How flat a spectrum is ranks forecast skill on real series, and real series follow it loosely

A preregistered test of the Kolmogorov-Szego predictability bound on 1,236 tree-ring, mood and machine-power series, against time-shuffled and phase-randomised controls

Zain Dana Harper

Version 1.0, 9 November 2026. Finding F76 in the findings ledger. Strong signal (tier B*).

Abstract

A classical result says how well any linear rule can forecast a stationary series one step ahead, using nothing but the series' spectrum: the flatter the spectrum, the less a forecast can win. We asked whether that ceiling, estimated on the first half of a real series, ranks and bounds the skill of a forecast on the second half. On 1,236 series (1,128 tree-ring chronologies, 106 experience-sampling mood series and two machine power logs) the design was hash-locked before any real-series statistic was computed. One minus spectral flatness ranked held-out one-step forecast skill with Spearman rho 0.414 (95% interval 0.361 to 0.461), against a time-shuffled control whose 95th percentile was -0.023. 80 of 1,236 series (6.5 percent, 5.2 to 8.0) beat the bound beyond estimation error, under the preregistered 10 percent kill. Real series tracked the bound less tightly than copies with the same spectra: rho 0.414 against 0.568 to 0.628 on 39 phase-randomised surrogate sets. A nonlinear analogue forecaster beat the bound in 6.4 percent of series, the same as its false rate of 6.3 percent on planted Gaussian series, so no predictability beyond the linear bound turned up. All six primary verdicts passed and both secondaries failed. The bound works as a rough ceiling with an error band, not as an exact number for one series.

1 What this means for a reader

Before you fit any forecasting model, you can compute from the data alone a rough upper limit on how much a linear forecast can help. On real tree-ring and mood series that limit sorted series by how forecastable they turned out to be, though less sharply than it does on idealised copies. Treat it as a ceiling with an error band: useful for deciding which series deserve modelling effort, unreliable as a promise about any one series.

2 Introduction and motivation

The power spectrum of a series says how its variance spreads over frequencies, from slow swings to fast jitter. A series with all its variance in a few frequencies, like the phases of the Moon, is easy to forecast. White noise spreads its variance evenly and cannot be forecast at all. Spectral flatness, the geometric mean of the spectrum over its arithmetic mean, puts that contrast on a scale from 0 to 1.

The link to forecasting is exact for one class of series. For a weakly stationary series, one whose mean and autocovariance do not change over time, the smallest one-step error variance any linear predictor can reach equals the exponential of the average log spectrum. We call this the Kolmogorov-Szego bound. Divided by the variance, that error is the flatness, so one minus the flatness is the most variance a linear one-step forecast can remove. For Gaussian series the best linear predictor is the best predictor of any kind, so the bound then covers every forecaster.

Davis and Jones [1] turned the formula into an estimator: the average log periodogram plus a constant estimates the log of the innovation variance. Luati, Proietti and Reale [2] later generalised that estimator into a

family they call the variance profile. In machine learning, Goerg [3] built Forecastable Component Analysis on a related quantity, the entropy of the normalised spectrum, and searched for the most forecastable combination of several series. Wang, Klee and Roos [4] proposed a spectral predictability score as a check made before any model is built and reported that it tracked the performance of several forecasting models on the M5 retail data.

Those works use spectral measures as indices of forecastability. We asked a narrower question about the bound itself. On real series that are not exactly stationary or Gaussian, does one minus flatness, estimated on the first half, rank the held-out skill of a linear forecast on the second half, and does it hold as a ceiling once estimation error is allowed for? And does it do so as well on real data as on Gaussian copies with the same spectra? The searches behind this section, run on 9 November 2026, did not return a test that sets real series against their own phase-randomised copies in this way. That is what the searches returned; it does not show that no such test exists.

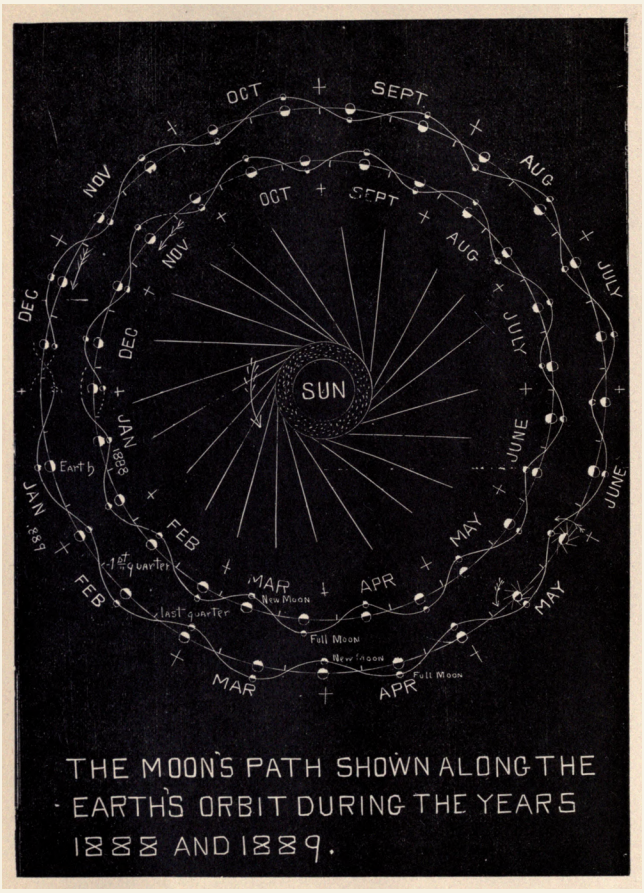


Figure 1. Robert P. Traxler, plate on page 30, "The Moon's path shown along the Earth's orbit during the years 1888 and 1889", from The Principles of Mechanics as Applied to the Solar System (San Francisco, 1889) [8]. Traxler drew his plates to argue for a repelling force from the Sun, his own theory in 1889, which physics did not adopt. The Moon's monthly wave along the orbit is the kind of series whose spectrum sits in a few lines and whose next step is close to certain. Public domain; scan from the Internet Archive, University of California Libraries.

3 Preregistration and design

The preregistration was written on 29 October 2026 and hash-locked at 00:35:53 UTC on 30 October. Before the lock, two things had run: a smoke test on scrambled data, with every real series permuted in time and only

counts and timings printed, and a calibration of the bound estimator on synthetic autoregressive series only. No statistic of an unscrambled real series was computed before the lock.

One change came before the lock and is disclosed in the locked file. The first estimator, a Welch spectrum with a log-bias correction, underestimated the true bound on synthetic series by 0.13, 0.12, 0.06 and 0.02 at lengths 60, 150, 300 and 1000. A Hann-tapered periodogram with the Euler-gamma log correction gave 0.07, 0.04, 0.02 and 0.01. The tapered periodogram became the primary estimator and both Welch forms became secondaries.

The decision the study informs was written in the same file: whether the flatness estimator is adopted into our study tooling as a predictability bound. It is adopted only if the planted check, the ranking test and the ceiling test all pass.

Kill conditions, without discretion:

- The bound's rank correlation with skill at or below the 95th percentile of the time-shuffled control.
- More than 10 percent of series beating the bound beyond estimation error.

The full set of preregistered verdicts and their rules is in Appendix A.

4 Data and provenance

Three kinds of series were read, without modification.

- Tree rings: 1,128 ring-width index series from the International Tree-Ring Data Bank [6], standard chronologies for sites in the conterminous United States, each at least 200 years long with no gaps. Median length 398 years. These are public data from NOAA's National Centers for Environmental Information, with credit to the contributors of each chronology. The site list was the one built for an earlier study of ours.
- Mood: 106 experience-sampling mood series, each at least 100 prompts long, in prompt order, from four openESM datasets [7] (Westhoff, Bailon, Neubauer and Bayer), each released under CC BY 4.0. Median length 113 prompts.
- Machine power: CPU package and GPU watts from one workstation, the longest unbroken run on a 1 second grid, 20,145 seconds each. These two logs are not public.

Tree-ring indices arrive already standardised by their chronology method, which removes slow growth trends before this study sees them. Mood series have night gaps that the prompt order ignores. Series whose training or test half was constant would have been dropped and counted; none was.

5 Methods

The bound

Each series of length n is split in half by time. On the training half, demeaned, the spectrum $S(\omega)$ is estimated with a tapered periodogram I_j at the Fourier frequencies ω_j. The Kolmogorov-Szego formula gives the smallest linear one-step error variance V_1 through

$$\ln V_1 = \frac{1}{2\pi} \int \ln S(\omega) d\omega \tag{1}$$

with the integral taken over frequencies from $-\pi$ to π, and the Davis-Jones form estimates the right-hand side from the periodogram as

$$\ln V_1 \approx \frac{1}{m} \sum_j \ln I_j + \gamma \tag{2}$$

where m is the number of frequencies used and $\gamma \approx 0.5772$ is Euler's constant, which corrects the downward

bias of a log periodogram. With V the variance of the training half, the flatness and the bound are

$$r = \frac{V_1}{V}, b = 1 - r \tag{3}$$

The forecast and its skill

An autoregressive model is fitted by least squares to the mean-centred training half. Its order p runs from 0 to p_{max}, the smaller of 20 and one tenth of the training length, and is chosen by AIC on a common sample, then refitted. On the test half the model forecasts each value from the true past, with the end of the training half as context. The one-step skill is

$$s = 1 - \frac{M}{V_{test}} \tag{4}$$

where M is the mean squared one-step error on the test half and V_{test} its variance. Using the test half's own variance means a level shift between halves counts against skill.

Ranking

The ranking test is the Spearman rho between b and s over all series pooled, with a 95 percent interval from 1,000 bootstrap resamples of series.

The ceiling and estimation error

Skill estimated on a finite half and a bound estimated on the other half both carry error, so a series can beat the point bound by chance. The rule for a real breach uses the series' own surrogates. For each series, 39 phase-randomised surrogates of the whole series, with the mean restored, go through the same pipeline, in the surrogate-data tradition of Theiler and colleagues [5]. With excess $d = s - b$, a violation is a series whose d beats the excess of at least 38 of its 39 surrogates. The per-series p value is

$$p_i = \frac{k_i + 1}{40} \tag{5}$$

where k_i is the number of surrogates whose excess is at or above the real one. For a Gaussian stationary series with that spectrum, violations occur 2 times in 40, or 5 percent, by chance. The share of violating series is reported with a Wilson interval.

The nonlinear check

Because the bound covers nonlinear forecasts only for Gaussian series, an analogue forecaster was run beside the linear model. It works on the standardised series, chooses an embedding dimension from 1, 2, 3, 4 and 6 on the last quarter of the training half, uses the square root of the library size as the number of neighbours, and takes its library from the training half only, at most the last 4,000 states. Its violations are counted by the same surrogate rule.

Multiple testing

A Benjamini-Hochberg correction at 0.05 runs over five p values: the pooled rho, the tree-ring rho, the mood rho, the binomial test that linear violations exceed 5 percent, and the same test for the analogue forecaster.

6 Controls and planted checks

Two controls and two planted checks ran with the same pipeline.

The time-shuffled control permutes every series in time, which destroys both the spectrum and any forecastable structure, and recomputes the pooled rho; 100 draws. Its 95th percentile was -0.023, the cut a real rho must clear. The p value against it is counted as a p value from draws, so with 100 draws the smallest

possible value is 1 in 101, or 0.0099.

The phase-surrogate control keeps each series' spectrum and makes it Gaussian and stationary. On those sets the bound should rank skill well and the violation rule should fire 5 percent of the time. Over the 39 surrogate sets the pooled rho ran from 0.568 to 0.628 (5th to 95th percentile, mean 0.595), and the mean violation rate of each set against the other 38 was 5.1 percent. The violation rule is calibrated.

Planted linear series: 300 Gaussian autoregressive series of order 1 to 4 with random stable roots and lengths drawn from the real tree and mood lengths, whose true bound is known analytically. The estimated bound correlated 0.925 with skill and 0.946 with the true bound, with a mean bias of -0.020 and a mean absolute error of 0.077. Their violation rate was 4.3 percent (2.5 to 7.3).

Planted nonlinear series: 50 noisy logistic maps of length 400, a deterministic nonlinear system that a linear forecast handles poorly. The analogue forecaster beat the linear bound in all 50, with median skill 0.987 against a median bound of 0.084, while the linear forecast violated in 6 percent. The nonlinear check can see nonlinear predictability when it is there.

7 Results

All 1,236 series were kept. Figure 2 sets the ranking against both controls and Table 1 gives the main numbers.

One minus spectral flatness ranked held-out one-step forecast skill with Spearman rho 0.414 (95% interval 0.361 to 0.461), against a time-shuffled control whose 95th percentile was -0.023. The ranking passed the Benjamini-Hochberg correction. A separate permutation of the pairing between bound and skill gave p 0.0005.

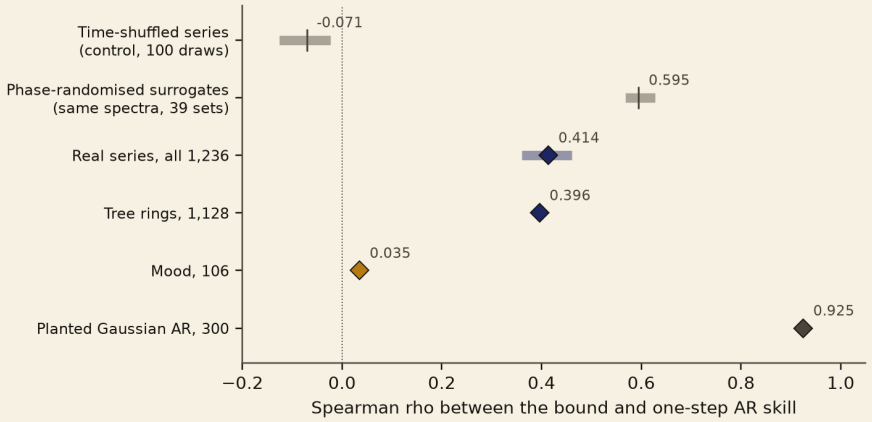


Figure 2. Rank correlation between the bound and one-step AR skill. Grey bars are the 5th to 95th percentile of each control, with ticks at the mean; the blue bar is the bootstrap interval of the pooled real value. Real series sit far above the time-shuffled control and below every phase-randomised surrogate set. Drawn from the run's result file.

Real series tracked the bound less tightly than copies with the same spectra: rho 0.414 against 0.568 to 0.628 on 39 phase-randomised surrogate sets. The real value sits below every one of the 39 sets. Gaussian stationary series with the same spectra let the bound rank skill better than the real series do.

80 of 1,236 series (6.5 percent, 5.2 to 8.0) beat the bound beyond estimation error, under the preregistered 10 percent kill. The rate sits above the 5 percent expected by chance (binomial p 0.013, which passed the correction), yet no single series survives a Benjamini-Hochberg correction across series: 0 of 1,236. Figure 3 shows the rates. Without the surrogate rule the bound would look much weaker as a ceiling: AR skill exceeded the point bound in 36.4 percent of real series and in 45.7 percent of planted Gaussian ones, where the bound is true by construction.

A nonlinear analogue forecaster beat the bound in 6.4 percent of series, the same as its false rate of 6.3 percent

Table 1. Main results on 1,236 series. Intervals are 95 percent: bootstrap over series for rho, Wilson for rates.

Quantity	Value	Interval or comparison
Spearman rho, bound against AR skill	0.414	0.361 to 0.461
Time-shuffled control, 95th percentile	-0.023	p 0.0099, 100 draws
Phase-surrogate sets, rho	0.595 mean	0.568 to 0.628, 39 sets
Violations, AR forecast	80 of 1,236, 6.5 percent	5.2 to 8.0
Violations, analogue forecast	79 of 1,236, 6.4 percent	5.2 to 7.9
Median bound	0.203	
Median AR skill	0.129	
Median excess, skill minus bound	-0.063	
AR skill above the point bound	36.4 percent	planted AR: 45.7 percent

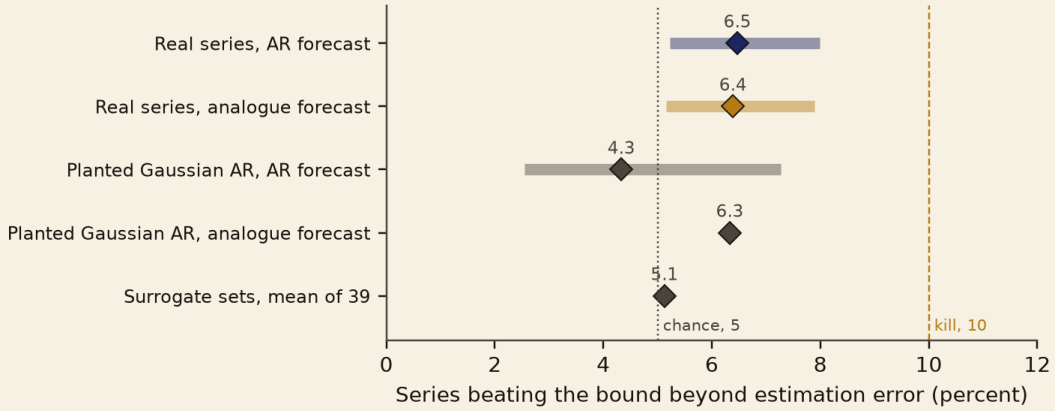


Figure 3. Share of series whose excess skill beats at least 38 of 39 of their own surrogates, with Wilson intervals where the result file gives one. The dotted line is the 5 percent chance rate and the dashed line the 10 percent kill.

on planted Gaussian series, so no predictability beyond the linear bound turned up. Its median skill was 0.044 below the linear model's. The secondary verdict for nonlinear excess failed.

Table 2 breaks the result down by kind of series.

Table 2. Results by kind of series. Rho is Spearman between bound and AR skill; the shuffle cut is the 95th percentile of that group's time-shuffled control.

Group	Series	Rho	Shuffle cut	Median bound	Median skill	Violations
Tree rings	1,128	0.396	-0.030	0.213	0.146	6.1 percent
Mood	106	0.035	0.098	0.024	-0.072	8.5 percent
Machine power	2	none	none	0.638	0.838	2 of 2

Tree rings carry the pooled result: rho 0.396, p 0.0099, passing the correction, with a median AR order of 2. Mood gives rho 0.035 (p 0.19, failing the correction), because its bound, median 0.024, leaves almost nothing to rank, and its median skill is below zero. Both power series beat the bound. Their load drifts between the two halves, so they are not stationary, and the bound does not apply to them.

Two secondaries ran as planned. The spectral exponent alone, the slope of the log spectrum against log frequency on the training half, ranked skill at rho 0.394 against 0.414 for flatness, so that verdict failed; the two measures correlate at 0.697 and the result file gives no interval on their difference, so we treat them as close. On sensitivities, the Welch estimator with segments of one eighth of the series gave rho 0.399 with a lower median bound of 0.099, a second Welch rule gave 0.441, and linear detrending in place of demeaning gave 0.423.

A later round of our own, run on 30 October 2026 with its own preregistration, tested four drift-aware versions of the bound on a split of the same tree-ring and mood series, choosing on 649 development series and testing on 585. None beat the original, which was chosen. On the 585 test series the original gave rho 0.404 against 0.595 on surrogates. The gap between real and surrogate series repeated on a split of the same data, not on new series.

8 Discussion

The bound does two things on real series. It ranks them: series with peaked spectra in the first half are the ones a linear model forecasts best in the second half. And, read with its own surrogate envelope, it holds as a ceiling for about 94 percent of series. The cheap number is useful for triage, the use Wang and colleagues [4] propose for their spectral score.

The part worth more work is the gap. Real series rank less well than Gaussian stationary series with identical spectra. Three explanations compete, and this study does not separate them. Fitting a model on half a series adds estimation error, which the surrogates share but may not share in full. Real series drift, as the power logs do in the extreme, and a bound estimated on one half then describes the other half less well. And real series may carry non-Gaussian structure, though the analogue forecaster found none that it could use. The violation rate sitting slightly above 5 percent fits the same picture: a small excess spread thinly across many series and not concentrated in a few series that clearly break the bound.

The surrogate rule matters for anyone who uses such a bound. Compared against the point bound, skill looked like a breach in a third of real series and almost half of planted ones. Compared against each series' own surrogates, breaches fell to the chance rate in planted data and slightly above it in real data.

9 Limitations

- Tree rings are 91 percent of the series and arrive already standardised, so the pooled rho may come partly from the mix of groups. Mood contributes almost no range of bound to rank.
- The study does not separate estimation error, drift and non-Gaussian structure as causes of the gap between real and surrogate series.
- The bound is a ceiling only after estimation error is allowed for; as a point value it was beaten in 36 percent of real series.
- Only one-step forecasts were tested. Nothing here covers longer horizons.
- The analogue forecaster finding no excess says nothing about stronger nonlinear models.
- The power logs are not public and are not stationary across halves.
- The bound itself is classical, and close heuristics exist; what is tested here is how it behaves on real series against their surrogates.
- The study was not blind to the kinds of data, and scoring was done by our own code.

10 Verified

Verified: on 9 November 2026 the series counts by group, the pooled and group rho values with their intervals and shuffle cuts, the violation counts, rates and Wilson intervals, the binomial p values, the medians, the share above the point bound, every control percentile and mean, the planted linear and nonlinear values, the secondaries and sensitivities, and all eight verdicts were re-read from the run's result file; they match the findings ledger. The preregistration's SHA-256 and the result file's SHA-256 match the hashes recorded beside them. The follow-up values (649 and 585 series, rho 0.404 and 0.595, the original bound chosen) were re-read from the later round's result file. Both data figures were drawn by a script that reads only the result file. The study was not re-run for this paper. The literature in the introduction was found by searches run on 9

November 2026 and read at the level of abstracts, search summaries and publisher pages; the full texts of the cited papers were not read.

11 Corrections

- 29 October 2026, before the lock: the first bound estimator underestimated the true bound on synthetic data, and the tapered periodogram became primary. Disclosed in the preregistration and in the preregistration section here.
- 30 October 2026: a follow-up round tested drift-aware versions of the bound on a split of the same series; none beat the original. Added to the results here.
- 9 November 2026: first published version.

12 Reproduction

- Data: ITRDB standard chronologies for the conterminous United States of at least 200 years with no gaps; the four openESM mood datasets named above, series of at least 100 prompts in prompt order. The two power logs are not public.
- Split each series in half by time. On the training half, demeaned, compute the Hann-tapered periodogram and eq. 2, then the bound of eq. 3.
- Fit an AR model by least squares with order chosen by AIC up to the smaller of 20 and one tenth of the training length; score eq. 4 on the test half from the true past.
- Build 39 phase-randomised surrogates of each whole series, run the same pipeline, and flag violations by eq. 5 at 38 of 39.
- Correlate bound and skill across series with Spearman rho; run 100 time-shuffled controls and a 1,000-resample bootstrap.
- Preregistration SHA-256: fe8ddb061ffc02c3f22d89222fb6b5e20baf402e128c930a85242f161207dae7.
  Plan SHA-256: 4b0268f175706e5803a21cd91d320ebb96e62f3fe97a67639030c3a7bd2c3d26.
  Run 20261030T003557Z. Result file SHA-256: d7c106b84548d9b9859404325c8529c1aaef2192 3c5ebe3ffa0b2d1229c66c28.
- Cost: the analysis took 272 seconds on one desktop computer. Machine energy over the run window was 47.1 kJ, shared with other work on the same machine.

13 Appendix A. The preregistered verdicts

Each verdict is a set of rules written before the run. A verdict passes only when every rule holds. Values are read from the run's result file.

14 Glossary

Each term is defined where it is first used in the text. The same definitions, with the field each comes from and what each excludes, are collected here.

- **power spectrum**. How a series' variance is spread over frequencies, from slow swings to fast jitter. Field: time-series analysis and signal processing. Excludes the timing or phase of the swings, which the spectrum does not record.
- **spectral flatness**. The geometric mean of a power spectrum divided by its arithmetic mean; it is 1 for white noise and near 0 when a few frequencies hold most of the variance. Field: signal processing (speech coding) and the theory of stationary processes. Excludes flatness of the series itself in time, and measures that use the spectrum's entropy instead of its geometric mean.

Table 3. Preregistered verdicts and the rules that decided each one.

Verdict	Deciding rule	Value	Holds
Planted linear series	rho of estimate with skill at least 0.7; with true bound at least 0.8; violations at most 0.10	0.925; 0.946; 0.043	yes
Planted nonlinear series	analogue violations at least 0.5; AR violations at most 0.10	1.0; 0.06	yes
Flatness tracks skill	rho above shuffle 95th percentile (-0.023); passes the correction; bootstrap lower bound above 0	0.414; yes; 0.361	yes
Bound holds	AR violation rate at most 0.10	0.065	yes
Phase surrogates keep the bound	surrogate 5th percentile above shuffle 95th percentile; mean surrogate violation rate at most 0.10	0.568; 0.051	yes
Adopt the flatness unit	every rule of the planted linear, tracking and holding verdicts	all hold	yes
Nonlinear beats the linear bound, secondary	analogue violations above 0.10; above its planted rate (0.063); passes the correction	0.064; yes; yes	no
Exponent alone, secondary	absolute rho of spectral exponent with skill at least the flatness rho (0.414)	0.394	no

- **weakly stationary**. A series whose mean and autocovariance do not change over time. Field: probability and time-series analysis. Excludes series with trends, level shifts or drifting variance, such as a machine whose load changes between the halves.
- **Kolmogorov-Szego bound**. The smallest one-step error variance any linear predictor can reach on a weakly stationary series, equal to the exponential of the average log spectrum; one minus the flatness is therefore the most variance a linear one-step forecast can remove. Field: the theory of weakly stationary processes. Excludes nonlinear predictors on non-Gaussian series, multi-step forecasts, and series whose statistics drift over time.
- **preregistration**. A dated written plan for a study, fixed and hash-locked before any result is seen: the question, the data, the statistic, the controls, the forecasts and the kill conditions. Field: research method (meta-science). Excludes a plan revised after looking at the outcome, and a protocol written only for the record after the run.
- **kill condition**. A rule written in the preregistration that, when met, decides against the claim with no room for judgement after the run. Field: research method (our usage, after Popper's falsification). Excludes reasons found after the run for setting a result aside.
- **verdict**. The pass or fail of one preregistered rule set against the run's numbers, read off by code. Field: research method (our usage). Excludes an interpretation written after reading the numbers.
- **tapered periodogram**. The squared Fourier transform of a series after its ends are faded with a Hann window, used as a raw estimate of the power spectrum at each Fourier frequency. Field: spectral estimation. Excludes smoothed or averaged spectrum estimates such as Welch's method, which we report only as secondaries.
- **autoregressive model**. A linear forecast of the next value from a weighted sum of the last p values, with the weights fitted by least squares; p is the order. Field: time-series statistics. Excludes models with moving-average terms, external inputs or nonlinear terms.
- **AIC**. The Akaike information criterion, a fit score that adds a penalty of two per fitted parameter, used here to choose the autoregressive order. Field: statistical model selection. Excludes choice by held-out error, which would use the test half.
- **one-step skill**. One minus the mean squared error of one-step-ahead forecasts on the test half, divided by the test half's variance; 0 matches forecasting the mean and 1 is perfect. Field: forecast verification. Excludes skill at longer horizons and skill measured on the data used to fit the model.

- **Spearman rho**. The correlation between the ranks of two quantities across cases; 1 means the same order, 0 no relation of order. Field: statistics (rank correlation). Excludes agreement in size, since two quantities can share an order and differ in scale.
- **phase-randomised surrogate**. A copy of a series with the same power spectrum and random Fourier phases, so it keeps the autocorrelation and is Gaussian and stationary by construction. Field: nonlinear time-series analysis (surrogate data). Excludes shuffling in time, which also destroys the spectrum.
- **violation of the bound**. A series whose excess skill over the bound beats that of at least 38 of its own 39 phase-randomised surrogates, which happens to 5 percent of Gaussian stationary series by chance. Field: our usage in this study. Excludes a series whose skill merely exceeds the point bound, which estimation error makes common.
- **Wilson interval**. A 95 percent interval for a proportion that stays inside 0 and 1 and behaves well when the count is small. Field: statistics (binomial proportions). Excludes intervals from resampling.
- **analogue forecaster**. A forecast that finds the k past states most like the present one and averages what came next; it can follow nonlinear structure a linear model misses. Field: nonlinear time-series prediction (method of analogues). Excludes trained nonlinear models such as neural networks.
- **Benjamini-Hochberg correction**. A rule that adjusts a set of p values so that, among the tests called significant, the expected share of false calls stays at or below a chosen rate, here 0.05. Field: statistics (multiple testing). Excludes corrections that control the chance of any false call at all, such as Bonferroni.
- **control**. A version of the data with the structure under test removed (for example shuffled or pooled), scored the same way as the real data, so the real score has something to beat. Field: statistics and experimental design. Excludes a comparison with a different dataset or with a published number.
- **planted check**. Data built so that the right answer is known, run through the same analysis before the real data, to show that the analysis can see the effect when it is there. Field: research method; called a positive control in experimental science. Excludes any check made on the real data, which cannot show what the analysis misses.
- **p value (from draws)**. The share of control draws scoring at or above the real score, with one added to the count and to the number of draws, so it is never zero. Field: statistics. Excludes a p value from a textbook formula that assumes a distribution.

15 References

[1] Davis, H. T. and Jones, R. H. (1968). Estimation of the innovation variance of a stationary time series. Journal of the American Statistical Association 63 (321), 141-149. doi:10.1080/01621459.1968.11009229

[2] Luati, A., Proietti, T. and Reale, M. (2012). The variance profile. Journal of the American Statistical Association 107 (498), 607-621.

[3] Goerg, G. M. (2013). Forecastable component analysis. Proceedings of the 30th International Conference on Machine Learning, PMLR 28 (2), 64-72. arXiv:1205.4591

[4] Wang, R., Klee, S. and Roos, A. (2025). Time series forecastability measures. arXiv:2507.13556

[5] Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. and Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. Physica D 58, 77-94. doi:10.1016/0167-2789(92)90102-S

[6] International Tree-Ring Data Bank. NOAA National Centers for Environmental Information, World Data Service for Paleoclimatology. Standard chronologies, retrieved October 2026. https://www.ncei.noaa.gov/products/paleoclimatology/tree-ring

[7] openESM experience-sampling datasets 0008 Westhoff, 0018 Bailon, 0062 Neubauer and 0076 Bayer. Zenodo, CC BY 4.0. doi:10.5281/zenodo.17347510, doi:10.5281/zenodo.22763396, doi:10.5281/zenodo.17347974, doi:10.5281/zenodo.22911022

[8] Traxler, R. P. (1889). The Principles of Mechanics as Applied to the Solar System. San Francisco: C. A. Murdock and Co. Internet Archive item principlesofmech00traxiala.