Computational Macrohistory: Operational Framework
State Space, Variables, and Dynamic Equations

Stefano Angeli
Foundations Institute of Computational Social Science (FICSS)
stefano.angeli@ficss.institute

July 2026

Abstract

This paper translates the axiomatic foundations of Computational Macrohistory (CMH), established in Document I, into an operational mathematical structure for empirical application. It defines a 25-dimensional state space organised into five categories of variables: demographic (D), economic (E), political (P), social (S), and psychological-collective (Ψ), with canonical definitions, measurement protocols, and documented data limitations for each. System evolution is specified through a coupled system of stochastic differential equations, interpreted throughout in the Itô sense, whose drift terms are theoretically motivated by structural-demographic and collective-action theory rather than fitted post hoc. Critical events are modelled through threshold functions that yield probability distributions, never point forecasts. Relative to the first version of this framework, three developments are integrated. The four theorems of Document I now supply quantitative admissibility diagnostics, including the aggregation threshold with the Herding Threshold correlation ceiling and the reflexivity contraction condition. The three-level prediction hierarchy of the Koopman programme (event prediction, regime probability, spectral structure persistence) locates each framework output at its proper epistemic level. The empirical results of the Arab Spring case studies inform a corrected canonical youth-bulge definition, a graded ladder of operationalisation levels connecting the full dynamic system to the reduced Systemic Stress Index, and the specification of the planned fifteen-variable SSI extension.

Keywords: computational macrohistory, state space, dynamic systems, structural-demographic variables, stochastic differential equations, systemic stress index, political instability

JEL Classification: C02, C63, N00, D74

Suggested citation: Angeli, S. (2026). *Computational Macrohistory: Operational Framework. State Space, Variables, and Dynamic Equations* (FICSS Working Paper No. WP-2026-002, v2.2). Foundations Institute of Computational Social Science. https://doi.org/10.5281/zenodo.21256304

1 Introduction

Document I of this series (Angeli, 2026a) established eight admissibility axioms (A1 through A8) that any valid Computational Macrohistory model must satisfy, and derived from them four formal theorems: a predictive-decay bound with a saturating accuracy law, an aggregation threshold incorporating intra-class correlation, a fixed-point condition for reflexive predictions, and a calibration decomposition anchoring falsifiability to the Murphy partition of the Brier score. Those results specify what CMH can legitimately claim and where its boundaries lie. They do not specify how to construct a working model. This paper supplies that construction: a state space, a set of measurable variables, a system of evolution equations, an event model, and the calibration and simulation machinery required to generate falsifiable probabilistic output. The relationship between the two documents is that of foundations to architecture. The axioms constrain which structures are permissible; this paper specifies one such structure in operational detail.

The first version of this framework appeared in January 2026, before the empirical and theoretical developments that have since reshaped the CMH research programme. Four of those developments motivate the present revision. First, the theorems now proved in Document I convert several qualitative compliance requirements into quantitative diagnostics: the critical population threshold of Axiom A1 is no longer a stipulated hyperparameter but an explicit function of variance, precision, and intra-class correlation, bounded by the Herding Threshold beyond which aggregation fails at any scale; the reflexivity constraint of A6 reduces to an estimable contraction constant; and A8 compliance is now assessed through a standardised reliability-resolution-uncertainty decomposition. Second, the two Arab Spring case studies (Angeli, 2026b, 2026c) operationalised a reduced five-component Systemic Stress Index against real historical data, exposing both the discriminating power and the specification limits of the empirical proxy, and revealing a definitional inconsistency in the youth-bulge variable that this version corrects canonically. Third, the methodological survey of prediction-horizon extension (Angeli, 2026d) introduced a three-level hierarchy of prediction types, distinguishing event prediction, regime probability, and spectral structure persistence, which now organises the epistemic status of every output the framework produces. Fourth, the Koopman-CMH programme (Angeli, 2026e) proved a conditional correspondence between the ergodicity axiom and the spectral structure of the associated Koopman operator and carried out a first empirical test on a Western European panel, opening a spectral extension of the operational framework that the first version could not anticipate.

The revision also removes material that Document I now covers. The earlier version restated the axioms at length, reproduced their motivations, and duplicated substantial portions of the foundational argument. Version 2.2 assumes Document I as read, cites its theorems where they carry operational weight, and confines itself to what is genuinely operational: definitions, equations, thresholds, calibration procedures, and the verification that the resulting structure inherits the epistemological constraints of the axiomatic base.

The comprehensive specification developed here may appear disproportionate to the quality and quantity of available historical data. The disproportion is deliberate. CMH trades model simplicity for epistemic discipline: the explicit machinery of state spaces, stochastic equations, and calibration protocols forces the researcher to articulate assumptions that a simpler model would leave implicit. The full framework functions as a reference architecture. Practical applications typically use reduced versions, and Section 2.6 formalises the ladder of such reductions, but a reduced model should be adopted in the knowledge of what it omits, not by default.

One interpretive caution frames everything that follows. Any resemblance between CMH-derived dynamical regimes and specific historical episodes is interpretive, not inferential. The framework generates abstract state-space trajectories and probability distributions; mapping these onto particular historical events involves additional interpretive layers that the mathematical structure does not license. Such mappings can be heuristically valuable, and the case studies of this series use them as illustrative test cases, but they do not constitute historical explanation or retrodiction.

2 The state space

2.1 Definition and mathematical structure

The state of a large-scale social system at time t is represented by the vector

$$\mathbf{X}(t) = [\mathbf{D}(t), \mathbf{E}(t), \mathbf{P}(t), \mathbf{S}(t), \mathbf{\Psi}(t)]^\top \in \mathbb{R}^{25} \tag{1}$$

where each sub-vector $\mathbf{D}, \mathbf{E}, \mathbf{P}, \mathbf{S}, \mathbf{\Psi} \in \mathbb{R}^5$ collects five variables from one of the five fundamental categories. The full component listing, with canonical definitions and primary data sources, is given in Section 3 and summarised in Appendix A.

2.2 The five fundamental categories

The categorical structure reflects both theoretical considerations and empirical practice in the structural-demographic tradition. Demographic variables (D) capture the size, growth, age structure, spatial distribution, and movement of populations; they form the material foundation of any society, change slowly on scales of years to decades, and constrain every other subsystem. Economic variables (E) capture output, inequality, price dynamics, employment, and growth, combining slow structural trends with faster cyclical components. Political variables (P) capture regime type, regime durability, civil liberties, corruption, and fiscal state capacity; they can remain stable for decades and then shift abruptly during transitions. Social variables (S) capture the structure of social relations through fragmentation, education, connectivity, trust, and religiosity; they evolve slowly but shape crisis dynamics profoundly. Psychological-collective variables (Ψ) capture aggregate mental states such as regime legitimacy, economic expectations, social stress, national identity, and protest propensity; they are the hardest to measure and the fastest to move, and they mediate the path from structural condition to collective behaviour.

The categories are not independent; Section 4 formalises their couplings, and the decomposition organises the state space without asserting any separability of the underlying dynamics.

2.3 Dimensionality

The choice of five variables per category balances completeness against parsimony, measurability, and interpretability. The variables must capture the major factors that the structural-demographic literature identifies as relevant to macro-political dynamics; a materially larger set would multiply parameters beyond what historical data can support and invite overfitting; every variable must be operationalisable with available data at least for the post-1950 period; and the state space must remain legible to historians and political scientists, not only to mathematicians. No finite variable set captures all aspects of social reality, and the claim is correspondingly modest: twenty-five dimensions suffice to model the dynamics relevant to CMH's predictive goals, namely the probability of critical events such as political instability, regime transition, and systemic crisis. Whether they do suffice is an empirical question that the falsifiability apparatus of Axiom A8 keeps permanently open.

2.4 Observable and latent variables

The variables divide into three epistemic classes. Directly observable variables are measurable through objective indicators with minimal interpretation: population aggregates, GDP, unemployment rates, fiscal revenue. Proxy-observable variables require interpretation but rest on established measurement traditions: survey-based trust, attainment-based education, usage-based connectivity. Latent constructs, which comprise the entire Ψ category, represent aggregate psychological states that can only be inferred from bundles of indicators: legitimacy from confidence surveys and compliance behaviour, protest propensity from stated willingness and observed mobilisation. The Ψ variables are included because structural conditions alone do not determine outcomes; collective perception and willingness to act mediate the path from structure to event. Their latent character introduces measurement uncertainty that every downstream analysis must carry explicitly, and the data-quality assessment of Appendix A grades them accordingly.

2.5 Temporal resolution

The state vector is defined at discrete time points with standard resolution $\Delta t = 1$ year, reflecting the reporting frequency of most macro-indicators, the years-to-decades scale of the dynamics CMH addresses, and computational tractability. Quarterly resolution is possible for economic variables in the modern period, and decadal resolution is appropriate for very long-run historical analysis. The dynamic equations of Section 4 are written in continuous time and implemented as discrete-time difference equations for simulation, a translation whose numerical treatment Section 8 specifies.

2.6 Levels of operationalisation

Not every CMH analysis deploys the full apparatus, and the empirical papers of this series have made it necessary to formalise the ladder of legitimate reductions. Four levels are distinguished, ordered by decreasing structural content.

Level 1 is the full dynamic system: all twenty-five variables, the complete coupled equation set of Section 4, calibrated parameters, and Monte Carlo simulation generating probabilistic trajectories and event distributions. This is the reference architecture, and no published CMH study has yet implemented it in full.

Level 2 is a reduced dynamic subsystem: a subset of variables retaining explicit evolution equations and coupling structure, simulated forward in time.

Level 3 is static structural analysis: the position of a system in state space, or in a projection of it, evaluated at a point in time or compared across systems, without forward simulation. Threshold comparisons and cross-country structural rankings belong here.

Level 4 is scalar index operationalisation: a weighted aggregate of a small variable subset, standardised and combined into a single stress measure, used for retrospective classification or comparative description. The Systemic Stress Index in its current five-component implementation (Section 4.7) is a Level 4 operationalisation, and the Arab Spring studies (Angeli, 2026b, 2026c) are explicit that their claims are confined to that level: they test whether the framework's core structural signals are detectable in historical data, not whether the full dynamic system is empirically adequate.

The ladder carries an epistemic discipline. Claims licensed at one level do not transfer upward: a discriminating Level 4 index says nothing about the adequacy of the Level 1 equations, and conversely a failure at Level 4 falsifies the reduced specification, not the framework. Every CMH publication states its operationalisation level explicitly.

3 The fundamental variables

Four criteria govern inclusion. A variable must have a documented relationship to political instability or systemic crisis in the scholarly literature; it must be operationalisable with available data at least for the post-1950 period; it must exhibit meaningful cross-country and temporal variation; and it must capture something not fully captured by the other variables. This section presents each category in turn, discussing theoretical roles and measurement limits in the text and collecting formal definitions in compact tables. Appendix A consolidates the full specification, and the replication repository documents source-level detail, retrieval protocols, and imputation rules.

3.1 Demographic variables (D)

The demographic sub-vector captures population structure and dynamics. Population growth (D_1) strains resources when rapid, presages fiscal challenge when negative, and generates future youth bulges with a lag that the dynamic equations model explicitly. The youth bulge (D_2) is the single most consistently instability-linked demographic quantity in the literature (Goldstone, 1991; Urdal, 2006): a large cohort of young adults, combined with limited economic opportunity, supplies the demographic raw material of mobilisation.

The canonical CMH definition of D_2 is the proportion of the population aged 15 to 24. This is a change from the first version of this framework, which admitted 15 to 29 as the default range, and it corrects an inconsistency that propagated into the first empirical study: WP-2026-003 used the 15 to 29 range, a published specification that cannot be altered retrospectively and is documented in that paper's errata. The narrower range is now binding for all CMH work for two reasons. It aligns the numerator population of D_2 with the reference population of youth unemployment (E_4), which ILO statistics define on the 15 to 24 range, so that the interaction term between demographic pressure and blocked opportunity operates on a single cohort. And it sharpens the lag structure: the canonical downstream parameters are a bulge persistence time τ_2 of 8 to 12 years, a fertility-to-bulge lag τ_{lag} of 15 to 18 years, and an empirical ceiling $D_2^{\max}$ of 20 to 25 per cent of total population. Analyses using other ranges must declare them and must not describe the result as the CMH youth bulge.

Urbanisation (D_3) concentrates population in settings that facilitate collective action and information diffusion, and rapid urbanisation can outpace infrastructure and governance capacity. Net migration (D_4) signals societal stress through exit and creates social tension through entry, but it carries a structural data limitation that must be documented wherever the variable is used: the underlying UN Population Division figures are quinquennial estimates, frequently derived residually rather than observed, and annual series are interpolations. For this reason D_4 is excluded from the Full Modern twenty-variable configuration and enters only analyses that explicitly justify its inclusion and carry its uncertainty. The dependency ratio (D_5) measures the load of the non-working-age population on the working-age population, with youth-heavy and age-heavy dependency carrying different implications.

Variable	**Definition**	**Unit**	**Primary sources**
D_1 Population growth	Annual change in total population	% per year	UN WPP; World Bank WDI
D_2 Youth bulge	Population aged 15 to 24 over total population	%	UN WPP age structure
D_3 Urbanisation	Urban population over total population	%	UN WUP; World Bank WDI
D_4 Net migration	Net migrants per 1,000 population (quinquennial, interpolated)	per 1,000	UN WPP
D_5 Dependency ratio	Population aged 0 to 14 and 65 plus over population 15 to 64	%	UN WPP

3.2 Economic variables (E)

The economic sub-vector captures material conditions. GDP per capita (E_1, PPP-adjusted) sets absolute living standards; very low levels correlate with state fragility, and rapid change in either direction can destabilise. The Gini coefficient (E_2) is the central inequality measure of structural-demographic theory: high or rising inequality generates grievance, delegitimises elites, and recurs across historical instability episodes (Turchin, 2016; Piketty, 2014). Measurement varies between income and consumption and between pre-tax and post-tax bases; the SWIID standardisation is the CMH default. Inflation (E_3) erodes savings and fixed incomes, and rapid price increases have historically triggered protest. Youth unemployment (E_4), defined on the labour force aged 15 to 24 in alignment with D_2, is the direct measure of blocked opportunity for the cohort most likely to mobilise; the conjunction of high D_2 and high E_4 is among the strongest structural signals the framework tracks. Economic growth (E_5) shapes expectations and legitimacy, with growth collapse after a period of rising expectations especially destabilising.

Variable	**Definition**	**Unit**	**Primary sources**
E_1 GDP per capita	PPP-adjusted output per person	intl. dollars	World Bank WDI; Penn World Tables; Maddison Project
E_2 Gini coefficient	Income or consumption inequality	index 0 to 1	SWIID; UNU-WIDER WIID
E_3 Inflation	Annual change in consumer price index	% per year	World Bank WDI; IMF WEO
E_4 Youth unemployment	Unemployed share of labour force aged 15 to 24	%	ILO ILOSTAT
E_5 Growth rate	Annual change in real GDP	% per year	World Bank WDI; IMF WEO

3.3 Political variables (P)

The political sub-vector captures governance and power. The Polity score (P_1) locates the regime on the autocracy-democracy spectrum from minus 10 to plus 10; the robust empirical finding that intermediate regimes are the most unstable (Hegre et al., 2001; Fearon and Laitin, 2003) enters the event model of Section 5 through an anocracy indicator, and enters the empirical SSI through the anocracy stress transformation of Section 4.7. Regime duration (P_2) counts years since the last major regime change; new regimes are vulnerable and consolidation takes time (Svolik, 2012), while very long-lived regimes accumulate rigidities. The civil liberties index (P_3) measures repression and its relaxation, both of which bear on instability in opposite directions over different horizons. Corruption (P_4) delegitimises regimes and corrodes state capacity, and is measurable only through perception-based indices whose longitudinal comparability is limited. Within CMH, P_4 is oriented as corruption intensity: the source indices score cleanliness and are inverted on ingestion, $P_4 = 10 - \text{CPI}/10$, so that higher values denote more corruption.

Fiscal state capacity (P_5) is defined as total government revenue as a share of GDP, the state's demonstrated ability to extract resources and implement policy, which structural-demographic theory places at the centre of stability dynamics (Turchin and Nefedov, 2009). A measurement rule introduced by the Arab Spring studies is now canonical: for rentier and petrostate economies, where hydrocarbon rents flow to the state outside the tax system, tax revenue alone systematically understates capacity, and total government revenue including resource revenue must be used. The narrower tax-to-GDP measure remains acceptable only for economies without significant resource rents, and the choice must be declared.

Variable	Definition	Unit	Primary sources
P_1 Polity score	Regime type, autocracy to democracy	index, minus 10 to 10	Polity V; V-Dem
P_2 Regime duration	Years since last major regime change	years	Polity V; Archigos
P_3 Civil liberties	Expert-coded liberties assessment	index 1 to 7	Freedom House; V-Dem
P_4 Corruption	Perceived corruption intensity (inverted CPI scale)	index 0 (clean) to 10 (highly corrupt)	Transparency International; WGI
P_5 State capacity	Total government revenue over GDP (incl. resource revenue for rentier states)	% of GDP	ICTD GRD; IMF GFS

3.4 Social variables (S)

The social sub-vector captures the structure of social relations. Ethnic fractionalisation (S_1), the probability that two randomly selected individuals belong to different ethnic groups, conditions both the difficulty of governance and the pattern of mobilisation, with effects mediated by political arrangements and inter-group inequality (Alesina et al., 2003; Fearon, 2003); it is effectively static in available data. Mean years of schooling (S_2) raises productivity, participation, and expectations, and education expanding faster than the economy's absorption capacity produces the educated unemployed, a classic elite-aspirant grievance channel. Internet penetration (S_3) is the connectivity variable: it lowers coordination costs, accelerates information cascades, and has been implicated in every major protest wave since the mid-2000s. Social trust (S_4) and religiosity (S_5) are survey-based measures of cohesion and identity resources whose effects are context-dependent and whose historical coverage is thin.

Variable	Definition	Unit	Primary sources
S_1 Ethnic fractionalisation	$1 - \sum_i p_i^2$ over group shares p_i	index 0 to 1	Alesina et al.; Fearon; EPR
S_2 Education	Mean years of schooling, adults 25 plus	years	Barro-Lee; UNESCO UIS
S_3 Internet penetration	Population share using the internet	%	ITU; World Bank WDI
S_4 Social trust	Share agreeing most people can be trusted	%	WVS; regional barometers
S_5 Religiosity	Share reporting religion very important	%	WVS; Pew; Gallup

3.5 Psychological-collective variables (Ψ)

The Ψ sub-vector represents latent aggregate mental states, inferred from indicator bundles rather than measured directly. Regime legitimacy (Ψ_1) is the population's aggregate perception that the political system deserves obedience; it functions as a stock accumulated over time and subject to shocks, and it is the pivot of the coupling structure, since structural crises can be weathered when legitimacy is high while moderate pressure topples regimes when it is low. Economic optimism (Ψ_2) captures aggregate expectations; the destabilising pattern is not low expectation but rising expectation followed by disappointment. The social stress index (Ψ_3) aggregates psychological strain through behavioural proxies such as suicide rates and negative-experience surveys, and may lead political manifestation of discontent. National identity strength (Ψ_4) conditions both cohesion and the feasibility of separatism. Protest propensity (Ψ_5) is the population's mobilisation potential, part attitudinal and part behavioural, measured jointly from stated willingness and observed protest events.

All five are graded lowest in the data-quality assessment of Appendix A. Their coverage is thin before the survey era, their cross-cultural comparability is contested, and every use of them propagates that uncertainty. They remain in the state space because omitting them would build into the framework the false claim that structure alone determines outcomes.

Variable	**Definition**	**Unit**	**Primary sources**
Ψ_1 Regime legitimacy	Perceived rightfulness of the political system	index 0 to 100	WVS; Gallup; regional barometers
Ψ_2 Economic optimism	Aggregate expectations of future conditions	index 0 to 100	Consumer confidence; Gallup
Ψ_3 Social stress	Aggregate psychological strain	index 0 to 100	WHO; Gallup negative experiences
Ψ_4 National identity	Strength of national identification	index 0 to 100	WVS; ISSP
Ψ_5 Protest propensity	Willingness to engage in political protest	index 0 to 100	WVS; ACLED; GDELT

3.6 Interactions

The twenty-five variables are connected by strong theoretical and empirical dependencies, within and across categories: youth bulge combined with youth unemployment raises protest propensity, inequality erodes legitimacy, weak fiscal capacity threatens regime survival, connectivity amplifies mobilisation, corruption drains effective capacity. Section 4 formalises these relationships as coupling terms; the resulting feedback structure makes the system genuinely dynamical, and the nonlinear responses that Section 5 models as critical events emerge from the interaction structure rather than from any single variable.

4 Dynamic equations

4.1 The master equation

The evolution of the state vector follows the CMH master equation in its compact form,

$$\frac{d\mathbf{X}}{dt} = F(\mathbf{X}, \boldsymbol{\theta}, \varepsilon(t)) \tag{2}$$

where F is the locally Lipschitz drift field required by Axiom A4, $\boldsymbol{\theta}$ is the structural parameter vector, and $\varepsilon(t)$ is a schematic placeholder indicating that the system is subject to stochastic forcing. This compact notation is purely symbolic; the rigorous stochastic formulation is given below by the corresponding Itô jump-diffusion equation, where the stochastic contributions are represented explicitly through Wiener and Poisson processes. A notational convention applies throughout this paper and is stated here, at the point where stochastic dynamics first enter: all stochastic differential equations are interpreted in the Itô sense, in accordance with the FICSS notation standard. Where the diffusion structure is used explicitly, the informal term $\varepsilon(t)$ expands into its Itô components,

$$d\mathbf{X} = F(\mathbf{X}, \boldsymbol{\theta})\, dt + \boldsymbol{\sigma}(\mathbf{X}, t)\, d\mathbf{W}(t) + \mathbf{J}(\mathbf{X}, t)\, d\mathbf{N}(t) \tag{3}$$

with $\boldsymbol{\sigma}(\mathbf{X}, t) \in \mathbb{R}^{25\times m}$ the state-dependent diffusion matrix scaling an m-dimensional Wiener process $\mathbf{W}(t)$, and $\mathbf{J}(\mathbf{X}, t) \in \mathbb{R}^{25\times k}$ the jump-size matrix scaling a k-dimensional Poisson counting process $\mathbf{N}(t)$ that models rare discrete shocks. Structural time delays, most importantly the fertility-to-bulge lag τ_{lag}, are part of the empirical specification of F and appear as delayed coupling terms within it, not as separate arguments of the master equation.

The drift decomposes into within-category dynamics and cross-category coupling,

$$F = \begin{bmatrix} F_D \\ F_E \\ F_P \\ F_S \\ F_\Psi \end{bmatrix} + \mathbf{C}(\mathbf{X}) \tag{4}$$

This decomposition is organisational rather than operational: the equations of Sections 4.3 to 4.6 present each row's total drift, with within-category and coupling contributions combined, and the split into F_{cat} and $\mathbf{C}(\mathbf{X})$ is recovered by assigning each term to the category of its argument variables. A methodological positioning is worth making explicit. CMH operates on aggregate observables and is agnostic to micro-level agent rules. Unlike agent-based models, which simulate individual interactions and derive macro-patterns emergently, CMH takes macro-variables as primitives and models their dynamics directly, a deliberate choice that follows from Axiom A1: individual behaviour lies outside the framework's scope. Two conventions hold throughout Section 4. Adjustment rates α, coupling coefficients β, and damping coefficients γ, δ are non-negative, with signs written explicitly in each equation and two stated exceptions noted where they arise (β_{E_2,E_5} in Section 4.4, κ_i in Section 4.9). And $\xi_i(t)$ denotes the scalar white-noise component of $\varepsilon(t)$ acting on variable i, with intensity σ_i, under the Itô reading declared above.

A remark on dimensional coherence. The equations of this section combine variables of differing natural units (percentages, indices on 0 to 1, indices on 0 to 100). The coupling coefficients β absorb these dimensional conversions and carry the appropriate compound units. The explicit specification of those units, variable by variable and coefficient by coefficient, is maintained in the replication repository. What follows preserves the natural measurement scales of each variable, so that the equations are written in the same units in which the variables are observed and reported.

4.2 From mechanisms to equations

Every drift term derives from an established theoretical mechanism rather than from ad hoc specification. Three traditions supply the mechanisms. Structural-demographic theory (Goldstone, 1991; Turchin, 2003; Turchin and Nefedov, 2009) contributes the core cycle of population pressure, elite overproduction, and state fiscal strain. Resource-competition models in the Malthusian and Lotka-Volterra lineage contribute carrying-capacity constraints and extraction dynamics. Collective-action theory (Granovetter, 1978; Kuran, 1991; Lohmann, 1994) contributes threshold effects and information cascades: grievance alone is insufficient, mobilisation capacity is required, and individual willingness depends nonlinearly on perceived participation of others. Table 1 records the mapping from mechanism to mathematical term.

Table 1: Mechanism-to-equation mapping.

Mechanism	Theoretical source	Mathematical representation
Population-resource pressure	Malthus; Turchin (2003)	Logistic growth with carrying capacity
Blocked mobility and grievance	Goldstone (1991); Urdal (2006)	Youth bulge term $\beta_{E_4,D_2}(D_2 - D_2^{\text{ref}})$ in E_4 dynamics
Elite overproduction	Goldstone (1991); Turchin (2016)	Elite Pressure Index $\mathcal{E}$, Section 4.4
State fiscal strain	Turchin and Nefedov (2009)	Revenue-drain balance in P_5 dynamics
Legitimacy erosion	Goldstone (2001)	Inequality and corruption terms in Ψ_1 dynamics
Mobilisation threshold	Granovetter (1978); Kuran (1991)	Logistic event functions with quadratic stress term
Information cascade	Kuran (1991); Lohmann (1994)	Connectivity coupling $S_3 \to \Psi_5$

Transparency also requires stating what the equations do not capture. Ideological dynamics enter only indirectly through the Ψ variables; exceptional individual leadership violates A1 and lies outside scope; international intervention is modelled only as exogenous shock; and technological discontinuities that alter the mode of production or communication can invalidate parameter stability, an A2 boundary that structural-break testing must monitor. These exclusions follow from the axiomatic boundaries rather than from oversight, and they mark the edge of the framework's applicability.

4.3 Demographic dynamics

Population growth relaxes toward an economically conditioned equilibrium,

$$\begin{aligned} \frac{dD_1}{dt} &= \alpha_{D_1}\,(D_1^{\text{eq}}(\mathbf{E}) - D_1) + \sigma_{D_1}\,\xi_{D_1}(t), \\ D_1^{\text{eq}}(\mathbf{E}) &= D_1^{\text{base}} + \beta_{D_1,E_1}\log(E_1/E_1^{\text{ref}}) - \beta_{D_1,E_3}\max(0,\,E_3 - E_3^{\text{thresh}}) \end{aligned} \tag{5}$$

capturing demographic momentum, the mortality-reducing effect of development, and the fertility-suppressing and emigration-inducing effect of severe inflation. The positive income dependence operates through the mortality channel and is calibrated for the framework's empirical domain; at high income the fertility channel reverses the net effect, a regime absorbed into the context-dependent D_1^{base} rather than modelled through a non-monotone form. The youth bulge is governed by cohort inertia. The cohort aged 15 to 24 at time t was born between $t-24$ and $t-15$, so $D_2(t)$ is a functional of fertility 15 to 24 years earlier. Its evolution is modelled as a delayed relaxation with explicit time lag and stochastic forcing,

$$\frac{dD_2}{dt} = \alpha_{D_2}\,(D_2^{\text{eq}}(D_1(t-\tau_{\text{lag}})) - D_2) + \sigma_{D_2}\,\xi_{D_2}(t) \tag{6}$$

with entry lag $\tau_{\text{lag}} = 15$ to 18 years and relaxation rate $\alpha_{D_2} \approx 1/\tau_2$ calibrated to the bulge persistence $\tau_2 = 8$ to 12 years. The stochastic term $\sigma_{D_2}\,\xi_{D_2}(t)$ captures fluctuations in cohort survival, unmodelled migration of young adults, and data revision noise in the underlying age-structure estimates. On the annual discretisation the delay is handled by state augmentation: the lagged values of D_1 enter as additional components of a finite-dimensional Markov state, so that the Jacobian analysis of Section 5.3 and the spectral connection of Section 6 apply to the augmented system. Urbanisation follows logistic growth modulated by development,

$$\frac{dD_3}{dt} = \alpha_{D_3}\, D_3 \left(1 - \frac{D_3}{D_3^{\max}}\right) \left(E_1/E_1^{\text{ref}}\right)^{\beta_{D_3,E_1}} + \sigma_{D_3}\, \xi_{D_3}(t) \tag{7}$$

where the development multiplier $(E_1/E_1^{\text{ref}})^{\beta_{D_3,E_1}}$ is positive for all admissible incomes, equals one at the reference, and reduces to the linear modulation $1 + \beta_{D_3,E_1} \log(E_1/E_1^{\text{ref}})$ to first order; $D_3^{\max}$ is the asymptotic urbanisation ceiling (Appendix C). Net migration relaxes toward an equilibrium set jointly by economic push factors, political conditions, and regional pull factors (subject to the D_4 data caveat of Section 3.1),

$$\frac{dD_4}{dt} = \alpha_{D_4}\, (D_4^{\text{eq}}(\mathbf{E}, \mathbf{P}) - D_4) + \sigma_{D_4}\, \xi_{D_4}(t) \tag{8}$$

where D_4^{eq} is documented in the replication materials and the regional pull channel enters through its Class B parameterisation rather than as an explicit argument. The dependency ratio follows mechanically from the age-structure dynamics and is treated as a deterministically projected quantity, updated annually from the UN demographic projections conditional on D_1 and D_2. These projections supply the forward path of D_5 directly, and no separate stochastic differential equation is required.

4.4 Elite dynamics and the economic subsystem

CMH does not model elite counts directly, since they resist measurement, but captures elite overproduction through its observable manifestations. Following Turchin (2016), competition for scarce elite positions manifests as concentration of resources among successful competitors, frustration of aspirants who cannot find appropriate positions, and mounting claims on state resources. The Elite Pressure Index aggregates these proxies,

$$\mathcal{E}(t) = \omega_1 E_2(t) + \omega_2 E_4(t) + \omega_3 \max\left(0,\, 1 - P_5(t)/P_5^{\text{ref}}\right) \tag{9}$$

with weights ω_i calibrated on historical episodes of documented elite overproduction and recorded in the replication repository. The max operator constrains the fiscal-claim component to non-negative values: when state capacity exceeds its reference, the resulting fiscal slack may suppress elite pressure but does not generate negative pressure, which would be semantically incoherent for an index whose name encodes the presumption of non-negativity. High $\mathcal{E}$ indicates elite overproduction conditions and feeds the fiscal dynamics below.

Inequality evolves through growth patterns, redistribution capacity, and a mean-reverting political limit,

$$\frac{dE_2}{dt} = \alpha_{E_2} \left[\beta_{E_2,E_5}(E_5 - E_5^{\text{eq}}) + \beta_{E_2,P_5}(P_5^{\text{ref}} - P_5) - \gamma_{E_2}(E_2 - E_2^{\text{eq}})\right] + \sigma_{E_2}\, \xi_{E_2}(t) \tag{10}$$

where the first term admits Kuznets-type dependence of distribution on growth, whose direction is itself context-dependent, so that the sign of β_{E_2,E_5} is estimated as a Class B parameter and constitutes a stated exception to the non-negativity convention of Section 4.1, the second encodes the failure of weak states to redistribute, and the third captures mean-reversion toward a context-dependent equilibrium $E_2^{\text{eq}} > 0$, the political pressure that extreme inequality generates against itself (Piketty, 2014; Turchin, 2016). Youth unemployment responds to growth, cohort pressure, and education beyond absorption capacity, with mean-reversion toward a structural rate,

$$\begin{aligned} \frac{dE_4}{dt} = \alpha_{E_4} \big[&- \beta_{E_4,E_5}(E_5 - E_5^{\text{eq}}) + \beta_{E_4,D_2}(D_2 - D_2^{\text{ref}}) \\ &+ \beta_{E_4,S_2}(S_2 - S_2^{\text{absorb}}) - \gamma_{E_4}(E_4 - E_4^{\text{eq}})\big] + \sigma_{E_4}\, \xi_{E_4}(t) \end{aligned} \tag{11}$$

the last two terms formalising the educated-unemployed channel and the structural rate E_4^{eq} to which the variable reverts exactly when growth sits at its equilibrium, the bulge at its reference, and education at absorption capacity. Here D_2^{ref} is the reference bulge level and S_2^{absorb} the labour-market absorption capacity expressed in mean schooling years; both are Class B parameters. GDP per capita, inflation, and growth follow standard reduced-form dynamics with shock exposure, specified in the replication materials.

Context-dependent equilibria (E_2^{eq}, E_4^{eq}, E_5^{eq}, P_4^{eq}, P_5^{eq}, Ψ_1^{eq}, Ψ_5^{eq}, $P_3^{\mathrm{eq}}(P_1, \Psi_1)$, D_4^{eq}, and the Elite Pressure Index weights ω_i) are Class B parameters: they vary across countries or structural regimes but are stable within a regime. Their canonical central values and admissible ranges are release-versioned in the replication repository alongside the Class C and Class D parameters. The superscript $^{\mathrm{ref}}$ denotes a fixed normalisation reference, distinct from the dynamic equilibrium $^{\mathrm{eq}}$: P_5^{ref} normalises the fiscal component of $\mathcal{E}$ and the redistribution terms, while P_5^{eq} is the level toward which P_5 reverts.

4.5 State capacity and the political subsystem

Fiscal capacity links the economic and political subsystems,

$$\frac{dP_5}{dt} = \alpha_{P_5}\left[\beta_{P_5,E_5}E_5 + \beta_{P_5,E_1}\log(E_1/E_1^{\mathrm{ref}}) - \beta_{P_5,\mathcal{E}}\,\mathcal{E} - \delta_{P_5}(P_5 - P_5^{\mathrm{eq}})\right] + \sigma_{P_5}\,\xi_{P_5}(t) \tag{12}$$

with growth enlarging the revenue base, development raising administrative capacity, elite pressure draining resources through rent-seeking and patronage, and mean-reversion toward the structural fiscal capacity P_5^{eq} for institutional depreciation. Regime duration increments annually and resets to zero on regime change, whose probability the event model of Section 5 supplies. Civil liberties relax toward an equilibrium conditioned on regime type and legitimacy, and corruption rises with weak capacity and high inequality against its own mean reversion,

$$\begin{aligned}
\frac{dP_3}{dt} &= \alpha_{P_3}\,[P_3^{\mathrm{eq}}(P_1, \Psi_1) - P_3] + \sigma_{P_3}\,\xi_{P_3}(t),\\
\frac{dP_4}{dt} &= \alpha_{P_4}\left[\beta_{P_4,P_5}(P_5^{\mathrm{ref}} - P_5) + \beta_{P_4,E_2}E_2 - \gamma_{P_4}(P_4 - P_4^{\mathrm{eq}})\right] + \sigma_{P_4}\,\xi_{P_4}(t).
\end{aligned} \tag{13}$$

4.6 Psychological-collective dynamics

Legitimacy integrates structural performance and decays under grievance,

$$\frac{d\Psi_1}{dt} = \alpha_{\Psi_1}\,[\beta_{\Psi_1,E_5}E_5 - \beta_{\Psi_1,E_2}E_2 - \beta_{\Psi_1,P_4}P_4 + \beta_{\Psi_1,S_4}S_4 - \gamma_{\Psi_1}(\Psi_1 - \Psi_1^{\mathrm{eq}})] + \sigma_{\Psi_1}\,\xi_{\Psi_1}(t) \tag{14}$$

so that growth builds the legitimacy stock while inequality and corruption erode it, trust buffers it, and the stock relaxes toward a regime-specific equilibrium. Economic optimism tracks growth and its recent trajectory, with the J-curve pattern of rising expectation followed by disappointment entering through the derivative dependence. Social stress rises with unemployment and inflation and falls with growth. Protest propensity aggregates grievance and capability,

$$\frac{d\Psi_5}{dt} = \alpha_{\Psi_5}\,[\beta_{\Psi_5,\Psi_3}\Psi_3 + \beta_{\Psi_5,S_3}S_3 - \beta_{\Psi_5,\Psi_1}\Psi_1 - \gamma_{\Psi_5}(\Psi_5 - \Psi_5^{\mathrm{eq}})] + \sigma_{\Psi_5}\,\xi_{\Psi_5}(t) \tag{15}$$

with connectivity supplying the coordination capability that collective-action theory requires, legitimacy suppressing mobilisation, and mean-reversion toward the baseline protest propensity in the absence of structural stressors. National identity evolves slowly and enters primarily as a modulator of the separatism pathway in the event model.

4.7 The Systemic Stress Index

The Systemic Stress Index aggregates pressures across subsystems into a single measure of instability risk. In its full theoretical form,

$$SSI(t) = \sum_i s_i \, w_i \, z_i(t) \tag{16}$$

where the sum runs over the nine components listed here, the z_i are standardised component values, $s_i \in \{+1, -1\}$ is the theoretical sign of each component, and the w_i are non-negative weights summing to one. The composition and signs follow Goldstone's (1991) state-breakdown model: demographic pressure ($+z(D_2)$), economic grievance ($+z(E_2)$, $+z(E_4)$, $+z(E_3)$), performance relief ($-z(E_5)$), state strength ($-z(P_5)$), legitimacy ($-z(\Psi_1)$), and mobilisation pressure ($+z(\Psi_3)$, $+z(\Psi_5)$). No single variable suffices; the conjunction matters, and both signs and weights are derived from theory rather than fitted, for reasons the empirical record has since vindicated. The relationship between stress and risk is nonlinear, modelled through the threshold enhancement

$$SSI^{\text{eff}}(t) = \min\left(SSI^{\max},\ SSI(t) + \kappa_{\text{SSI}} \cdot \max\left(0,\ SSI(t) - SSI^{\text{thresh}}\right)^2\right) \tag{17}$$

which captures the acceleration of instability dynamics once multiple stressors align, Goldstone's revolutionary situation in mathematical form. The saturation bound $SSI^{\max}$ prevents the quadratic term from producing numerically unbounded values when SSI greatly exceeds the threshold, and its value is fixed at +3.0 in the standard calibration (three component-standard-deviation units above the reference mean), a level beyond which further increases carry no additional discriminating information. SSI^{eff} serves descriptive and monitoring purposes, regime classification and threshold surveillance at Level II; the event model of Section 5.2 operates on the raw SSI and estimates its own nonlinearity from the event record. The two nonlinearities are alternative representations for different uses, never stacked.

The empirical implementation to date is a Level 4 reduction of this object. The five-component SSI used in the Arab Spring studies (Angeli, 2026b, 2026c) combines youth bulge, inequality, youth unemployment, an anocracy stress transformation of the Polity score, and internet penetration,

$$SSI_5(t) = 0.15\, z(D_2) + 0.25\, z(E_2) + 0.25\, z(E_4) + 0.25\, z(A) + 0.10\, z(S_3), \qquad A = 1 - \frac{|P_1|}{10} \tag{18}$$

where the anocracy transformation A assigns maximum stress to pure anocracy ($P_1 = 0$) and minimum to consolidated regimes of either type, standing in for the legitimacy and capacity variables that the full index requires but historical data do not yet support at scale. Two empirical lessons from that implementation now constrain the framework. The index's discriminating power depends on the theoretically motivated weight on anocracy stress: under atheoretical equal weighting the composite loses discrimination, which confirms that theory-derived weighting carries substantive content rather than serving as a convenience. And the misclassified cases of the eleven-country study each traced to an identifiable omission of the five-variable reduction rather than to random error, which is the informative-failure pattern that Section 7.5 formalises.

The planned extension is SSI-15, a fifteen-variable implementation over a panel of at least thirty validated country-cases, moving directly from SSI-5 without an intermediate stage. Its preconditions are fixed in advance of any data contact: theory-derived weights specified before estimation, a resolved cross-regional standardisation strategy, and a country sample balanced between event and non-event cases. Until those preconditions are met, SSI-15 remains a specification awaiting its data.

4.8 Coupling and feedback structure

The cross-category couplings of the equations above generate the feedback architecture that determines system stability. Reinforcing loops destabilise: low growth erodes legitimacy, eroded legitimacy raises instability risk, and instability suppresses growth; inequality feeds elite pressure, elite pressure drains state capacity, and weak capacity fails to redistribute. Balancing loops stabilise: high stress can force reform that relieves it, repression can purchase temporary stability at the price of long-run legitimacy, and crisis can circulate elites and relieve overproduction pressure. The relative strength of the two loop families, encoded in the calibrated coupling coefficients, determines whether a given structural configuration sits in a stable basin or near a boundary, and the bifurcation analysis of Section 5.4 makes that determination operational.

4.9 Stochastic components and reflexivity

Every equation carries a noise term whose intensity may be state-dependent. The canonical heteroskedasticity is stress-amplified volatility,

$$\sigma_{\Psi_5}(\mathbf{X}) = \sigma_{\Psi_5}^{\text{base}}\,(1 + \alpha_\sigma \max(0,\, SSI(t))) \tag{19}$$

reflecting the empirical regularity that fluctuations widen as systems approach critical regions; the truncation keeps the intensity positive when the standardised index is negative, since the untruncated linear form would drive it below zero for sufficiently unstressed systems. Discrete shocks enter through the jump term of the Itô form of Section 4.1: economic crises as negative jumps in growth and output, disasters as joint output and stress shocks, external conflicts as multi-variable impacts, and leadership transitions as duration resets with legitimacy consequences. Arrival rates and impact distributions are calibrated from historical shock frequencies.

Reflexivity, the modification of the system by knowledge of predictions, is implemented through the Ψ variables as

$$\Psi_i(t^+) = \Psi_i(t^-) + G_i(P(\text{event}),\ \text{salience}) \tag{20}$$

with G_i a bounded feedback function as Axiom A6 requires. The canonical specification is

$$G_i(P(\text{event}),\ \text{salience}) = \kappa_i \cdot \tanh(\eta_i \cdot P(\text{event}) \cdot \text{salience}) \tag{21}$$

where κ_i controls the maximum displacement and carries the direction of the reflexive response for variable i, negative for legitimacy-type variables and positive for mobilisation-type variables, so that κ_i is signed and constitutes, alongside β_{E_2,E_5}, the second stated exception to the non-negativity convention of Section 4.1; η_i governs the sensitivity to the prediction-salience product; and the hyperbolic tangent enforces the boundedness that A6 demands through a smooth response function. The product $|\kappa_i|\,\eta_i$ bounds the local slope of G_i and thus feeds the empirical estimate of the contraction constant introduced below. Theorem 3 of Document I supplies the operational diagnostic: when the induced map on predictive distributions is a contraction under the Wasserstein-1 metric, with Lipschitz constant $\kappa < 1$, a unique self-consistent prediction exists and iterative belief dynamics converge to it; when $\kappa \geq 1$ the system is reflexively unstable and point predictions are not licensed, though the probability of entering such a regime may still be estimated. The constant κ is estimable from natural experiments in which predictive information was released exogenously, and CMH analyses of high-salience predictions must report the estimated or bounded κ alongside the prediction. For most applications, where predictions are not widely known, salience is low and reflexive effects are negligible perturbations.

5 Critical events and thresholds

5.1 Taxonomy

CMH predicts the probability of critical events, organised in a severity hierarchy of decreasing frequency. This event taxonomy is distinct from both the operationalisation ladder of Section 2.6 and the prediction hierarchy of Section 6, and the three should not be conflated.

Table 2: Taxonomy of critical events.

Severity	Event class	Operational definition
1	Instability episode	Mass protest above 1% of population; political violence above 100 casualties per year; government crisis
2	Regime transition	Polity change of 3 points or more; revolution (overthrow via popular mobilisation); coup (overthrow via elite or military action)
3	State failure	Civil war onset above 1,000 battle deaths per year; loss of territorial control or government dissolution

5.2 Threshold functions

Event probabilities are logistic functions of the state,

$$P(\text{Event}_k \mid \mathbf{X}(t)) = \frac{1}{1 + \exp\left(-\left(\beta_{k,0} + \boldsymbol{\beta}_k^\top \mathbf{X}(t)\right)\right)} \tag{22}$$

with event-specific coefficient vectors. The instability specification operates on the stress index and its dynamics,

$$P(\text{instability} \mid \mathbf{X}) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1\, SSI + \beta_2\, SSI^2 + \beta_3\, \Delta SSI))} \tag{23}$$

where the quadratic term implements the mobilisation threshold of collective-action theory (Granovetter, 1978; Kuran, 1991) and the difference term $\Delta SSI = SSI(t) - SSI(t-1)$ captures the destabilising effect of rapidly rising stress independent of its level. Regime change combines structural stress with regime-specific vulnerability,

$$P(\text{regime change} \mid \mathbf{X}) = \frac{1}{1 + \exp\Big(-\Big(\gamma_0 + \gamma_1\, SSI + \gamma_2\,(1 - \Psi_1/100) + \gamma_3 \mathbb{1}_{\text{anocracy}} + \frac{\gamma_4}{1+P_2}\Big)\Big)} \tag{24}$$

where the anocracy indicator activates for $-5 \leq P_1 \leq 5$, encoding the inverted-U regularity (Hegre et al., 2001; Fearon and Laitin, 2003), and the duration term encodes new-regime vulnerability (Svolik, 2012), with the unit offset regularising the hazard at regime age zero, where the unregularised reciprocal of the first version would diverge. Conditional on regime change, the mode depends on mobilisation capacity: high protest propensity, connectivity, and youth bulge favour revolution over coup, while repression shifts the balance toward elite action (recalling that higher P_3 denotes fewer liberties on its source scale),

$$P(\text{revolution} \mid \text{regime change}) = \frac{1}{1 + \exp(-(\delta_0 + \delta_1 \Psi_5 + \delta_2 S_3 + \delta_3 D_2 - \delta_4 P_3))}. \tag{25}$$

The outputs are instantaneous annual probabilities. Cumulative probabilities over horizon T follow from forward simulation,

$$P(\text{Event within } T) = 1 - \prod_{\tau=0}^{T-1} (1 - P(\text{Event} \mid \mathbf{X}(t+\tau))) \tag{26}$$

with future states generated by the ensemble machinery of Section 8. The product treats the annual probabilities as discrete hazards along a single simulated trajectory; because the hazards are state-dependent, the product is evaluated per ensemble member and then averaged across the ensemble, since the expectation of a product of state-dependent hazards does not equal the product of their expectations. Under this construction the widening of the predictive distribution with horizon, mandated by Axiom A7, is automatic rather than appended.

5.3 Bifurcations and early warning

The coupled system can exhibit multiple equilibria, and as stress parameters drift, stable and unstable equilibria can collide and annihilate in a saddle-node bifurcation, leaving only a high-instability attractor. The mathematical signature is that the dominant real eigenvalue of the Jacobian approaches and crosses zero, rendering the Jacobian singular at the bifurcation point. Approaching such a boundary, the system displays the generic precursors of critical transitions (Scheffer et al., 2009): critical slowing down, with recovery time $\tau_{\text{recovery}} = -1/\lambda_{\text{dom}}$ diverging as λ_{dom}, the real part of the dominant Jacobian eigenvalue, approaches zero from below, together with rising variance and rising autocorrelation. The framework monitors these as early-warning indicators, estimating the local Jacobian from recent trajectory segments and flagging proximity to a regime boundary when λ_{dom} rises above a calibrated negative threshold.

An interpretive caution from the first version bears repeating with emphasis. The bifurcation analysis describes formal dynamical properties of the model. Applied to empirical data, it yields information about proximity to regime boundaries as defined by the model, which does not automatically translate into claims about real-world tipping points. The language of critical transitions is exact within the model and metaphorical outside it, and users should resist reifying model-internal concepts as direct descriptions of historical mechanism.

6 The prediction hierarchy and the spectral extension

The most consequential theoretical development since the first version of this framework is the recognition that the question of predictive horizon cannot be posed without first specifying the type of prediction. WP-2026-005 (Angeli, 2026d) introduced a three-level hierarchy of the CMH prediction space, and every output of the operational framework now carries its level explicitly.

At the first level sits event prediction: the trajectory-level forecast of where the system will be at time $t+\tau$. This is the prediction type that Axiom A7 governs directly. Theorem 1 of Document I establishes that its accuracy follows the saturating law $A(\tau) = A_\infty + (A_0 - A_\infty)e^{-\lambda\tau}$, with $\lambda \approx \lambda_{\max}$ inside the exponential window of the chaotic regime and λ set by the Kramers escape rate in the multistable noise-driven regime, the two clauses of Axiom A5. The instantaneous event probabilities of Section 5.2, propagated forward, are Level I objects, and their useful quantitative horizon is five to fifteen years.

At the second level sits regime probability: the forecast that the system will be operating in a qualitatively defined dynamical regime, stable, stressed, or approaching transition, over a future interval. Regime forecasts decay far more slowly because they require only coarse-grained correctness, and because the CMH state space possesses a natural time-scale separation: demographic variables evolve over fifteen to thirty years, structural economic variables over five to fifteen, political and psychological variables progressively faster. The slow manifold parameterised by demographic and long-run economic variables carries regime-level information on decadal scales, and the SSI, which loads principally on slow variables, is best understood as a coordinate on that manifold. The empirical finding that structural stress indices discriminate among outcomes over five-to-fifteen-year windows (Angeli, 2026c) is a Level II result.

At the third level sits spectral structure persistence: the identification of invariant or near-invariant cyclic structures in the dynamics, together with estimates of their periods. WP-2026-006 (Angeli, 2026e) formalised this level through the Koopman operator associated with the master equation and proved the Conditional Koopman-A2 Correspondence (its Theorem 1), which links the conditioned ergodicity of Axiom A2, together with the structural-causality condition of A3, to the spectral decomposition of that operator. Its first empirical test, on an eight-country Western European panel spanning 1820 to 1913 and 1945 to 2020, returned a detectability-limited null result: the data could not resolve Level III modes at the available window lengths, which bounds what current historical panels can detect without constituting evidence against the existence of such modes. A refinement from that work now binds the operational framework: in a stochastic system with non-degenerate diffusion, which is what A4 and A5 jointly commit CMH to, persistent structure does not mean eigenvalues on the unit circle. Phase diffusion pushes every oscillatory mode strictly inside the disc, and the operational definition of a Level III mode is an eigenvalue $\lambda_j = e^{-\gamma_j + i\omega_j}$ whose decay time $1/\gamma_j$ is long relative to the resolvable observation window; operationally, WP-2026-006 fixes the criterion at a decay time exceeding three times the longest period resolvable from the structurally stationary window. Spectral persistence therefore means slow decay rather than exact invariance, and searching for exactly unimodular stochastic Koopman eigenvalues is a category error.

The hierarchy disciplines communication. A statement such as a 35 per cent probability of severe instability within a five-year window is Level I or II and must carry its confidence interval, sample size, validation metric, and stated limits. A statement that a society occupies a particular phase of a multi-decadal structural cycle is Level III and claims nothing about specific events. The framework's ambition of horizon extension, from the five-to-fifteen-year quantitative range toward multi-decadal phase statements, operates entirely at Levels II and III; the Lyapunov Wall at Level I is a structural feature of the dynamics, not an obstacle awaiting better methods.

7 Parameters, calibration, and uncertainty

7.1 Parameter classes

The equations of Section 4 contain parameters of four kinds, each with its own calibration discipline. Structural constants (Class A) reflect stable relationships unlikely to vary across contexts, such as demographic adjustment rates and the signs of basic couplings; they are fixed from the literature or from theoretical constraint. Context-dependent parameters (Class B) vary across countries or periods but are stable within a structural regime, such as baseline growth rates, equilibrium values (E_2^{eq}, E_4^{eq}, E_5^{eq}, P_4^{eq}, P_5^{eq}, Ψ_1^{eq}, Ψ_5^{eq}, $P_3^{\mathrm{eq}}(P_1, \Psi_1)$, D_4^{eq}), and Elite Pressure Index weights $(\omega_1, \omega_2, \omega_3)$; they are estimated separately per context from the historical record of the structural regime in question. Coefficients (Class C) quantify relationship strengths, including SSI weights, threshold-function coefficients, and coupling strengths; they are estimated from pooled cross-country panels by maximum likelihood or Bayesian methods, with the standing exception that SSI weights are theory-derived rather than fitted, for the overfitting reasons Section 4.7 documents. Noise parameters (Class D) govern stochastic intensity, shock arrival rates, and shock magnitude distributions; they are estimated from residual analysis and historical shock frequencies. The canonical central values and admissible ranges of all Class B, C, and D parameters are release-versioned in the replication repository.

7.2 Staged calibration

Calibration proceeds in four stages. Single-variable dynamics are estimated first by time-series methods, with fit statistics reported. Cross-variable couplings are then estimated by panel vector autoregression, with Granger tests identifying significant relationships and structural-break tests guarding the A2 boundary. Event-model parameters are estimated by logistic regression on historical events, using post-1950 data where the variables are reliably measured. A holdout period, typically post-2000, is reserved throughout for out-of-sample validation, comparing predicted probabilities against observed event frequencies through the calibration apparatus that Theorem 4 of Document I standardises: reliability, resolution, and uncertainty components of the Brier score, reported alongside AUC-ROC and log-likelihood.

7.3 Uncertainty propagation

Every parameter estimate carries a distribution, frequentist interval or Bayesian posterior, and predictive uncertainty must reflect parameter uncertainty and stochastic dynamics jointly. The mechanism is Monte Carlo over both: each ensemble member draws a parameter vector from its posterior, simulates a trajectory under independently drawn noise and shocks, and computes event probabilities; the reported prediction is the resulting distribution, summarised by mean and credible intervals, never a point value. This double propagation is what allows the framework to honour A5 and A7 simultaneously, since the ensemble spread widens with horizon by construction.

7.4 Sensitivity and robustness

Local sensitivity is assessed through partial derivatives of outputs with respect to parameters, and global sensitivity through Sobol variance decomposition, identifying which parameters drive output uncertainty and therefore where data investment matters most. Robustness is assessed against alternative variable operationalisations, alternative calibration samples, alternative functional forms, and leave-one-variable-out perturbations. The Arab Spring studies established the template: six alternative SSI weighting schemes and full leave-one-component-out analysis, reported regardless of whether they flattered the baseline specification.

7.5 Informative failure

A feature separating CMH from curve-fitting is that negative results carry scientific content. A calibrated model that fails to identify stable regimes, or produces Brier scores indistinguishable from the climatological baseline, is evidence against the specification and must be reported as such. Diagnosis distinguishes four failure sources: parameter estimation error from insufficient data, variable selection error from a wrong subset of the twenty-five, functional form error from misspecified equations, and framework boundary violation where the system fails an axiom. Only the last falsifies the framework; the first three falsify model instantiations within it, and the diagnostics that Document I attaches to each axiom (the Theorem 2 threshold and Herding Threshold for A1, structural-break tests for A2, the Theorem 3 contraction constant for A6, the Theorem 1 bound for A7, the Theorem 4 decomposition for A8) make the boundary determination itself quantitative. Required reporting for every CMH analysis includes calibration metrics with intervals, comparison against naive base-rate nulls, cases where the model underperforms the baseline, and structural-break test results. The misclassification analysis of the eleven-country study, in which each error traced to a specific documented omission of the reduced specification, is the working example of this discipline.

8 Simulation architecture

The continuous-time system of Section 4 is integrated by the Euler-Maruyama scheme for Itô stochastic differential equations (Kloeden and Platen, 1992), of strong order one half and weak order one,

$$\mathbf{X}(t+\Delta t) = \mathbf{X}(t) + F(\mathbf{X}(t), \boldsymbol{\theta})\,\Delta t + \boldsymbol{\sigma}(\mathbf{X}(t), t)\,\mathbf{Z}_t\,\sqrt{\Delta t} + \Delta\mathbf{J}_t \tag{27}$$

with $\mathbf{Z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$ an independent standard normal draw in $\mathbb{R}^m$, $\Delta\mathbf{J}_t$ the aggregate of jumps arriving in $(t, t+\Delta t]$ under the calibrated compound Poisson shock process, and the Milstein correction available where the state dependence of the diffusion warrants raising the strong order to one. The standard step is one year. Bounded variables (urbanisation, Gini, Polity, the index-valued Ψ components) are enforced by projection onto their admissible ranges after each step, a pragmatic device whose boundary bias is material only when trajectories crowd the bounds, a condition the diagnostic output flags; on the compact admissible domain the projection enforces, all drifts are Lipschitz and the discretised system is well posed. The standard ensemble comprises one thousand members, each drawing parameters, noise, and shocks independently as Section 7.3 specifies; ensemble members are embarrassingly parallel, and full runs for a fifty-year horizon complete in seconds to minutes on commodity hardware. Outputs comprise trajectory percentiles, event probabilities with credible intervals, derived indices including SSI and the Elite Pressure Index, and diagnostic quantities including residuals and Jacobian eigenvalues for the early-warning monitors of Section 5.3. All simulations run under seeded random number generation, and the seed is stored with the results.

Initial conditions $\mathbf{X}(t_0)$ are set from observed historical data at the simulation start date t_0 for all variables for which data exist. For variables with missing observations at t_0, the nearest available observation within a two-year window is used; where no observation falls within that window, the variable is initialised at its estimated regime-specific equilibrium value, and the ensemble draws that value from the posterior distribution of the corresponding Class B parameter to propagate the initialisation uncertainty. The fraction of variables requiring equilibrium initialisation is reported alongside the simulation results. For retrospective validation runs, initial conditions must be drawn exclusively from data available at or before t_0, never from information that postdates the simulation start. The delayed drift of D_2 additionally requires the observed D_1 history over $[t_0 - \tau_{\text{lag}}, t_0]$, which the historical demographic record supplies.

Replication follows the FICSS convention binding since WP-2026-006: the canonical paper PDF is deposited on Zenodo with a permanent DOI, while all replication materials, including data files, scripts, codebooks, parameter tables, and the source-level documentation of the variable specifications in Section 3, reside in the public repository of the FICSS GitHub organisation, one sub-folder per study.

9 Axiom compliance, scope, and limitations

9.1 Compliance

The operational framework inherits the constraints of Document I, and the theorems proved there now make the verification quantitative rather than declarative. Table 3 summarises the correspondence.

Table 3: Axiom compliance of the operational framework.

Axiom	**Framework implementation**	**Quantitative diagnostic (Document I)**
A1 Aggregation	All 25 variables are population-level aggregates	Threshold $N_c = \sigma^2(1-\rho)/(\epsilon^2\alpha - \sigma^2\rho)$; Herding Threshold $\rho^* = \epsilon^2\alpha/\sigma^2$ (Theorem 2; symbols in the sense of Document I)
A2 Ergodicity	Calibration valid within structural regimes; Class B parameters per context	Structural-break tests on panel estimates
A3 Causality	Explicit causal couplings; events conditioned on structural state	Conditional independence testing of event-state dependence
A4 Continuity	Locally Lipschitz drift; shocks as jumps in a continuous system	Lipschitz verification on drift terms
A5 Indeterminacy	Stochastic terms in every equation; all outputs distributional	Lyapunov or Kramers estimation per regime
A6 Reflexivity	Bounded feedback through Ψ variables; explicit tanh specification	Contraction constant κ (Theorem 3)
A7 Decay	Ensemble spread widens with horizon by construction	Saturating law $A(\tau) = A_\infty + (A_0 - A_\infty)e^{-\lambda\tau}$ (Theorem 1)

Axiom	Framework implementation	Quantitative diagnostic (Document I)
A8 Falsifiability	Testable probability distributions; out-of-sample validation	Murphy decomposition: REL, RES, UNC (Theorem 4)

Two entries deserve comment. The A1 diagnostic is stronger than population size alone: Theorem 2 shows that intra-class behavioural correlation above the Herding Threshold defeats aggregation at any scale, so that CMH applications to periods of polarisation cascade or synchronised collective behaviour must estimate ρ and verify feasibility before proceeding. The A7 entry strengthens the first version's qualitative horizon-widening claim: the saturating form bounds what any Level I prediction can claim and supplies the climatological baseline against which skill is measured.

9.2 Scope

The framework is valid for large-scale social systems with populations above the A1 threshold, in practice on the order of millions; for quantitative horizons of one to twenty years, with longer horizons admissible only as conditional regime envelopes at Level II or phase statements at Level III; for political instability, regime transition, and systemic crisis as outcome classes; and for systems without singular dominating actors or unprecedented structural transformation. Small populations, individual behaviour, horizons beyond fifty years, and genuinely unprecedented events lie outside scope, and the framework makes no claims there.

9.3 Limitations

The limitations divide into data, model, and theory. On data, the Ψ variables are poorly measured before the survey era, pre-1950 coverage is sparse for most variables, D_4 is interpolated from quinquennial estimates, and P_5 requires the rentier-state substitution of Section 3.3. On model, twenty-five variables may omit relevant factors, the functional forms are approximations, and the parameter space is large relative to the historical record, so that overfitting is a live risk which the framework counters through out-of-sample validation, null-model comparison, Bayesian regularisation, theory-derived rather than fitted index weights, and the mandatory reporting of negative results; even so, apparent predictive success can reflect in-sample fitting, and only the accumulating out-of-sample record can settle the question. On theory, structural-demographic mechanisms may not transfer to all contexts, modern information technology may be altering mobilisation dynamics in ways the current couplings understate, and the international system enters only as exogenous shock. These limitations define the framework's precision and applicability rather than invalidating it, and several, most visibly the reduced SSI specification, are the explicit targets of the next papers in the series.

10 Conclusion

This paper has specified the operational content of Computational Macrohistory: a 25-dimensional state space with canonical variable definitions and documented measurement limits, a theoretically derived system of coupled Itô stochastic differential equations, a probabilistic event model built on threshold functions, a graded ladder of operationalisation levels connecting the full system to the reduced empirical index, and the calibration, uncertainty, and simulation machinery that turns the structure into falsifiable output. Relative to its first version, the framework is now anchored at both ends: downward to the theorems of Document I, which convert axiom compliance into quantitative diagnostics, and upward to the prediction hierarchy and spectral programme of the later papers, which locate each output at its proper epistemic level and chart the extension of the framework's horizon beyond the event-prediction range.

The division of labour across the series is now settled. Document I states what CMH may claim; this paper states how a claiming model is built; the case studies test whether the framework's structural signals are detectable in historical data at the operationalisation levels the data support; and the Koopman programme develops the spectral level at which the framework's longest-horizon content resides. The nearest empirical milestone is the SSI-15 validation panel, whose preconditions Section 4.7 fixes; the nearest theoretical milestone is the integration of the spectral diagnostics into the early-warning apparatus of Section 5.3. Both will test this framework in the only way Axiom A8 recognises: against data, with the failures reported alongside the successes.

References

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2), 155–194.

Angeli, S. (2026a). *Computational Macrohistory: Axiomatic Foundations for a Quantitative Science of Large-Scale Social Systems* (FICSS Working Paper No. WP-2026-001, v3.0). Foundations Institute of Computational Social Science. https://doi.org/10.5281/zenodo.21178550

Angeli, S. (2026b). *Structural Preconditions of the Arab Spring: A Three-Country Proof of Concept* (FICSS Working Paper No. WP-2026-003). Foundations Institute of Computational Social Science. https://doi.org/10.5281/zenodo.18848734

Angeli, S. (2026c). *Structural Preconditions of Political Instability in MENA, 2000-2012: An Eleven-Country Analysis* (FICSS Working Paper No. WP-2026-004). Foundations Institute of Computational Social Science. https://doi.org/10.5281/zenodo.19661257

Angeli, S. (2026d). *Beyond the Lyapunov Wall: Extending the Predictive Horizon of Computational Macrohistory* (FICSS Working Paper No. WP-2026-005). Foundations Institute of Computational Social Science. https://doi.org/10.5281/zenodo.20111272

Angeli, S. (2026e). *Koopman-CMH: A Formal Treatment of Spectral Structure Persistence in Historical Dynamics* (FICSS Working Paper No. WP-2026-006). Foundations Institute of Computational Social Science. https://doi.org/10.5281/zenodo.20679000

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8(2), 195–222.

Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review*, 97(1), 75–90.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1), 27–38.

Goldstone, J. A. (1991). *Revolution and Rebellion in the Early Modern World*. University of California Press.

Goldstone, J. A. (2001). Toward a fourth generation of revolutionary theory. *Annual Review of Political Science*, 4(1), 139–187.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6), 1420–1443.

Hegre, H., Ellingsen, T., Gates, S., & Gleditsch, N. P. (2001). Toward a democratic civil peace? Democracy, political change, and civil war, 1816-1992. *American Political Science Review*, 95(1), 33–48.

Kloeden, P. E., & Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer.

Kuran, T. (1991). Now out of never: The element of surprise in the East European revolution of 1989. *World Politics*, 44(1), 7–48.

Lohmann, S. (1994). The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989-91. *World Politics*, 47(1), 42–101.
Marshall, M. G., & Gurr, T. R. (2020). *Polity V Project: Political Regime Characteristics and Transitions, 1800-2018*. Center for Systemic Peace.
Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.
Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.
Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260), 53–59.
Svolik, M. W. (2012). *The Politics of Authoritarian Rule*. Cambridge University Press.
Turchin, P. (2003). *Historical Dynamics: Why States Rise and Fall*. Princeton University Press.
Turchin, P. (2016). *Ages of Discord: A Structural-Demographic Analysis of American History*. Beresta Books.
Turchin, P., & Nefedov, S. A. (2009). *Secular Cycles*. Princeton University Press.
Urdal, H. (2006). A clash of generations? Youth bulges and political violence. *International Studies Quarterly*, 50(3), 607–629.

A Consolidated variable specification

Table A.1 consolidates the canonical specification of the 25-dimensional state space. Full source-level detail, including retrieval protocols, interpolation rules, and imputation documentation, is maintained in the replication repository.

Table A.1: The CMH state space: canonical specification.

Var.	Name	Definition	Unit	Typical range	Primary source	Quality tier
D_1	Population growth	Annual change in total population	% per year	-2 to $+4$	UN WPP	A
D_2	Youth bulge	Population aged 15 to 24 over total	%	12 to 25	UN WPP	A
D_3	Urbanisation	Urban share of population	%	10 to 100	UN WUP	A
D_4	Net migration	Net migrants per 1,000 (quinquennial, interpolated)	per 1,000	-20 to $+20$	UN WPP	C
D_5	Dependency ratio	Non-working-age over working-age population	%	40 to 100	UN WPP	A
E_1	GDP per capita	PPP-adjusted output per person	intl. dollars	500 to 100,000	WDI; Maddison	A
E_2	Gini coefficient	Income or consumption inequality	index 0 to 1	0.25 to 0.65	SWIID	B
E_3	Inflation	Annual CPI change	% per year	-2 to $+20$	WDI; IMF WEO	A
E_4	Youth unemployment	Unemployed share of labour force 15 to 24	%	5 to 50	ILO ILOSTAT	B
E_5	Growth rate	Annual real GDP change	% per year	-10 to $+15$	WDI; IMF WEO	A
P_1	Polity score	Autocracy-democracy spectrum	-10 to $+10$	full range	Polity V; V-Dem	B
P_2	Regime duration	Years since last major regime change	years	0 to 100+	Polity V	B

Var.	Name	Definition	Unit	Typical range	Primary source	Quality tier
P_3	Civil liberties	Expert-coded liberties	index 1 to 7	full range	Freedom House	B
P_4	Corruption	Perceived corruption intensity (inverted CPI scale)	0 (clean) to 10 (corrupt)	full range	TI CPI; WGI	B
P_5	State capacity	Total government revenue over GDP	% of GDP	10 to 50	ICTD GRD; IMF GFS	B
S_1	Ethnic fractionalisation	$1 - \sum_i p_i^2$	index 0 to 1	0 to 0.93	Alesina; Fearon; EPR	C
S_2	Education	Mean years of schooling, 25 plus	years	0 to 14	Barro-Lee; UNESCO	B
S_3	Internet penetration	Population share using the internet	%	0 to 100	ITU	A
S_4	Social trust	Generalised trust share	%	5 to 70	WVS	C
S_5	Religiosity	Religion very important share	%	20 to 95	WVS; Pew	C
Ψ_1	Regime legitimacy	Perceived rightfulness of system	0 to 100	full range	WVS; Gallup	D
Ψ_2	Economic optimism	Aggregate expectations	0 to 100	full range	Confidence indices	D
Ψ_3	Social stress	Aggregate psychological strain	0 to 100	full range	WHO; Gallup	D
Ψ_4	National identity	Strength of national identification	0 to 100	full range	WVS; ISSP	D
Ψ_5	Protest propensity	Mobilisation potential	0 to 100	full range	WVS; ACLED	D

Quality tiers grade reliability and coverage: tier A denotes standardised methodology with near-complete modern coverage, tier B good reliability with comparability caveats, tier C moderate reliability with coverage gaps, tier D latent constructs with survey-era coverage only. Post-1990 coverage is good to excellent for tiers A and B; pre-1960 coverage is poor for all tiers except the core demographic aggregates. Gaps of one to two years are linearly interpolated, gaps of three to five years spline-interpolated with a medium uncertainty flag, and longer gaps handled by multiple imputation with a high uncertainty flag; every imputed value carries its method, uncertainty estimate, and auxiliary-information source in the assembled dataset.

B Key derivations

B.1 The Systemic Stress Index

The SSI construction proceeds in three steps. Each component is standardised against regional reference parameters, $z_i(t) = (X_i(t) - \mu_i^{\text{ref}})/\sigma_i^{\text{ref}}$, where the reference means and dispersions are estimated from the regional historical record over a stated window (for the MENA implementation, 1960 to 2010) to ensure regional comparability; sensitivity of the resulting index to reference-parameter perturbation of one standard deviation must be reported. The standardised components are aggregated linearly with theory-derived signs s_i and non-negative weights summing to one, as specified in Section 4.7. The nonlinear enhancement of Section 4.7 is then applied above the calibrated threshold. Linear aggregation is chosen over multiplicative alternatives because the additive structure preserves the interpretability of component contributions, and because the nonlinearity of the stress-to-risk relationship is introduced downstream, through the quadratic term of the event model, where it is estimated from the event record rather than imposed inside the index, with the enhancement term of Section 4.7 as its descriptive counterpart at Level II. Pairwise interaction terms between components are a documented candidate refinement, deferred until the event sample can support their estimation.

For the five-component empirical implementation, the anocracy stress transformation replaces the unavailable legitimacy and capacity components. The transformation $A = 1 - |P_1|/10$ maps the Polity score onto a stress scale that is maximal at pure anocracy and minimal at consolidated autocracy or democracy, encoding the inverted-U vulnerability regularity in index form. The weight structure (0.25 each on inequality, youth unemployment, and anocracy stress; 0.15 on youth bulge; 0.10 on connectivity) reflects the interaction model of structural-demographic theory, in which economic grievance and regime vulnerability are comparably important, demographic pressure operates with a lag through the labour market, and connectivity facilitates rather than generates mobilisation. The robustness analyses of the Arab Spring studies show that discrimination survives regime-emphasis reweighting and fails under equal or economic-emphasis weighting, a pattern the theory predicts: without the anocracy component a fully autocratic rentier state with high demographic and economic stress scores is indistinguishable from a vulnerable anocracy with the same scores.

B.2 The Elite Pressure Index

Elite overproduction is unobservable directly. The index of Section 4.4 combines its three measurable manifestations, resource concentration (E_2), aspirant frustration (E_4), and fiscal claim pressure (the non-negative shortfall of P_5 from its reference, enforced by the max operator of Section 4.4), with weights calibrated on historical episodes of documented elite overproduction. An extended variant adds the education-absorption gap ($S_2 - S_2^{\text{absorb}}$) as a fourth manifestation where schooling data support it, capturing credentialed aspirants specifically. Interpretation is ordinal within a structural regime: the index ranks elite pressure across time and across comparable systems, and its absolute level carries no meaning across structural breaks.

B.3 Event probability functions

The logistic form of the event model follows from the threshold character of collective action. Let latent instability pressure be $Y^* = \beta_0 + \boldsymbol{\beta}^\top \mathbf{X} + u$ with logistic disturbance u; the event occurs when Y^* crosses zero, yielding the logistic probability of Section 5.2. The quadratic stress term enters because the mapping from individual thresholds to aggregate mobilisation is itself nonlinear (Granovetter, 1978): as stress rises, the distribution of individual participation thresholds is traversed at an accelerating rate once the cascade region is reached, and the quadratic term is the lowest-order polynomial representation of that acceleration. The duration term in the regime-change specification represents consolidation as a hazard declining hyperbolically with regime age, regularised at age zero by the unit offset and consistent with the empirical survival pattern of new regimes. Coefficients are estimated by penalised logistic regression on the post-1950 event record, with a Firth-type penalty (Firth, 1993) correcting the small-sample bias that the rarity of severe events would otherwise induce.

B.4 Stability analysis

Local stability is assessed through the Jacobian $\mathbf{A}(\mathbf{X}^*) = \partial F/\partial \mathbf{X}$ evaluated at equilibrium. For the two-variable elite-state subsystem obtained by holding the slow demographic backbone fixed, the trace-determinant conditions yield the boundary in the $(\mathcal{E}, P_5)$ plane at which the stable equilibrium loses hyperbolicity, and the saddle-node condition $\det \mathbf{A} = 0$ locates the fold. The early-warning quantities of Section 5.3 are the empirical counterparts: dominant eigenvalue estimated from locally linearised recent trajectory, recovery time as its negative reciprocal, and variance and autocorrelation as model-free corroborators. The estimation is performed on detrended residuals within a moving window, and the window length trades bias against variance in the usual way; the replication materials fix the defaults.

C Reference parameter values

Reference parameters are release-versioned in the replication repository; Table C.1 records the canonical central values and admissible ranges for the demographic block and the index machinery, which are the quantities fixed by theory rather than estimation.

Table C.1: Canonical reference parameters.

Parameter	**Meaning**	**Central value**	**Range**
τ_{lag}	Fertility-to-bulge lag	16 years	15 to 18
τ_2	Bulge persistence time	10 years	8 to 12
$D_2^{\max}$	Empirical youth-bulge ceiling (15 to 24 definition)	22%	20 to 25
$D_3^{\max}$	Asymptotic urbanisation ceiling	95%	90 to 98
E_3^{thresh}	Inflation stress threshold	15% per year	10 to 20
SSI^{thresh}	Nonlinear enhancement threshold	+0.5	+0.3 to +0.7
κ_{SSI}	Enhancement coefficient	0.5	0.3 to 0.8
$SSI^{\max}$	SSI saturation bound	+3.0	fixed
M	Standard ensemble size	1,000	500 to 10,000
Δt	Standard integration step	1 year	0.25 to 1

D Mathematical notation summary

Symbol	**Meaning**
$\mathbf{X}(t)$	State vector in $\mathbb{R}^{25}$
$\mathbf{D}, \mathbf{E}, \mathbf{P}, \mathbf{S}, \boldsymbol{\Psi}$	Category sub-vectors in $\mathbb{R}^5$
$F(\mathbf{X}, \boldsymbol{\theta}, \varepsilon(t))$	Master-equation vector field (compact form)
$\boldsymbol{\theta}$	Structural parameter vector
$\varepsilon(t)$	Informal stochastic term (Itô components: $\boldsymbol{\sigma}\, d\mathbf{W} + \mathbf{J}\, d\mathbf{N}$)
$\boldsymbol{\sigma}(\mathbf{X}, t)$	Diffusion matrix, $\mathbb{R}^{25\times m}$
$\mathbf{W}(t)$	Vector Wiener process
$\mathbf{J}(\mathbf{X}, t),\ \mathbf{N}(t)$	Jump-size matrix ($\mathbb{R}^{25\times k}$) and k-dimensional Poisson counting process
$\xi_i(t)$	Scalar white-noise component of $\varepsilon(t)$, intensity σ_i
$\alpha, \beta, \gamma, \delta$	Adjustment rates, coupling, damping, and event-model coefficients (non-negative; stated exceptions β_{E_2,E_5}, κ_i)
$\mu_i^{\text{ref}}, \sigma_i^{\text{ref}}$	Reference mean and dispersion for standardisation (Appendix B.1)
$\mathbf{Z}_t,\ \Delta\mathbf{J}_t$	Standard normal draw and aggregated jumps per integration step
$\mathbf{C}(\mathbf{X})$	Cross-category coupling
$\mathcal{E}(t)$	Elite Pressure Index
$SSI(t),\ SSI^{\text{eff}}(t)$	Systemic Stress Index and its nonlinear enhancement
A	Anocracy stress transformation, $1 - \lvert P_1\rvert/10$

Symbol	Meaning
$z_i(t)$	Standardised component value
s_i	Theoretical sign of SSI component i, $s_i \in \{+1, -1\}$
$w_i,\ \omega_i$	SSI weights; Elite Pressure Index weights
$\tau_{\text{lag}},\ \tau_2$	Fertility-to-bulge lag; bulge persistence time
$\mathbb{1}_{\text{anocracy}}$	Indicator for $-5 \le P_1 \le 5$
κ	Reflexivity contraction constant (Theorem 3, Document I)
κ_i, η_i	Reflexivity feedback parameters (signed maximum displacement, sensitivity)
$\rho,\ \rho^*$	Intra-class correlation; Herding Threshold
N_c	Critical population threshold (Theorem 2, Document I)
$A(\tau),\ A_\infty$	Predictive accuracy at horizon τ; climatological baseline
$\lambda_j = e^{-\gamma_j + i\omega_j}$	Stochastic Koopman eigenvalue (decay rate γ_j, frequency ω_j)
λ_{dom}	Dominant Jacobian eigenvalue (early warning)
$\mathbf{A}(\mathbf{X}^*)$	Jacobian at equilibrium
M	Monte Carlo ensemble size
Δt	Integration time step
$SSI^{\max}$	SSI saturation bound
$D_2^{\max},\ D_3^{\max}$	Youth-bulge ceiling; urbanisation ceiling
Superscripts	$^{\text{ref}}$ fixed normalisation reference; $^{\text{eq}}$ dynamic equilibrium (Class B); $^{\text{thresh}}$ activation threshold; $^{\text{absorb}}$ absorption capacity; $^{\text{base}}$ baseline value

Document II of the Computational Macrohistory Series

Foundations Institute of Computational Social Science
`https://ficss.institute`

Contact: stefano.angeli@ficss.institute