Structural stress and political instability in the MENA region: A computational macrohistory analysis of the Arab Spring across eleven countries, 2000–2012

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

July 2026

Abstract

This paper extends the Computational Macrohistory (CMH) framework to eleven countries of the Middle East and North Africa over 2000 to 2012, addressing the structural preconditions of the Arab Spring. Building on the three-country proof of concept that preceded it, we construct a five-component Systemic Stress Index (SSI) over demographic pressure, inequality, youth unemployment, regime type, and connectivity. The SSI is a reduced proxy for testing whether the framework's structural signals are detectable in historical data; it does not implement the full dynamic system. At the 2010 benchmark the index ranks the instability countries above the stability countries in twenty of twenty-eight pairs, an area under the curve of 0.714 with an exact permutation probability of 0.158, and separates the two groups by 0.290 index units. Threshold classification assigns eight of eleven correctly; with the threshold fixed in advance that accuracy carries an exact probability of 0.197, but against its own threshold-optimised null it sits at the median and carries no evidential weight. The ranking is stable across weighting schemes except under heavy connectivity weighting. Leave-one-component-out analysis locates the discrimination in youth unemployment and the anocracy transformation of the Polity score, which carry half the index weight; the other three contribute nothing or degrade it, and the inequality series is model-imputed across the panel. Three misclassified cases, Jordan, Egypt, and Syria, are examined, each a specifiable limit of the specification. The paper states the conditions under which the expanded programme, at thirty or more validated country-cases, would constitute statistical validation.

Keywords: Arab Spring, computational macrohistory, Systemic Stress Index, structural-demographic theory, political instability, MENA, anocracy, youth unemployment, regime change

JEL Classification: C02, C63, N00, D74

Suggested citation: Angeli, S. (2026). *Structural stress and political instability in the MENA region: A computational macrohistory analysis of the Arab Spring across eleven countries, 2000–2012* (FICSS Working Paper No. WP-2026-004, v2.3). Foundations Institute of Computational Social Science. https://doi.org/10.5281/zenodo.21625582

1 Introduction

On 17 December 2010, Mohamed Bouazizi, a twenty-six-year-old street vendor in Sidi Bouzid, set himself on fire in protest against police harassment. Within four weeks President Zine El Abidine Ben Ali had fled Tunisia. Within two months Hosni Mubarak had resigned in Egypt; within ten, Muammar Gaddafi had been killed in Libya, and uprisings of varying intensity had swept through more than a dozen countries. The Arab Spring was a system-level event, a near-simultaneous cascade of political mobilisation across a region characterised for decades by a combination of authoritarian stability and persistent structural grievance. What gave Bouazizi's act its catalytic force were the conditions that had been accumulating for years before the spark arrived.

How much of the outcome was structurally foreseeable? The question belongs to a research programme that applies quantitative and computational methods to large-scale sociopolitical transitions. Structural-demographic theory (Turchin 2003, 2016; Goldstone 2011) has argued, with considerable empirical support, that revolutionary crises emerge from the convergence of demographic pressure, economic grievance, and political fragility rather than from isolated trigger events, however catalytic.

This paper applies the Computational Macrohistory framework (Angeli 2026a, 2026b) to that question systematically. CMH is a formal, axiomatic extension of structural-demographic theory that represents historical systems as trajectories in a twenty-five-dimensional state space and models their dynamics through differential equations and probabilistic inference.[1] It does not predict specific events; it characterises the conditions under which critical transitions become more or less probable. We employ a reduced five-component operationalisation of the framework, the Systemic Stress Index (SSI), to assess whether measurable structural indicators over 2000 to 2010 systematically distinguished countries that subsequently experienced revolutionary transition or civil war from those that did not. The SSI is a Level 4 operationalisation designed to establish whether the framework's core structural signals are empirically detectable in historical data with currently available variables.

This is the fourth paper in the FICSS working paper series on Computational Macrohistory. Document I established the axiomatic foundations. Document II developed the operational framework, the state space, and the dynamic equations. Document III presented a three-country proof of concept, Tunisia, Egypt, and Saudi Arabia, showing that the SSI could be operationalised with real-world data and produced interpretable results consistent with theoretical expectations. This paper extends that analysis to eleven countries, expanding both the geographic coverage and the analytical ambition. With $N = 11$ the study remains firmly within exploratory case study analysis and short of statistical validation, and should be read strictly as exploratory quantitative historical work. The expanded sample nonetheless permits threshold analysis, robustness testing, and the identification of systematic patterns unavailable at $N = 3$.

2 Theoretical framework

2.1 Structural-demographic theory and its extensions

The theoretical foundation of this paper is structural-demographic theory as developed by Goldstone (1991, 2011) and systematically formalised by Turchin (2003, 2016). The theory holds that large-scale political crises, revolutions, state collapses, and civil wars, are not primarily products of ideological conflict, charismatic leadership, or contingent triggers, but of structural

[1]The three documents preceding this one are as follows. Document I (Angeli 2026a) establishes the eight axiomatic foundations, A1 through A8, and their epistemological justification. Document II (Angeli 2026b) develops the twenty-five-dimensional state space [D, E, P, S, Ψ], the variable taxonomy with full operational definitions, the dynamic equation system, and the four-level operationalisation hierarchy. Document III (Angeli 2026c) presents the three-country proof of concept, Tunisia, Egypt, and Saudi Arabia, that preceded this study. All are deposited on Zenodo with permanent DOIs.

conditions that accumulate over generational timescales: population growth outpacing economic productivity, expansion of elite cohorts competing for a fixed set of high-status positions, fiscal strain on the state, and erosion of institutional legitimacy. When several stress channels converge, systems become brittle, and the trigger that causes collapse is often far less significant than the vulnerability it ignites. Not all scholars share this emphasis; work in political sociology and international relations has stressed agency, contingency, and external intervention in shaping revolutionary outcomes, and Kuran (1991) argued that revolutionary cascades are intrinsically unpredictable.[2] This paper does not seek to resolve that debate. It asks a narrower question: whether structural indicators measured quantitatively before 2011 co-varied systematically with subsequent instability outcomes across the MENA region.

Computational Macrohistory extends structural-demographic theory instead of replacing it. CMH adopts the core causal mechanism, structural accumulation as the driver of critical transitions, and adds three contributions. It supplies a formal axiomatic foundation, Axioms A1 through A8, that fixes the epistemological commitments of the framework, including the probabilistic treatment of outcomes (A5), bounded reflexivity as an endogenous and modellable feedback mechanism (A6), and institutionalised falsifiability (A8).[3] It proposes a twenty-five-dimensional state space that extends the structural-demographic variable set to include psychological-collective variables (Ψ) capturing population-level perceptual and attitudinal states that mediate between structural conditions and collective action. And it specifies the dynamics of state-space trajectories as a system of differential equations, opening the way to trajectory analysis, stability classification, and bifurcation detection. None of these extensions is employed directly here. The SSI operationalises the core mechanisms of structural-demographic theory in a replicable scalar index. This is a Level 4 operationalisation of the framework (Angeli 2026b): the theoretical architecture specifies what the index measures, what it does not, and which epistemological claims an analysis of this kind can support.

2.2 The Arab Spring in the structural-demographic literature

The Arab Spring has attracted substantial attention from researchers in the structural-demographic tradition. Goldstone (2011) argued that the configuration of high youth bulges, anocratic regimes, and frustrated elite aspirations in the MENA region made it particularly vulnerable to revolutionary crisis after 2000. Urdal (2006) had previously demonstrated a consistent relationship between youth bulge size and political violence across a large cross-national sample. Korotayev and Zinkina (2011) applied a demographic-structural analysis to the Egyptian revolution directly, identifying the conjunction of youth bulge and urbanisation among its structural preconditions.

[2] Kuran's (1991) central contribution is the concept of preference falsification: under authoritarian conditions individuals publicly express support for a regime they privately oppose, making true political sentiment invisible to observers, including other citizens. The result is a structural disconnect between underlying discontent and its visible expression. When conditions shift and a critical mass begins to reveal true preferences publicly, cascades can occur with extreme rapidity, making revolutions appear unpredictable even when structurally overdetermined. The mechanism is not incompatible with structural-demographic theory, since the two operate at different analytical levels: the structural account identifies the conditions under which the threshold for preference revelation falls far enough for cascades to become probable, while Kuran describes the micro-level mechanism by which they unfold once triggered. Axiom A5 formalises the complementarity by holding that structural accumulation raises the probability of critical transitions without determining their timing or triggering mechanism.

[3] The three axioms cited here are defined in Document I (Angeli 2026a). A5, endogenous indeterminacy, holds that the trajectory of a historical system is not fully determined by its initial conditions and structural parameters, and that genuine probabilistic indeterminacy is endogenous to complex social systems and is not an artefact of measurement error; the axiom prohibits point predictions and mandates probabilistic inference. A6, bounded reflexivity, holds that agents within a historical system can observe and respond to descriptions of that system's state, introducing bounded and modellable feedback between measurement and dynamics. A8, computational falsifiability, requires that empirical claims derived from the framework be expressed in computationally implementable form specifying in advance what evidence would falsify them; for this study, A8 requires that the index formula and classification rule be stated before results are reported, as they are in Section 3.3, and that the analysis be transparent about its epistemological status.

The emphasis on anocracy, regimes intermediate between full democracy and full autocracy, as a specific vulnerability condition draws on Vreeland (2008) and is central to the CMH operationalisation. Work on conflict prediction more broadly (Fearon and Laitin 2003; Hegre et al. 2013) has identified overlapping structural variables, income per capita, population size, rough terrain, as consistent correlates of civil war onset, providing an independent empirical tradition against which the SSI results can be read.[4] On digital communication, Howard and Hussain (2013) argued for a significant enabling effect of social media on protest coordination during the Arab Spring; subsequent research has qualified that claim substantially, and the negative finding reported in Section 6.2 adds to the sceptical body of evidence.

A separate tradition has constructed composite instability indices calibrated specifically to the Arab Spring. Korotayev, Issaev, Malkov, and Shishkina (2014) developed a multiplicative instability index across nineteen Arab countries, aggregating five sub-indicators: conflict potential, socially disaffected youth, regime capacity to manage tensions, historical immunity to internal conflict, and external influence. Their index, calibrated through nonlinear regression on the 2011 events themselves, achieved a coefficient of determination of 0.91 against an ordinal scale of observed resistance to destabilisation. The methodological proximity to this study is substantial, since both aggregate structural factors into a scalar score and evaluate discrimination across the same regional sample, but the differences matter. Their sub-indicators rely on ordinal expert assessments in place of publicly available quantitative data; their multiplicative functional form, with exponents fitted to the calibration sample, carries an overfitting risk that additive theory-weighted specifications are designed to avoid; and the inclusion of subjective factors such as historical immunity and external influence, while it may improve within-sample fit, renders the index non-replicable without access to the same expert panel. The SSI sacrifices within-sample explanatory power in exchange for complete reproducibility and explicit falsifiability, a trade-off that Axiom A8 mandates. Korotayev and colleagues also found that poverty levels and corruption showed no statistically significant correlation with destabilisation outcomes in their sample, and that their youth-related sub-indicator, while contributing to overall fit, carried lower individual significance than regime capacity and conflict potential. Those findings converge with the leave-one-component-out results of Section 6.2.

Complementary to the index-construction approach, Gaub (2012) proposed an analytical framework distinguishing four levels of instability causation: conditions, the long-term structural root causes that change gradually; catalysts, emerging dynamics that multiply tensions over shorter timescales; triggers, one-time events that animate actual dislocation; and the state capacity to manage these pressures. Her comparative analysis of the six principal Arab Spring crises shows that root causes were broadly shared across the region while the catalysts that turned latent conditions into active instability were highly case-specific: rapid expansion of university graduates without corresponding labour-market absorption in Tunisia, food price inflation reaching nineteen per cent in Egypt, a five-year drought that halved water resources and displaced millions of rural inhabitants into Syrian cities, and an extreme demographic acceleration in Yemen, where the population doubled in thirteen years. Gaub argues that the principal analytical failure preceding the Arab Spring was not the overlooking of structural conditions, which were well documented, but the failure to identify the catalysts that gave those conditions their disruptive temporality. That distinction maps onto the difference between the static SSI, which captures

[4]The two conflict-prediction traditions cited are methodologically distinct. Fearon and Laitin (2003) employ cross-sectional logistic regression on a large country-year panel covering 1945 to 1999 to identify structural correlates of civil war onset, emphasising conditions that favour insurgency, rough terrain, large population, low income per capita, over grievance-based variables; their approach is correlational and retrospective. Hegre et al. (2013) develop a probabilistic forecasting model trained on 1970 to 2009 data and used to generate out-of-sample predictions for 2010 to 2050, with explicit calibration and uncertainty quantification. The SSI differs from both. It is mechanistically grounded in structural-demographic theory instead of atheoretically correlational, uses a theoretically selected variable set in place of a broad specification search, and is evaluated retrospectively on a single regional episode and not on a global panel. The comparison is positional and not competitive: the index occupies a distinct place between detailed case analysis and large-N statistical prediction.

conditions, and the dynamic CMH specification toward which the programme is moving, which would capture catalysts as first-derivative terms in the state-space trajectory. It also names a structural limitation of this analysis: the SSI identifies vulnerability but cannot determine timing, because catalysts and triggers are contingent and not structural, a constraint consistent with Axiom A7 on predictive decay.

A third strand has examined the demographic mechanisms underlying instability with attention to youth bulge dynamics. Schomaker and Wentzel (2014), extending Hirschman's (1970) exit-voice framework to regime change, tested the hypothesis that youth bulges produce revolutionary pressure only in combination with restricted emigration opportunities, a configuration they term locked-in under pressure. Using logistic regression on a panel of ninety-two countries over 1986 to 2005, they found that the interaction variable significantly increased the probability both of regime change in general and of democratic transition in particular, with an odds ratio of 2.56 at $p < 0.05$. Their finding that a Muslim-majority dummy showed no significant effect in any specification bears directly on a MENA-focused sample, since it addresses the concern that a regionally concentrated analysis might conflate structural with cultural or religious factors. Their treatment of emigration as a safety valve that reduces revolutionary pressure by channelling youth discontent toward exit instead of voice identifies a variable absent from the current SSI specification, one that may account for part of the residual variance in countries where emigration is particularly constrained, such as Yemen, or particularly available, such as the Maghreb states with established migration corridors to Europe.

What distinguishes this analysis from prior work is methodological ambition and not theoretical claim. We seek to construct a replicable, formally specified index that assigns each country a scalar stress value and to assess whether that value, computed from pre-event data, discriminates outcomes, and not merely to argue that structural conditions were relevant. That operationalisation discipline is what Axiom A8 requires: claims must be expressed in a form that permits falsification.

3 Data and methods

3.1 Country selection and outcome classification

The eleven countries analysed here, Algeria, Bahrain, Egypt, Jordan, Kuwait, Morocco, Oman, Saudi Arabia, Syria, Tunisia, and Yemen, are the principal MENA states affected by or adjacent to the Arab Spring protests of 2010 and 2011 for which systematic data on all five variables are available across the full 2000 to 2012 window. Libya and Iraq were excluded for incomplete pre-event series; both experienced major instability in the period, but the gaps would require more imputation than this analysis accepts. The Palestinian Territories lack internationally comparable state-level statistics for several variables.

Outcome classification is assigned at country level on the most severe political development observed during 2011 and 2012. Six categories are used: revolution, meaning regime transition following mass mobilisation (Tunisia, Egypt); civil war, meaning transition to armed internal conflict (Yemen, Syria); protest contained, meaning major protest without regime change (Algeria, Oman); reform avoided, meaning substantive policy or constitutional concession without regime change (Jordan, Morocco); protest suppressed, meaning forcible suppression of organised protest with external support (Bahrain); and stable, meaning no major protest (Saudi Arabia, Kuwait). For threshold and robustness analysis a binary instability variable collapses revolution and civil war into a single positive class ($N = 4$) against all remaining outcomes ($N = 7$).

The binary classification is a deliberate analytical choice with non-trivial implications that should be stated explicitly. Yemen and Syria are grouped with Tunisia and Egypt, but the mechanisms that produced their outcomes differ substantially: all four regimes were challenged, and in neither of the first two did a clean revolutionary transition occur. Syria in particular, where the Assad regime survived and the conflict evolved into a multi-actor civil war with sectarian

and geopolitical dimensions, is a fundamentally different political dynamic from the Tunisian or Egyptian cases. Collapsing these into one binary category serves analytical tractability; it carries no theoretical claim that revolutions and civil wars share identical preconditions. A more refined analysis would distinguish the outcome types, and we identify this as a priority for future versions of the index. Section 7 returns to the distinction.

3.2 Variables

Five variables are employed, drawn from the twenty-five-dimensional CMH state space (Angeli 2026b). Selection criteria were theoretical relevance, data availability, and prior validation in the structural-demographic literature. Data were collected from primary compilations: the World Bank World Development Indicators for D_2, E_4 and S_3, the UNU-WIDER WIID Companion dataset for E_2, and the Polity5 dataset for P_1. Indicator codes, dataset versions, access dates, and the aggregation applied to the demographic series are documented in the codebook of the replication materials (Appendix C), and Section 8.2 discusses the consequences of the imputation model underlying E_2.

Table 1: SSI component variables.

Symbol	Variable	CMH category	Source	Theoretical role
D_2	Youth bulge, ages 15 to 24, % of population	Demographic	World Bank WDI, age-band series	Demographic pressure
E_2	Gini coefficient $\times$ 100	Economic	UNU-WIDER WIID Companion	Income inequality and grievance
E_4	Youth unemployment, ages 15 to 24, %	Economic	World Bank WDI, ILO modelled estimates	Frustrated expectations
P_1	Polity V score, -10 to $+10$	Political	Polity5 dataset, variable polity2	Regime vulnerability
S_3	Internet penetration, % of population	Social	World Bank WDI, ITU source	Coordination capacity

Note: sources are World Bank (2026), UNU-WIDER (2022), Polity V Project (2021) and, for S_3, the ITU series distributed through the World Development Indicators. The E_2 series is drawn from a model-imputed compilation and every value is an estimate rather than a survey observation; Kuwait carries a single constant value of 40.41 across all thirteen years. Section 8.2 documents the imputation and its consequences. The expected sign on the index is positive for every component, higher values indicating greater instability pressure, with the exception of S_3, which is theoretically ambiguous: internet penetration enables protest coordination but correlates strongly with national income in this sample, which makes it a potential confound in high-income autocracies (Section 6.1). The youth bulge is defined over ages 15 to 24, the canonical band for the D_2 variable throughout the CMH series.

3.3 The Systemic Stress Index

The index is a weighted sum of standardised component scores. Each component is transformed into a z-score against MENA regional reference parameters, taken unchanged from the three-country study that precedes this one (Angeli 2026c, Table 5) and release-versioned in the replication repository. Using conventional instead of sample-estimated reference parameters is a deliberate

methodological choice: estimating parameters from the same sample used for assessment would introduce circularity and inflate apparent performance. The parameters appear in Table 2.

The component weights reflect the theoretical hierarchy established in the structural-demographic literature. The economic grievance variables E_2 and E_4 each receive 0.25, consistent with the theory's emphasis on frustrated economic expectations as a primary driver of mobilisation (Goldstone 1991; Turchin 2016). Institutional vulnerability, the anocracy stress transformation of P_1, likewise receives 0.25, in line with the anocracy hypothesis (Vreeland 2008), under which intermediate regimes are structurally most exposed to revolutionary challenge. Demographic pressure D_2 receives 0.15, reflecting Urdal's (2006) finding that youth bulge effects are mediated through economic and political variables and do not operate independently. Internet penetration S_3 receives 0.10, the lowest weight, reflecting its more distal and contested role in protest coordination. A note on component independence is warranted: the additive specification does not require orthogonality among components, and moderate correlations are expected on theoretical grounds, for instance between the youth bulge D_2 and youth unemployment E_4, which share a demographic substrate. The weights accordingly express theoretical priorities within the structural-demographic framework rather than statistically independent contributions to variance.

Table 2: SSI reference parameters, MENA regional reference values (component-specific windows; see text).

Component	**Variable**	**Reference mean** μ	**Reference dispersion** σ	**Weight** w
D_2	Youth bulge (%)	22.5	4.5	0.15
E_2	Gini $\times$ 100	40.0	8.0	0.25
E_4	Youth unemployment (%)	18.0	8.0	0.25
$A(P_1)$	Anocracy stress	0.5	0.3	0.25
S_3	Internet penetration (%)	15.0	20.0	0.10

Source: Angeli (2026c), Table 5. The reference window differs by component for reasons of data availability: 1960 to 2010 for the youth bulge, 1990 to 2010 for youth unemployment, 2000 to 2010 for internet penetration, and a long-run estimated distribution for the Gini coefficient, while the anocracy parameters describe a theoretical reference distribution over the Polity scale rather than an observed sample of country-years. The parameters are conventional reference values fixed before data contact, not moments estimated from the present sample.

The anocracy stress transformation converts the Polity V score into a measure of institutional vulnerability, hereafter A,

$$A = 1 - \frac{|P_1|}{10} \tag{1}$$

The transformation is symmetric around zero: fully consolidated democracies ($P_1 = +10$) and full autocracies ($P_1 = -10$) both yield $A = 0.0$, while pure anocracies ($P_1 = 0$) yield $A = 1.0$. Its theoretical basis is the anocracy hypothesis (Goldstone 2011; Vreeland 2008), under which intermediate regimes are most vulnerable to mass mobilisation because they lack both the legitimacy resources of democracies and the coercive capacity of consolidated autocracies. The inverted-V functional form, maximum vulnerability at the centre and minimum at the extremes, is a Level 4 operationalisation[5] conceptually analogous to Vreeland's (2008), though that work

[5]The four-level operationalisation hierarchy is defined in Document II (Angeli 2026b). Level 1 is the raw theoretical model, the full system with all twenty-five variables, dynamic equations, and stochastic terms. Level 2 is the dynamic system, a tractable differential equation specification that retains temporal dynamics on a reduced variable set. Level 3 is static structural analysis, the position of a system in the state space or in a projection of it, evaluated at a point in time or compared across systems without forward simulation; threshold comparisons

decomposes the Polity score into sub-components rather than using the aggregate directly.[6]

The full index follows the signed form of Document II,

$$SSI(t) = \sum_{i=1}^{5} s_i \, w_i \, z_i(t), \qquad z_i(t) = \frac{X_i(t) - \mu_i}{\sigma_i}, \qquad s_i \in \{+1, -1\} \tag{2}$$

where s_i carries the theoretical sign of component i. All five components take $s_i = +1$ in the baseline specification, with weights and reference parameters as in Table 2, so that

$$SSI(t) = 0.15\, z(D_2) + 0.25\, z(E_2) + 0.25\, z(E_4) + 0.25\, z(A) + 0.10\, z(S_3) \tag{3}$$

Positive values indicate above-average systemic stress relative to the MENA historical baseline; negative values indicate below-average stress. The sign assignment on S_3 is the one the theory leaves open, and Section 6.2 reports that this sample rejects it. Document II also specifies a threshold-enhanced variant of the index for regime surveillance; the classification exercise here uses the raw weighted sum throughout, because the enhancement is calibrated for monitoring and its parameters are not identifiable on eleven cases. Appendix B states the calculation procedure in full, and every value reported in this paper is reproducible from the data of the replication repository and the parameters of Table 2.

4 Results

4.1 Country-level trajectories, 2000 to 2012

Figure 1 presents the index time series for all eleven countries across the full observation window. Several structural patterns stand out.

and cross-country structural rankings belong to this level. Level 4 is scalar index operationalisation, a weighted aggregate of standardised component scores, which is what the SSI implements. Each level trades theoretical completeness for empirical tractability. The present dataset, eleven countries, five variables and no repeated-measures estimation, supports Level 4 for the index itself and Level 3 for the threshold and ranking analysis of Section 5.

[6] Vreeland (2008) decomposes the Polity IV composite score into two sub-components, competitiveness of political participation and openness and competitiveness of executive recruitment, arguing that the composite conflates dimensions with opposite effects on civil war risk and that the participation sub-component alone drives the anocracy effect on internal conflict. The aggregate transformation is retained here for three reasons. The Polity V dataset provides aggregate scores with continuous annual coverage for all eleven countries across the window, whereas sub-component series have more frequent gaps. For standardisation against MENA historical baselines, the aggregate score supplies a consistent one-dimensional input without requiring separate reference parameters for each sub-component. And the aggregate preserves comparability with Document III, which used the same transformation. The Vreeland decomposition is identified as a methodological improvement for the fifteen-variable index specified in Document II, pending fuller sub-component coverage.

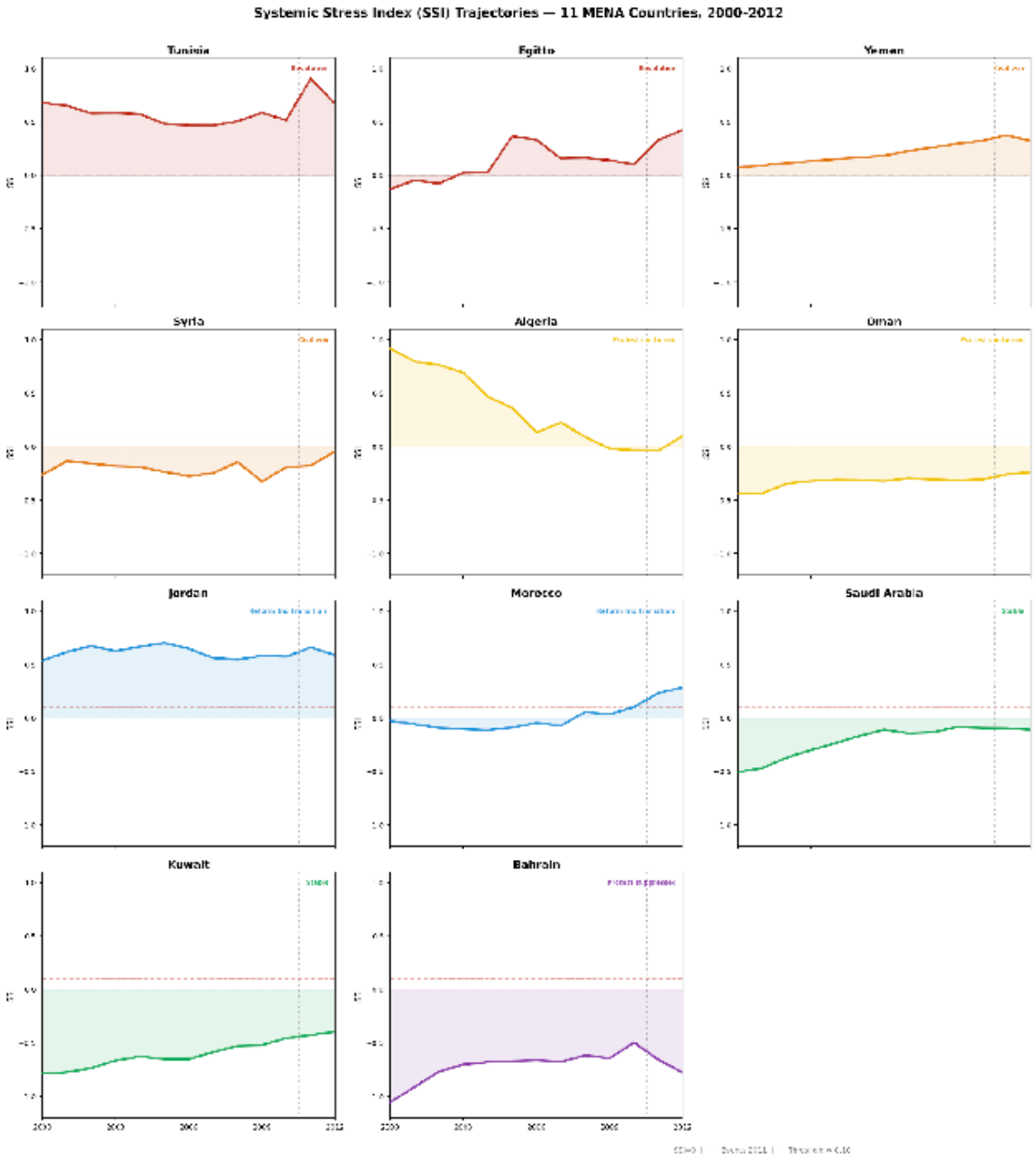

Figure 1: SSI trajectories for eleven MENA countries, 2000 to 2012. Colours indicate outcome category. The dashed horizontal line marks $SSI = 0$, the dotted vertical line marks event onset in 2011, and the dash-dot line marks the classification threshold at approximately 0.10.

Among the instability countries, Tunisia displays the most clearly elevated trajectory, holding above 0.45 throughout the decade and peaking at 0.904 in 2011, the post-event year in which youth unemployment surged following the transition. Yemen drifts steadily upward, from 0.074 in 2000 to 0.324 in 2010, driven by the gradual accumulation of demographic and economic stress. Egypt presents a more mixed picture, fluctuating between -0.130 and 0.365 over the decade, reflecting the incomplete anocracy signal and the moderate Gini coefficient discussed in Section 7. Syria's trajectory is negative throughout, an anomaly examined in Section 7.3.

Among the stability cases, Algeria displays the most dramatic movement: an index of 0.918 in 2000, the highest value in the entire panel, declining to -0.034 by 2010 as youth unemployment fell substantially during the post-conflict economic recovery of the mid-2000s. The trajectory illustrates the point: high structural stress is a necessary but not sufficient condition for instability,

and declining stress can remove the conditions that make mobilisation likely even from an elevated baseline. The Gulf states, Bahrain, Kuwait, and Saudi Arabia, together with Oman, cluster in strongly negative territory throughout the period, consistent with their oil-funded co-optation capacity, low youth unemployment, and, in Bahrain and Kuwait, high internet penetration that works against the index's instability signal.

4.2 Group-level trajectories

Figure 2 presents mean trajectories for the instability group, Tunisia, Egypt, Yemen, and Syria ($N = 4$), and the stability group, the remaining seven countries, with one-standard-deviation bands. Group separation is visible throughout the period and becomes most consistently pronounced over 2005 to 2010, the years in which the structural preconditions of the Arab Spring were intensifying. At the 2010 benchmark the instability group mean stands at 0.188 against −0.102 for the stability group, a separation of 0.290 index units.

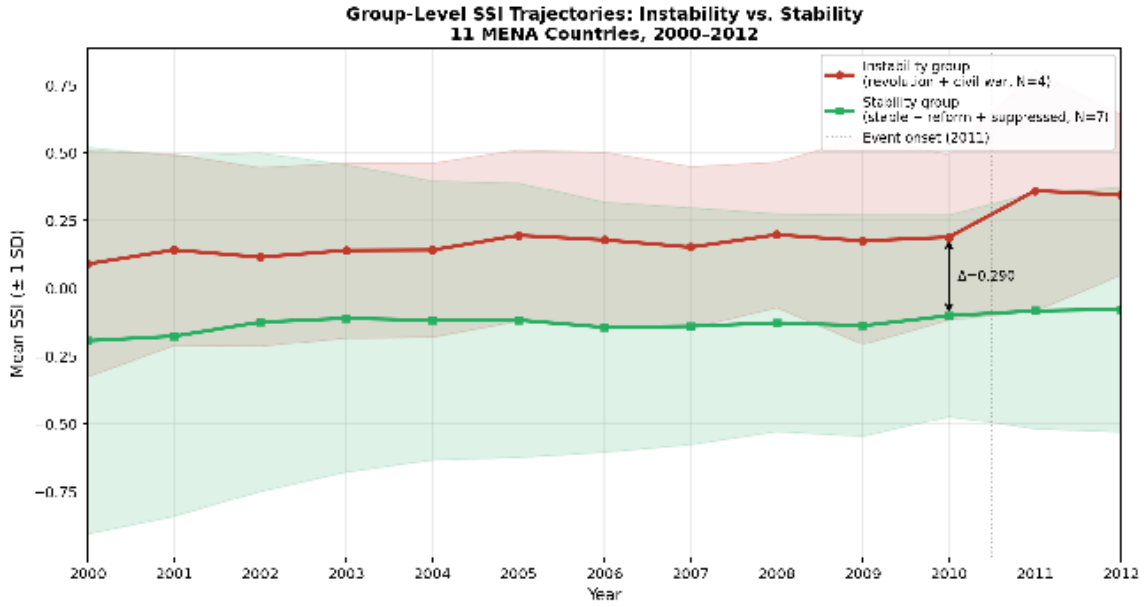

Figure 2: Group-level mean SSI trajectories, 2000 to 2012. Red marks the instability group, revolution and civil war; green marks the stability group. Shaded bands indicate one standard deviation within each group; the arrow indicates group separation at 2010.

4.3 The 2010 pre-event benchmark

Table 3 and Figure 3 present the index values for all eleven countries at 2010, ranked in descending order and coded by outcome. The distribution falls into three zones. All five countries above 0.10, Jordan, Tunisia, Yemen, Morocco, and Egypt, experienced protest or transition, though protest also occurred below the threshold, in Algeria and Oman, so the zone marks a concentration and not a boundary of occurrence. Countries between −0.10 and 0.10, Algeria and Saudi Arabia, form a transitional band. Countries below −0.10, Syria, Oman, Kuwait, and Bahrain, present either structural stability or, in Syria's case, the anomaly of Section 7.3.

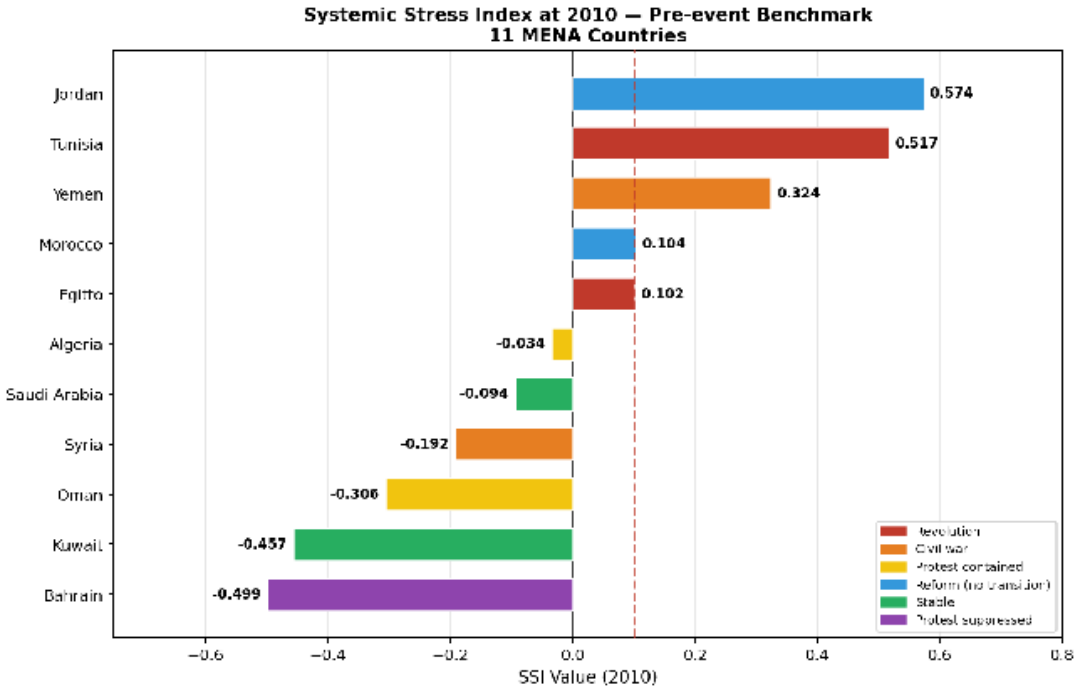

Figure 3: SSI rankings at 2010. Horizontal bars are coloured by outcome; the dashed line marks the classification threshold at approximately 0.10.

Table 3: SSI at 2010: country rankings and outcome classification.

Country	**SSI (2010)**	**Outcome**	**Mean SSI, 2000 to 2010**	**Binary**
Jordan	0.574	Reform avoided	0.613	0
Tunisia	0.517	Revolution	0.554	1
Yemen	0.324	Civil war	0.184	1
Morocco	0.104	Reform avoided	−0.037	0
Egypt	0.102	Revolution	0.096	1
Algeria	−0.034	Protest contained	0.400	0
Saudi Arabia	−0.094	Stable	−0.238	0
Syria	−0.192	Civil war	−0.214	1
Oman	−0.306	Protest contained	−0.337	0
Kuwait	−0.457	Stable	−0.636	0
Bahrain	−0.499	Protest suppressed	−0.718	0

Note: Binary = 1 for revolution or civil war, 0 otherwise. Mean SSI is computed over the pre-event window 2000 to 2010.

Figure 4 decomposes each country's 2010 index into its five weighted component contributions, exposing the structural drivers behind each score. Youth unemployment is the dominant positive contributor among the highest-stress countries, accounting for 0.377 index units in Jordan, 0.362 in Tunisia, and 0.206 in Egypt. Anocracy stress is the critical differentiator: Yemen and Algeria both carry the highest contribution observed in the cross-section, 0.250 against a theoretical maximum of 0.417, despite moderate overall levels, while the Gulf monarchies and Syria, with consolidated Polity scores, show negative contributions on that component. The pair illustrates what the symmetry of the transformation does and does not assert. Algeria in 2010 scores

$P_1 = +2$ and Yemen $P_1 = -2$, so the two receive identical institutional vulnerability despite sitting on opposite sides of the democratic-autocratic axis. The functional form treats distance from the centre as the vulnerability-relevant quantity and direction as immaterial, which is the anocracy hypothesis stated strictly. Whether that indifference survives decomposition of the Polity aggregate is the question Note 6 defers.

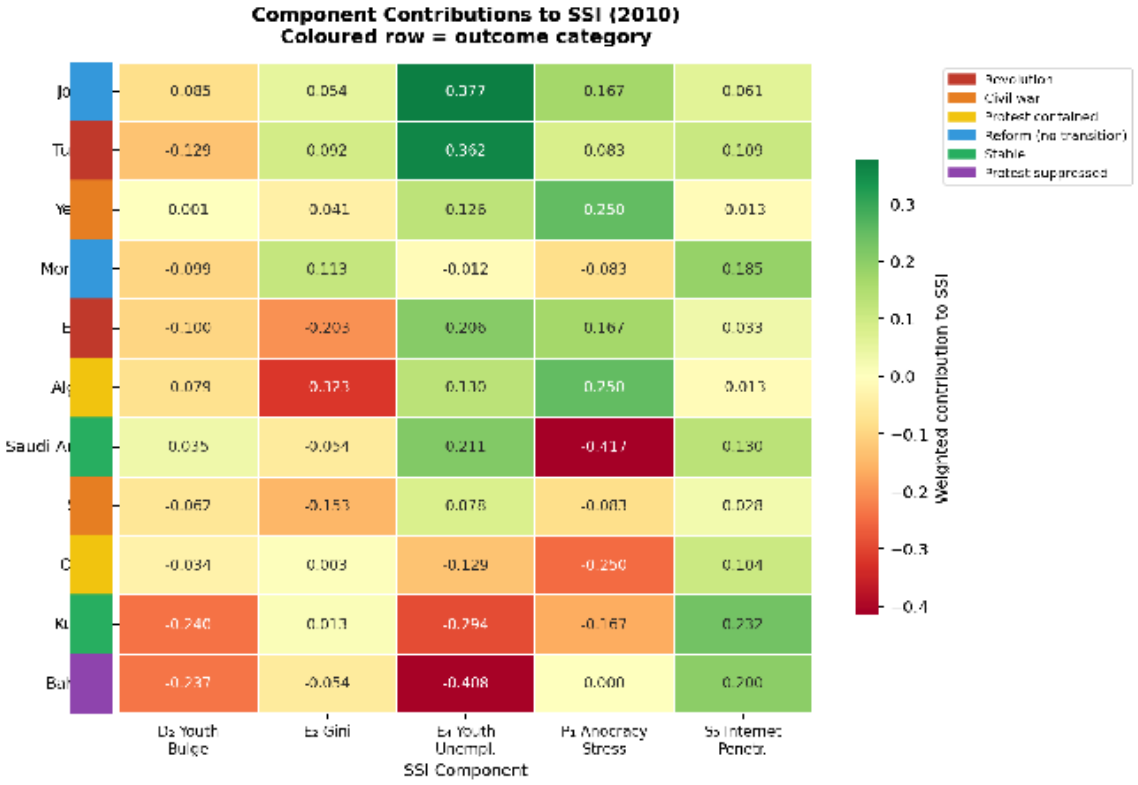

Figure 4: Weighted component contributions to the 2010 SSI. Warm colours indicate positive contributions to instability stress, cool colours negative contributions. Countries are ordered by SSI ranking as in Table 3; row shading indicates outcome category.

5 Threshold analysis

To assess discriminative capacity systematically we scan all candidate thresholds derived from the observed 2010 distribution. For each threshold τ, countries with $SSI \geq \tau$ are classified as instability-predicted and the remainder as stability-predicted, and accuracy is computed against the binary outcome.

The scan returns a maximum descriptive accuracy of 72.7%, eight correct classifications out of eleven, achieved across two distinct intervals, $\tau \in [-0.033, 0.102]$ and $\tau \in [0.105, 0.323]$. The two intervals produce different classification matrices: the lower one records Egypt as a true positive and Morocco as a false positive, while the upper one records Morocco as a true negative and moves Egypt to the false negatives. Both yield identical overall accuracy. Between the intervals, at $\tau \approx 0.103$, accuracy falls to 63.6%, seven of eleven, because Egypt at 0.102 and Morocco at 0.104 straddle the boundary. We adopt $\tau \approx 0.10$ as the canonical reference because it falls in the lower interval, where Egypt is classified as instability, the assignment consistent with the observed outcome.[7] The threshold is a descriptive classification boundary within this sample, not

[7]Maximum accuracy of eight of eleven is achieved across two intervals, $\tau \in [-0.033, 0.102]$ and $\tau \in [0.105, 0.323]$. Between them, at $\tau \approx 0.103$, accuracy falls to seven of eleven because Egypt at 0.102 and Morocco at 0.104 straddle the boundary. The value $\tau \approx 0.10$ is adopted within the lower interval because it classifies Egypt as instability,

a predictive parameter, and its stability across the range from -0.033 to 0.323 is itself informative about the discriminative structure of the data.

At $\tau \approx 0.10$ the index returns three true positives, Tunisia, Yemen, and Egypt; five true negatives, Algeria, Saudi Arabia, Oman, Kuwait, and Bahrain; two false positives, Jordan and Morocco; and one false negative, Syria. Table 4 gives the classification matrix.

Table 4: Classification matrix at threshold $\tau \approx 0.10$.

	Predicted instability	**Predicted stability**
Actual instability ($N = 4$)	TP = 3 (Tunisia, Yemen, Egypt)	FN = 1 (Syria)
Actual stability ($N = 7$)	FP = 2 (Jordan, Morocco)	TN = 5 (Algeria, Saudi Arabia, Oman, Kuwait, Bahrain)

Note: Accuracy = 8/11 = 72.7%. At $N = 11$ the figure is descriptive only. No p-values or confidence intervals are reported; see Section 8.1.

Two naive baselines are conventionally cited at this point. A random classifier assigning the instability label with probability 4/11 achieves an expected accuracy of 65/121, about 53.7%, and a majority-class classifier that always predicts stability achieves 7/11 = 63.6%, exploiting only the class imbalance. Neither comparison is admissible for the figure reported above, because the threshold that produces eight of eleven was selected by scanning the sample while both baselines are threshold-free. The appropriate null for a threshold-optimised statistic is the distribution of that same statistic under random assignment of the outcome labels, and with 330 possible assignments of four instability labels across eleven countries it is enumerable exactly.

Under that null, a threshold chosen to maximise accuracy reaches eight of eleven or better in 165 of the 330 assignments. The observed accuracy therefore sits precisely at the median of its own null distribution, and taken alone it separates the index from chance not at all. Two statistics that do not depend on in-sample threshold selection carry the discrimination instead. Holding the threshold fixed at 0.10 in advance, eight of eleven or better occurs in 65 of the 330 assignments, an exact one-sided probability of 0.197. The rank-ordering statistic, 20 of 28 correct instability-stability pairs and an area under the curve of 0.714 against a chance expectation of 14 of 28,[8] occurs in 52 of the 330 assignments, an exact one-sided probability of 0.158. We therefore take pairwise rank ordering as the primary discrimination measure throughout and read the 72.7% figure as a descriptive summary of the classification matrix. At $N = 11$ neither probability is presented as a hypothesis test; both are reported because the enumeration is exact, and omitting it would leave the threshold statistic looking stronger than it is.

The three misclassifications carry distinct theoretical interpretations. Jordan's 0.574, the highest value in the sample, reflects genuine structural stress: youth unemployment at 30.1%, a Gini of 41.7, and an anocratic Polity score. That Jordan did not experience revolution is attributable to factors outside the current specification, among them the Hashemite monarchy's preemptive reform strategy, substantial Gulf financial transfers, and the country's geopolitical position as a buffer state. Egypt sits at the threshold itself, at 0.102, illustrating the boundary ambiguity inherent in any scalar classification. Syria presents the most revealing anomaly and is treated in Section 7.3.

consistent with the observed outcome. The choice is descriptively motivated and affects no substantive finding.

[8]The correct-pairs metric counts, across the twenty-eight instability-stability country pairs, the cases in which the instability country carries the higher index at 2010. A value of 28/28 indicates perfect rank ordering and 14/28 corresponds to chance. The metric is formally equivalent to the Wilcoxon-Mann-Whitney U statistic normalised by the number of pairs, which is in turn equal to the area under the receiver operating characteristic curve. It is the appropriate non-parametric discrimination measure for small samples in which parametric estimation of the area is unreliable. The baseline achieves 20/28 = 0.714, interpretable as follows: given a randomly selected instability-stability pair, the index ranks the instability country above the stability country in 71.4% of cases.

6 Robustness analysis

6.1 Alternative weighting schemes

We evaluate five alternative weighting schemes alongside the baseline to establish whether the results depend on the particular weight assignment. The schemes, labelled equal, economic, regime, demographic, and coordination, appear in Table 5. All satisfy the constraint that the weights sum to one.

Table 5: Alternative weighting schemes and discrimination metrics.

Scheme	$w(D_2)$	$w(E_2)$	$w(E_4)$	$w(A)$	$w(S_3)$	**Separation**	**Correct pairs**
Baseline	**0.15**	**0.25**	**0.25**	**0.25**	**0.10**	**0.290**	**20/28**
Equal	0.20	0.20	0.20	0.20	0.20	0.142	17/28
Economic	0.08	0.35	0.35	0.15	0.07	0.298	20/28
Regime	0.10	0.18	0.18	0.40	0.14	0.301	22/28
Demographic	0.40	0.15	0.25	0.10	0.10	0.256	18/28
Coordination	0.10	0.20	0.20	0.20	0.30	0.031	15/28

Note: Separation is the difference between the instability-group and stability-group mean index at 2010. Correct pairs counts the instability-stability pairs in which the instability country carries the higher index, out of $4 \times 7 = 28$. The baseline is shown in bold for reference.

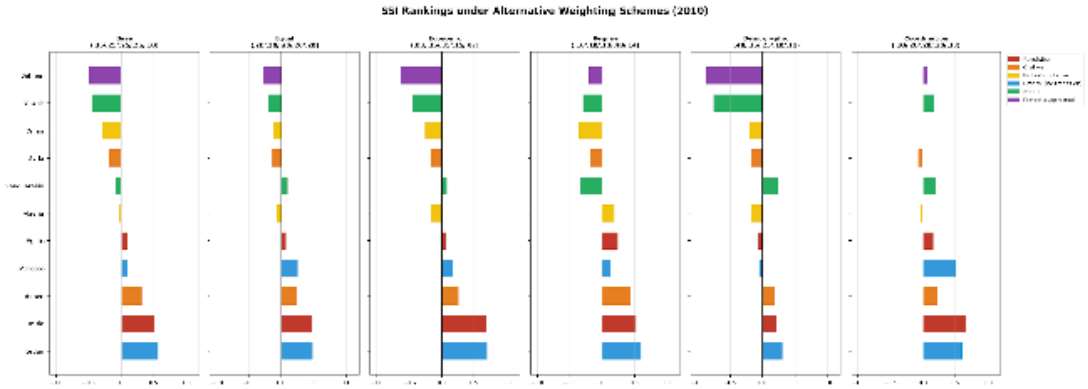

Figure 5: SSI rankings under six weighting schemes at 2010. Each panel shows one weight configuration. Country order is held constant across panels by baseline ranking, descending, so that each scheme's reordering of the same countries is directly visible. Colours indicate outcome category.

Two findings warrant attention. The coordination scheme, which assigns thirty per cent of total weight to internet penetration, produces the worst discrimination in the panel, a separation of 0.031 and fifteen of twenty-eight correct pairs, indistinguishable from chance. High internet penetration correlates strongly with income in this sample: Kuwait at 61.4%, Bahrain at 55.0%, and Saudi Arabia at 41.0% were among the region's most connected states by 2010, and none experienced instability. Under heavy internet weighting the index conflates diffusion capacity with structural privilege, reversing the intended causal mechanism. This is a useful negative result. Internet penetration alone is not a reliable indicator of instability in this sample, a finding that aligns with the scholarly scepticism about simple internet-mobilisation arguments that has developed since 2011 and that suggests the variable is a confounder rather than a predictor in high-income authoritarian contexts, at least in its current single-variable operationalisation.

A mechanical feature of the standardisation compounds the substantive confound. Two of the five components are non-stationary over their reference windows in ways that leave their standardised values nearly uniform across the 2010 cross-section, and in opposite directions. Internet penetration is standardised against a mean of 15 and a dispersion of 20 drawn from a window, 2000 to 2010, that coincides with the steepest phase of its diffusion; a terminal-year cross-section drawn from a decade of monotone growth necessarily sits in the upper tail of its own reference distribution, and the standardised values at 2010 accordingly run from -0.132 to 2.320 with a cross-sectional mean of 0.961. The youth bulge is the mirror image over a much longer window, 1960 to 2010, whose MENA peak falls in the 1980s and 1990s: at 2010 the standardised values reach a maximum of 0.233 with a mean of -0.623. Both components accordingly enter the cross-section as near-constant offsets with little discriminative content, and the connectivity offset is negatively correlated with each of the three stress components, at -0.617 with the youth bulge, -0.619 with youth unemployment and -0.603 with anocracy stress. Axiom A2 conditions historical comparability on the absence of structural breaks between the reference window and the period analysed; for S_3 the break is internal to the reference window itself, since the window coincides with the diffusion event, and the condition accordingly fails by construction. The point is distinct from the income confound: the confound explains why connectivity points the wrong way, while the reference window explains why it does so with almost no variance to offer in exchange.

The regime scheme, which places high weight on anocracy stress, achieves the best pairwise discrimination at twenty-two of twenty-eight while holding separation comparable to the baseline. This is consistent with the theoretical expectation that institutional vulnerability is the most proximate driver of revolutionary onset among the variables currently in the model. Its weakness is symmetrical: it penalises consolidated autocracies heavily, so that countries such as Syria, where institutional consolidation masked deep fragility, are misclassified even further under it.

6.2 Leave-one-component-out analysis

We assess the marginal contribution of each component by excluding it in turn and rescaling the remaining weights proportionally to sum to one. Figure 6 reports group separation and correct pairwise ordering under each exclusion.

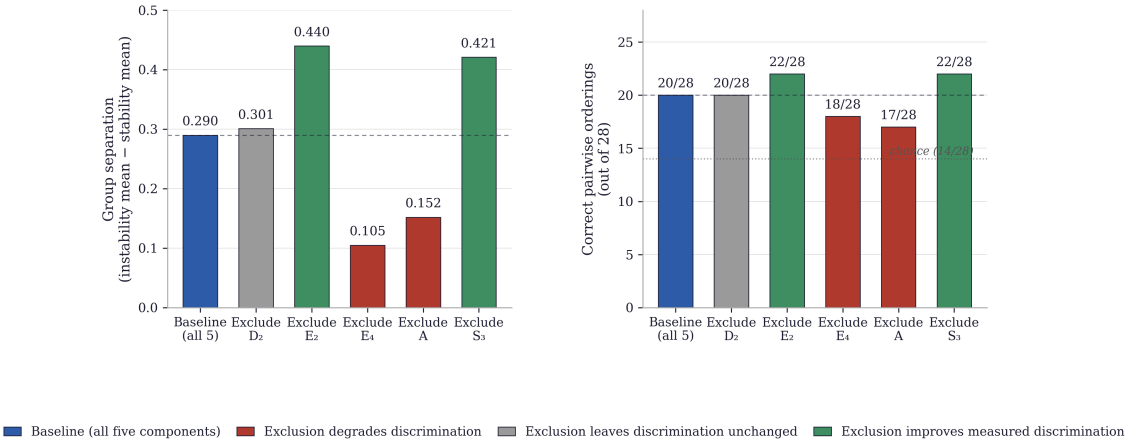

Figure 6: Leave-one-component-out analysis, all five components. The left panel reports the group separation metric, the right panel correct pairwise rank ordering out of 28, with the chance benchmark of 14/28 marked. Blue marks the baseline; red marks exclusions that substantially degrade discrimination (E_4, A); green marks exclusions that improve measured discrimination (E_2, S_3); grey marks the exclusion that leaves discrimination materially unchanged (D_2).

The pattern is sharply asymmetric across the five components. Excluding youth unemployment

reduces separation from 0.290 to 0.105, a fall of 64 per cent, and correct pairs from twenty to eighteen. Excluding anocracy stress reduces separation to 0.152 and correct pairs to seventeen, the lowest value across all exclusions. Excluding the youth bulge leaves correct pairs at twenty and raises separation marginally to 0.301. Excluding the Gini coefficient or internet penetration improves measured discrimination in both cases to twenty-two correct pairs, with separation rising to 0.440 and 0.421 respectively.

Read together with the single-component ordering statistics, this locates the discrimination narrowly. Youth unemployment and anocracy stress each rank nineteen of twenty-eight pairs correctly on their own and twenty-one in combination, counting strict wins only; the anocracy component alone produces three tied pairs, which would raise its count to 20.5 under a half-credit convention, and no other single-component count is affected by ties. The youth bulge ranks fifteen, which is chance. The Gini ranks eleven and internet penetration five, both below chance and the latter substantially so. Half the index weight therefore carries the discrimination and the other half does not, and the five-component baseline at twenty of twenty-eight is outperformed within its own robustness analysis by three simpler specifications. We retain the baseline because its weights were fixed by theory before contact with the data. Adopting a better-performing variant after seeing the outcomes would reproduce precisely the calibration circularity that Section 2.2 identifies in expert-weighted alternatives and that Axiom A8 exists to prevent. The honest statement of the result is accordingly narrower than robustness: the specification contains two components that discriminate and three that do not, and the pre-specification discipline is what stops that finding from being converted into a better headline figure.

Read through the sign convention of Section 3.3, the ordering statistic of five of twenty-eight is a rejection of the assignment $s_{S_3} = +1$ in this sample. The component orders the cases, and orders them the wrong way. Whether the correct treatment is $s_{S_3} = -1$ or exclusion altogether cannot be settled on eleven cases, and the question is referred to the expanded specification.

The Gini result reflects two features of the component. Its variation across the MENA sample is limited relative to the reference distribution, with most countries clustered between 33 and 44 and generating weak standardised values. More fundamentally, and as Section 8.2 documents, the series carries almost no independent annual information. The internet penetration result echoes the weighting analysis: with S_3 excluded, the Gulf states' inflated connectivity scores no longer dampen the discrimination between high-stress and low-stress countries.

These results do not establish that inequality and connectivity are theoretically irrelevant. They indicate that the operationalisation of both in the current five-variable specification fails to capture the mechanisms they are meant to represent. The Gini finding receives independent corroboration from Korotayev et al. (2014), who found that poverty levels and corruption, variables closely related to distributional outcomes, showed no significant correlation with destabilisation in their nineteen-country Arab sample and were consequently dropped from their index. Their explanation parallels ours: income distribution across MENA countries is relatively compressed by global standards and comparable to several Western European states, which limits discriminative power within a regionally homogeneous sample. The convergence of two methodologically distinct approaches, one multiplicative and expert-calibrated, the other additive and theory-derived, on the same finding strengthens the reading that the problem lies in how inequality is measured and not in its theoretical relevance. The Gini captures the static distribution of income at a point in time but not the dynamics that matter for grievance mobilisation, specifically the perceived trajectory of economic status and the experience of downward social mobility among educated youth. A measure of Gini change or of downward mobility among educated youth might perform substantially better, and both are directions for the expanded programme.

A further and more fundamental factor in E_2's weak performance is the nature of the source series, documented in Section 8.2. The Gini component is drawn from a model-imputed compilation with complete country-year coverage, and for nine of the eleven countries the 2010 value on which the benchmark classification rests is an imputed figure and not an anchor observation.

Kuwait carries a single constant value across all thirteen years, Syria a constant value from 2007 onward, and Bahrain and Saudi Arabia share an identical series in every year. A component constructed as a smooth function of two or three anchor points per country cannot by construction carry information about a single event year, which locates the weak performance of E_2 in the measurement and leaves the causal role of inequality untested in this study.

7 Anomalous cases and the limits of the current specification

Three countries are misclassified at $\tau \approx 0.10$: Jordan as a false positive, Egypt as a marginal case correctly classified only by a narrow margin, and Syria as a false negative. Each carries distinct theoretical implications, discussed here in ascending order of analytical significance.

7.1 Egypt: threshold sensitivity and deterioration trajectories

Egypt's 0.102 places it at the threshold boundary, the second-lowest value among the four instability countries and the fifth-highest overall. Its correct classification is numerically contingent: a threshold of 0.103 would place it among the stable. The boundary position reflects two structural features that distinguish Egypt from Tunisia. Its Gini coefficient was substantially lower, 33.5 against 43.0 for Tunisia in 2010, pulling the E_2 contribution to -0.203. Its Polity score held at -3 from 2005 onward, a partial anocracy classification short of the deeper authoritarian consolidation of Saudi Arabia or Syria.

The index cannot capture the trajectory of deterioration, and the Egyptian series illustrates the limitation concretely. The index reads 0.163 in 2008, 0.141 in 2009, and 0.102 in 2010: on its own measurements, Egypt was moving away from the instability zone in the years immediately preceding the uprising. That decline is not evidence of improving conditions. Inflation, food prices, and informal-sector unemployment rose sharply over 2009 and 2010 in ways that annual Gini and youth unemployment statistics capture only partially, so the components that would have registered the deterioration are either absent from the specification or measured too coarsely to move. A first-difference augmentation of the index, adding ΔSSI as a second classification criterion, would therefore not have rescued the Egyptian case: the deterioration that mattered was invisible to the index in levels and equally invisible in differences.[9] Moving to a dynamic specification requires adding variables whose derivatives carry the relevant signal, rather than appending a derivative term to the existing five.

The Mubarak regime's succession crisis, which became acute after 2008, introduced a political vulnerability, elite fragmentation and uncertainty about institutional continuity, that has no analogue in the current variable set. The catalytic dimension that Gaub (2012) identifies illustrates the same gap: food price inflation reached approximately nineteen per cent in 2010, with dairy, fats, and fruit prices rising between January 2010 and January 2011 from declining rates to increases of twenty-seven, twenty-eight, and fourteen per cent respectively. That accelerating deterioration operated on a population that had experienced prior improvement, poverty rates having fallen to 16.7 per cent in 2000 before rising again after 2005, generating the frustrated-expectations dynamic that structural-demographic theory identifies as a primary mobilisation driver. Survey evidence collected in the months following Mubarak's fall is consistent with the weighting the specification places on economic components, with the caveat that its timing makes it a retrospective account rather than a pre-event measurement. In the second wave of the Arab

[9] Formally, a first-difference augmentation would define $\Delta SSI(t) = SSI(t) - SSI(t-1)$ and apply a dynamic classification rule of the form $SSI(t) \geq \tau$ or $[SSI(t) \geq \tau - \delta$ and $\Delta SSI(t) > 0]$, where δ is a tolerance margin and the second clause captures countries approaching the threshold on a deteriorating trajectory. The Egyptian series over 2007 to 2010 reads 0.158, 0.163, 0.141, 0.102, so $\Delta SSI(2009 \text{ to } 2010) = -0.039$ and $\Delta SSI(2007 \text{ to } 2010) = -0.056$. The second clause never fires, and the augmented rule reproduces the level-based classification exactly. The deterioration documented in Section 7.1 is therefore invisible to the index in differences as well as in levels, which locates the deficiency in the variable set rather than in the static functional form.

Barometer, fielded in Egypt in June 2011, the state of the economy was named the country's principal challenge by 82 per cent of respondents against 7 per cent for security and for corruption, and three quarters located the essential feature of democracy in socio-economic provision and not in political rights, with 6 per cent naming free elections (Tessler, Jamal, and Robbins 2012). A Pew survey fielded in late March and early April 2011 found 82 per cent rating improved economic conditions as very important against 55 per cent for regular honest elections, though a fair judiciary rated almost as highly as the economy at 79 per cent, and a forced choice between a good democracy and a strong economy split the sample evenly at 47 to 49 (Pew Research Center 2011). Axiom A6 applies to the evidence as much as to the object: attitudes measured after a successful uprising are shaped by it, and the Arab Barometer report notes the resulting optimism in its own responses. The individual-level record therefore illustrates the specification's economic emphasis without independently confirming it. Egypt's correct classification is therefore partly fortuitous, and its analytical value lies in identifying the variables a more complete specification would need.

7.2 Jordan: structural stress contained by institutional capacity

Jordan is perhaps the most theoretically clarifying case in the sample. With the highest index value at 2010, 0.574, and the highest pre-event decade mean, 0.613, Jordan registers greater structural stress than Tunisia on the demographic, labour-market and institutional components, and lower stress on inequality and connectivity, yet experienced only limited protest and no regime transition. A superficial reading would treat this as model failure. A more careful reading treats it as an informative limit condition.

The index correctly identifies that Jordan was under severe structural stress. Youth unemployment at 30.1% was among the highest in the sample, and the anocracy score placed the regime among the institutionally most vulnerable. What the index cannot model is the set of mechanisms through which the Jordanian monarchy absorbed and contained that stress: a preemptive reform strategy, with constitutional amendments announced before protest movements could coalesce and a cabinet reshuffle in February 2011; substantial financial transfers pledged by Gulf Cooperation Council states during 2011; and a form of Hashemite legitimacy drawing on both tribal and religious authority in ways that cross-national institutional measures do not register. A regional demonstration effect also operated: the Jordanian state had time to observe the Egyptian and Tunisian trajectories and adjust its response, a form of strategic reflexivity that Axiom A6 identifies as a source of indeterminacy in structural predictions. The case receives indirect corroboration from Korotayev et al. (2014), whose index assigns Jordan a moderate instability score despite high conflict potential and youth pressure, because their regime-capacity and historical-immunity sub-indicators capture the containment mechanisms the SSI omits. Both indices identify Jordan as structurally stressed; the multiplicative index, by including qualitative moderators, produces a score more consistent with the observed outcome. Adding a state capacity variable would accordingly improve classification disproportionately for cases of this type.

Jordan is a theoretically clarifying case. Structural stress reached levels that produced revolution in other institutional contexts and was managed. A structural model identifies vulnerability; outcome depends on factors the model does not capture. Structural stress is necessary but insufficient, and institutional capacity can contain severe pressure when competently deployed.

7.3 Syria: the critical anomaly

Syria is the most analytically challenging case and the one that most directly exposes the limits of the current specification. With an index of -0.192 at 2010, suggesting low structural stress relative to the MENA historical baseline, Syria went on to experience the most catastrophic outcome in the sample: a civil war that by late 2015 had killed more than 250,000 people by

the estimate then in use at the United Nations, displaced close to half the country's population between refugees and the internally displaced, and drawn direct military intervention from at least six foreign states, Russia, Iran, the United States, Turkey, Saudi Arabia, and Qatar (United Nations General Assembly 2015; Blanchard, Humud, and Nikitin 2015).

The anomaly has two dimensions. The first is data quality. Syria's Gini in the panel is an imputed series, rising from 37.7 in 2000 to a maximum of 38.7 in 2004, declining to 35.1 by 2007, and then constant at that value through 2012. The six-year plateau is an artefact of the imputation model described in Section 8.2 and not an observation of stable inequality, and the underlying survey record was in any case collected under conditions that made accurate income measurement unreliable in a state with substantial off-books patronage flows. There are strong grounds to infer that the figures understate actual inequality, particularly after the economic liberalisation programme of the mid-2000s that enriched regime-connected business elites while cutting agricultural and industrial subsidies, though the inference rests on the qualitative record and not on any measurement available here. The distribution of economic gains was intensely regional and sectarian: the western cities benefited substantially while the eastern, predominantly Sunni rural periphery, which became the cradle of the uprising, experienced sustained marginalisation that aggregate Gini statistics cannot capture. The limitation is not unique to Syria. Any state with substantial data opacity and off-books income flows will be systematically underrepresented by components that depend on official survey data, and the index should be applied with particular caution to authoritarian states whose official statistics are likely to be unreliable.

Beyond the measurement problem, Syria experienced a catalytic environmental shock invisible to the index. A drought entering its fifth consecutive year by 2010 had halved the country's water resources between 2002 and 2008, pushing an estimated two to three million people into extreme poverty and driving large-scale internal migration from rural areas into cities already under economic strain (Gaub 2012). Syria, which had maintained agricultural self-sufficiency, was suddenly unable to ensure food security, and fruit and vegetable prices rose by approximately twenty-seven per cent in the year preceding January 2010, an inflation that continued to rise through the remainder of the year (Gaub 2012). These environmental, migratory, and inflationary dynamics operated beneath the resolution of the index's annual aggregate variables, compounding the measurement problem with a genuine structural omission.

The second dimension is theoretical. The Assad regime's Polity score of -6 in 2010 classifies it as a consolidated authoritarian state, yielding an anocracy stress of 0.4 and a negative contribution to the index. The classification is formally correct and substantively misleading. The Syrian Ba'ath regime was not a stable, institutionalised autocracy in the manner of the Saudi monarchy or the pre-reform Kuwaiti state. It was a second-generation personalised regime that had consolidated power through sectarian networks, in particular Alawi military dominance, pervasive internal surveillance, and the systematic destruction of the civil society institutions that might otherwise have served as organisational infrastructure for opposition. A regime of this type, internally hollowed while externally presenting as consolidated, is in fact highly brittle: it suppresses the visible signs of instability while accumulating the conditions for catastrophic failure.

Syria's misclassification is therefore a systematic gap in the specification and not random error. The index has no variable for coercive capacity, regime internal cohesion, or the vulnerability of personalised autocracies to elite defection. Adding a measure of military expenditure as a share of GDP, or of the degree of personalisation of power drawing on the regime-type coding of Geddes, Wright, and Frantz (2014) that distinguishes personalist from single-party and military autocracies, would likely improve performance substantially on cases of this type. The extension is methodologically tractable rather than theoretically insurmountable.

8 Limitations and the path to statistical validation

8.1 Sample size and inferential constraints

The most visible constraint is $N = 11$. At this sample size the standard apparatus of inference is unavailable: confidence intervals and logistic regression coefficients are either undefined or too imprecise to inform. Exact permutation probabilities are computable, and Section 5 reports them, but they rest on an enumeration of 330 assignments and do not substitute for a sample large enough to estimate a distribution. The 72.7% figure is a descriptive statement about classification in this sample. It is useful to fix a vocabulary for the epistemic status of empirical work under the framework, and we propose one here; on that classification the present analysis is a Type B exploratory application, and the falsifiability requirement of Axiom A8 is met in the form appropriate to that type, discussed below.[10]

A deeper constraint follows from the axioms and not from the sample. Document I in its third version derives a sample-size condition for the statistical regularity that Axiom A1 asserts, and shows that achievable precision is bounded below by a variance floor proportional to the intra-class correlation among the units under study, with the required sample diverging as that correlation approaches a herding threshold. The Arab Spring is a regional cascade in which the eleven units are neither independent nor weakly coupled, and the mechanism of preference revelation described in Note 2 is that high-correlation regime stated explicitly. The limitation is therefore not only that eleven cases are few. A cascade of this kind admits a floor on achievable precision that further cases drawn from the same episode would not lower, which means expansion must diversify across episodes and periods and not merely add countries to this one.

A second consequence concerns the form in which compliance with Axiom A8 can be claimed. Document I version 3.0 makes the requirement operationally specific, assessing compliance through the decomposition of a probability forecast into reliability, resolution and uncertainty alongside the aggregate Brier score. That apparatus does not apply to the present output. The SSI is a scalar position in a reduced state space and not a calibrated probability, and mapping it into the unit interval would require estimating the transformation on the same eleven cases used for assessment, which is the circularity that Section 3.3 rejects in its choice of reference parameters. What A8 requires of a Type B study, and what this study delivers, is the weaker and well-defined obligation that the index formula, the component weights and the classification rule be fixed and stated before results are reported, as they are in Section 3.3, and that the analysis be transparent about its epistemological status. The calibration decomposition belongs to the validation study of at least thirty cases, where a probability model can be fitted on one partition and assessed on another.

Located in the three-level prediction hierarchy that Document II adopts (Angeli 2026b), the discrimination reported here is a Level II result. The Roman numeral matters: this hierarchy grades the epistemic status of a claim and is distinct from the Arabic-numbered ladder of operationalisation levels invoked above, which grades the tractability of an implementation. A Level II claim concerns the probability that a system occupies an instability-prone regime over a window of several years, and makes no claim at the level of individual event prediction, which the framework places beyond reach for reasons Axiom A7 states.

The path to statistical validation requires a minimum of thirty validated country-cases with

[10] We introduce here a three-part classification of study types for empirical applications of the framework, offered as a series convention. Type A studies are theoretical derivations or formal proofs requiring no empirical data; Documents I and II are of this kind. Type B studies are exploratory empirical applications with fewer than thirty country-cases; they may produce descriptive results and generate hypotheses but cannot support statistical inference or falsification claims under Axiom A8. Type C studies are statistical validation exercises with at least thirty country-cases, complete variable coverage, and documented outcomes, and constitute the minimum standard for falsifiability claims. This study is Type B by design, positioned as the intermediate step between the three-country proof of concept, also Type B, and the future Type C validation study. The thirty-case threshold separating Type B from Type C is the one Document II fixes for the fifteen-variable validation panel.

complete data across all five variables and documented outcomes. The most tractable expansion strategy adds historical cases of structural-demographic crisis outside the MENA region: sub-Saharan African state collapses in the 1990s, post-Soviet political transitions between 1989 and 1993, and Latin American democratic breakdowns in the 1970s. A broader historical dataset would also constitute a more stringent test, since it would require the index to generalise across diverse regional and temporal contexts instead of a single regional episode. Each expansion introduces data quality challenges and forces the reference-parameter question, whether MENA-specific baselines should give way to global historical averages as the sample diversifies. That question is an instance of Axiom A2, which conditions historical comparability on the absence of structural breaks between the reference window and the period analysed, and Section 6.1 shows that the condition already fails within the present sample for one component.

8.2 Data quality

The Gini component requires a fuller account than the other four. The series is drawn from the UNU-WIDER WIID Companion dataset, a product distinct from the WIID proper: where the WIID reports survey observations with the gaps they leave, the Companion supplies complete annual country coverage by model imputation. Every value of E_2 in this panel is therefore an estimate produced by that model, and the panel's apparent completeness on this component is a property of the source and not of the underlying survey record.

Three consequences follow. First, for nine of the eleven countries the 2010 value that enters the benchmark classification lies on a smooth interpolation between two or three anchor points and not at an anchor. Algeria's thirteen-year series is a single straight line, Tunisia's is two segments meeting at 2006, Kuwait carries one constant value throughout, and Syria is constant from 2007 to 2012. Second, Bahrain and Saudi Arabia carry identical values in every year of the panel to six decimal places, which indicates that the imputation assigns the two states the same predicted series, neither having usable survey coverage. Third, the component's temporal variation is largely deterministic, which bears directly on the leave-one-component-out result of Section 6.2 and supplies its mechanism.

We report the component as the source constructs it and have verified that the classification, the threshold intervals and the rank ordering are insensitive to substantial perturbation of the affected series. The Bahraini Gini would have to rise by roughly ten points to alter a single pairwise ordering and by nineteen to alter its own classification, and neither magnitude is plausible. The limitation is not specific to this study. Any application of a distributional component to states with sparse survey infrastructure inherits the imputation model of whichever compilation supplies the series, and the appropriate responses are to name the source product precisely and to prefer measures of distributional change, which imputation of this kind smooths away, in future specifications.

The remaining four components are taken directly from primary compilations with no imputation applied here. The demographic component requires an aggregation step, documented in Appendix B, because the source reports age bands as shares of the population of the same sex and not of the total.

8.3 Omitted variables

The discussion of misclassified cases identifies three theoretically relevant variables omitted from the current specification: coercive capacity, particularly military willingness to fire on protesters; elite cohesion and regime type, personalised against institutionalised autocracy as operationalised by Geddes, Wright, and Frantz (2014); and external financial transfers, the rentier-state mechanism. Each is measurable in principle and each requires additional data collection. The full theoretical index of Document II (Angeli 2026b), which in its complete form carries nine components, includes state capacity as a partial proxy for the first two; its planned

validation successor, the fifteen-variable SSI-15 discussed in Section 9, would incorporate the coercive-capacity and personalisation measures directly, but data availability at country-year level across a large historical sample remains the binding constraint. The leave-one-component-out results also indicate that the Gini coefficient in its current static form contributes noise rather than signal in this sample. The finding concerns the measurement and says nothing about the theoretical relevance of inequality. Future specifications should prioritise measures of inequality change, relative deprivation indices, or subjective economic wellbeing data such as those available from the Arab Barometer over static distributional snapshots.

Two further omissions merit discussion in light of the comparative literature. The first is what Korotayev et al. (2014) term immunity to internal conflict: countries that have experienced large-scale violence in the recent past, Algeria's civil war of the 1990s being the salient case in this sample, tend to exhibit reduced protest propensity during subsequent destabilisation waves. The mechanism works as follows: recent collective experience of catastrophic conflict raises the perceived cost of revolutionary mobilisation relative to the status quo, a form of societal risk aversion that Korotayev operationalises as a dampening factor. Algeria's trajectory is consistent with the hypothesis: the highest index in the panel at 2000, 0.918, followed by the sharpest decline over the decade, to -0.034 by 2010. While falling youth unemployment accounts for much of that decline, the residual stability of the Algerian regime during the 2011 protests, despite substantial remaining structural stress, may owe something to a conflict-memory effect the index does not capture.

The second is the availability of emigration as an alternative to political mobilisation. Schomaker and Wentzel (2014) show empirically that the combination of youth bulge and restricted migration opportunity significantly raises revolutionary probability, a finding they read through Hirschman's (1970) exit-voice framework, in which emigration acts as a safety valve that reduces domestic pressure by channelling discontent toward exit. The variable is absent from the current specification but could be operationalised as a net emigration rate or as a binary indicator of legal migration access to high-income destinations. It is potentially relevant to residual variance in countries where emigration corridors are particularly constrained, Yemen and to some degree Egypt, or particularly open, the Maghreb states with established labour migration to France and southern Europe.

8.4 Static approximation

The index is a static, level-based measure. It does not capture the rate of change in structural conditions, temporal lags between structural accumulation and political mobilisation, or dynamic feedback between political events and structural variables. The Egyptian case of Section 7.1 illustrates the limitation and also its precise shape: the deterioration that mattered was not visible in the index's own levels or differences, because the variables carrying it were outside the specification. The CMH framework at Level 2, the dynamic system $d\mathbf{X}/dt = F(\mathbf{X}, \boldsymbol{\theta}, \varepsilon(t))$, is explicitly dynamic; the SSI is a Level 4 approximation. The compact form is invoked here symbolically only, following the convention of Document II, where $\varepsilon(t)$ denotes the stochastic forcing whose expanded form is a system of stochastic differential equations interpreted throughout the series in the Itô sense. No stochastic integral is evaluated in this paper. Moving from the static to the dynamic specification is the primary technical objective of the next phase of the programme, and the Egyptian case indicates that the move must expand the variable set as well as add derivative terms.

A related technical constraint concerns the anocracy transformation. The functional form $A = 1 - |P_1|/10$ introduces a point of non-differentiability at $P_1 = 0$. For the static Level 4 specification this is immaterial, since only pointwise values are computed. The transition to a Level 2 dynamic system, however, requires smooth functions for numerical integration, and standard solvers can produce spurious oscillations or instability at non-differentiable points. Future dynamic specifications will therefore require a smooth alternative specification, not an

approximation of the piecewise form, for instance the quadratic

$$A = 1 - \left(\frac{P_1}{10}\right)^2, \tag{4}$$

which preserves the inverted-V shape, the range $[0, 1]$, and the symmetry around $P_1 = 0$ while guaranteeing infinite differentiability.

9 Conclusion

This paper has presented the most extensive quantitative application of the Computational Macrohistory framework to date, extending the three-country proof of concept of Document III to eleven countries. A formally specified, theory-derived index of structural stress classified eight of eleven MENA countries correctly at the pre-event benchmark, with a group-level separation of 0.290 index units between the instability and stability clusters. The ranking is stable across alternative weighting schemes except where connectivity is heavily upweighted, in which case discrimination collapses because internet penetration is a confounder in this sample, a negative result that adds to the evidence qualifying simple internet-mobilisation arguments.

The three misclassifications specify the limits of the framework rather than undermining it. Jordan shows that structural stress is necessary but not sufficient for instability, and that institutional capacity, external support, and strategic adaptation can contain even severe pressure. Egypt shows the boundary sensitivity of scalar threshold classification and, more sharply, that a deterioration the index cannot see in levels it also cannot see in differences: the remedy is a wider variable set, not merely a derivative term. Syria shows that the five-variable specification cannot distinguish stable autocracy from fragile personalised repression, a gap that is theoretically important and practically tractable through measures of regime personalisation and coercive capacity.

The leave-one-component-out analysis locates the discrimination in youth unemployment and anocracy stress, consistent with structural-demographic theory's emphasis on frustrated expectations and institutional vulnerability as the proximate drivers of revolutionary onset. It also shows that the other three components, carrying half the index weight between them, contribute nothing or degrade the result, and that the pre-specified baseline is outperformed within its own robustness analysis by three simpler variants. We report this without adopting the better variants, because the weights were fixed before contact with the data and post hoc selection would forfeit exactly the falsifiability the exercise is built to preserve. The constructive content of the negative result lies in specifying what inequality and connectivity would need to become, as measures, before either could contribute positively.

The broader framework remains to be demonstrated empirically. This analysis establishes that its core structural signals are detectable in the Arab Spring panel; it does not validate the full twenty-five-variable dynamic system. The path from exploratory application to statistical validation runs through a substantially expanded dataset, at least thirty country-cases across additional historical periods, the fifteen-variable index specified in Document II, extended to incorporate coercive capacity and regime personalisation, and a transition from static classification to dynamic trajectory analysis. What the present results establish is that the framework produces empirically tractable and analytically informative output at the current frontier of available data, and that the systematic analysis of its failures specifies what additional data, variables, and sample size the next stage requires.

The anticipatory architecture of the programme, theory preceding full empirical validation, is deliberate. Many of the variables required to test the framework fully have been systematically measured only for the past forty to fifty years, and data systems for psychological-collective variables such as mass sentiment and political trust are only now reaching the coverage that cross-national comparative analysis requires. What this paper shows is that even a substantially

reduced operationalisation produces discriminatively useful results, and that examining its failures generates theoretically productive hypotheses. We invite replication and extension using the materials of Appendix C.

References

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). *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.18646832

Angeli, S. (2026c). *Computational Macrohistory: Exploratory empirical application. The Arab Spring as a preliminary test case for structural-demographic theory* (FICSS Working Paper No. WP-2026-003, v3.1). Foundations Institute of Computational Social Science. https://doi.org/10.5281/zenodo.21440130

Blanchard, C. M., Humud, C. E., & Nikitin, M. B. D. (2015). *Armed conflict in Syria: Overview and U.S. response* (CRS Report No. RL33487). Congressional Research Service. https://www.everycrsreport.com/reports/RL33487.html

Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review, 97*(1), 75–90. https://doi.org/10.1017/S0003055403000534

Gaub, F. (2012). *Understanding instability: Lessons from the Arab Spring* (AHRC Public Policy Series No. 9). Arts and Humanities Research Council.

Geddes, B., Wright, J., & Frantz, E. (2014). Autocratic breakdown and regime transitions: A new data set. *Perspectives on Politics, 12*(2), 313–331. https://doi.org/10.1017/S1537592714000851

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

Goldstone, J. A. (2011). Understanding the revolutions of 2011: Weakness and resilience in Middle Eastern autocracies. *Foreign Affairs, 90*(3), 8–16.

Hegre, H., Karlsen, J., Nygård, H. M., Strand, H., & Urdal, H. (2013). Predicting armed conflict, 2010–2050. *International Studies Quarterly, 57*(2), 250–270. https://doi.org/10.1111/isqu.12007

Hirschman, A. O. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Harvard University Press.

Howard, P. N., & Hussain, M. M. (2013). *Democracy's Fourth Wave? Digital Media and the Arab Spring*. Oxford University Press.

Korotayev, A. V., Issaev, L. M., Malkov, S. Yu., & Shishkina, A. R. (2014). The Arab Spring: A quantitative analysis. *Arab Studies Quarterly, 36*(2), 149–169.

Korotayev, A., & Zinkina, J. (2011). Egyptian revolution: A demographic structural analysis. *Entelequia: Revista Interdisciplinar, 13*, 139–165.

Kuran, T. (1991). Now out of never: The element of surprise in the East European revolution of 1989. *World Politics, 44*(1), 7–48. https://doi.org/10.2307/2010422

Pew Research Center. (2011). *Egyptians embrace revolt leaders, religious parties and military, as well*. Pew Research Center, Global Attitudes Project. https://www.pewresearch.org/global/2011/04/25/egyptians-embrace-revolt-leaders-religious-parties-and-military-as-well/

Polity V Project. (2021). *Polity5: Political regime characteristics and transitions, 1800–2018*. Center for Systemic Peace.

Schomaker, R., & Wentzel, D. (2014). *Determinants of the emergence (and success) of regime changes: Theory and empirical findings* (Working Paper). FÖV Speyer and HS Pforzheim.

Tessler, M., Jamal, A., & Robbins, M. (2012). New findings on Arabs and democracy. *Journal of Democracy, 23*(4), 89–103. https://doi.org/10.1353/jod.2012.0066

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.

United Nations General Assembly. (2015). *Statement by the President of the General Assembly at the informal briefing to the General Assembly by the Special Envoy of the Secretary-General for Syria* [Press statement]. United Nations. https://www.un.org/en/ga/president/70/pdf/statements/20151119-mistura.pdf

UNU-WIDER. (2022). *World Income Inequality Database (WIID) Companion dataset (wiid-country), version 30 June 2022*. United Nations University World Institute for Development Economics Research. https://doi.org/10.35188/UNU-WIDER/WIIDcomp-300622

Urdal, H. (2006). A clash of generations? Youth bulges and political violence. *International Studies Quarterly, 50*(3), 607–629. https://doi.org/10.1111/j.1468-2478.2006.00416.x

Vreeland, J. R. (2008). The effect of political regime on civil war: Unpacking anocracy. *Journal of Conflict Resolution, 52*(3), 401–425. https://doi.org/10.1177/0022002708315594

World Bank. (2026). *World Development Indicators*. The World Bank Group. Series SP.POP.1519 and SP.POP.2024 by sex, SL.UEM.1524.ZS, and IT.NET.USER.ZS, accessed January 2026. https://databank.worldbank.org/source/world-development-indicators

A Complete SSI panel, 2000 to 2012

Table 6 gives the full panel of computed index values for all eleven countries across all thirteen years. Values are computed from the deposited data panel and the reference parameters of Table 2 by the procedure of Appendix B, and are reported to three decimal places. Bold values mark the post-event years 2011 and 2012.

Table 6: Complete SSI panel, all countries, 2000 to 2012.

Year	Algeria	Bahrain	Egypt	Jordan	Kuwait	Morocco	Oman	Saudi Arabia	Syria	Tunisia	Yemen
2000	0.918	−1.056	−0.130	0.537	−0.787	−0.027	−0.435	−0.505	−0.263	0.679	0.074
2001	0.798	−0.915	−0.046	0.619	−0.775	−0.057	−0.436	−0.470	−0.135	0.652	0.093
2002	0.767	−0.769	−0.076	0.674	−0.737	−0.092	−0.347	−0.372	−0.158	0.580	0.113
2003	0.691	−0.704	0.024	0.627	−0.666	−0.105	−0.321	−0.300	−0.182	0.583	0.133
2004	0.469	−0.682	0.027	0.668	−0.628	−0.114	−0.309	−0.235	−0.189	0.574	0.151
2005	0.362	−0.672	0.365	0.704	−0.652	−0.089	−0.312	−0.167	−0.237	0.482	0.167
2006	0.132	−0.662	0.333	0.649	−0.653	−0.046	−0.322	−0.109	−0.276	0.468	0.185
2007	0.226	−0.677	0.158	0.562	−0.588	−0.070	−0.293	−0.144	−0.249	0.467	0.228
2008	0.091	−0.618	0.163	0.546	−0.530	0.057	−0.306	−0.134	−0.144	0.506	0.262
2009	−0.019	−0.645	0.141	0.583	−0.523	0.036	−0.315	−0.084	−0.326	0.586	0.297
2010	−0.034	−0.499	0.102	0.574	−0.457	0.104	−0.306	−0.094	−0.192	0.517	0.324
2011	−0.038	−0.659	**0.330**	**0.661**	−0.427	**0.234**	−0.262	−0.096	−0.178	**0.904**	**0.374**
2012	**0.101**	−0.780	**0.426**	**0.589**	−0.395	**0.286**	−0.240	−0.110	−0.044	**0.667**	**0.324**

Note: Bold values for 2011 and 2012 are post-event observations included for completeness. The core analysis uses the pre-event window 2000 to 2010. Values are computed by the formula of Section 3.3.

B Calculation procedure

The demographic component requires an aggregation step that the source series do not supply directly, because the World Development Indicators report the 15 to 19 and 20 to 24 age bands as shares of the population of the same sex and not of the total. For each country-year the youth bulge is therefore obtained by reconstructing sex-specific totals from the absolute age-band counts, applying the sex-specific band shares, and dividing the combined youth count by the combined total:

$$D_2 = 100 \times \frac{\sum_g \left(p^g_{\text{15-19}} + p^g_{\text{20-24}}\right) N^g}{\sum_g N^g}, \qquad N^g = N^g_{\text{0-14}} + N^g_{\text{15-64}} + N^g_{\text{65+}} \tag{5}$$

where g ranges over the two reported sexes, p^g denotes the reported band share of the population of sex g expressed as a fraction, and N^g the total population of that sex. The procedure reproduces all 143 panel values to within 0.015 percentage points.

The index for a country-year is then computed in three steps. The Polity score is transformed to anocracy stress, $A = 1 - |P_1|/10$. Each of the five components is standardised as $z = (X - \mu)/\sigma$ against the reference parameters of Table 2. The standardised components are combined as $SSI = 0.15\, z(D_2) + 0.25\, z(E_2) + 0.25\, z(E_4) + 0.25\, z(A) + 0.10\, z(S_3)$. Every value reported in this paper, including all robustness variants and the group separation and correct-pairs metrics, is reproducible to the reported precision from the deposited data panel, the parameters of Table 2, and the weight specifications of Sections 3.3 and 6.1, using any spreadsheet or statistical software. All values were re-verified by independent recomputation from the raw panel for this version.

C Replication materials

Data and code for replication are available in the FICSS umbrella repository at `https://github.com/FICSS-Institute/computational-macrohistory/tree/main/CMH-Arab-Spring-11countries`. The folder contains the raw eleven-country panel in CSV format, the codebook documenting indicator codes, dataset versions, access dates, the reproducible source queries and the demographic aggregation of Appendix B, the source files as downloaded, and the figures of this paper. Replication is deterministic: the procedure of Appendix B applied to the deposited panel reproduces every reported value. Figures 1 to 6 of this version correspond to the files `fig1_ssi_trajectories.png`, `fig2_group_trajectories.png`, `fig3_ssi_2010_ranking.png`, `fig4_component_decomposition.png`, `fig5_weighting_schemes.png`, and `fig6_leave_one_out.png`.

Document IV of the Computational Macrohistory Series

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

Contact: stefano.angeli@ficss.institute