

Detecting Slow Cycles in Historical Panels

An In Vitro Study

Stefano Angeli

Foundations Institute of Computational Social Science (FICSS), Milano

`stefano.angeli@ficss.institute` ORCID 0009-0007-6643-2307

Version 1.0, October 2026

WP-2026-007 · DOI 10.5281/zenodo.22952295

Abstract

Secular cycles, oscillations of political instability with periods of one to three centuries, are a central claim of structural-demographic theory and the target of the Koopman programme of this series (WP-2026-005, WP-2026-006). Before a historical test can be read, one must know whether it could have seen a cycle if there were one. We answer this question in vitro. We plant a slow cycle of known period and coherence in simulated societies, and we measure the power of a size-calibrated pooled Hankel-DMD test against a null twin without the cycle. We also measure the forecast value of the detected cycle for crisis onsets 10 to 200 years ahead. The design, the hypotheses and the decision rules were preregistered, and the verification was run once on fresh seeds. For a 200-year cycle with coherence time three periods and moderate observation noise, observed at decadal cadence with a 70-year delay window, power 0.8 needs about 350–390 pooled decadal columns, however they are split between societies and years: 32×200 years, 8×500 , $4 \times 1,000$ or $2 \times 2,000$. The strict historical panel of independent states since 1789 sits at the edge, with power 0.79 [0.73, 0.84] (a post hoc interval; the preregistered rule counts detections, §6). A noisy proxy of the slow driver lifts it to 1.00. The persistence criterion used in WP-2026-006 has power at most 0.07 on its own premise. Once detected, the cycle's phase beats a renewal model by 0.16 of Brier skill at 100 years, as preregistered; in an exploratory comparison it falls short of a hazard model on the observed state, which keeps positive skill up to 200 years. In vitro, long-horizon skill comes from the observed state and the model class, not from spectral detection as such.

Keywords: Computational Macrohistory, Koopman operator, Dynamic Mode Decomposition, secular cycles, statistical power, size-calibrated test, forecast skill, in vitro

JEL Classification: C12, C15, C63, N00

FICSS Working Paper Series No. WP-2026-007. Angeli, S. (2026). Detecting Slow Cycles in Historical Panels: An In Vitro Study. FICSS Working Paper WP-2026-007. Foundations Institute of Computational Social Science (FICSS), Milano. DOI: 10.5281/zenodo.22952295

1 Introduction

WP-2026-005 distinguished three levels of the CMH prediction space: events, regime probabilities, and the slow spectral structure of a society's dynamics. It argued that the third level is where the predictive horizon can extend beyond the Lyapunov wall, and it proposed a first empirical test. WP-2026-006 gave that test a formal treatment and ran it on eight Western European countries. It found no persistent mode. Version 1 of WP-2026-006 already read that negative result as a statement about detectability, not existence, on a post hoc power analysis. Version 2 adds the

evidence of the simulation study reported here (its note N1): the test used had negligible power on the paper’s own premise.

This paper supplies the missing piece: the power curves. The question is practical. How much history does a test need to see a slow cycle if there is one? And what is the cycle worth, once seen, for forecasting crises? We answer it *in vitro*, where the truth is known. A simulated world contains a planted cycle of known period and coherence, and a null twin has the same marginal behaviour and no cycle. We preregistered the design and the decision rules, and ran the verification once; the protocol, frozen before any verification run, is deposited with this paper together with the register of its later changes (see Data and code).

The results fall into three groups.

- The requirement for detection, expressed in pooled decadal columns, and how noise, regularity, period, shocks, cadence and the choice of observed variables move it.
- The replacement of the persistence criterion of WP-2026-006 by a size-calibrated test, with the evidence that the former cannot work.
- The forecast value of a detected cycle. The detected phase beats a renewal model but not a hazard model on the observed state.

The paper is in vitro throughout. Its historical counterpart, a preregistered backcast of crisis probabilities on V-Dem, is the subject of WP-2026-008.

Related work. Three literatures meet here.

- *Cycles in instability series.* In a spectral analysis of political violence in the United States, 1780–2010, Turchin (2012) reports a secular cycle with a 50-year oscillation superimposed on it. Li et al. (2027) reduce a structural-demographic model to three equations and prove a Hopf bifurcation (§2.4).
- *Detecting a cycle against noise.* Climate science tests a spectral peak against a red-noise background (Mann and Lees 1996; Torrence and Compo 1998). Dynamic mode decomposition estimates the Koopman spectrum from data (Tu et al. 2014), but sensor noise biases its estimates, and corrections have been proposed (Dawson et al. 2016; Hemati et al. 2017); near the unit circle the sampling theory is that of local-to-unity autoregression (Phillips 1987). Our test combines pooled Hankel DMD with an optimal hard threshold on the singular values (Gavish and Donoho 2014), and it replaces the red-noise null with a null twin that has the same marginal behaviour (§2.2).
- *Forecasting instability.* Political instability has been forecast two years ahead with a global model (Goldstone et al. 2010), and armed conflict up to three years ahead (Hegre et al. 2019) and to 2050 (Hegre et al. 2013). The seismic analogue, a renewal model on the time since the last event, is the Brownian Passage Time of long-term earthquake forecasts (Matthews 2002; Ellsworth et al. 1999). We ask what a detected slow mode adds to such forecasts at horizons of decades to two centuries.

2 The simulated world

2.1 Societies with a planted cycle

Each simulated society is a Seldon-0 society of $N = 200$ agents of two types (40 and 160). Each agent is in one of two states, cooperating or not, and switches at rates that depend on the other agents’ states, on an institutional legitimacy variable L and on a resource variable R , which in turn respond to the agents. Its fast dynamics are demographic noise, with relaxation times of about 1–10 years. The model is stochastic, not chaotic: within a basin, trajectory predictability is lost over the fast mixing time.

The slow dynamics come from a common bias $s(t)$, the “climate” of the society, which we drive with a phase-diffusion oscillator:

$$s(t) = \bar{s} + A \cos \varphi(t), \quad d\varphi = \omega dt + \sqrt{2D} dW, \quad \omega = 2\pi/P, \quad D = 1/\tau_c.$$

For this oscillator $e^{i\varphi}$ is an exact Koopman eigenfunction of the joint Markov process, with eigenvalue $\lambda = e^{(i\omega - D)\Delta}$ per step Δ . The period P and the coherence time τ_c of the planted mode are therefore known exactly, which is what a power study needs.

The oscillator drives the society across two folds. Its cooperative state disappears at $s \approx 2.41$ and its collapsed state at $s \approx -2.62$, and between them the society is bistable. With a trough at $s_{lo} = -3.0$ and a crest s_{hi} just above the upper fold, the society falls into crisis near the crests and is forced to recover near the troughs. A **crisis** is the set $\{L < 0.2 \text{ and } S < 0.2\}$, where S is the share of cooperating agents, with a two-year dwell; an **onset** is an entry into it. In every world s_{hi} is set by bisection so that the long-run onset rate is 0.30 per century per society. Base rates are thus equal across worlds, and differences in power or skill reflect the cycle, not the frequency of crises. One model time unit is one year.

2.2 Worlds and the null twin

world	P (years)	τ_c (years)	role
B	200	600 ($\tau_c/P = 3$)	base
Reg1, Reg10	200	200, 2,000 ($\tau_c/P = 1, 10$)	regularity
P150, P300	150, 300	3P	period
Shock	200	600	extraordinary shocks at 0.002 per society-year
ME	250	750	model error: forecasters trained in B
R6	30	200	the premise of WP-2026-006 §6.7, inside one basin

Table T1. Calibration of the worlds (development seeds, 200 societies $\times$ 5,000 years, $dt = 0.025$ year). Onset rates are per century per society; the target is $0.30 \pm 3\%$.

world	s_{hi} (cycle)	onset rate	s_{hi} (null twin)	onset rate	time in crisis (cycle / null)
B	2.4875	0.299	2.6000	0.294	0.245 / 0.297
Reg1	2.5125	0.308	2.6000 (null of B)	0.294	0.261 / 0.297
Reg10	2.4781	0.299	2.6000 (null of B)	0.294	0.240 / 0.297
P150	2.4906	0.302	2.6125	0.296	0.179 / 0.278
P300	2.5125	0.298	2.6250	0.308	0.381 / 0.383
Shock	2.4875 (as B)	0.327 (not calibrated)	2.6000, with shocks	0.316	0.275 / 0.315
R6	2.3 (fixed)	0.0004	2.3 (fixed)	0.0000	0.010 / 0.005
ME	2.4875 (as B)	0.288	—	—	0.300 / —

Null-twin coherence times $\tau_{c,0} = P/(2\pi\sqrt{3})$: 13.8 years (P150), 18.4 (B), 27.6 (P300), 2.8 (R6).

A test of persistence needs a world that differs from B only in persistence. Each cycle world has a **null twin** with the same period and trough and the same arcsine marginal law of s . Its coherence time is $\tau_{c,0} = P/(2\pi\sqrt{3})$, the largest for which the spectrum of $\cos \varphi$ has no peak at a positive frequency: two Lorentzians of half-width D centred at $\pm\omega$ have a maximum away from

zero if and only if $\omega > D/\sqrt{3}$. The phase still rotates on average, but its spectrum has no peak; s_{hi} is recalibrated to the same crisis rate. Two other nulls were tried in the pilot and rejected, because they could not match both the crisis rate and the time spent in crisis: phase diffusion without rotation, and an Ornstein–Uhlenbeck driver with the same marginal.

2.3 What the historian sees

The observer sees three aggregate indicators, S, R and L, sampled every Δ_{obs} years ($\Delta_{\text{obs}} = 10$ is the primary cadence, $\Delta_{\text{obs}} = 1$ the secondary). Each indicator carries measurement noise $\sigma_{\text{obs}} \cdot \text{sd}_{\text{pool}} \cdot \varepsilon$, where sd_{pool} is the indicator’s pooled standard deviation and ε is standard normal, with $\sigma_{\text{obs}} \in \{0, 0.5, 1\}$; 0.5 is primary. Crisis onsets are observed exactly, with their dates, as historians record them; only the indicators are noisy. The secondary observation set **O4** adds a noisy proxy of the slow driver, $s + 0.5 \cdot \text{sd}(s) \cdot \varepsilon$, the analogue of a structural-demographic indicator. A panel is n independent societies observed for T years after a burn-in of $\max(400, 2P)$ years.

2.4 A mechanism from structural-demographic theory

The planted oscillator is a device, not a theory, but slow modes of this kind arise in structural-demographic models. Li et al. (2027) reduce the compartmental model behind Turchin’s Multi-Path Forecasting framework to three equations, for radicals, violence and social unhappiness. They prove a Hopf bifurcation: as the rate of economic decline crosses a threshold, a stable focus gives way to a limit cycle. The cycle needs the “cathartic” release of pressure by violence; without it, the reduced system cannot oscillate. The analysis is deterministic. With noise, the two sides of the threshold become the two forms of a slow Koopman mode, the third level of WP-2026-005. The CMH axioms already name this hierarchy in the formalisation of axiom A7 (WP-2026-001, p. 9), and their version 4, in preparation, restates A7 as predictability by level. Above the threshold, the limit cycle becomes a phase-diffusing cycle, the case of world B. Below it, noise excites the focus and produces quasi-cycles whose coherence time grows without bound as the threshold is approached. In a finite record the two are hard to tell apart. Testing our detector on that model is a natural robustness check; we have run it only in an exploratory form (§8).

3 The test

The quantity of interest is the modulus of the Koopman eigenvalue that belongs to the cycle. A noisy panel of societies gives an estimate of it through dynamic mode decomposition on delay coordinates. We use three estimators on the same data:

- E1, one Hankel-DMD per society, with the median of the band statistic;
- E2, canonical DMD on the pooled Hankel columns of all societies;
- E3, total-least-squares DMD on the same pooled columns.

The *band statistic* M is the largest modulus among the eigenvalues whose period lies in $[100, 400]$ years. The test rejects when M exceeds the 95% quantile of its distribution in the null twin, estimated from 400 null panels with the same n , T , cadence and noise. The test is therefore size-calibrated by construction, and its power is the rejection rate over 200 cycle panels.

The persistence criterion of WP-2026-006 (D1) declares a mode persistent when its modulus exceeds $1 - \Delta/T_{\text{usable}}$. It has no calibration. We run it next to M on the same panels. In world R6, which reproduces the premise of WP-2026-006 §6.7, D1 fires in at most 7% of panels at every $N_{\text{eff}} \leq 1,530$ (H1, refuted as preregistered). It fires in 0–2% of null panels. Its rate falls to zero as records lengthen, because the estimated modulus is biased below a threshold that rises with T .

In world B, D1 does not discriminate at all: at (32, 100) with $\sigma_{\text{obs}} = 0$ it fires in 56% of cycle panels and 52% of null panels, and its hits are never near $P = 200$.

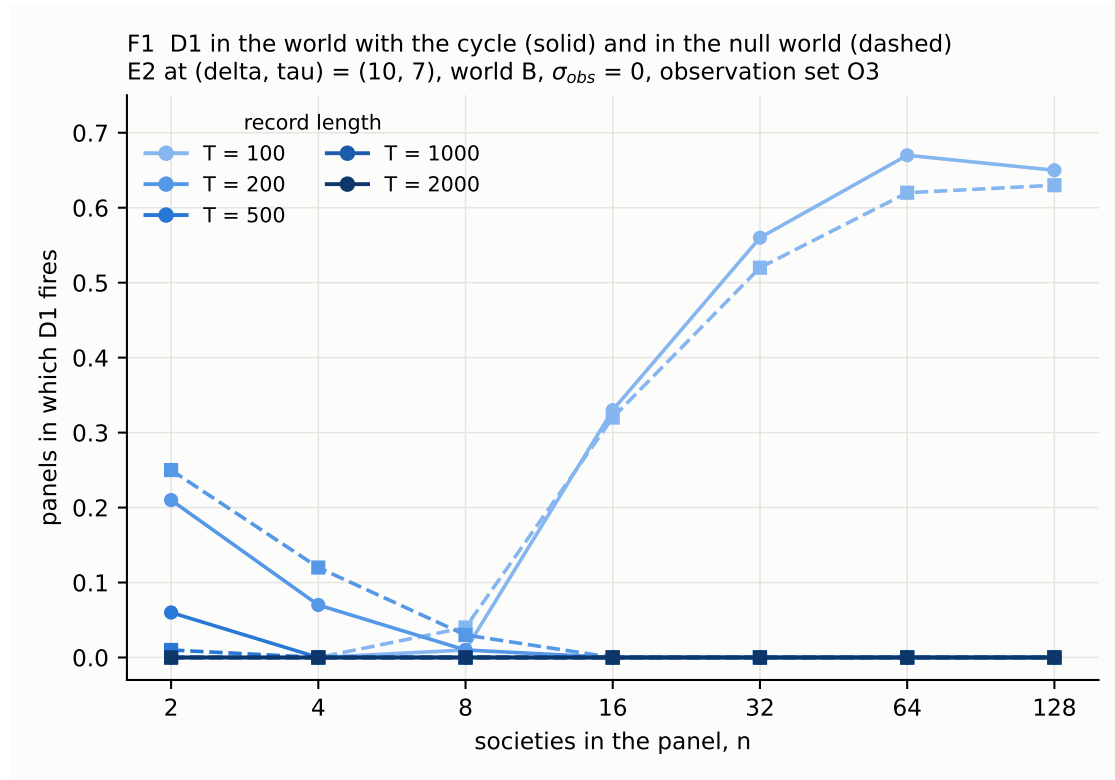


Figure F1. Share of panels in which the persistence criterion D1 fires, in world B (solid) and in its null twin (dashed), against the number of societies n , one curve per record length T . Estimator E2 at decadal cadence with a 70-year window, $\sigma_{\text{obs}} = 0$, observation set O3.

4 How much data

4.1 Power surfaces

Figure F2 shows the power of the three estimators in world B with moderate noise ($\sigma_{\text{obs}} = 0.5$), at the primary configuration: decadal cadence and a delay window of seven steps, 70 years. It is plotted against $N_{\text{eff}} = n(T/10 - 8)$, the number of pooled decadal Hankel columns. Pooling pays at every short record. With 16 societies observed for 200 years, the pooled estimator E2 detects the cycle in 79% of panels and the per-society estimator E1 in 46%. With 32 societies the two are 98% and 76%. TLS debiasing (E3) does not help at moderate noise: 47% at (16, 200). It matters at high noise, where E3 is the only estimator whose power grows monotonically with n (§5).

4.2 The requirement in one number

For $T \geq 200$ the 0.8 contour of E2 is close to a line of constant N_{eff} . E2 first reaches power 0.8 at $(n, T) = (32, 200)$, $(8, 500)$, $(4, 1,000)$ and $(2, 2,000)$, with $N_{\text{eff}} = 384$, 336, 368 and 384 (Figure F3). Table T2 gives the smallest n on the grid at which E2 reaches 0.8, for every world and noise level; Appendix Table A1 gives the same for all three estimators. Both tables are deposited as CSV files (Data and code). **A 200-year cycle of moderate regularity needs about 350–390 pooled decadal columns, however they are split between societies and years.** At $T = 100$, half a period, no panel up to 128 societies reaches 0.8 (0.77 at $n = 128$).

Table T2. Smallest number of societies n on the grid $(2, 4, \dots, 128)$ at which E2 reaches power 0.8; observation set O3, decadal cadence, 70-year window. “> 128”: not reached on the

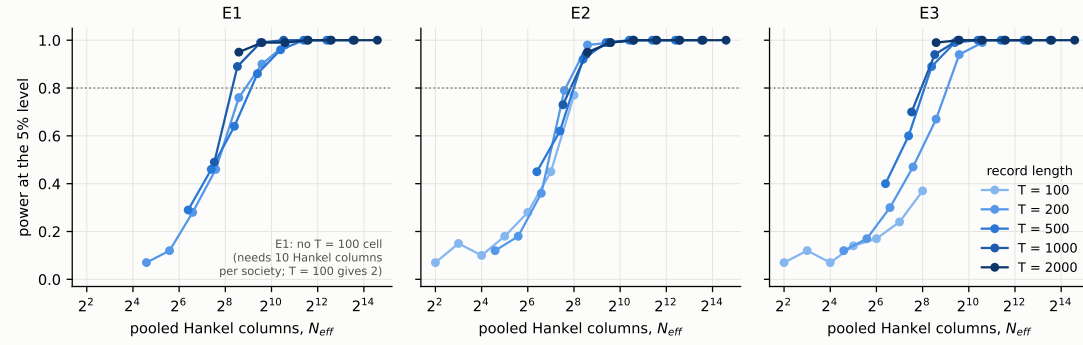
F2 power against N_{eff} , world B, $\sigma_{\text{obs}} = 0.5$, estimator at $(\delta, \tau) = (10, 7)$ 

Figure F2. Power of the three estimators E1, E2 and E3 in world B, $\sigma_{\text{obs}} = 0.5$, against N_{eff} , one curve per record length T ; decadal cadence, 70-year window, observation set O3. E1 has no curve at $T = 100$: by the preregistered rule, E1 is not computed when a society has fewer than 10 Hankel columns, and at $T = 100$ each society has 2 ($T/10 - 8$).

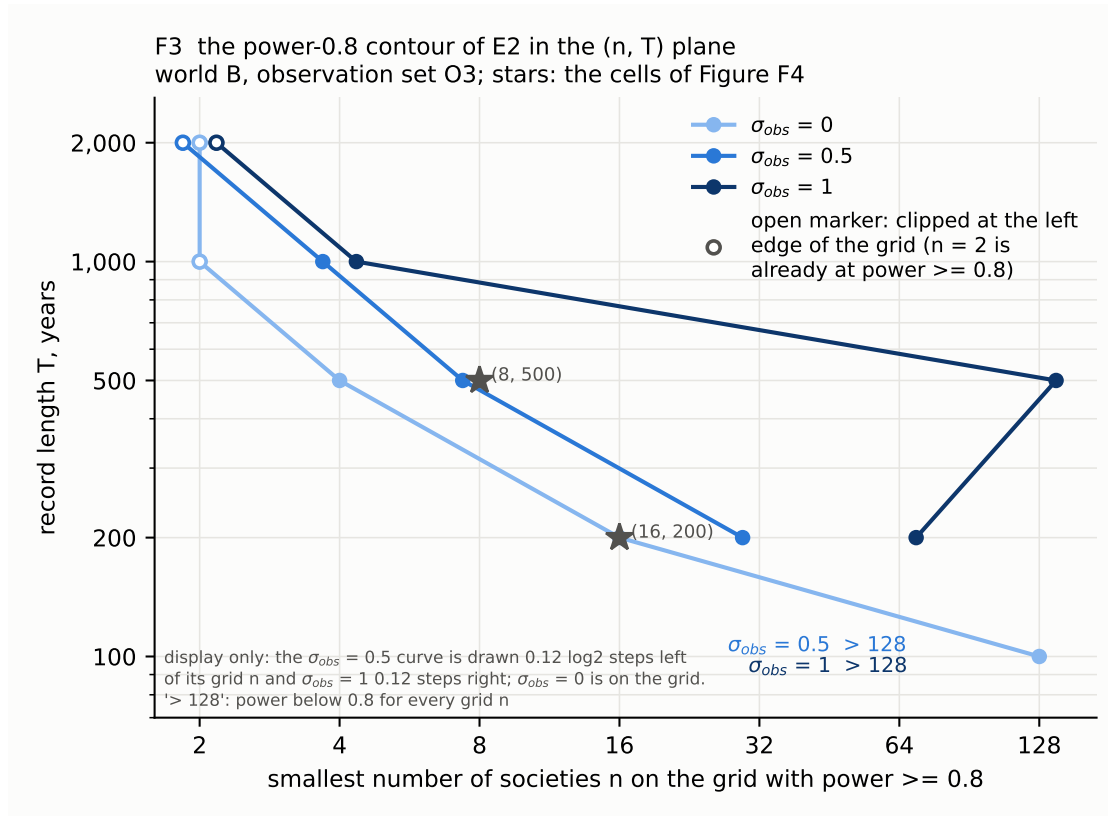


Figure F3. The power-0.8 contour of E2 in the (n, T) plane, world B, observation set O3, for $\sigma_{\text{obs}} = 0$, 0.5 and 1: for each T , the smallest n on the grid at which the power reaches 0.8, as in Table T2. Where no n on the grid reaches 0.8 the curve stops; a point at the left edge of the grid ($n = 2$) is drawn with an open marker. As the note in the figure says, the $\sigma_{\text{obs}} = 0.5$ curve is drawn 0.12 log2 steps left of its grid n and the $\sigma_{\text{obs}} = 1$ curve 0.12 steps right, for display only; $\sigma_{\text{obs}} = 0$ is on the grid. The stars mark the two cells of Figure F4.

grid. At $\sigma_{\text{obs}} = 1$ the power of E2 is not monotone in n (§5), so there the entry is the first n that reaches 0.8, not a threshold.

world	σ_{obs}	T = 100	T = 200	T = 500	T = 1,000	T = 2,000
B	0	128	16	4	2	2
B	0.5	> 128	32	8	4	2
B	1	> 128	64	128	4	2
Reg1	0	> 128	32	16	8	2
Reg1	0.5	> 128	64	64	32	4
Reg1	1	> 128	> 128	128	4	32
Reg10	0	128	16	4	2	2
Reg10	0.5	128	32	8	2	2
Reg10	1	> 128	64	8	2	2
P150	0.5	> 128	> 128	64	32	16
P300	0.5	> 128	32	8	4	2
Shock	0.5	128	32	8	4	2

This number is not comparable with the “ $N_{\text{eff}} \approx 1,000$ ” of WP-2026-006 §6.7. That figure counted annual observations, and with annual cadence and a 7-year window the same estimator has no power at all (§4.4).

4.3 How precision scales

Two preregistered hypotheses concern the scaling. The first is the rate of precision. In the cells with a reliable detection rate (no noise, $T \in \{1,000, 2,000\}$, correct-detection rate ≥ 0.9), the standard deviation of the estimated modulus falls as $N_{\text{eff}}^{-\gamma}$ with $\gamma = 0.474$ (95% interval [0.441, 0.506], 13 cells). This is the $N^{-1/2}$ rate of WP-2026-006’s own Proposition 4.1. It excludes the N^{-1} scaling of the calculation in its §6.7, which WP-2026-006 version 2 withdraws.

The second is whether N_{eff} is a sufficient statistic for power, that is, whether a year of record is worth as much as a society. The preregistered answer is no. In a logistic regression of detection on $\log N_{\text{eff}}$ and $\log T$ over all cells with $n \geq 4$, the coefficient of $\log T$ is $+0.337$ [0.154, 0.520]. At equal N_{eff} , longer records help. A post hoc decomposition, which we report as exploratory, locates this effect in the $T = 100$ cells. There the 70-year window leaves two columns per society and the record covers half a period. Restricted to $T \geq 200$ the coefficient is -0.312 [−0.604, −0.021]. Beyond one period of record, at equal N_{eff} , breadth is slightly better than length, and N_{eff} is roughly the currency of detection. (*This reading needs a preregistered replication on fresh seeds before it can be used as evidence.*)

4.4 Cadence and window

The window must match the period. At (16, 200), annual cadence with a 7-year window, the configuration of WP-2026-006, gives power 0.17. Annual cadence with a 70-year window gives 0.40, and decadal cadence with the same 70-year window gives 0.79. With annual cadence and a 7-year window, power never exceeds 0.40 anywhere up to 128 societies $\times$ 2,000 years. With the 70-year window at annual cadence it reaches 0.8 only at (64, 1,000) and (16, 2,000). The number of columns is not enough: they must span enough time to hold a phase of the cycle, and at decadal cadence each column does.

5 What moves the requirement

Table T3 and Figure F4 give the power of E2 at two cells near the contour, (16, 200) and (8, 500), as each factor is varied from world B.

factor	level	(16, 200)	(8, 500)
noise σ_{obs}	0 / 0.5 / 1.0	0.83 / 0.79 / 0.51	0.99 / 0.92 / 0.77
regularity τ_c/P	1 / 3 / 10	0.38 / 0.79 / 0.79	0.69 / 0.92 / 0.97
period P	150 / 200 / 300	0.20 / 0.79 / 0.53	0.23 / 0.92 / 0.81
shocks	none / 0.002 per year	0.79 / 0.70	0.92 / 0.93
observation set	O3 / O4	0.79 / 1.00	0.92 / 1.00

F4 power of E2 at the two cells of the contour as each factor of Table T3 is varied from world B, observation set O3, $\sigma_{\text{obs}} = 0.5$

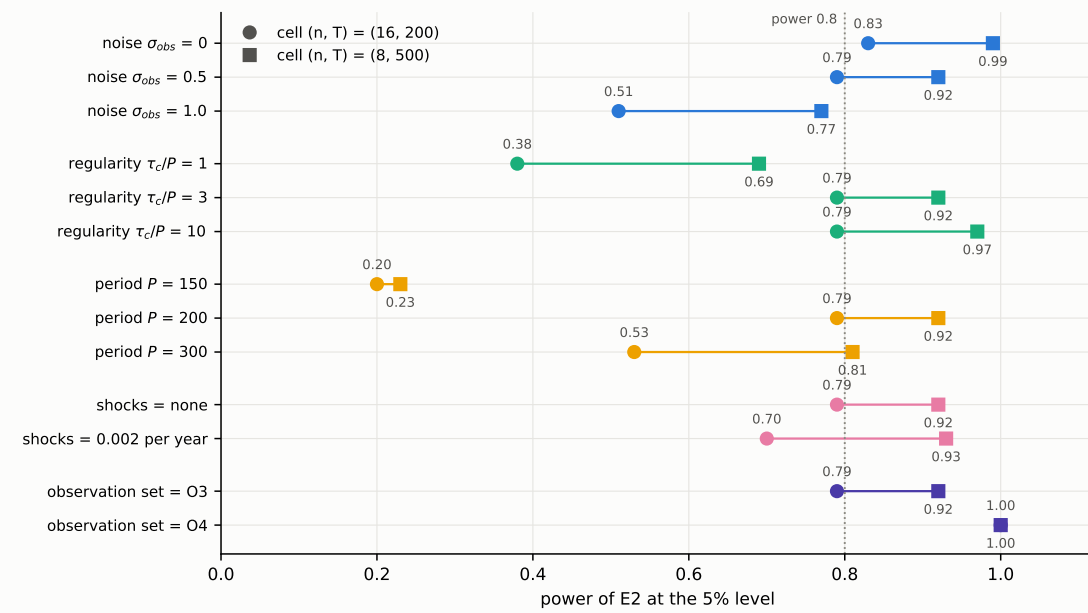


Figure F4. Power of E2 at $(n, T) = (16, 200)$ and $(8, 500)$ as each factor of Table T3 is varied from world B ($\sigma_{\text{obs}} = 0.5$, decadal cadence, 70-year window, observation set O3).

- **Noise** costs power mostly on short records: at $(16, 200)$ power falls from 0.83 without noise to 0.51 at $\sigma_{\text{obs}} = 1$, while at $(8, 500)$ it stays at 0.77. At $\sigma_{\text{obs}} = 1$ the power of E2 is not monotone in n on long records (in world B at $T = 2,000$: 0.84, 0.83, 0.78 and 0.62 for $n = 2$ to 16, then 1.00 at $n = 32$). The rank chosen by the Gavish–Donoho threshold and the retained modes change with the size of the panel. E3, the TLS estimator, is monotone there and reaches 1.00. Debiasing matters most when noise is high.
- **Regularity.** A cycle whose coherence time equals its period (Reg1) needs about four times the columns of world B: the smallest n at power 0.8 is 64 at $T = 200$ and 500, 32 at $T = 1,000$ and 4 at $T = 2,000$. More regularity than $\tau_c/P = 3$ helps little at $T = 200$ (0.79 in both B and Reg10).
- **Period.** A 150-year cycle is hard for E2 at every record length (0.20 and 0.23): the estimator places its modulus well below the truth (median 0.880 against 0.978), a projection bias of the finite delay window. E3 does better (0.85 at $(16, 500)$). A 300-year cycle needs records at least as long as the period (0.53 at $T = 200$, 0.81 at $T = 500$).
- **Extraordinary shocks** at a rate of one per 500 society-years cost little: $0.79 \rightarrow 0.70$ at $(16, 200)$, no change at $(8, 500)$.
- **What is observed matters more than how much.** Adding a noisy proxy of the slow driver (O4) lifts the strict historical panel from the edge to certainty: 1.00 at $(16, 200)$, with the modulus estimated without bias (0.984 against a true 0.9835 per decade). With

O4, eight societies over 200 years suffice (0.91), and two over 500 (0.93). The proxy here is the driver itself plus noise, so this is an upper bound. The direction is nonetheless clear: an indicator close to the slow variable, a structural-demographic proxy, is worth more than many more countries.

Two factors could not be assessed. Independent gaps in 30% of observations leave only 3–24 pooled columns per cell under the rule that builds Hankel columns inside fully observed segments. The printed powers are noise, which reflects a limit of that rule, not the clustered gaps of historical records. The attenuation of the modulus by measurement noise (hypothesis H4) meets its preregistered rule, but the rule measures the wrong quantity on this model. Without noise the attenuation is larger than with it (0.063–0.067 against 0.028–0.029), because the dominant bias is the projection of a nonlinear model on a finite Hankel dictionary. The noise-induced, sample-size-independent bias that WP-2026-006 version 2 describes in its item N3 rests on a separate linear benchmark.

6 Historical reach

Where do real panels sit on this map? The strictest historical panel we can defend is that of independent states in V-Dem v16 (Coppedge et al. 2026), observed at decadal cadence over continuous spans. Twenty-one states have at least 200 continuous years of independence. Before the power tables were computed, we fixed which cells of the grid such a panel can reach:

- the **primary rule** counts only continuous, independent records: 100 years with up to 32 states, and 200 years with up to 16;
- a **sensitivity rule** also counts periods of dependence, and gives 32 states over 200 years.

Table T4 sets these cells against their power.

Table T4. Reachable cells of the strict historical panel against the power of E2 (world B, $\sigma_{\text{obs}} = 0.5$, decadal cadence, 70-year window). Wilson 95% intervals where shown: post hoc arithmetic on $R = 200$ panels, exploratory; the preregistered rule decides on the count of detections, not on the interval.

n	T	N_{eff}	power	rule
2 / 4 / 8 / 16 / 32	100	4–64	0.07 / 0.15 / 0.10 / 0.18 / 0.28	primary
2 / 4 / 8	200	24–96	0.12 / 0.18 / 0.36	primary
16	200	192	0.79 [0.73, 0.84]	primary (maximum)
64 / 128	100	128 / 256	0.45 / 0.77	sensitivity
32	200	384	0.98 [0.95, 0.99]	sensitivity (maximum)

The best primary cell detects a 200-year cycle in 158 of 200 panels, two short of the 160 that the preregistered threshold of 0.8 requires. The rule was written to decide, and it decides: the strict panel does not reach the threshold. It is a documented limit at the edge, not far from it: the interval contains 0.8 (a post hoc Wilson interval, exploratory; the rule decides on the count), and the critical value carries its own sampling error from 400 null panels. The sensitivity rule would give 0.98, but it has no decisional value, because it treats colonial and occupied periods as if they carried the dynamics of an independent state. The historical panel therefore sits where the contour of §4.2 passes.

Can longer historical series move the panel inside the contour? A preregistered inventory of the long series that can actually be obtained answers no. The inventory covers the Maddison

Project Database (Bolt and van Zanden 2024), the Bank of England’s millennium dataset (Bank of England n.d.), Clio-Infra (n.d.), Allen’s European wages and prices (Allen 2001), the Global Price and Income History series for England and Istanbul (Global Price and Income History Group n.d.), and Schmelzing’s long series of interest rates and inflation (Schmelzing 2020). A unit counts if it has two quantitative state series of different kinds at decadal or finer resolution over the same span. Under the primary counting rules, 20 countries qualify over 200 years and 6 over 500 years; the variants of the rules give at most 23 and 7, and none qualifies over 1,000 years. Two of the rules (book titles and Allen’s real wages are excluded) were fixed after a first count was seen, and one reading was added last. No variant changes the verdict when no territory counts twice. Only a reading that counts a city and its country as two societies, such as London and England, would pass the thresholds, in three variants of four; the pooled test needs distinct societies. The best cell they reach outside the strict panel is (4, 500), with power 0.62. This is a limit of the data, not of the method. The inventory, and the historical backcast of crisis probabilities that the same study preregistered, will be reported in WP-2026-008 (in preparation). As §5 shows, what would change the verdict is not more countries but indicators close to the slow variable.

7 What a detected cycle is worth

Detection is a scientific question; forecasting is a practical one. Suppose the cycle has been detected: how much does it help to forecast a crisis 10 to 200 years ahead? We trained five forecasters on a panel of 64 societies observed for 1,000 years ($\sigma_{\text{obs}} = 0.5$), and scored them on 1,000 fresh societies per replicate, with ten replicates per world; for reasons of run time, F6 is scored on the first 200 of them. The target is at least one crisis onset in the next H years, and the score is the Brier skill against the base rate F1 (Gneiting and Raftery 2007). The forecasters are:

- **F2**, a seismic renewal model: a Brownian Passage Time on the time elapsed since the last onset (Matthews 2002; Ellsworth et al. 1999), the model behind long-term earthquake forecasts (Working Group on California Earthquake Probabilities 2003, 2008);
- **F3**, a hazard model on the observed state: a logistic regression on the three indicators and their 30-year changes;
- **F4**, the Koopman phase: a logistic regression on the phase of the band eigenvector estimated by E2;
- **F6**, the true model class, with parameters estimated by simulated minimum distance;
- **F5**, the oracle, which knows the model, the parameters and the hidden phase.

Table T5. Brier skill against the base rate, world B, training panel (64, 1,000), $\sigma_{\text{obs}} = 0.5$: mean, with 2.5–97.5% of the pooled bootstrap draws at 100 and 200 years. H^* is the largest horizon whose lower bound is positive. F6 is scored on the first 200 test societies of each replicate, the other forecasters on all 1,000. Figure F5 plots the same skill against H .

forecaster	$H = 10$	30	60	100	200	H^*
F2 renewal	−0.023	−0.049	−0.046	−0.041 [−0.254, +0.084]	−0.084 [−0.255, +0.034]	< 10
F3 observed state	+0.818	+0.284	+0.183	+0.173 [+0.130, +0.214]	+0.080 [+0.043, +0.116]	200
F4 Koopman phase	+0.218	+0.116	+0.108	+0.113 [+0.074, +0.160]	+0.070 [+0.035, +0.105]	200

forecaster	H = 10	30	60	100	200	H*
F6 true class, estimated (200 societies)	+0.787	+0.349	+0.297	+0.286 [+0.099, +0.421]	+0.104 [−0.012, +0.207]	100
F5 oracle	+0.967	+0.630	+0.521	+0.417 [+0.345, +0.480]	+0.121 [+0.070, +0.175]	200

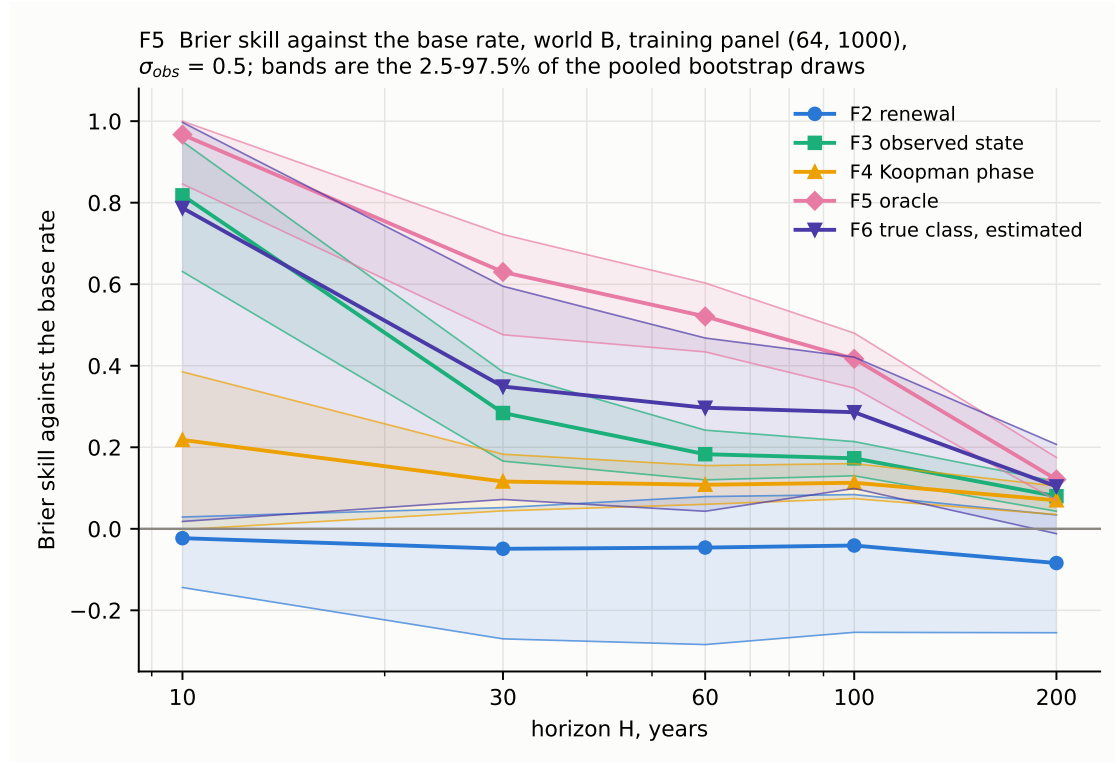


Figure F5. Brier skill against the base rate by horizon H for F2–F6, world B, training panel (64, 1,000), $\sigma_{\text{obs}} = 0.5$, as in Table T5. The bands are the 2.5–97.5% range of the pooled bootstrap draws.

Four findings follow.

The detected phase beats renewal. The preregistered comparison, F4 against F2 at 100 years, gives $+0.156$ [$+0.036$, $+0.348$]. F4 found a band eigenvalue in every replicate, with a median period of 227 years.

It does not beat the observed state. F4 is below F3 at every horizon up to 100 years and not distinguishable from it at 200; at 100 years the difference is -0.060 [-0.104 , -0.019]. The sign is the same in the seven other cells, worlds and training regimes we checked (-0.038 to -0.059 at 100 years). This comparison was not preregistered, and we report it as an exploratory reading that needs a replication on fresh seeds. Its meaning is nonetheless plain: once the observed state and its recent changes are known, the phase of the slow mode, as F4 extracts it, carries no further information about the next crisis.

Renewal fails in vitro. F2 never beats the base rate, and it is badly calibrated at long horizons: its forecasts between 0.56 and 0.74 come true 35–38% of the time. The reason is structural. Onsets are locked to the phase of the cycle, with a crisis in some cycles and none in others, so the intervals between onsets are neither independent nor unimodal. A renewal model on elapsed time alone is misspecified.

The horizon is set by the observed state and by the training record. F3 keeps positive skill up to 200 years, the edge of the grid. The correct model class has the higher mean skill at 30–100 years, but the difference is not significant with 200 test societies ($F6 - F3 = +0.096 [-0.032, +0.244]$ at 100 years, paired on the 200 test societies where F6 is scored), and it pays a large cost of estimation: on the same 200 societies the oracle is ahead of it by $+0.157 [+0.056, +0.265]$ at 100 years. Both are post hoc paired differences, exploratory. The bounds of these intervals, and of the $F4 - F3$ interval above, are Monte Carlo estimates from a bootstrap: rerunning the deposited script reproduces them within 0.002. The means of Table T5 give $+0.113$ and $+0.131$, because there F3 and F5 are scored on all 1,000 societies and F6 on 200. Under model error (world ME, with period 250 years in the test societies) F6 falls below F3 ($+0.176$ against $+0.215$). Training length caps the horizon: with 200 years of training record, no outcome at 200 years has been seen, and $H^* \leq 100$ for every fitted forecaster.

8 Limits and robustness

One model family. Every result comes from one kind of simulated society, Seldon-0 with a single planted oscillator. The planted mode is an exact Koopman eigenfunction; a real slow dynamic need not be one. It may be a slowly drifting equilibrium, a sequence of regimes, or a cycle whose period changes over time. The requirement of §4.2 is a statement about this family.

Independent societies. Our panels are independent draws. Historical polities are not: they share shocks, trade, wars and ideas. Cross-sectional dependence lowers the effective number of units, so for real panels the requirement of §4.2 is a lower bound.

The crisis is defined on the observed indicators. An onset is an entry into $\{L < 0.2, S < 0.2\}$, and L and S are two of the three indicators the forecasters see. This favours F3 at short horizons. At 60 years and beyond the advantage is smaller, but F3 still leads F4.

One use of the phase. F4 uses the phase alone. A hazard model on phase and state together was not part of the preregistered design and was not run; it could do better than either.

What could not be assessed. Independent gaps in 30% of the observations leave too few complete Hankel columns under our rule. The attenuation of the modulus by measurement noise is dominated, in this model, by the projection bias of a finite dictionary (§5).

A structural-demographic model. The planted oscillator is a device. A natural robustness check is the reduced structural-demographic model of Li et al. (2027), with noise, above and below its Hopf threshold. With the published parameters, its period near the threshold is about 4–5.5 model time units; the model has no calendar scale, so the check requires a choice of time scale. We ran this test only in an exploratory form. We used development seeds and did not follow the preregistered protocol. So it is not evidence, and we report no numbers from it. Its null is also different from the one in §3. Here each series is compared with its own AR(1) surrogate, not with a null twin (§2.2). Against this null, a detector of the same type rejected on both sides of the threshold. So a rejection did not point to a limit cycle, or to any specific mechanism. The line between quasi-cycles and a limit cycle stays open. The script and its exploratory output are in the deposit of this paper.

9 Consequences for history

Three consequences follow, each conditional on the model family of §8.

Detection is an observation problem more than a sample problem. The strict panel of independent states sits at the edge of detection for a 200-year cycle (0.79). What moves it decisively is not more units but a better observed variable: a noisy proxy of the slow driver lifts the same panel to certainty (§5). For historical research, the data worth collecting are therefore indicators close to the slow variable, such as the structural-demographic indicators of elite overproduction, popular immiseration and state fiscal distress, rather than more countries

with the same political indices. The long series that can be obtained today do not change the verdict (§6).

The persistence criterion of WP-2026-006 must be replaced by a size-calibrated test. The criterion has at most 7% power on its own premise, and it does not discriminate a cycle from its null twin (§3). A historical test of slow cycles should be preregistered only in a cell where a calibrated test has power of at least 0.8, against a null that matches the marginal behaviour of the data.

A detected cycle is not the route to long-horizon forecasts. In vitro, long-horizon skill comes from the observed state, from the right model class, and from a training record long enough to have seen outcomes at the horizon; it does not come from the phase of a detected mode (§7). A historical forecasting layer should therefore be a hazard model on the observed state, with elapsed time as a candidate addition, rather than a pure renewal model. Its preregistered backcast on V-Dem, and what it finds, are the subject of WP-2026-008 (in preparation).

Data and code

The following are deposited with this working paper on Zenodo ([doi:10.5281/zenodo.22952295](https://doi.org/10.5281/zenodo.22952295)): the simulation and analysis code, with the fixed seeds of the preregistered runs (development seeds and the verification seeds, used once at a tagged version of the code); the CSV files of Tables T2 and A1, with the script that computes them from the verified power table; and the preregistered protocol. The protocol is deposited in its frozen version 1.0 (commit 92f47d3, 24 September 2026, 12:27 CEST), with the register of the changes made after the freeze and before the first verification run: version 1.1 (commit 46ac731, 13:37) filled the calibration appendix and logged four deviations, and version 1.2 (commit 5ba6593, 15:11) fixed the bound on what historical data can reach (used in §6 of this paper) and logged two more. The body of the protocol (hypotheses, decision rules, estimators, forecasters, scores) did not change after the freeze. The code was tagged at the commit of version 1.2, and the verification runs followed, once, between 16:09 and 20:34 the same day. Each version is deposited with its SHA-256 checksum and the diffs between versions. The script and output of the exploratory check (§8) are in a separate folder of the deposit, marked as exploratory. The same folder holds the output of the linear benchmark that WP-2026-006 version 2 cites in its note N3. Its script is in the code. The paired differences of §7 are recomputed by a script in a folder marked as post hoc; it was reconstructed after the verification, because the original was not kept, and it gives the point estimates exactly and the bounds within 0.002.

Use of AI tools

I used AI tools in this work. Claude models (Anthropic) helped me write and check the code, run the simulations, and draft and revise the text. DeepSeek models did smaller tasks, such as literature searches and first checks. GPT (OpenAI) took part in one review discussion. All tasks and checks are logged in the laboratory records. I checked the results and made all research choices. I am responsible for the content.

Appendix

Table A1. As Table T2, for the three estimators: E1 (one Hankel-DMD per society, median statistic), E2 (pooled canonical DMD), E3 (pooled total-least-squares DMD). “—”: cell not computed (E1 is not computed when a society has fewer than 10 Hankel columns; at $T = 100$ each society has 2).

world	σ_{obs}	estimator	T = 100	T = 200	T = 500	T = 1,000	T = 2,000
B	0	E1	—	32	8	4	2
B	0	E2	128	16	4	2	2
B	0	E3	128	16	8	4	2
B	0.5	E1	—	64	16	4	2
B	0.5	E2	> 128	32	8	4	2
B	0.5	E3	> 128	64	8	4	2
B	1	E1	—	64	16	4	2
B	1	E2	> 128	64	128	4	2
B	1	E3	> 128	64	8	4	2
Reg1	0	E1	—	64	16	8	2
Reg1	0	E2	> 128	32	16	8	2
Reg1	0	E3	> 128	64	16	4	2
Reg1	0.5	E1	—	64	32	8	4
Reg1	0.5	E2	> 128	64	64	32	4
Reg1	0.5	E3	> 128	64	16	4	2
Reg1	1	E1	—	64	16	8	4
Reg1	1	E2	> 128	> 128	128	4	32
Reg1	1	E3	> 128	64	16	8	2
Reg10	0	E1	—	32	8	2	2
Reg10	0	E2	128	16	4	2	2
Reg10	0	E3	128	16	4	2	2
Reg10	0.5	E1	—	64	16	4	2
Reg10	0.5	E2	128	32	8	2	2
Reg10	0.5	E3	> 128	32	8	2	2
Reg10	1	E1	—	64	16	4	2
Reg10	1	E2	> 128	64	8	2	2
Reg10	1	E3	> 128	64	8	4	2
P150	0.5	E1	—	> 128	32	8	8
P150	0.5	E2	> 128	> 128	64	32	16
P150	0.5	E3	> 128	128	16	8	4
P300	0.5	E1	—	128	64	32	4
P300	0.5	E2	> 128	32	8	4	2
P300	0.5	E3	> 128	64	8	4	2
Shock	0.5	E1	—	64	16	4	2
Shock	0.5	E2	128	32	8	4	2
Shock	0.5	E3	> 128	64	8	2	2

References

- Allen, R. C. (2001). The great divergence in European wages and prices from the Middle Ages to the First World War. *Explorations in Economic History* 38(4), 411–447. doi:10.1006/exeh.2001.0775. Data used: the wage and price files on the author’s research pages (summary files of labourers’ and craftsmen’s wages, and the city files), <https://www.nuffield.ox.ac.uk/people/sites/allen-research-pages/>, downloaded 25 September 2026.
- Angeli, S. (2026a). Computational Macrohistory: Axiomatic Foundations for a Quantitative Science of Large-Scale Social Systems. FICSS Working Paper WP-2026-001. doi:10.5281/zenodo.18288164.
- Angeli, S. (2026b). Beyond the Lyapunov Wall: Extending the Predictive Horizon of Computational Macrohistory. A Methodological Roadmap. FICSS Working Paper WP-2026-005,

- version 2. doi:10.5281/zenodo.20111271.
- Angeli, S. (2026c). Koopman–CMH: A Formal Treatment of Spectral Structure Persistence in Historical Dynamics. FICSS Working Paper WP-2026-006, version 2. doi:10.5281/zenodo.20678999.
- Angeli, S. (in preparation). The historical counterpart of WP-2026-007: a preregistered backcast of crisis probabilities on V-Dem. FICSS Working Paper WP-2026-008.
- Bank of England (n.d.). A millennium of macroeconomic data for the UK, version 3.1 (updated to 2016), <https://www.bankofengland.co.uk/statistics/research-datasets>, downloaded 25 September 2026.
- Bolt, J., and van Zanden, J. L. (2024). Maddison-style estimates of the evolution of the world economy: A new 2023 update. *Journal of Economic Surveys* 39(2), 631–671 (published online 3 April 2024; print issue April 2025). doi:10.1111/joes.12618. Data used: Maddison Project Database 2023, version 1.0, DataverseNL, doi:10.34894/INZBF2, downloaded 25 September 2026.
- Clio-Infra (n.d.). Clio-Infra database, compact data files of 15 indicators, <https://clio-infra.eu/data/>, downloaded 25 September 2026. Documentation on DataverseNL; for GDP per capita, Bolt, J., and van Zanden, J. L., The long view on economic growth: New estimates of GDP, doi:10.34894/I7YZIV.
- Coppedge, M., et al. (2026). V-Dem Dataset v16. Varieties of Democracy (V-Dem) Project. doi:10.23696/vdemds26.
- Dawson, S. T. M., Hemati, M. S., Williams, M. O., and Rowley, C. W. (2016). Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids* 57(3), 42. doi:10.1007/s00348-016-2127-7.
- Ellsworth, W. L., Matthews, M. V., Nadeau, R. M., Nishenko, S. P., Reasenber, P. A., and Simpson, R. W. (1999). A physically-based earthquake recurrence model for estimation of long-term earthquake probabilities. US Geological Survey Open-File Report 99-522. doi:10.3133/ofr99522.
- Gavish, M., and Donoho, D. L. (2014). The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory* 60(8), 5040–5053. doi:10.1109/TIT.2014.2323359.
- Global Price and Income History Group (n.d.). Data files England 1209–1914 (Clark; data as of 10 April 2006) and Istanbul 1469–1914 (Pamuk and Arroyo Abad), <https://gpih.ucdavis.edu/Datafilelist.htm>, downloaded 25 September 2026.
- Gneiting, T., and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102(477), 359–378. doi:10.1198/016214506000001437.
- Goldstone, J. A., Bates, R. H., Epstein, D. L., Gurr, T. R., Lustik, M. B., Marshall, M. G., Ulfelder, J., and Woodward, M. (2010). A global model for forecasting political instability. *American Journal of Political Science* 54(1), 190–208. doi:10.1111/j.1540-5907.2009.00426.x.
- Hegre, H., Allansson, M., Basedau, M., et al. (2019). ViEWS: A political violence early-warning system. *Journal of Peace Research* 56(2), 155–174. doi:10.1177/0022343319823860.
- Hegre, H., Karlsen, J., Nygård, H. M., Strand, H., and Urdal, H. (2013). Predicting armed conflict, 2010–2050. *International Studies Quarterly* 57(2), 250–270. doi:10.1111/isqu.12007.
- Hemati, M. S., Rowley, C. W., Deem, E. A., and Cattafesta, L. N. (2017). De-biasing the dynamic mode decomposition for applied Koopman spectral analysis of noisy datasets. *Theoretical and Computational Fluid Dynamics* 31(4), 349–368. doi:10.1007/s00162-017-0432-2.
- Li, M., Li, Z., Ma, J., Fernandes, R., and Wang, M. (2027). Nonlinear dynamics of political instability in a reduced structural-demographic model. *Journal of Applied Analysis and Computation* 17(2), 681–705. doi:10.11948/20260182.
- Mann, M. E., and Lees, J. M. (1996). Robust estimation of background noise and signal detection in climatic time series. *Climatic Change* 33(3), 409–445. doi:10.1007/BF00142586.

- Matthews, M. V. (2002). A Brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America* 92(6), 2233–2250. doi:10.1785/0120010267.
- Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* 74(3), 535–547. doi:10.1093/biomet/74.3.535.
- Schmelzing, P. (2020). Eight centuries of global real interest rates, R-G, and the ‘suprasecular’ decline, 1311–2018. Bank of England Staff Working Paper No. 845. Data used: the data file of the paper (“This version: 1.2 January 2020”), from the Bank of England page of the working paper, downloaded 25 September 2026.
- Torrence, C., and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* 79(1), 61–78. doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2.
- Tu, J. H., Rowley, C. W., Luchtenburg, D. M., Brunton, S. L., and Kutz, J. N. (2014). On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics* 1(2), 391–421. doi:10.3934/jcd.2014.1.391.
- Turchin, P. (2012). Dynamics of political instability in the United States, 1780–2010. *Journal of Peace Research* 49(4), 577–591. doi:10.1177/0022343312442078.
- Working Group on California Earthquake Probabilities (2003). Earthquake probabilities in the San Francisco Bay region: 2002–2031. US Geological Survey Open-File Report 03-214. doi:10.3133/ofr03214.
- Working Group on California Earthquake Probabilities (2008). The Uniform California Earthquake Rupture Forecast, version 2 (UCERF 2). US Geological Survey Open-File Report 2007-1437. doi:10.3133/ofr20071437.
- Angeli (2026a–c) are lettered within this paper; in the text the series is cited by working-paper number.*

Document VII of the Computational Macrohistory Series
Foundations Institute of Computational Social Science
<https://ficss.institute>

License: CC-BY 4.0 International
Contact: stefano.angeli@ficss.institute

This document may be cited, shared, and adapted with attribution under the Creative Commons Attribution 4.0 International License.