FICSS WORKING PAPER | WP-2026-006

Koopman–CMH:

A Formal Treatment of Spectral Structure Persistence in Historical Dynamics

Galen Fontaise

Abstract

This paper proves the Koopman–A2 Correspondence (Theorem 1) — the formal result announced as Conjecture 1 of WP-2026-005 — and conducts the first pre-registered empirical test of its substantive antecedent. Theorem 1 shows that shared ergodic-class membership implies shared leading Koopman spectrum, conditional on Axioms A2, A3, A4, and a forcing-homogeneity assumption that refines the antecedent stated in WP-2026-005.

The empirical test applies a canonical Hankel-DMD estimator to a panel of eight Western European polities across two structurally stationary sub-windows (1820–1913, 1945–2020). The pre-registered binary criterion rejects Question 1 in both sub-windows: no eigenvalue beyond the invariant-measure cluster meets the persistence threshold. A post-hoc power analysis finds effectively zero power at the available effective sample size — the post-war effective sample size (N_{eff} = 210) falls below the operator's parameter count — so the verdict constrains detectability, not existence. A credible test requires $N_{eff} \approx$ 1,000.

Two methodological results give the paper positive content. The invariant-measure mode is recovered with the directional attenuation Proposition 4.2 predicts, on the one mode whose true value the axiomatic structure fixes a priori. A Juglar-band correspondence recurs across the inter-war break but is not statistically distinguishable from noise (37% match rate in surrogate null panels). The paper recommends canonical Tu et al. (2014) DMD with Gavish-Donoho truncation as the default for macrohistorical panel data.

Keywords: *Koopman operator; dynamic mode decomposition; Computational Macrohistory; spectral persistence; pre-registered empirical test; falsifiability*

JEL Classification: C02, C63, N00, D74

Suggested citation: Fontaise, G. (2026). Koopman–CMH: A Formal Treatment of Spectral Structure Persistence in Historical Dynamics. FICSS Working Paper WP-2026-006. ORCID: 0009-0007-6643-2307.

Introduction

1.1 The empirical question

WP-2026-005 formulated the empirical question of macrohistorical spectral regularity in operator-theoretic terms and developed the Koopman framework within the Computational Macrohistory programme of WP-2026-001 through WP-2026-004; readers new to the programme are directed there for the motivating context and to WP-2026-001 for the axiomatic foundations. The framework's central object is the stochastic Koopman operator $\mathcal{K}\_\theta$ associated with the macrohistorical dynamics in a structural regime θ, and its central empirical prediction — Conjecture 1 of WP-2026-005 — is that the spectrum of $\mathcal{K}\_\theta$ contains modes whose decay times exceed the time-scales the historical record offers, modes the framework calls Level III. If Conjecture 1 holds, the long-run dynamical structure of macrohistorical systems is encoded in finitely many slowly-decaying spectral modes that survive structural transitions, are at least partly shared across structurally comparable polities, and are in principle estimable from panel data. Conjecture 1 is a positive empirical prediction and admits a falsificationist test; the present paper proves the conditional theorem that underlies it and conducts the first pre-registered empirical test.

1.2 The test conducted here

Pre-registered in §5 is a multi-task Hankel-DMD analysis on a Western European panel of eight countries, with state vector of five variables spanning the four substantive CMH dimensions D, E, P, S — population, GDP per capita, the V-Dem Liberal Democracy Index, the V-Dem Electoral Democracy Index, and the COW Composite Index of National Capability. The panel divides into two structurally stationary sub-windows: 1820–1913 covers the long nineteenth century between the post-Napoleonic settlement and the First World War, and 1945–2020 covers the post-war European order. The two sub-windows together provide approximately 1,250 listwise-complete country-years across 170 panel-years of observation, which is the longest joint coverage the available data sources support for a state vector of this composition.

The empirical question is posed as a pre-registered falsification test, and we are explicit about its register. We do not ask whether persistent spectral structure *exists* in macrohistorical dynamics; we ask whether, on a pre-specified eight-country panel and under a pre-specified estimator and threshold, such structure is *detectable*. The distinction is not rhetorical: as the power analysis of §6.4 and §6.7 shows, the test's ability to reject the null is bounded by the effective sample size, and a rejection of Q1 is therefore in the first instance a statement about detectability on this design. We carry this qualification from here to the conclusion rather than introducing it post hoc.

Three questions are pre-registered. Question 1 asks whether the multi-task estimator returns eigenvalues at or near the unit circle after debiasing, against a stringent persistence threshold derived from the longest period the sub-window can resolve. Question 2 asks whether persistent modes, if identified by Q1, are shared across countries in the panel rather than country-specific. Question 3 asks whether the periods of identified persistent modes correspond to documented bands in the macrohistorical and economic-history literatures. Q1 is binary, with a fixed threshold; Q2 and Q3 are conditional on Q1 and inherit its binary character through their own pre-registered criteria.

Specification in §5 preceded the assembly of any panel data, execution in §6 proceeded without modification of its criteria, and reporting in this paper followed regardless of the direction in which the results fell. The pre-registration commitment is the operational content of Axiom A8 of the framework as specified in WP-2026-001 §3 — the falsifiability axiom — and the discharge of that commitment is part of what the present paper offers as a contribution to the empirical macrohistorical literature. The paper is the first in the CMH programme to discharge an A8 obligation on a substantive empirical test rather than on the framework's foundational construction or on a foundational case study.

In both sub-windows the test rejected its pre-registered Q1. Some pre-registered tests reject and some do not, and the pre-registration's contract is to report the result regardless. The body of the paper takes up what the rejection does and does not entail, what the test produced as ancillary findings that the binary verdict does not exhaust, and what design refinements the test's own results suggest for the next iteration of the empirical programme.

1.3 What the paper contributes

Beyond the discharge of the A8 obligation on Conjecture 1, the paper offers four contributions of varying scope.

The empirical findings comprise three distinct sub-results. The invariant-measure mode of the multi-task estimator is recovered with the finite-sample attenuation that Proposition 4.2 of §4 predicts, on a benchmark whose true value the framework's axiomatic structure fixes a priori — the first empirical observation in the programme that one of the framework's quantitative claims is borne out by data. The sub-threshold spectral structure in the five-to-nine year band recurs across the inter-war structural break with matched periods and matched moderate moduli, an observation that the paper records and offers two contrasting readings of without adjudicating. The robustness diagnostics document a substantive violation of the conditional-independence hypothesis of Proposition 4.1 in the post-war sub-window and an at-the-edge violation of the local-normality hypothesis of Theorem 1 in both sub-windows, with operational consequences for the effective sample size and the perturbation bounds.

A second contribution is formal. The proof of Theorem 1 in §3 and the explicit Assumption 3.1 introduced there refine the antecedent of Conjecture 1 as stated in WP-2026-005 §4.5. The original statement attributed the implication "shared structural regime ⇒ shared Koopman spectrum" to Axioms A2 and A4. Lemma 3.2a shows that the step from shared structural regime to shared transition kernel — without which the Koopman operators cannot be identified — requires A3 read in closure form and a separate forcing-homogeneity assumption (Assumption 3.1). The Koopman–A2 Correspondence therefore rests on A2, A3, A4, and Assumption 3.1 jointly, and the paper renders the full hypothesis set inspectable. Assumption 3.1 in particular becomes a target of empirical diagnosis in its own right, addressed by the cross-country residual correlation diagnostic of §6.4.

There is a methodological yield as well. The development of the empirical pipeline revealed three things about estimator behaviour on macrohistorical panel data. The canonical DMD construction of Tu et al. (2014) with Gavish-Donoho rank truncation is markedly better-behaved than the unprojected ridge variants that the §5 pre-registration's wording permitted. The TLS-DMD variant of Hemati et al. (2017) is over-conservative on the present panel, for reasons traceable to assumption mismatches between its derivation and the panel's noise structure. The forward-backward variant of Dawson et al. (2016) exhibits an overcorrection problem that places its eigenvalues outside the unit circle, in formal contradiction with the contraction property of the underlying operator. The paper recommends the canonical pipeline with Gavish-Donoho truncation as the default for similar applications and identifies the conditions under which the alternatives can and cannot be trusted to provide informative debiasing.

Last in scope is a procedural one. The full trail of estimator implementations that preceded the canonical pipeline, including the bug-level diagnoses that motivated each revision, is documented in the project's GitHub repository (Appendix B), with the salient quantities summarised in §6.1 and §8. The aim is to make the trajectory inspectable rather than implicit, in keeping with the A8 commitment that subsequent users of the methodology should be able to identify both what was done and what was reconsidered along the way. The applied dynamical-systems literature has a paucity of such trajectories in published form, and the present paper offers one.

The paper presupposes WP-2026-005, to which it is the announced companion; the introductory apparatus is not duplicated here. The structure is as follows: §2 sets up the formal apparatus, §3 proves Theorem 1, §4 specifies the estimator, §5 pre-registers the test, §6 reports the results, §7 and §8 discuss them, and §9 concludes; Appendix A collects the proofs, and Appendix B points to the replication materials on GitHub.

The Stochastic Koopman Operator on the CMH State Space

2.1 Why the operator must be stochastic

WP-2026-005 §4.1 presented the Koopman operator for deterministic systems. For CMH the deterministic construction is unavailable, and the reason is axiomatic rather than technical: Axiom A5 asserts an irreducible stochastic component in the macrohistorical dynamics, and building the predictive architecture on a deterministic operator would place it in direct contradiction with A5. The operator the present paper requires is the stochastic Koopman operator.

The starting point is the dynamical form given by Axiom A4. A4 specifies that the macrohistorical state evolves according to

$$(dX)/(dt) = F(X, \theta, \varepsilon(t)),$$

with F locally Lipschitz in its first argument and $\varepsilon(t)$ a stochastic process. For the purposes of the operator-theoretic treatment it is convenient to work with the discrete-time transition induced by this dynamics over a fixed sampling interval — one year, the natural observational unit of the CMH datasets — rather than with the continuous-time generator directly. Integrating the A4 dynamics over one unit interval yields a stochastic transition: the state X_{t+1} is a random variable whose distribution is determined by X_t, by the structural parameters θ, and by the law of the noise process. This transition is described completely by a Markov transition kernel

$$P\_\theta(x, A) = Pr(X_{t+1} \in A \mid X_t = x),$$

the probability that the system, currently at state x, lands in the measurable set A one step later. The subscript θ records that the kernel depends on the structural regime; this dependence will carry the entire weight of the cross-country argument in §3, and it is worth fixing notation for it now. The reduction from the continuous A4 dynamics to a discrete-time Markov kernel is standard and requires only that the noise process have the weak-dependence properties already assumed elsewhere in the CMH axiomatic system; the one-year kernel P_θ inherits its regularity from the Lipschitz property of F.

Throughout this paper, the Itô interpretation of all stochastic integrals is adopted, and the discrete-time kernel P_θ is the Itô–Euler one-step discretisation of the continuous-time dynamics. The empirical content of the paper rests entirely on the discrete-time kernel itself, and the Itô-vs-Stratonovich question matters only through the discretisation: under the Itô convention the kernel is well defined and the operator constructed from it is a genuine contraction. The alternative Stratonovich reading would change the drift term of the continuous-time generator by the familiar correction and, with it, the fine structure of the discretised kernel; the qualitative results of this paper are insensitive to the choice, but the quantitative mode estimates of §6 are not, and the convention must therefore be stated and held fixed.

2.2 The observable space and the operator

With the transition kernel P_θ in hand, the stochastic Koopman operator is defined as the conditional-expectation operator associated with it. For an observable function g — a real-valued function of the macrohistorical state, an element of the function space specified below — the operator $\mathcal{K}\_\theta$ acts by

$$(\mathcal{K}\_\theta\, g)(x) = \mathbb{E}[\, g(X_{t+1}) \mid X_t = x \,] = \int g(y)\, P\_\theta(x, dy).$$

The value of $\mathcal{K}\_\theta\, g$ at a state x is the expected value of the observable one step into the future, given that the system is at x now. Two features of this definition deserve emphasis because they are exactly the features that make the framework compatible with the CMH axioms rather than in tension with them.

The first is that $\mathcal{K}\_\theta$ is linear in g, and linear without qualification. For any observables g and h and any scalars α and β, $\mathcal{K}\_\theta(\alpha g + \beta h) = \alpha \mathcal{K}\_\theta g + \beta \mathcal{K}\_\theta h$, because conditional expectation is linear. This linearity holds regardless of how nonlinear the underlying dynamics F may be and regardless of the presence of the stochastic forcing. It is worth being explicit about what has and has not been achieved here, because the point is easy to overstate. The nonlinearity of the dynamics has not been removed; it has been relocated. It now resides in the structure of the operator $\mathcal{K}\_\theta$ — in its spectrum and its eigenfunctions — rather than in the equation of motion. What linearity buys is access to spectral analysis: an operator, unlike a nonlinear map, can be decomposed into eigenvalues and eigenfunctions, and that decomposition is what §2.3 and §2.4 exploit.

The second feature is that the stochasticity has been absorbed into the operator without being suppressed. The deterministic Koopman operator acts by composition, g *mapsto* $g \circ \Phi$; the stochastic operator acts by conditional expectation, g *mapsto* $\mathbb{E}[g(X_{t+1}) \mid X_t = \cdot]$. When the noise vanishes the second reduces to the first, so the construction is a genuine generalisation. But when the noise does not vanish — the case A5 insists upon — the operator describes the evolution of observables in expectation, averaging over the irreducible stochastic component. This is the precise sense in which the framework respects A5: it does not claim that the future state is determined, only that the *expected value* of an observable evolves linearly and deterministically. The distinction between a determinate trajectory and a determinate law for the expectation of observables is the conceptual hinge of the whole paper, and §2.4 returns to it.

The operator must act on a definite function space, and the choice of that space is where Axiom A2 enters. A2 (conditional historical ergodicity) asserts that, for a fixed structural regime θ, the macrohistorical system possesses a well-defined invariant statistical description: there exists a probability distribution to which time averages of observables converge. In the language of the present construction, A2 guarantees the existence of an invariant probability measure μ_θ for the kernel P_θ — a measure satisfying $\int P\_\theta(x, A)\, \mu\_\theta(dx) = \mu\_\theta(A)$ for every measurable set A, so that a system distributed according to μ_θ remains so distributed under the dynamics. A2 is the axiom that licenses speaking of this measure at all. The natural function space on which to study $\mathcal{K}\_\theta$ is then $L^2(\mu\_\theta)$, the space of observables that are square-integrable against the invariant measure. The choice is not a matter of analytical convenience. Without A2 there is no μ_θ, without μ_θ there is no $L^2(\mu\_\theta)$, and without that space the spectral decomposition on which the entire Level III architecture rests has no domain on which to live. The dependence runs in one direction and it is worth stating plainly: A2 is the existence condition for the object that this paper analyses.

On $L^2(\mu\_\theta)$ the operator $\mathcal{K}\_\theta$ has a property that will be used repeatedly in §4: it is a contraction, meaning that it never increases the $L^2(\mu\_\theta)$ norm of an observable, $\| \mathcal{K}\_\theta g \| \leq \| g \|$ for every g. The proof is a one-line consequence of the conditional Jensen inequality together with the invariance of μ_θ, and it is given in Appendix A. The contraction property is what guarantees that the spectrum of $\mathcal{K}\_\theta$ is confined to the closed unit disc of the complex plane — a fact that gives the eigenvalue classification of §2.3 its meaning, since it tells us that there are no eigenvalues outside the disc to worry about and that the boundary of the disc is the meaningful dividing line.

2.3 Spectral decomposition

Because $\mathcal{K}\_\theta$ is a bounded linear operator on the Hilbert space $L^2(\mu\_\theta)$, the machinery of spectral theory applies to it. The structure of its spectrum, however, depends on a mixing assumption, and this is the point at which the "weak mixing and finite-variance noise" conditions named in WP-2026-005's statement of Conjecture 1 acquire their formal role. Mixing is, in fact, the dynamical content that A2 only partially supplies: A2 in its CMH formulation is an *operational* ergodicity assumption, a statement that time averages converge, and this is weaker than mathematically rigorous ergodicity. For the spectral decomposition to take the form the paper needs, the system must satisfy weak mixing — a strengthening of ergodicity which requires that correlations between the system's state at well-separated times decay, not necessarily monotonically, but in the averaged sense characteristic of weak mixing. The finite-variance condition on the noise ensures that the observables of interest, and their images under the operator, remain within $L^2(\mu\_\theta)$. These two conditions are assumptions, not theorems; they are stated here as part of the hypothesis set that Theorem 1 will inherit, and their empirical plausibility for historical systems is discussed among the limitations in §8.

Under weak mixing the spectrum of $\mathcal{K}\_\theta$ separates into three parts — the trivial eigenvalue $\lambda = 1$ carried by the invariant measure, the transient modes strictly inside the unit disc, and the persistent structure near the circle — and WP-2026-005 §4.2 described the three classes and their correspondence with the three-level prediction hierarchy, whose predictive horizons Table 1 of that paper tabulates. The stochastic setting requires a refinement of the third class, the part of the spectrum that the framework calls *Level III*. A naïve reading would place Level III on the unit circle itself, identifying it with eigenvalues $|\lambda_j| = 1$ other than the trivial $\lambda = 1$. A more careful reading is required for stochastic systems with non-degenerate diffusion, which is the setting Axioms A4 and A5 jointly commit to: A4 supplies the dynamical form, A5 the irreducibility of the stochastic forcing. For an SDE $dX/dt = F(X, \theta, \varepsilon(t))$ with non-degenerate noise, the discrete-time transition kernel P_θ generally possesses a continuous spectrum in the interior of the unit disc, and the *only* point eigenvalue on the unit circle itself is the trivial $\lambda = 1$ corresponding to the invariant measure. Putative oscillatory modes with deterministic-skeleton frequency ω are subject to *phase diffusion* (as Document II §4.7 develops in the context of the SSI's dynamic structure): stochastic perturbations cause the phase of the mode to accumulate variance linearly in time, with the consequence that the corresponding stochastic Koopman eigenvalue takes the form $\lambda_j = e^{-\gamma_j + i\omega_j}$ with $\gamma_j > 0$ set by the magnitude of the diffusion. Even an oscillator whose deterministic skeleton possesses a perfect limit cycle has stochastic Koopman eigenvalues strictly inside the disc, with $|\lambda_j| < 1$, and the modulus deficit $1 - |\lambda_j|$ is set by the phase-diffusion rate. Searching for stochastic Koopman eigenvalues with $|\lambda_j| = 1$ exactly is therefore a category error: the appropriate object is not modes on the unit circle but modes whose decay rate is small relative to the time-scale of the observation window.

The framework redefines Level III accordingly. A mode is Level III in $\mathcal{K}\_\theta$ when its eigenvalue $\lambda_j = e^{-\gamma_j + i\omega_j}$ has decay rate γ_j small enough that $1/\gamma_j$ exceeds the relevant macrohistorical time-scale. Specifically, with T_{max} the longest period resolvable from a structurally stationary observation window of length T, a mode is Level III when its decay time $\tau_{decay,j} = 1/\gamma_j$ exceeds $3\ T_{max}$ — the threshold that makes the mode operationally indistinguishable from neutrally stable within the resolution the data support. Under this redefinition, "spectral structure persistence" does not require modes on the unit circle but modes whose decay is slow relative to the window. Whether such modes exist in macrohistorical dynamics is the empirical question of §5 and §6. What §2 establishes is that, *if* they exist, they sit in a neighbourhood of the unit circle whose width is set by the phase-diffusion rate of the underlying SDE — not at the circle itself, in any stochastic system worth modelling. The boundary case $\gamma_j = 0$ is recovered only in the deterministic-skeleton limit, which the framework does not assume; the practical question is one of how slowly γ_j decays toward zero as system noise is reduced, not whether $\gamma_j = 0$ exactly.

One technical assumption must be added, and unlike the parts of the spectral picture described so far it does not follow from weak mixing alone. Theorem 1 will require that the leading modes — the K eigenvalues nearest the unit circle, those carrying the persistent and slowly-decaying structure the paper is after — be separated from the rest of the spectrum by a gap. Formally, there is a spectral gap $\delta > 0$ such that the K leading eigenvalues are at distance at least δ from the remainder of the spectrum. This gap is what makes "the leading modes" a well-defined collection rather than an arbitrary truncation, and §4 will show that it is also what makes them stably estimable from finite data. The decision taken in the outline, and confirmed for this draft, is to carry the spectral gap as an explicit hypothesis of Theorem 1 rather than to derive it from more primitive conditions. Deriving it would require committing to structural properties of the map F — quantitative bounds on its form

— that CMH deliberately leaves open, and a derived gap would buy a marginally cleaner theorem statement at the cost of a much stronger and less defensible assumption. Stated as a hypothesis, the gap is transparent: it is a condition that the historical data either approximately satisfy or do not, and §6 reports the relevant diagnostic.

2.4 The evolution of eigenfunctions, and the reconciliation with A5

The reason the spectral decomposition matters is the simplicity of the temporal evolution it produces. Let φ_j be an eigenfunction of $\mathcal{K}\_\theta$ with eigenvalue λ_j, so that $\mathcal{K}\_\theta\ \varphi_j = \lambda_j\ \varphi_j$. Applying the definition of the operator and then iterating, the conditional expectation of this observable t steps into the future, given the present state x_0, is

$$\mathbb{E}[\ \varphi_j(X_t)\ \mid\ X_0 = x_0\ ] = \lambda_j^{\,t}\ \varphi_j(x_0).$$

The expected future value of an eigen-observable is its present value multiplied by a power of the eigenvalue. For a transient mode, $|\lambda_j| < 1$ with $1 - |\lambda_j|$ substantial relative to the time-scale of interest, the factor $\lambda_j^{\,t}$ shrinks rapidly and the expectation relaxes to zero. For a Level III mode, $\lambda_j = e^{-\gamma_j + i\omega_j}$ with γ_j small relative to the time-scale of interest, the factor $\lambda_j^{\,t} = e^{-\gamma_j t} e^{i\omega_j t}$ has modulus $e^{-\gamma_j t}$ that decays slowly on that time-scale: the expectation oscillates at angular frequency ω_j with amplitude decay slow enough to be operationally persistent within the observation window. Since any observable in $L^2(\mu\_\theta)$ can be expanded in eigenfunctions, the evolution of an arbitrary observable's expectation is a superposition of these elementary behaviours — rapidly-decaying terms from the deep interior of the disc and slowly-decaying oscillating terms from a neighbourhood of the unit circle. This is the precise sense in which the framework delivers "spectral structure persistence": the persistence is the slow attenuation, on the time-scale of the observation window, of the near-unit-circle terms in this superposition. The boundary case of *exact* persistence ($\gamma_j = 0$, $|\lambda_j| = 1$) is recovered only in the deterministic-skeleton limit, which the framework's commitment to non-degenerate stochastic forcing (A4 with A5) excludes; the framework's empirical content concerns slow-decay modes, not exact-persistence ones.

WP-2026-005 §4.5 stated the reconciliation with Axiom A5 discursively; the identity just derived is its formal content, and the resolution follows from it. A5 is a statement about the trajectory X_t; the Koopman identity is a statement about $\mathbb{E}[\varphi_j(X_t)\ \mid\ X_0]$, a distinct mathematical object. A persistent mode says that the *expectation* of a certain observable oscillates; it says nothing about where the system actually is at any future time, and it leaves the realised trajectory as indeterminate as A5 requires. This is why the construction had to be stochastic from the outset: in a deterministic Koopman framework the expectation and the trajectory would coincide, and the reconciliation would be unavailable.

The same distinction refines, rather than contradicts, Axiom A7: the geometric decay A7 describes applies with full force to the transient modes, while the Level III modes decay at the much slower rate $e^{-\gamma_j t}$ set by phase diffusion, bounded also by the eventual breakdown of structural stationarity. The consequences for the CMH predictive architecture are taken up in §7.

2.5 From invariant measure to transition kernel: what A2 provides and what it does not

The construction so far has used Axiom A2 for one specific purpose: to guarantee the existence of the invariant measure μ_θ, and with it the function space $L^2(\mu\_\theta)$ on which the operator is studied. It is necessary to be exact about the limits of what A2 delivers, because the cross-country argument of §3 will need more than A2 can supply, and the integrity of that argument depends on the gap being identified openly rather than papered over.

A2, in its CMH formulation, is a statement about the convergence of time averages: for a system in a fixed structural regime, the long-run time average of any macro-observable converges to a fixed expectation. Formally this is the statement that the invariant measure μ_θ exists and that the system's empirical distribution converges to it. What A2 does not state — and what cannot be extracted from it by any reading of its formalisation — is anything about the *transition kernel* P_θ. The invariant measure and the transition kernel are distinct objects, and the gap between them is not a technicality. Two Markov dynamics with entirely different transition kernels can share the same invariant measure: the invariant measure constrains the kernel but does not determine it.

The kernel and the invariant measure are distinct objects, and the stochastic Koopman operator is built from the former: A2 supplies the latter, not the former. The bridge between them — the ingredient that licenses the inference from shared structural regime to shared transition kernel — is the content of §3.1, where it is formulated as a named condition (Definition 3.2) together with Assumption 3.1 and used in the proof of Lemma 3.2a. The refined axiomatic attribution of Conjecture 1 that results — A2, A3 in closure form, A4, and Assumption 3.1 jointly, rather than A2 and A4 alone as WP-2026-005 stated — is made precise there.

Theorem 1: Conditional Koopman–A2 Correspondence

3.1 Preliminaries: ergodic class, and the bridge condition

The purpose of §3 is to prove, within the CMH axiomatic system, the connection that WP-2026-005 anticipated as Conjecture 1 — the Koopman–A2 Correspondence — as a conditional implication: shared ergodic class ⇒ shared Koopman spectrum, the countries differing only in the amplitudes with which the shared modes are excited. The substantive claim, that the countries of a particular empirical panel do belong to a common ergodic class, is empirical and no deductive argument can establish it; §3.3 states the falsification criterion and §5–§6 test it.

Two definitions are needed before the theorem can be stated. The first fixes what it means for two countries to belong to the same ergodic class. The second is the formal version of the kernel-sharing condition whose necessity was established in §2.5; it is stated here as a named condition, derived from Axioms A3 and A4, and the discussion following the definition is where the identifiability assumption flagged in the §2 hand-off is introduced and made explicit.

> **Definition 3.1 (Ergodic class).** Two countries c and c', with structural parameter vectors $\theta^{(c)}$ and $\theta^{(c')}$, belong to the same *ergodic class* over an observation window W if their structural parameters satisfy the regime-stability conditions of Axiom A2 throughout W and induce the same invariant measure: $\mu_{\theta^{(c)}} = \mu_{\theta^{(c')}}$.

Definition 3.1 is the direct operationalisation of A2's notion of "same statistical class". By A2, a country in a stable structural regime possesses a well-defined invariant measure; two countries are in the same ergodic class when those measures coincide. The definition deliberately fixes the ergodic class through the invariant measure alone, because that is exactly what A2 supplies — no more. Whether sharing an invariant measure is enough to share a Koopman operator is the question §2.5 showed must be answered separately, and the next definition is the answer.

> **Definition 3.2 (Structural Kernel-Sharing condition, SKS).** A family of countries $\{c\}$ satisfies the *Structural Kernel-Sharing condition* over an observation window W if any two members sharing a structural parameter vector — $\theta^{(c)} = \theta^{(c')}$ — also share the one-step Markov transition kernel up to equivalence: $P_{\theta^{(c)}} = P_{\theta^{(c')}}$ as transition kernels on the CMH state space.

The SKS condition is the formal bridge between the invariant measure, which A2 governs, and the transition kernel, which the Koopman operator is built from. It is not an additional axiom; the claim of this paper, and the content of Lemma 3.2a below, is that SKS is a *derived* consequence of Axioms A3 and A4 read jointly — subject, however, to one identifiability assumption that must be stated explicitly rather than left implicit, and subject also to a strengthening of A3 beyond its original CMH formulation. The reasoning is as follows. A4 specifies that the macrohistorical dynamics take the form $dX/dt = F(X, \theta, \varepsilon(t))$, with the structural parameters θ entering as an argument of the generating function F. A3 as formalised in CMH Document I §3.3 is an *existence* claim: there exists a structural cause producing the macrohistorical dynamics. For SKS to follow, the existence reading is insufficient — what is needed is the stronger *closure* form: that the generating relation specified by A4 is the *whole* of the dynamics, with no unmodelled country-specific primitive operating alongside it. We distinguish the two readings explicitly. **A3-weak** is the original existence claim of CMH Document I. **A3-strong** is the closure form: macrohistorical evolution emerges from causal relations among structural variables and from nothing outside them. Theorem 1 requires A3-strong, not A3-weak. The distinction matters because an empirical rejection of Theorem 1's antecedent on a given panel may indicate either that the panel lies outside any shared ergodic class (countries differ in θ) or that A3-weak holds but A3-strong does not on that panel (countries share θ but possess unmodelled primitive dynamics distinguishing them); the test cannot distinguish these without auxiliary information. Taken together, A3-strong and A4 say that the law of transition is a function of θ and of the law of the noise process alone: fix the structural regime and fix the noise law, and the transition kernel is thereby fixed. Two countries with $\theta(^c) = \theta(^{c'})$ therefore transition between states by the same law, which is the SKS condition.

The identifiability assumption is the qualification "and fix the noise law". A3 and A4 jointly determine the transition kernel as a function of the structural parameters *and* the law of the stochastic forcing $\varepsilon(t)$. For SKS to follow, it must be the case that countries sharing a structural regime also share the law of the forcing. We state this as a formal assumption.

> **Assumption 3.1 (Forcing-homogeneity).** *There exists a measurable functional G, common to all countries in the panel, such that the law of the stochastic forcing for country c is* $\mathscr{L}\_\varepsilon(^c) = G(\theta(^c))$. *In particular, two countries with* $\theta(^c) = \theta(^{c'})$ *have* $\mathscr{L}\_\varepsilon(^c) = \mathscr{L}\_\varepsilon(^{c'})$.

Assumption 3.1 says that the noise law is itself a function of the structural regime — that no country-specific noise law operates as an independent dynamical input alongside the structural parameter. Two countries in the same structural regime have, by Assumption 3.1, the same forcing law; the qualifier "more weakly but sufficiently" of an earlier formulation is absorbed into this single functional dependency.

The assumption is plausible within the CMH framework. A1 (statistical aggregation) already implies that the idiosyncratic component of macrohistorical noise is the residue of individual-level fluctuations averaged over a population above the critical scale, and such a residue is naturally regime-dependent rather than country-specific. But plausibility is not derivation, and the honest course is to carry Assumption 3.1 explicitly as a hypothesis of Theorem 1. With it, SKS is a theorem (Lemma 3.2a); without it, SKS would have to be assumed directly, and the theorem would rest on a less primitive foundation. Carrying Assumption 3.1 is the price of deriving SKS from the axiom system rather than postulating it, and it is a price worth paying because the assumption is inspectable: §6 reports a diagnostic — the comparison of residual noise structure across countries within a putative ergodic class — that bears directly on whether it holds.

3.2 Statement and proof

> **Theorem 1 (Conditional Koopman–A2 Correspondence).** *Let c and c' be two countries observed over a window W, with structural parameter vectors* $\theta(^c)$ *and* $\theta(^{c'})$. *Assume:*

(i) the regime-stability conditions of Axiom A2 hold for both countries throughout W, so that invariant measures $\mu\theta(^c)$ *and* $\mu\theta(^{c'})$ *exist;*

(ii) the panel is a macrohistorical system in the sense of CMH (Axioms A3 and A4 apply), and Assumption 3.1 (Forcing-homogeneity) holds across the panel;

(iii) the stochastic forcing is weak mixing with finite second moments;

(iv) the stochastic Koopman operators admit a spectral gap $\delta > 0$ separating the K leading eigenvalues from the remainder of the spectrum;

(v) the stochastic Koopman operators are normal on the K-dimensional subspace spanned by the leading eigenfunctions.

Then, if c and c' belong to the same ergodic class in the sense of Definition 3.1, their stochastic Koopman operators coincide, $\mathcal{K}\theta(^{c}) = \mathcal{K}\theta(^{c'})$, and consequently they share the leading eigenvalues $\{\lambda_j\}_{j=1}K$ and the leading eigenfunctions $\{\varphi_j\}_{j=1}K$. The two countries' trajectories differ only in the amplitude coefficients $\{b_j(^{c})\}$ and $\{b_j(^{c'})\}$ with which the shared modes are excited, these being determined by the respective initial conditions.

A word on hypothesis (v), which is new relative to the formulation anticipated in WP-2026-005. The local-normality condition is a regularity property of the operator on its leading invariant subspace: the leading eigenfunctions, equipped with the *L²(μ_θ)* inner product, form a near-orthogonal system rather than one in which they are nearly aligned with each other. The condition is standard in the operator-theoretic treatments of DMD that this paper builds on, and is the default rather than a strong restriction in that literature. Its role is technical and confined: it enters only in the perturbation bound of Lemma 3.4, where it ensures that the Lipschitz constant relating perturbations of the operator to displacements of the leading spectral projection is set by the spectral gap alone, rather than acquiring a multiplicative factor — the *pseudospectral condition number* of the leading eigenvectors — that can be large for highly non-normal operators (Trefethen and Embree 2005, §34; Davis–Kahan extensions to non-normal operators in Bhatia 2007, §VII.3). For systems whose Koopman operator is far from normal on the leading subspace, the conclusion of Theorem 1 still holds *qua* identity of operators, but the quantitative form of the perturbation control degrades and the empirical programme of §6 inherits a wider uncertainty in the recovered modes. The condition is inspectable: §6 reports the condition number of the empirical leading eigenvectors as a diagnostic, and a high value would qualify any positive Level III finding by flagging it as derived under conditions where the gap-based perturbation control is loose. The honest treatment is to name (v) among the hypotheses and to inspect it in the data, rather than to bury it in the proof.

The proof proceeds through four lemmas. The first two establish that membership in a common ergodic class forces, on the one hand, a shared invariant measure and, on the other, a shared transition kernel — the two halves of the gap identified in §2.5. The third converts a shared kernel into a shared operator and hence a shared spectrum. The fourth supplies the consequence about leading eigenfunctions and is the lemma that §4 will draw on. Full proofs are given in Appendix A; the arguments are summarised here in the form that makes the logical structure visible.

> **Lemma 3.1 (Invariant measure under A2).** *Under hypothesis (i), each country possesses a unique invariant measure μθ, and membership in a common ergodic class is equivalent, by Definition 3.1, to equality of these measures.* The existence of *μ_θ* is the content of A2 as formalised in §2.2; uniqueness follows from the weak-mixing hypothesis (iii), since a weakly mixing Markov kernel admits at most one invariant probability measure. The lemma is little more than a restatement of A2 in the operator-theoretic vocabulary, but it is worth isolating because it makes explicit that the *only* thing a common ergodic class delivers directly is the shared measure — the shared kernel must be obtained separately, which is the work of the next lemma.
>
> **Lemma 3.2a (Shared kernel from A3, A4, and Assumption 3.1).** *Under hypothesis (ii) — Axioms A3 and A4 applying to the panel, together with Assumption 3.1 — two countries with $\theta(^{c}) = \theta(^{c'})$ satisfy the Structural Kernel-Sharing condition: $P\theta(^{c}) = P\theta(^{c'})$.* The argument is the formalisation of the reasoning in §3.1. By A4 the one-step transition kernel is the law of the time-one map of the stochastic differential system *dX/dt = F(X,θ,ε)*; this law is a

functional of the generating function F, the structural parameter θ, and the law of the forcing ε. By A3 there is no further input: the dynamics are generated by the structural configuration and contain no primitive component unaccounted for by it. Hence the kernel is a function $P = \Pi(\theta, \mathscr{L}\_\varepsilon)$ of the structural parameter and the forcing law alone. Assumption 3.1 then gives $\mathscr{L}\varepsilon(^c) = \mathscr{L}\varepsilon(^{c'})$ whenever $\theta(^c) = \theta(^{c'})$, since both equal $G(\theta(^c)) = G(\theta(^{c'}))$ under the common functional G. Therefore $P\theta(^c) = \Pi(\theta(^c), \mathscr{L}\_\varepsilon) = \Pi(\theta(^{c'}), \mathscr{L}\_\varepsilon) = P\theta(^{c'})$. The equivalence qualification in Definition 3.2 absorbs the freedom to relabel the state space; the substantive content is that no country-specific dynamical degree of freedom survives the conditioning on θ.

The argument requires a specific reading of A3. CMH Document I §3.3 formalises A3 as the existence claim $P(E \mid X) \neq P(E)$ for some structural variable — that is, the assertion that the macrohistorical dynamics admit at least one structural cause. The use of A3 in Lemma 3.2a is stronger: it reads A3 as a closure claim, that the structural configuration *exhausts* the generative inputs to the dynamics, leaving room only for the noise law and not for an unmodelled country-specific component. This stronger reading is implicit in CMH Document I §3.3 — A3's operational implication "Confounders: must be identified and controlled" presupposes closure of the causal system — but the present paper makes the closure use explicit. The reader who prefers the minimal existence reading of A3 should treat the closure form as a separate auxiliary assumption that combines with A3 to license the argument; either way, the substantive content is the same.

Lemma 3.2a also refines the attribution of Conjecture 1 in WP-2026-005. WP-2026-005 §4.5 stated Conjecture 1 as following from Axiom A2 together with Axiom A4. Lemma 3.2a shows that this attribution is incomplete on two counts: A2 secures the invariant measure (Lemma 3.1) and A4 supplies the dynamical form, but the step from shared structural regime to shared transition kernel — the step without which the Koopman operators cannot be identified — requires A3 (in the closure-form reading discussed above) and Assumption 3.1 in addition. The Koopman–A2 Correspondence therefore rests on A2, A3, A4, and Assumption 3.1 jointly. This is a refinement of the originating statement rather than a contradiction of it: a methodological roadmap names the principal axiom in play, and a companion paper, in carrying out the proof, identifies the complete hypothesis set. The refinement strengthens the result, in that it makes the full deductive dependency visible and therefore checkable, and it makes Assumption 3.1 — the auxiliary hypothesis whose status the original formulation left implicit — a target of empirical inspection in its own right.

Lemma 3.2b (Shared invariant measure, consistency). *Under the hypotheses of Theorem 1, a shared transition kernel implies a shared invariant measure, consistently with Definition 3.1.* This lemma runs in the direction opposite to the logical need — the theorem requires shared kernel, and Lemma 3.2a already supplies it — but it is included for a reason of internal coherence. It verifies that the two characterisations of "same ergodic class" do not come apart: if A3 and A4 force the kernel to be shared, then the invariant measure, being uniquely determined by a weakly mixing kernel, is shared as well, so Definition 3.1's measure-based criterion is automatically met. Without Lemma 3.2b there would be a latent worry that the kernel-based and measure-based notions of ergodic class might be inconsistent; the lemma discharges that worry. The proof is immediate from the uniqueness established in Lemma 3.1.

Lemma 3.3 (Shared kernel implies shared operator and spectrum). *If $P\theta(^c) = P\theta(^{c'})$, then $\mathcal{K}\theta(^c) = \mathcal{K}\theta(^{c'})$ as operators on the common space $L^2(\mu\_\theta)$, and in particular the two operators have the same spectrum.* The stochastic Koopman operator was defined in §2.2 entirely through the transition kernel: $(\mathcal{K}\_\theta\, g)(x) = \int g(y)\, P\_\theta(x, dy)$. Two operators defined by the same kernel and acting on the same function space are the same operator. The function space is indeed common: by Lemma 3.2b the invariant measures coincide, so $L^2(\mu\theta(^c)) = L^2(\mu\theta(^{c'}))$. Identical operators have identical spectra and identical eigenfunctions. This lemma is the short formal hinge of the theorem, and it is short because the substantive work was done earlier — in §2.2's decision to build the operator from the kernel, and in Lemma 3.2a's demonstration that the kernel is shared.

> **Lemma 3.4 (Leading eigenfunctions are well defined and stable under perturbation).** *Under the spectral-gap hypothesis (iv), the K leading eigenvalues and their eigenfunctions form a well-defined collection. Stability of the leading spectral structure under perturbations of size η in operator norm is controlled by two distinct constants, depending on the regularity of the leading eigenstructure:*

(a) Under hypothesis (v) (local normality on the leading subspace), the leading spectral projection is displaced by at most LDK · η in Hilbert–Schmidt norm, where LDK = C / δ depends on the absolute spectral gap δ alone (Davis–Kahan form, with universal constant C depending only on K).

(b) Without hypothesis (v), the leading eigenvalues themselves are still displaced by at most LBF · η, where LBF = κ(V) is the condition number of the matrix V of leading right-eigenvectors with respect to the inner product on $L^2(\mu\_\theta)$ (Bauer–Fike form, Bhatia 2007, §VIII.1). The leading spectral projection is correspondingly displaced, but the projection bound is governed by a Dunford–Taylor contour integral of the resolvent rather than by a simple linear function of κ(V) and the gap. The placeholder upper-bound expression LBF,Π ∼ κ(V)/δ_{ps}, with δ_{ps} the pseudospectral gap, is an asymptotic scaling parameterisation valid in the limit of small operator-norm perturbations under bounded resolvent norms on the leading-cluster contour. Under empirically observed non-normality of moderate magnitude (κ(V) ∼ 20), the exact projection-displacement bound depends nonlinearly on the geometry of the leading-cluster contour and on the resolvent norm $\|(z - \mathcal{K}\_\theta)^{-1}\|$ along that contour, both of which can produce projection-error magnitudes substantially larger than κ(V)/δ_{ps} would suggest. The sharp pseudospectral form of this bound, with the resolvent norm computed numerically along Γ and integrated to obtain the true projection-error magnitude, is part of the WP-2026-007 refinement; the present paper uses the eigenvalue-displacement form of Bauer–Fike (which is sharp under diagonalisability alone) and treats the projection-displacement bound as a placeholder for the sharper bound that the methodological extension will provide.

The well-definedness in (a) and (b) is what the gap buys: with *δ > 0*, the *K* leading eigenvalues are unambiguously separated from the rest, and "the leading modes" denotes a specific finite collection rather than an arbitrary truncation. Form (a) is the sharpest available under normality and is the appropriate bound when hypothesis (v) is empirically satisfied. Form (b) is the appropriate bound when (v) fails empirically, as it does in the present panel (§6.4 reports *κW ≈ 20* in both sub-windows). The two bounds are not in conflict: (a) is a strengthening of (b) under (v), and (b) is the form that any analysis under empirically observed non-normality must use. Lemma 3.4 plays no role in the equality asserted by Theorem 1 itself, which is exact; its role is to make the theorem *usable*. Theorem 1 says the true operators share an exact leading spectral structure; Lemma 3.4 says that an estimator which recovers the operator only approximately will still recover the leading structure approximately, with an error controlled by the gap and (if non-normality is present) by the eigenvector condition number. It is therefore the bridge to §4, where the operator is known only through a finite-data estimate.

Proof of Theorem 1. Assume *c* and *c′* belong to the same ergodic class. By Definition 3.1 this means they share a structural regime satisfying A2, with $\mu\theta^{(c)} = \mu\theta^{(c')}$; Lemma 3.1 records that this measure exists and is unique under hypothesis (iii). By hypothesis (ii), Lemma 3.2a applies and yields a shared transition kernel, $P\theta^{(c)} = P\theta^{(c')}$. Lemma 3.3 converts the shared kernel into a shared operator on the common function space, $\mathcal{K}\theta^{(c)} = \mathcal{K}\theta^{(c')}$, and hence into a shared spectrum and shared eigenfunctions. By hypotheses (iv)–(v) and Lemma 3.4, the *K* leading eigenvalues $\{\lambda_j\}_{j=1}K$ and eigenfunctions $\{\varphi_j\}_{j=1}K$ form a well-defined collection, common to the two countries since the operators are identical. Finally, expanding each country's observed trajectory in the shared eigenfunction basis, the only country-dependent quantities are the coefficients of the expansion — the amplitudes $b_j^{(c)}$ and $b_j^{(c')}$ — which are fixed by the projection of each country's initial condition onto the shared basis. The shared dynamical architecture, country-specific excitation: this is the conclusion of the theorem. ■

> **Remark 3.1 (Approximate homogeneity).** Theorem 1 assumes exact forcing homogeneity (Assumption 3.1). The post-war panel violates this ($\bar{\rho} \approx 0.193$, §6.4): the shared-propagator interpretation of Corollary 3.1 must be read under an approximate form of the theorem, in which Lemma 3.2a's exact kernel identity weakens to a bound of order the homogeneity violation. A version of the theorem with quantified approximate homogeneity is deferred to WP-2026-007. The Q1 verdict of §6.2 is robust to the violation because it rests on point estimates of moduli, not on confidence intervals; the violation and the test's power limitations are independent qualifications and compound (§6.4, §6.7).

3.3 The corollary, and the residual empirical claim

> **Corollary 3.1 (The shared-propagator model is correctly specified).** *Under the hypotheses of Theorem 1, the multi-task estimation model of §4 — a single shared propagator across all countries of an ergodic class, with country-specific amplitude coefficients — is the correctly specified statistical model, not an approximation adopted for tractability.*

The corollary is the payoff of the theorem for the empirical programme, and it is worth stating why it is more than a restatement. Multi-task estimation, in which a single operator is fitted jointly across a panel rather than one operator per country, could be motivated on purely pragmatic grounds: the per-country time series are short, pooling buys statistical power, and the cost is some bias if the countries are not really alike. Read that way, multi-task estimation is a bias–variance compromise. Theorem 1 changes its status. If the theorem's hypotheses hold and the countries are in a common ergodic class, then there genuinely is one operator, and fitting one operator is not a compromise but the correct specification; the per-country alternative is then the misspecified model, fitting N distinct operators where one exists and thereby spending degrees of freedom on structure that is not there. This inversion — multi-task as correct specification rather than convenient approximation — is what licenses the joint estimator of §4 to be read as an estimator *of the shared Koopman operator* rather than merely as a regularised panel fit. It also sets up the falsification logic, because it tells us exactly what the failure of the substantive claim would look like in the data.

This is the point at which the conditional theorem and the substantive empirical claim must be separated cleanly, and the separation stated as a testable proposition. Theorem 1 is an implication: shared ergodic class ⇒ shared Koopman spectrum with persistent leading modes. The empirical programme cannot test an implication; it can only test the consequent on a given panel, and through the consequent draw inferences about the antecedent. The substantive claim of Conjecture 1 is that the antecedent — the conjunction of Definition 3.1 holding for the panel, Assumption 3.1 holding across the panel, and A3 (in closure-form) being satisfied by the panel's dynamics — is true of a specific historical panel. Corollary 3.1 supplies a model under which the consequent has empirical content: the multi-task shared-propagator estimator. The falsification logic for the substantive claim must be built on what that estimator can deliver from a fixed-window panel, and the data-side prerequisite for the model to be informative at all is that the estimator recover *some* persistent spectral structure beyond the trivial invariant-measure cluster — otherwise the question of whether the structure is shared across countries (the multi-task-vs-independent comparison) does not arise. §5.4 therefore operationalises the test through three sequential questions: first, whether persistent modes exist in the estimated spectrum (Q1); second, conditional on Q1, whether the multi-task estimator fits the panel as well as the per-country alternative on those modes (Q2); third, conditional on Q1, whether the periods of the identified modes correspond to historically documented cycles (Q3). The pre-registered falsification criterion is therefore on the conjunction Q1 ∧ Q2 ∧ Q3, with rejection at any sequential stage closing the question at that stage. §6.2 reports the result on Q1; §6.6 reports what is observable from the data when Q1 rejects.

A final observation places this structure within the CMH axiomatic system rather than merely beside it. Empirical failure of the substantive claim would not be a failure of Theorem 1 — the theorem is a proved implication and is not at risk from data. It would be a finding that the antecedent is false. A rejection of the substantive claim is, strictly, a rejection of the conjunction "Definition 3.1 holds for the panel ∧ Assumption 3.1 holds ∧ A3 holds in its closure form", against the alternative under which at least one fails. The test, as constructed in §5–§6, cannot discriminate among these conjuncts; what it discriminates is the conjunction as a whole. The framework-level inference to A2 specifically — A2 being the axiom that underwrites Definition 3.1 via the existence and uniqueness of the invariant measure — is therefore conditional on the auxiliary hypotheses being assumed correct. Assumption 3.1 is inspectable through the cross-country residual correlation diagnostic that §6.4 reports; A3 in closure form is harder to inspect directly, but Definition 3.1's measure-based characterisation is the most natural target if Assumption 3.1 is independently judged plausible. The disciplinary self-scrutiny that Axiom A8 was written to require operates on the whole hypothesis set: a rejection prompts re-examination of A2 first, but also of Assumption 3.1 and of the closure reading of A3, with the diagnostics of §6 partitioning the burden of explanation among them. The theorem secures the deduction; the data are left genuinely free to refute the antecedent, and the framework is then required to identify which component of the antecedent the data have refuted.

The Multi-Task Hankel-DMD Estimator

4.1 From an exact operator to a finite-data estimate

Theorem 1 is a statement about an object that is never directly observed. The stochastic Koopman operator $\mathcal{K}\_\theta$ is infinite-dimensional, defined through a transition kernel that is itself unknown; what the historical record provides is a finite collection of state vectors sampled annually across a panel of countries. The task of this section is to construct an estimator that maps that finite record to an approximation of the leading spectral structure of $\mathcal{K}\_\theta$, and to establish what can be proved about the relationship between the estimate and the object it estimates.

It is worth being explicit about a distinction that WP-2026-005 left implicit and that the precision of a companion paper requires. The Koopman operator and Dynamic Mode Decomposition are not the same thing, and they do not stand in the relation of theory to application so much as in the relation of estimand to estimator. The Koopman operator is the mathematical object — exact, infinite-dimensional, defined by the dynamics. DMD is a procedure that takes data and returns a finite matrix whose spectrum is, under conditions that must be stated and cannot be assumed, an approximation of part of the spectrum of that operator. Conflating the two invites a specific error: treating a property of the DMD output as though it were automatically a property of the underlying dynamics. The whole burden of §4.4 is to identify the conditions under which the inference from estimator to estimand is licensed, and the size of the gap that remains even when it is.

The estimator developed here has three components, each addressing a distinct feature of the CMH empirical setting: the limited temporal depth of annual historical series, the multi-country panel structure, and the choice of the observable functions in which the operator is represented. The three combine into what WP-2026-005 named the multi-task Hankel-DMD estimator. §4.2 and §4.3 construct the components and fix the observable dictionary; §4.4 states the statistical properties; §4.5 addresses missing data.

4.2 Component I: Hankel augmentation and the delay embedding

The Hankel augmentation strategy, and the short-series problem it addresses, are described in WP-2026-005 §4.4: when the usable series length is comparable to the state dimension, the standard DMD estimate of A is ill-conditioned, and delay augmentation enlarges the effective observable space without requiring additional observations. What WP-2026-005 did not supply is the justification appropriate to the stochastic setting; we add it here.

The Hankel augmentation addresses this by enlarging the effective observable space without requiring additional observations. In place of the raw state vector *X(t)*, the estimator works with the delay-augmented vector

$$XH(t) = [\ X(t)^{\top}op, X(t-1)^{\top}op, dots, X(t-\tau)^{\top}op\ ]^{\top}op,$$

which stacks the current state together with its values at τ preceding years. The columns *XH(t)*, assembled across *t*, form a Hankel matrix, and DMD is applied to this augmented matrix rather than to the raw one. The justification is operator-theoretic: the delay coordinates enlarge the dictionary of observables, bringing the span of the represented functions closer to the span of the true Koopman eigenfunctions and so reducing the projection error inherent in representing an infinite-dimensional operator by a finite matrix. More precisely, the delay-augmented vector spans a Krylov-style subspace generated by the iterates $\{x, \mathcal{K}\_\theta\, x, \mathcal{K}\_\theta^2\, x, \ldots, \mathcal{K}\_\theta\hat{}\tau\, x\}$ of the original observables under the Koopman operator: each additional delay coordinate adds one term to a Krylov sequence in $L^2(\mu\_\theta)$, and the span of $\tau+1$ such terms is the appropriate finite-dimensional approximation to the leading invariant subspace under the standard Krylov-projection error bounds (Korda and Mezić 2018, §4; Tu et al. 2014, §6).

Because the dynamics are stochastic (Axiom A5), the classical topological justifications for delay embedding — Takens' theorem and its successors — do not apply directly: they concern deterministic flows on compact manifolds. The Hankel augmentation is justified here on operator-theoretic rather than topological grounds, as Krylov-subspace enrichment of the observable space on which the Koopman operator acts; the augmented coordinates supply additional observables that improve the conditioning of the finite-section approximation without any appeal to deterministic reconstruction.

The choice of the delay τ is a bias–variance decision, and it should be stated as one rather than fixed by rule of thumb. A larger τ enlarges the observable dictionary and reduces the projection bias — the systematic error from representing $\mathcal{K}\_\theta$ in too small a function space. But a larger τ also consumes the series: an augmented matrix built from a series of length T with delay τ has only $T - \tau$ usable columns, so each year added to the delay is a column removed from the sample, raising the variance of every estimated quantity. The optimal τ balances these, and the balance depends on the periods of the modes one is trying to resolve. For the demographic and economic cycles that CMH targets, with hypothesised periods of twenty to a hundred years, a delay in the range of five to ten years is appropriate — long enough to enrich the dictionary materially but short enough to leave the bulk of the series intact. §6.4 reports the sensitivity of the mode estimates to τ within this range; the estimator's conclusions should not, and in a well-behaved case will not, depend sharply on the precise value chosen.

4.2a Identifying the invariant-measure cluster

Under Axiom A2 the dynamics possess an invariant measure whose corresponding Koopman eigenvalue is unity; in finite samples this appears as one or more near-unit eigenvalues forming the invariant-measure cluster. The persistence test concerns modes *beyond* this cluster, so the cluster must be identified before the test is applied. We identify it by period resolvability. An estimated eigenvalue *hatλ* is assigned to the invariant-measure cluster if it is real and positive, or if its implied period $T = 2\pi/|arghat\lambda|$ exceeds the sub-window length (so that no full cycle is observable and the mode is indistinguishable from a low-frequency drift toward the invariant measure). A mode is admissible as a *substantive persistent mode* only if it is oscillatory with a resolvable period, $2 \leq T \leq Tw_{in}d_{o}w$, and has modulus at or above the persistence threshold. This replaces the informal exclusion used in the previous version and makes the test reproducible: the verdict no longer depends on a visual judgment of which near-unit eigenvalues belong to the cluster. Cross-references: §5.4 and §6.2 apply this criterion operationally.

4.3 Component II: the multi-task shared propagator

The second component is the one that Theorem 1 directly licenses, and the logical dependence is worth making explicit because it changes how the component should be understood. The multi-task formulation imposes a single propagator matrix A across all countries of the panel:

$$XH^{(c)}(t+1) \approx A\, XH^{(c)}(t) \quad \textit{for every country } c,$$

with the country-specific information confined to the amplitude coefficients $b_j^{(c)}$ with which each country's trajectory excites the shared modes. The estimator for A is the solution of a stacked least-squares problem: the one-step pairs $(XH^{(c)}(t),\ XH^{(c)}(t+1))$ from all countries and all years are pooled into a single regression, regularised to control the conditioning of the pooled design matrix, and A is the common map fitted to the pool. The amplitudes are then recovered country by country by projecting each national trajectory onto the eigenbasis of the fitted A.

The multi-task formulation and its motivation are described in WP-2026-005 §4.4. By Corollary 3.1 this is not a modelling compromise: under the theorem's hypotheses the shared propagator is the correctly specified model, the per-country alternative is the misspecified one, and pooling converts Nc short series into one effective sample for estimating the shared spectrum.

A point of estimator design must be flagged here, because it is where the correct-specification reading could be quietly betrayed. The regularisation that controls the conditioning of the pooled regression introduces a bias toward whatever the regulariser's null target is — typically toward a propagator with smaller eigenvalue moduli. Left uncontrolled, this would interact with the bias discussed in §4.4 and compound it. The estimator therefore uses a regularisation whose strength is selected by an out-of-sample criterion on held-out country-years rather than fixed a priori, so that the regulariser is doing the work of conditioning the problem and not the separate work of shrinking the spectrum. §4.4 returns to this.

The third component is the observable dictionary. The baseline is the linear identity on the Hankel-augmented state, as pre-registered in §5.5 on the short-series grounds that WP-2026-005 §4.4 sets out; the kernel extension is deferred to WP-2026-007.[1]

4.4 Statistical properties of the estimator

This subsection states what can be proved about the relationship between the multi-task Hankel-DMD estimate and the leading spectral structure of $\mathcal{K}\_\theta$. Three properties matter: that the estimator converges to the right object as the sample grows, that its error at finite sample size has a characterisable and — this is the substantive finding — directional bias, and that the leading modes it targets are identifiable at all. A fourth result, on effective sample size under panel dependence, qualifies the convergence rate of the first when the panel is exposed to common shocks. Full proofs are deferred to Appendix A; the statements and their consequences are given here.

> **Proposition 4.1 (Consistency).** *Under the hypotheses of Theorem 1 — so that the shared-propagator model is correctly specified — together with the weak-mixing and finite-second-moment conditions on the forcing and the conditional independence of country trajectories given the structural regime θ, the leading spectrum of the multi-task Hankel-DMD estimator converges in probability to the leading spectrum of* $\mathcal{K}\_\theta$*. Convergence holds in the*

[1] Kernel DMD (Williams et al. 2015), discussed in WP-2026-005 §4.4, was pre-registered as a robustness check but not executed on the present panel (§8).

> *regime $Nc \to \infty$ at fixed series length T, and separately in the regime $T \to \infty$ at fixed Nc. The rate in the $Nc \to \infty$ regime is $O_p(Nc^{-1/2})$ at the nominal pooled sample size when conditional independence holds; under panel dependence the rate is the same in form but the effective sample size is reduced, as made precise by Proposition 4.4 below.*

The conditional independence hypothesis deserves a brief comment, because its presence in the proposition is not cosmetic and its plausibility in the empirical setting is not automatic. The hypothesis is the natural panel-data counterpart of the weak-mixing condition for a single trajectory: weak mixing controls temporal dependence within a country, conditional independence given θ controls cross-country dependence given the structural regime. Under it, the pooled sample behaves asymptotically as an independent sample of size $Nc \cdot (T\text{-}\tau\text{-}1)$ and the multi-task estimator inherits the standard parametric rate. Without it — that is, if countries within a putative ergodic class share stochastic components beyond their common θ, as may happen when a panel of polities is jointly exposed to global shocks — the rate is degraded by an effective-sample-size factor. Whether the hypothesis holds for a given panel is an empirical question, and Proposition 4.4, which makes the panel-dependent rate precise, also gives the diagnostic that bears on it. For the Western European panel of §5 this question is live and cannot be assumed away: §6.4 reports the diagnostic and interprets the precision claims of §6 against the effective rather than the nominal sample size.

The two regimes of Proposition 4.1 are not of equal practical relevance. Historical series have a length fixed by the calendar and the demand of structural stationarity; what a research programme can grow is the number of countries in the panel. The $Nc \to \infty$ regime is therefore the one that matters for CMH, and it is available only because of the multi-task formulation: a per-country estimator has no *Nc* to grow, and its consistency would rest entirely on $T \to \infty$, a limit the historical record does not offer. Consistency in the country dimension is the formal counterpart of Corollary 3.1's claim that pooling is correct specification — it is correct specification that makes the pooled limit converge to the operator rather than to some country-averaged artefact.

The convergence statement of Proposition 4.1 invokes Lemma 3.4. Lemma 3.4 established that the leading spectral projection of the true operator is Lipschitz-stable under operator-norm perturbations, with a constant set by the spectral gap δ — under the local-normality hypothesis (v) of Theorem 1. The consistency proof supplies the other half: the multi-task Hankel-DMD estimator converges to $\mathcal{K}\_\theta$ in an operator-norm sense on the represented subspace, at a rate that the proof in Appendix A makes explicit. Composing the two — estimator converges to operator, leading projection is stable under that convergence — yields convergence of the *leading* structure specifically, which is the structure Theorem 1 is about.

> **Proposition 4.2 (Finite-sample bias and its direction).** *At finite sample size with measurement noise on the observed state, the multi-task Hankel-DMD estimator is biased. When the observed state is contaminated by measurement noise, the bias is systematic and directional: the estimated eigenvalues are attenuated, their moduli displaced inward toward the origin of the complex plane relative to the true moduli. The magnitude of the attenuation increases with the measurement-noise variance and decreases with the effective sample size; its leading-order form is given in Appendix A.*

This proposition is the substantive finding of §4, and it requires comment beyond its statement, because the direction of the bias is not neutral with respect to the paper's central claim. The mechanism is standard in errors-in-variables regression and is well documented in the DMD literature: measurement noise enters both the predictor and the response matrices of the underlying regression, and the resulting bias pulls the estimated eigenvalues inward — the same attenuation that, in ordinary regression, pulls a slope coefficient toward zero. The consequence for CMH is specific. Level III prediction is the claim that there exist neutrally stable modes, eigenvalues with modulus at or very near one. An inward-attenuating bias displaces estimated moduli *away* from the unit circle and *toward* the interior. A genuinely neutral mode, $|\lambda_j| = 1$, will be estimated with modulus below one, and so will present as a slowly decaying transient rather than as a persistent oscillation. The bias, in other words, runs in the direction of *understating* spectral persistence — of making the system look more dissipative, more forgetful, less Level-III-amenable, than it is.

There are two things to say about this, and they pull in opposite directions. The first is reassuring and should be stated plainly: a bias that understates persistence is conservative with respect to the paper's thesis. If the uncorrected estimator nonetheless finds modes at or near the unit circle, those modes are if anything more credible than the point estimates suggest, because the bias was working against their detection. A positive Level III finding from a biased-conservative estimator is a robust finding. The second is the warning: the same conservatism means the uncorrected estimator will produce false negatives. A real neutral mode of modest amplitude can be attenuated below the detection threshold and missed entirely, and a programme that reported "no persistent modes found" from an uncorrected estimator would be reporting, in part, an artefact of its own bias. The estimator therefore must incorporate a debiasing step — the total-least-squares formulation of DMD (Askham and Kutz, 2018), or the forward–backward correction (Dawson et al., 2016), both of which target the errors-in-variables mechanism directly — and §6 reports results with and without it, so that the contribution of the correction is visible rather than buried. The honest summary is that the bias direction is a friend to positive findings and an enemy of negative ones, and the empirical programme must be built so that a null result cannot be an unexamined consequence of the estimator's own attenuation.

> **Proposition 4.3 (Identifiability of the leading modes).** *The K leading modes are identifiable from the multi-task Hankel-DMD estimator under three jointly necessary conditions: a rank condition on the pooled, delay-augmented data matrix, ensuring the represented observable space is genuinely K-dimensional and not collapsed by collinearity; the spectral-gap condition $\delta > 0$ of Theorem 1 hypothesis (iv), ensuring the leading modes are separated from the remainder and so are a well-defined target; and a temporal-extent condition, that the usable series length after delay augmentation span at least three full periods of the slowest mode to be resolved.* The third condition deserves a remark, because WP-2026-005 stated it as an empirical rule of thumb — three to five cycles for reliable identification — and it appears here as a derived condition rather than an asserted one. The derivation makes its content precise: with fewer than roughly three periods in the usable record, the frequency of the slowest mode and its growth rate are not separately identifiable from the data, because too few cycles have elapsed to distinguish a slow oscillation from a slow trend. The "*3T*" minimum-series-length criterion of WP-2026-005 §5 is the empirical shadow of this identifiability condition, and stating it as a condition rather than a heuristic is what allows §5 to treat the choice of estimation window as a formal requirement rather than a matter of judgement.

The choice of K — how many leading modes to retain — was left open in §2.3 and must now be pinned down, because Proposition 4.3 makes identifiability conditional on K and so K cannot be a free parameter tuned after the fact. The estimator selects K by the spectral gap itself: K is the number of eigenvalues lying above the largest gap in the sorted spectrum of the fitted propagator, so that the retained modes are exactly those the gap condition certifies as well-separated. This makes K a feature of the estimate rather than a researcher choice, which is the property needed to keep the identifiability claim of Proposition 4.3 from becoming circular. §5.4 records K among the quantities fixed before the substantive test is run.

> **Proposition 4.4 (Effective sample size under panel dependence).** *Suppose the conditional independence hypothesis of Proposition 4.1 fails: the one-step residuals across countries, after conditioning on the structural regime, have non-zero pairwise correlation $\rho_{cc'}$ averaged over country pairs. Let N denote the pooled column count of the Hankel-augmented predictor matrix — the actual number of pooled one-step observations, which after pre-processing per Appendix B differs from the theoretical $N_c \cdot (T - \tau - 1)$ by an amount depending on the country-specific span heterogeneity of the panel. Then the rate constant of Proposition 4.1 degrades by a factor reflecting both the cross-country dependence and the within-country temporal dependence retained by the weak-mixing process. At leading order in cross-country correlation,*

$$N_eff \approx (N)/(1 + overline\rho) \cdot (Nc - 1),$$

where $\bar{\rho}$ is the average pairwise residual correlation across countries. The full effective sample size, accounting also for within-country mixing-rate corrections, is derived in Appendix A.5. Confidence statements about the estimated leading spectrum must be expressed on N_eff rather than on the nominal pooled count N; the diagnostic overline$\hat{\rho}$, computed from the fitted residuals, is reported in §6.4 together with the corresponding N_eff values.

Proposition 4.4 is a routine consequence of the clustered-error analysis of panel data adapted to the operator-norm convergence of Proposition 4.1; its derivation is given in Appendix A.5. The reason for stating it in the main text rather than confining it to the appendix is that the practical content of the proposition is not technical but operational: it changes how the precision of the empirical estimates of §6 is reported, and it makes the assumption that underlies Proposition 4.1's rate inspectable rather than implicit. A panel in which $\bar{\rho}$ is close to zero behaves, for the purposes of spectral estimation, as a panel of *Nc* independent draws; a panel in which $\bar{\rho}$ is appreciable behaves, in the limit $\bar{\rho} \to 1$ of full common-component dependence, as a single effective draw regardless of *Nc*. The Western European panel will lie somewhere between these extremes — closer to the first within a structurally stable sub-window, closer to the second across European-scale shocks even within such a window — and the empirical claims of §6 will be calibrated accordingly.

4.5 Missing data

Historical panels are incomplete, and the incompleteness is concentrated in exactly the early decades where the long modes most need observations. WP-2026-005 committed the companion paper to addressing this, and the treatment must be explicit because the choice of missing-data strategy is not neutral with respect to the spectral estimates — each strategy distorts the estimated spectrum in its own direction, and a strategy chosen for convenience can manufacture or destroy the very modes the paper is testing for.

Three strategies are available, and the estimator uses all three, in a deliberate order. The conservative baseline is listwise restriction: the substantive test is run on the subset of country-years for which every selected variable is observed without interpolation. This baseline sacrifices sample size — sometimes a great deal of it — but it has the decisive virtue that its spectral estimates cannot be artefacts of imputation, because nothing has been imputed. The second strategy is multiple imputation, which restores the discarded country-years at the cost of a known risk: imputation models carry their own correlation structure, and that structure can imprint spurious modes on the spectrum, oscillations that are properties of the imputation model rather than of history. The third is low-rank matrix completion, which carries a subtler and more dangerous version of the same risk: completing a data matrix under a low-rank assumption is, almost by construction, the imposition of low-dimensional spectral structure, and using a low-rank completion to then test for low-dimensional spectral structure is close to circular. The danger is real enough that matrix completion is used here only as an outer robustness check, never as a primary analysis, and §6.4 reports its results with that status explicitly marked.

The protocol, then, is sequential and its order encodes the epistemics. The baseline result is the listwise one, because it is the one that cannot be an imputation artefact. Multiple imputation is reported as a sensitivity analysis: if it agrees with the baseline, the agreement buys back the discarded sample with confidence; if it disagrees, the disagreement is itself a finding about the fragility of the estimate. Matrix completion sits furthest out, a check on the check. A Level III conclusion is credited only if it survives the listwise baseline; the imputation-based analyses can strengthen a conclusion or flag its fragility, but they cannot establish one on their own. §5.3 operationalises this protocol on the specific structure of the Western European panel.

Empirical Test: The Western European Panel, 1820–2020

5.1 Panel construction

WP-2026-005 §5 specified the dataset selection criteria and the rationale for a Western European panel. The eight polities — France, the United Kingdom, Germany, Italy, Spain, the Netherlands, Sweden, and Belgium (FRA, GBR, DEU, ITA, ESP, NLD, SWE, BEL) — and the two sub-windows, 1820–1913 and 1945–2020 with the two World Wars excluded by a strict cut, implement those criteria; eight countries rather than ten because, by Proposition 4.4, adding the tightly coupled Scandinavian candidates would raise $\bar{\rho}$ without raising N_eff. The nominal total is 170 country-years per country, 1,360 across the panel; §5.2 tabulates listwise-complete coverage by country. The specification in §5 preceded the assembly of any panel data, in discharge of the A8 obligation on Conjecture 1.

The pre-war sub-window deserves a candid remark. Even with the World Wars excluded, the 1820–1913 period is not free of structural shifts of the kind that A2 worries about — Italian unification (1861) and German unification (1871) reshape the political map, the 1848 revolutions affect multiple countries simultaneously, the Franco-Prussian War (1870–1871) and the formation of the Third Republic introduce regime changes, and the Belle Époque post-1880 is a structurally distinct period from the pre-1850 era of restoration regimes. The pre-war sub-window is therefore not "structurally stationary" in any strong sense; it is, more honestly, a window in which structural change is more gradual than in the war years, with no single discontinuity of comparable magnitude to the two World Wars. Modes resolvable in this sub-window will be those whose periods are short enough to permit identification within sub-periods between major shifts — practically, modes with periods up to roughly thirty to forty years. Longer modes nominally identifiable by Proposition 4.3 (T) over a 94-year window are at risk of being structural-shift artefacts rather than dynamical features. §6.4 reports a cross-window robustness check that addresses this concern directly: a mode that appears in both sub-windows is more credible than a mode that appears only in one. The post-war sub-window 1945–2020 is itself not free of shifts (decolonisation, EU integration, German reunification 1990) but they are smaller in magnitude than the inter-war discontinuities and the sub-window is more genuinely stationary in the technical sense.

Five categories of country-year-variable triples are systematically unavailable for historical reasons rather than from data-collection failure, and the empirical execution made the boundaries concrete: Italian V-Dem indices before 1862 (the first post-unification year V-Dem codes), all variables for Belgium before 1830 (independence), Belgian Maddison gdppc before 1846, the COW Composite Index of National Capability for Germany 1945–1954, and CINC for all panel countries 2017–2020 (the v6.0 release endpoint). The Germany case warrants explanation. The Federal Republic of Germany was formed in 1949 but did not achieve full sovereignty in the Correlates of War system until the Deutschlandvertrag of 5 May 1955; consequently no CINC value is assigned to any German state for the years 1945–1954. The harmonised panel therefore uses COW code 255 (unified Germany) for 1820–1944 and 1990–2016, COW code 260 (Federal Republic) for 1955–1989, and flags 1945–1954 and 2017–2020 as systematically unavailable. This is a property of the data, not a methodological choice — but it does mean that Germany contributes fewer listwise-complete country-years to the post-war sub-window than the other panel members, with consequences for the spectral estimates discussed in §6.4.

5.2 Variables and the state-space dimension

The state vector X for the panel is five-dimensional, with one or two variables in each of the four CMH dimensions D, E, P, S. The variable set is the following.

For the demographic dimension D: population (in thousands), from the Maddison Project Database 2023. Population is the simplest D variable available with full panel coverage over the entire 1820–2020 window and is the canonical demographic state variable in CMH treatments.

For the economic dimension *E*: GDP per capita in 2011 USD PPP, from the Maddison Project Database 2023. GDP per capita rather than total GDP, because the panel includes countries of very different sizes and the level rather than the absolute aggregate is the comparable structural quantity.

For the political dimension *P*: the Liberal Democracy Index and the Electoral Democracy Index, both from V-Dem v16. Two P variables rather than one reflects two considerations: P is the dimension in which the Koopman framework's primary predictions of Conjecture 1 were originally formulated, so it deserves two-variable resolution; and the two V-Dem indices are conceptually distinct components of the political regime, capturing institutional features that move on different time scales. The Liberal Democracy Index emphasises liberal-constitutional dimensions (rule of law, minority protections), while the Electoral Democracy Index emphasises electoral-procedural dimensions; modes operating on one but not the other would be lost if only one variable were retained.

For the social/structural dimension *S*: the Composite Index of National Capability (CINC) from the Correlates of War National Material Capabilities dataset v6.0. CINC is a composite of military, demographic, and economic capabilities expressed as a country's share of the global total; it is the canonical S variable in the international-relations literature and has annual coverage from 1816 to 2016, with the caveat about Germany 1945–1954 noted in §5.1.

The Ψ dimension is excluded for the reasons stated in WP-2026-005 §5 (third criterion); §8 records the consequence for inferential reach and §7.2 the design that lifts it.

The state-space dimension is therefore $n = 5$. With the Hankel delay $\tau = 7$ specified in §5.5, the augmented dimension is $n(\tau+1) = 40$. By the design considerations of §4.3 and the rank condition Proposition 4.3 (R), the pooled panel must have effective sample size $N_{eff} \geq 10 \cdot 40 = 400$ for the spectral estimates to be well-conditioned.

The harmonised panel produced by the data-collection pipeline contains the following counts at the time of writing. Of 1,360 nominal country-years, 1,250 are listwise complete across all five variables — coverage 91.9 percent. By sub-window: the 1820–1913 sub-window contributes 684 listwise-complete country-years out of 752 nominal (90.9 percent), the 1945–2020 sub-window contributes 566 out of 608 (93.1 percent). Both sub-windows are well above the 400-country-year threshold, and within sub-window the design satisfies the rank condition by a substantial margin. The Germany sub-window-2 coverage of 62 country-years is the lowest individual-country count and is discussed in §6.4 as a sensitivity element; the other seven countries contribute between 68 and 94 listwise-complete years.

The realised totals are summarised in the following coverage matrix.

Country	**Sub-window 1 (1820–1913)**	**Sub-window 2 (1945–2020)**	**Combined**
France	94/94	72/76	166/170
United Kingdom	94/94	72/76	166/170
Germany	94/94	62/76	156/170
Italy	52/94	72/76	124/170
Spain	94/94	72/76	166/170
Netherlands	94/94	72/76	166/170
Sweden	94/94	72/76	166/170
Belgium	68/94	72/76	140/170

Total	**684/752**	**566/608**	**1,250/1,360**

A remark on variable transformations. GDP per capita is non-stationary in levels (it grows secularly); the estimator works with its log-difference, the annual log growth rate, which is stationary within each structural regime. Population is similarly log-differenced. The two V-Dem indices are bounded in *[0,1]* and treated as locally stationary in levels within each sub-window. CINC is a share variable in *[0,1]* and is treated as locally stationary in levels. These transformations are the same that the standard panel-econometric literature applies to these variables and require no special justification beyond the stationarity demand inherited from §2.2 — the operator-theoretic framework requires the represented observables to lie in $L^2(\mu\_\theta)$, which fails for variables that grow without bound.

5.3 Missing data: operationalising the §4.5 protocol

The missing-data protocol of §4.5 is implemented on the panel as follows. The three sources have non-trivial coverage gaps over the window, and the gaps are not randomly distributed; they cluster in the systematic-unavailability categories enumerated in §5.1 (Italian V-Dem pre-1862, Belgian variables pre-1830, Belgian Maddison gdppc pre-1846, German CINC 1945–1954, all CINC 2017–2020). The protocol distinguishes three categories of country-year-variable triples, and the categorisation drives the analysis ordering.

A triple is *fully observed* if the source records a value. The set of fully observed triples defines the *listwise baseline* panel: the substantive test is run, in its primary form, on the 1,250 listwise-complete country-years tabulated in §5.2.

A triple is *imputable* if the value is missing but a multiple-imputation model could be plausibly fitted from neighbouring observations. The set of fully observed plus imputable triples defines the *multiply-imputed* panel, used for the sensitivity analysis.[2]

A triple is *systematically unavailable* if the source has no observation and the absence reflects a substantive issue rather than a data-collection gap — a country that did not exist, a regime change that makes the variable definitionally inapplicable, a territorial discontinuity that makes pre-and-post values incomparable. The five categories enumerated in §5.1 fall here. Systematically unavailable triples are flagged, with the historical reason, in the data-collection prospect in the replication repository (Appendix B), and are excluded from both the listwise and multiply-imputed panels. They cannot be imputed because imputation would attribute structure where none exists.

The matrix-completion sensitivity check, the outermost of the three layers in §4.5, is applied as a single robustness analysis in §6.4 rather than as a primary analytic path. Its results are reported alongside the listwise baseline and the multiple-imputation sensitivity, with the explicit note that any conclusion drawn from it is conditional on the low-rank assumption that the completion imposes — an assumption that is, by design, sympathetic to the framework's central prediction and therefore not the strongest possible test of it.

5.4 The three falsification questions and the pre-specified criteria

[2]Multiple imputation is by chained equations with twenty imputations, country fixed effects, year dummies, and the other panel variables as predictors; the spectral estimator is run on each imputation and the results combined using Rubin's rules adapted to spectral quantities. The Maddison Project Database 2023, unlike earlier releases, does not surface the i_cgdppc interpolation flag at the country-year level, so the panel cannot distinguish Maddison-interpolated from Maddison-observed values within the formally-fully-observed set; the reconstruction quality is treated as a general caveat in §8.

Theorem 1 admits three empirical tests, each corresponding to a distinct question that the data can answer about the conditional claim's antecedent. Each is operationalised with a pre-specified criterion for acceptance or rejection. These criteria are fixed before the data are processed; they are not adjusted in light of §6's results.

Question 1 — Existence of persistent modes. Does the multi-task Hankel-DMD estimator, applied to the listwise baseline panel, return any eigenvalues with modulus at or near the unit circle after debiasing? The criterion is quantitative. After debiasing per the TLS-DMD scheme specified in §5.5 (with the forward–backward variant of Remark A.3 reported as a secondary check in §6.4), an eigenvalue is classified as *persistent* if it is admissible under the period-resolvability rule of §4.2a and its estimated modulus $|\hat{\lambda}_j|$ exceeds the threshold $1 - 1/(3\ T_{max})$, where T_{max} is the longest period resolvable from the sub-window length per Proposition 4.3 (T). We commit to this single form throughout, reporting T_{max} from the embedding-corrected usable length: $T_{max} \approx 31.3$ years pre-war and ≈ 25.3 years post-war, giving thresholds 0.989 and 0.987 respectively. The factor of three follows the applied DMD literature's convention that reliable spectral identification requires the observation window to span at least three full periods of the slowest mode of interest (Tu et al. 2014, §3.2; WP-2026-005 §5); the threshold is the linear approximation of $exp(-1/(3\ T_{max}))$, calibrated so that a mode at the threshold has exact decay time $3\ T_{max}$ to first order in $1/T_{max}$, with approximation error of order $1/T_{max}^2$ ($\sim 0.01\%$). The earlier production-code rounding $1 - 1/93 \approx 0.989$ differs from the exact $1 - 1/(3 \cdot 31.3) = 0.9894$ by $\approx 4 \times 10^{-4}$ and affects no verdict; we adopt the exact form here. The choice of threshold operationalises "neutrally stable on the window's time scale": a mode meeting it has, at the resolution the data support, a decay time exceeding the sub-window itself. Rejection criterion for Question 1: if zero admissible eigenvalues meet the threshold after debiasing in both sub-windows, Level III spectral persistence is rejected for this panel.

The binary threshold's stringency is a design choice and not a derivation from first principles: a factor of two or five in place of three would produce a different threshold without changing the formal structure of the test, and the choice of three is justified by convention rather than by a power-analysis argument. The threshold is also stringent relative to the redefined Level III concept of §2.3, which identifies Level III modes as those with decay time $1/\gamma_j$ exceeding $3\ T_{max}$ — exactly the threshold the binary Q1 operationalises. The pre-registered Q1 is therefore a stringent operationalisation of the Level III definition under stochastic dynamics: it asks whether the data return modes whose decay rate is bounded below the phase-diffusion floor of the conjectured underlying SDE on the time-scale of the sub-window. It does *not* ask whether the data return modes with $|\lambda_j| = 1$ exactly, which would be a category error in a stochastic framework (§2.3); the threshold is the operationalisation that closes the gap between the discrete-data binary criterion and the continuous-time Level III definition. To make the underlying continuum visible alongside the binary verdict, §6.2 reports the *estimated decay time* of the leading mode in each sub-window as a continuous quantity, $hat\tau d_e c_a y = -1/log|hat\lambda_1|$, regardless of whether the threshold is crossed. The continuous quantity is the natural target of any subsequent re-test under a different threshold or a stratified one (Remark 7.1), and reporting it here preserves the information that the binary criterion compresses. The interpretation of a sub-threshold $hat\tau d_e c_a y$ — whether it constitutes evidence against Conjecture 1 or merely evidence of the test's limited power on this panel — is taken up in §6.7 and §8.

Question 2 — Cross-country sharing of the modes. Are the persistent modes shared across the panel, as Theorem 1 predicts, or do they differ country-by-country? The test is the joint-fit versus independent-fit comparison foreshadowed in Corollary 3.1 and §3.3. The multi-task estimator (one shared A across the panel) is compared with an independent-fit estimator (one $A^{(c)}$ per country) on the same listwise baseline panel. The criterion is the out-of-sample log-predictive score on a held-out test partition: for each country, the last 20 percent of country-years are reserved as a test set, the two estimators are fitted to the remaining 80 percent, and each is scored on its one-step-ahead predictive log-likelihood on the test set. The shared-propagator model is *not rejected* if its mean test-set log-likelihood is within one standard error of the independent-fit model's; it is *rejected* if its test-set log-likelihood is more than two standard errors below the independent-fit model's. The intermediate region (between one and two standard errors below) is reported as inconclusive, with the gap quantified rather than dichotomised. The standard-error band is computed by stationary bootstrap of country-year blocks with 1000 replications, block length set to $2 \lceil hat\tau IAC \rceil$ years where $hat\tau IAC$ is the integrated autocorrelation time of the residuals from the *multi-task* fit (the null model in the comparison), so that the block length is invariant under the comparison direction.

The asymmetry of the criterion — "within one SE" for acceptance, "more than two SE below" for rejection — is deliberate and follows the logic of correct specification. Under Theorem 1, the shared-propagator model should match the independent-fit model up to estimation noise, because they are fitting the same true operator with different effective sample sizes; an apparent advantage of the independent-fit model is itself informative only if it exceeds noise. The strict acceptance band guards against accepting on the strength of noise that happens to align; the lenient rejection band guards against rejecting on the strength of noise that happens to favour the unrestricted model.

Question 3 — Correspondence with documented historical cycles. If persistent modes are found and they are shared across the panel, do their estimated frequencies correspond to cycles documented in independent historical and economic literature? This is the question of substantive coherence: a Koopman mode is more credible as a description of macrohistorical dynamics if its period matches independently characterised phenomena (Kondratiev waves around 50 years, demographic cycles of 20–25 years for fertility echo effects, secular cycles in the cliometric literature) than if its period corresponds to no known phenomenon. The test cannot be quantitative in the same way as Questions 1 and 2 — there is no formal statistic for "correspondence" — but it can be made disciplined. Each identified persistent mode is matched, before §6 examines specific candidates, against three pre-specified period bands: short cycles (10–25 years, demographic and business cycle range), medium cycles (25–60 years, Kondratiev range), long cycles (60–120 years, secular cycles in the Goldstone–Turchin tradition). A mode is classified as *historically corresponding* if its estimated period $T_j = 2\pi/\omega_j$ falls within one of these bands and its eigenfunction loadings (which variables it concentrates on) are consistent with the documented mechanism for that band. The pre-specified band structure prevents post-hoc fitting: the matching cannot be moved after the modes are estimated.

The criterion for Question 3 is presence rather than absence: at least one persistent mode whose period and loading structure match an independently documented cycle would support the historical-correspondence claim. The absence of any such match would be a more diagnostic finding than the presence of a partial one, because it would suggest the persistent modes are statistical artefacts disconnected from substantive history. A mode classified as persistent (Question 1) and shared (Question 2) but historically uncorrespondent (Question 3) would be a striking finding requiring its own interpretation — possibly a real but undocumented macrohistorical cycle, possibly a methodological artefact that survived all three filters.[3]

5.5 Pre-specified parameter values

Several quantitative choices that affect the estimator's behaviour cannot be left to be made when the data are inspected; recording them here is the procedural form of pre-registration for parameter selection.

The Hankel delay is $\tau = 7$ years. The choice sits in the middle of the 5–10 year range argued for in §4.2 and is motivated by the resolution requirements of the modes of interest: $\tau = 7$ is sufficient to support frequency identification of modes with periods down to roughly 14 years (the Nyquist limit of the augmented sample),

[3]Q3 is calibrated by Q1's persistence cut: only modes that pass Q1 enter, so the effective candidate count is small. For panels where Q1 admits a larger set, a Bonferroni-style correction across the three pre-specified bands would be required to control the false-positive rate of "at least one match in any band"; the present test relies on the Q1 conditioning to keep the comparison count low. For larger panels in WP-2026-007 and successors this calibration will need to be revisited.

which covers the lower end of the demographic-cycle range, while leaving 87 usable columns per country in the pre-war sub-window and 69 in the post-war sub-window.

The regularisation strength λ for the multi-task ridge regression is selected by 10-fold cross-validation over countries (each fold withholds one country) with the predictive log-likelihood as the criterion. The cross-validation is run within the listwise baseline panel only; the selected λ is then applied unchanged to the multiply-imputed and matrix-completed panels in §6.4.

> **Pre-registration note.** The CV criterion as written here is the criterion the §5.5 pre-registration commits to. The production implementation (v3 of the estimator pipeline, Appendix B) replaced predictive log-likelihood with predictive RMSE per state dimension during execution, in response to a numerical pathology in the log-likelihood criterion documented in the repository implementation history (Appendix B): the per-dimension Gaussian log-likelihood is dominated by Hankel-augmented dimensions with near-zero residual variance at $\lambda = 0$, making the criterion monotone in λ and uninformative for selection. The change is recorded here as an explicit pre-registration deviation rather than buried in the pipeline trail of §8. The substantive content of the §6 results does not depend on the CV criterion choice because the Gavish-Donoho rank selection is the operative dimensionality control in the canonical pipeline; the deviation is disclosed for completeness under the A8 commitment to inspectable pre-registration.

The number of leading modes K is determined by the spectral gap, as fixed in Proposition 4.3 and §4.4. The selection criterion is the largest gap in the sorted modulus spectrum of the fitted propagator, with the gap measured on a logarithmic scale to handle the eigenvalue clustering near the origin. The pre-registered cutoff is a multiplicative factor of *1.5* between consecutive sorted moduli on this log scale, used as a screening heuristic; the absolute spectral gap $\delta = |hat\lambda K| - |hat\lambda K_{+1}|$ that enters the perturbation bound of Lemma 3.4 is a distinct quantity, and §6 reports both. The multiplicative cutoff is a screening rule, not the theoretical gap that licenses Lemma 3.4's Lipschitz bound: two eigenvalues at *0.9* and *0.6* satisfy the *1.5* factor but have an absolute gap of *0.3*, which is the quantity that enters the operator-theoretic perturbation control. The distinction matters because a positive screening result is necessary but not sufficient for stable leading-subspace identification; §6.1 reports the absolute gap alongside the multiplicative factor for both sub-windows. If no log-gap exceeding *1.5* appears in the sorted spectrum, no K is reported and the conditions of Proposition 4.3 (G) are recorded as failed for the panel.

Debiasing is performed by total-least-squares DMD (Askham and Kutz, 2018). The forward–backward correction of Dawson et al. (2016) is reported as a secondary check in §6.4. The uncorrected standard DMD estimator is also reported, for transparency about the magnitude of the bias correction.

The observable dictionary is the linear (Hankel-augmented identity) dictionary as the primary specification, per §4.3's choice for the short-series regime. The Gaussian-kernel DMD variant of Williams et al. (2015), with a kernel width set by the median heuristic on the pooled state vectors, is pre-registered as a robustness check in §6.4, contingent on the availability of a canonical Mercer-style feature-space truncation implementation in the multi-task panel setting. As §6.4 and §8 record, no such canonical implementation was developed within the timeline of this paper; the kernel-DMD robustness check is therefore reported as not executed rather than as failed, and is deferred to WP-2026-007 where the panel extension provides a more favourable N_{eff} for kernel methods to be informative.

Empirical Results

6.1 Estimator implementation and effective rank

Implementation of the multi-task Hankel-DMD estimator specified in §4 and parametrised by §5.5 follows the canonical DMD pipeline of Tu et al. (2014), Algorithm 2: pooled construction of the predictor and response matrices from per-country Hankel-augmented states, economy singular-value decomposition of the pooled predictor, truncation at the effective rank, and computation of the reduced propagator on the truncated subspace. The effective rank itself is data-driven, not tuned: the universal hard-threshold of Gavish and Donoho (2014) on the singular values of the pooled predictor matrix selects r as the number of singular values exceeding $\omega(\beta) \cdot \sigma_{me}d_{ian}$, with β the aspect ratio of the predictor matrix and $\omega(\beta)$ the polynomial coefficient tabulated for the unknown-noise setting. The full pipeline and the replication code are in the project's GitHub repository (Appendix B).

The implementation went through three revisions in the course of the empirical execution, and a note on the development trajectory is owed here. The first two used unprojected ridge regression in the full Hankel-augmented dimension and produced numerically incoherent spectra. The first version exhibited eigenvalues with $|\lambda|$ approaching 3, in formal contradiction with the contraction property of the stochastic Koopman operator on $L^2(\mu\_\theta)$ established in §2.2. The bug was traced to two distinct sources: a TLS-DMD implementation that lacked the rank truncation step canonical to the method, and a cross-validation criterion based on Gaussian log-likelihood that was unusable because residuals on certain Hankel-augmented dimensions are numerically zero at $\lambda = 0$. The second revision corrected the TLS step but left the log-likelihood criterion in place; the bug-fix did not resolve the monotone-in-λ cross-validation pattern. The third revision moved to the canonical Tu et al. pipeline, replaced the log-likelihood criterion with predictive RMSE, and produced the well-conditioned spectra reported below. The full trail is documented in the repository implementation history (Appendix B) because the §5.5 pre-registration committed to "multi-task Hankel-DMD" without specifying which of the several variants in the applied DMD literature was intended, and the canonical pipeline is the one the literature treats as default.

Applied to the listwise baseline panel, the Gavish-Donoho threshold selects effective rank $r = 16$ for the 1820–1913 sub-window (pooled $N = 612$, Hankel-augmented dimension $d = 40$) and $r = 15$ for the 1945–2020 sub-window ($N = 494$, $d = 40$). The truncation is substantial — roughly 60 percent of the nominal Hankel-augmented dimension is below the noise floor — and this is itself a result. It says that the represented state-space has a much lower effective dimension than the nominal Hankel-augmented dimension would suggest, which is consistent with the picture that motivates the multi-task formulation in the first place: the panel countries share a leading low-dimensional dynamical structure, and the high-dimensional residual is largely common observational noise.

Cross-validation over the rank parameter, performed leave-one-country-out, produces a curve that is informative in itself and worth dwelling on because it surfaces a methodological point with broader implications. For the pre-war sub-window, the predictive RMSE is minimised at $r = 2$ and rises sharply through $r = 4$, then partially recovers at higher ranks. The minimum at $r = 2$ corresponds to a one-step prediction that retains only the invariant-measure mode and its conjugate — the prediction "next year is like this year, on average" performs better one-step than any richer model. The divergence between this CV-selected rank ($r = 2$) and the Gavish-Donoho-selected rank ($r = 16$) is a factor of eight, and it is not a numerical artefact: the CV criterion targets one-step forecast accuracy, while the Gavish-Donoho criterion targets correct dimensionality of the signal subspace. These are two distinct questions with different answers on the same data, and the divergence between them illustrates why predictive RMSE is the wrong criterion for *spectral* identification. The leading two modes carry the bulk of the one-step predictive content because conditional expectation is dominated by the invariant-measure-like behaviour at the one-step horizon, but the spectral structure of interest in §6.2–§6.3 lives in the lower modes that the gap-based rank selection retains. For the post-war sub-window the cross-validation curve is non-monotone with a minimum at $r = 20$, indicating that higher ranks provide marginal predictive improvement; this also does not reflect spectral identification but is a different artefact —

when noise is plentiful, additional dimensions can absorb noise without harming one-step prediction, but they do not correspond to identifiable spectral structure. The Gavish-Donoho rank, fixed in §4.3 as the criterion for the rank-condition (R) of Proposition 4.3, is therefore the rank used for spectral analysis below; the cross-validation curves are reported as supplementary information and as a diagnostic that any future application of DMD to historical panel data should anticipate. The methodological takeaway is that future pre-registrations should be explicit about which criterion drives rank selection — predictive performance is appropriate for forecasting applications, while signal-noise separation is appropriate for spectral identification — and not assume that the two criteria agree.

6.2 Question 1: persistence under the pre-registered criterion

Applying the criterion of §5.4, with the invariant-measure cluster identified by the period-resolvability rule of §4.2a, the test rejects Q1 in both sub-windows: no mode with a resolvable period attains the persistence threshold. We state immediately the qualifier that governs the interpretation of this verdict, namely that the post-hoc power of the test against finite-modulus alternatives above the threshold is effectively zero at the available effective sample size (§6.4). The rejection is thus a rejection of detectability; the surrogate analysis below (§6.7) shows that the one sub-threshold correspondence of interest is not distinguishable from noise.

For the pre-war sub-window the estimator returns sixteen eigenvalues, of which the leading two are $|hat\lambda_1| = |hat\lambda_2| = 0.9936$, with $hat\lambda_{1,2} = 0.9936 \pm 0.0034\, i$. The remaining fourteen eigenvalues have moduli ranging from 0.91 down to 0.32, all strictly inside the unit disc. The two leading eigenvalues are a complex-conjugate pair with vanishingly small imaginary part, which is the operator-theoretic signature of the invariant measure: they correspond to the constant function on $L^2(\mu\_\theta)$ and a near-zero-frequency partner that emerges from the finite-sample estimation of a unit eigenvalue. Two eigenvalues meet the persistence threshold 0.989; both belong to the null-mode cluster. No eigenvalue with non-trivial frequency content (defined as $|Im(hat\lambda)| > 10^{-6}$ and $|arg(hat\lambda)| > 10^{-4}$ — that is, distinguishable from the near-zero-frequency null cluster at machine-tolerance levels) meets the threshold. Under the pre-registered criterion of §5.4, **Question 1 is rejected for the pre-war sub-window**: the spectrum contains no neutrally stable mode beyond the invariant measure. The leading non-trivial-frequency mode has $|hat\lambda_3| = 0.91$, corresponding to a decay time $hat\tau d_e c_a y = -1/log(0.91) \approx 10.6$ years — well below the sub-window length of 94 years and below the binary persistence threshold's implicit decay-time floor of $3\ T_{max} = 93$ years.

For the post-war sub-window fifteen eigenvalues are returned. The leading is $|hat\lambda_1| = 0.9703$, again the invariant-measure mode but with a larger gap to unity than in the pre-war estimate. The remaining fourteen have moduli from 0.95 down to 0.30. No eigenvalue meets the threshold 0.987, including the null mode itself. Under the pre-registered criterion, **Question 1 is rejected for the post-war sub-window**. The leading non-trivial-frequency mode has $|hat\lambda_2| = 0.95$, corresponding to a decay time $hat\tau d_e c_a y \approx 19.5$ years — again well below the sub-window length of 76 years and the implicit decay-time floor of $3\ T_{max} = 75$ years.

Condition (G) of Proposition 4.3, the spectral-gap condition, is satisfied in both sub-windows on the *multiplicative* screening rule of §5.5 (factor ≥ 1.5 between consecutive log-moduli). The largest log-gap factors are 1.76 for the pre-war estimate and 1.80 for the post-war estimate, both above the pre-registered cutoff. The absolute spectral gap $\delta = |hat\lambda K| - |hat\lambda K_{+1}|$ — the quantity that enters the Lipschitz bound of Lemma 3.4 — is $\delta_{pre} \approx 0.04$ for $K=15$ in the pre-war fit and $\delta_{post} \approx 0.05$ for $K=14$ in the post-war fit. Both are positive but small in absolute terms; the perturbation control of Lemma 3.4 is therefore weak in absolute units, which §6.4 takes up under the local-normality diagnostic. The fact that the screening gap condition is satisfied while no eigenvalue clears the persistence threshold is the substantive content of the rejection: the data support a well-defined leading subspace of dimension $K = 15$ (pre-war) and $K = 14$ (post-war), but the eigenvalues populating that subspace are all transient on the time-scale of the sub-window. The leading structure exists; it is not Level-III in the operational sense the pre-registration required.

The threshold's stringency interacts with the observation window — a 76-year panel cannot distinguish a decay time of two hundred years from one of two thousand — and §8 records the resulting resolution ceiling; §6.3 returns to the distinction between Level III persistence and slowly-decaying structure.

6.3 Sub-threshold spectral structure: a post-hoc descriptive

Question 1 was framed as a binary test against a fixed threshold, but the spectra of §6.2 contain structure below that threshold that the binary framing does not exhaust. The leading subspaces identified by the gap criterion — fifteen modes pre-war, fourteen post-war — contain several oscillatory modes with moduli in the *0.55* to *0.95* range and periods that fall within the period bands pre-specified for Question 3 in §5.4. This sub-threshold structure is not a Level-III finding and does not bear on the binary verdict of §6.2. Its examination here is a post-hoc descriptive of the leading subspace, declared as such, and the criteria used to organise it are pre-registered (the three period bands of §5.4) while the threshold is not (the persistence threshold of §5.4 has been failed by every mode under discussion).

Within the post-war spectrum an oscillatory mode at period $T = 123.8$ years with modulus *0.9537* falls just above the long-cycle band *60* to *120* years, corresponding to the secular-cycle range of the cliometric literature. A mode at $T = 33.2$ years with modulus *0.829* falls within the medium-cycle band; a mode at $T = 24.6$ years with modulus *0.875* at the boundary between medium and short bands; a mode at $T = 8.9$ years with modulus *0.741* within the short cycle band and at the upper end of the documented Juglar range; and a mode at $T = 5.7$ years with modulus *0.693* at the lower edge of the short band. Of the fourteen modes in the leading post-war subspace, five fall within or at the boundary of the pre-specified period bands. The pre-war spectrum is less differentiated by this criterion: the highest-modulus oscillatory modes after the null cluster have periods of *7.15* years (modulus *0.784*) and *7.70* years (modulus *0.651*), both within or just below the short-cycle band, with no medium- or long-cycle modes resolvable within the sub-window's T_{max} of approximately *31* years.

The cross-window comparison is the more interesting analysis, and it produces a finding that the binary persistence test missed. A mode is taken to be cross-window corresponding if its period in one sub-window matches a period in the other to within twenty percent and its modulus matches to within *0.10*. Under this criterion three matches emerge, all in the band *5* to *9* years: the pre-war mode at $T = 7.15$ years, modulus *0.784*, corresponds to two post-war modes at $T = 8.92$ years, modulus *0.741*, and $T = 5.73$ years, modulus *0.693*; the pre-war mode at $T = 7.70$ years, modulus *0.651*, corresponds to the post-war mode at $T = 8.92$ years, modulus *0.741*. No cross-window matches are produced in any other pre-specified band.

The substantive content of the cross-window correspondence in the *5–9* year band requires careful statement. The matched modes do not meet the pre-registered persistence threshold and so do not constitute Level-III persistence in the operational sense the paper has committed to. What they do constitute is something weaker but not nothing: a band of frequencies in which the leading spectrum of the multi-task estimator returns modes of similar period and similar moderate modulus across a structural break of thirty-one years, two World Wars, and the dissolution and reconstruction of the European state system. The matching is not derived from a model that imposes continuity across the break — the two sub-window fits are entirely independent — and the matched moduli sit comfortably below the threshold of the persistence test, so the matches are not artefacts of the persistence criterion either. The pre-registered band structure of §5.4 makes the post-hoc framing disciplined: the bands were chosen before the data were examined, and the modes either fall within them or they do not. They fall within the Juglar-cycle band, which is the documented range of investment-cycle oscillations in the economic history literature. The surrogate analysis of §6.7 finds a correspondence of this strength in roughly 37% of calibrated null panels; the pattern is therefore not statistically distinguishable from noise, and the reading developed below is offered only as a hypothesis the data permit, not as evidence for the model.

A reading that holds the framework's pre-registered commitments while making sense of this pattern is the following. Level-III persistence as defined in §5.4 is the strongest form of the spectral-structure-persistence prediction of Conjecture 1; what the data here suggest is the existence of slowly-decaying modes in the Juglar band, modes whose decay time exceeds the resolution of either sub-window separately but is not infinite on the time-scale of the panel. The cross-window match indicates that the same dynamical regularity, attenuated by realised decay over the sub-window, recurs in the post-war regime: the period is preserved, the modulus is restored from the within-sub-window decay by re-equilibration to a new structural configuration, and the overall pattern is consistent with a band of weakly-persistent modes that the binary persistence threshold cannot detect. This reading is offered as a hypothesis the data do not contradict; the surrogate analysis of §6.7, however, finds a correspondence of equal strength in roughly 37% of null panels, so the pattern is not statistically distinguishable from noise and cannot be advanced as evidence for the framework. The disciplined test of it would require a panel with a longer continuous observation window or, equivalently, a panel that bridges multiple structural epochs and tests cross-epoch correspondence as a primary test rather than as a post-hoc descriptive — the design discussed in §7.

A formal concern that applies specifically to the cross-window matching analysis must be flagged here. The matching criterion depends on the period $T_j = 2\pi/|arg(hat\lambda_j)|$, which is determined by the *phase* of the estimated eigenvalue. Remark A.6 in Appendix A.4 establishes that the Hankel-augmented DMD estimator on noisy data produces a bias with a phase-displacement component in addition to the modulus-attenuation component characterised by Proposition 4.2, with the magnitude of the phase displacement depending on the quadratic form $\psi_j{}^t op\ S\_\tau\ \varphi_j$ for the block-shift matrix S_τ. For modes with substantial local non-normality (which §6.4 identifies as common in the leading subspace of the present panel), this phase-displacement bias can produce systematic distortions in the period estimates $hat\ T_j$ that affect the cross-window matching criterion. The §6.7 surrogate analysis absorbs this concern at first order — null panels and data panels are processed through the identical pipeline and inherit the same phase-displacement structure — but a second-order asymmetric phase warping cannot be ruled out, as Remark A.6's second paragraph develops. The matching reported here should therefore be read with the understanding that the matched periods $hat\ T_j$ are point estimates subject to a phase-displacement bias whose magnitude the present paper does not separately characterise; a controlled decomposition is part of the WP-2026-008 simulation programme.

6.4 Robustness diagnostics

Several robustness diagnostics on the primary estimates were pre-specified in §5.4. The headline diagnostics are reported here; the full numerical detail is in the replication repository (Appendix B).

The previous version defended the binary verdict by appeal to hypothesis (v) of Theorem 1 (local normality) and a Bauer–Fike bound on the eigenvalue perturbation. We withdraw that argument. The leading eigenspace is empirically far from normal — the condition number of the leading-subspace eigenvector matrix is $\kappa W = 19.2$ for the pre-war fit and *21.7* for the post-war fit, and the maximum per-mode misalignment between left- and right-eigenvectors reaches *0.89* pre-war and *0.90* post-war — and a Bauer–Fike radius scaling with this conditioning is uninformative at a threshold so near the unit circle. We therefore estimate the sampling variability of the verdict directly, by leave-one-country-out (LOO) resampling, which absorbs the non-normality rather than bounding around it.

Refitting the canonical estimator on each of the eight seven-country sub-panels yields, in the post-war window, a leading substantive (resolvable-period) mode of modulus *0.955* with a LOO standard deviation of *0.002* (range *0.952–0.958*), never approaching the threshold *0.987*. In the pre-war window the near-unit moduli that appear under resampling (*0.96–0.99*) are invariant-measure cluster members under the period-resolvability rule of §4.2a — real-positive, or carrying implied periods of order 10^3 years on a 94-year window — while the leading mode with a resolvable period lies below *0.94*. Counting, in each fold, the modes that are simultaneously at or above the persistence threshold and oscillatory with a resolvable period ($2 \leq T \leq Tw_{in}d_ow$), the count is zero in all sixteen folds across both sub-windows. The binary verdict is in this sense robust to panel composition, and the robustness is established by a sampling estimate rather than by a perturbation bound. A confidence-interval treatment of Q1 under non-normality remains outside the pre-registered design and is left to the WP-2026-007 pseudospectral refinement; hypothesis (v) is retained as a stated assumption that the verdict no longer requires.

The effective sample size underlying this fit is itself the binding constraint. With residual cross-country correlation $barρ = 0.193$ in the post-war window, the effective sample size is $N_eff = N_{nom}/(1 + barρ(Nc - 1)) = 494/2.35 \approx 210$, against an operator of rank fifteen and hence 225 free parameters: the post-war operator is under-determined ($N_eff/r^2 \approx 0.93$). The pre-war window, with $barρ = 0.044$, is adequately determined ($N_eff \approx 467$, $N_eff/r^2 \approx 1.82$). The post-war under-determination is the structural source of the null power reported in §6.2.

Examining the conditional-independence hypothesis of Proposition 4.1 requires the average pairwise cross-country residual correlation $bar\hat{ρ}$, where the residual is the one-step DMD-fit residual on the projected leading subspace. The pre-war estimate is $bar\hat{ρ} = 0.044$ with standard deviation *0.114* across the twenty-eight country pairs and individual pairwise correlations ranging from *-0.27* to *+0.22*. The post-war estimate is $bar\hat{ρ} = 0.193$ with standard deviation *0.155* and pairwise correlations ranging from *-0.17* to *+0.45*. The pre-war panel sits close to the conditional-independence regime that Proposition 4.1 assumed; the post-war panel substantively violates it, in a direction consistent with European-scale common shocks operating jointly on the panel. Translated through Proposition 4.4, the effective sample size is $N_eff = 467$ for the pre-war estimates against $N_{nominal} = 612$, and $N_eff = 210$ for the post-war estimates against $N_{nominal} = 494$ — a deflation factor of *1.31* for the pre-war fit and *2.35* for the post-war fit. Confidence statements about the post-war leading spectrum should be reported on $N_eff = 210$, which corresponds to roughly fourteen effective country-years per leading mode. The width of any error bar scales as $N_eff^{-1/2}$, so post-war precision is approximately *1.53* times worse than the nominal pooled count would suggest.[4]

Two alternative debiasing schemes were specified for comparison in §5.5. The total-least-squares variant with SVD-projection of Hemati et al. (2017) selects a much sharper truncation than the standard estimator — three modes for the pre-war sub-window and six modes for the post-war sub-window — and recovers essentially only the null-mode cluster. The pre-war fit's three retained modes consist of the invariant-measure eigenvalue at $|hatλ| = 0.9936$, its near-zero-frequency complex conjugate at $|hatλ| = 0.9936$, and a single low-modulus mode at $|hatλ| = 0.31$ with no oscillatory content; the post-war fit's six modes consist of the null-mode pair at $|hatλ| = 0.97$, two slowly-decaying real modes at $|hatλ| = 0.84$ and *0.78*, and two real modes below *0.5*, again with no oscillatory structure beyond the null cluster. This is the conservatism Proposition 4.2 hinted at: a debiasing correction that attempts to symmetrise the noise structure between predictor and response, applied to a panel where the noise structure is partly common across the panel rather than independent, ends up over-truncating and pruning the leading subspace down to the modes that survive the most aggressive shrinkage. The

[4]The N_eff values rest on Proposition 4.4's equicorrelation assumption: $barρ$ enters as a single scalar averaging pairwise correlation across countries and time. Common shocks plausibly concentrate in sub-periods (the oil shocks of the 1970s, the EMS coordination of the 1980s, the European debt crisis of the 2010s), and N_eff is a concave function of the correlation profile, so the true effective sample size under time-varying correlation would be smaller than the equicorrelation calculation states. The post-war $N_eff = 210$ should therefore be read as an upper bound; a year-clustered block-bootstrap would give a more conservative estimate but departs from the pre-registered quantity.

Hemati estimator returns no oscillatory modes beyond the null cluster in either sub-window, which is consistent with the §6.2 verdict on a stricter side: a debiasing estimator more conservative than the canonical one finds no persistent structure at all. The forward-backward variant of Dawson et al. (2016), in contrast, exhibits the overcorrection problem that Hemati et al. (2017) §V documented for noisy panel data: applied to the canonical-projected representation, it produces moduli with $|hat\lambda| > 1$ in several leading positions, $|hat\lambda|_{max} = 2.04$ for the pre-war fit and 1.63 for the post-war fit, in formal contradiction with the contraction property and indicating that the debiasing correction has overshot. The forward-backward estimator is therefore not informative on the persistence question in this setting. Its behaviour on the present panel confirms the fragility documented by Hemati et al. (2017) §V for noisy panel data and reinforces the case for using the canonical Tu et al. (2014) pipeline as the default for macrohistorical applications, a methodological recommendation §7.3 develops.

A caveat accompanies the kernel-DMD variant of Williams et al. (2015), pre-registered as a robustness check in §5.5. The implementation used a non-canonical truncation strategy — Gavish-Donoho on the Gram-matrix singular values rather than the Mercer-style feature-space truncation Williams et al. (2015) prescribe — and produced spectra with $|hat\lambda|_{max}$ on the order of 15, numerically incoherent with the contraction property. The failure should be read as a consequence of the truncation choice rather than as evidence against the kernel-DMD method itself: a canonical implementation has not been tested on this panel, and the present paper records the kernel-DMD robustness check as *not executed* rather than as *failed*. A canonical kernel-DMD test on macrohistorical panel data is deferred to WP-2026-007, where the panel extension provides a more favourable N_{eff} for kernel methods to be informative.

A cross-window comparison was reported as a substantive finding in §6.3. The comparison also bears on the identifiability conditions of Proposition 4.3 rather than on the stability of the multi-task estimate per se: identifiability requires (T), that the usable per-country series length exceeds three times the period of the slowest mode resolvable in the window. The post-war modes at $T = 33.2$ years and $T = 24.6$ years have no pre-war counterparts within the matching tolerance. For the pre-war sub-window, $T_{max} \approx 31$ years per condition (T), which means that a hypothetical mode at $T = 33.2$ years would have been formally unresolvable in the pre-war fit even if present in the underlying dynamics. The absence of a pre-war counterpart is therefore consistent both with the post-war modes being genuinely post-war features of the panel's dynamics and with their being modes of the underlying dynamics that the pre-war window was structurally unable to detect. Discrimination between these two possibilities requires a longer pre-war sub-window — the panel extension of §7.2 — and is not within the present test's scope.

6.5 Empirical confirmation of the attenuation bias

Proposition 4.2 of §4.4 stated that finite-sample DMD on noisy data exhibits a systematic attenuation bias: estimated eigenvalues have moduli displaced inward from the unit circle relative to their true values, with the magnitude of the attenuation increasing with measurement-noise variance and decreasing with effective sample size. The proposition was proved in Appendix A.4 in operator-norm form and characterised in §4.4 as conservative with respect to the paper's central thesis — a bias that understates persistence is friendly to positive findings and inimical to negative ones.

An empirical test of Proposition 4.2 is provided by the leading eigenvalue of the canonical DMD estimator that the paper did not pre-register but that the data make available. The mode in question is the invariant-measure mode, whose true eigenvalue is exactly unity: it is the unit eigenvalue of $\mathcal{K}\_\theta$ guaranteed by Axiom A2 and the existence of μ_θ, with the constant function as its eigenfunction. Unlike every other mode in the spectrum, this one has a known true value, which makes it a natural benchmark for the bias correction.

Estimated values of $|hat\lambda_{null}|$ are 0.9936 for the pre-war sub-window and 0.9703 for the post-war sub-window. The displacement from unity is 0.0064 in the pre-war estimate and 0.0297 in the post-war estimate —

a roughly five-fold difference between the two sub-windows. The per-mode form of Proposition 4.2 specialised to the invariant-measure mode predicts $\mathbb{E}[|hat\lambda_{null}|] - 1 = -\sigma\_\varepsilon^2 \cdot \psi_{null}{}^{t}op\ CXX^{-1}\ \varphi_{null} / N_{eff}$, with φ_{null} and ψ_{null} both proportional to the constant function. The relevant effective sample sizes are $N_{eff} = 467$ for the pre-war fit and $N_{eff} = 210$ for the post-war fit, a ratio of approximately *2.2*. The observed attenuation ratio of approximately *4.6* exceeds the ratio of effective sample sizes by a factor of two. The discrepancy is consistent with the noise structure differing between the two sub-windows — the post-war panel has both lower N_{eff} and higher residual variance per dimension after standardisation — but the paper does not claim that this attribution is itself an empirical test: the residual noise variance σ_ε^2 was not estimated independently ex ante, and the quadratic form $\psi_{null}{}^{t}op\ CXX^{-1}\ \varphi_{null}$ was not computed from data. The leading-order *direction* of Proposition 4.2 — inward displacement from the true value with magnitude scaling inversely with N_{eff} and directly with σ_ε^2 — is consistent with the empirical pattern; the *magnitude* is consistent within an order of magnitude but a calibrated quantitative test would require quantifying σ_ε^2 and the quadratic form per sub-window.

This is a benchmark observation rather than a controlled test. Proposition 4.2 holds for any specific mode under its hypotheses; specialising it to the invariant-measure mode yields a per-mode prediction whose constants depend on quantities not estimated in the present paper, and the constants in the proof were given to leading order rather than in calibrated form. But the qualitative direction is confirmed by an eigenvalue whose true value the framework's axiomatic structure fixes a priori, and this is the kind of internal-consistency check the framework permits but does not promise. It is reported here as an unanticipated directional confirmation of a theoretical prediction, with appropriate caveats about the absence of pre-registration on this specific check and the absence of magnitude calibration. A future paper that uses Proposition 4.2 to correct estimated moduli upward — applying a bias correction calibrated on simulated data with known generative spectra, with appropriate uncertainty propagation — could test whether the corrected sub-threshold modes lift over the persistence threshold; this is the most direct route from the present null to a potential positive finding that the paper's own results suggest.

6.6 Questions 2 and 3: formally precluded, informatively reported

Questions 2 and 3 are formally precluded by the negative Q1 verdict; we report the joint-model fit as descriptive diagnostics only. The 15% reduction in one-step predictive RMSE for the pooled model relative to country-specific fits (*0.247* joint against *0.291* independent post-war; *0.211* against *0.224* pre-war) is consistent with mere similarity of the national dynamics and does not, in the absence of a positive Q1, bear on the identity-of-modes claim of Theorem 1. It should not be read as confirmatory.

For Question 3, the band matching of §6.3 already addresses the substantive content. The five post-war modes whose periods fall within or at the boundary of the pre-specified bands constitute the informative version of the Question 3 analysis, with the caveat that the modes in question are sub-threshold and the historical-correspondence claim of §5.4 was conditional on the persistence threshold being met.

6.7 Surrogate analysis and power calculation

The Juglar-band correspondence that recurs across the inter-war break must be read against its null distribution. A surrogate analysis on calibrated null panels finds a correspondence of comparable strength in approximately 37% of replicates. The observed correspondence is therefore not statistically distinguishable from noise, and we do not interpret it as evidence for the model. We record it as a sub-threshold observation only; these results are placed here among the primary results of §6 rather than as a robustness afterthought, because they govern the inferential weight that can be placed on the §6.3 finding.

The surrogate panels were generated as follows. Each country trajectory $X(^c)_t \in \mathbb{R}^5$ for $t = 1, \ldots, T$ (with $T = 94$ pre-war, $T = 76$ post-war) was drawn from a per-variable AR(1) plus common-shock process: for each variable $j \in \{1, \ldots, 5\}$ and time t,

$$X(^c)_{j,t} = \rho_j X(^c)_{j,t-1} + \sqrt{(1 - overline\rho)W} \cdot \varepsilon(^c)_{j,t} + \sqrt{(overline\rho)W} \cdot \eta_{j,t},$$

with ρ_j the per-variable AR(1) coefficient ($\rho_1 = 0.10$ for log-differenced population, $\rho_2 = 0.15$ for log-differenced GDP, $\rho_3 = \rho_4 = 0.85$ for V-Dem libdem and polyarchy in levels, $\rho_5 = 0.90$ for CINC); $\varepsilon(^c)_{j,t} \sim \mathcal{N}(0,1)$ are country-specific idiosyncratic innovations; and $\eta_{j,t} \sim \mathcal{N}(0,1)$ are common-shock innovations shared across all countries. The mixing weight $\bar{\rho}W$ is set to the observed cross-country residual correlation ($\bar{\rho}W = 0.044$ pre-war, $\bar{\rho}W = 0.193$ post-war). The canonical DMD pipeline and the cross-window matching procedure (Appendix B) are applied with effective rank $r = 15$ on each of $N_{sim} = 1000$ surrogate panel pairs under the pre-registered tolerances.

Under this null, *54.3%* of surrogate panel pairs produce at least one cross-window match in the 5-to-9 year period band; the joint event "at least one match in the band AND best matched $|hat\lambda| \geq 0.78$" occurs in *37.4%* of surrogate panels. The cross-window correspondence reported in §6.3 is not unusual under this null. The conservative reading — that the match is broadly consistent with what panels of this complexity produce by chance — has the stronger claim on the empirical record. The framework-consistent reading of §7.1 remains possible, but is demoted by this analysis to the less probable of the two readings. The *37%* figure is a point estimate under one parametric null; a phase-randomised variant is identified as a robustness check for WP-2026-007's extended panel programme.

The power calculation for Q1 is reported in the same section because it qualifies the binary verdict in the same way. The §6.5 invariant-measure mode attenuation provides a calibration of the noise scale, since the true value of that mode is fixed by Axiom A2 at exactly unity. The calibrated noise scale gives $2 \cdot SE = 0.0128$ pre-war and *0.0594* post-war, both exceeding the threshold's distance from unity ($1/(3\ T_{max}) = 0.0106$ pre-war and *0.0132* post-war). The Q1 test has effectively zero power against any alternative hypothesis of finite-modulus persistence above the threshold. This is a limitation of the pre-registered design — the absence of a pre-registered power analysis is itself a limitation, as §8 records — and it means the negative Q1 verdict, taken alone, is a statement about detectability, not about existence.

To translate this into the §7.2 panel-extension programme: detection of a mode with decay time $\tau d_e c_a y$ at 95% confidence requires $N_{eff} \geq \sigma\_\varepsilon^2 \cdot \psi^t op\ CXX^{-1}\ \varphi\ /\ [(1/(3\ T_{max}) - 1/\tau d_e c_a y)\ /\ 2]$. For the pre-war sub-window, detection of a mode with decay time 200 years requires $N_{eff} \geq 1060$, a *2.3*-fold increase over the current $N_{eff} = 467$; for the post-war sub-window, the same target requires $N_{eff} \geq 1530$, a *7.3*-fold increase over the current *210*. The panel-extension targets of §7.2 — twenty to twenty-five countries — bring the post-war N_{eff} to roughly *1000* on plausible cross-country residual-correlation assumptions, and the pre-war sub-window to several thousand, which is the regime in which the Q1 test acquires meaningful power. The negative Q1 verdict reported here is therefore not evidence against Conjecture 1's substantive content; it is, in large part, a statement about the power of the test design at the panel size and noise structure the present test had available. The §7.2 extensions are not contingencies against the rejection but the natural next step that the power analysis identifies as the precondition for a power-adequate test.

Three caveats on the power calculation should be stated explicitly. First, the noise scale $\sigma\_\varepsilon^2 \cdot \psi^t op\ CXX^{-1}\ \varphi$ is calibrated from the observed attenuation of the invariant-measure mode rather than estimated from independent pilot data; the sampling variability of that calibration is not propagated into the standard error on $|hat\lambda|$. This is a plug-in calibration on the same data the test examines, and it understates uncertainty in the same direction that a post-hoc analysis tends to: the required N_{eff} values above should therefore be read as *lower bounds* on what an adequately powered test would need, not as point estimates. A proper power analysis would draw the noise calibration from independent pilot data or from a simulation study calibrated to the documented measurement-error properties of V-Dem, Maddison, and COW, and would return values potentially substantially larger than those reported here.

Second, the calibration is built on the invariant-measure mode, whose right- and left-eigenvectors are the constant function and therefore have unit alignment by construction; the local-normality inflation of §6.4 (worst-mode misalignment up to *0.90*) does not apply to the null mode itself. The Level III modes that the Q1 test would identify, if they existed, would not enjoy that protection: for any leading mode whose right-left misalignment is close to the worst observed, the effective standard error on $|hat\lambda_j|$ would be inflated by approximately $1/|\langle u_j, v_j \rangle|$, with the consequence that the required N_{eff} for detection of that mode would scale roughly quadratically in the inflation factor. The power calculation above is therefore optimistic for the modes the test was designed to detect, in addition to being optimistic by construction; the actual power against Level III alternatives is plausibly lower than the present calculation states by a margin that the present panel cannot quantify but that the WP-2026-007 pseudospectral refinement of Lemma 3.4 will make precise.

Third, the noise model assumed by Proposition 4.2 — additive iid Gaussian noise — does not hold for the §5 panel, as Remark A.5 noted. Departures from iid Gaussian noise affect the leading-order constants of the bias formula and therefore the calibration above; the direction of the effect is not fixed by the proof and would require either a controlled simulation study or independent noise characterisation to determine. The qualitative content of the analysis — that the precision floor exceeds the threshold distance and that detection of plausibly persistent modes requires N_{eff} substantially above the present panel — is robust to these three caveats, but the specific numerical thresholds are not, and the reader should treat them accordingly.

Both post-hoc analyses are recorded in the replication archive (Appendix B) with the scripts that produced them, so the reader can verify the simulation parameters and rerun the calculations. They do not displace the §6.2 verdict within its pre-registered terms; they qualify its inferential reach beyond those terms. The qualification points in the same direction as the within-pre-registered findings of §6.4 on N_{eff} deflation and hypothesis (v) non-normality: the test, as constructed, is more conservative than its binary verdict alone makes evident.

Discussion

7.1 The sub-threshold pattern and what it suggests about the operator

Two readings of the §6.3 cross-window correspondence in the Juglar band were available: a framework-consistent reading, on which the matched modes are slowly-decaying structure whose decay times fall in the resolution gap that §8 identifies, and a conservative reading, on which the match is a coincidence of independently fitted local minima. The post-hoc surrogate test reported in §6.7 substantially shifts the balance. Under a calibrated null with no planted persistent oscillation, panels of comparable effective rank and per-variable persistence structure produce cross-window matches of comparable strength roughly one time in three. The §6.3 finding is therefore not unusual under that null, and the conservative reading — that the match is broadly consistent with what panels of this complexity produce by chance — has the stronger claim on the empirical record. The framework-consistent reading remains *possible* on the present evidence, but the surrogate analysis demotes it from "the most natural reading" to "one of two readings, the less probable on present evidence". What the pattern does still establish, regardless of which reading is correct, is that the multi-task estimator on this panel returns a sub-threshold spectral feature in the Juglar band that the pre-registered tolerances of §5.4 register as cross-window stable. The substantive content has shifted: if the conservative reading holds — and the surrogate test gives weight to it — this is a feature of the estimator and a methodological lesson, and the null-distribution validation flagged by the present analysis should be added to the standard robustness suite for spectral-correspondence claims on historical panels in any subsequent paper. If the framework-consistent reading nonetheless holds, it does so on weaker evidence than §6.3 considered alone would suggest, and the design refinements of §7.2 are the appropriate place to translate it into a stronger test against a proper null.

7.2 The programme of empirical extension

A test that rejects its pre-registered question and produces sub-threshold structure in a documented band of frequencies is also a test whose design has been informed about how to do better. The present test is one observation; the programme of which it forms a part has several routes forward, each of which the §6 results have helped to identify with more precision than they could be specified before the test was run. Five extension directions are worth describing in turn, in approximate order of expected yield-to-effort.

Panel extension is the most direct route. The Western European panel of eight countries on 1820–1913 and 1945–2020 produced an effective sample size of 467 for the pre-war fit and 210 for the post-war fit (as obtained from the full Appendix A.5 derivation of Proposition 4.4, including within-country mixing-rate corrections), on a Hankel-augmented state space of dimension 40 truncated to effective rank 15 or 16. Extending the panel to roughly twenty to twenty-five countries — adding Central and Eastern Europe with explicit treatment of the 1917–1991 Soviet break, North America, parts of East Asia, and selected Latin American countries — would on plausible assumptions about cross-country residual correlation push the effective sample size above 1,000 for the post-war fit and well above that for a pre-war fit extended back to 1800 or earlier on the GDP and population dimensions where Maddison coverage permits. The Gavish-Donoho effective rank scales sub-linearly in N, so the truncation would not increase proportionally, and the rank-to-sample-size ratio would improve from approximately 14:1 on the present panel to perhaps 60:1 or better. At that resolution the persistence threshold of §5.4 becomes feasible to clear for modes that on the present panel sit a few percentage points below it.

A complementary extension widens the sub-window. The pre-war sub-window of the present test is 94 years; Maddison and V-Dem coverage permits 1700–1913 for the subset of countries with longer documentary records (Britain, the Netherlands, France, Sweden), at the cost of significantly higher imputation density for the earliest decades and reduced coverage uniformity. A panel-balanced extension to 1750–1913 — 164 years — would substantially loosen the resolution condition (T) of Proposition 4.3 and would in particular bring Kondratiev-band modes within the resolution limit, which on the present panel they are not for the pre-war sub-window. Whether the trade-off between window length and imputation density is favourable requires a separate analysis the paper does not undertake, but it is the next most plausible extension after country count.

Dimensional enrichment is a third extension: a complementary test on the 1990–2025 sub-window with text-analytic and survey-based Ψ variables would address the dimensional projection limit recorded in §8 at the cost of forgoing the long-window analysis. The two designs answer different questions and belong jointly in the programme.

Methodological refinement is a fourth extension. The present test used the canonical Tu et al. (2014) pipeline as primary estimator and the SVD-projected TLS-DMD of Hemati et al. (2017) and the forward-backward DMD of Dawson et al. (2016) as alternatives. §6.4 and §8 documented that the latter two behave poorly on this panel — TLS-DMD is over-conservative, forward-backward overcorrects beyond the unit circle. A dedicated methodological study that simulates panels from known generative spectra and compares the canonical pipeline against optimised DMD (Askham and Kutz 2018, §5), sparsity-promoting DMD (Jovanović et al. 2014), and other recent variants would identify the estimator best suited to macrohistorical panel data and would in particular adjudicate whether the §6.5 attenuation pattern can be corrected at the level of individual modes rather than only in operator-norm aggregate. This is a candidate WP-2026-008.

Theoretical refinement is the fifth extension. The local-normality hypothesis (v) of Theorem 1 was at the edge of empirical violation on the present panel and the conditional-independence hypothesis of Proposition 4.1 was substantively violated post-war (§6.4, §8). A refined Theorem 1' that replaces hypothesis (v) with a pseudospectral gap condition and that weakens Proposition 4.1's conditional independence to a quantitative tolerance on *barρ* — both achievable within the framework's mathematical apparatus — would tighten the framework's interpretive coverage of the kind of panel the present test examined. The refinements belong in the CMH monograph rather than in a working paper, but the present empirical results are the most direct motivation the framework has yet had to formulate them.

These five extensions are not independent. Panel extension is the foundation; sub-window extension complements it; the Ψ extension answers a different question that the long-window programme will not; the methodological study informs all of the above; the theoretical refinement integrates the lessons of all four. The next paper in the programme, tentatively WP-2026-007, would conduct the panel extension and the sub-window extension jointly. The order of the others is more flexible; what matters is that each has been identified concretely enough by the present test for its specification to begin.

For macrohistorical panel data the canonical Tu et al. (2014) DMD pipeline with Gavish–Donoho rank truncation is recommended in preference to TLS-DMD and forward–backward variants, and residual cross-country correlation must be measured and reported because it controls the effective sample size. On the present design a credible test of Conjecture 1 requires that the effective sample size exceed the operator's parameter count with margin: targeting $N_eff/r^2 \approx 3$ implies N_eff of order 675, and with margin of order one thousand, against the 210 available post-war. This is the concrete condition a future panel extension must meet.

> **Remark 7.1 (Methodological lessons).** Four lessons from the execution feed the framework's iterative refinement. The Hankel delay $\tau = 7$ proved over-conservative: a Gavish-Donoho effective rank of 15 to 16 against an augmented dimension of 40 indicates that τ in the 4-to-5 range should be the default, with $\tau = 7$ reserved for the longest-cycle bands. The persistence threshold should be stratified by period band in adequately powered designs, since a long-period mode can be Level III in the framework-relevant sense yet fall below a threshold calibrated for short-period modes. Multi-task pooling outperformed independent country fits by 15 percent in one-step RMSE even under the post-war dependence violation; pooling gains and effective-sample-size deflation coexist, and the framework's exposition should make this dual character explicit. Finally, pre-registrations should name the DMD variant explicitly (Tu canonical, Askham-Kutz optimised, Jovanović sparsity-promoting) and report *barρ* alongside the headline estimates.

7.3 Relations to neighbouring literatures

WP-2026-005 §§1–2 positioned the CMH framework against neighbouring literatures; the present results enter three specific conversations.

Cliodynamics is the first. The empirical test reported here did not produce evidence for the dominant structural cycles that Cliodynamics central works identify — the secular-trend Demographic-Structural cycle of approximately 200 years (Turchin and Nefedov) or the fathers-and-sons generational cycle of 50 years (Turchin 2018). The post-war fit's long-period mode at 124 years falls between these and clears neither band; the medium-period modes at 24 and 33 years are below the generational-cycle range. The pre-war fit produces no medium- or long-period modes within the sub-window's resolution limit. A reading consistent with the Cliodynamics framing would attribute this to the present test's state vector — D-E-P-S without disaggregation of inequality and elite-mobilisation variables that Cliodynamics central works treat as load-bearing — and would predict that a state vector enriched along those lines might recover the relevant cycles. The CMH framework and the Cliodynamics framework are not adjudicated on this panel, and the present results do not constitute evidence against the Cliodynamics findings as such; they constitute one negative test that the framework can offer as part of its developing differentiation from the alternative programme.

A second conversation is with applied DMD literature. The §6.4 finding that TLS-DMD of Hemati et al. (2017) is over-conservative on this panel and that forward-backward DMD of Dawson et al. (2016) overcorrects is consistent with cautionary observations in the applied dynamical-systems community (e.g., Sashidhar and Kutz 2022, who flagged the overcorrection problem in social-data applications). The present paper adds an explicit empirical case in macrohistorical panel data and recommends the canonical Tu et al. pipeline with Gavish-Donoho truncation as the recommended default for similar applications. The methodological study flagged for WP-2026-008 in §7.2 would convert this recommendation into a systematic comparison.

A third conversation is with economic-history work on business cycles. The Juglar cycle (5–11 years in classical formulations) is the most-studied medium-term oscillation in industrial-era economic history, and its persistence across two-century records has been debated since Schumpeter (1939). The present test recovers sub-threshold spectral structure within this band that survives the inter-war break, but it does so on a state space that includes only one explicitly economic variable (GDP per capita), with the remaining structure carried by demographic, political, and military-capacity variables that the classical Juglar literature does not treat as central. A reading consistent with this is that the multi-task estimator detects the Juglar-band signature as a shared oscillation in the joint dynamics of the four-dimensional state, including but not driven by GDP. Whether this interpretation holds requires variable-by-variable decomposition of the mode shapes, which the present test does not perform and which is a candidate diagnostic for the next paper in the programme.

7.4 Synthesis

The pre-registered test rejected. The sub-threshold Juglar correspondence recurred but is not statistically distinguishable from noise. The attenuation bias was confirmed in direction on the invariant-measure mode. The programme of extension is identified in §7.2. The present paper records one observation; the programme continues.

Limitations

Three categories of limitation constrain the inferential reach of the §6 results. They are concurrent, not alternative: a reader who finds the headline verdict less authoritative for any one of them is reading the paper correctly.

Design limits. The state vector covers four of the five CMH dimensions (D, E, P, S) but not Ψ (psychological-collective), which is unavailable pre-1990 in panel-comparable form; any Level III mode with substantial Ψ-projection is invisible to the present test. Each dimension is represented by a single variable — population, GDP per capita, two V-Dem indices, CINC — reducing sub-dimensional resolution. Both political variables come from V-Dem, so the political-dimension claims inherit any systematic V-Dem coding bias. The S dimension terminates with COW NMC v6.0 at 2016, excluding the panel's final four years. These design limits are programmatic: they reflect what two-century coverage and panel uniformity jointly permit in 2026, not methodological failures.

Theoretical scope conditions. Theorem 1's local-normality hypothesis (v) is at the edge of empirical violation in both sub-windows: the maximum per-mode misalignment between right- and left-eigenvectors is 0.89–0.90, approximately nine times the gap-only Lipschitz bound, and a pseudospectral treatment of Lemma 3.4 is deferred to WP-2026-007. The conditional-independence hypothesis of Proposition 4.1 is substantively violated post-war (*barρ* ≈ *0.193*, deflating N_{eff} from 494 to 210). The Q1 binary verdict is robust to both violations — it rests on point-estimates of moduli, not confidence intervals — but any confidence-interval treatment of the leading subspace would require the §7.2 refinements first. The sub-window length imposes a resolution ceiling: a mode with decay time between the sub-window length and three times that length is below the §5.4 threshold regardless of its existence in the dynamics, and the Q1 rejection does not extend to that range.

Estimator limits. The pre-registration committed to "multi-task Hankel-DMD" without naming the variant; two unprojected-ridge implementations were developed before the canonical Tu et al. (2014) pipeline was adopted, each producing spectra that violated the contraction property and were therefore discarded (§6.1, Appendix B). The cross-validation criterion shifted from Gaussian log-likelihood (pre-registered) to predictive RMSE (production) because the log-likelihood criterion proved monotone in the regularisation parameter on the present panel; the deviation is a pre-registration limitation. The kernel-DMD robustness check was not executed because the non-canonical truncation strategy produced a numerically incoherent spectrum; it is deferred to WP-2026-007. No pre-registered power analysis was conducted: the §6.7 post-hoc calculation is not a substitute, and future CMH pre-registrations should include a power calculation as a fixed component against a stylised null calibrated from pilot data.

The limitations compound. Future papers in the programme address each in turn — §7.2 specifies the extensions — but the present paper operates within all of them simultaneously.

Conclusion

WP-2026-006 delivers what WP-2026-005 announced: a proof that shared ergodic class implies shared Koopman spectrum (Theorem 1, conditional on A2, A3, A4, and Assumption 3.1 jointly), and an empirical test of that theorem's substantive antecedent on a pre-registered Western European panel.

The test rejected. Under the persistence threshold of §5.4, no mode beyond the invariant-measure cluster was detectable in either sub-window. The post-hoc power analysis of §6.7 finds effectively zero power against finite-modulus alternatives at the available effective sample size: the rejection is a verdict on detectability, not on existence, and the broader claim of Conjecture 1 is not refuted by a test that cannot have distinguished the alternative from the null.

Three substantive findings accompany the verdict. The invariant-measure mode is recovered with the directional attenuation Proposition 4.2 predicts — the first piece of directional empirical support for a quantitative framework claim. The sub-threshold spectral structure in the Juglar band recurs across the inter-war break but is not statistically distinguishable from noise under the §6.7 surrogate analysis. The post-war conditional-independence violation (*barρ* ≈ *0.193*) reduces the effective sample size to 210 and establishes $N_{eff} \approx 1{,}000$ as the concrete target for a test with meaningful power.

The methodological yield is independent of the verdict: canonical Tu et al. (2014) DMD with Gavish-Donoho truncation is the recommended default for macrohistorical panel data; TLS-DMD and forward-backward DMD as currently formulated are not.

The programme continues with the extensions specified in §7.2. The present paper records one observation, pre-registered, executed, and disclosed in its actual direction. That sequence is the operational content of Axiom A8.

APPENDIX A — Proofs and technical details

A.1 Notation and standing conventions

Throughout this appendix, $(\Omega, \mathscr{F}, \mathbb{P})$ denotes the underlying probability space, and the CMH state space is the measurable space $(\mathcal{X}, \mathscr{B})$ with $\mathcal{X} \subseteq \mathbb{R}^n$, n the dimension of the (observed) state vector. For a country c in a structural regime θ, the discrete-time Markov process $\{X_t(^c)\}_t \in \mathbb{Z}$ takes values in $\mathcal{X}$ and is governed by the transition kernel $P\_\theta : \mathcal{X} \times \mathscr{B} \to [0,1]$ obtained from the one-year integration of the A4 dynamics under the Itô convention (§2.1). The invariant probability measure of P_θ, when it exists and is unique, is denoted μ_θ, and $L^2(\mu\_\theta)$ denotes the Hilbert space of μ_θ-square-integrable real-valued observables on $\mathcal{X}$, with inner product $\langle f, g \rangle = \int fg \ \ d\mu\_\theta$ and norm $\|\cdot\|$. The stochastic Koopman operator is

$$(\mathcal{K}\_\theta\, g)(x) = \mathbb{E}[g(X_1) \mid X_0 = x] = \int g(y)\, P\_\theta(x, dy), \quad g \in L^2(\mu\_\theta).$$

All inequalities and convergence statements involving operators are understood in the appropriate norm, specified locally.

For ease of reference, the principal symbols used throughout the paper are collected in Table A.1.

Table A.1 — Principal notation.

Symbol	Meaning	Domain
$\mathcal{X}$	CMH state space	subset of $\mathbb{R}^n$, dimension $n=5$ for the §5 panel
θ	Structural parameter vector indexing dynamical regime	finite-dimensional
μ_θ	Invariant probability measure under structural regime θ	σ-finite, finite
P_θ	One-step transition kernel under regime θ	Markov kernel on $\mathcal{X}$
$\mathcal{K}\_\theta$	Stochastic Koopman operator under regime θ	bounded on $L^2(\mu\_\theta)$
λ_j	j-th eigenvalue of $\mathcal{K}\_\theta$ (sorted by $\lvert\lambda_j\rvert$)	complex, $\lvert\lambda_j\rvert \leq 1$
φ_j, ψ_j	right- and left-eigenvectors of λ_j, normalised by $\psi_j{}^t op\ \varphi_j = 1$	in $L^2(\mu\_\theta)$
K	Number of leading eigenvalues retained	positive integer, set by spectral gap
δ	Absolute spectral gap: $\lvert\lambda K\rvert$ - $\lvert\lambda K_{+1}\rvert$	non-negative
T	Per-country observation series length (calendar years)	$T = 94$ pre-war, $T = 76$ post-war
T_{max}	Longest period resolvable from sub-window, set to $T/3$ in the pre-registered Q1 criterion	calendar years
T_j	Period of mode j: $T_j = 2\pi / \lvert arg(\lambda_j)\rvert$	calendar years
$\tau d_e c_a y,\ _j$	Decay time of mode j: $-1/log\lVert\lambda_j\rVert$	calendar years
τ	Hankel embedding delay (pre-registered: $\tau = 7$)	positive integer

Nc	Number of countries in the panel (*Nc = 8*)	positive integer
N	Pooled column count of the Hankel-augmented predictor matrix	positive integer (*N = 612* pre-war, *494* post-war)
N_{eff}	Effective sample size under panel dependence (Prop 4.4)	positive (N_{eff} = *467* pre-war, *210* post-war)
$\bar{\rho}$	Average pairwise cross-country residual correlation	in *[-1, 1]*
overline$\hat{\rho}$	Sample estimate of $\bar{\rho}$ from multi-task fit residuals	reported in §6.4
κW	Condition number of the matrix of leading right-eigenvectors	≥ *1*, reported in §6.4
σ_ε²	Variance of additive measurement noise on the observed state	non-negative
CXX	Population covariance matrix of the Hankel-augmented state	symmetric positive-definite
$\hat{\cdot}$	Sample-estimate counterpart of any population quantity	matches domain

The persistence threshold is $|hat\lambda| \geq 1 - 1/(3 \cdot T_{max})$ with T_{max} the longest resolvable period ($T_{max} = T_{usa}b_{le}/3$, Proposition 4.3). We commit to this single form. The values are *0.989* (pre-war, $T_{max} \approx 31.3$) and *0.987* (post-war, $T_{max} \approx 25.3$); the production-code rounding $1 - 1/93 \approx 0.989$ differs from the exact $1 - 1/(3 \cdot 31.3) = 0.9894$ by $\approx 4 \times 10^{-4}$ and the embedding-corrected usable length is used throughout. No verdict depends on the rounding or on the raw-versus-corrected choice: the substantive resolvable-period modes (≤ *0.94* pre-war, *0.955* post-war) lie below the threshold under every convention.

A note on what is and is not proved here. The propositions and lemmas reproduced below are those of §3 and §4. The proofs are complete in the sense that no step is asserted without justification, but several steps invoke standard results from ergodic theory, spectral perturbation theory, and time-series econometrics by citation rather than by derivation. Where a citation is given the reader should consult the source for the underlying proof; where no citation is given the argument is self-contained. The intent is to make every step inspectable, not to reproduce textbook material.

A.2 Auxiliary results

Two auxiliary lemmas are used repeatedly and are stated first.

> **Lemma A.1 (Contraction on $L^2(\mu\_\theta)$).** *The stochastic Koopman operator $\mathcal{K}\_\theta$ is a contraction on $L^2(\mu\_\theta)$: for every $g \in L^2(\mu\_\theta)$, $\|\mathcal{K}\_\theta\, g\| \leq \|g\|$.*
>
> *Proof.* By the conditional Jensen inequality applied to the convex function *x mapsto x²*,

$$[\mathbb{E}(g(X_1) \mid X_0 = x)]^2 \leq \mathbb{E}[g(X_1)^2 \mid X_0 = x] \quad \mu\_\theta\text{-a.e.}$$

Integrating both sides against μ_θ and using the invariance of μ_θ under P_θ — namely $\int \mathbb{E}[g(X_1)^2 \mid X_0 = x]\, d\mu\_\theta(x) = \int g^2\, d\mu\_\theta$ — yields $\|\mathcal{K}\_\theta\, g\|^2 \leq \|g\|^2$.

A direct corollary is that the spectrum of $\mathcal{K}\_\theta$ is contained in the closed unit disc of $\mathbb{C}$. This is the result invoked in §2.3 to justify confining the spectral analysis to $\{|\lambda| \leq 1\}$.

Lemma A.2 (Uniqueness of the invariant measure under weak mixing). *If P_θ admits an invariant probability measure μ_θ and is weakly mixing with respect to it, then μ_θ is the unique invariant probability measure for P_θ.*

Proof. Standard. Weak mixing implies ergodicity, and an ergodic Markov kernel admits at most one invariant probability measure: any invariant measure that is absolutely continuous with respect to μ_θ must equal μ_θ by ergodicity, and any invariant measure that is singular with respect to μ_θ contradicts the weak-mixing-induced decay of correlations. See Walters (1982, Theorem 1.6) or Eisner et al. (2015, §4.4).

A.3 Proofs of §3 lemmas and Theorem 1

Lemma 3.1 (Invariant measure under A2). *Under hypothesis (i) of Theorem 1, each country possesses a unique invariant measure μ_θ, and membership in a common ergodic class is equivalent, by Definition 3.1, to equality of these measures.*

Proof. Existence is the content of Axiom A2 as formalised in §2.2: under the regime-stability conditions of A2, the invariant probability measure μ_θ exists. Uniqueness is Lemma A.2 applied under hypothesis (iii). The equivalence claim is Definition 3.1.

Lemma 3.2a (Shared kernel from A3, A4, and Assumption 3.1). *Under hypothesis (ii) — Axioms A3 and A4 applying to the panel, together with Assumption 3.1 — two countries with $\theta(^c) = \theta(^{c\prime})$ satisfy $P\theta(^c) = P\theta(^{c\prime})$.*

Proof. By Axiom A4, the continuous-time dynamics governing country c are $dX(^c)/dt = F(X(^c), \theta(^c), \varepsilon(^c)(t))$ with F locally Lipschitz in the first argument. Integrating these dynamics over the unit interval $[t, t+1]$ under the Itô convention yields a stochastic flow $\Phi(^c)_{t,t+1}$ whose law, conditional on $X_t(^c) = x$, defines the one-step transition kernel

$$P\theta(^c)(x, A) = \mathbb{P}(\Phi(^c)_{t,t+1}(x;\, \varepsilon(^c)) \in A \mid X_t(^c) = x), \quad A \in \mathscr{B}.$$

The flow $\Phi(^c)$ is a measurable functional of three inputs: the generating function F (common to all countries by A4), the structural parameter $\theta(^c)$, and the noise sample path $\varepsilon(^c)$ over $[t, t+1]$. Its conditional law is therefore a functional of F, $\theta(^c)$, and the *law* of $\varepsilon(^c)$, which we denote $\mathscr{L}\_\varepsilon(^c)$:

$$P\theta(^c) = \Pi(F,\, \theta(^c),\, \mathscr{L}\_\varepsilon(^c))$$

for some measurable functional Π. Axiom A3, read in the closure-form (A3-strong) introduced in §3.1, asserts that the macrohistorical dynamics are generated by, and only by, the causal structure among structural variables; no further primitive dynamical input is admitted. In the present notation A3-strong amounts to the statement that the functional Π has no fourth, country-specific argument — there is no $\xi(^c)$ that distinguishes the dynamics of two countries with identical F, θ, and $\mathscr{L}\_\varepsilon$. By Assumption 3.1, $\mathscr{L}\_\varepsilon(^c) = G(\theta(^c))$ for a common measurable functional G, so $\mathscr{L}\_\varepsilon(^c) = \mathscr{L}\_\varepsilon(^{c\prime})$ whenever $\theta(^c) = \theta(^{c\prime})$. Substituting,

$$P\theta(^c) = \Pi(F,\, \theta(^c),\, \mathscr{L}\_\varepsilon) = \Pi(F,\, \theta(^{c\prime}),\, \mathscr{L}\_\varepsilon) = P\theta(^{c\prime}),$$

which is the SKS condition of Definition 3.2.

A note on the equivalence qualification in Definition 3.2. The equality $P\theta(^c) = P\theta(^{c\prime})$ holds as identity of measurable functionals on $\mathscr{X} \times \mathscr{B}$ under the canonical labelling of the state space. The qualification "up to equivalence" absorbs the freedom to relabel state coordinates by a measurable isomorphism, which does not affect the spectral structure of the operator built from the kernel.

Lemma 3.2b (Shared invariant measure, consistency). *Under the hypotheses of Theorem 1, a shared transition kernel implies a shared invariant measure.*

Proof. If $P\theta(^c) = P\theta(^{c\prime})$, then any measure invariant under the first is invariant under the second. By Lemma A.2, each kernel has a unique invariant probability measure, so $\mu\theta(^c) = \mu\theta(^{c\prime})$.

Lemma 3.3 (Shared kernel implies shared operator and spectrum). *If $P\theta(^c) = P\theta(^{c'})$, then $\mathcal{K}\theta(^c) = \mathcal{K}\theta(^{c'})$ as operators on the common space $L^2(\mu\_\theta)$, with identical spectrum.*

Proof. The function space is common by Lemma 3.2b. On that common space, $(\mathcal{K}\_\theta\ g)(x) = \int g(y)\ \ P\_\theta(x, dy)$ depends on the kernel alone, so identical kernels define identical operators. Identical operators have identical spectra and identical eigenspaces.

Lemma 3.4 (Leading eigenfunctions are well defined and stable under perturbation). *Assume hypothesis (iv) of Theorem 1: the spectrum of $\mathcal{K}\_\theta$ admits a gap $\delta > 0$ separating the K leading eigenvalues — those of largest modulus — from the remainder of the spectrum. Assume further (hypothesis (v) of Theorem 1) that $\mathcal{K}\_\theta$ is normal on the closed subspace spanned by the leading eigenfunctions. Let ΠK denote the spectral projection onto this subspace, and let $mathcal\tilde{K}$ be a bounded operator on $L^2(\mu\_\theta)$ with $\|mathcal\tilde{K} - \mathcal{K}\_\theta\|_{op} \leq \eta$. Then provided $\eta < \delta/2$, $mathcal\tilde{K}$ admits a spectral projection $\tilde{\Pi}K$ onto its K leading eigenvalues, and*

$$\|\tilde{\Pi}K - \Pi K\|_{op} \leq (2\eta)/(\delta - 2\eta).$$

Proof sketch. The result is the standard gap-controlled spectral-projection bound of Davis–Kahan type for operators, extended from the self-adjoint Hilbert-space setting to the normal-operator setting in which it applies on the leading subspace by hypothesis (v). The functional-calculus proof proceeds by writing the spectral projection as a contour integral around the leading eigenvalues, $\Pi K = (1)/(2\pi\ i) \oint\_\Gamma\ (z - \mathcal{K}\_\theta)^{-1}\ \ dz$, where Γ is a closed contour enclosing the K leading eigenvalues and separated from the remainder of the spectrum by at least $\delta/2$. The resolvent identity $(z - mathcal\tilde{K})^{-1} - (z - \mathcal{K}\_\theta)^{-1} = (z - mathcal\tilde{K})^{-1}(mathcal\tilde{K} - \mathcal{K}\_\theta)(z - \mathcal{K}\_\theta)^{-1}$, together with the bound $\|(z - \mathcal{K}\_\theta)^{-1}\|_{op} \leq 2/\delta$ for $z \in \Gamma$ when $\mathcal{K}\_\theta$ is normal on the leading subspace, yields the stated inequality. The condition $\eta < \delta/2$ ensures that $mathcal\tilde{K}$ has no eigenvalues crossing Γ, so the leading projection of the perturbed operator is well defined and is recovered by the same contour. Full details follow Kato (1995, §IV.3.5) and Bhatia (2007, §VII.3).

Remark A.1 (On the local-normality assumption). Hypothesis (v) of Theorem 1, invoked in the proof above, is local — it is required only on the K-dimensional leading subspace, not on the whole operator — which is a substantially weaker condition than global normality. The honest treatment is to include it among the hypotheses of Theorem 1, as is done in the post-revision statement of §3.2. For systems whose Koopman operator is not normal on the leading subspace, the appropriate replacement is a bound involving the eigenvalue condition number (the *pseudospectral* gap; see Trefethen and Embree, 2005), and the Lipschitz constant in Lemma 3.4 acquires a multiplicative factor equal to that condition number. §6 reports a diagnostic — the condition number of the empirical leading eigenvectors — that bears on whether the assumption is met in the data.

Proof of Theorem 1. Combine Lemmas 3.1, 3.2a, 3.2b, 3.3, and 3.4 as in the main-text proof of §3.2. The role of Lemma 3.4 in the theorem itself is to certify that the shared leading spectral structure is *well defined* as a finite collection of eigenvalues and eigenfunctions, the conclusion the theorem asserts. Its role in subsequent results — Proposition 4.1 in particular — is to certify that an estimator approximating the operator approximates this same leading structure.

A.4 Proofs of §4 propositions

The DMD literature contains a number of consistency and bias results for the standard estimator under various sampling regimes. The propositions of §4 are the extensions of these to the multi-task Hankel-augmented setting required for CMH. The proofs are given in the form of reductions to the standard results: each proposition is shown to follow from a corresponding standard theorem under a verifiable modification of the sampling structure. This is the cleanest way to make visible which parts of the proof are new — the multi-task pooling and the Hankel augmentation — and which parts are inherited.

Proposition 4.1 (Consistency). *Under the hypotheses of Theorem 1, the leading spectrum of the multi-task Hankel-DMD estimator converges in probability to the leading spectrum of* $\mathcal{K}\_\theta$ *as* $Nc \to \infty$ *at fixed T (with* $T \geq \tau + K + 1$*), and separately as* $T \to \infty$ *at fixed Nc. The rate of convergence in the* $Nc \to \infty$ *regime is* $O_p(Nc^{-1/2})$ *for the leading eigenvalues, with a constant depending on the spectral gap* δ*, the delay* τ*, the effective noise variance, and the Hilbert–Schmidt norm of the residual projection error of the Hankel-augmented dictionary.*

Proof. The argument has three steps.

Step 1 — Pooled-data large-sample structure. Under Theorem 1's hypotheses, the shared-propagator model is correctly specified (Corollary 3.1). The multi-task Hankel-DMD estimator solves the regularised least-squares problem

$$\hat{A} = argminA \sum_{c=1}^{N^c} \sum_t \| XH^{(c)}(t+1) - A\, XH^{(c)}(t) \|^2 + \lambda \|A\|F^2,$$

where $\|\cdot\|F$ is the Frobenius norm and $\lambda > 0$ is a regularisation parameter selected by the cross-validation procedure of §4.3. Under the SKS condition (Lemma 3.2a) the country-indexed pairs $(XH^{(c)}(t), XH^{(c)}(t+1))$ are identically distributed across c for t in the strictly stationary interior of each country's series, and weakly dependent within each country by the weak-mixing hypothesis (Theorem 1, hypothesis (iii)). The pooled sample of size $Nc \cdot (T - \tau - 1)$ therefore satisfies the moment and dependence conditions of the M-estimation framework for weakly dependent panels (Hayashi, 2000, §6; for the time-series case with mixing, Doukhan, 1994, §2).

Step 2 — Consistency of $\hat{A}$ *for the projected operator.* Let $\mathcal{K}\_\theta[H]$ denote the compression of $\mathcal{K}\_\theta$ to the finite-dimensional subspace spanned by the Hankel-augmented observable dictionary — formally, $\mathcal{K}\_\theta[H] = \Pi[H]\ \mathcal{K}\_\theta\ \Pi[H]$ where $\Pi[H]$ is the $L^2(\mu\_\theta)$-orthogonal projection onto the dictionary. By a multi-task extension of the consistency result of Tu et al. (2014, Theorem 1), adapted to the pooled-panel sampling structure of Step 1, $\hat{A} \to \mathcal{K}\_\theta[H]$ in operator norm at rate $O_p(Nc^{-1/2})$ as $Nc \to \infty$. The proof uses standard M-estimator arguments: the regularised least-squares objective has population minimiser $\mathcal{K}\_\theta[H]$ under correct specification, the empirical objective converges uniformly to its population analogue on bounded operator balls under the weak-mixing-plus-finite-second-moment hypotheses, and the rate follows from a central-limit-theorem argument for α-mixing arrays (Doukhan, 1994, Theorem 2.1).

Step 3 — From the projected operator to the leading spectrum. The estimand of Step 2 is $\mathcal{K}\_\theta[H]$, not $\mathcal{K}\_\theta$. The gap between them is the *projection error*, the loss incurred by representing the infinite-dimensional operator in a finite-dimensional dictionary. The Hankel augmentation enriches the observable dictionary as a Krylov subspace generated by the iterates $\{x, \mathcal{K}\_\theta\, x, \mathcal{K}\_\theta^2\, x, \ldots, \mathcal{K}\_\theta\hat{}\tau\, x\}$ of the base observables (Korda and Mezić 2018, §4); under the standard Krylov-projection bounds for bounded operators on Hilbert spaces, the projection error on the leading K-dimensional invariant subspace tends to zero as τ grows, provided the leading eigenfunctions admit polynomial approximation in the chosen Krylov basis. The rate depends on the spectral gap and on the smoothness of the eigenfunctions, and is bounded by standard polynomial-approximation arguments on $L^2(\mu\_\theta)$. Under this condition Lemma 3.4 transfers operator-norm closeness into closeness of the leading spectral projection at the rate stated. For finite τ, a residual projection error remains and enters the stated rate constant; this is the "Hilbert–Schmidt norm of the residual projection error" of the proposition. Composing Steps 2 and 3 via Lemma 3.4 yields the result. The $T \to \infty$ regime at fixed Nc uses the within-country mixing CLT of Doukhan (1994, Theorem 2.1) applied to each country's series separately and then aggregated; the rate is again $O_p((Nc \cdot T)^{-1/2})$, with Nc entering as a constant prefactor rather than as the asymptotic dimension.

Remark A.2 (Two technical caveats). The consistency rate $O_p(Nc^{-1/2})$ in Step 2 requires (a) that the regularisation parameter λ tends to zero faster than $Nc^{-1/2}$, which is automatic under the cross-validation selection of §4.3 in correctly-specified settings (Györfi et al., 2002, §13.2), and (b) that the pooled design matrix have eigenvalues bounded away from zero in expectation — the rank condition of Proposition 4.3. Both conditions are inherited from the M-estimator framework rather than imposed afresh, but their dependence on the data structure means the rate constant degrades as the pooled design becomes ill-conditioned. In practice this matters: a panel that is "wide" (large Nc) but in which countries are not genuinely independent samples from the ergodic class — Western European polities sharing common shocks, say — has an effective Nc smaller than its nominal one, and the asymptotic rate overstates the precision actually available. §6.4 reports the effective-sample-size diagnostic of Proposition 4.4 that flags this; the proposition is now stated in the main text rather than only in this appendix.

Proposition 4.2 (Finite-sample bias and its direction). *At finite sample size with measurement noise on the observed state, the multi-task Hankel-DMD estimator is biased. The leading-order bias of the estimated eigenvalue* $\hat{\lambda}_j$ *is*

$$\mathbb{E}[\hat{\lambda}_j] - \lambda_j = -\lambda_j \cdot frac\sigma\_\varepsilon^2 \cdot \psi_j{}^t op\ CXX^{-1}\ \varphi_j N_eff + o(N_eff^{-1}),$$

where σ_ε^2 is the variance of the additive measurement noise on the state observations, CXX is the population covariance matrix of the Hankel-augmented state, φ_j and ψ_j are the right and left eigenvectors of the population propagator A at eigenvalue λ_j normalised by $\psi_j{}^t op\ \varphi_j = 1$, and N_eff is the effective pooled sample size of Proposition 4.4. The quadratic form $\psi_j{}^t op\ CXX^{-1}\ \varphi_j$ is real and positive under the local-normality hypothesis (v) of Theorem 1. The bias is then multiplicative in λ_j and negative in sign, so $|\mathbb{E}[\hat{\lambda}_j]| < |\lambda_j|$ at leading order: the estimated eigenvalue moduli are attenuated inward toward the origin.

Averaging over the K leading modes, with the conventions above and assuming the leading K eigenvectors form a biorthogonal system, yields the trace form

$$(1)/(K) \sum_{j=1} K\ fracmathbbE[\hat{\lambda}_j] - \lambda_j\lambda_j = -frac\sigma\_\varepsilon^2 \cdot mathrmtr(\Pi K\ CXX^{-1}\ \Pi K) N_eff + o(N_eff^{-1}),$$

where ΠK is the projection onto the leading K-dimensional invariant subspace. The trace form is the averaged version of the per-mode bias; for any individual mode the per-mode form is the operative one, and §6.5 applies it to the invariant-measure mode whose φ, ψ are both proportional to the constant function. The statement above is the form of the bias under the assumption of independent measurement noise on the predictor and response matrices; for the Hankel-augmented setting in which X_- and X_+ share overlapping state vectors, an additional cross-covariance contribution to the bias appears, as Remark A.6 below characterises. The proof that follows establishes the form under the independent-noise assumption; Remark A.6 then extends the characterisation to the Hankel-augmented case, showing that the modulus-attenuation conclusion is preserved for modes with real biorthogonal eigenvector structure (in particular for the invariant-measure mode of §6.5) and acquires a phase-displacement contribution for non-normal modes.

Proof. The argument is the errors-in-variables bias calculation for ordinary least squares (Greene, 2018, §8.5), adapted to the multi-task Hankel-DMD pooled regression. Let XH^* denote the noise-free Hankel-augmented state and $XH = XH^* + \varepsilon$ the observed state, with ε independent of XH^* and $Var(\varepsilon) = \sigma\_\varepsilon^2 I$. *Under the simplifying assumption that the noise vectors $\varepsilon XH(_t)$ are independent across t* — the assumption Remark A.6 below relaxes — the pooled regression solves

$$\hat{A} = (\sum (XH(t+1))(XH(t))^t op)(\sum (XH(t))(XH(t))^t op)^{-1}$$

ignoring regularisation, which is asymptotically negligible by Remark A.2(a). The numerator is unbiased for A CXX at leading order; the denominator estimates $CXX + \sigma\_\varepsilon^2 I$ rather than CXX, by independence of ε. The bias of $\hat{A}$ is therefore

$$\mathbb{E}[\hat{A}] = A \cdot CXX\ (CXX + \sigma\_\varepsilon^2 I)^{-1} + o(N_eff^{-1}) = A \cdot (I - \sigma\_\varepsilon^2\ CXX^{-1}) + o(N_eff^{-1}),$$

expanding the inverse to first order in σ_ε^2. The leading-order bias of $\hat{A}$ is thus $-\sigma\_\varepsilon^2 A\ CXX^{-1}$. Passing to the spectrum via first-order eigenvalue perturbation theory (Stewart and Sun, 1990, Theorem IV.2.3) for the leading eigenvalues λ_j of A, with right and left eigenvectors φ_j, ψ_j satisfying $\psi_j{}^t op\ \varphi_j = 1$,

$$\mathbb{E}[\hat{\lambda}_j] - \lambda_j = \psi_j{}^t op\ (-\sigma\_\varepsilon^2 A\ CXX^{-1})\ \varphi_j + o(N_e ff^{-1}) = -\sigma\_\varepsilon^2\ \lambda_j \cdot \psi_j{}^t op\ CXX^{-1}\ \varphi_j + o(N_e ff^{-1}),$$

using $A\ \varphi_j = \lambda_j\ \varphi_j$, which is the stated per-mode form. The quadratic form $\psi_j{}^t op\ CXX^{-1}\ \varphi_j$ is positive whenever φ_j and ψ_j are sufficiently aligned with each other in the CXX^{-1} inner product; under hypothesis (v) of Theorem 1 this holds for the leading modes. The averaged trace form follows by summing over j and dividing by K, using biorthogonality $\sum_j \varphi_j\ \psi_j{}^t op = \Pi K$ on the leading invariant subspace. The negative sign and the multiplicative-in-λ_j structure are the substantive features:

$$|\mathbb{E}[\hat{\lambda}_j]| = |\lambda_j| \cdot |1 - \sigma\_\varepsilon^2 \cdot \psi_j{}^t op\ CXX^{-1}\ \varphi_j / N_e ff| < |\lambda_j|$$

to leading order whenever the parenthesised factor is strictly less than one, which holds for $\sigma\_\varepsilon^2 > 0$ and finite $N_e ff$ under hypothesis (v).

Remark A.3 (Debiasing strategies). The bias of Proposition 4.2 is asymptotically removable. Total-least-squares DMD (Askham and Kutz, 2018) solves a different regression problem in which noise is attributed symmetrically to the predictor and the response, producing a consistent estimator of A under the same noise structure. The forward–backward DMD of Dawson et al. (2016) achieves the same end by averaging the standard DMD estimator with its time-reversed counterpart; under stationarity of the underlying process the two biases are equal and opposite at leading order, and their average is unbiased at order $N_e ff^{-1}$. §6 reports estimates under both corrections and under the uncorrected standard estimator, so that the reader can see directly the magnitude of the bias correction and verify that any reported Level III conclusion is robust to the choice of debiasing scheme.

Remark A.4 (Direction of the bias under non-normality). The conclusion that the bias is inward-attenuating — $|\mathbb{E}[\hat{\lambda}_j]| < |\lambda_j|$ — depends on the positivity of the quadratic form $\psi_j{}^t op\ CXX^{-1}\ \varphi_j$, which is licensed by hypothesis (v) of Theorem 1 (local normality on the leading K-dimensional eigenspace). When hypothesis (v) is violated and the left and right eigenvectors ψ_j, φ_j become substantially non-aligned, the sign of the quadratic form is not fixed by the proof: a purely real attenuation can in principle become a complex-valued displacement of $hat\lambda_j$ in the complex plane, with the modulus inward-displaced only on average across modes. The averaged trace form of Proposition 4.2 retains the sign because $tr(\Pi K\ CXX^{-1}\ \Pi K) > 0$ for any non-degenerate CXX and any projection ΠK, but the per-mode statement weakens to "approximately inward" rather than "strictly inward". The conservative argument of §4.4 — that an attenuating bias is friendly to positive findings — therefore remains intact in averaged form on the leading subspace, but the per-mode interpretation requires explicit reading of hypothesis (v) on the mode in question. §6.4 inspects hypothesis (v) empirically; the operational consequence for the present test is that the directional check of §6.5 on the invariant-measure mode applies in the regime where hypothesis (v) holds for that specific mode, which it does (the invariant-measure eigenvector is the constant function, for which the right- and left-eigenvector alignment is unity by construction).

Remark A.5 (Noise model assumptions in Proposition 4.2). The proof of Proposition 4.2 assumes additive, iid, Gaussian measurement noise on the Hankel-augmented state observations, with covariance $\sigma\_\varepsilon^2\ I$. The actual measurement-error structure of the §5 panel is unlikely to satisfy these conditions in full. V-Dem indices arise from expert coding with known coder-effects (Pemstein et al. 2024) that produce non-Gaussian error structure and partial cross-country correlation in measurement error. Maddison historical reconstructions of pre-1870 GDP per capita have larger relative measurement error in older years and exhibit heteroskedasticity over time. COW CINC is itself a composite of six components, each with its own measurement structure. The errors-in-variables calculation that produces the leading-order bias formula remains qualitatively informative under these deviations from the iid Gaussian assumption — the direction of the bias is robust to most realistic noise structures, by the general analysis of Schennach (2016, §3) — but the leading-order constants and the specific functional form of the bias are not. The empirical confirmation in §6.5 is therefore a *directional* check of the proposition, not a calibrated quantitative test

> of its formula. A controlled simulation study of the bias under realistic noise structures, calibrated to the specific measurement-error properties of V-Dem, Maddison, and COW, is identified as a candidate diagnostic for the methodological study flagged in §7.2 (WP-2026-008).
>
> **Remark A.6 (Hankel-induced noise cross-covariance and complex-valued bias).** The proof of Proposition 4.2 as stated assumes that the measurement noise in the Hankel-augmented predictor matrix X_- and response matrix X_+ are independent, an assumption that is appropriate for the raw state vector $X(t)$ under iid measurement noise but is *not* appropriate for the Hankel-augmented vector $XH(t)$. The Hankel augmentation by construction makes $XH(t)$ and $XH(t+1)$ share τ overlapping state vectors: if $XH(t) = [X(t), X(t-1), \ldots, X(t-\tau)]^top$, then $XH(t+1) = [X(t+1), X(t), X(t-1), \ldots, X(t-\tau+1)]^top$, with τ entries in common.

Under iid measurement noise $\varepsilon_s \sim \mathcal{N}(0, \sigma\_\varepsilon^2 I_n)$ on the raw state $X(s)$ with n the dimension of the base state vector, the implied noise on $XH(t)$ has the block-Toeplitz covariance

$$Var(\varepsilon XH(_t)) = \sigma\_\varepsilon^2 I_n(_{\tau+1}),$$

which preserves the diagonal form on a single Hankel column. However, the cross-covariance between the noise on consecutive Hankel columns is non-zero:

$$Cov(\varepsilon XH(_t), \varepsilon XH(_{t+1})) = \sigma\_\varepsilon^2 \cdot S\_\tau,$$

where S_τ is the $n(\tau+1) \times n(\tau+1)$ shift matrix that maps row k to row k-n for $k > n$ and to zero otherwise — a block-Toeplitz off-diagonal capturing the τ entries of overlap between $XH(t)$ and $XH(t+1)$. In the standard OLS-style derivation of the DMD bias, the term $\mathbb{E}[X\_+ X\_-^top]$ acquires a contribution from this cross-covariance: instead of the simple $\mathbb{E}[X\_+ X\_-^top] = A\ CXX$ assumed in the proof, the correct expression is

$$\mathbb{E}[X\_+ X\_-^top] = A\ CXX + \sigma\_\varepsilon^2 \cdot S\_\tau.$$

The DMD estimator becomes

$$\hat{A} = (A\ CXX + \sigma\_\varepsilon^2\ S\_\tau)\ (CXX + \sigma\_\varepsilon^2\ I)^{-1},$$

which differs from the form in the proof of Proposition 4.2 by the additive term $\sigma\_\varepsilon^2\ S\_\tau$ in the numerator. Two consequences follow.

First, the cross-covariance term shifts the bias in a direction that depends on the structure of S_τ relative to the eigenvector basis of A. For modes whose right- and left-eigenvectors φ_j, ψ_j are such that $\psi_j^top\ S\_\tau\ \varphi_j$ has a non-zero imaginary part, the bias acquires a *phase-displacing* component: the estimated eigenvalue is rotated in the complex plane rather than purely shrunk in modulus. The pure-attenuation reading of Proposition 4.2 holds only when $\psi_j^top\ S\_\tau\ \varphi_j$ is real, which is the case under local-normality (hypothesis v) and for modes with real-valued eigenfunctions. For general modes under non-normality of the magnitude observed in §6.4 ($\kappa W \approx 20$), the bias can rotate the estimated phase $arg(hat\lambda_j)$ relative to the true $arg(\lambda_j)$ by a non-negligible angle, in addition to attenuating the modulus.

Second, the invariant-measure mode used as the empirical-confirmation benchmark in §6.5 is *not* affected by the phase-displacement issue. The invariant-measure eigenfunction is the constant function on $L^2(\mu\_\theta)$, with right- and left-eigenvectors identical and real-valued. The quadratic form $\psi_{null}^top\ S\_\tau\ \varphi_{null}$ is therefore real-valued by symmetry, and the bias on $|hat\lambda_{null}|$ remains a real-valued attenuation rather than a complex-valued displacement. The empirical confirmation of §6.5 — that the invariant-measure mode is recovered with modulus inward of unity, in the direction predicted by Proposition 4.2 — is therefore not contradicted by Remark A.6. The remark applies in full strength to the Level III modes the test would have identified, if they existed: for those modes, the Hankel-induced cross-covariance would produce both modulus attenuation and phase displacement, and the §6.5 calibration of the noise scale from the invariant-measure mode alone does not capture the phase-displacement component. The WP-2026-008 controlled simulation study identified in Remark A.5 should include the Hankel-cross-covariance term explicitly in its bias characterisation.

The qualitative content of Proposition 4.2 — that finite-sample DMD on noisy Hankel-augmented data produces eigenvalue estimates displaced from the true values in a direction set by the measurement-noise covariance and the eigenvector structure — survives Remark A.6 intact. What changes is the specific form of the displacement: under iid noise on raw states with Hankel augmentation, the displacement is a combination of modulus attenuation and phase rotation rather than pure modulus attenuation. The Level III binary criterion of §5.4 is expressed on $|hat\lambda|$ alone, so the phase component does not affect the Q1 verdict directly; the cross-window matching criterion of Q3, however, does depend on $arg(hat\lambda)$ via the period $T_j = 2\pi / |arg(hat\lambda_j)|$, and the §6.3 cross-window match in the Juglar band could in principle be affected by phase-displacement bias of the magnitude this remark identifies. The surrogate analysis of §6.7 incorporates this risk implicitly — the null panels are subject to the same Hankel-induced phase bias as the data panels, and the comparison is between the data match-rate and the null match-rate under identical bias structure, so the bias cancels at first order in the cross-window match probability.

A second-order caveat applies. The first-order bias-cancellation argument holds in the regime where the measurement-noise amplitude σ_ε^2 is small relative to the leading singular values of the state covariance CXX. When $\sigma\_\varepsilon^2 / \sigma_{min}(CXX)$ is not small — which is the regime the present panel may occupy on its trailing modes, since the Gavish–Donoho threshold is itself calibrated to where the signal-to-noise ratio approaches unity — higher-order interactions between the noise and the non-normal eigenvector structure can produce asymmetric phase warping that does *not* cancel cleanly between data and null panels. The direction and magnitude of the asymmetry depend on the specific geometry of the eigenvector basis under non-normality, and a closed-form characterisation of the second-order term is not attempted here. The controlled simulation study identified in Remark A.5 (WP-2026-008), calibrated to the specific measurement-error properties of V-Dem, Maddison, and COW, should explicitly include a sensitivity analysis on σ_ε^2 at multiple magnitudes to characterise where the first-order argument breaks down and a more careful surrogate construction becomes necessary. For the present paper the first-order argument is what is operative, with the explicit acknowledgement that it is first-order; a direct decomposition of the phase-displacement contribution to the §6.3 match would require the controlled simulation study just identified.

> **Proposition 4.3 (Identifiability of the leading modes).** *The K leading modes are identifiable from the multi-task Hankel-DMD estimator if and only if the following three conditions hold jointly: (R) the pooled Hankel design matrix has rank at least K with probability tending to one; (G) the spectral-gap condition $\delta > 0$ of Theorem 1 hypothesis (iv) holds, with K defined as the number of eigenvalues above the gap; and (T) the usable per-country series length satisfies $T - \tau - 1 \geq 3\ T_{max}$, where T_{max} is the period of the slowest mode to be resolved.*
>
> *Proof.* Necessity. (R) is necessary because if the design has rank less than K, the dictionary spans a space of dimension less than K and cannot represent K distinct eigenfunctions. (G) is necessary because absent a gap, the "K leading eigenvalues" is not a well-defined target — any clustering of eigenvalues without a gap is arbitrary and unstable. (T) is necessary because the eigenvalue $\lambda_j = e^i\omega_j$ with period $T_j = 2\pi/\omega_j$ is identifiable from the spectrum of $\hat{A}$ only if the data contain enough cycles for the frequency ω_j to be distinguishable from a slowly drifting trend; a record shorter than three periods admits a family of (ω_j, ρ_j) pairs producing the same observed pattern, with growth-rate variation absorbing frequency variation. The lower bound $3\ T_{max}$ is the standard time-series identification threshold for oscillatory regressors (Hannan, 1973).

Sufficiency. Under (R), (G), and (T), the proof of Proposition 4.1 applies and $\hat{A}$ converges to $\mathcal{K}\_\theta[H]$ in operator norm. The K leading eigenvalues of $\mathcal{K}\_\theta[H]$, defined by the gap (G), are recovered by Lemma 3.4 up to the perturbation $\eta = \|\hat{A} - \mathcal{K}\_\theta[H]\|_{op}$, which tends to zero by Proposition 4.1. (T) ensures that the recovered eigenvalues' frequency components are themselves identified, not merely the leading subspace. The estimator's gap-based selection of K from the empirical spectrum is consistent with the true K provided the empirical gap converges to the true gap, which follows from operator-norm convergence and continuity of the spectrum on normal operators (Kato, 1995, §IV.3.5).

A.5 Derivation of Proposition 4.4 (effective sample size)

Proposition 4.4 of §4.4 was stated in the main text in terms of its operational content; this section provides the derivation. The result is the panel-data adaptation of clustered-standard-error logic (Cameron and Miller, 2015) to the operator-norm convergence of Step 2 in the proof of Proposition 4.1.

> *Proof.* Let N denote the pooled column count of the Hankel-augmented predictor matrix, with $N = \sum c\ Nc$ where Nc is the per-country usable count after Hankel augmentation. Under conditional independence given θ, the variance of the pooled cross-product estimator $\hat{C}XX = N^{-1} \sum c,_t\ XH(^c)(t)\ XH(^c)(t)^t op$ scales as N^{-1}, and the central-limit step in Step 2 of the proof of Proposition 4.1 delivers $O_p(N^{-1}/^2)$ at the nominal sample size. When cross-country residual correlations are non-zero with average pairwise correlation $\bar{\rho}$, the variance of $\hat{C}XX$ acquires an additional term from the cross-country covariances. Standard cluster-robust calculations yield

$$Var(\hat{C}XX) \propto frac1 + overline\rho(Nc^* - 1)N,$$

where Nc^* = number of country clusters in the panel (eight for the present panel), so that the effective sample size — the size of an independent sample yielding the same variance — is

$$N_eff = (N)/(1 + overline\rho)(Nc^* - 1).$$

Propagating this through the resolvent identity used in Lemma 3.4 yields the rate-constant degradation of Proposition 4.4. With $N = 612$ and $\bar{\rho} = 0.044$ for the pre-war sub-window, $N_eff = 612 / (1 + 0.044 \cdot 7) = 468$. With $N = 494$ and $\bar{\rho} = 0.193$ for the post-war sub-window, $N_eff = 494 / (1 + 0.193 \cdot 7) = 210$. These are the values reported in §6.4.

Appendix B — Implementation and Replication

B.1–B.6 Replication materials

All replication materials — data sources and harmonisation log (B.1), pre-processing pipeline (B.2), estimator implementation history with bug-level diagnoses (B.3), canonical DMD pipeline details (B.4), cross-validation for rank selection (B.5), and robustness diagnostics implementation (B.6) — are documented in full in the project's GitHub repository: https://github.com/FICSS-Institute/computational-macrohistory/tree/main/CMH-Koopman-Western-Europe. The repository README reproduces the section structure of the present appendix and links to the corresponding scripts and logs. The salient quantities from B.3–B.5 that bear on the §6 results are tabulated in §6.1, §6.4, and §8 of the main text.

B.7 Software environment and replication artefacts

The empirical analysis was conducted in Python 3.12 on a Linux x86-64 system, with NumPy 1.26, SciPy 1.13, and pandas 2.2; no dependencies beyond the standard scientific Python stack are required. The repository directory contains the harmonised panel (data/WP-2026-006_WesternEurope_Panel_v0-1.csv, 1,360 country-years × 5 variables, with codebook), the panel-construction and validation scripts with the harmonisation log, the production estimator (scripts/WP-2026-006_estimator_v3.py), the robustness diagnostics script, the verbatim run logs, and the post-hoc surrogate and power analyses of §6.7. The behaviour of the discarded v1 and v2 estimator implementations is reproducible from v3 by the targeted modifications documented in the repository README (omission of the SVD-projection step for v1; Gaussian log-likelihood cross-validation criterion for v2). Analysis from the CSV is reproducible without internet access once the repository has been cloned: approximately 30 seconds of compute time for the canonical estimator and 45 seconds for the full robustness suite, with no specialised hardware. The paper is deposited canonically on Zenodo with a permanent DOI; replication materials are hosted on GitHub under MIT license, consistent with the FICSS convention separating paper archival (Zenodo) from data and code distribution (GitHub).

References

Askham, T., and J. N. Kutz. 2018. "Variable Projection Methods for an Optimized Dynamic Mode Decomposition." *SIAM Journal on Applied Dynamical Systems* 17 (1): 380–416. doi:10.1137/M1124176.

Bhatia, R. 2007. *Perturbation Bounds for Matrix Eigenvalues*. Classics in Applied Mathematics 53. Philadelphia: SIAM.

Bolt, J., and J. L. van Zanden. 2024. "Maddison-Style Estimates of the Evolution of the World Economy: A New 2023 Update." GGDC Research Memorandum GD-205. Groningen: University of Groningen.

Cameron, A. C., and D. L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317–372. doi:10.3368/jhr.50.2.317.

Davis, C., and W. M. Kahan. 1970. "The Rotation of Eigenvectors by a Perturbation. III." *SIAM Journal on Numerical Analysis* 7 (1): 1–46. doi:10.1137/0707001.

Dawson, S. T. M., M. S. Hemati, M. O. Williams, and C. W. Rowley. 2016. "Characterizing and Correcting for the Effect of Sensor Noise in the Dynamic Mode Decomposition." *Experiments in Fluids* 57 (3): 1–19. doi:10.1007/s00348-016-2127-7.

Doukhan, P. 1994. *Mixing: Properties and Examples*. Lecture Notes in Statistics 85. New York: Springer. doi:10.1007/978-1-4612-2642-0.

Eisner, T., B. Farkas, M. Haase, and R. Nagel. 2015. *Operator Theoretic Aspects of Ergodic Theory*. Graduate Texts in Mathematics 272. Cham: Springer. doi:10.1007/978-3-319-16898-2.

Gavish, M., and D. L. Donoho. 2014. "The Optimal Hard Threshold for Singular Values is *4/√(3)*." *IEEE Transactions on Information Theory* 60 (8): 5040–5053. doi:10.1109/TIT.2014.2323359.

Greene, W. H. 2018. *Econometric Analysis*. 8th ed. New York: Pearson.

Hannan, E. J. 1973. "The Asymptotic Theory of Linear Time-Series Models." *Journal of Applied Probability* 10 (1): 130–145. doi:10.2307/3212501.

Hayashi, F. 2000. *Econometrics*. Princeton, NJ: Princeton University Press.

Hemati, M. S., C. W. Rowley, E. A. Deem, and L. N. Cattafesta. 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.

Jovanović, M. R., P. J. Schmid, and J. W. Nichols. 2014. "Sparsity-Promoting Dynamic Mode Decomposition." *Physics of Fluids* 26 (2): 024103. doi:10.1063/1.4863670.

Juglar, C. 1862. *Des Crises Commerciales et de leur Retour Périodique en France, en Angleterre et aux États-Unis*. Paris: Guillaumin.

Kato, T. 1995. *Perturbation Theory for Linear Operators*. 2nd ed. Classics in Mathematics. Berlin: Springer. doi:10.1007/978-3-642-66282-9.

Korda, M., and I. Mezić. 2018. "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator." *Journal of Nonlinear Science* 28 (2): 687–710. doi:10.1007/s00332-017-9423-0.

Pemstein, D., K. L. Marquardt, E. Tzelgov, Y. Wang, J. Medzihorsky, J. Krusell, F. Miri, and J. von Römer. 2024. "The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data." V-Dem Working Paper 21, 9th ed. University of Gothenburg, Varieties of Democracy Institute.

Sashidhar, D., and J. N. Kutz. 2022. "Bagging, Optimized Dynamic Mode Decomposition for Robust, Stable Forecasting with Spatial and Temporal Uncertainty Quantification." *Philosophical Transactions of the Royal Society A* 380 (2229): 20210199. doi:10.1098/rsta.2021.0199.

Schennach, S. M. 2016. "Recent Advances in the Measurement Error Literature." *Annual Review of Economics* 8: 341–377. doi:10.1146/annurev-economics-080315-015058.

Schmid, P. J. 2010. "Dynamic Mode Decomposition of Numerical and Experimental Data." *Journal of Fluid Mechanics* 656: 5–28. doi:10.1017/S0022112010001217.

Schumpeter, J. A. 1939. *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process*. 2 vols. New York: McGraw-Hill.

Singer, J. D. 1987. "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816–1985." *International Interactions* 14 (2): 115–132. doi:10.1080/03050628808434695.

Stark, J. 1999. "Delay Embeddings for Forced Systems. I: Deterministic Forcing." *Journal of Nonlinear Science* 9 (3): 255–332. doi:10.1007/s003329900072.

Stewart, G. W., and J.-G. Sun. 1990. *Matrix Perturbation Theory*. Boston: Academic Press.

Takens, F. 1981. "Detecting Strange Attractors in Turbulence." In *Dynamical Systems and Turbulence, Warwick 1980*, edited by D. A. Rand and L.-S. Young, 366–381. Lecture Notes in Mathematics 898. Berlin: Springer. doi:10.1007/BFb0091924.

Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. 1992. "Testing for Nonlinearity in Time Series: The Method of Surrogate Data." *Physica D: Nonlinear Phenomena* 58 (1–4): 77–94. doi:10.1016/0167-2789(92)90102-S.

Trefethen, L. N., and M. Embree. 2005. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton, NJ: Princeton University Press.

Tu, J. H., C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz. 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. 2003. *Historical Dynamics: Why States Rise and Fall*. Princeton, NJ: Princeton University Press.

Turchin, P. 2018. *Ages of Discord: A Structural-Demographic Analysis of American History*. Chaplin, CT: Beresta Books.

Turchin, P., and S. A. Nefedov. 2009. *Secular Cycles*. Princeton, NJ: Princeton University Press.

Walters, P. 1982. *An Introduction to Ergodic Theory*. Graduate Texts in Mathematics 79. New York: Springer. doi:10.1007/978-1-4612-5775-2.

Williams, M. O., I. G. Kevrekidis, and C. W. Rowley. 2015. "A Data-Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition." *Journal of Nonlinear Science* 25 (6): 1307–1346. doi:10.1007/s00332-015-9258-5.