Forecasts drift, mood steps or drifts in many people, and five prompts a day miss fast emotion

A preregistered run of change-point, drift and switching tests on forecast questions and phone-survey mood series, with AR(1) surrogates, in which 11 of 20 predictions did not hold and one test could not fire

Zain Dana Harper

Version 1.0, 14 November 2026. Finding F38 in the findings ledger. Signal (tier B), mixed.

Abstract

Most statistics for a series of measurements assume the process behind it holds still. We built tests for three ways it can move (a step, a slow drift, a switch between two states), checked each on synthetic series with known answers, and ran them on crowd forecasts and on mood reported several times a day. Twenty predictions were fixed in advance. The ledger scores 9 of the 20 as held and 11 as not held, and one of the 9 held only because its rule could never fire. In 238 forecast questions the drift test fired on 97 percent and 96 percent ended nearer the answer than they began, by a median 0.24 in probability, which is what a crowd converging on an outcome is expected to do. On the two slider datasets the change-point test fired in 22 and 40 percent of people and the drift test in 26 and 43 percent, against about 5 percent per test for a stationary AR(1) series. On the 1 to 7 and 1 to 5 scales the raw change-point rates of 47 and 70 percent fell to 18 and 12 percent once ties were broken, so most of those detections came from the rating scale. The switching test used 19 surrogates, so its smallest p value was 0.05 and the rule "p below 0.05" could never fire; no claim about switching is made. Simulated people whose mood relaxes in 5, 15 or 30 minutes, sampled five times a day for 20 days with no measurement noise, were fitted at the sampling floor or with a noise share of 0.8 or more in 89, 98 and 89 percent of 100 runs, against a preregistered bar of 90 percent.

1 What this means for a reader

A crowd forecast moves steadily toward the right answer as the question nears its end, so its average over the whole period describes no moment of it. Mood reported on a phone is steadier, but in a large minority of people it steps or drifts over the weeks of a study by more than their own hour-to-hour carryover explains. A study that prompts five times a day also cannot see feelings that rise and fade within the hour: they show up as noise. More than half of what we predicted about these data was wrong, and one of our tests was built so that it could not give a positive answer, which we report as a defect.

2 Introduction

A series is stationary when its mean and its autocovariance do not change over time, a term from time-series analysis that excludes trends, level shifts and drifting variance. Measures such as an autocorrelation or a spectrum summarise a whole record in a few numbers, and those numbers describe the process only if it is stationary. A recent primer on non-stationarity in psychological time series [4] warns, in the summary we read, that ignoring a trend biases estimates and distorts their sampling distributions.

Mood research meets the problem in experience sampling, a design in which people answer short questions on a phone several times a day. Such data are sparse and unevenly spaced, and how spacing biases models fitted to them is a studied question [7]. A simulation study found that popular time-series tools lose the dynamics of a toy emotion model once measurements are as sparse as experience sampling, and its authors suspected a

mismatch between prompts hours apart and processes that unfold over seconds to minutes [5]. A Monte Carlo study of continuous-time autoregressive models found estimates degraded when the process was observed more slowly than it ran [6].

We asked three things. Which of the processes we hold data on are stationary, by tests that can tell a step from a drift? How widely do people differ in the time their mood takes to relax? And what is the fastest mood process a five-a-day schedule can measure? The decision this informed was which of our own stationary measures may be applied to which data. Our searches returned the work above and no published share of experience-sampling participants whose series fail a stationarity test; that is what two searches returned and no proof that none exists.

3 Preregistration and design

The preregistration, a dated plan fixed and hashed before any result is seen, was written on 29 October 2026 and its SHA-256 recorded at 15:54:42 UTC that day, before any measure was run on any real series. Before writing it we had seen file headers, counts of people and prompts, record spans, and the results of the synthetic recovery runs. It fixes 20 predictions, each scored held or not held as written by our own code and reading of the result files, with nothing re-scored after the run and every unplanned analysis labelled exploratory. Each scored prediction is a verdict. Appendix A lists all 20.

4 Data and provenance

- Good Judgment Project survey forecasts, years 1 to 4, CC0 1.0 [1]. We kept closed binary questions open for at least 60 days with forecasts on at least 60 days: 238 questions. Each series is the daily median probability given to the outcome that occurred, with empty days carried forward.
- Four datasets from the openESM database [2], each CC BY 4.0: 0008 (valence on a 0 to 100 slider), 0018 (valence on a slider from -50 to 50), 0062 (happiness, 1 to 7) and 0076 (positive affect, 1 to 5). People with at least 60 usable prompts and 3 distinct values were kept: 113, 70, 34 and 373, in total 590. Only distributions across people are reported.
- Forest Inventory and Analysis remeasurements for Oregon and Washington [3], a United States Government work in the public domain: annual mortality share and survivor growth by remeasurement year, 14 and 12 years.
- Two looping animations rendered by our own software, 288 and 144 frames, tiled four times. They are a known periodic control and are not public data.

5 Methods

The three tests

Each test compares a series with surrogates of itself: simulated series from an AR(1) process fitted to it, in which each value is the mean plus a fraction φ of the previous value's distance from the mean plus Gaussian noise e_t,

$$x_t = \mu + \varphi(x_{t-1} - \mu) + e_t \tag{1}$$

with the same length, mean, standard deviation and φ as the real series. A surrogate keeps the memory between neighbouring values and has no step, drift or switching. For a statistic that is k times matched or beaten among N surrogates,

$$p = \frac{1 + k}{N + 1} \tag{2}$$

so the smallest possible value is $1/(N + 1)$.

- Change-point test: the number of changes in mean or variance found by a penalised search (PELT, penalty $3 \ln n$, segments of 5 or more), against 99 surrogates.
- Drift test: a model whose AR coefficient follows a random walk, $\varphi_t = \varphi_{t-1} + \varepsilon_t$, fitted by a Kalman filter; the likelihood ratio against a fixed coefficient, against 49 surrogates.
- Switching test: a two-state Markov-switching AR(1) against a single AR(1), likelihood ratio, against 19 surrogates.

A series is labelled switching if the third test gives p below 0.05 with at least 3 switches, else shift if the first does, else drift if the second does, else stationary. In a stationary AR(1) world each test fires about 5 percent of the time.

Timescales

For each person we fitted an Ornstein-Uhlenbeck time constant τ to the real prompt times. The model is a mood that relaxes toward its mean, observed with independent noise, so two prompts d hours apart correlate as

$$r(d) = (1 - \rho)e^{-d/\tau} \tag{3}$$

where ρ is the share of variance that is noise. A person is at the sampling floor when the fitted τ is shorter than their median gap between prompts within a day; there the time constant is not measured.

The aliasing simulation

Aliasing, a term from signal processing, is what happens when a process changes faster than it is sampled: the fast change is recorded as something else, here as noise. We simulated people from eq. 3 with $\rho = 0$ and true τ from 5 minutes to 48 hours, on a schedule of 20 days with 5 prompts a day, 100 runs per value, and fitted them as real people were fitted.

6 Controls and planted checks

Every measure first had to recover known values from synthetic series, which is a planted check: data built so the right answer is known. Over 40 seeds, the median error had to be below the 5th percentile of a shuffled pairing of recovered and planted values and at most half its median. In the recovery file 15 measures pass every rule. Three are recorded as failing and were not used for claims: a spectral slope on five-a-day sampling (ratio 1.04), variance shares in three bands on records under 90 days, and a fit of two timescales in one person (ratios 0.63 and 0.89).

Two controls, estimators that should fail, passed at first. A single spectral peak passed the test for a frequency sweep (ratio 0.44) and a single AR coefficient passed the test for a drifting one (0.49), because each correlates with the endpoints. The rule was changed so that every moving case is also scored on the movement itself, where both fail at ratio 1.00. A third control showed that a switching model fitted to a shuffled series still recovers the gap between its two levels from the histogram alone; only the durations of the states need time order.

The change-point search was also run on pure noise, 200 series per setting. It reported a change in 2 to 6 percent of white-noise series, and in 83 to 100 percent of AR(1) series with φ of 0.6 or more. That is why no real series is read from its raw change count and each is judged against AR(1) surrogates.

7 Results

The ledger scores 9 of the 20 as held and 11 as not held, and one of the 9 held only because its rule could never fire.

The defect

The switching test used 19 surrogates, so its smallest p value was 0.05 and the rule "p below 0.05" could never fire; no claim about switching is made. The switching share is 0 in every dataset by construction, and the prediction that it would be under 15 percent held for that reason alone. An exploratory rerun with "p at most 0.05" and at least 3 switches flagged 47 to 67 percent of people in the mood datasets and 54 percent of forecast questions. Those rates cannot be read as switching: Gaussian surrogates do not copy a bounded, skewed rating scale, and the control above shows the model reads level gaps from the histogram. A surrogate that keeps each series' own distribution of values is the test this needs, and it was not run.

Forecasts

In 238 forecast questions the drift test fired on 97 percent and 96 percent ended nearer the answer than they began, by a median 0.24 in probability, which is what a crowd converging on an outcome is expected to do. Only 1.7 percent of questions were labelled stationary. We had predicted that steps would be the most common label, on the idea that news arrives as jumps; drift was, at 78 percent of questions against 20 percent for shift. The median question ran 126.5 days.

Mood

Table 1 and Figure 1 give the labels and test rates. On the two slider datasets the change-point test fired in 22 and 40 percent of people and the drift test in 26 and 43 percent, against about 5 percent per test for a stationary AR(1) series. These shares rest on 113 and 70 people and have no intervals. We had predicted that more than half the people in every dataset would be labelled stationary, and that held in one dataset of four.

Table 1. Labels by priority (switching, shift, drift, stationary) and the rate at which each test fired at p below 0.05. Switching is 0 throughout because of the defect.

Series	n	Stationary	Shift	Drift	Change-point test	Drift test
Mood 0008, slider	113	0.58	0.22	0.19	0.22	0.26
Mood 0018, slider	70	0.36	0.40	0.24	0.40	0.43
Mood 0062, 1 to 7	34	0.32	0.47	0.21	0.47	0.35
Mood 0076, 1 to 5	373	0.25	0.70	0.05	0.70	0.17
Forecast questions	238	0.02	0.20	0.78	0.20	0.97

On the 1 to 7 and 1 to 5 scales the raw change-point rates of 47 and 70 percent fell to 18 and 12 percent once ties were broken, so most of those detections came from the rating scale. A coarse scale gives runs of identical answers (a median 45 and 46 percent of consecutive prompts tie in these two datasets), a run of identical values has zero variance, and a Gaussian surrogate never produces one. Dequantising, adding uniform noise one scale step wide, removes the ties; it was an exploratory correction made after the run. On the sliders, where 3 to 4 percent of consecutive prompts tie, the rates barely moved (0.22 to 0.24 and 0.40 to 0.39).

Timescales and the sampling floor

Table 2 gives the fitted time constants. The median lies between 2 and 48 hours in each dataset, and across all 590 people the 5th to 95th percentile range runs from 0.54 to 109 hours, a factor of about 200. Between 13 and 40 percent of people sit at the sampling floor. We had predicted at least 20 percent in every dataset, which held in two.

Simulated people whose mood relaxes in 5, 15 or 30 minutes, sampled five times a day for 20 days with no measurement noise, were fitted at the sampling floor or with a noise share of 0.8 or more in 89, 98 and 89 percent of 100 runs, against a preregistered bar of 90 percent. The prediction missed its bar by one run in a hundred at two of the three values. Its second rule held: the median correlation between consecutive prompts

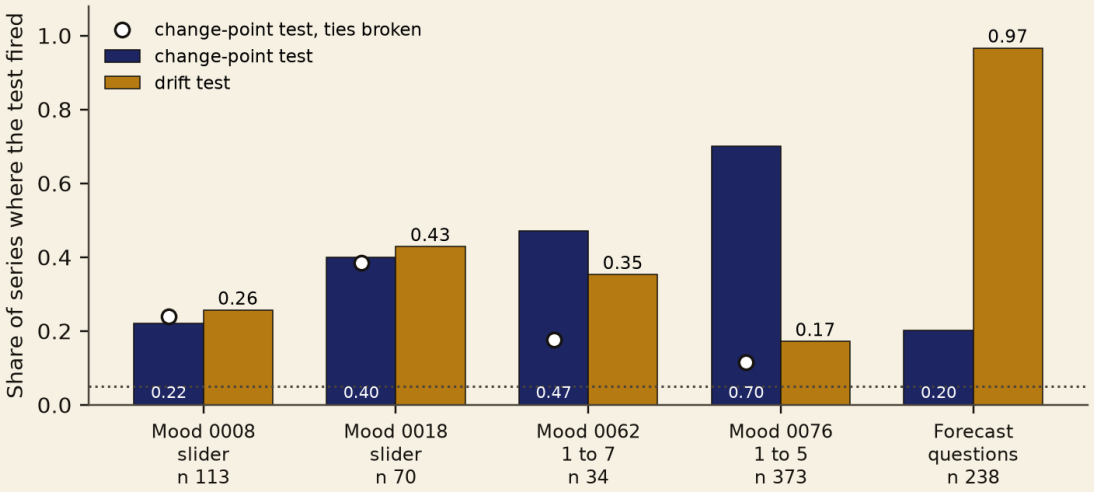


Figure 1. Share of series on which the change-point test and the drift test fired, with the change-point rate after dequantising (open circles, mood only) and the 5 percent rate expected of a stationary series (dotted line). Drawn from the study's result files.

Table 2. Fitted time constants in hours, the share of people at the sampling floor, and the median gap between prompts within a day.

Dataset	n	Median	5th to 95th percentile	At floor	Gap
0008	113	10.0	1.6 to 72	0.23	3.2
0018	70	17.0	1.8 to 809	0.13	3.4
0062	34	13.5	1.2 to 346	0.18	2.0
0076	373	3.5	0.29 to 49	0.40	2.7
All	590	5.9	0.54 to 109	none	none

was 0.04, -0.02 and 0.01. Figure 2 shows the whole sweep. A true time constant of 1 hour gave the same 89 percent; at 3 hours the share fell to 22 percent, and at 12 and 48 hours to zero, with fitted values of 11.5 and 37.6 hours. The schedule measures time constants of about 3 hours and longer and folds everything faster into one floor. Sampling theory predicts this; the simulation sizes it for this schedule.

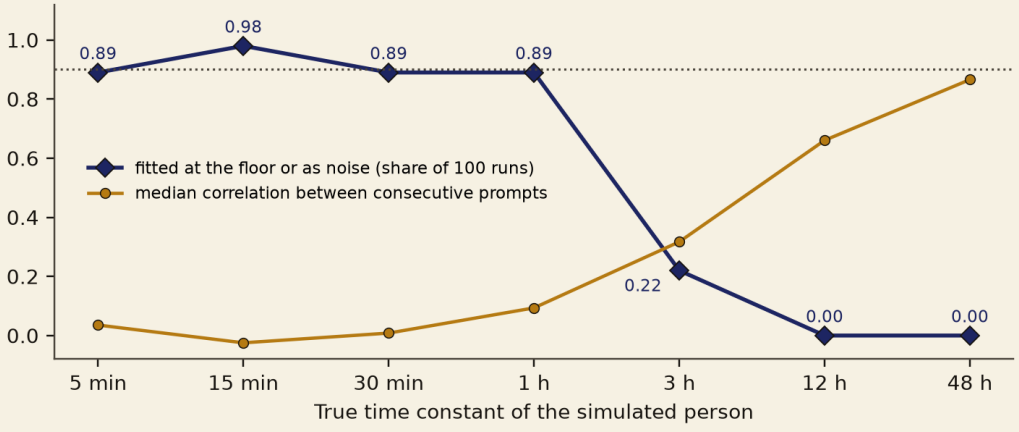


Figure 2. Aliasing by simulation: for each true time constant, the share of 100 simulated people fitted at the sampling floor or with a noise share of 0.8 or more (navy), and the median correlation between consecutive prompts (ochre). The dotted line is the preregistered bar of 0.9. Drawn from the study's result files.

Three further timescale predictions did not hold. A model with two timescales beat one on each person's held-out prompts for 50 to 60 percent of people, where we had predicted at most half; people simulated with one timescale on the same prompt times picked two 46 to 51 percent of the time, so the comparison is close to a coin flip and supports no claim either way. In the 37 people with records of 60 days or more, the median share of variance arriving within a day was 0.72 (interquartile range 0.64 to 0.84) against a predicted range of 0.3 to 0.7, a narrow miss. A two-component mixture was preferred over one on that share (BIC -20.3 against -14.8), against our prediction of a continuum; a cluster of people at the floor beside a spread above it would give the same preference, and that was not tested. As predicted, 27 percent of those 37 people could not be told from their own shuffled series.

The other series

The two animation loops met two of their four predictions: every change point repeated at the same frame in the next loop, and every feature found repeated with the loop. A mode-finding method missed one loop's period because its window of 80 frames was shorter than the 288-frame loop. The forest inventory series, 14 and 12 annual values, are too short for any of the tests; a first run labelled them drifting through a software fault on short series, which was fixed and the faulty output kept. Oregon's mortality share was 0.077 to 0.096 from 2011 to 2020 and 0.125, 0.143, 0.153 and 0.178 from 2021 to 2024. Each year measures a different panel of plots, so this is a description and no test.

8 Discussion

The forecast result is the cleanest and the least surprising. A crowd's probability for the outcome that happens rises over a question's life, slowly more often than in jumps. Any stationary summary of a forecast series, such as one autocorrelation for the whole question, averages over a process that is changing throughout.

For mood the supported statement is narrower than our write-up first made it. On continuous sliders, a quarter to two fifths of people show a level change or a drifting coefficient beyond an AR(1) with Gaussian noise, each test taken alone. A rejection means "more than AR(1) with a Gaussian marginal", which a long-memory or skewed process also produces. On coarse rating scales the change-point test mostly detected the scale. A researcher fitting one autoregressive model per person to several weeks of slider data should expect it to be wrong in form for a sizeable minority, and should not use a variance-based change-point cost on a 5-point scale.

The sampling result agrees with the simulation literature cited above [5,6]: a schedule of prompts hours apart cannot recover a process that runs in minutes. Our addition is the size for one common design. With five prompts a day, a person whose feelings turn over within the hour is indistinguishable from a person who is half noise. Measuring that person needs a continuous signal.

The defect has a general form. A surrogate count sets the smallest p value a test can return, and a threshold at or below that value makes the test silent. The count has to be chosen from the threshold.

9 Limitations

- Eleven of 20 predictions did not hold, and one that held could not have failed. A further one, that the inventory series are too short to test, is a statement about series length.
- The switching test could not fire. No switching result exists, in either direction.
- The headline shares have no intervals. They rest on 113, 70, 34 and 373 people and 238 questions.
- The surrogates copy mean, variance and lag-1 memory only, with Gaussian noise. Bounded probabilities and rating scales sit outside that family.
- The tie correction and the switching rerun were exploratory and made after the results were seen.
- The stationarity tests order prompts and ignore the gaps between them, including nights.

- The aliasing result comes from simulation of one model and one schedule, and follows from sampling theory. It missed its own bar of 90 percent at two of three values.
- That forecasts move toward the answer is expected, and the series was defined as the probability of the realised outcome.
- Forest inventory years are different plot panels; those numbers are descriptive.
- Scoring was done by our own code and our own reading of the result files, and was not blind.

10 Verified

Verified: on 14 November 2026 the label shares, test rates, forecast rise share and median, tie shares and dequantised rates, exploratory switching rates, time-constant percentiles, floor shares, held-out comparison shares, within-day share and mixture values, every aliasing value, the inventory series, the loop results, the noise false-positive rates and the recovery ratios were re-read from the study's result files. They match the findings ledger, our append-only dated list of recorded results, with one exception given in the corrections. The preregistration's SHA-256 was recomputed and matches the recorded hash. The held and not-held scores were read from the study's results text and checked against the result files. Dataset licences were read from our download record. Both figures were drawn by a script that reads only the result files. The study was not re-run for this paper. The literature was found by two searches run on 14 November 2026 and read at the level of search snippets; one further work is cited from an earlier search note of ours at the same level.

11 Corrections

- 29 October 2026: found during the run. The switching test's 19 surrogates made its rule unable to fire; a fault on series shorter than the drift model's burn-in labelled four inventory series as drifting and was fixed; two controls passed an endpoint-only rule and the rule was tightened.
- 31 October 2026: a review of our ledger found that the 97 percent quoted for forecast drift is the rate of the drift test and that the drift label covers 78 percent of questions. This paper gives both.
- 14 November 2026: the ledger counts 16 measures passing recovery. The recovery file marks 15 as passing every rule, and this paper uses 15.
- 14 November 2026: narrowed the mood statement. The ledger says a quarter to two fifths of people on sliders step or drift; the files give that range for each test taken alone (0.22 and 0.40; 0.26 and 0.43), and 42 and 64 percent of people carry either label.
- 14 November 2026: first published version.

12 Reproduction

- Data: Good Judgment Project survey forecasts, Harvard Dataverse, CC0 1.0, doi:10.7910/DVN/BPCDH5; openESM datasets 0008, 0018, 0062 and 0076, CC BY 4.0, doi:10.3758/s13428-026-03112-y for the database; USDA Forest Service FIA DataMart, public domain, https://apps.fs.usda.gov/fia/datamart/CSV/.
- Build each forecast series as the daily median probability of the realised outcome. Keep mood series of 60 or more prompts with 3 or more distinct values.
- For each series fit eq. 1, draw surrogates, and compute eq. 2 for the change-point count (99 surrogates), the drift likelihood ratio (49) and the switching likelihood ratio (19; use 39 or more to make the 0.05 rule able to fire).
- Fit eq. 3 to each person's real prompt times. Simulate eq. 3 on 20 days of 5 prompts for the aliasing sweep.
- Preregistration SHA-256: cf81972b4975844362218d58ca863efe286f70748879a3064f8b33bb647e8190. The study was run by its own scripts, so the files record no plan hash and no run identifier.

13 Appendix A. The preregistered verdicts

Each prediction is scored as written. Values are read from the study's result files.

Table 3. The 20 preregistered predictions and their outcomes.

Prediction	Value	Outcome
Mood: stationary label over half in every dataset	0.58, 0.36, 0.32, 0.25	not held
Mood: shift or drift above 14 percent in two datasets and under 40 in all	0.42, 0.64, 0.68, 0.75	not held
Mood: switching under 15 percent	0, by the defect	held, could not fail
Forecasts: at least 60 percent not stationary	0.98	held
Forecasts: shift the most common label	shift 0.20, drift 0.78	not held
Forecasts: at least 75 percent rise toward the answer	0.96	held
Loops: top mode at a multiple of the loop frequency within 1 percent	one loop 0.5 percent off; the other 5.3 percent off	not held
Loops: change points repeat at the same phase	22 of 22 in one loop; none found in the other	held
Loops: drift path repeats, r above 0.8	1.00 and -0.35 (a flat path)	not held
Loops: every reported feature repeats with the loop	all repeat	held
Median time constant 2 to 48 h; pooled spread at least 1.5 decades	3.5 to 17.0 h; 2.3 decades	held
At least 20 percent at the floor in each dataset	0.23, 0.13, 0.18, 0.40	not held
Aliasing: at least 90 percent at floor or noise; lag-1 under 0.1	0.89, 0.98, 0.89; 0.04, -0.02, 0.01	not held
Two timescales win for at most half; gain under 0.02 nats	0.55, 0.60, 0.50, 0.56; gains under 0.001	not held
Within-day variance share between 0.3 and 0.7	0.72	not held
One mixture component preferred	two preferred	not held
At least 20 percent of long records match their shuffle	0.27	held
Inventory series too short to test	14 and 12 years	held, by series length
Tree growth slower in the second period in both states	-0.01 and 0.00 inches a year	not held
Mortality share highest from 2021 on in one state	2024 and 2022	held

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.

- **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.
- **experience sampling**. A study design in which people answer short questions on a phone several times a day, so that mood is recorded close to the moment it is felt. Field: psychology (methods). Excludes one-off questionnaires and recall of past weeks.
- **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.
- **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.
- **surrogate series**. A simulated series that keeps chosen properties of a real one, here its length, mean, standard deviation and lag-1 memory, and nothing else, so a statistic on the real series can be compared with what those properties alone produce. Field: time-series statistics. Excludes the real series' distribution

of values, which a Gaussian surrogate does not copy.

- **AR(1) process**. A series in which each value is the mean plus a fixed fraction of the previous value's distance from the mean, plus fresh random noise. Field: time-series statistics. Excludes steps, drifting coefficients, switching between states and long memory.
- **change point**. A time at which the mean or variance of a series changes; here found by a penalised search that adds a change only when it improves the fit by more than a fixed penalty. Field: time-series statistics. Excludes gradual change, which the drift test reads.
- **drift test**. A comparison of an autoregressive model whose coefficient follows a random walk with one whose coefficient is fixed, by likelihood ratio. Field: time-series statistics, our usage. Excludes a drift of the mean level alone with a constant coefficient, which this test does not target.
- **switching test**. A comparison of a model with two hidden states, each with its own level and spread, that the series moves between, with a single AR(1), by likelihood ratio. Field: econometrics (Markov-switching models). Excludes a one-way step, which the change-point test reads.
- **Ornstein-Uhlenbeck time constant**. The time over which a quantity that relaxes toward its mean loses about 63 percent of a deviation, fitted here from each person's prompts at their real times with a separate noise term. Field: stochastic processes. Excludes a second, slower or faster timescale in the same person, which this fit does not model.
- **sampling floor**. The condition in which a person's fitted time constant is shorter than their median gap between prompts within a day, so the value is not measured. Field: our usage, from sampling theory. Excludes a claim that the person's mood is fast; noise gives the same reading.
- **aliasing**. The misreading that results when a process changes faster than it is sampled, so that the fast change appears in the record as something else, here as noise. Field: signal processing. Excludes measurement error in the instrument, which is present even with dense sampling.
- **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.
- **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.
- **dequantising**. Adding uniform random noise one scale step wide to values recorded on a coarse scale, so that identical neighbouring answers stop being exactly equal. Field: signal processing. Excludes any recovery of the finer value the person would have given on a continuous scale.
- **findings ledger**. The append-only, dated list of every result we have recorded, numbered F1, F2 and on; corrections are added as new dated lines and old entries are never rewritten. Field: our records. Excludes a curated list of successes.

15 References

[1] Mellers, B. and others. Good Judgment Project data: survey forecasts, years 1 to 4, and question file. Harvard Dataverse, CC0 1.0. doi:10.7910/DVN/BPCDH5

[2] Siepe, B. S., Haslbeck, J. M. B., Kloft, M., Büchner, A., Zhang, Y., Fried, E. I. and Heck, D. W. (2026). Introducing openESM: a database of openly available experience sampling datasets. Behavior Research Methods 58, 240. doi:10.3758/s13428-026-03112-y. Datasets 0008, 0018, 0062 and 0076, CC BY 4.0.

[3] USDA Forest Service. Forest Inventory and Analysis DataMart, Oregon and Washington tables. Public domain. https://apps.fs.usda.gov/fia/datamart/CSV/

[4] Ryan, O. and others. Non-stationarity in time-series analysis: modeling stochastic and deterministic trends. Multivariate Behavioral Research. doi:10.1080/00273171.2024.2436413

[5] Recovering within-person dynamics from psychological time series (2021). Multivariate Behavioral Research. doi:10.1080/00273171.2021.1896353. The search result did not show the authors.

[6] Consequences of sampling frequency on the estimated dynamics of AR processes using continuous-time models. PubMed record 37428727. The search result did not show the authors.

[7] De Haan-Rietdijk, S. and others (2017). Discrete- vs. continuous-time modeling of unequally spaced experience sampling method data. PubMed Central record PMC5655034.