 6.2% and 6.9%; in the surest third it says 73.4% and 74.6% and the range held 69.0% and 68.2% (the registered routine gave 68.3% for ETH; the difference is the routine). In routine (i), the plug-in calibration slopes are 0.38 and 0.40, the predictive ones 0.68 and 0.67; with the true variance the slopes are 1.55 and 1.48 and the gaps −1.1 (−2.0 to −0.2) and −1.5 (−2.3 to −0.4). We have no simulation calibrated to the coins, so we cannot say how much of this is the method; the forecast's volatility memory is the UTC time of day, which carries the US session (12:00 to 16:00 UTC, up to 2.5 times the mean variance) and the funding times (the first five minutes after 00:00, 08:00 and 16:00 UTC carry 1.2 to 1.5 times the variance of other hours' first five minutes). The ranking skill at the same time of day (routine ii) is 0.681 and 0.684, and 0.728 and 0.740 without the scaling by the day's activity so far. On the coins the fitted model (WIN's six facts, fitted on 2020–2022) does add to the calculation in 2023: 0.735 and 0.791 against 0.679 and 0.685. That is one year and two coins, but it is evidence against a general claim that the random-walk probability exhausts the information in the day's path.

Three index markets, hourly. For the E-mini S&P 500, the E-mini Nasdaq-100 and the DAX only hourly bars of 2026 were available, so volatility comes from each bar's high and low (Parkinson, 1980). The method's own gap on simulated random walks of seven and nine hourly bars is −1.6 and −1.75 points. Net of it, the range held 5.3 points more often than stated on the S&P (0.8 to 9.9), 4.5 on the Nasdaq (−0.1 to 8.9) and 2.3 on the DAX (−1.8 to 6.4); the registered mark was 5 points; under the registered routine the S&P's figure was 5.6 points, which is the registered outcome (failed), and the 5.3 here is exploratory. On WIN cut into hourly bars the same procedure gives +1.4. The calculation ranks readings at 0.668, 0.662 and 0.657. One year of hourly bars from a free source, with a blunter volatility measure, does not settle whether US index futures hold their range more often than a random walk allows.

S7. A one-hour version

A separate registered test (13 October 2026) asked a shorter question on WIN: will the next hour (twelve five-minute candles) make no new session high and no new session low? Its two judged questions, whether candle size alone tells a range, failed; the version below was

reported with it, not judged. Readings are every five minutes from 10:00 in 2025 and 2026 (38,897 readings, successive hours overlapping); the years had been examined in earlier work. The range held for the next hour 60–62% of the time. The distance from the price to the nearer edge of the range ranks the readings with an AUC of 0.83 pooled over times of day when it is counted in points, and 0.87–0.88 when it is counted in expected candles (the distance divided by a fitted forecast of the next five-minute candle's size, fitted to 2024). By fifths of that distance, nearest first, the range held for the next hour 14%, 37–40%, 65–67%, 85–88% and 98% of the time. These AUCs are pooled over times of day and are not comparable with the same-time skill of the paper; the version uses a fitted forecast and has no calibration analysis.

S8. The simulated sessions in full

Each simulated session has 565 one-minute candles (09:00 to 18:24, the long WIN session); each candle is built from 24 Brownian sub-steps, and its high and low are the extremes of those sub-steps and the previous close. The standard deviation of a minute's move is

$$\sigma_{d,\,m} = \exp(h_d)\,\sqrt{\pi_m}\,\exp\left(z_{d,\,k(m)}\right)$$

- h follows an AR(1) across sessions with coefficient 0.72 (measured on WIN after removing a 250-session level) and stationary standard deviation 0.23;
- π_m (the profile) is WIN's measured average squared one-minute change by minute of the session, normalised to mean one (the opening minutes reach eleven times the mean, the closing auction minute about three times);
- z is an AR(1) with coefficient 0.5 over the session's half-hour blocks, stationary standard deviation 0.32, centred so that $\mathrm{E}[\exp(2z)] = 1$. It is the part of the remaining variance that the morning cannot forecast.

The two dispersions were chosen from a small grid so that the simulated readings reproduce the standard deviation of log daily realised variance (0.55 simulated and real) and of the log ratio of the true remaining variance to its forecast (0.66 real; 0.66 simulated). The same pipeline as for the data then runs unchanged: the readings and their filters, the 40-session memory, the volatility forecast, the three probabilities and every measure. Where a bootstrap interval is given for a simulated quantity it measures simulation noise only (20,000 sessions).

Market-calibrated simulations with short sessions. For each market a 4×3 grid of the two dispersions (daily 0.20, 0.23, 0.26, 0.29; within-day 0.29, 0.32, 0.35; 1,000 sessions each; every point recorded in `revision/round3.json`) was searched for the pair closest to that market's two targets. WIN: 0.23 and 0.32 (simulated 0.54 and 0.67 against 0.55 and 0.66), with sessions of 565 and 535 minutes and WIN's intraday pattern. WDO: 0.26 and 0.32 (0.59 and 0.67 against 0.60 and 0.68), with sessions of 570 and 540 minutes and WDO's pattern. Sessions come in blocks of 250, the last 28% of each block short, so that the 20-session memory mixes the two lengths only at the changes, as B3's daylight-saving schedule does; the daily volatility level restarts at each block. Each market's run has 12,000 sessions. Their mean log forecast error is $+0.03$ and $+0.02$, against -0.08 in the data.

The bias-matched variant takes the reference simulation's readings and multiplies the forecast variance by exp(0.115), the difference between the data's mean log forecast error and the simulation's, before computing the plug-in probability; nothing else changes.

The hourly version's simulation is simpler and was registered with that test: sessions of seven or nine hourly bars, each built from 600 Brownian steps at constant volatility, run through the hourly pipeline (Parkinson volatility per bar). Its own gap was −3.3 and −3.5 points under the registered 400-step routine, which overstated the probabilities by up to about two points in that setting, and −1.6 and −1.75 with the exact probability. The registered outcome is the one under the registered routine (the S&P's 5.6 points, failing the 5-point mark); the 5.3 points of Section S6, with the exact probability, is exploratory.

The alternatives of the power analysis use the same generator with one change each: *momentum* adds to each minute a drift equal to 0.15 times the mean one-minute move of the previous five minutes (variance ratio of 30-minute to 5-minute moves 1.17); *mean reversion* uses −0.15 (variance ratio 0.88); *sticky extremes* adds a drift of 0.3 minute-standard-deviations away from the running high (low) in the minute after the last close came within one minute-standard-deviation of it.

References

The works cited here are listed in full in the paper's references: Andersen, Bollerslev, Diebold & Labys (2000); Aït-Sahalia, Mykland & Zhang (2005); Baltussen, Da, Lammers & Martens (2021); Bandi & Russell (2008); Gao, Han, Li & Zhou (2018); Hansen & Lunde (2006); Osler (2003); Parkinson (1980); Zhang, Mykland & Aït-Sahalia (2005).