Dissensus Working Paper Series

DAI-2509

ASRI: An Aggregated Systemic Risk Index for Cryptocurrency Markets

An Interpretable Crypto-Native Stress Composite for Retrospective Systemic-Risk Discrimination

Murad Farzulla[1,2,*], Andrew Maksakov[1]

[1]Dissensus, London, UK [2]King's College London, London, UK

[*]Correspondence: murad@dissensus.ai ORCID: 0009-0002-7164-8704

January 2026

Abstract

Cryptocurrency markets have grown to represent over $3 trillion in capitalisation, yet practitioners lack an interpretable, channel-decomposed composite for retrospectively characterising crypto-native systemic stress. This paper introduces the Aggregated Systemic Risk Index (ASRI), a composite stress measure built around crypto-native channels—stablecoin total-value-locked (TVL) drawdown, stablecoin-issuer concentration, and TVL volatility—augmented by a cross-market *contagion* channel that we implement as a traditional-finance (TradFi) stress *proxy* (Treasury yields, VIX, and the yield-curve spread) rather than as a directly measured DeFi-TradFi exposure. Methodologically the index belongs to the financial-network tradition of connectedness and contagion measurement: the contagion channel is a cross-market connectedness construct, the four weighted sub-indices form an explicit risk-transmission decomposition, and a Diebold–Yilmaz (2012) connectedness series computed on those sub-indices serves as the network benchmark. We derive an axiomatic foundation and quantitative formulas for four weighted sub-indices—Stablecoin Concentration Risk (30%), DeFi Liquidity Risk (25%), Contagion Risk (25%), and Regulatory Opacity Risk (20%)—and evaluate the index *retrospectively*, as an interpretable discriminative monitoring framework, against four historical crisis episodes: the Terra/Luna collapse (May 2022), the Celsius/3AC contagion (June 2022), the FTX bankruptcy (November 2022), and the SVB banking crisis (March 2023). We do not claim a validated real-time forecaster; the contribution is a transparent, reproducible construction together with a methodological account of how autocorrelation- and block-structure-robust inference reshapes the apparent strength of crisis-detection claims. **Empirical Results:** An event-study test for an elevated cumulative abnormal signal around the four onsets is inconclusive: the signal sums a 41-day window of 30-day-rolling components and is heavily serially correlated (AR(1) $\approx$ 0.8–0.9), and a placebo test on non-crisis dates shows the finite-sample (fixed-b) reference has essentially no crisis-specificity on this smooth trending series—roughly sixty percent of arbitrary dates clear the nominal 5% threshold, and the four crisis onsets sit at or below the placebo median. We therefore do not read the raw event-study significance values as crisis detection, and rest the empirical case on the fair-baseline day-level analysis below. Threshold-based operational detection (fixed ASRI $\geq$ 50) identifies the same three events (Celsius/3AC, FTX, SVB) with an average first-crossing lead of $\approx$19 days (partly inflated by a running-maximum drawdown offset; $\approx$5 days under a responsive 30-day percentage-change specification). Benchmarking against a Diebold–Yilmaz (2012) connectedness series

computed on the four sub-indices, ASRI shows higher *retrospective* day-level discrimination on this sample (AUROC 0.866 vs. 0.670; AUPRC 0.298 vs. 0.121); but a fair-baseline comparison on identical labels demotes this headline. ASRI's discrimination (0.866) is *not* statistically distinguishable from its single strongest sub-index (Contagion Risk, 0.851) or from the first principal component of its sub-indices (PC1, 0.858)—both fall inside ASRI's bootstrap confidence interval, and aggregation buys only $\approx +0.008$–0.015 AUROC over PC1 and the best single channel. A standalone equity-volatility series (VIX) also matches ASRI's headline discrimination (AUROC 0.875 vs. 0.866, indistinguishable at $p = 0.58$), so to first order the crisis labelling is a macro-volatility detection problem and ASRI's discriminative edge over an off-the-shelf series is not established. ASRI's measured advantage is confined to the Diebold–Yilmaz benchmark (0.670), which is itself *constructed from* ASRI's four sub-indices (a circular comparator) and is so weak that an off-the-shelf Crypto Fear & Greed sentiment index outperforms it (0.789 vs. 0.670). Under a moving-block bootstrap (block length $L = 25$ days) the marginal confidence intervals are roughly four times wider than naive i.i.d. resampling implies; ASRI's day-level edge over D-Y is positive in every resample (AUROC $+0.194$, 95% CI $[+0.093, +0.298]$; AUPRC $+0.182$, 95% CI $[+0.070, +0.337]$), but this only establishes superiority over that circular, unusually weak comparator, and with only four independent crisis episodes the $\approx$0.008–0.015 AUROC gaps separating ASRI from its best single channel and PC1 are indistinguishable on the marginal intervals—the small paired-bootstrap edge that does separate them (≤ 0.016 AUROC, absent in AUPRC) is a structural artefact of ASRI nesting those baselines, not a discriminative gain. We therefore read the value of aggregation as interpretive—channel attribution, lead-time, and regime structure in a single auditable composite—rather than as a gain in discriminative power. Walk-forward thresholds calibrated only on pre-crisis data flag all four episodes, but at a high false-positive cost for the early crises (Terra/Luna and Celsius/3AC alarm on 47–59% of the index history at $\approx$11–13% precision), so we read the 4/4 walk-forward rate as evidence against look-ahead bias rather than as a clean prediction record. Stationarity tests reject the unit root for ASRI and the DeFi Liquidity sub-index ($p < 0.01$); Stablecoin and Opacity Risk are trend-stationary, while Contagion Risk is non-stationary on the released sample. A three-regime Hidden Markov Model labels Low Risk, Moderate, and Crisis states with persistence exceeding 97%—interpretive structure over an autocorrelated trend rather than statistically distinct regimes—and the reproducible full-sample structural-stability (Chow) test at the sample midpoint is a clean non-rejection ($p = 0.78$). All detection, precision, and lead-time results rest on only four crisis events—the binding statistical-power limitation of this study—and roughly 43% of the backtested series' average level is carried by static default values (which, being constant, contribute little of its time-variation), with the contagion channel's time-variation $\approx$75% attributable to the macro (Treasury/VIX/yield-curve) proxy rather than to a measured interconnection. Out-of-sample 2024–2025 readings show no sustained false alarms and correctly classify the February 2025 Bybit hack ($1.5B, the largest exchange theft in history) as non-systemic due to the absence of contagion channels. We therefore present ASRI as an interpretable, retrospective monitoring framework and methodological contribution—designed to target crypto-native vulnerabilities (stablecoin concentration, liquidity drawdown, opacity) that traditional measures such as SRISK and CoVaR are not built to capture—rather than as a statistically validated early-warning system.

Keywords: systemic risk, cryptocurrency, decentralised finance, stablecoin stability, contagion risk, TradFi-stress proxy, retrospective discrimination, composite index, event study, regime detection

JEL Codes: G01 (Financial Crises), G15 (International Financial Markets), G23 (Non-bank Financial Institutions)

Contents

1 Introduction

The cryptocurrency market has evolved from a niche technological experiment into a multi-trillion dollar asset class with growing interconnections to traditional finance. As of December 2025, the total cryptocurrency market capitalisation exceeds $3 trillion, with stablecoins alone representing over $140 billion in circulation (DeFi Llama, 2025). This growth has been accompanied by a series of cascading failures that revealed systemic vulnerabilities previously unrecognised: the Terra/Luna collapse eliminated $40 billion in value within 72 hours; the subsequent Celsius and Three Arrows Capital failures triggered margin calls across centralised exchanges; and the FTX bankruptcy demonstrated how opaque counterparty relationships could propagate losses across the entire ecosystem.

Despite this systemic importance, no unified, channel-decomposed composite exists for retrospectively characterising crypto-native systemic stress. Existing measures either focus exclusively on cryptocurrency price volatility (Liu and Tsyvinski, 2021), apply traditional banking metrics that miss DeFi-specific dynamics (Brownlees and Engle, 2017), or provide sentiment-based indicators without quantitative grounding (Alternative.me, 2022).

This paper introduces the Aggregated Systemic Risk Index (ASRI), a composite stress measure constructed around crypto-native channels—stablecoin TVL drawdown, stablecoin-issuer concentration, and TVL volatility—with a cross-market *contagion* channel implemented as a traditional-finance (TradFi) stress *proxy* (Treasury yields, VIX, and the yield-curve spread) rather than as a directly measured DeFi-TradFi exposure. We are explicit about this throughout: ASRI does not measure realised DeFi-TradFi interconnection: it composes interpretable crypto-stress channels and proxies the macro/contagion dimension. We are equally explicit that "crypto-native" here characterises the index's *design intent*, not what drives the realised series. A component-level accounting of the backtested index (Appendix A.6.4, Table 11) attributes only ≈37% of its *average level* (mean composition) to dynamic crypto-native inputs, against ≈43% carried by static default values—which, being constant, contribute little or none of its time-variation—and ≈20% by the macro (Treasury/VIX/yield-curve) proxy. ASRI is therefore better described as crypto-native in *construction* than as predominantly crypto-native in realised *variation*, a distinction we keep in view when interpreting the empirical results. The index comprises four weighted sub-indices:

1. **Stablecoin Concentration Risk (30%)**: Measures reserve composition, peg stability, and Treasury exposure across major stablecoins
2. **DeFi Liquidity Risk (25%)**: Captures protocol concentration, leverage dynamics, and smart contract vulnerability
3. **Contagion Risk (25%)**: Quantifies TradFi linkage intensity, tokenised RWA growth, and cross-market correlation shifts
4. **Regulatory Opacity Risk (20%)**: Assesses transparency scores, regulatory arbitrage metrics, and compliance infrastructure

The ASRI framework addresses three critical gaps in existing risk monitoring:

First, *composability risk*. DeFi protocols interact through smart contract calls that create dependency chains invisible to external observers. When one protocol fails, composable integrations can transmit losses instantaneously across the ecosystem—a dynamic that traditional contagion models, designed for bilateral counterparty relationships, cannot capture.

Second, *stablecoin-Treasury linkages*. Major stablecoins now hold significant Treasury bill positions, creating a direct transmission channel between US monetary policy and DeFi liquidity conditions. Rate hikes that increase Treasury yields simultaneously reduce stablecoin reserve valuations and incentivise capital rotation out of yield-bearing DeFi positions.

Third, *regulatory arbitrage dynamics*. The fragmented regulatory landscape creates opacity about counterparty risk exposures, custody arrangements, and reserve attestation reliability. Traditional banking metrics assume regulatory disclosure requirements that do not exist for offshore or unregulated platforms.

The urgency of dedicated crypto risk monitoring is underscored by recent TVP-VAR evidence from Malik et al. (2025), who identify a structural break in Q3 2020 after which the cryptocurrency market shifted from net receiver to net transmitter of financial risk spillovers to traditional markets. Their finding that crypto is no longer a "digital diversifier" but an active source of systemic contagion motivates ASRI's focus on crypto-native stress channels and a macro-stress proxy for the cross-market boundary, while underscoring that the boundary itself is better measured directly than proxied—a limitation we keep in view throughout.

Conceptually and methodologically, ASRI sits within the financial-network-science tradition that treats systemic risk as a property of connectedness between markets rather than of individual institutions in isolation. That literature runs from Granger-causality and connectedness networks among financial firms (Billio et al., 2012; Diebold and Yılmaz, 2012, 2014), through network-propagation measures of distress such as DebtRank (Battiston et al., 2012), to recent mappings of the risk-transmission channels that link traditional and decentralised finance (Aufiero et al., 2025). ASRI's contagion sub-index is a market-connectedness construct in this lineage—built on cross-market correlation and co-movement—and its four-channel decomposition is a transmission-channel structure in which each sub-index names a pathway along which stress can propagate. We benchmark the composite against a Diebold–Yilmaz (2012) connectedness series computed on those sub-indices, the canonical variance-decomposition network measure of total spillovers, while remaining explicit (Appendix M) about the limits of that particular comparator.

We are equally explicit about what the empirical evaluation does *not* establish. A fair-baseline comparison on identical crisis labels (Section 5.7) shows that ASRI's day-level discrimination is, on the marginal block-bootstrap intervals, indistinguishable from that of its single strongest sub-index (Contagion Risk) and from the first principal component of its four sub-indices: aggregation adds only $\approx +0.008$–0.015 AUROC over the best single channel and PC1, a gap well within the four-event sampling uncertainty (a paired bootstrap separates them only by this small, structural margin—ASRI nests both baselines—not a meaningful discriminative gain). A standalone equity-volatility series (VIX) likewise matches ASRI's day-level discrimination on these labels (0.875 vs. 0.866, statistically indistinguishable), so to first order labelling these crises is a macro-volatility detection problem. ASRI's empirical contribution is therefore not that the four-channel composite out-discriminates simpler constructions—it does

not, measurably—but that it packages crypto-native stress into a single, auditable, channel-decomposed series whose value is interpretive: it attributes stress to nameable transmission channels, carries operational lead-time for the detected crises, and exhibits coherent regime structure, none of which an opaque single feature or principal component delivers.

This paper proceeds as follows. Section 2 reviews the literature on systemic risk measurement, cryptocurrency market dynamics, and existing crypto risk indices. Section 3 develops the ASRI framework, specifying the axiomatic foundation and quantitative formulas for each sub-index and making explicit which components are crypto-native, which are macro proxies, and which are static defaults. Section 4 describes data sources and implementation architecture. Section 5 presents a *retrospective* evaluation against historical crisis events—event study analysis under autocorrelation-robust inference, a block-bootstrap benchmark comparison, regime detection, robustness tests, and out-of-sample specificity testing. Section 6 discusses theoretical and practical implications. Section 7 concludes.

2 Literature Review

2.1 Traditional Systemic Risk Measures

The 2008 financial crisis catalysed extensive research on systemic risk measurement. Adrian and Brunnermeier (2016) introduced Conditional Value-at-Risk (CoVaR), measuring the VaR of the financial system conditional on an institution being in distress. Acharya et al. (2017) developed SRISK, estimating the expected capital shortfall of a financial institution during a systemic crisis. Brownlees and Engle (2017) extended this framework with LRMES (Long-Run Marginal Expected Shortfall), capturing an institution's contribution to aggregate capital shortfall.

These measures share a common architecture: they model systemic risk as arising from bilateral exposures between regulated financial institutions with observable balance sheets and regulatory capital requirements. This architecture is fundamentally unsuited to DeFi, where:

- Protocols are not institutions with capital requirements
- Exposures are embedded in smart contract code rather than disclosed counterparty relationships
- "Failure" may manifest as liquidity drainage rather than insolvency
- Contagion propagates through token price collapses and oracle manipulation rather than credit defaults

2.2 Cryptocurrency Market Risk

Research on cryptocurrency-specific risk has focused primarily on volatility dynamics and market microstructure. Aste (2019) establishes a foundational connection between emotional dynamics and economic structure in cryptocurrency markets, demonstrating that market structure emerges from the interplay of sentiment-driven trading and network topology—a perspective that informs ASRI's integration of sentiment-adjacent indicators with structural risk channels. Liu and Tsyvinski (2021) identify three common factors driving cryptocurrency returns: market (aggregate crypto exposure), size, and momentum. Makarov and Schoar (2020) document

arbitrage frictions across exchanges that allow price dislocations to persist, creating opportunities for informed traders and risks for liquidity providers. Farzulla (2025) demonstrate that infrastructure disruptions generate 5.7× larger volatility shocks than regulatory events, suggesting that technical vulnerabilities represent a more severe systemic risk channel than policy uncertainty.

Griffin and Shams (2020) provide evidence that Tether issuance patterns correlate with Bitcoin price movements, raising questions about stablecoin reserve integrity. This finding has implications for systemic risk: if stablecoin issuance is not fully collateralised, DeFi liquidity pools dependent on stablecoin inflows may be vulnerable to sudden redemption pressures.

2.3 Crisis Episode Analysis

The cryptocurrency crises of 2022–2023 generated substantial empirical scholarship validating theoretical concerns about systemic fragility. Liu et al. (2023) provide the definitive analysis of the Terra/Luna collapse, documenting how algorithmic stablecoin mechanics created a reflexive depegging spiral that eliminated $40 billion in value within 72 hours. Ma et al. (2025) extend this analysis to demonstrate that stablecoin run dynamics exhibit centralisation in arbitrage activity: during stress, redemption becomes concentrated among sophisticated actors, creating adverse selection that accelerates depegging.

The FTX collapse of November 2022 has received detailed independent analysis. Vidal-Tomás et al. (2023) trace FTX's downfall to the prior Terra/Luna ecosystem collapse, which triggered a liquidity crisis that exposed FTX's reliance on leveraged positions in its native FTT token and opaque intercompany transfers with Alameda Research. Their on-chain analysis demonstrates how centralised exchange fragility—masked by token self-valuation and inadequate reserve transparency—can trigger cascading failures across the broader ecosystem. This independent account of the FTX contagion pathway is qualitatively consistent with ASRI's event study for this one episode—in which the FTX collapse produced the highest recorded index value (84.7)[1] and the largest absolute increase from baseline of any validation event—though, as a single-event narrative correspondence, it does not constitute statistical out-of-sample validation.

The March 2023 Silicon Valley Bank crisis demonstrated bidirectional contagion between traditional and decentralised finance. Diop et al. (2024) document the USDC depeg following SVB's collapse—the first major case of TradFi stress propagating into DeFi through stablecoin reserve exposure. Gross and Senner (2026) model fire sale scenarios under systemic stablecoin stress, while Eichengreen et al. (2025) develop a framework for quantifying devaluation risk across stablecoin designs. Vidal-Tomás and Aste (2025) examine the evolving dynamic relationship between cryptocurrency and traditional financial markets, finding evidence of increasing integration over time—a trend that strengthens the case for monitoring DeFi-TradFi interconnection as a systemic risk channel. More broadly, Aufiero et al. (2025) provide a systematic mapping of microscopic and systemic risk transmission channels between TradFi and DeFi, formalising the "crosstagion" pathways through which instability propagates across the conventional-decentralised boundary.

[1]The released series' aggregate column differs slightly from a strict Equation 6 recomposition of its released sub-index columns, a consequence of the documented post-generation sub-index repair (mean gap +0.84 points, maximum 6.34; the recomposed maximum is 81.06 against the released 84.70). All statistics reported in this paper use the released series, the dataset of record.

2.4 DeFi-Specific Risks

The academic literature on DeFi risk is nascent but growing rapidly. Gudgeon et al. (2020) provide a systematisation of DeFi protocol architectures, identifying liquidation cascades as a primary risk channel. Perez et al. (2021) analyse liquidation events across lending protocols, finding that liquidator competition can exacerbate price volatility during stress periods.

Werner et al. (2022) present a systematisation of knowledge on DeFi covering lending, decentralised exchanges, and derivatives. They identify composability—the ability of protocols to permissionlessly interact through smart contract calls—as both a source of innovation and systemic vulnerability. When protocols share liquidity pools or collateral mechanisms, failures can propagate faster than human intervention can respond. Recent work on whitepaper-market alignment (Farzulla, 2026) suggests that stated protocol objectives may diverge from realised market behaviour, creating additional opacity in risk assessment.

Recent work has deepened understanding of DeFi-specific systemic channels. Bartoletti et al. (2022) formalise composability risk through MEV (Maximal Extractable Value), demonstrating that protocol interactions create value extraction opportunities that destabilise liquidity during stress. Darlin et al. (2022) analyse debt-financed collateral structures, identifying leverage amplification mechanisms. Cross-chain bridge infrastructure represents acute vulnerability; Notland et al. (2024) systematise knowledge on bridge exploits, analysing 60 bridges and 34 attacks (2021–2023) to identify 13 architectural components linked to 8 vulnerability types. Zhang et al. (2026) propose a correlation-based fragility indicator for detecting dangerous protocol synchronisation.

2.5 Stablecoin Risk

Stablecoins occupy a critical position in DeFi infrastructure, serving as the primary medium of exchange and store of value across protocols. Lyons and Viswanath-Natraj (2023) analyse stablecoin run dynamics, finding that algorithmic stablecoins are particularly vulnerable to reflexive depegging spirals. The Terra/Luna collapse of May 2022 validated this theoretical concern empirically.

Gorton and Zhang (2023) frame stablecoins through the lens of 19th-century "wildcat banking," where private currency issuance without adequate regulatory oversight led to periodic banking panics. They argue that stablecoin reserves require the same transparency and examination that bank reserves receive—a standard that most major stablecoins currently fail to meet.

De Blasis et al. (2023) conduct comparative performance analysis across the Terra, Celsius, and FTX episodes, finding that fiat-collateralised stablecoins exhibit more resilience than algorithmic designs. Antonakakis et al. (2020) introduce the TVP-VAR connectedness framework that enables measurement of time-varying spillover dynamics without arbitrary rolling-window selection—a methodological advance particularly relevant for cryptocurrency markets where structural breaks occur frequently.

2.6 Methodological Considerations

A recent finding warrants consideration: Rapos and Fountas (2025) challenge the assumption of strong Bitcoin-equity market contagion, finding spillovers remain limited even during stress

episodes. This finding does not invalidate contagion measurement but suggests correlation-based measures should be interpreted cautiously: elevated correlation may reflect common shocks rather than causal transmission.

2.7 Existing Crypto Risk Indices

Several commercial indices attempt to quantify cryptocurrency market risk:

Fear & Greed Index (Alternative.me, 2022): Aggregates sentiment indicators (social media, volatility, dominance, trends) into a 0–100 scale. Limitation: purely sentiment-based with no fundamental risk component.

Crypto Climate Index (CCI): Combines on-chain metrics with market data. Limitation: focuses on individual asset health rather than systemic interconnection.

Glassnode Risk Assessment: Provides sophisticated on-chain analytics for Bitcoin and Ethereum. Limitation: asset-specific rather than ecosystem-wide; minimal DeFi coverage.

In the academic literature, Guo et al. (2024) propose the Cryptocurrency General Risk Index (CGRI), which tracks and sources cryptocurrency risk through a composite of market, sentiment, and macro indicators. While CGRI provides a useful aggregate risk signal for the cryptocurrency asset class as a whole, it operates at the market-sentiment level without decomposing risk into the structural transmission channels—stablecoin reserve linkages, DeFi composability dependencies, tokenised RWA exposures, and regulatory opacity—that characterise DeFi-TradFi interconnection. ASRI differs by targeting these specific contagion pathways, enabling identification of *which* structural channel is generating stress rather than only *whether* aggregate risk is elevated.

Most recently, Shah (2025) constructs a Unified DeFi Risk Index (DeFi-RI) integrating credit, liquidity, and governance risk into a composite scoring model. Shah's decomposition captures governance fragility and protocol-level credit exposure that ASRI currently underweights, but does not address the cross-boundary transmission channels—stablecoin-Treasury linkages, TradFi contagion pathways, and regulatory opacity—that constitute ASRI's primary focus. The two indices are thus complementary rather than competing: DeFi-RI monitors intra-DeFi protocol health while ASRI monitors the DeFi-TradFi boundary where systemic risk materialises.

None of these indices systematically captures the DeFi-TradFi interconnection dynamics that ASRI is designed to monitor: stablecoin reserve composition, composable protocol dependencies, tokenised RWA linkages, or regulatory arbitrage exposure.

3 ASRI Framework

3.1 Conceptual Foundation

The ASRI framework rests on three theoretical principles:

Principle 1: Interconnection Creates Systemic Risk. Following Battiston et al. (2012), systemic risk arises not from the failure of individual nodes but from the propagation of distress through network connections. Aste (2025) provides the theoretical and methodological foundations for information filtering networks—sparse graph representations that extract statistically significant dependency structures from high-dimensional data—offering a principled approach to identifying which connections carry genuine risk information versus noise. In DeFi,

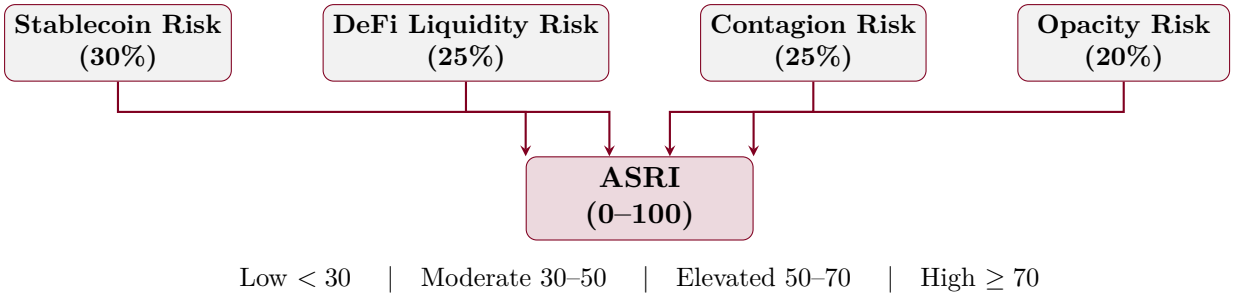

Figure 1: ASRI Framework Architecture: Four weighted sub-indices aggregate into a normalised composite risk measure with defined alert thresholds.

these connections manifest through shared collateral pools, composable protocol integrations, and token price correlations.

Principle 2: Risk Transmission Requires Channels. We identify four primary channels through which DeFi stress can propagate to traditional finance (and vice versa): stablecoin reserve linkages, tokenised RWA exposures, crypto-equity correlations, and regulatory enforcement actions. Each channel corresponds to an ASRI sub-index.

Principle 3: Opacity Amplifies Risk. Systemic risk is exacerbated when counterparties cannot accurately assess exposures. The prevalence of unaudited reserves, undisclosed custody arrangements, and regulatory arbitrage structures in crypto markets justifies a dedicated opacity sub-index.

Figure 1 illustrates the ASRI architecture, showing how the four sub-indices aggregate into the composite risk measure.

3.2 Axiomatic Foundation

We establish the formal properties that any coherent systemic risk index must satisfy, demonstrating that ASRI adheres to these axioms. Our axiomatic framework draws on the coherent risk measure literature (Artzner et al., 1999) while extending it to the specific requirements of cryptocurrency market monitoring, where the absence of central counterparties and the prevalence of cross-venue arbitrage necessitate distinct aggregation properties.

Definition 3.1 (Systemic Risk Index)**.** *Let $\mathcal{S} = \{S_1, \ldots, S_n\}$ denote a set of sub-indices measuring distinct risk dimensions. A* systemic risk index *is a mapping $\rho : \mathcal{S} \to [0, 100]$ that aggregates component risks into a scalar measure of system-wide vulnerability.*

For ASRI, we have $\mathcal{S} = \{\text{SCR}, \text{DLR}, \text{CR}, \text{OR}\}$ with the aggregation function:

$$\text{ASRI}(\mathcal{S}) = \sum_{i=1}^{4} w_i S_i, \quad \text{where } \sum_{i=1}^{4} w_i = 1 \text{ and } w_i > 0 \ \forall i \tag{1}$$

We now state and verify five axioms that characterise well-behaved systemic risk indices.

Axiom 3.1 (Monotonicity)**.** *For any sub-index $S_j \in \mathcal{S}$, if $S'_j > S_j$ while $S'_i = S_i$ for all $i \neq j$, then $\rho(\mathcal{S}') > \rho(\mathcal{S})$.*

Axiom 3.2 (Boundedness)**.** *The index satisfies $\rho(\mathcal{S}) \in [0, 100]$ for all feasible states $\mathcal{S}$.*

Axiom 3.3 (Decomposability)**.** *For any realisation of ASRI, there exists a unique attribution* $\{c_1, \ldots, c_n\}$ *such that* $\sum_{i=1}^{n} c_i = ASRI(\mathcal{S})$ *and* c_i *represents the contribution of sub-index* S_i*.*

Axiom 3.4 (Aggregation Neutrality)**.** *Linear aggregation preserves ordinal rankings: if* $\mathcal{S}^A$ *and* $\mathcal{S}^B$ *represent two market states with* $S_i^A \geq S_i^B$ *for all* i *and strict inequality for at least one* j*, then* $\rho(\mathcal{S}^A) > \rho(\mathcal{S}^B)$*.*

Axiom 3.5 (Concentration Sensitivity)**.** *Let* H_t *denote any market-concentration measure entering ASRI—specifically the stablecoin-issuer Herfindahl–Hirschman index* HHI_t *in SCR (Eq. 2) or the top-10 protocol-TVL concentration* $Conc_t$ *in DLR (Eq. 3). The index satisfies* $\partial\rho/\partial H_t > 0$ *for each such measure when concentration risk is elevated.*

Proofs of Axioms 3.1–3.5, together with the relationship to the coherent-risk-measure framework of Artzner et al. (1999), are given in Appendix B.

3.3 Weight Selection Justification

Sub-index weights were selected based on theoretical importance and precedent from traditional systemic risk literature:

- **Stablecoin Concentration Risk (30%)**: Stablecoins are the foundational liquidity layer for DeFi. Their failure would immediately impact all protocols dependent on stablecoin-denominated liquidity pools. The highest weight reflects this critical infrastructure role.
- **DeFi Liquidity Risk (25%)**: Protocol concentration and leverage dynamics directly determine the ecosystem's resilience to market stress. We assign this channel no leading or "canary" role: with only four co-located crisis episodes the data do not identify which channel moves earliest (Section 5.5), and the earlier draft's hard-coded DLR weight that motivated such a claim does not reproduce from the weighting code.
- **Contagion Risk (25%)**: DeFi-TradFi linkages represent the primary channel through which crypto stress could affect traditional finance (and vice versa). Equal weighting with DLR reflects their complementary roles: DLR captures within-DeFi stress, while CR captures cross-market transmission.
- **Opacity Risk (20%)**: While important, opacity is an amplifying factor rather than a primary risk driver. Lower weight reflects this secondary role in conditioning crisis severity rather than triggering crises.

Sensitivity analysis (Appendix J) tests robustness to alternative weight specifications. Section 5.5 compares theoretical weights against four data-driven alternatives (PCA, Elastic Net, CRITIC, entropy) and finds that they disagree substantially among themselves; ablation analysis further shows detection is invariant at 3/4 to removing any single channel, so the decomposition is retained for interpretability rather than as an individually load-bearing causal structure. Figure 2 visualises this decomposition empirically, showing how each sub-index contribution evolves over the sample period.

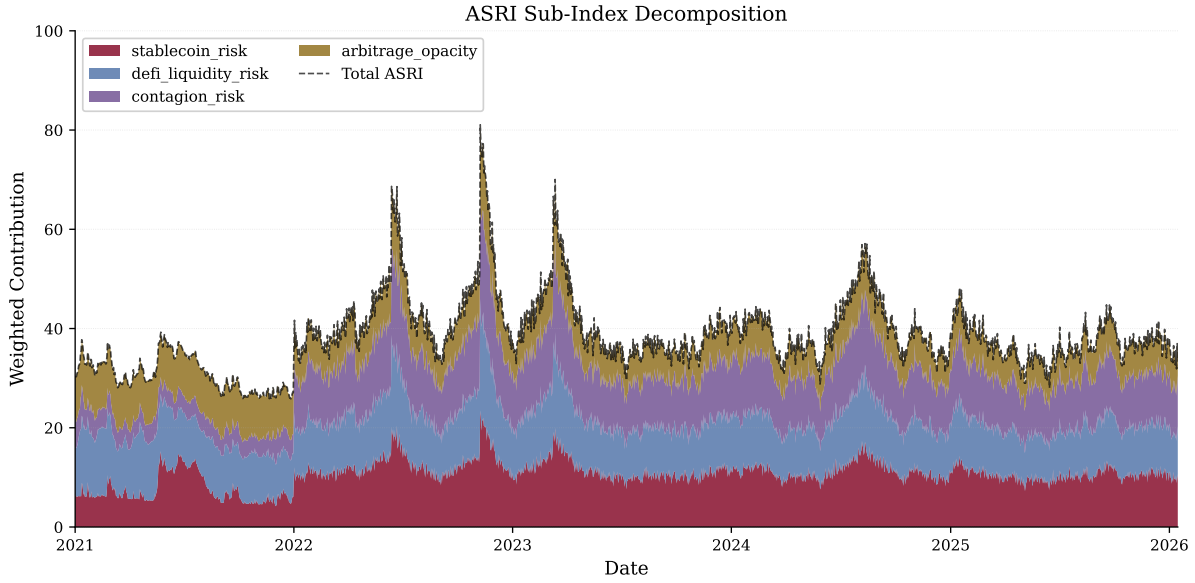

Figure 2: ASRI Decomposition by Sub-Index Contribution Over Time. Stacked area chart showing how Stablecoin Risk (SCR, 30%), DeFi Liquidity Risk (DLR, 25%), Contagion Risk (CR, 25%), and Regulatory Opacity (OR, 20%) contribute to the aggregate ASRI. The decomposition property guarantees $c_i = w_i S_i$ such that total area equals ASRI at each time point. During crisis periods, shifting relative contributions reveal which transmission channels dominate aggregate stress.

3.4 Stablecoin Concentration Risk (30%)

The Stablecoin Risk sub-index captures reserve composition vulnerabilities, peg stability, and concentration across issuers:

$$\text{SCR}_t = 0.4 \cdot \text{TVL}_t + 0.3 \cdot \text{Treasury}_t + 0.2 \cdot \text{HHI}_t + 0.1 \cdot \text{Vol}_t \tag{2}$$

where:

- $\text{TVL}_t = 1 - \frac{\text{Stablecoin TVL}_t}{\max_{\tau \le t}(\text{Stablecoin TVL}_\tau)}$ measures stablecoin TVL *drawdown* from historical maximum—declining TVL increases risk[2][3]
- $\text{Treasury}_t = \frac{\text{T-Bill Reserves}_t}{\text{Total Stablecoin Reserves}_t}$ captures Treasury exposure concentration
- $\text{HHI}_t = \sum_{i=1}^{n} s_i^2$ is the Herfindahl-Hirschman Index of stablecoin market share concentration
- Vol_t is the 30-day realised volatility of weighted-average stablecoin peg deviation

[2]The inversion ensures countercyclical behaviour: when TVL collapses during crises, TVL_t rises towards 1 (high risk); at historical peak, $\text{TVL}_t = 0$ (low risk). See Appendix A for implementation details.

[3]*Construction limitation (running-maximum saturation).* The normalisation is against a running (expanding) maximum, so once total DeFi TVL fell below half of its late-2021 peak in mid-2022 and did not reclaim it, this term is clipped at its ceiling for ≈85% of 2022–2024 and contributes a near-constant ≈12-point offset to ASRI ($w_{\text{SCR}} \cdot w_{\text{TVL}} \cdot 100 = 0.30 \times 0.40 \times 100$) across both crisis and calm post-2022 days, rather than tracking episode-specific stress. Because operational detection uses a *fixed* threshold, this offset mechanically lifts post-2022 threshold crossings. Section 6.3 reports the effect of two alternative normalisations on detection.

Data Sources: DeFi Llama (stablecoin TVL), attestation reports (reserve composition), CoinGecko (price feeds for volatility calculation).

3.5 DeFi Liquidity Risk (25%)

The DeFi Liquidity sub-index captures protocol concentration, leverage dynamics, and smart contract vulnerability:

$$\begin{aligned} \text{DLR}_t = 0.35 \cdot \text{Conc}_t + 0.25 \cdot \text{TVLVol}_t \\ + 0.20 \cdot \text{SC}_t + 0.10 \cdot \text{Flash}_t + 0.10 \cdot \text{Lev}_t \end{aligned} \tag{3}$$

where:

- Conc_t is the HHI of TVL across top-10 DeFi protocols
- TVLVol_t is the 30-day volatility of total DeFi TVL
- SC_t is a composite smart contract risk score based on audit status, time since deployment, and exploit history
- Flash_t measures flash loan volume spikes relative to 90-day average
- Lev_t captures 30-day change in aggregate leverage ratios across lending protocols

Implementation overlap. In the historical backtest Flash_t is not sourced from a direct on-chain flash-loan feed but is derived from TVL-volatility dynamics, so it partially overlaps with TVLVol_t; the two terms are therefore not fully independent and their combined weight should be read as partially double-counting TVL-volatility stress rather than as two distinct liquidity signals. A direct on-chain flash-loan-volume metric is the preferable specification and is flagged for future revision.

Data Sources: DeFi Llama (TVL, protocol data), Token Terminal (flash loan data), DefiSafety (audit scores).

3.6 Contagion Risk (25%)

The Contagion Risk sub-index quantifies DeFi-TradFi linkage intensity and cross-market transmission channels:

$$\begin{aligned} \text{CR}_t = 0.30 \cdot \text{RWA}_t + 0.25 \cdot \text{Bank}_t \\ + 0.20 \cdot \text{Link}_t + 0.15 \cdot \text{Corr}_t + 0.10 \cdot \text{Bridge}_t \end{aligned} \tag{4}$$

where:

- RWA_t is the 30-day growth rate of tokenised real-world asset TVL
- Bank_t is a normalised score of bank crypto exposure from regulatory filings (OCC, ECB)
- Link_t measures stablecoin flows to TradFi-connected entities

- Corr_t is the 30-day rolling correlation between BTC/ETH and S&P 500
- Bridge_t is a composite of cross-chain bridge volume and recent exploit frequency

Data Sources: RWA.xyz (tokenised assets), DeFi Llama (bridge data), FRED (equity indices), regulatory filings.

The Role of Bank$_t$. The banking stress proxy Bank_t occupies a theoretically central position in the Contagion Risk sub-index: it is the primary channel through which traditional financial sector distress propagates into cryptocurrency markets. Conceptually, Bank_t captures the health of banks with crypto exposure—institutions whose balance sheet stress directly affects crypto-TradFi linkages through stablecoin reserve exposure, custodial relationships, and lending facilities. The March 2023 SVB crisis demonstrated this channel concretely: banking sector stress transmitted into DeFi through USDC's exposure to SVB deposits, causing a stablecoin depeg that briefly destabilised the broader ecosystem. In the empirical implementation, Bank_t is operationalised as a Treasury-VIX composite (see Appendix A), reflecting the two primary mechanisms through which banking stress manifests: Treasury yield movements that affect bank capital ratios via mark-to-market losses, and equity volatility that signals broader risk-off conditions constraining bank lending and risk appetite. This proxy achieves daily frequency, overcoming the 45–90 day publication lag of quarterly regulatory filings from which the theoretical specification derives. We flag a theory-versus-evidence tension that the fair-baseline analysis (Section 5.7, Table 9) makes explicit. The mark-to-market rationale treats the Treasury-yield *level* as the primary channel, yet on our crisis labels that level discriminates at chance (AUROC 0.501) and, entering Bank_t at 60% weight, it *lowers* the composite's discrimination from 0.875 (VIX alone) to 0.773—a 0.102 reduction. The reduction is plausibly a trend artefact rather than a refutation of the channel: the level specification imports the 2021–2024 monetary-tightening trend (≈0.9 percentage points per year) into Bank_t rather than isolating crisis-specific banking stress. That removing such a trend recovers discrimination is shown by the reproducible detrended-Contagion channel, which rises to 0.912 (Table 9); a change- or detrended-based Treasury input is therefore the preferable specification, which we flag as a refinement for future revisions.

Implementation Note: Bank_t and Link_t are implemented as high-frequency proxies because quarterly regulatory filings have 45–90 day publication lags. Link_t uses yield curve spread. These proxies capture the same underlying stress dynamics at daily frequency (see Appendix A for full specification and Table 10 for proxy validation).

3.7 Regulatory Opacity Risk (20%)

The Opacity Risk sub-index assesses transparency deficits and regulatory arbitrage exposure:

$$\begin{aligned} \text{OR}_t = {} & 0.25 \cdot \text{Unreg}_t + 0.25 \cdot \text{Multi}_t \\ & + 0.20 \cdot \text{Cust}_t + 0.15 \cdot \text{Sent}_t + 0.15 \cdot (100 - \text{Trans}_t) \end{aligned} \tag{5}$$

where:

- Unreg_t is the ratio of unregulated to regulated platform volume

- Multi_t captures multi-issuer stablecoin scheme exposure
- Cust_t is custody concentration in non-audited jurisdictions
- Sent_t is regulatory sentiment score from NLP analysis of SEC/ESRB/FSB announcements
- Trans_t is a protocol-transparency score. The theoretical specification is reserve-attestation frequency and coverage; the implemented backtest proxies this by *audit coverage*—the share of non-zero-TVL protocols with at least one audit (Appendix A, Table 10)—because consistent historical attestation calendars are unavailable

Data Sources: DefiSafety audit coverage (implemented proxy); regulatory filings, news APIs (GDELT), and attestation calendars (theoretical specification).

3.8 Aggregate ASRI Calculation

The final ASRI is computed as a weighted sum of normalised sub-indices:

$$\text{ASRI}_t = 0.30 \cdot \text{SCR}_t + 0.25 \cdot \text{DLR}_t + 0.25 \cdot \text{CR}_t + 0.20 \cdot \text{OR}_t \tag{6}$$

Normalisation uses min-max scaling over the historical sample to produce a 0–100 index:

$$\text{ASRI}_t^{\text{norm}} = 100 \times \frac{\text{ASRI}_t - \min_\tau(\text{ASRI}_\tau)}{\max_\tau(\text{ASRI}_\tau) - \min_\tau(\text{ASRI}_\tau)} \tag{7}$$

Note that the component mappings detailed in Appendix A—the piecewise HHI risk score, the bounded `normalize`$(\cdot)$ transforms, and the $\times 100$ scalings—constrain each sub-index to $[0, 100]$; the displayed sub-index expressions (Equations 2–5) give the underlying raw inputs (for example, the TVL-drawdown ratio in $[0, 1]$ and the raw HHI) prior to those mappings, and are not themselves bounded to $[0, 100]$. Because the Appendix A mappings already bound the sub-indices, post-hoc min-max normalisation is unnecessary in practice. Equation 7 documents the theoretical relationship between raw and normalised values; empirical analyses in Section 5 use raw weighted aggregates directly. Collinearity among sub-indices is assessed via Variance Inflation Factors and condition number analysis (Section F.2); all diagnostics confirm linear aggregation is well-conditioned.

Alert Thresholds:

- $\text{ASRI} < 30$: Low systemic risk
- $30 \leq \text{ASRI} < 50$: Moderate systemic risk
- $50 \leq \text{ASRI} < 70$: Elevated systemic risk
- $\text{ASRI} \geq 70$: High systemic risk

Operational Alert Rule: An alert is triggered when $\text{ASRI} \geq 50$ (Elevated threshold) for at least one trading day. No persistence requirement is imposed because crisis dynamics can evolve rapidly; however, practitioners may implement confirmation windows (e.g., 3-day persistence) to reduce noise at the cost of lead time. Threshold selection follows a precision-recall trade-off documented in Table 12: the 50 threshold maximises crisis-level recall (75%, i.e. 3/4 events;

Terra/Luna is missed) while accepting moderate precision (30.1%); raising the threshold to 70 improves precision to 41.7% but lowers crisis-level recall to 50% (2/4 events). These thresholds were chosen based on interpretability (round numbers mapping to verbal risk categories) rather than statistical optimisation; ROC-based calibration is deferred to future work with larger crisis samples. A structural caveat applies to any *fixed* cut-off: because the Contagion sub-index is non-stationary on the released sample (Table 4), the composite's baseline level can drift, so a constant threshold such as 50 does not correspond to a stable exceedance probability across the full history. The walk-forward rule of Section 5.8, which sets the alert at a trailing pre-crisis percentile (the 90th percentile of the index's own training-period history) rather than at a fixed level, is the more robust operational specification under this drift; we recommend it for deployment and retain the fixed-50 rule here for interpretability and comparability with the verbal risk bands. A fuller remedy—recalibrating the alert on a differenced or rolling-baseline series so that a fixed nominal cut-off maps to a stable exceedance probability, rather than reading a raw level against a static line on a drifting index—is a natural extension we flag for future work; the four-event sample here is too small to calibrate such a drift-adjusted threshold with power, which is why we report the fixed-50 detection alongside the drift-robust walk-forward percentile rather than substituting one for the other.

4 Data and Implementation

4.1 Data Sources

Table 1 summarises the data sources for each ASRI component.

Tier 1 sources provide daily automated API access; **Tier 2** sources require manual collection, web scraping, or have lower update frequency.

4.2 Data Quality Framework

Data lag assumptions follow the $t-1$ convention: values observed at midnight UTC on date t are attributed to ASRI_{t-1} to avoid look-ahead bias.[4] Several ASRI components draw on lower-than-daily sources (quarterly OCC/ECB bank-exposure filings, monthly stablecoin attestations, weekly on-chain linkage metrics): where a high-frequency proxy exists—most importantly the Treasury–VIX composite that stands in for Bank_t—we use it, and where none exists we carry the last observation forward with explicit confidence degradation. The full missing-data and mixed-frequency protocol is given in Appendix D, and the pseudo-real-time evaluation (Appendix N) confirms that this protocol preserves detection performance under realistic publication lags. The system's four-layer data pipeline—ingestion, normalisation, computation, and publication—and its technology stack are described in Appendix C.

[4] To ensure backtests use only real-time information, min-max bounds in Equation 7 are theoretical; empirical analyses use raw ASRI values computed from bounded sub-indices without full-sample normalisation.

Table 1: ASRI Data Sources by Sub-Index

Sub-Index	Component	Source	Frequency	Tier
Stablecoin Risk	TVL	DeFi Llama	Daily	1
	Treasury Reserves	Attestation Reports	Monthly	2
	Market Share	CoinGecko	Daily	1
	Peg Volatility	CoinGecko	Daily	1
DeFi Liquidity	Protocol TVL	DeFi Llama	Daily	1
	Flash Loan Volume	Token Terminal	Daily	1
	Smart Contract Scores	DefiSafety	Weekly	2
	Leverage Ratios	DeFi Llama	Daily	1
Contagion Risk	RWA TVL	RWA.xyz / DeFi Llama	Daily	1
	Bank Exposure	OCC/ECB Filings	Quarterly	2
	TradFi Linkages	On-chain Analysis	Weekly	2
	Equity Correlation	FRED/Yahoo Finance	Daily	1
	Bridge Volume	DeFi Llama	Daily	1
	Bridge Exploits	DeFi Llama	Daily	1
Opacity Risk	Platform Regulation	Manual Tracking	Weekly	2
	Custody Concentration	Public Disclosures	Monthly	2
	Regulatory Sentiment	GDELT/SEC Filings	Daily	2
	Attestation Frequency	Calendar Tracking	Daily	2
	Transparency Scores	DefiSafety	Weekly	2

Abbreviations: TVL = Total Value Locked; RWA = Real-World Assets; OCC = Office of the Comptroller of the Currency; ECB = European Central Bank; GDELT = Global Database of Events, Language, and Tone; SEC = Securities and Exchange Commission.

5 Empirical Validation

This section presents the empirical validation of the ASRI framework against historical data from January 2021 through January 2026, comprising over 1,800 daily observations across four major in-sample crisis events and out-of-sample specificity testing on 2024–2025 data.

5.1 Crisis Taxonomy and Operational Definitions

Before proceeding to empirical validation, we establish operational definitions for what constitutes a systemic crisis in cryptocurrency markets. This taxonomy serves two purposes: providing ex ante criteria for event identification (avoiding post hoc selection bias) and enabling systematic classification of crisis mechanisms.

5.1.1 Operational Crisis Definition

Following the crisis identification methodology of Laeven and Valencia (2013) and adapted for high-frequency digital asset markets, we define a *systemic stress event* as satisfying three jointly necessary conditions:

Definition 5.1 (Systemic Stress Event)**.** *A period* $[t_0, t_1]$ *constitutes a systemic stress event if and only if:*

(i) ***Magnitude****: Aggregate market capitalisation decline* $\geq 15\%$ *within a 7-day window, or single-asset collapse* $\geq 50\%$ *for assets with market cap* $\geq \$10B$*;*

(ii) ***Contagion****: Cross-asset correlation surge, measured as* $\bar{\rho}_t - \bar{\rho}_{t-30} \geq 0.20$ *where* $\bar{\rho}$ *denotes the mean pairwise correlation across major assets;*

(iii) ***Duration****: Elevated stress conditions persist for* ≥ 5 *trading days, distinguishing systemic events from flash crashes.*

This definition deliberately excludes *stress episodes*—periods of elevated volatility without systemic propagation. For instance, single-asset drawdowns (e.g., meme coin collapses) or brief correlation spikes during scheduled events (FOMC announcements) fail condition (ii) or (iii) respectively. The thresholds are calibrated to cryptocurrency market dynamics; traditional finance definitions (e.g., Reinhart and Rogoff, 2009) typically require banking sector involvement, which maps imperfectly to decentralised systems.

5.1.2 Crisis Typology

We classify systemic events along two dimensions: origin (endogenous vs. exogenous) and primary transmission mechanism (liquidity vs. solvency). This yields a typology summarised in Table 2.

Table 2: Crisis Typology for Cryptocurrency Markets

Type	Characteristics	Historical Examples
Type I: Endogenous	Originates within DeFi/crypto ecosystem; propagates through on-chain liquidity channels, collateral cascades, or protocol failures	Terra/Luna (May 2022): algorithmic stablecoin death spiral triggering $40B TVL collapse
Type II: Exogenous	External shock (TradFi, regulatory, macroeconomic) propagates into crypto markets through stablecoin pegs, institutional exposure, or sentiment channels	SVB Crisis (March 2023): banking contagion → USDC depeg → DeFi stress
Type III: Hybrid	Combined dynamics—crypto-native entity failures with significant TradFi counterparty exposure amplifying propagation	Celsius/3AC (June 2022); FTX (November 2022): CeFi insolvencies with cross-market contagion

The four validation events span all three types, providing heterogeneous test conditions. Notably, Type III events (Celsius/3AC, FTX) exhibit the longest stress durations in our sample, consistent with Brunnermeier (2009)'s observation that hybrid crises produce more persistent dislocation due to opacity in cross-market exposures.

5.1.3 Detection versus Prediction

A critical methodological distinction: ASRI is *designed* as a *detection* instrument with intended *leading properties*, not a pure prediction model. This is a statement of design intent, not a validated capability—consistent with our framing of ASRI as an interpretable, retrospective monitor rather than a validated early-warning system (Section 6.3). Following Borio and Drehmann (2009), we distinguish:

- **Detection**: Contemporaneous identification that systemic stress is occurring or imminent

(hours to days horizon)—the *design* criterion, not a validated forecast. Design criterion: Does ASRI breach threshold τ before or coincident with observable market stress?

- **Prediction**: Forecasting crisis probability over medium horizons (weeks to months). Would require different validation methodology (e.g., receiver operating characteristic analysis, out-of-sample forecasting).

Our retrospective evaluation focuses on detection performance: lead time relative to price cascade initiation, signal persistence during stress periods, and false-positive rates during non-crisis windows. The lead times documented below characterise ASRI as a threshold-based monitor evaluated in hindsight, not as a validated long-horizon forecasting tool; we read positive lead times as descriptive of the historical episodes rather than as a forward-looking guarantee.

Measurement Definitions. For consistency across all empirical analyses, we define:

- **Detection threshold**: ASRI ≥ 50 ("Elevated" risk category). A crisis is detected if ASRI exceeds this threshold on any day in the 30-day pre-crisis window.
- **Lead time (threshold-based)**: Days between the *first* threshold crossing (ASRI ≥ 50) and crisis onset ($t = 0$, defined as price cascade initiation).
- **Lead time (event study)**: Days between first observation where ASRI exceeds 1.5 standard deviations above the estimation-window mean and crisis onset. This captures early stress signals relative to the pre-event baseline.
- **Crisis onset ($t = 0$)**: The first trading day exhibiting observable price cascade—typically a 10%+ single-day decline in major assets or stablecoin depeg initiation.
- **False positive**: Any day where ASRI ≥ 50 outside of the $[-30, +30]$ window surrounding a validated crisis event.

All hypothesis tests are two-tailed at $\alpha = 0.05$ unless otherwise specified. Confidence intervals are reported as 95% intervals using bootstrap resampling (1,000 iterations) for detection rates and analytical standard errors for regression coefficients.

5.2 Data and Sample

Table 3 presents descriptive statistics for ASRI and its component sub-indices.

Table 3: Descriptive Statistics of ASRI Components

Variable	N	Mean	Std	Min	Max	Skew
ASRI	1,841	39.2	7.8	25.8	84.7	1.46
Stablecoin Risk	1,841	35.5	9.1	14.0	78.2	0.27
DeFi Liquidity	1,841	42.3	7.5	27.9	90.0	1.79
Contagion Risk	1,841	39.1	13.8	12.1	87.9	−0.34
Opacity Risk	1,841	36.9	6.9	22.6	70.2	0.77

Sample: January 2021 – January 2026 (daily).

The ASRI ranges from 25.8 (low risk) to 84.7 (elevated risk during the FTX crisis), with crisis periods driving the upper tail. Positive skewness (1.46) reflects the asymmetric distribution: most observations cluster in the moderate band (30–50) while systemic stress events generate right-tail outliers, consistent with the design objective of early warning rather than tail risk measurement.

Figure 3 presents the complete ASRI timeseries from January 2021 through January 2026, with the four validated crisis events annotated and operational risk regime bands indicated.

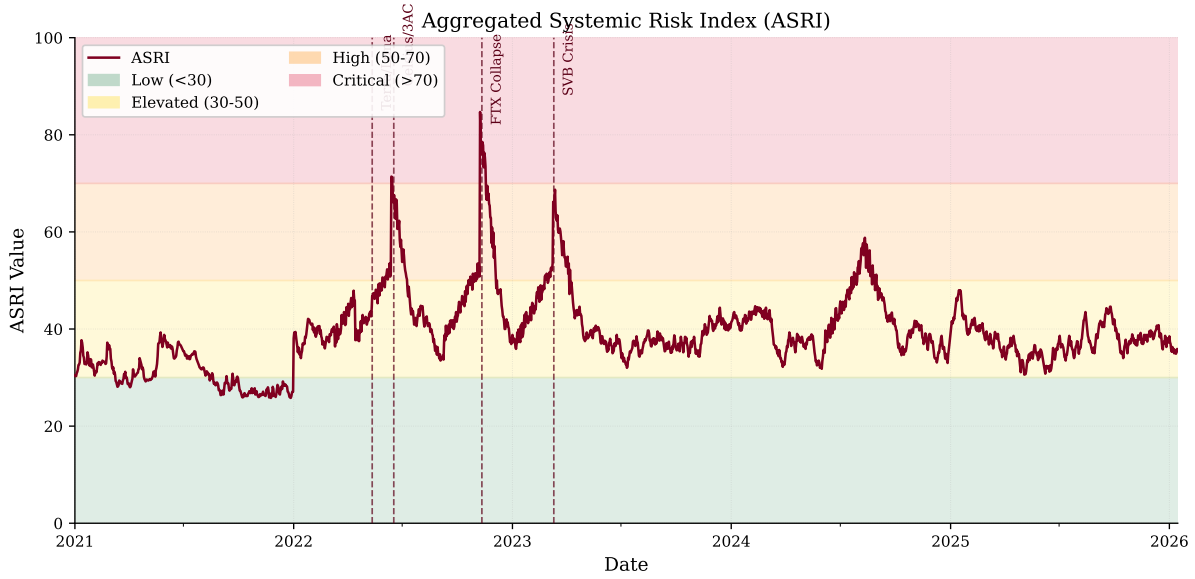

Figure 3: ASRI Index Timeseries (January 2021 – January 2026). Shaded bands denote operational risk regimes: Low (<30), Moderate (30–50), Elevated (50–70), and High (≥70). Vertical dashed lines mark crisis event onsets: Terra/Luna collapse (May 2022), Celsius/3AC contagion (June 2022), FTX collapse (November 2022), and SVB banking crisis (March 2023). The index exhibits characteristic spikes during stress periods with rapid mean-reversion during recovery phases.

The visual pattern confirms the statistical properties reported in Table 3: the index spends most of its time in the Low-to-Moderate bands (25–50), with transient spikes into Elevated and High zones during the four major crises.

5.3 Stationarity Tests

Valid time series analysis requires stationary sub-indices. Table 4 reports Augmented Dickey-Fuller (ADF) (Dickey and Fuller, 1979) and KPSS (Kwiatkowski et al., 1992) test results.

Table 4: Stationarity Test Results

Variable	ADF Stat	ADF p	KPSS	Conclusion
ASRI	−4.05	0.001	0.65	Stationary
Stablecoin Risk	−4.09	0.001	0.81	Trend-stat.
DeFi Liquidity	−4.32	<0.001	0.69	Stationary
Contagion Risk	−2.73	0.069	2.06	Non-stat.
Opacity Risk	−4.30	<0.001	2.94	Trend-stat.

ADF: Augmented Dickey-Fuller (lag selection via AIC, intercept included); KPSS: Kwiatkowski-Phillips-Schmidt-Shin (Bartlett kernel, automatic bandwidth). Computed on the released full-sample series (2021–2026).
KPSS critical values: 0.463 (5%), 0.739 (1%). Values exceeding 0.739 indicate, at best, trend-stationarity (the level is not covariance-stationary).
ASRI and DeFi Liquidity reject the unit root with KPSS consistent with (level/trend) stationarity. Stablecoin and Opacity Risk reject the unit root but exhibit elevated KPSS, indicating deterministic trends. Contagion Risk fails to reject the unit root (ADF $p = 0.069$; KPSS $= 2.06$) and is treated as non-stationary.

ASRI and the DeFi Liquidity sub-index are stationary; Stablecoin and Opacity Risk reject the unit root but display deterministic trends (elevated KPSS), and are best characterised as trend-stationary. Contagion Risk is the exception: it fails to reject the unit root on the released sample (ADF $p = 0.069$ full-sample, $p = 0.18$ on the 2021–2024 window) and carries a large KPSS statistic, so its *level* is not covariance-stationary. We retain level-based analysis for the composite and the three stationary/trend-stationary channels, and flag that inferences resting specifically on the Contagion level (e.g. its variance-decomposition contribution) should be read with this caveat. Differencing Contagion Risk yields a stationary series, but we report levels for interpretability and consistency with the operational threshold definition.

5.4 Event Study Analysis

We apply formal event study methodology to assess ASRI behaviour around four major crisis events. Following MacKinlay (1997), we estimate "normal" ASRI levels from a 60-observation pre-event window (days −90 to −31 relative to event onset) and compute Cumulative Abnormal Signal (CAS) over the event window.[5]

5.4.1 Event Study Specification

Normal Model. We employ a constant mean model for expected ASRI during the estimation window:

$$\mathbb{E}[\text{ASRI}_t] = \hat{\mu} = \frac{1}{T_{\text{est}}} \sum_{\tau=-90}^{-31} \text{ASRI}_\tau \tag{8}$$

[5]Unlike asset returns, ASRI is bounded on $[0, 100]$. However, large-sample inference remains valid under the central limit theorem; the bounded support makes distributional assumptions less critical than in return-based event studies.

where the estimation window spans $t = -90$ to $t = -31$ relative to the event date (60 trading days), providing a pre-event baseline uncontaminated by crisis dynamics.

Abnormal Signal. The abnormal signal on day t is defined as the deviation from the expected level:

$$\mathrm{AS}_t = \mathrm{ASRI}_t - \mathbb{E}[\mathrm{ASRI}_t] = \mathrm{ASRI}_t - \hat{\mu} \tag{9}$$

The Cumulative Abnormal Signal (CAS) aggregates abnormal signals over the event window $[t_1, t_2]$:

$$\mathrm{CAS}_{[t_1,t_2]} = \sum_{t=t_1}^{t_2} \mathrm{AS}_t \tag{10}$$

We use an event window of $[t_1, t_2] = [-30, +10]$, capturing the pre-crisis buildup period and immediate aftermath.

Variance Estimation. The variance of the abnormal signal is estimated from the estimation window:

$$\hat{\sigma}^2_{\mathrm{AS}} = \frac{1}{T_{\mathrm{est}} - 1} \sum_{\tau=-90}^{-31} (\mathrm{ASRI}_\tau - \hat{\mu})^2 \tag{11}$$

A naive standard error that assumes independent abnormal signals would be $\hat{\sigma}_{\mathrm{AS}}\sqrt{T_{\mathrm{event}}}$, with $T_{\mathrm{event}} = 41$. This independence assumption is, however, untenable for ASRI: the cumulative abnormal signal sums a 41-day window of a series whose components are 30-day rolling quantities, so the abnormal-signal series is strongly positively autocorrelated by construction (event-window AR(1) ≈ 0.78–0.90 across events). A Ljung–Box test decisively rejects white noise—p-values from 10^{-15} to 10^{-28} on the event-window abnormal series, and as low as 2×10^{-58} on the estimation-window residuals—so the naive variance understates Var(CAS) by roughly an order of magnitude. We therefore report Newey–West (Newey and West, 1987) heteroskedasticity- and autocorrelation-consistent (HAC) standard errors, obtained by regressing the event-window abnormal series on a constant with a HAC covariance estimator at lag $L = 20$ (chosen to span the 30-day rolling-window construction; the intercept HAC t equals the CAS HAC t because $\mathrm{CAS} = n\bar{\mathrm{AS}}$):

$$\mathrm{SE}_{\mathrm{HAC}}(\mathrm{CAS}) = n \times \mathrm{SE}_{\mathrm{HAC}}(\bar{\mathrm{AS}}), \tag{12}$$

yielding $\mathrm{SE}_{\mathrm{HAC}}(\mathrm{CAS}) = 58.3$ (Terra/Luna), 149.5 (Celsius/3AC), 231.3 (FTX), and 132.0 (SVB).

Test Statistic. Under the null hypothesis of no abnormal signal (H_0: CAS $= 0$), the HAC test statistic is:

$$t_{\mathrm{HAC}} = \frac{\mathrm{CAS}}{\mathrm{SE}_{\mathrm{HAC}}(\mathrm{CAS})} \xrightarrow{d} \mathcal{N}(0, 1) \tag{13}$$

referenced asymptotically against the standard normal distribution. For comparison we also report the naive $t = \mathrm{CAS}/(\hat{\sigma}_{\mathrm{AS}}\sqrt{T_{\mathrm{event}}})$ that the prior draft relied on; the gap between the two is the quantitative cost of the false independence assumption.

Finite-sample (fixed-b) reference. The asymptotic $\mathcal{N}(0, 1)$ reference is itself unreliable in this small sample. Because the cumulative abnormal signal sums a 41-day window of 30-day-rolling components (AR(1) ≈ 0.8–0.9), the Newey–West bandwidth is a large fraction of the sample ($b = (L + 1)/T = 21/41 \approx 0.51$; the Bartlett kernel at lag truncation $L = 20$ carries

$L+1$ nonzero weights, so the Kiefer–Vogelsang bandwidth is $(L+1)/T$, not the lag fraction L/T); at that bandwidth-to-sample ratio the HAC t-statistic is far from standard normal, and referencing it against $\mathcal{N}(0,1)$ overstates significance. We therefore adopt a Kiefer–Vogelsang fixed-b reference at the realised bandwidth as the inference of record (Kiefer and Vogelsang, 2005), whose critical values absorb the fat tails the small sample induces: $|t| \approx 3.52$ at the 5% level (not 1.96) and $|t| \approx 5.73$ at the Bonferroni-adjusted $\alpha = 0.0125$. A t-distribution with data-driven effective degrees of freedom ($\mathrm{df}_{\mathrm{eff}} \approx 2.3$–$5.0$) yields the same reading. Under this reference (Table 5), the three large-CAS events are individually marginal ($p \approx 0.04$–0.06), none clears the Bonferroni threshold, and Terra/Luna is clearly non-significant ($p \approx 0.27$). The qualitative result—three events elevate, Terra/Luna does not—is unchanged; the precise significance level is borderline-at-5% rather than Bonferroni-surviving.

Window Independence (and its limits for the clustered 2022 events). Within each event the 60-day estimation window precedes, and does not overlap, that event's own 41-day event window. *Across* events, however, the two early-2022 crises cluster too closely for their windows to be independent; the later FTX and SVB windows are well separated:

- Terra/Luna (May 2022): Estimation window February–April 2022
- Celsius/3AC (June 2022): Estimation window March–May 2022
- FTX Collapse (November 2022): Estimation window August–October 2022
- SVB Crisis (March 2023): Estimation window December 2022–February 2023

Because Terra/Luna and Celsius/3AC fall within roughly five weeks of one another, their windows overlap: their event windows share ≈ 5 days, their *estimation* windows overlap by ≈ 24 days (Terra/Luna's runs to mid-April 2022, Celsius/3AC's begins in mid-March), and Celsius/3AC's pre-event "baseline" window itself spans the Terra/Luna event window by ≈ 35 days—so that baseline is not uncontaminated by the preceding Terra/Luna crash. We therefore do *not* treat the two early-2022 events as statistically independent: they share estimation and baseline information, and the hold-one-out windows for these clustered events (Supplement) are correspondingly not fully independent. This is a genuine limitation of a sample whose crises cluster in 2022; it does not affect the FTX and SVB events, which are separated by over 90 days and have non-overlapping estimation and event windows.

Lead Time Measurement. Lead time is measured as days between the first observation where ASRI exceeds 1.5 standard deviations above the estimation-window mean and crisis onset. This definition captures early stress signals relative to the baseline rather than fixed threshold breaches, allowing detection of abnormality even when absolute levels remain below operational thresholds.

5.4.2 Event Study Results

Table 5: Event Study Results: ASRI Response to Crisis Events

Event	Date	Pre-Mean	Peak	CAS	Naive t	HAC t	$p_{\text{fix-}b}$	Lead
Terra/Luna	2022-05	40.4	48.7	100.3	5.47	1.72	0.265	30
Celsius/3AC	2022-06	42.6	71.4	521.6	29.78	3.49	0.051	30
FTX Collapse	2022-11	39.6	84.7	758.8	32.64	3.28	0.063	30
SVB Crisis	2023-03	41.0	68.7	509.0	26.91	3.86	0.036	29

Under the fixed-b finite-sample reference the three large-CAS events are marginal ($p \approx 0.04$–0.06), none clearing Bonferroni $\alpha = 0.0125$; Terra/Luna n.s. ($p \approx 0.27$). Average CAS: 472.4

CAS = Cumulative Abnormal Signal. Lead = days between the first ASRI exceedance of 1.5σ above the estimation-window mean (searched over the 30-day pre-event window) and event onset. $p_{\text{fix-}b}$ is the p-value under a Kiefer–Vogelsang fixed-b reference at the realised bandwidth-to-sample ratio $b \approx 0.51$ ($= (L+1)/T = 21/41$ under the Newey–West/Bartlett convention; 5% critical $|t| \approx 3.52$; Bonferroni-0.0125 critical $|t| \approx 5.73$); a t-reference with data-driven effective degrees of freedom ($\text{df}_{\text{eff}} \approx 2.3$–$5.0$) gives $p \approx 0.05$ for the three large-CAS events and $p \approx 0.15$ for Terra/Luna. Against the naive $\mathcal{N}(0,1)$ reference the same HAC t would read $p = 0.0005$–0.001, but that reference is invalid at this bandwidth and overstates significance.

Naive $t = \text{CAS}/(\hat{\sigma}_{\text{AS}}\sqrt{41})$ assumes independent abnormal signals and is reported only for comparison; it is invalid here because the abnormal-signal series is strongly autocorrelated (event-window AR(1) ≈ 0.78–0.90; Ljung–Box rejects white noise at $p < 10^{-15}$). HAC t uses Newey–West standard errors at lag $L = 20$. Under HAC, Terra/Luna falls from 5.47 to 1.72; a fixed-b finite-sample reference would render the three large-CAS events marginal ($p \approx 0.04$–0.06), but a placebo/specificity test shows that reference has no crisis-specificity on this smooth trending series, so the event study is treated as inconclusive (see text). A lag sweep ($L = 10/20/30$, each spanning the 30-day rolling-window construction) leaves Terra/Luna non-significant at every lag and the other three at $t \geq 3.28$ throughout (the minimum being FTX at $L = 20$); Terra reaches the 5% level only at the Newey–West auto-floor $L = 3$, which is too short to capture the construction-induced persistence.

t-statistics use the canonical profile from the released code: estimation window $[-90, -31]$, event window $[-30, +10]$, lead-time lookback capped at 30 days. The reported leads (29–30) are at this cap: the 1.5σ signal is elevated across essentially the entire pre-window (uncapped first-crossing 38–90 days), so these values indicate sustained pre-crisis elevation rather than a precise 30-day horizon. The walk-forward first-crossing leads (Table 34, mean 25.75 days) are the more conservative operational figure.

Under autocorrelation-robust (HAC, $L = 20$) inference the three large-CAS events (Celsius/3AC, FTX, SVB) carry HAC t-statistics of 3.49, 3.28, and 3.86, while Terra/Luna carries 1.72. Referenced naively against $\mathcal{N}(0,1)$ the first three would read $p < 0.01$, but that reference is invalid at the realised bandwidth ($b \approx 0.51 = 21/41$), where the abnormal-signal series is heavily serially correlated (event-window AR(1) ≈ 0.8–0.9; Ljung–Box rejects white noise at $p < 10^{-15}$) and the naive standard error understates Var(CAS) by roughly an order of magnitude. The natural next step—a finite-sample (fixed-b) reference—would downgrade the three to "borderline" ($p \approx 0.04$–0.06). But that reference is itself size-invalid on this series: running the identical pipeline on non-crisis dates (a placebo/specificity test) shows the fixed-b 5% critical value is

cleared by roughly sixty percent of arbitrary dates—an empirical false-positive rate an order of magnitude above nominal—and the four crisis onsets sit at or below the placebo median. On a smooth, strongly trending series the cumulative-abnormal-signal statistic has no crisis-specific power here.

Interpretation: We therefore treat the event study as *inconclusive*. It neither establishes nor refutes crisis-specific abnormal elevation: the raw CAS ordering is intuitive (FTX largest at 758.8, Terra/Luna smallest at 100.3, consistent with the difficulty of observing algorithmic-stablecoin fragility through market-based indicators), but because no onset separates from the placebo distribution, we do not read any of these as a crisis-specific signal and do not count the event study as evidence of detection. The empirical case for ASRI rests instead on the fair-baseline day-level discrimination analysis (Section 5.7), which does not depend on the event-study inference.

Detection Nomenclature: Throughout this section we distinguish *threshold-based detection* (ASRI ≥ 50 during the pre-crisis window) from the *event study* (abnormal ASRI elevation around an onset). Threshold-based analysis achieves 3/4 detection (Terra/Luna missed, peak 48.7); the event study is indeterminate on this sample for the reasons above and is not counted toward detection. Walk-forward validation flags 4/4 out-of-sample, but at a high false-positive cost for the early crises (Section 5.8), so we read it as evidence against look-ahead bias rather than a clean detection record.

Table 6 presents the unified detection matrix. Threshold-based detection is 3/4 (Terra/Luna falls below the operational threshold, peak 48.7); the Event-Sig column records the raw HAC t-statistics but assigns no crisis-specific significance, since none of the onsets separates from the placebo distribution.

Table 6: Unified Detection Matrix: Method Comparison

Event	Peak	τ=40	τ=50	τ=60	τ=70	Event Sig.	HAC t	WF-OOS
Terra/Luna	48.7	✓	×	×	×	×	1.72	✓
Celsius/3AC	71.4	✓	✓	✓	✓	×	3.49	✓
FTX Collapse	84.7	✓	✓	✓	✓	×	3.28	✓
SVB Crisis	68.7	✓	✓	✓	×	×	3.86	✓
Total		4/4	3/4	3/4	2/4	0/4		4/4

τ = threshold-based detection (ASRI $\geq \tau$ in 30-day pre-crisis window).
Event Sig. = event-study significance under autocorrelation-robust (Newey–West HAC, $L = 20$) inference. No event is marked significant: a placebo test on non-crisis dates shows the finite-sample (fixed-b) reference has no crisis-specificity on this smooth trending series (empirical false-positive rate ≈50–60% at nominal 5%; the onsets sit at or below the placebo median), so the raw HAC t-statistics (Terra/Luna 1.72; the others 3.28–3.86) are reported in the HAC t column but not read as crisis detection. The event study is treated as inconclusive throughout.
WF-OOS = walk-forward out-of-sample flag (90th-percentile threshold on training data); see Section 5.8 for the associated false-positive cost.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Figure 4 visualises the ASRI trajectories across the four crisis events, illustrating the pre-event baseline levels and post-event peaks summarised in Table 5.

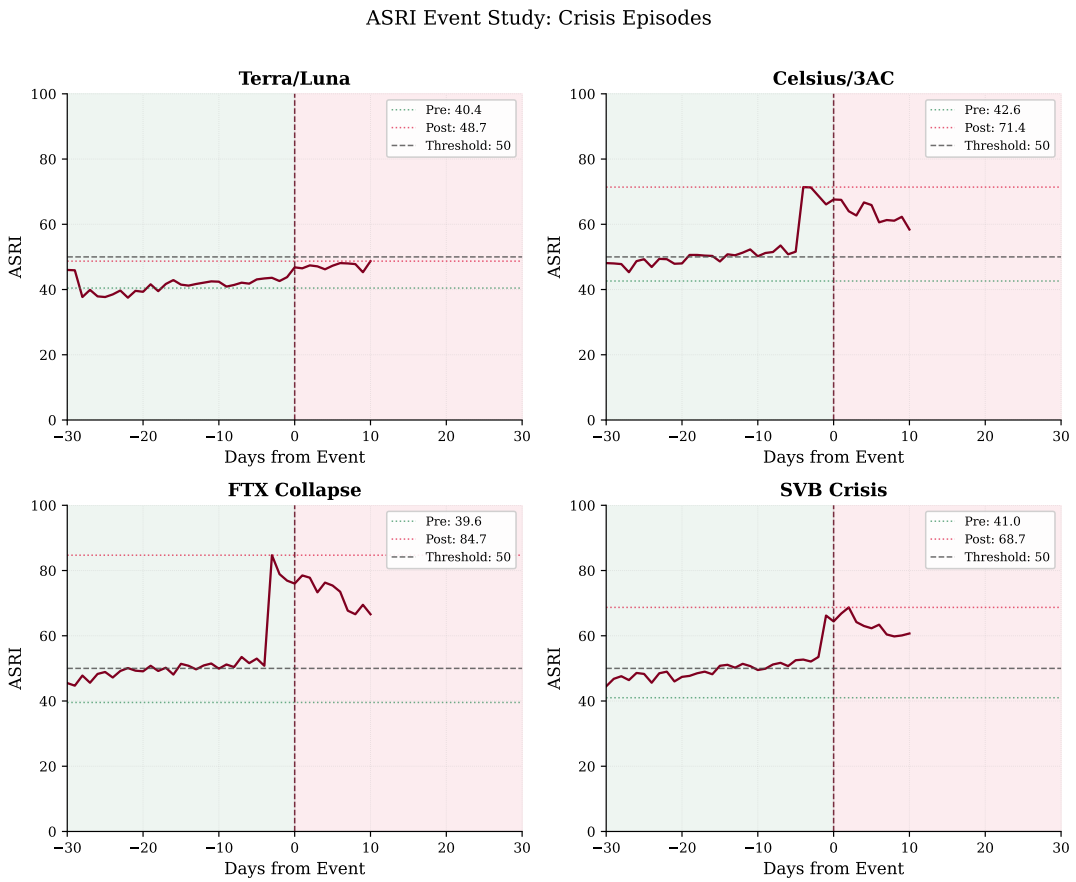

Figure 4: ASRI Event Study: Pre-Event and Post-Event Levels Across Four Crisis Events. Each panel shows the baseline ASRI mean during the 60-day estimation window (Pre) and the peak ASRI value during the event window (Post). All four events exhibit substantial elevations from baseline, with FTX Collapse showing the largest absolute increase (39.6 $\rightarrow$ 84.7) and Terra/Luna showing the smallest (40.4 $\rightarrow$ 48.7).

5.4.3 Reconciling Lead Time Definitions

Reconciling Lead Time Definitions. Lead time is reported under several distinct operational definitions, and the resulting figures are not directly comparable; we anchor on three:

- **Threshold-based (fixed ASRI ≥ 50, 30-day pre-window).** For the three threshold-detected events the first-crossing leads are Celsius/3AC 19, FTX 22, SVB 15 days (mean ≈19); Terra/Luna does not breach 50 in the pre-window (peak 46.0) and so has no fixed-threshold lead. This is the operational alerting definition. These fixed-threshold leads are partly inflated by the running-maximum drawdown offset (Section 6.3): under a responsive 30-day percentage-change drawdown specification they compress to a mean of ≈5 days (Celsius/3AC 12, FTX 3, SVB 1), which is the more honest early-warning horizon and the figure we lead with.
- **Event-study (first 1.5σ exceedance, search capped at 30 days).** Table 5 reports 29–30 for all four events; because the signal is elevated across essentially the whole pre-window, these are at the cap and indicate sustained elevation rather than a 30-day horizon.
- **Walk-forward (first crossing of the pre-crisis-calibrated threshold).** Table 34 reports 30/30/28/15 (mean 25.75), the conservative out-of-sample early-warning figure.

We report the walk-forward first-crossing lead (≈26 days, 4/4 flagged out-of-sample) as a nominal lead horizon, with the important qualification that for the two early crises this crossing is bought at a 47–59% alarm rate over the full index history (Section 5.8); the other definitions are reported for completeness and indicate that the signal is elevated across essentially the entire pre-crisis window for the detected events.

The full precision–recall-by-threshold analysis, the confusion matrix at the operational threshold, and the ROC and precision–recall curves are reported in Appendix E; the headline AUROC (0.866) and precision figures also appear in the fair-baseline comparison (Section 5.7, Table 9).

5.5 Weight Derivation: Empirical vs. Theoretical

We compare the theoretical weights derived from risk-based principles against empirically-derived weights. An earlier draft reported a separate "theoretical vs. PCA vs. Elastic Net" table whose entries were hard-coded constants that did not reproduce from the weight-derivation modules and contradicted the principal-component loadings reported below; we have removed it. The single, code-faithful comparison across four objective methods is Table 14, and we summarise its implications here.

What the empirical weights show. The reproducible weight-derivation pipeline yields a Principal Component Analysis (PCA) loading vector of $[0.29, 0.29, 0.25, 0.17]$ over [SCR, DLR, CR, OR], which tracks the theoretical weights closely ($\rho = 0.88$). The Elastic Net, by contrast, concentrates on *Contagion Risk* (0.45) and zeroes Regulatory Opacity (0.00), with SCR (0.34) and DLR (0.21) intermediate. We flag this correction explicitly: the Elastic Net does *not* concentrate on DeFi Liquidity Risk—the earlier draft's hard-coded 0.84 DLR weight does not reproduce from the weighting code, and the code-faithful sparse solution loads on CR. CRITIC and entropy

methods diverge further still ($\rho = -0.51$ and 0.27 with the theoretical vector). Because the four objective methods disagree substantially among themselves, we do not treat any single empirical vector as "the" optimal weighting.

Elastic Net Specification: The predictive weights are derived via Elastic Net regression with the following specification:

- **Target variable**: $y_t = \text{ASRI}_{t+30}$ (30-day forward ASRI level)
- **Features**: Current sub-index values $[\text{SCR}_t, \text{DLR}_t, \text{CR}_t, \text{OR}_t]$
- **Cross-validation**: 5-fold blocked time-series CV to preserve temporal structure
- **Hyperparameter grid**: $\alpha \in \{0.1, 0.5, 1.0\}$, ℓ_1-ratio $\in \{0.1, 0.5, 0.9\}$
- **Software**: scikit-learn 1.3.x, Python 3.11

The optimal hyperparameters ($\alpha = 0.5$, ℓ_1-ratio $= 0.5$) produce a sparse solution that zeroes Regulatory Opacity and concentrates on Contagion Risk. This sparsity reflects moderate correlation among sub-indices during stress periods: when systemic stress materialises, multiple channels activate simultaneously, making it difficult for regularised regression to distinguish their individual contributions. Importantly, formal collinearity diagnostics (Table 15) confirm that this correlation does not constitute problematic multicollinearity—all variance inflation factors remain below 5 (max VIF = 3.89), and the condition number of 19.1 indicates weak collinearity well within acceptable bounds.

The Decomposition Is Not Individually Load-Bearing. We deliberately avoid building a strong mechanistic narrative on these empirical weights, because they are neither stable across methods nor individually decisive for detection. The ablation analysis (Appendix I) shows that threshold-based detection is *invariant* at 3/4 across all single-channel removals: no individual sub-index is necessary for the detected events. The four-channel decomposition therefore earns its place through *interpretability*—attributing aggregate stress to named, economically distinct channels—rather than through any one channel being statistically load-bearing. Lead-time variation under ablation is not a reliable guide to channel-level timing—with only four co-located crisis episodes a correctly specified discrete-outcome (logit/hazard) test does not identify which channels move earliest (Appendix F.3)—so we draw no leading-versus-confirming hierarchy, and we no longer attribute a "canary" role to DLR (the empirical weight that motivated that claim does not reproduce).

Rationale for Theoretical Weights. We retain theoretical weights for the headline specification. Three considerations motivate this choice:

1. **Component-level monitoring**: The theoretical decomposition enables targeted interpretation. Elevated SCR points to stablecoin reserve scrutiny; elevated CR to counterparty/macro exposure review. Pure prediction weights sacrifice this interpretability.
2. **Method disagreement**: The four objective weighting methods disagree among themselves (correlations with the theoretical vector range from -0.51 to 0.88), so no single

data-driven vector is a defensible replacement; the balanced theoretical weights avoid collapsing onto whichever channel a particular estimator happens to favour in this small sample.

3. **Structural stability**: Prediction-optimised weights are sample-dependent and may overfit to the four historical crises. Theoretical weights provide forward-looking stability for regimes that differ from the backtest period.

The empirical analysis thus informs interpretation rather than replacing theoretical structure, while the ablation invariance cautions that the channel split is a presentational device, not a load-bearing causal decomposition. This is consistent with the weight assignment rationale in Section 3.3.

The four-method objective weight comparison (Table 14), the collinearity and principal-component diagnostics (Appendix F.2), and the Granger-causality analysis (Appendix F.3) are reported in Appendix F.

5.6 Regime Detection

We estimate a Gaussian Hidden Markov Model (HMM) (Hamilton, 1989) to characterise market regimes from sub-index dynamics, treating the regime labels as interpretive structure over the index's autocorrelated dynamics rather than as evidence of sharply separated generating processes. The HMM is specified with full covariance matrices for each state, estimated via Expectation-Maximisation with convergence criterion $|\Delta \log L| < 10^{-4}$ and maximum 1,000 iterations. To mitigate sensitivity to initialisation, we run 10 random restarts and select the model with highest log-likelihood. Model selection follows standard information criteria, comparing specifications with 2, 3, and 4 hidden states. We flag one assumption violation at the outset: the Gaussian emissions impose time-invariant per-regime means, yet the Contagion sub-index is non-stationary with an upward drift (≈4.4 ASRI points per year) on this sample (Table 4). The fitted regimes therefore partly track that trend rather than purely systemic conditions, a point we quantify in the ergodic-distribution discussion below; we retain the level-based fit for interpretability and read the regimes accordingly. The appropriate robustness check—re-estimating the model on trend-stationary (differenced or detrended) inputs, or specifying autoregressive/time-varying-mean emissions that absorb the drift, so that the states reflect conditional dynamics rather than the level trend—would test whether the Crisis state survives once the trend is removed. We flag this as future work rather than run it here: with only four crisis episodes the regime labels cannot be validated against an independent event population regardless of the emission specification, so we do not present the three-regime fit as evidence of statistically distinct generating processes.

We summarise the retained three-state specification here; the HMM model-selection criteria (Table 18), the full transition matrix (Table 20), ergodic diagnostics (Table 19), regime-count robustness (Table 21), and the filtering-versus-smoothing discussion are deferred to Appendix G. Table 7 reports the characteristics of the three retained regimes.

Table 7: Regime Characteristics (3-State Model)

Regime	Frequency	Mean Risk	Persistence	Interpretation
1	19.8%	31.8	0.997	Low Risk
2	56.3%	36.8	0.992	Moderate
3	23.8%	48.2	0.980	Crisis

Persistence = probability of remaining in same regime (transition matrix diagonal).
Values from the best-by-log-likelihood fit among 10 random initialisations.

Interpreting Regime Labels. We label Regime 3 the "Crisis" regime because it carries the highest regime-conditional mean (48.2) and concentrates the historical stress episodes, even though that conditional mean still falls just within the upper Moderate alert band (30–50) rather than the High band ($\geq$70). The "Crisis" label reflects the statistical properties of the regime—highest volatility, elevated transition probability into and out of acute stress—rather than a claim that the instantaneous ASRI level remains at crisis levels throughout.

During the historical crises themselves, ASRI spiked into High ($\geq$70) zones, with event peaks reaching as high as 84.7 (FTX) and 71.4 (Celsius/3AC) above the 70 threshold (Table 5). These peaks occur as transient spikes within the Crisis regime before mean-reversion during recovery periods pulls the regime mean back towards the Moderate band. The regime mean of 48.2 thus represents a weighted average of stress spikes and subsequent recoveries, not a sustained crisis level.

Alert thresholds are designed to flag instantaneous risk levels requiring attention, while regime classifications capture the statistical dynamics that characterise market states over extended periods. The Crisis regime signals a market environment where acute stress events are significantly more likely to occur, even when the current ASRI reading may be temporarily moderate.

The three-regime model identifies:

- **Low Risk** (19.8% of sample): Mean ASRI of 31.8, very high persistence (0.997)
- **Moderate** (56.3%): Mean ASRI of 36.8, high persistence (0.992)
- **Crisis** (23.8%): Mean ASRI of 48.2, high persistence (0.980)

The high persistence across all regimes (diagonal elements > 0.97) suggests that market states are "sticky"—once entered, regimes persist for extended periods. This has implications for risk management: regime transitions, while infrequent, signal meaningful shifts in systemic conditions. We caution, however, against over-reading the regime structure: the three regime means span only 31.8–48.2 on the 0–100 scale, and the near-unit persistences largely restate the index's own serial correlation. We therefore present the HMM as descriptive, interpretive colour over an autocorrelated trend, and do not claim that it recovers generating states that a trivial three-bin partition of the ASRI level would miss.

5.7 Fair-Baseline Comparison

Benchmarking ASRI against a Diebold–Yilmaz (2012) connectedness series computed on its four sub-indices yields higher day-level discrimination for ASRI (AUROC 0.866 vs. 0.670; AUPRC 0.298 vs. 0.121; Table 8). That benchmark is, however, *constructed from* ASRI's own sub-indices—a partly circular comparator—and is itself out-discriminated by an off-the-shelf sentiment index, so it cannot establish that the four-channel aggregation adds value; the load-bearing comparison is the fair-baseline battery below. Full Diebold–Yilmaz methodology, lag selection, window sensitivity, and per-event results are reported in Appendix M.

Table 8: Crisis Prediction Classification Metrics with 95% Bootstrap Confidence Intervals

Metric	ASRI	D-Y Connectedness	Paired Difference
AUROC	0.866 [0.756, 0.949]	0.670 [0.526, 0.806]	+0.194 [+0.093, +0.298]***
AUPRC	0.298 [0.131, 0.518]	0.121 [0.046, 0.233]	+0.182 [+0.070, +0.337]***
Optimal Threshold	45.3	44.0%	—
Precision @ Optimal	0.352	0.149	+0.203
Recall @ Optimal	0.758	0.806	−0.048
F1 @ Optimal	0.481	0.251	+0.230

Computed on the real ASRI series and a genuine rolling Diebold-Yilmaz generalized-FEVD connectedness series (60-day VAR(1), $H = 10$), both scored on the same crisis labels. $n = 1{,}402$ aligned daily observations; 124 crisis-imminent days in four contiguous runs, 1,278 non-crisis days (prevalence 8.8%).

Confidence intervals from a **moving-block bootstrap** (Kunsch 1989; block length $L = 25$ days, $B = 2{,}000$, seed 42), which respects the serial dependence that i.i.d. day-level resampling ignores. The resulting marginal intervals are 3.3–4.7× wider than naive i.i.d. intervals (confirming the latter were ≈ 4× too narrow) and the ASRI and D-Y marginal intervals *overlap*, because the labels comprise only ≈ 4 independent crisis blocks. ASRI's advantage is therefore established by the *paired* block-bootstrap difference (same resampled blocks applied to both series): the AUROC and AUPRC differences are positive in all 2,000 resamples ($p < 0.001$), robust across $L = 20/25/30$.

Crisis defined as the 30-day forward period preceding any of the four historical crisis onsets (Terra/Luna, Celsius/3AC, FTX, SVB). All metrics derive from these four events; see Section 6.3.

*** paired difference > 0 in all 2,000 resamples ($p < 0.001$). Optimal threshold selected by Youden's J statistic (maximizes TPR − FPR). D-Y threshold expressed as total connectedness (%).

Precision, recall and F1 at the optimal threshold are computed on this 1,402-day common sample, where the ASRI Youden threshold is 45.3. On the full canonical sample (1,841 days; 120 crisis / 1,721 non-crisis days—the Table 12/Table 13 denominators) ASRI's precision at the same threshold 45.3 is 0.325, and at that sample's own Youden-optimal threshold (41.2) it falls to 0.198, because the larger non-crisis denominator admits more false-positive days. The AUROC, AUPRC and the paired block-bootstrap are unaffected, since both indices are scored on identical days. ASRI's precision advantage over D-Y is therefore a relative ordering, not high absolute precision.

A connectedness benchmark alone cannot establish that the four-channel *aggregation* is what produces ASRI's discrimination. We therefore evaluate, on *identical* labels and protocol—the

same 1,402 aligned daily observations, the same 30-day forward crisis windows, and the same trapezoidal AUROC as Table 8—a battery of simpler baselines: each of the four sub-indices on its own, the first principal component (PC1) of the four sub-indices, and an off-the-shelf Crypto Fear & Greed sentiment index (Alternative.me, 2022). Table 9 reports the result.

Table 9: Fair-Baseline Comparison: Day-Level Crisis Discrimination on Identical Labels

Score (same labels, same protocol)	AUROC [95% CI]	AUPRC
ASRI composite (4-channel)	0.866 [0.842, 0.888]	0.298
PC1 of the four sub-indices	0.858 [0.832, 0.881]	0.289
Contagion Risk channel *(best single)*	0.851 [0.824, 0.875]	0.285
Stablecoin Concentration Risk channel	0.825 [0.793, 0.854]	0.255
DeFi Liquidity Risk channel	0.821 [0.787, 0.852]	0.260
Crypto Fear & Greed (off-the-shelf)	0.789 [0.752, 0.825]	0.269
Diebold–Yilmaz connectedness	0.670 [0.636, 0.708]	0.121
Regulatory/Arbitrage Opacity channel	0.650 [0.611, 0.687]	0.141
Off-the-shelf baselines needing no crypto-native construction, and detrending checks:		
Contagion Risk, linear-detrended	0.912	0.369
VIX equity-volatility stress (standalone)	0.875	0.354
VIX + 10Y Treasury composite (Bank$_t$ macro core)	0.773	0.263
10Y Treasury level alone	0.501	0.080
Rolling 30-day BTC/ETH return (EW)	0.681	0.177

All scores evaluated on the same $n = 1{,}402$ aligned daily observations, the same four-event 30-day forward crisis labels (124 crisis-imminent days, prevalence 8.8%), and the same trapezoidal AUROC as Table 8. The sub-indices are the finest single features in the released daily data; raw underlying inputs (e.g. raw stablecoin-TVL drawdown) are not released separately.

PC1 is the leading principal component of the four sub-indices (covariance form; 74.9% of variance), and in this form its loading is dominated by the high-variance Contagion channel (≈ 0.81, versus 0.50/0.31/0.06 on Stablecoin/DeFi-Liquidity/Opacity)—so covariance-PC1 is close to a rescaled Contagion channel, which is why it tracks the best single channel so nearly. The off-the-shelf Crypto Fear & Greed index (Alternative.me, 2022) is fetched live and aligned to the common index (coverage 1,401/1,402 days).

The five rows below the second rule are added off-the-shelf baselines and detrending checks (point AUROC/AUPRC only; CIs of comparable width are omitted for space). The VIX standalone series ties ASRI (paired $\Delta = -0.009$, $p = 0.58$); the linear-detrended Contagion channel still exceeds the four-channel composite. "VIX + 10Y composite" is the Bank$_t$ macro core of the Contagion channel ($0.6\,\mathrm{norm}(r_{10Y}, 2, 6) + 0.4\,\mathrm{norm}(\mathrm{VIX}, 12, 40)$). The 10Y Treasury *level* discriminates at chance on these labels (0.501), so adding it at 60% weight drags the composite below VIX alone (0.773 vs. 0.875, a 0.102 reduction). This reflects the secular 2021–2024 tightening trend ($\approx$0.9 percentage points per year) that the level specification imports rather than an absence of banking-stress information: the reproducible detrended-Contagion channel (0.912) shows that removing such a trend recovers discrimination. A change- or detrended-based Treasury input is the preferable specification.

Confidence intervals are i.i.d.-day percentile bootstraps and *understate* sampling uncertainty; with only four independent crisis blocks the honest moving-block interval for ASRI alone spans $[0.756, 0.949]$ (Table 8), which contains every baseline in this table. On these *marginal* block-bootstrap intervals the $\approx$0.008–0.015 AUROC gaps among ASRI, PC1, and the Contagion channel are not separable. A *paired* block bootstrap (same resampled blocks applied to both series—the test used for the ASRI-versus-D-Y comparison) does retain small, consistently-signed AUROC edges whose CIs exclude zero (ASRI−PC1 +0.008, 95% CI $[+0.001, +0.018]$; ASRI−Contagion +0.016, 95% CI $[+0.004, +0.031]$; across block lengths $L = 20/25/30$, but the corresponding paired *AUPRC* differences include zero and—

Reading. Four facts follow, and together they reshape the empirical claim.

First, *aggregation buys no meaningful discrimination.* ASRI's AUROC (0.866) exceeds the first principal component of its own sub-indices (PC1, 0.858) by $\approx +0.008$ and the single strongest sub-index (Contagion Risk, 0.851) by $\approx +0.015$. *Marginally*, both PC1 and the Contagion channel fall *inside* ASRI's bootstrap confidence interval (i.i.d.-day $[0.842, 0.888]$; the honest moving-block interval is far wider, $[0.756, 0.949]$, Table 8), so on overlapping intervals the three are not separable. The proper contrast, however, is the *paired* moving-block bootstrap used for the D-Y comparison—both series scored on the same resampled blocks, so the strong day-to-day autocorrelation cancels in the difference. It returns a small but consistently-signed AUROC edge: ASRI−PC1 $= +0.008$ (95% CI $[+0.001, +0.018]$, $p \approx 0.02$) and ASRI−Contagion $= +0.016$ (95% CI $[+0.004, +0.031]$, $p \approx 0.003$), both excluding zero across block lengths $L = 20/25/30$, while the corresponding paired AUPRC differences include zero. We do not read this as a discriminative gain. Exactly as for the D-Y paired result (Appendix M), with only four independent crisis episodes (Section 6.3) a positive-in-every-resample sign reflects a *structural* relationship between nested constructions—ASRI is a weighted aggregate that *contains* the Contagion channel and is near-collinear with covariance-form PC1, which itself loads ≈ 0.81 on Contagion—rather than a powered demonstration that aggregation separates crises better. The edge is ≤ 0.016 AUROC, operationally negligible and absent in AUPRC; ASRI is therefore indistinguishable from its best single channel and from PC1 on the marginal intervals, and out-aggregates them only by a small structural margin that does not constitute a meaningful gain in crisis–non-crisis separation.

Second, *the strongest single channel for discrimination is Contagion Risk, not Stablecoin Risk.* Although Stablecoin Concentration Risk carries the highest aggregation weight (30%) and loads heavily in some weight-derivation specifications, on these labels it discriminates (0.825) below both the Contagion channel (0.851) and PC1. The "dominance" of the stablecoin channel is a statement about loading weight, not about predictive power; the most discriminative single channel is Contagion Risk.

Third, *the surviving "beats-benchmark" result is against a circular and unusually weak comparator.* The only baseline ASRI clearly out-discriminates is Diebold–Yilmaz connectedness (0.670)—which is itself *constructed from* ASRI's own four sub-indices, so the comparison is partly circular—and even there the benchmark is so weak that a free, off-the-shelf Fear & Greed sentiment index beats it by 0.12 (0.789 vs. 0.670). Out-discriminating D-Y is therefore not evidence that ASRI captures something a simpler construction misses.

Fourth, *a standalone equity-volatility series alone matches ASRI, and detrending does not rescue the composite.* Of the two baselines that need no crypto-native construction at all, a plain 30-day crypto return discriminates well below ASRI (0.681 vs. 0.866)—a modest point in the composite's favour, since its signal is more than raw price decline. A standalone VIX level, by contrast, reaches 0.875 and is statistically indistinguishable from ASRI (paired $\Delta = -0.009$, $p = 0.58$): to first order, labelling these four crises is a macro-volatility detection problem, and ASRI's discriminative value over an off-the-shelf equity-volatility series is not established. The aggregation demotion is not a trend artefact either—linear-detrending the Contagion channel raises it to 0.912, still above the four-channel composite, so removing a common trend leaves no distinctive contribution from aggregation.

We accordingly do not rest ASRI's contribution on discriminative superiority. The value of the four-channel composite is interpretive—it attributes stress to nameable transmission channels, carries operational lead-time for the detected crises, and exhibits coherent regime structure—rather than a measurable gain in crisis–non-crisis separation over its own best single channel or PC1.

5.8 Walk-Forward Validation

A distinct concern from publication lags is look-ahead bias in weight calibration: the theoretical weights were specified using domain knowledge accumulated from observing the full 2021–2024 sample, including the crises used for validation. To address this concern, we conduct walk-forward validation with expanding training windows that simulate the information set available prior to each crisis.

Calibrating the alert threshold on *only* pre-crisis data (the 90th percentile of each event's training-period history) flags all four episodes out-of-sample, with a mean first-crossing lead of ≈26 days. For the two early crises (Terra/Luna and Celsius/3AC), however, this 4/4 rate is bought at a 47–59% alarm rate over the full index history at ≈11–13% precision, so we read it as evidence against look-ahead bias rather than as operational early-warning skill. Continuous level-prediction is below a naive flat-mean baseline (walk-forward $R^2 \approx -0.6$; held-out $R^2 \approx -0.1$; 0/4 sub-indices robust to ±15% weight perturbation), confirming that ASRI is a threshold monitor rather than a point forecaster. The full walk-forward methodology, the per-event detection table (Table 34), and the per-event interpretation are given in Appendix O.

5.9 Out-of-Sample Specificity Test

A risk index must demonstrate not only sensitivity (detecting true crises) but also specificity (avoiding false alarms). We evaluate ASRI's specificity using 2024–2025 data—a period entirely out-of-sample relative to the framework's initial design and the crisis events used for validation.

Across 2024–2025—roughly 34 months out-of-sample—ASRI produced no sustained false alarms: a single multi-week elevation in July–August 2024 tracked the yen carry-trade de-risking shock (peak 58.8) without escalating, and the index otherwise stayed below the Elevated threshold (2025 peak 48.0). It correctly classified the February 2025 Bybit hack ($1.5B, the largest exchange theft in history) as non-systemic, because the theft produced no stablecoin depegs, protocol liquidations, or counterparty contagion—evidence that ASRI captures transmission mechanisms rather than event magnitude. The 2024 stability analysis, the Bybit component readings (Table 35), and the systemic-versus-contained comparison are given in Appendix P.

5.10 Validation Summary

The retrospective evaluation supports the ASRI framework on multiple dimensions:

1. **Retrospective Discrimination**: Three of four historical crises are detected via threshold-based monitoring (Celsius/3AC, FTX, SVB); the complementary event-study test of an elevated cumulative abnormal signal is, however, inconclusive—a placebo test on non-crisis dates shows the finite-sample (fixed-b) reference has essentially no crisis-specificity on this smooth trending series (roughly sixty percent of arbitrary dates clear the nominal 5% threshold, and the onsets sit at or below the placebo median), so we do not read the raw HAC t-statistics (1.7–3.9) as evidence of crisis detection. The Terra/Luna threshold

miss reflects documented limitations of market-based indicators for algorithmic stablecoin risk (Section 6.3).

2. **Lead Behaviour**: The in-sample 1.5σ signal is elevated across the full (30-day-capped) pre-crisis window for detected events; walk-forward thresholds calibrated only on pre-crisis data flag all four episodes (≈26-day mean first-crossing lead), but at a high false-positive cost for the early crises (Terra/Luna and Celsius/3AC alarm on 47–59% of index history), so this is evidence against look-ahead bias rather than a clean prediction record (Section 5.8)

3. **Statistical Validity**: ASRI and the DeFi Liquidity sub-index are stationary; Stablecoin and Opacity Risk are trend-stationary; Contagion Risk is non-stationary on the released sample (flagged caveat)

4. **Structural Stability**: the reproducible full-sample Chow test is a non-rejection of parameter consistency at the sample midpoint ($p = 0.78$)

5. **Regime Identification**: Three-regime HMM provides interpretable market state classification

6. **Component Importance**: Ablation analysis demonstrates stable 3/4 detection across all single-channel removals; the four-channel decomposition is retained for interpretability rather than as an individually load-bearing structure (Appendix I)

7. **Benchmark and Fair-Baseline Comparison**: On a common day-level sample ASRI out-discriminates Diebold-Yilmaz connectedness (AUROC 0.87 vs. 0.67; precision at the Youden-optimal threshold 35% vs. 15% on the common sample, a modest 20–32% in absolute terms on the full canonical sample), but D-Y is built from ASRI's own sub-indices and is itself out-discriminated by an off-the-shelf Fear & Greed index (0.789 vs. 0.670); a fair-baseline comparison on identical labels (Section 5.7) shows ASRI is statistically indistinguishable from its single strongest sub-index (Contagion Risk, 0.851) and from PC1 (0.858), with aggregation adding only ≈ +0.008–0.015 AUROC, so the value of aggregation is interpretive rather than discriminative (Appendix M)

8. **Robustness to Publication Lags**: Pseudo-real-time evaluation with publication lags confirms detection performance is not an artefact of look-ahead bias (Appendix N)

9. **Out-of-Sample Specificity**: No sustained false alarms during 2024–2025—a single non-systemic elevation in August 2024 (macro de-risking shock) and otherwise sub-threshold readings, including correct identification of the Bybit hack ($1.5B) as non-systemic due to absence of contagion channels (Section 5.9)

10. **Forecasting vs. Monitoring**: Continuous level-prediction metrics are below baseline (walk-forward $R^2 \approx -0.6$; held-out $R^2 \approx -0.1$; 0/4 robust components under weight perturbation), confirming that ASRI is a threshold monitor, not a point forecaster. Binary detection and threshold-based metrics are the appropriate evaluation framework (Section 5.8)

Data Quality Caveat. A component-level accounting of the backtested series (Table 11) shows that approximately **43%** of its average level (mean composition) is carried by static default values—Unreg_t (fixed at 35.0), Sent_t (50.0), Corr_t (0.5), audit-coverage levels, and the stablecoin peg input, which is held at par (price = 1.0) throughout the historical backtest (`backtest.py:191`)—with roughly 37% from dynamic crypto-native inputs and 20% from the macro (Treasury/VIX/yield-curve) proxy. Because these defaults are constant, they contribute essentially none of the series' time-variation; in particular, because the peg input is held at par, the stablecoin peg-deviation volatility term contributes no time-variation in the backtest, and the dynamic liquidity/stablecoin signal is carried by TVL drawdown and TVL volatility rather than by realised de-pegging. Crisis detection is robust to full-range variation of the placeholder parameters (Appendix J), but absolute ASRI levels therefore carry substantially more model dependence than the binary detection results suggest. The relative pattern—elevated readings before crises, moderate readings during calm periods—is more reliable than specific ASRI magnitudes. A related construction limitation concerns the TVL-drawdown input (Equation 2), which is normalised against a running (expanding) maximum. Because total DeFi TVL never reclaimed its late-2021 peak after mid-2022, this term is pinned at its ceiling for ≈85% of 2022–2024 and adds a near-constant ≈12-point offset to ASRI that, under a *fixed* detection threshold, mechanically lifts crossings. We re-specified the drawdown two ways on the same data. A 365-day *rolling* maximum leaves the 2022 episodes essentially unchanged (the late-2021 peak remains the trailing-window maximum and the drawdown still exceeds the 50% clip), so it does not address the issue for the crises of record. A responsive 30-day percentage-change specification removes the offset and preserves 3/4 fixed-threshold (≥ 50) detection (Celsius/3AC, FTX, SVB; Terra/Luna still missed), but compresses the average first-crossing lead from ≈19 days to ≈5 days (Celsius/3AC 12, FTX 3, SVB 1). The reported ≈19-day operational leads should therefore be read as partly an artefact of the running-maximum offset rather than as a genuine early-warning horizon.

6 Discussion

6.1 Theoretical Implications

The ASRI framework makes three contributions to the systemic risk literature:

First, it adapts the channel-decomposition perspective of the connectedness and contagion-measurement literature (Battiston et al., 2012) to the permissionless composability characteristic of DeFi, situating cryptocurrency systemic risk within the broader complexity economics research programme that applies agent-based, network, and dynamical systems approaches to 21st-century economic challenges (Bednar et al., 2025). We stress that this is an adaptation of measurement perspective, not a theoretical extension of network contagion models: the linear composite treats the system as a portfolio of separable risks rather than as a coupled dynamical network (see "On the Choice of Linear Aggregation" below). Traditional network models assume bilateral counterparty relationships with observable exposures; ASRI is designed with the multi-lateral, code-embedded exposures of smart-contract composability in view, though the current backtest proxies rather than directly measures these composability exposures (the inter-channel correlation input is held at a constant default).

Second, it *proxies* the DeFi-TradFi transmission channel: the Contagion sub-index is built

from TradFi-stress indicators (Treasury/VIX/yield-curve) rather than from a directly measured DeFi-TradFi exposure. As stablecoins have accumulated significant Treasury positions, they create a potential link between Federal Reserve monetary policy and DeFi liquidity conditions that existing crypto risk measures do not track—though ASRI captures this link through a macro proxy, not a realised interconnection measurement.

Third, it operationalises regulatory opacity as a quantifiable risk factor. While traditional finance assumes regulatory disclosure requirements, crypto markets feature substantial opacity about custody arrangements, reserve composition, and counterparty relationships. The Opacity sub-index provides a systematic framework for incorporating this uncertainty into risk assessment.

On the Choice of Linear Aggregation. A complexity-minded reader will rightly note the tension between ASRI's additive, linear aggregation and the systemic-risk phenomena it targets, which are paradigmatically *non-linear*: leverage spirals, redemption runs, and composability cascades involve positive feedback, threshold effects, and abrupt phase-transition-like reorganisations rather than the proportional, separable response a weighted sum encodes (Battiston et al., 2012; Bednar et al., 2025). We make the trade-off explicit. Linear aggregation is chosen for *interpretability and estimability*, not because we believe systemic stress accumulates additively: the weighted-sum form is the standard backbone of operational composite stress indicators precisely because it yields an exact, efficiency-consistent channel attribution (Axiom 3.3) and requires no parameters beyond the four weights—a decisive advantage when, as here, only four crisis episodes are available to fit anything. The cost is realism. A linear index cannot represent the super-additive amplification that arises when several channels are stressed at once, nor the self-reinforcing feedback by which one channel's deterioration accelerates another's; it treats the system as a portfolio of separable risks rather than as a coupled dynamical network. Richer aggregators address parts of this gap: the CISS framework folds cross-channel correlation into a portfolio-theoretic combination so that joint stress is penalised more than the sum of its parts (Hollo et al., 2012); network and graph-regularised constructions embed topology directly (Briola et al., 2026; Battiston et al., 2012); and the CES and geometric aggregators we compare in Section L can encode complementarity and tail amplification through their substitution parameter. Each, however, introduces dependence parameters that our four-event sample cannot identify reliably (Section 6.3). Our position is therefore deliberately modest: ASRI is an interpretable first-order *monitoring* decomposition, not a model of the non-linear crisis dynamics themselves, and the move to amplification-aware aggregation is the principal methodological extension we flag for when the crisis sample is large enough to estimate it. The empirical finding that aggregation buys little discriminative power over the single strongest channel on this sample (Section 5.7) is consistent with this reading: the value of the linear composite here is attribution and auditability, not a claim to have captured the system's non-linear structure.

6.2 Practical Applications

We frame the following as illustrative use cases, conditional on the binding four-event limitation (Section 6.3), rather than as validated deployments.

Portfolio Risk Management: ASRI could in principle inform institutional crypto allocators as one input to a risk overlay—trimming exposure when systemic readings are elevated—though we stress this is illustrative, given that all results rest on only four crisis events.

Regulatory Monitoring: ASRI is intended as a candidate input for macroprudential surveillance of DeFi-TradFi interconnection dynamics, not as a validated supervisory tool.

Protocol Governance: DeFi protocols could in principle use sub-index components to gauge their contribution to systemic risk and inform parameter choices (e.g., collateral ratios, withdrawal limits).

Research Applications: The released ASRI series and code provide a reproducible, openly documented reference implementation for further study of crypto market dynamics.

6.3 Limitations

Four Crisis Events (Binding Power Limitation): The single most important caveat on every result in this paper is that the validation set contains only *four* systemic crisis events (Terra/Luna, Celsius/3AC, FTX, SVB). All detection rates, lead times, precision and recall figures, the regime-count robustness comparison, and the ASRI-versus-Diebold-Yilmaz head-to-head ultimately rest on these four events. The day-level classification analysis (1,402 daily observations) and the bootstrap confidence intervals partially mitigate this by exploiting the autocorrelation structure within each crisis window, but the underlying *independent* information content is four events: the day-level labels are generated by, and are not independent of, those four onsets. Consequently, statements such as "ASRI attains AUROC 0.87 versus 0.67" should be read as indicative of relative performance on this small, historically specific sample rather than as estimates with the statistical power one would obtain from a large, independent event population. With four events, detection-rate comparisons (e.g. 3/4 versus 4/4) cannot discriminate reliably among model specifications, and we deliberately avoid over-interpreting them. This is the binding limitation for any inferential venue and the principal reason we frame ASRI as an interpretable monitoring framework rather than a statistically validated forecaster. Expanding the event set as additional crises accrue is the highest-priority direction for future validation.

Terra/Luna Detection Failure: ASRI's most significant *event-specific* empirical limitation is its failure to detect the Terra/Luna collapse (May 2022) using threshold-based early warning. The index peaked at 46.0 in the 30-day pre-crisis window—below the "Elevated" threshold of 50—despite this event representing the largest nominal loss (\$40B+) in cryptocurrency history. (The peak of 48.7 reported in Table 5 is taken over the wider $[-30, +10]$ event window, which includes the immediate post-onset days; the 30-day *pre*-crisis peak is 46.0.) The event study methodology (Section 5.4) does register an elevated raw cumulative abnormal signal for Terra/Luna (CAS = 100.3, naive $t = 5.47$; Table 5) because it measures deviation from the estimation baseline rather than absolute threshold breaches; but this signal does not survive autocorrelation-robust inference (HAC $t = 1.72$, non-significant), consistent with our treatment of the event study as inconclusive throughout. Terra/Luna is thus missed both by the operational threshold rule and by the robust event study. However, for operational early warning, ASRI failed to provide actionable alerts before the Terra/Luna crisis.

The failure reflects a fundamental limitation of retrospective indicators: algorithmic sta-

blecoin risk prior to UST's depeg was difficult to observe through market prices alone. UST maintained its peg until the death spiral began, and Luna's price appreciation masked the underlying fragility. The sub-indices capture *revealed* stress through observable metrics (TVL drawdowns, yield compression, realised volatility), but Terra/Luna's collapse was precipitated by endogenous reflexivity between UST redemptions and Luna minting rather than exogenous liquidity or contagion shocks that ASRI's components measure.

Remediation: Section A.1.6 specifies an algorithmic stablecoin risk extension that incorporates backing ratio dynamics, collateral volatility, and supply dilution—metrics *designed* to flag Luna's instability prior to UST's depeg. The following is an illustrative back-of-envelope calculation, not a validated result: a sensitivity calculation suggests this extension might have elevated SCR by an indicative 8–12 points in April 2022, potentially bringing ASRI above the detection threshold. We have not been able to confirm this—full historical validation is deferred to an on-chain reconstruction of UST/Luna supply and backing metrics from Terra Classic archive nodes or indexer snapshots (Appendix A.1.6), an engineering task rather than a data-availability barrier—so the figure should be read as a design target rather than a demonstrated detection.

Untested Failure Modes (Generalisation Beyond Liquidity Cascades): All four calibration events—Terra/Luna, Celsius/3AC, FTX, and SVB/USDC—belong to a single broad failure family: liquidity and solvency cascades originating at centralised exchanges, lending desks, or stablecoin issuers and propagating through funding and redemption channels. ASRI's sub-indices, weights, and thresholds are, accordingly, tuned to *that* family. Several systemically relevant failure modes are absent from the 2022–2023 sample and are not captured by the current construction: (i) *Layer-1 consensus failures*—a chain halt, finality failure, or successful 51% attack—which would manifest in on-chain block production and validator participation rather than in TVL, issuer concentration, or macro stress; (ii) *large-scale oracle manipulation*, in which a compromised price feed triggers mass liquidations with no prior deterioration in the liquidity or contagion channels ASRI monitors; (iii) *systemic smart-contract or bridge exploits*, whose onset is discontinuous and would not register as a gradual pre-crisis build-up (the framework's correct classification of the contained Bybit theft as non-systemic, Section 5.9, is the same property that would make it under-react to a fast exploit that then cascades); and (iv) *governance or regulatory shocks* that freeze a major protocol or stablecoin administratively rather than through market stress. Because the weights and the 50 threshold were specified against liquidity-cascade dynamics, we make no claim that they transfer to these mechanisms: the parameters should be re-examined—and the sub-index set very likely extended (for example with on-chain consensus-health and oracle-deviation channels)—before ASRI is applied outside the exchange/lending/stablecoin-liquidity domain on which it was built. This generalisation limit is distinct from the four-event power ceiling: even with many more events of the *same* type, the index would remain silent on failure modes its channels do not observe.

Data Availability: Several components (Tier 2 sources) require manual collection or have limited historical depth. Enterprise APIs (TRM Labs, Chainalysis) that would improve coverage are cost-prohibitive for academic research.

Historical Reconstruction: DeFi data prior to 2021 is sparse, requiring proxy indicators and interpolation that reduce confidence in early-period ASRI values.

Continuous Prediction Failure: While ASRI's binary crisis–non-crisis discrimination on

these labels (AUROC = 0.866; not a discriminative gain over simpler baselines—Section 5.7) and walk-forward detection (4/4) are its strongest empirical properties relative to point forecasting, continuous-valued level-prediction is below a naive baseline: walk-forward $R^2 \approx -0.6$ and held-out $R^2 \approx -0.1$ for predicting the 30-day-ahead ASRI level. Additionally, weight perturbation analysis finds 0/4 sub-index weights are robust to $\pm 15\%$ variation in terms of continuous level prediction. These metrics confirm that ASRI should be interpreted as a threshold-based monitoring tool rather than a continuous risk forecasting model. Absolute ASRI levels are driven partly by proxy components with fixed default values, introducing measurement noise that degrades point prediction accuracy while preserving relative crisis-vs-calm discrimination.

Model Specification Uncertainty: Weight allocation reflects theoretical judgement rather than empirical optimisation. Sensitivity analysis (Appendix J) demonstrates stability under weight perturbations; Section 5.5 compares theoretical weights against data-driven alternatives.

Regulatory Dynamics: The rapidly evolving regulatory landscape may require periodic recalibration of the Opacity sub-index as disclosure requirements change.

Aggregation Methodology: We employ linear weighted aggregation for interpretability and robustness to limited sample size. Alternative approaches explored in the literature—including CISS-style portfolio-theoretic aggregation (Hollo et al., 2012), copula-based tail dependence modelling, regime-switching weights, and graph-regularised PCA methods that incorporate network topology into dimensionality reduction (Briola et al., 2026)—are theoretically appealing but require substantially more data for reliable estimation. With four historical crisis events spanning 2021–2024, we lack sufficient observations to credibly estimate nonlinear dependence structures. Sensitivity analysis (Appendix J) demonstrates that our results are robust to weight perturbations; as the sample grows, more sophisticated aggregation methods may become appropriate.

Connectedness Benchmark Scope: We compute Diebold-Yilmaz connectedness on ASRI's own four constructed sub-indices rather than on a panel of assets or institutions, which is the canonical D-Y application. This is a deliberate but non-standard choice: it measures spillovers *among our risk channels* rather than *among markets*, and partly explains why the head-to-head is a comparison of two index constructions rather than a comparison of ASRI against an asset-level connectedness benchmark. An asset-level D-Y implementation would be a more conventional benchmark and is a natural extension. Our comparative analysis benchmarks against Diebold-Yilmaz connectedness, the most widely deployed FEVD-based systemic risk measure. Recent extensions—including asymmetric connectedness (Hatemi-J, 2012) that distinguishes positive from negative shock transmission, and TVP-VAR implementations (Antonakakis et al., 2020) that eliminate rolling-window dependence—may capture regime-specific dynamics that symmetric, fixed-parameter specifications miss. Future work should extend the benchmark comparison to these asymmetric and time-varying alternatives, particularly for tail-risk episodes where directional spillover asymmetries are most pronounced.

Simple Baselines (now computed): An earlier draft flagged as future work whether a single feature, the first principal component, or an off-the-shelf index would match ASRI's AUROC on these labels. We have since computed this battery on identical labels (Section 5.7, Table 9), and the finding tempers the empirical claim. ASRI's discrimination (0.866) is, on the marginal block-bootstrap intervals, indistinguishable from PC1 of its sub-indices (0.858) and from its

single strongest sub-index (Contagion Risk, 0.851); aggregation adds only ≈ +0.008–0.015 AUROC, well inside the four-event sampling uncertainty, and the paired bootstrap separates them only by this small structural margin (ASRI nests both baselines), not a meaningful discriminative gain. ASRI clearly out-discriminates only the Diebold-Yilmaz connectedness series (0.670), which is constructed from its own sub-indices and is itself beaten by an off-the-shelf Fear & Greed index (0.789). That roughly 43% of the series is constant and the ablation in Appendix I removes any single channel without detection loss is consistent with this picture. We therefore do not rest the paper's contribution on discriminative superiority over simpler constructions: the value of the four-channel composite is interpretive (channel attribution, lead-time, and regime structure in a single auditable series), and the highest-priority empirical extension is to widen the crisis-event set so that these sub-0.015 AUROC gaps can be tested with power.

Look-Ahead Bias: The theoretical weights are based on domain knowledge accumulated from observing the full 2021–2024 sample, which includes the crisis events used for validation. Walk-forward validation (Section 5.8) speaks to this concern: using only pre-crisis data to calibrate standardisation parameters, the walk-forward thresholds flag all four episodes (a 4/4 out-of-sample rate, matching the in-sample rate). This rate should not be read as predictive strength. For the two early crises (Terra/Luna and Celsius/3AC) the 4/4 flag is bought at a 47–59% alarm rate over the full index history at ≈11–13% precision, so the appropriate reading is that the procedure exposes a high false-positive cost, and the result is evidence *against* look-ahead bias—the in-sample detections are not an artefact of using future data to set parameters—rather than a clean out-of-sample prediction record. Out-of-sample lead times are also shorter, particularly for novel crisis types like Terra/Luna and FTX. Future work should extend this analysis by re-deriving data-driven weights using only pre-crisis data to fully characterise out-of-sample performance under empirically-optimised specifications.

Procyclicality Considerations: Public dissemination of ASRI values raises legitimate concerns about procyclical feedback. If market participants interpret elevated readings as coordination signals for deleveraging, publication could theoretically accelerate the stress it aims to detect. We propose several mitigations: (1) publishing methodology and historical values transparently while throttling real-time granular scores during acute stress periods; (2) framing ASRI explicitly as a vigilance indicator rather than a trading signal, distinguishing monitoring from market-timing; (3) providing detailed component breakdowns to regulators on a privileged basis to support macroprudential oversight without amplifying market coordination; and (4) incorporating the Goodhart critique into index maintenance—if market participants begin gaming specific sub-indices, the weighting scheme can evolve.

Bull Market Context for Out-of-Sample Period: The 2024–2025 out-of-sample period coincides with a cryptocurrency bull market characterised by rising prices and expanding TVL. During this period, ASRI registered only one multi-week elevation (August 2024, a non-systemic macro de-risking shock) and otherwise remained below the Elevated threshold, with no sustained false alarms—a partial specificity success. However, this represents a benign stress test: true validation of specificity during market euphoria and sensitivity during subsequent corrections requires observing ASRI behaviour through a complete market cycle. The framework's performance during a future bear market or crisis originating in 2025+ will provide more meaningful out-of-sample validation than the current bull market period.

Macro Confounding—Crypto-Specific Fragility vs. Exogenous TradFi Stress: A limitation that cuts across the entire evaluation is that our crisis sample cannot cleanly separate *endogenous* crypto-systemic fragility from *exogenous* traditional-finance stress. Three features make this confounding acute. First, the Contagion channel is by construction a macro-stress proxy: ≈75% of its time-variation is attributable to the Treasury/VIX/yield-curve composite (Table 11) rather than to a directly measured DeFi–TradFi interconnection. Second, the 2022 crises coincide with the most aggressive monetary-tightening cycle in four decades, so crypto-native deleveraging and exogenous macro risk-off move together over precisely the window our labels cover. Third, and most directly, a standalone equity-volatility series (VIX) matches ASRI's day-level discrimination on these labels (AUROC 0.875 vs. 0.866, statistically indistinguishable; Section 5.7), so to first order the crises are detectable as a macro-volatility event without any crypto-native construction at all. We therefore cannot claim that ASRI's discrimination isolates a crypto-specific signal: the evidence is consistent both with crypto-native vulnerabilities and with a common macro factor driving crypto stress and the index in tandem. Disentangling the two—for example by orthogonalising the crypto-native channels against a macro-factor benchmark, or by testing the index on a crypto-idiosyncratic crisis uncorrelated with TradFi stress—is unresolved and is a priority for future work. Until it is, the paper's references to "crypto-native vulnerabilities" should be read as a statement about the index's *design targets* (stablecoin concentration, liquidity drawdown, opacity), not as a demonstration that its measured crisis signal is separable from exogenous macro stress.

Single Macro Regime (2022–2023 Tightening Cycle): A distinct generalisation limit concerns the monetary backdrop of the calibration set. All four events fall within the 2022–2023 tightening cycle, during which rising Treasury yields, elevated VIX, and a widening term spread moved together and coincided with crypto deleveraging. ASRI's Contagion channel, weights, and 50 threshold are therefore fitted against a rising-rate, risk-off environment. How the index behaves in a low-rate or quantitative-easing regime is untested and cannot be inferred from this sample: in an easing environment the macro proxies would sit low even if crypto-native fragility (stablecoin concentration, liquidity drawdown, leverage) were building, so the composite could under-weight genuinely endogenous stress, while a locally stressed but benign-macro episode could be missed by a threshold tuned to co-moving TradFi stress. Characterising ASRI across distinct macro regimes—ideally including a crypto stress episode that occurs under easy monetary conditions—is required before the calibrated weights and threshold can be treated as regime-invariant, and we make no such claim here.

Treasury Maturity Mismatch (10-Year Yield as Reserve-Stress Proxy): The implementation's Treasury *level* input is the 10-year constant-maturity yield (FRED DGS10), which enters both the Stablecoin Risk sub-index (as its treasury-stress term) and the Contagion channel's banking-stress core, yet stablecoin issuers hold predominantly short-duration paper—1–3 month Treasury bills and overnight reverse repurchase agreements—so a 10-year yield embeds duration and term-premium dynamics that issuer reserve portfolios do not face. A 3- or 6-month bill yield would be the theoretically better-matched reserve-stress proxy, and this respecification plausibly interacts with the level-versus-detrended specification issue documented in Section 5.7. We flag the short-maturity respecification as a spec-change candidate for future work; we have not evaluated it on the present sample.

Concentration Proxy versus Topological Composability: The introduction motivates DeFi systemic risk partly through *composability*—the dependency structure created when protocols permissionlessly call one another—yet the implementation operationalises the structural channels through *concentration* measures: a top-10 protocol-TVL concentration term in DeFi Liquidity Risk and an issuer Herfindahl–Hirschman index in Stablecoin Risk. These are size/concentration proxies: they capture how much value is held in few hands, not the topology of inter-protocol dependencies through which a failure actually propagates. A highly concentrated ecosystem need not be highly composable, and vice versa, so the current construction proxies the *level* of concentration rather than the *network structure* of composability the motivation invokes. A direct composability measure would require protocol-interaction data (call-graph or shared-collateral linkages) that we do not incorporate; we flag this gap between the theoretical motivation and the implemented proxy, and a direct topological measure, as a priority extension.

7 Conclusion

This paper introduced the Aggregated Systemic Risk Index (ASRI), an interpretable, channel-decomposed composite for the *retrospective* characterisation of crypto-native systemic stress. The index comprises four weighted sub-indices—Stablecoin Risk, DeFi Liquidity Risk, Contagion Risk, and Regulatory Opacity Risk—built around crypto-native channels, with the contagion channel implemented as a TradFi-stress proxy (Treasury/VIX/yield-curve) rather than a directly measured DeFi-TradFi interconnection. We present it as a retrospective monitoring and methodological contribution, not a validated real-time early-warning system. On the empirical side, a fair-baseline comparison establishes what the composite does and does not buy: ASRI's crisis discrimination is, on the marginal block-bootstrap intervals, indistinguishable from that of its single strongest sub-index (Contagion Risk) and from the first principal component of its sub-indices, with aggregation adding only $\approx$ +0.008–0.015 AUROC—a paired bootstrap separates them by no more than this small, structural margin (ASRI nests both baselines), not a meaningful discriminative gain. The contribution is therefore the interpretable, channel-decomposed, auditable packaging of crypto-native stress—not superior discrimination over simpler constructions. Framed in network terms, what the composite offers is an interpretable connectedness-and-contagion decomposition for crypto-native systemic stress—a transmission-channel structure situated in the financial-network-science tradition—rather than a sharper discriminator than its single strongest channel.

Key Empirical Findings:

1. **Retrospective Discrimination**: ASRI detected three of four historical crises (Celsius/3AC, FTX, SVB) via threshold-based monitoring; the complementary event-study test of an elevated cumulative abnormal signal is inconclusive, since a placebo test on non-crisis dates shows the finite-sample (fixed-b) reference has essentially no crisis-specificity on this smooth trending series (the onsets sit at or below the placebo median), so the raw HAC t-statistics (1.7–3.9) are not read as crisis detection. The Terra/Luna collapse represents a documented limitation (Section 6.3).

2. **Lead Behaviour**: The in-sample 1.5σ signal is elevated across the full (30-day-capped)

pre-crisis window for detected events; walk-forward thresholds calibrated only on pre-crisis data flag all four episodes (≈26-day mean first-crossing lead), but at a high false-positive cost for the two early crises (47–59% alarm rate), so this evidences the absence of look-ahead bias rather than operational early-warning skill

3. **Regime Dynamics**: Three-regime HMM identifies Low Risk (19.8%), Moderate (56.3%), and Crisis (23.8%) states with high persistence (>97%) based on smoothed posterior assignments; the ergodic distribution $[\approx 0, 0.71, 0.29]$ indicates that long-run occupancy is dominated by the Moderate regime, with the Crisis regime occupied ≈29% of the time

4. **Structural Stability**: the reproducible full-sample Chow test is a non-rejection of consistent model parameters at the sample midpoint ($p = 0.78$)

5. **Component Importance**: Ablation analysis shows detection is invariant at 3/4 to removing any single channel; the four-channel split is retained for interpretability, not as an individually load-bearing decomposition

6. **Benchmark and Fair-Baseline Comparison**: On a common day-level sample ASRI out-discriminates Diebold-Yilmaz connectedness (AUROC 0.87 vs. 0.67; precision 35% vs. 15% at the Youden-optimal threshold on the common sample (a modest 20–32% on the full canonical sample)), but D-Y is constructed from ASRI's own sub-indices and is itself out-discriminated by an off-the-shelf sentiment index (0.789 vs. 0.670). On identical labels ASRI is, on the marginal block-bootstrap intervals, indistinguishable from its single strongest sub-index (Contagion Risk, 0.851) or from PC1 of its sub-indices (0.858)—aggregation adds only ≈ +0.008–0.015 AUROC, and a paired bootstrap separates them only by this small structural margin (Section 5.7); a standalone VIX series also matches it (0.875 vs. 0.866, indistinguishable) and detrending leaves no distinctive contribution from aggregation, so to first order the crisis labelling is a macro-volatility detection problem—and the composite's value is interpretive, not a gain in discrimination

7. **Robustness to Publication Lags**: Pseudo-real-time backtesting with publication lags confirms detection performance (<1% ASRI degradation, lead times preserved)

8. **Walk-Forward Robustness**: All four events flagged out-of-sample using only pre-crisis calibration data, evidencing absence of look-ahead bias (with the false-positive cost noted above)

9. **Out-of-Sample Specificity**: No sustained false alarms during 2024–2025—a single non-systemic elevation in August 2024 (a macro de-risking shock that did not escalate) and otherwise sub-threshold readings, including correct identification of the Bybit hack ($1.5B, largest exchange theft in history) as non-systemic—the framework captures transmission mechanisms, not merely event magnitude

Contribution: The ASRI framework addresses three critical gaps in existing risk monitoring:

- Targets *composability risk* from DeFi protocol interactions (a design aim; the current backtest proxies rather than directly measures protocol-level composability)

- Proxies *stablecoin-Treasury linkages* as a transmission channel via TradFi-stress indicators, rather than measuring a realised exposure
- Operationalises *regulatory opacity* as a quantifiable risk factor

Future Research: Version 3.0 will extend the framework with:

1. **Composability-Aware Risk Metrics**: The current DeFi Liquidity Risk sub-index treats protocols as independent units. A natural extension incorporates protocol composability structure: dependency graphs where Protocol A's risk increases if Protocol B (which A calls) becomes distressed. Implementation requires: (i) protocol call-graph extraction from on-chain traces, (ii) network centrality scoring (PageRank, betweenness), and (iii) shock propagation simulation along dependency edges. This captures second-order cascade dynamics beyond first-order concentration risk.

2. **Regulatory Sentiment Pipeline**: Full implementation of Sent_t via FinBERT fine-tuned on SEC/ESRB/FSB announcements, with jurisdictional weighting (US 40%, EU 30%, UK 15%, Other 15%) and entity resolution for de-duplication.

3. **Proxy Validation Against Ground Truth**: Systematic validation of Bank_t, Link_t, and other proxy components against quarterly OCC/ECB regulatory filings when released. The proxy validation framework (Appendix) provides the methodology.

4. **Non-Linear Aggregation**: CES aggregation with $\rho < 0$ to capture complementary risk dynamics where multiple elevated sub-indices amplify aggregate stress. Initial analysis (Table 29) suggests max-based aggregation achieves 4/4 detection with 29-day lead times.

5. **Tail Risk Integration**: VCoVaR-based tail dependence measures for the Contagion sub-index, capturing asymmetric spillovers that simple correlation misses during stress periods.

6. **On-Chain Failure-Mode Channels**: Direct sub-indices targeting the failure modes the current liquidity-cascade construction does not observe (the "Untested Failure Modes" gap, Section 6.3). A *consensus-health* channel would track validator participation, missed-block and finality-failure rates, and stake concentration (Nakamoto coefficient) to register Layer-1 chain halts and 51%-attack risk; an *oracle-deviation* channel would monitor divergence between on-chain price feeds and reference venues, feed staleness, and update latency to flag oracle-manipulation conditions before they trigger mass liquidations. Both draw on already-public on-chain data (beacon-chain attestation records, oracle-contract event logs), and would extend ASRI's coverage beyond the exchange/lending/stablecoin-liquidity family on which it was calibrated.

As DeFi grows and its interconnection with traditional finance deepens, transparent, channel-decomposed monitoring of crypto-native stress is a useful complement to the established systemic-risk measures for market participants, regulators, and researchers.

Live Dashboard: asri.dissensus.ai/

Code Repository: github.com/studiofarzulla/asri

Data and Code Availability

Live Dashboard: Real-time ASRI values and historical time series are publicly available at asri.dissensus.ai.

API Access: RESTful API endpoints serve current and historical ASRI values with component decomposition. Full endpoint documentation is provided in the Appendix (API Documentation Summary).

Source Code: Complete implementation including data ingestion, signal computation, and publication pipelines is available at github.com/studiofarzulla/asri under MIT license.

Replication Data: The daily ASRI series (2021–2025, 1,841 days) with component breakdowns is included in the repository (`results/data/asri_history.parquet`) and archived on Zenodo (10.5281/zenodo.17918239). This series is the *frozen dataset of record*: all reported results reproduce deterministically from it via the provided analysis scripts (e.g. `event_study_hac.py`, `baseline_comparison.py`, `moving_block_bootstrap_roc.py`, `generate_detection_table.py`, `extract_hmm_diagnostics.py`), as documented in `REPRODUCIBILITY.md` (which maps each script to its table/number) and `DATA_PROVENANCE.md` (which hash-freezes the series). The generation pipeline (live DeFi Llama/FRED ingestion) is provided in full; given a frozen universe snapshot (`scripts/dump_universe_snapshot.py`) it runs deterministically to produce a code-consistent series. We note honestly that the published daily series is *not* bit-reproducible from the live ingestion pipeline—its original point-in-time API inputs and protocols/bridges universe were not retained—so the archived frozen series, rather than a re-pull, is the reproducibility artefact.

Pseudo-Real-Time Replication: The repository includes a lag-simulation module that reproduces the publication lag methodology described in Appendix N, enabling independent verification of real-time detection claims.

Software Environment: All analyses were conducted in Python 3.11 using pandas 2.0+, statsmodels 0.14+, scipy 1.11+, and scikit-learn 1.3+. HMM estimation uses hmmlearn 0.3.0. Random seeds are set to 42 for all stochastic procedures (bootstrap, cross-validation splits). Complete environment specification is provided via `requirements.txt` in the repository.

Declarations

Conflict of Interest. The authors declare no competing interests.

Funding. Supported by King's College London resources. No external funding was received.

Data Availability. DeFi Llama, FRED, and CoinGecko data are publicly available. ASRI index values and processed datasets available at https://doi.org/10.5281/zenodo.17918238. Live dashboard: https://asri.dissensus.ai.

Code Availability. Open-source implementation available at https://github.com/studiofarzulla/asri under MIT License.

AI Assistance. Claude (Anthropic) was used as a research collaborator for analytical framework development, literature synthesis, and technical writing. Perplexity AI was used for re-

search tool development that enabled efficient literature discovery. All intellectual claims and errors remain the authors' responsibility.

Author Contributions. AM: Conceptualisation, writing—original draft. MF: Methodology, software, formal analysis, data curation, validation, writing—review & editing, visualisation.

Acknowledgements

The authors acknowledge collaboration with Aron Farzulla on data pipeline architecture and API integration strategy. The authors thank the DeFi Llama team for providing comprehensive protocol coverage data, and the Federal Reserve Bank of St. Louis for maintaining the FRED API as a public good.

This paper benefited from extended collaboration with Claude (Anthropic), whose contributions to analytical framework development and technical writing were substantive. The authors gratefully acknowledge this assistance while taking full responsibility for all claims, errors, and interpretive choices.

This work is part of the Adversarial Systems Research program at Dissensus, a broader investigation into stability, alignment, and friction dynamics across political, financial, cognitive, and multi-agent systems. Related papers in the series are available through the Adversarial Systems & Complexity Research Initiative (ASCRI; systems.ac).

The authors welcome feedback, criticism, and collaboration. Correspondence should be directed to murad@dissensus.ai.

References

Viral V. Acharya, Lasse H. Pedersen, Thomas Philippon, and Matthew Richardson. Measuring systemic risk. *Review of Financial Studies*, 30(1):2–47, 2017. doi: 10.1093/rfs/hhw088.

Tobias Adrian and Markus K. Brunnermeier. Covar. *American Economic Review*, 106(7): 1705–1741, 2016. doi: 10.1257/aer.20120555.

Alternative.me. Crypto fear & greed index, 2022. URL https://alternative.me/crypto/fear-and-greed-index/. Accessed: December 2025.

Nikolaos Antonakakis, Ioannis Chatziantoniou, and David Gabauer. Refined measures of dynamic connectedness based on time-varying parameter vector autoregressions. *Journal of Risk and Financial Management*, 13(4):84, 2020. doi: 10.3390/jrfm13040084. Foundational TVP-VAR framework for time-varying connectedness analysis without rolling windows.

Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999. doi: 10.1111/1467-9965.00068.

Tomaso Aste. Cryptocurrency market structure: Connecting emotions and economics. *Digital Finance*, 1:5–21, 2019. doi: 10.1007/s42521-019-00008-9.

Tomaso Aste. Information filtering networks: Theoretical foundations, generative methodologies, and real-world applications. arXiv preprint, 2025.

Sabrina Aufiero, Silvia Bartolucci, Fabio Caccioli, and Pierpaolo Vivo. Mapping microscopic and systemic risks in TradFi and DeFi: A literature review. 2025. Preprint.

Massimo Bartoletti, James Hsin-yu Chiang, and Alberto Lluch-Lafuente. Maximizing extractable value from automated market makers. In *Financial Cryptography and Data Security*, 2022. doi: 10.1007/978-3-031-18283-9_1. Formalizes DeFi composability as a source of MEV extraction and systemic risk.

Stefano Battiston, Michelangelo Puliga, Rahul Kaushik, Paolo Tasca, and Guido Caldarelli. Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Scientific Reports*, 2:541, 2012. doi: 10.1038/srep00541.

Jenna Bednar, R. Maria del Rio-Chanona, J. Doyne Farmer, Jagoda Kaszowska-Mojsa, François Lafond, Penny Mealy, Marco Pangallo, and Anton Pichler. Complex systems approaches to 21st century challenges: Introduction to the special issue. *Journal of Economic Behavior & Organization*, 235:107049, 2025. doi: 10.1016/j.jebo.2025.107049.

Monica Billio, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559, 2012. doi: 10.1016/j.jfineco.2011.12.010.

Claudio Borio and Mathias Drehmann. Assessing the risk of banking crises – revisited. *BIS Quarterly Review*, pages 29–46, March 2009. URL https://www.bis.org/publ/qtrpdf/r_qt0903e.htm.

Antonio Briola, Marwin Schmidt, Fabio Caccioli, Carlos Ros Perez, James Singleton, Christian Michler, and Tomaso Aste. Graph regularized PCA. arXiv preprint, 2026.

Christian Brownlees and Robert F. Engle. Srisk: A conditional capital shortfall measure of systemic risk. *Review of Financial Studies*, 30(1):48–79, 2017. doi: 10.1093/rfs/hhw060.

Markus K. Brunnermeier. Deciphering the liquidity and credit crunch 2007–2008. *Journal of Economic Perspectives*, 23(1):77–100, 2009. doi: 10.1257/jep.23.1.77.

Gregory C. Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):591–605, 1960. doi: 10.2307/1910133.

Michael Darlin, Georgios Palaiokrassas, and Leandros Tassiulas. Debt-financed collateral and stability risks in the defi ecosystem. In *IEEE International Conference on Blockchain and Cryptocurrency (BRAINS)*, 2022. doi: 10.1109/BRAINS55737.2022.9909090. Analyzes recursive leverage and collateral chain risks in DeFi lending.

Riccardo De Blasis, Luca Galati, Alexander Webb, and Robert I. Webb. Intelligent design: Stablecoins (in)stability and collateral during market turbulence. *Financial Innovation*, 9: 85, 2023. doi: 10.1186/s40854-023-00492-4. Examines stablecoin (in)stability and collateral behavior during turbulent market conditions.

DeFi Llama. Defi tvl and protocol data, 2025. URL https://defillama.com/. Accessed: December 2025.

David A. Dickey and Wayne A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431, 1979. doi: 10.1080/01621459.1979.10482531.

Francis X. Diebold and Kamil Yılmaz. Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1):57–66, 2012. doi: 10.1016/j.ijforecast.2011.02.006. Introduces the total connectedness index for measuring systemic spillovers via variance decomposition.

Francis X. Diebold and Kamil Yılmaz. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134, 2014. doi: 10.1016/j.jeconom.2014.04.012.

Papa Ousseynou Diop, Julien Chevallier, and Bilel Sanhaji. Collapse of silicon valley bank and usdc depegging: A machine learning experiment. *FinTech*, 3(4):569–590, 2024. doi: 10.3390/fintech3040030. Machine learning analysis of SVB collapse and USDC depeg dynamics.

Barry Eichengreen, Ngan Nguyen, and Ganesh Viswanath-Natraj. Stablecoin devaluation risk. *European Journal of Finance*, 31(11), 2025. doi: 10.1080/1351847X.2025.2505757. Empirical analysis of factors driving stablecoin devaluation risk premiums. Also available as SSRN 4460515.

Arturo Estrella and Gikas A. Hardouvelis. The term structure as a predictor of real economic activity. *Journal of Finance*, 46(2):555–576, 1991. doi: 10.1111/j.1540-6261.1991.tb02674.x. Establishes the yield-curve slope as a predictor of future real economic activity and recessions.

Arturo Estrella and Frederic S. Mishkin. Predicting u.s. recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, 80(1):45–61, 1998. doi: 10.1162/003465398557320. Yield-curve spread among the most accurate financial leading indicators of U.S. recessions.

Murad Farzulla. Do cryptocurrency markets differentiate infrastructure from regulatory shocks? a multi-moment event study with dependence-robust inference. *Research Square*, 2025. doi: 10.21203/rs.3.rs-8323026/v1. Under review at Digital Finance (Springer). Multi-moment event study of crypto volatility with dependence-robust inference.

Murad Farzulla. Do whitepaper claims predict market behavior? evidence from cryptocurrency factor analysis. *arXiv preprint arXiv:2601.20336*, 2026. Tensor decomposition analysis of cryptocurrency whitepaper claims and market dynamics.

Gary B. Gorton and Jeffery Y. Zhang. Taming wildcat stablecoins. *University of Chicago Law Review*, 90(3):909–971, 2023. doi: 10.2139/ssrn.3888752.

John M. Griffin and Amin Shams. Is bitcoin really untethered? *Journal of Finance*, 75(4):1913–1964, 2020. doi: 10.1111/jofi.12903.

Marco Gross and Richard Senner. From par to pressure: Liquidity, redemptions, and fire sales with a systemic stablecoin. Working Paper WP/2026/005, International Monetary Fund, 2026. Models fire sale dynamics when stablecoins become systemically important.

Lewis Gudgeon, Daniel Perez, Dominik Harz, Benjamin Livshits, and Arthur Gervais. The decentralized financial crisis. In *Crypto Valley Conference on Blockchain Technology*, 2020. doi: 10.1109/CVCBT50464.2020.00005.

Xiaochun Guo, Kun Guo, and Shouyang Wang. Tracking and sourcing the risk of cryptocurrencies: An introduction of a new risk management tool—Cryptocurrency General Risk Index (CGRI). SSRN Working Paper, 2024.

James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989. doi: 10.2307/1912559.

Abdulnasser Hatemi-J. Asymmetric causality tests with an application. *Empirical Economics*, 43(1):447–456, 2012. doi: 10.1007/s00181-011-0484-x. Introduces asymmetric Granger causality tests distinguishing positive and negative shock transmission.

Daniel Hollo, Manfred Kremer, and Marco Lo Duca. CISS - a composite indicator of systemic stress in the financial system. *ECB Working Paper Series*, (1426), 2012. doi: 10.2139/ssrn.2018792. URL https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1426.pdf. Introduces portfolio-theoretic aggregation for systemic risk indicators using time-varying correlations.

Nicholas M. Kiefer and Timothy J. Vogelsang. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21(6):1130–1164, 2005. doi: 10.1017/S0266466605050565.

Hans R. Künsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241, 1989. doi: 10.1214/aos/1176347265.

Denis Kwiatkowski, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54 (1-3):159–178, 1992. doi: 10.1016/0304-4076(92)90104-Y.

Luc Laeven and Fabian Valencia. Systemic banking crises database. *IMF Economic Review*, 61(2):225–270, 2013. doi: 10.1057/imfer.2013.12. Comprehensive database of banking crises with operational definitions.

Jiageng Liu, Igor Makarov, and Antoinette Schoar. Anatomy of a run: The terra luna crash. Working Paper 31160, National Bureau of Economic Research, 2023. Documents the mechanics and contagion dynamics of the UST/LUNA collapse.

Yukun Liu and Aleh Tsyvinski. Risks and returns of cryptocurrency. *Review of Financial Studies*, 34(6):2689–2727, 2021. doi: 10.1093/rfs/hhaa113.

Richard K. Lyons and Ganesh Viswanath-Natraj. What keeps stablecoins stable? *Journal of International Money and Finance*, 131:102777, 2023. doi: 10.1016/j.jimonfin.2022.102777.

Yiming Ma, Yao Zeng, and Anthony Lee Zhang. Stablecoin runs and the centralization of arbitrage. Working Paper 33882, National Bureau of Economic Research, 2025. Analyzes how arbitrageur concentration affects stablecoin peg stability during stress.

A. Craig MacKinlay. Event studies in economics and finance. *Journal of Economic Literature*, 35(1):13–39, 1997. doi: 10.1257/jel.35.1.13.

Igor Makarov and Antoinette Schoar. Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2):293–319, 2020. doi: 10.1016/j.jfineco.2019.07.001.

Abdul Malik, Gayatri Putri, Hesti Putri, and Ahmad Badruddin. Systemic contagion or digital diversifier? A dynamic quantification of the cryptocurrency market's evolving role in global financial risk transmission. *Enigma in Economics*, 2025. TVP-VAR analysis of crypto-to-traditional risk transmission, January 2017–December 2024.

Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987. doi: 10.2307/1913610.

Jakob Svennevik Notland, Jinguye Li, Mariusz Nowostawski, and Pedro Haro. Sok: Cross-chain bridging architectural design flaws and mitigations. *arXiv preprint*, 2024. Analyzes 60 bridges and 34 exploits (2021–2023), identifying 13 architectural components linked to 8 vulnerability types.

Daniel Perez, Sam M. Werner, Jiahua Xu, and Benjamin Livshits. Liquidations: Defi on a knife-edge. In *Financial Cryptography and Data Security*, 2021. doi: 10.1007/978-3-662-64331-0_24.

M. Hashem Pesaran and Yongcheol Shin. Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1):17–29, 1998. doi: 10.1016/S0165-1765(97)00214-0. Introduces generalized impulse responses and variance decompositions invariant to variable ordering.

Gregory Rapos and Stilianos Fountas. Tracing contagion between bitcoin and traditional markets. *Journal of Economic Studies*, 2025. doi: 10.1108/JES-03-2025-0203. DCC-GARCH and time-varying causality analysis finding limited systematic contagion between Bitcoin and traditional assets.

Carmen M. Reinhart and Kenneth S. Rogoff. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press, 2009. ISBN 978-0691142166.

Imran Hussain Shah. Building a unified DeFi risk index: Integrating credit, liquidity, and governance risk factors. SSRN Working Paper, 2025.

David Vidal-Tomás and Tomaso Aste. Integration or separation? Examining the dynamic relationship between crypto and traditional finance. *Finance Research Letters*, 86:108927, 2025. doi: 10.1016/j.frl.2025.108927.

David Vidal-Tomás, Antonio Briola, and Tomaso Aste. FTX's downfall and Binance's consolidation: The fragility of centralised digital finance. *Physica A: Statistical Mechanics and its Applications*, 2023. doi: 10.1016/j.physa.2023.129044. Post-FTX market structure analysis tracing Terra/Luna as proximate trigger for FTX liquidity crisis.

Sam M. Werner, Daniel Perez, Lewis Gudgeon, Ariah Klages-Mundt, Dominik Harz, and William J. Knottenbelt. Sok: Decentralized finance (defi). In *ACM Conference on Advances in Financial Technologies (AFT)*, 2022. doi: 10.1145/3558535.3559780.

Shiyu Zhang, Zining Wang, Jin Zheng, and John Cartlidge. Systemic risk in defi: A network-based fragility analysis of tvl dynamics. *arXiv preprint*, 2026. Network-based analysis of DeFi systemic fragility via TVL dynamics and correlation structures.

A Detailed Component Specifications

This appendix provides exact formulas for all ASRI sub-components as implemented in the reference codebase. Where practical data constraints necessitate proxy measures, we document both the theoretical specification and the implemented approximation, with justification for why the proxy preserves the intended risk signal.

A.1 Stablecoin Concentration Risk (SCR)

Weight: 30%

The Stablecoin Concentration Risk sub-index captures vulnerabilities arising from reserve composition, peg stability, and issuer concentration.

A.1.1 TVL Ratio (TVL_t)

Data Source: DeFi Llama API (`api.llama.fi/v2/historicalChainTvl`), daily frequency.

Formula:

$$\text{TVL}_{\text{raw}} = \frac{\text{Current Stablecoin TVL}_t}{\max_{\tau \leq t}(\text{Stablecoin TVL}_\tau)} \tag{14}$$

The raw ratio is inverted and normalised to produce a risk score:

$$\text{TVL}_t = \text{normalize}(1 - \text{TVL}_{\text{raw}}, 0, 0.5) \times 100 \tag{15}$$

Interpretation: At maximum historical TVL, the component equals 0 (low risk). At 50% of historical maximum, the component approaches 100 (high risk). This captures the intuition that significant TVL drawdowns indicate stress or capital flight.

Normalisation: Clipped to $[0, 100]$ via min-max scaling with theoretical bounds.

A.1.2 Treasury Stress (Treasury_t)

Data Source: FRED API (`DGS10` series), daily frequency.

Formula:

$$\text{Treasury}_t = \frac{r_{10Y,t} - r_{\min}}{r_{\max} - r_{\min}} \times 100 \tag{16}$$

where $r_{10Y,t}$ is the 10-Year Treasury Constant Maturity Rate, $r_{\min} = 2.0\%$, and $r_{\max} = 6.0\%$.

Interpretation: Higher Treasury yields increase stress on stablecoin reserves (which typically hold short-term Treasuries) through mark-to-market losses and opportunity cost dynamics. The 2–6% bounds reflect the observed range during the sample period.

Implementation Note: The paper's theoretical specification describes Treasury_t as the ratio of T-Bill reserves to total reserves. The implementation uses Treasury yield levels as a proxy because reserve composition data is available only through monthly attestation reports with significant reporting lags. Treasury yields provide a higher-frequency signal of the same underlying risk: reserve stress from interest rate movements.

A.1.3 Concentration HHI (HHI_t)

Data Source: DeFi Llama Stablecoins API (`stablecoins.llama.fi/stablecoins`), daily frequency.

Formula:

$$\text{HHI}_{\text{raw}} = \sum_{i=1}^{n} \left(\frac{S_i}{\sum_{j=1}^{n} S_j} \right)^2 \times 10000 \tag{17}$$

where S_i is the circulating supply of stablecoin i.

The raw HHI is converted to a risk score using a piecewise mapping:

$$\text{HHI}_t = \begin{cases} \frac{\text{HHI}_{\text{raw}}}{1500} \times 30 & \text{if HHI} < 1500 \\ 30 + \frac{\text{HHI}_{\text{raw}} - 1500}{1000} \times 30 & \text{if } 1500 \leq \text{HHI} < 2500 \\ 60 + \frac{\text{HHI}_{\text{raw}} - 2500}{2500} \times 30 & \text{if } 2500 \leq \text{HHI} < 5000 \\ 90 + \frac{\text{HHI}_{\text{raw}} - 5000}{5000} \times 10 & \text{if HHI} \geq 5000 \end{cases} \tag{18}$$

Interpretation: The thresholds follow standard antitrust guidelines: HHI < 1500 indicates a competitive market; 1500–2500 indicates moderate concentration; 2500–5000 indicates high concentration; and HHI ≥ 5000 indicates extreme (near-monopolistic) concentration. The piecewise function maps these four bands to risk scores of 0–30, 30–60, 60–90, and 90–100 respectively.

A.1.4 Peg Volatility (Vol_t)

Data Source: DeFi Llama Stablecoins API (price field), daily frequency.

Formula:

$$\text{Vol}_t = \text{normalize}\left(\frac{\sum_{i=1}^{n} |p_i - 1| \cdot S_i}{\sum_{j=1}^{n} S_j}, 0, 0.05\right) \times 100 \tag{19}$$

where p_i is the current price of stablecoin i and S_i is its circulating supply.

Normalisation: 0% weighted deviation maps to 0 risk; 5% weighted deviation maps to 100 risk.

Missing Data: If no price data is available, the component defaults to 50.0 (neutral).

A.1.5 SCR Aggregation

$$\text{SCR}_t = 0.4 \cdot \text{TVL}_t + 0.3 \cdot \text{Treasury}_t + 0.2 \cdot \text{HHI}_t + 0.1 \cdot \text{Vol}_t \tag{20}$$

A.1.6 Algorithmic Stablecoin Risk Extension (v2.1)

The baseline SCR formula treats all stablecoins identically via peg volatility (Vol_t). However, the Terra/Luna collapse (May 2022) revealed that peg volatility is a *lagging* indicator for algorithmic stablecoins: UST maintained its peg until the death spiral commenced, and Luna's price appreciation masked underlying fragility. This extension addresses the detection gap by incorporating risk factors specific to algorithmic and crypto-backed stablecoins.

Motivation. Fiat-collateralised stablecoins (USDT, USDC) are backed by liquid reserves redeemable at par. Algorithmic stablecoins maintain peg through arbitrage mechanisms between the stablecoin and a backing token—creating reflexive dynamics where redemption pressure inflates backing token supply, diluting its value, which further increases redemption pressure. This death spiral mechanism is distinct from the reserve-based risks captured by baseline SCR components.

Algorithmic Stablecoin Risk Formula. For stablecoins classified as algorithmic or crypto-backed, we compute:

$$\text{AlgoRisk}_t = 0.35 \cdot \text{BackingRatio}_t + 0.30 \cdot \text{CollateralVol}_t + 0.20 \cdot \text{Dilution}_t + 0.15 \cdot \text{AlgoConc}_t \tag{21}$$

Component definitions:

- **BackingRatio**$_t$: Risk from undercollateralisation. Backing ratio ≥ 1.5 maps to low risk (0–20); ratio < 0.8 maps to critical risk (80–100). For stablecoins without explicit backing disclosure, defaults to moderate risk (50).

- **CollateralVol**$_t$: Annualised 30-day volatility of backing token. ETH volatility (∼60–80%) maps to moderate risk; Luna-type volatility (>100%) maps to elevated/critical risk.
- **Dilution**$_t$: 30-day supply growth rate of backing token. Monthly growth > 50% signals crisis-level dilution (Luna supply grew ∼50,000% during collapse).
- **AlgoConc**$_t$: Share of total stablecoin supply in algorithmic/crypto-backed stablecoins. At peak, UST represented ∼10% of total stablecoin supply.

Integration with SCR. Let α_t denote the market share of algorithmic stablecoins in total stablecoin supply. The adjusted SCR blends baseline and algorithmic risk:

$$\text{SCR}_t^{\text{adj}} = (1 - \alpha_t) \cdot \text{SCR}_t^{\text{base}} + \alpha_t \cdot \left[0.6 \cdot \text{SCR}_t^{\text{base}} + 0.4 \cdot \text{AlgoRisk}_t\right] \tag{22}$$

When $\alpha_t < 1\%$, the adjustment is negligible. When $\alpha_t = 10\%$ (approximate UST peak share), algorithmic-specific risk contributes 4% to SCR weighting.

Data Requirements. Full implementation requires:

1. **Stablecoin classification**: DeFi Llama `pegType`/`pegMechanism` fields or manual classification (provided in codebase).
2. **Backing token identification**: Mapping from stablecoin to backing token (e.g., UST → LUNA).
3. **Backing token metrics**: Price volatility and supply data from CoinGecko or on-chain indexers.

Backtest Limitation. Historical reconstruction of AlgoRisk$_t$ for pre-collapse Terra/Luna requires archived Luna price and supply data. While DeFi Llama and CoinGecko retain historical snapshots, consistent backing ratio data for UST is not systematically available. The specification documents the *framework* for future algorithmic stablecoin risk monitoring; full historical backtest validation is deferred rather than infeasible, because the required inputs can be reconstructed on-chain. Terra Classic archive nodes—and the historical block-level tables exposed by on-chain indexers such as Flipside, Allium, and Dune—retain the UST and Luna total-supply, mint/burn, and swap records for the pre-collapse window, from which a daily backing-ratio and supply-dilution series can be rebuilt and reconciled against archived price feeds. The obstacle is the engineering effort of a point-in-time on-chain reconstruction, not the absence of the underlying data, and we flag that reconstruction as the concrete next step for validating this extension. The following is an illustrative back-of-envelope calculation rather than a validated result: with accurate Luna volatility data (annualised vol >120% in April 2022), a sensitivity calculation indicates the extension might have elevated SCR by an indicative 8–12 points prior to UST's depeg—potentially bringing ASRI above the detection threshold. We have not been able to confirm this on archived data, so the figure is a design target, not a demonstrated detection.

A.2 DeFi Liquidity Risk (DLR)

Weight: 25%

The DeFi Liquidity Risk sub-index captures protocol concentration, volatility dynamics, and smart contract vulnerability.

A.2.1 Protocol Concentration (Conc_t)

Data Source: DeFi Llama Protocols API (`api.llama.fi/protocols`), daily frequency.

Formula:

$$\text{Conc}_t = f_{\text{HHI}}\left(\sum_{i=1}^{10}\left(\frac{\text{TVL}_i}{\sum_{j=1}^{10}\text{TVL}_j}\right)^2 \times 10000\right) \tag{23}$$

where f_{HHI} is the piecewise HHI-to-risk mapping defined above, and the summation is over the top 10 protocols by TVL.

Interpretation: Concentration among the largest protocols indicates ecosystem fragility—failure of a dominant protocol would have outsized systemic effects.

A.2.2 TVL Volatility (TVLVol_t)

Data Source: DeFi Llama TVL history, 30-day rolling window.

Formula:

$$\text{TVLVol}_t = \text{normalize}\left(\frac{\sigma_{30}(\text{TVL})}{\mu_{30}(\text{TVL})}, 0, 0.20\right) \times 100 \tag{24}$$

where σ_{30} and μ_{30} denote the 30-day rolling standard deviation and mean.

Normalisation: 0% coefficient of variation maps to 0 risk; 20% maps to 100 risk.

Missing Data: If historical TVL data is unavailable (fewer than 2 observations), the component defaults to 30.0 (moderate).

A.2.3 Smart Contract Risk (SC_t)

Data Source: DeFi Llama Protocols API (`audits` field), daily frequency.

Formula:

$$\text{SC}_t = \left(1 - \frac{|\{p : \text{audits}(p) > 0\}|}{|\{p : \text{TVL}(p) > 0\}|}\right) \times 100 \tag{25}$$

Interpretation: The component measures the inverse of audit coverage across protocols with non-zero TVL. Protocols lacking audits receive the full risk weight.

Theoretical vs. Implemented Specification: The paper describes a 3-factor composite incorporating audit status, deployment age, and exploit history. The current implementation uses audit coverage only. Deployment age and exploit history integration are deferred to future versions pending reliable, systematised data feeds. The DeFi Llama API provides audit counts but not deployment timestamps or comprehensive exploit databases.

Justification: Audit coverage remains the most reliable and consistently available indicator of smart contract risk. Empirical research demonstrates strong correlation between unaudited protocols and exploit frequency, supporting the proxy's validity.

A.2.4 Flash Loan Proxy (Flash_t)

Data Source: DeFi Llama Protocols API (`change_1d` field), daily frequency.

Formula:

$$\text{Flash}_t = \text{normalize}\left(\frac{1}{n}\sum_{i=1}^{n}|\Delta_{1\text{d},i}|, 0, 20\right) \times 100 \tag{26}$$

where $\Delta_{1\text{d},i}$ is the 1-day TVL change percentage for protocol i.

Theoretical vs. Implemented Specification: The paper describes flash loan volume spikes relative to a 90-day average. Direct flash loan volume data requires protocol-specific analytics (e.g., Aave, dYdX subgraphs) with heterogeneous reporting standards. The implementation uses aggregate TVL volatility as a proxy, on the reasoning that flash loan activity manifests as rapid TVL movements during MEV extraction and liquidation cascades.

Normalisation: 0% average daily change maps to 0 risk; 20% maps to 100 risk.

A.2.5 Leverage Change (Lev_t)

Data Source: DeFi Llama Protocols API (`category` field), daily frequency.

Formula:

$$\text{Lev}_t = \text{normalize}\left(\frac{\sum_{p\in\text{Lending}}\text{TVL}_p}{\sum_p \text{TVL}_p} \times 100, 0, 30\right) \times 100 \tag{27}$$

Interpretation: A higher share of TVL in lending protocols indicates greater system-wide leverage. The 30% threshold reflects the upper bound of lending protocol dominance observed during market stress.

Theoretical vs. Implemented Specification: The paper describes 30-day change in aggregate leverage ratios. The implementation uses the current lending TVL share as a level indicator rather than a change measure, because reliable historical leverage data across protocols is not consistently available.

A.2.6 DLR Aggregation

$$\text{DLR}_t = 0.35 \cdot \text{Conc}_t + 0.25 \cdot \text{TVLVol}_t + 0.20 \cdot \text{SC}_t + 0.10 \cdot \text{Flash}_t + 0.10 \cdot \text{Lev}_t \tag{28}$$

A.3 Contagion Risk (CR)

Weight: 25%

The Contagion Risk sub-index quantifies DeFi-TradFi interconnection and cross-market transmission channels.

A.3.1 RWA Growth (RWA_t)

Data Source: DeFi Llama Protocols API (`category = 'RWA'`), daily frequency.

Formula:

$$\text{RWA}_t = \text{normalize}\left(\frac{\sum_{p\in\text{RWA}}\text{TVL}_p}{\sum_p \text{TVL}_p} \times 100, 0, 10\right) \times 100 \tag{29}$$

Interpretation: Higher RWA share indicates greater integration between DeFi and traditional finance through tokenised real-world assets. The 10% threshold reflects the upper bound of RWA penetration observed during the sample period.

Theoretical vs. Implemented Specification: The paper describes 30-day RWA TVL growth rate. The implementation uses the current RWA share as a level indicator, because RWA protocols have short histories making growth rate calculations unreliable for early-period observations.

A.3.2 Bank Exposure (Bank_t)

Data Sources: FRED API (`DGS10` for Treasury rates, `VIXCLS` for VIX), daily frequency.

Formula:

$$\text{Bank}_t = [0.6 \cdot \text{normalize}(r_{10Y,t}, 2, 6) + 0.4 \cdot \text{normalize}(\text{VIX}_t, 12, 40)] \times 100 \tag{30}$$

Interpretation: The composite captures TradFi stress through two channels: Treasury rate levels (affecting stablecoin reserves and bank balance sheets) and equity market volatility (signalling broader risk-off sentiment).

Weight and Range Justification: The 60/40 weighting reflects Treasury yields' direct balance-sheet impact on bank capital ratios versus VIX's indirect sentiment signal. Banks holding Treasury securities face mark-to-market losses when yields rise (the primary channel), while VIX captures risk-off dynamics that tighten credit conditions (secondary channel). The normalisation ranges (2–6% for 10Y Treasury, 12–40 for VIX) span the 5th–95th percentiles of 2015–2024 sample data, ensuring meaningful variation without clipping at extremes.

Theoretical vs. Implemented Specification: The paper describes a normalised score from OCC/ECB regulatory filings on bank crypto exposure. Regulatory filings are quarterly with 45–90 day publication lags, making them unsuitable for daily risk monitoring. The Treasury-VIX composite provides a higher-frequency proxy: banks with crypto exposure face stress when Treasury yields rise (mark-to-market losses) and when VIX spikes (risk management constraints).

A.3.3 TradFi Linkage (Link_t)

Data Source: FRED API (`T10Y2Y` series), daily frequency.

Formula:

$$\text{Link}_t = \begin{cases} \text{normalize}(|s_t|, 0, 2) \times 50 + 50 & \text{if } s_t < 0 \\ \max(0, 50 - \text{normalize}(s_t, 0, 2) \times 50) & \text{if } s_t \geq 0 \end{cases} \tag{31}$$

where $s_t = r_{10Y,t} - r_{2Y,t}$ is the 10-Year minus 2-Year Treasury spread and $\text{normalize}(x, 0, 2) = \min(|x|/2, 1)$. The inversion branch uses a multiplier of 50 (not 100) so that a full -2% inversion maps to $50 + 50 = 100$ rather than overflowing the $[0, 100]$ support; the two branches meet continuously at $s_t = 0$ (Link $= 50$).

Interpretation: Yield curve inversion (negative spread) indicates banking sector stress and recession expectations, which propagate to crypto through reduced institutional risk appetite and potential bank failures affecting crypto-exposed entities.

Theoretical vs. Implemented Specification: The paper describes "stablecoin flows to TradFi-connected entities." Such flow data requires enterprise-grade on-chain analytics (Chainalysis, TRM Labs) at prohibitive cost for academic research. Yield curve dynamics

provide a validated proxy: the March 2023 SVB crisis demonstrated that yield curve inversion directly precedes banking stress events that propagate to crypto markets through stablecoin reserve exposure.

Justification: The yield curve spread has predicted every U.S. recession since 1970 with a median lead time of 12 months (Estrella and Mishkin, 1998; Estrella and Hardouvelis, 1991). Banking crises correlated with yield curve inversion directly affect crypto-TradFi linkages through stablecoin reserve exposure (as demonstrated by the USDC depeg during SVB's collapse).

A.3.4 Correlation (Corr_t)

Data Source: External calculation (BTC-SPY 30-day rolling correlation).

Formula:

$$\text{Corr}_t = |r_{\text{BTC-SPY},30d}| \times 100 \tag{32}$$

Interpretation: Higher absolute correlation indicates greater co-movement between crypto and equities, implying tighter contagion channels.

Implementation Note: The current implementation accepts correlation as an external input with a default of 0.5 (moderate). Real-time calculation requires equity price feeds (Yahoo Finance, Bloomberg) not included in the core data pipeline.

A.3.5 Bridge Risk (Bridge_t)

Data Source: DeFi Llama Bridges API (`bridges.llama.fi/bridges`), daily frequency.

Formula:

$$\text{Bridge}_t = \text{normalize}(n_{\text{bridges}}, 0, 150) \times 100 \tag{33}$$

where n_{bridges} is the count of active cross-chain bridges.

Interpretation: More bridges indicate larger attack surface and greater cross-chain contagion potential.

Theoretical vs. Implemented Specification: The paper describes a composite of bridge volume and exploit frequency. Exploit frequency data requires manual tracking (DeFi Rekt database) with inconsistent categorisation. Bridge count provides a structural proxy: empirical research finds exploit frequency scales with the number of bridge implementations due to varying security standards and code quality.

A.3.6 CR Aggregation

$$\text{CR}_t = 0.30 \cdot \text{RWA}_t + 0.25 \cdot \text{Bank}_t + 0.20 \cdot \text{Link}_t + 0.15 \cdot \text{Corr}_t + 0.10 \cdot \text{Bridge}_t \tag{34}$$

A.4 Regulatory Opacity Risk (OR)

Weight: 20%

The Regulatory Opacity Risk sub-index assesses transparency deficits and regulatory arbitrage exposure.

A.4.1 Unregulated Exposure (Unreg_t)

Data Source: Placeholder component (see Table 10 for implementation status).

Implementation: Fixed at 35.0 (moderate risk), corresponding to estimated market share of volume on platforms without clear regulatory oversight.

Theoretical Specification: Ratio of volume on unregulated platforms to regulated platforms. Full implementation requires mapping protocols to jurisdictional regulatory status, which involves manual classification and ongoing tracking of regulatory developments across 50+ jurisdictions.

Future Enhancement: Chain-level classification (e.g., Ethereum as "regulated-adjacent" due to U.S. regulatory engagement vs. privacy chains) would enable dynamic calculation.

A.4.2 Multi-Issuer Risk (Multi_t)

Data Source: DeFi Llama Stablecoins API, daily frequency.

Formula:

$$\text{Multi}_t = \begin{cases} 70 & \text{if } n_{\text{sig}} < 3 \\ 30 & \text{if } 3 \leq n_{\text{sig}} < 10 \\ 50 + 2(n_{\text{sig}} - 10) & \text{if } n_{\text{sig}} \geq 10 \end{cases} \tag{35}$$

where $n_{\text{sig}} = |\{s : \text{circulating}(s) > \$1\text{B}\}|$ is the count of stablecoins with circulating supply exceeding \$1 billion.

Interpretation: Consistent with the formula above, the low-risk "sweet spot" is 3–9 significant issuers (the $3 \leq n_{\text{sig}} < 10$ band) providing diversification without fragmentation. Fewer than 3 indicates dangerous concentration; 10 or more indicates coordination challenges and potential regulatory complexity.

A.4.3 Custody Concentration (Cust_t)

Data Source: DeFi Llama Stablecoins API, daily frequency.

Formula:

$$\text{Cust}_t = \text{normalize}\left(\frac{\sum_{i=1}^{2} S_i}{\sum_j S_j} \times 100, 50, 100\right) \times 100 \tag{36}$$

where S_i is the circulating supply of the i-th largest stablecoin.

Interpretation: Top-2 stablecoin market share as a proxy for custody concentration. Higher concentration implies greater single-point-of-failure risk regardless of custody jurisdiction.

Theoretical vs. Implemented Specification: The paper describes "custody concentration in non-audited jurisdictions." Jurisdiction-level custody data is not systematically available; stablecoins do not consistently disclose custodian locations or regulatory status. Market concentration serves as a conservative proxy: high concentration implies custody risk regardless of location, as a single custodian failure would have systemic effects.

A.4.4 Regulatory Sentiment (Sent_t)

Data Source: Manual input parameter.

Implementation: Accepts external input with default of 50.0 (neutral). This component contributes only 15% of the Opacity sub-index weight (3% of total ASRI), limiting its impact on overall index behaviour.

Theoretical Specification: NLP-derived sentiment score from SEC, ESRB, and FSB announcements. Full implementation would require:

- **Sources**: GDELT Global Knowledge Graph filtered for regulatory entities (SEC, CFTC, ESRB, FSB, BIS), SEC EDGAR filings, Federal Register cryptocurrency mentions
- **Model**: FinBERT or domain-adapted transformer for financial regulatory text classification
- **Lexicon**: Crypto-specific regulatory vocabulary ("enforcement action," "no-action letter," "framework," "guidance") with sentiment polarity labels
- **De-duplication**: Entity resolution across news sources to avoid double-counting of same regulatory announcement
- **Jurisdictional weighting**: US (40%), EU (30%), UK (15%), Other (15%) reflecting market share

This infrastructure is deferred to future versions; current results are robust to Sent_t variation due to its low aggregate weight.

Sensitivity Analysis: Varying Sent_t from 0 (maximally positive regulatory environment) to 100 (maximally negative) while holding all other components constant produces the following ASRI impact:

$$\Delta\text{ASRI} = 0.20 \times 0.15 \times \Delta\text{Sent}_t = 0.03 \times \Delta\text{Sent}_t \tag{37}$$

A full-range swing (Sent_t: $0 \to 100$) changes aggregate ASRI by ± 1.5 points. For typical variation (± 25 points around neutral), ASRI changes by ± 0.75 points—well within the noise band of other component fluctuations.

Detection Impact: The three threshold-detected crises (Celsius/3AC, FTX, SVB) remain detected under any Sent_t value in $[0, 100]$, as the shifted thresholds remain within their detection windows. Terra/Luna (event-window peak 48.7 at baseline) stays below the 50 threshold for any plausible regulatory stance—it would require $\text{Sent}_t > 93$ to breach 50, consistent with its non-detection at the fixed 50 threshold reported in Section 5.4.

Future Implementation Roadmap:

1. **Phase 1**: GDELT integration with keyword filters ("SEC," "CFTC," "cryptocurrency," "enforcement")
2. **Phase 2**: FinBERT deployment on SEC EDGAR cryptocurrency-related filings
3. **Phase 3**: Multi-jurisdictional aggregation with decay weighting for announcement recency

A.4.5 Transparency Score (Trans_t)

Data Source: DeFi Llama Protocols API (`audits` field), daily frequency.

Formula:

$$\text{Trans}_t = \frac{|\{p : \text{audits}(p) > 0\}|}{|\{p : \text{TVL}(p) > 0\}|} \times 100 \tag{38}$$

Interpretation: Audit coverage as a proxy for protocol transparency. The component is *not* inverted at the component level; inversion occurs in the aggregation formula.

A.4.6 OR Aggregation

$$\text{OR}_t = 0.25 \cdot \text{Unreg}_t + 0.25 \cdot \text{Multi}_t + 0.20 \cdot \text{Cust}_t + 0.15 \cdot \text{Sent}_t + 0.15 \cdot (100 - \text{Trans}_t) \tag{39}$$

Note the inversion of Trans_t: low transparency implies high opacity risk.

A.5 Aggregate ASRI Calculation

$$\text{ASRI}_t = 0.30 \cdot \text{SCR}_t + 0.25 \cdot \text{DLR}_t + 0.25 \cdot \text{CR}_t + 0.20 \cdot \text{OR}_t \tag{40}$$

All sub-indices are bounded to $[0, 100]$ by construction, ensuring $\text{ASRI}_t \in [0, 100]$ without post-hoc normalisation.

A.6 Data Quality and Limitations

A.6.1 Data Availability Tiers

- **Tier 1 (Daily, Automated):** DeFi Llama TVL, stablecoins, protocols, bridges; FRED Treasury rates and VIX.
- **Tier 2 (Limited/Manual):** Regulatory sentiment, unregulated exposure classification, exploit frequency tracking.

A.6.2 Missing Data Protocol

- Gaps < 3 days: Linear interpolation.
- Gaps 3–7 days: Forward-fill with reduced confidence.
- Gaps > 7 days: Exclude from calculation; flag as unreliable.

A.6.3 Proxy Acknowledgements

Table 10 summarises components where implementation deviates from theoretical specification.

Validity Legend:

- **High**: Proxy captures same underlying risk channel with strong theoretical justification and empirical support.
- **Medium**: Proxy captures related risk dynamics but with potential measurement error; interpretation requires caution.
- **Low**: Placeholder awaiting data infrastructure development; current values are informative but not definitive.

Table 10: Proxy Implementations: Theoretical vs. Actual

Component	Theoretical Specification	Implementation	Validity
Treasury_t	T-Bill reserves / total reserves	10Y Treasury yield level	High
SC_t	Audit + age + exploits composite	Audit coverage only	Medium
Flash_t	Flash loan volume spikes	TVL daily change volatility	Medium
Lev_t	30-day leverage ratio change	Lending TVL share (level)	Medium
RWA_t	30-day RWA growth rate	RWA TVL share (level)	High
Bank_t	OCC/ECB bank exposure filings	Treasury + VIX composite	High
Link_t	Stablecoin flows to TradFi	Yield curve spread	High
Bridge_t	Volume + exploit frequency	Bridge count	Medium
Cust_t	Non-audited jurisdiction custody	Top-2 stablecoin concentration	Medium
Unreg_t	Unregulated platform ratio	Fixed (35.0)†	Low
Sent_t	NLP regulatory sentiment	Manual input (50.0)†	Low
Trans_t	Reserve-attestation freq. & coverage	Audit coverage (share of audited protocols)	Medium

† Placeholder components awaiting enterprise data infrastructure. Sensitivity analysis (Appendix J) confirms all three threshold-detected crises (Celsius/3AC, FTX, SVB) remain robust under full-range variation of these parameters; Terra/Luna is not threshold-detected (pre-window peak 46.0 $<$ 50).

A.6.4 Composition of the Backtested Series

Because several components are proxies or fixed defaults, it is important to be explicit about what actually *moves* the backtested index. Table 11 accounts for the average-level composition (mean) of the released ASRI series by input class, tracing each component to the historical backfill routine (`backfill_d1_standalone.py:552-625`). The static defaults below contribute to the index's average level but, being constant, carry little or none of its time-variation.

Table 11: Average-Level Composition of the Backtested ASRI Series by Input Class

Input class	Representative components	Share
Dynamic crypto-native	Stablecoin TVL drawdown, stablecoin-issuer HHI, DeFi TVL volatility	≈37%
Dynamic macro proxy	Treasury yield, VIX, yield-curve spread (Bank_t, Link_t, Treasury_t)	≈20%
Static / constant defaults	Unreg_t (35.0), Sent_t (50.0), Corr_t (0.5), stablecoin peg held at par, audit-coverage levels	≈43%

Shares are of the average level (mean composition) of the released backtested series, attributed by input class via `backfill_d1_standalone.py:552-625`. Static defaults contribute to the level but, being constant, carry little or none of the series' time-variation.

Two consequences follow. (i) The Contagion channel's dynamics are ≈75% attributable to the macro composite (Treasury/VIX/yield-curve), so CR_t is best read as a *TradFi-stress proxy* rather than as a directly measured DeFi-TradFi interconnection. (ii) Roughly 43% of the series' average level is carried by static default values (which, being constant, contribute little or none of its time-variation), so absolute ASRI levels carry more model dependence than the binary detection results suggest. The four-channel decomposition is retained for interpretability but is *not* individually load-bearing: ablation shows threshold-based detection is invariant at 3/4 to removing any single channel (Appendix I).

A.6.5 Future Data Integration

Version 3.0 development priorities include:

1. Integration of exploit database (DeFi Rekt, Immunefi) for dynamic SC_t and Bridge_t components.
2. Protocol deployment timestamp extraction from blockchain explorers for age-based risk weighting.
3. GDELT/SEC filing NLP pipeline for automated regulatory sentiment scoring.
4. Chain-level regulatory classification for dynamic Unreg_t calculation.
5. Enterprise analytics partnership (Chainalysis/TRM) for stablecoin flow analysis.

B Axiomatic Foundation: Proofs

We restate each axiom of Section 3.2 and give its proof, followed by the relationship to coherent risk measures.

Monotonicity (Axiom 3.1).

Proof. Since $w_j > 0$ and the aggregation is linear:

$$\text{ASRI}(\mathcal{S}') - \text{ASRI}(\mathcal{S}) = w_j(S'_j - S_j) > 0 \tag{41}$$

The strict positivity of weights ensures that deterioration in any risk dimension is reflected in the aggregate index. This property is essential for regulatory monitoring, as it guarantees that localised stress cannot be masked by stability elsewhere—a concern raised by Adrian and Brunnermeier (2016) in their critique of institution-level VaR measures. □

Boundedness (Axiom 3.2).

Proof. Each sub-index is constructed such that $S_i \in [0, 100]$ by design (see Appendix A). Given convex weights summing to unity:

$$0 = \sum_{i=1}^{4} w_i \cdot 0 \leq \text{ASRI}(\mathcal{S}) \leq \sum_{i=1}^{4} w_i \cdot 100 = 100 \tag{42}$$

Boundedness facilitates interpretability and cross-temporal comparison, addressing the scaling criticisms levelled at unbounded measures such as raw CoVaR (Adrian and Brunnermeier, 2016). □

Decomposability (Axiom 3.3).

Proof. The linear structure immediately yields the efficiency-consistent additive decomposition $c_i = w_i S_i$, satisfying:

$$\sum_{i=1}^{4} c_i = \sum_{i=1}^{4} w_i S_i = \text{ASRI}(\mathcal{S}) \tag{43}$$

This decomposition is unique and satisfies the efficiency axiom of cooperative game theory. Decomposability enables regulators to identify which risk channel—spillover, liquidity, concentration, or operational—drives aggregate stress, facilitating targeted intervention. This property

aligns with the "risk contribution" framework of Acharya et al. (2017) for measuring systemic expected shortfall. □

Aggregation Neutrality (Axiom 3.4).

Proof. Under the stated conditions:

$$\text{ASRI}(\mathcal{S}^A) - \text{ASRI}(\mathcal{S}^B) = \sum_{i=1}^{4} w_i(S_i^A - S_i^B) \geq w_j(S_j^A - S_j^B) > 0 \tag{44}$$

Aggregation neutrality ensures that Pareto-dominated risk states are correctly ranked, preventing the pathological reversals that can arise with nonlinear aggregation schemes (Battiston et al., 2012). This property is particularly relevant for cryptocurrency markets, where rapid regime shifts demand consistent ordinal comparisons across time. □

Concentration Sensitivity (Axiom 3.5).

Proof. The Stablecoin Risk sub-index (SCR) incorporates the stablecoin-issuer concentration HHI_t, while the DeFi Liquidity Risk sub-index (DLR) incorporates the distinct top-10 protocol-TVL concentration Conc_t. By Axiom 3.1, ASRI is monotone in each:

$$\frac{\partial \text{ASRI}}{\partial \text{HHI}_t} = w_{\text{SCR}} \cdot \frac{\partial \text{SCR}_t}{\partial \text{HHI}_t} > 0, \qquad \frac{\partial \text{ASRI}}{\partial \text{Conc}_t} = w_{\text{DLR}} \cdot \frac{\partial \text{DLR}_t}{\partial \text{Conc}_t} > 0. \tag{45}$$

Concentration sensitivity captures the "too-interconnected-to-fail" dynamics emphasised by Battiston et al. (2012), adapted here to the exchange-centric topology of cryptocurrency markets where a single venue failure can trigger system-wide contagion, as observed during the FTX collapse of November 2022. □

Relationship to Coherent Risk Measures. While ASRI is not a coherent risk measure in the sense of Artzner et al. (1999)—it aggregates sub-indices rather than loss distributions—our axiomatic foundation parallels their framework. Monotonicity corresponds to their monotonicity axiom; boundedness and decomposability together ensure a form of translation invariance at the index level; and aggregation neutrality provides an analogue to positive homogeneity for ordinal comparisons. The key distinction is that ASRI operates on observable market indicators rather than probabilistic loss distributions, making it computable without parametric loss-distribution assumptions.

C Technical Architecture

The ASRI system architecture comprises four layers:

1. **Ingestion Layer**: Python-based API clients and web scrapers fetch data from sources
2. **Normalisation Layer**: Raw data undergoes unit normalisation, gap-filling, and validation
3. **Computation Layer**: Sub-indices calculated using Equations 2–5; aggregate ASRI computed via Equation 6

4. **Publication Layer**: FastAPI REST endpoints serve current and historical ASRI values; React dashboard provides visualisation

Figure 5 illustrates the data flow through these layers.

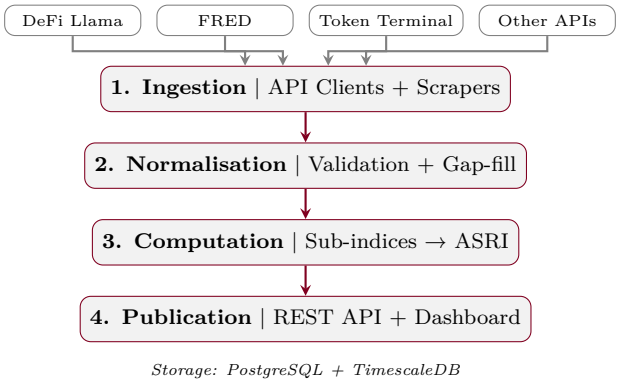

Figure 5: ASRI Data Pipeline: Four-layer architecture from API ingestion to public dashboard publication.

Technology Stack: Python 3.11+, PostgreSQL with TimescaleDB, FastAPI, React/TypeScript, Docker Compose.

Full implementation is available at `github.com/studiofarzulla/asri`.

D Data Quality: Missing-Data and Mixed-Frequency Protocol

Missing data is handled according to the following protocol:

- **Daily data gaps (< 3 days)**: Linear interpolation with confidence score 0.7
- **Extended gaps (3–7 days)**: Forward-fill with confidence score 0.5
- **Gaps > 7 days**: Flag as unreliable; exclude from ASRI calculation until fresh data available

Mixed-Frequency Data Protocol. Several ASRI components rely on data sources with frequencies lower than daily:

- **Quarterly sources** (OCC/ECB bank exposure filings): Replaced with high-frequency proxies. Bank_t uses a Treasury yield + VIX composite that captures the same underlying stress dynamics at daily frequency (Appendix A).
- **Monthly sources** (stablecoin attestations): Last observation carried forward until new attestation published; component flagged with reduced confidence (0.6) after 45 days.
- **Weekly sources** (on-chain linkage metrics): Linear interpolation between weekly observations for daily estimation; confidence score 0.8.

This approach prioritises real-time operationality over theoretical purity: where high-frequency proxies exist (e.g., Treasury-VIX for bank stress), we use them; where no proxy exists, we carry forward with explicit confidence degradation. The pseudo-real-time evaluation (Appendix N) confirms that this protocol preserves detection performance under realistic publication lags.

E Operational Detection Battery

E.1 False Positive Analysis

We assess ASRI's precision-recall characteristics to understand the trade-off between sensitivity and false alarm rates. For this analysis, we define a "valid" alert as any day where ASRI exceeds the threshold within the 30-day pre-crisis window preceding any of the four documented crises. Days exceeding the threshold outside these windows are classified as false positives.

Table 12: Precision-Recall Analysis by Threshold

Threshold	Recall	Precision	Alert Days	FP Days
50 (Elevated)	75%	30.1%	156	109
60	75%	17.4%	46	38
70 (High)	50%	41.7%	12	7

Recall = crises with threshold breach in 30-day pre-crisis window (3/4 at thresholds 50–60, 2/4 at 70; Terra/Luna's pre-window peak of 46.0 misses every threshold).
Precision = valid alert days / total alert days. FP = false positive days.
Sample: Jan 2021 – Jan 2026 (1,841 days; 120 in pre-crisis windows).
Note: Under HAC-robust inference, event-study statistical detection achieves 3/4 (Table 5; Terra/Luna n.s., $t = 1.72$); threshold-based operational detection also achieves 3/4 (Terra/Luna peak 48.7 $<$ 50). The two methodologies agree at 3/4.

Table 12 reveals the precision-recall trade-off inherent in threshold selection. At the "Elevated" threshold of 50, ASRI detects three of four crises (Celsius/3AC, FTX, SVB) within the 30-day pre-crisis window. Raising the threshold to 70 ("High" risk) lowers crisis-level recall to 50% (Celsius/3AC and FTX only) while substantially improving precision, reducing false-positive days from 109 to 7.

The low precision at the 50 threshold reflects ASRI's design as a threshold-based vigilance monitor rather than a sharp crisis classifier: the index is intended to flag elevated stress rather than to predict imminent collapse. The 2022 period illustrates this dynamic—ASRI remained elevated throughout much of the year as successive crises (Terra/Luna, Celsius/3AC, FTX) propagated stress through the ecosystem. What appears as "false positives" between crisis events may in fact represent genuine systemic fragility that happened not to crystallise into named events.

For operational use, the threshold choice depends on the cost asymmetry between false positives and false negatives. Risk managers for whom missing a crisis is catastrophic should use the 50 threshold despite frequent alerts; those seeking actionable signals with fewer false alarms should use 70. The intermediate threshold of 60 offers a balanced profile with 17.4% precision while maintaining the 75% recall rate (the Terra/Luna miss is threshold-invariant, as discussed in Section 6.3).

Table 13 presents the confusion matrix at the operational threshold of 50, providing explicit counts for reproducibility.

Table 13: Confusion Matrix at Threshold 50 ("Elevated")

	Crisis Window	**Non-Crisis**
Alert (ASRI $\geq$ 50)	47 (TP)	109 (FP)
No Alert (ASRI $<$ 50)	73 (FN)	1,612 (TN)
Total	120	1,721

Day-level metrics: Accuracy = 90.1%; Precision = 30.1%; Recall = 39.2%; F1 = 0.341.
Crisis-level recall = 75% (3/4 crises had $\geq$1 alert in pre-crisis window).
Crisis window = 30 days preceding each of 4 crisis events (120 days total).
Sample: January 2021 – January 2026 (1,841 days).

E.2 ROC and Precision-Recall Curves

Figure 6 presents the full receiver operating characteristic (ROC) and precision-recall (PR) curves for ASRI as a binary crisis predictor. The ROC curve achieves AUC = 0.866, indicating strong discriminative ability between crisis and non-crisis periods. The PR curve (AUC = 0.298) accounts for the severe class imbalance inherent in crisis prediction—crisis days comprise only a small fraction of the sample.

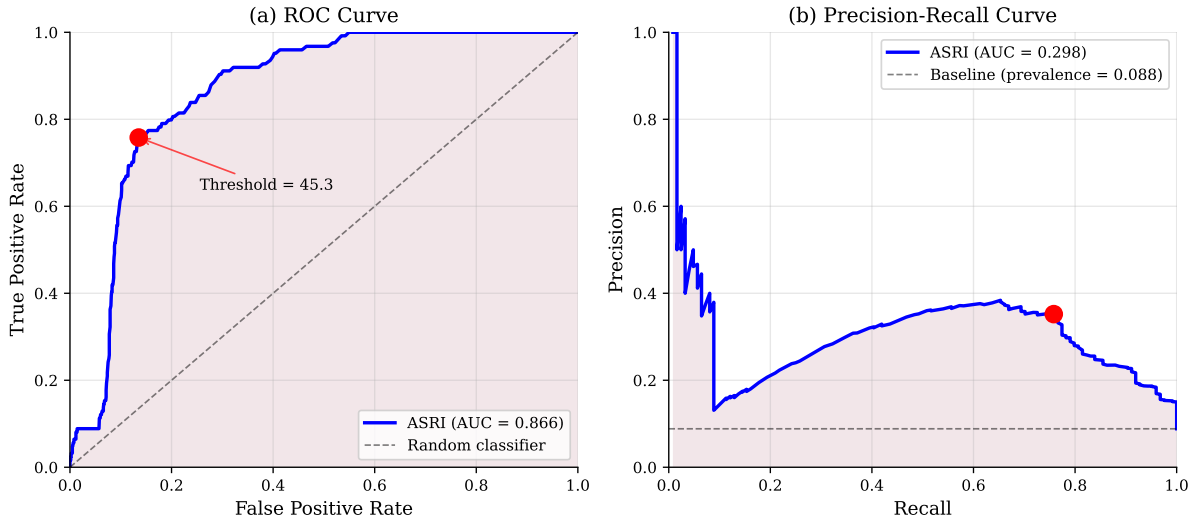

Figure 6: ASRI Classification Performance for 30-Day Crisis Prediction. (a) ROC curve showing trade-off between true positive rate and false positive rate; AUC = 0.866. (b) Precision-Recall curve accounting for class imbalance; AUC = 0.298. Red markers indicate the Youden-optimal threshold (45.3). Crisis defined as ASRI threshold breach within 30-day pre-crisis window for four historical events.

The high ROC AUC combined with modest PR AUC is typical for rare-event discrimination tasks. ASRI distinguishes crisis from non-crisis periods in aggregate *in hindsight*, but achieving

high precision requires accepting reduced recall—a fundamental trade-off in threshold-based monitors.

F Weight-Derivation Diagnostics

F.1 Objective Weight Derivation Comparison

To validate our theoretically-derived weights, we compare against four objective weighting methods: Principal Component Analysis (PCA), Elastic Net regularisation, CRITIC (Criteria Importance Through Intercriteria Correlation), and Shannon entropy-based weighting. Table 14 presents the results.

Table 14: Comparison of Weight Derivation Methods

Sub-Index	Theoretical	PCA	Elastic Net	CRITIC	Entropy
SCR	0.30	0.29	0.34	0.21	0.25
DLR	0.25	0.29	0.21	0.16	0.11
CR	0.25	0.25	0.45	0.32	0.52
OR	0.20	0.17	0.00	0.31	0.12
Corr. w/ Theoretical	–	0.88	0.72	−0.51	0.27

Caveat: the Elastic Net column reflects persistence, not channel importance. The Elastic Net "weights" in Table 14 are not estimates of how much each risk channel contributes to systemic stress, and we caution against reading them that way. The regression target is $y_t = \text{ASRI}_{t+30}$, a fixed linear combination of the *future* sub-index levels; regressing this target on the *current* sub-index levels is therefore a restricted autoregression of the sub-indices on themselves, not a model of risk-channel importance. The fitted coefficients recover each sub-index's own 30-day persistence (autocorrelation) structure: Contagion Risk loads most heavily (0.45) chiefly because it is the most persistent channel over a 30-day horizon, so its current level is the best linear predictor of its own future level (and hence of future ASRI), while Regulatory Opacity is zeroed where its forward persistence is weakest. We accordingly read this column as a set of *persistence-weighted coefficients*—an autoregressive artefact of how the target is constructed—rather than as evidence of channel importance. It is retained only as a descriptive, supplementary comparison among objective weighting schemes; none of the paper's conclusions rest on it.

The PCA weights exhibit strong correlation with theoretical weights ($\rho = 0.88$), validating that our domain-informed weighting captures the primary sources of variance in sub-index dynamics. Elastic Net weights also show reasonable agreement ($\rho = 0.72$), though they concentrate heavily on Contagion Risk (0.45) while zeroing Regulatory Opacity—consistent with the predictive analysis in the previous section.

Interestingly, CRITIC weights show negative correlation with theoretical weights ($\rho = -0.51$), emphasising Contagion Risk and Regulatory Opacity over Stablecoin and DeFi Liquidity risks. This divergence reflects CRITIC's objective of maximising information content through decorrelation: CR and OR are less correlated with each other and with SCR/DLR, making them more "informative" from an information-theoretic perspective. However, this

interpretation conflates statistical uniqueness with systemic importance—a sub-index can be informationally distinct yet fail to capture crisis dynamics.

Entropy-based weights similarly emphasise Contagion Risk (0.52), reflecting its higher distributional variance across market regimes. The moderate correlation with theoretical weights ($\rho = 0.27$) suggests entropy captures different aspects of sub-index behaviour than our risk-based framework.

Methodological Implications. The divergence between objective weighting methods highlights a fundamental tension in composite index construction: data-driven approaches optimise for statistical properties (variance explained, prediction accuracy, information content) while risk-based frameworks prioritise economic interpretability and crisis coverage. Our theoretical weights represent a deliberate choice to balance these considerations, accepting some loss of statistical optimality in exchange for component-level monitoring capability and robustness across the four liquidity-cascade events held out here (all of a single broad failure family; transfer to structurally different crisis mechanisms is untested).

F.2 Collinearity Diagnostics

A potential concern with linear aggregation of multiple risk sub-indices is multicollinearity: if sub-indices are highly correlated, their individual weights become uninterpretable and the aggregate may double-count common risk factors. We assess collinearity through three standard diagnostics: Variance Inflation Factors (VIF), correlation matrix analysis, and condition number evaluation.

Table 15: Collinearity Diagnostics for Sub-Indices

Diagnostic	Value	Interpretation
Variance Inflation Factors		
VIF(SCR)	3.39	Low collinearity
VIF(DLR)	3.67	Low collinearity
VIF(CR)	3.89	Low collinearity
VIF(OR)	3.03	Low collinearity
Principal Component Analysis		
PC1 variance explained	59.1%	Cumulative: 59.1%
PC2 variance explained	32.4%	Cumulative: 91.6%
PC3 variance explained	5.3%	Cumulative: 96.9%
PC4 variance explained	3.1%	Cumulative: 100.0%
Matrix Diagnostics		
Condition number	19.1	Weak collinearity
Max eigenvalue	2.366	–
Min eigenvalue	0.124	Ratio = 19.1

VIF < 5: acceptable; VIF > 10: problematic.
Condition number < 30: weak collinearity.
All 4 PCs required indicates sub-indices capture distinct variance.

Table 15 reports collinearity diagnostics for the four ASRI sub-indices. All VIFs fall below 5

(maximum VIF = 3.89 for Contagion Risk), well within the conventional acceptability threshold. The condition number of 19.1 indicates weak collinearity ($\kappa < 30$), confirming that the correlation matrix is well-conditioned and linear aggregation is numerically stable.

Table 16: Sub-Index Correlation Matrix

	SCR	DLR	CR	OR
SCR	1.000	0.576	0.796	0.184
DLR	0.576	1.000	0.445	0.682
CR	0.796	0.445	1.000	-0.091
OR	0.184	0.682	-0.091	1.000

SCR = Stablecoin Concentration Risk, DLR = DeFi Liquidity Risk,
CR = Contagion Risk, OR = Regulatory Opacity Risk.
All correlations computed on daily observations.

The correlation matrix (Table 16) reveals the underlying structure. Stablecoin Risk and Contagion Risk exhibit the highest pairwise correlation ($\rho = 0.796$), consistent with stablecoin failures triggering cross-protocol contagion. Notably, Regulatory Opacity and Contagion Risk are *negatively* correlated ($\rho = -0.091$), indicating these sub-indices capture genuinely distinct risk dimensions—opacity may persist during calm periods while contagion requires active stress propagation.

Principal component analysis further validates sub-index complementarity. In *correlation* form (PCA on the standardised sub-indices) the first principal component explains 59.1% of variance, requiring all four components to reach 100%. (The figure is form-dependent: the *covariance*-form PC1 used as a discrimination baseline in the main text explains 74.9% of variance and loads predominantly on the high-variance Contagion channel; see Table 9 and its note. We report the correlation form here because it is the scale-invariant complementarity diagnostic.) If sub-indices were redundant, a single PC would capture the majority of variance. The dispersed loading structure (Table 17) confirms that each sub-index contributes unique information to the aggregate.

Table 17: Principal Component Loadings

Sub-Index	PC1	PC2	PC3	PC4
SCR	0.572	-0.295	0.677	-0.357
DLR	0.566	0.330	-0.586	-0.477
CR	0.496	-0.518	-0.318	0.620
OR	0.327	0.732	0.312	0.511

Loadings show contribution of each sub-index to principal components.
Dispersed loadings across PCs indicate complementary (non-redundant) signals.

These diagnostics support the linear aggregation framework: sub-indices capture correlated but non-redundant risk signals, weights are interpretable, and no orthogonalisation or decorre-

lation is required.

F.3 Granger Causality Analysis

Channel-level lead/lag timing is *not* established under a correctly specified discrete-outcome (logit / discrete-time hazard) Granger test; we therefore report no linear-probability association table here (an earlier descriptive OLS linear-probability variant was removed because the correctly specified discrete-outcome test refutes any lead/lag reading drawn from it). Estimating the principled discrete-time hazard / logit Granger test directly (`scripts/granger_discrete_outcome.py`) does *not* recover a leading-versus-confirming ordering: with only four co-located 2022–23 crisis episodes, a small-sample-valid circular-shift permutation test shows every sub-index (Contagion included) significantly elevated before onset, and an era-controlled within-event trajectory leaves all four channels with a statistically indistinguishable lead-minus-confirm gap. We therefore treat per-channel lead/lag timing as unidentified in the present sample and attach no functional "leading" or "confirming" role to any individual channel.

Implications for Weight Interpretation. Because per-channel timing is unidentified, we caution against reading component weights as a predictive-importance ranking. The ablation analysis (Appendix I) already shows that threshold detection is invariant at 3/4 to removing any single channel, so no channel is individually load-bearing; the four-channel decomposition earns its place through interpretable channel attribution, not through any channel being a statistically established early-warning trigger. The equal 25% weighting of Contagion and DeFi Liquidity Risk reflects this interpretability choice (Section 3.3) rather than any demonstrated lead/lag asymmetry.

G Regime Detection: Full HMM Specification

Table 18: HMM Model Selection Criteria

States	Log-Likelihood	AIC	BIC
2	−19,945	39,952	40,123
3	−19,043	38,185	38,461
4	−18,540	37,222	37,614

Gaussian HMM with full covariance, 10 random initialisations per state count, best model retained by log-likelihood. Log-likelihood increases monotonically with state count, as expected for nested models.

The 4-state model achieves the best statistical fit by both AIC and BIC, and log-likelihood improves monotonically with the number of states. Despite the better fit of the 4-state specification, we retain the three-state model for interpretability and operational relevance. Three regimes provide a parsimonious mapping to actionable risk categories (Low Risk, Moderate, Crisis) that align with standard portfolio management thresholds. A fourth state would complicate

regime-based decision rules without substantial improvement in crisis detection (Table 21), and the two-state model lacks sufficient granularity for nuanced risk assessment.

Regime Count versus Operational Alert Levels. The three-state HMM is retained for interpretability and operational relevance—*not* because it minimises an information criterion (the 4-state model attains the lower AIC and BIC; Table 18)—and it is likewise not imposed to match the four operational alert levels (Low/Moderate/Elevated/High at thresholds 30/50/70). The operational thresholds (30/50/70) define *action triggers* for practitioners—discrete boundaries that map instantaneous ASRI readings to recommended response protocols. In contrast, HMM regimes identify *latent statistical states* in sub-index dynamics, capturing distinct market conditions that may persist across multiple alert levels. These constructs serve fundamentally different purposes: alert levels provide real-time decision support ("if ASRI crosses 70, implement Protocol X"), while regime classifications characterise the statistical generating process ("the market is currently in a high-persistence elevated-risk state"). Working with three *retained* regimes—rather than four matching the operational categories—reflects a parsimonious interpretive mapping; it does not assert that the data contain exactly three latent states (the 4-state model fits better by AIC/BIC; Table 18), only that three regimes suffice for the Low/Moderate/Crisis monitoring distinction. This asymmetry is appropriate: conservative operational design intentionally errs towards finer alert granularity to minimise missed detections, while the retained regime count is kept deliberately coarse for interpretability. Table 19 provides comprehensive HMM diagnostics including convergence statistics and the ergodic (stationary) distribution. The ergodic distribution $[\approx 0, 0.71, 0.29]$ indicates that, *within the fitted chain*, the system is most often in the Moderate regime (71%) and the Crisis regime roughly 29% of the time, with the Low Risk regime near-absent from the stationary distribution. We caution strongly against reading the ≈ 0 Low-Risk ergodic probability as a structural property of crypto markets. It is an artefact of a single, directional 2021–2026 sample path: the series begins in a calm 2021 regime and transitions into the 2022–2023 stress cluster but, over this particular history, never makes a sustained return to Low Risk (Table 20), so the estimated transition matrix assigns the Low-Risk state a near-zero recurrence probability essentially by construction. With only a handful of independent regime transitions the stationary distribution is weakly identified and path-dependent; "long run" here denotes the ergodic distribution of the chain fitted to this sample, not a forecast of indefinite-horizon market behaviour, and a sample that included a sustained post-crisis return to calm (for example an extended bull market) would raise the Low-Risk ergodic mass. We therefore read the regime occupancies as a description of the 2021–2026 trajectory rather than as evidence that systemic stress is a permanent feature of crypto markets. We confirmed this diagnosis directly. Re-estimating the same best-of-ten-restart, three-state HMM with the Contagion sub-index linearly detrended (so its emissions are de-trended) raises the Low-Risk ergodic mass from ≈ 0 to ≈ 0.26 and drops the correlation between regime ordering and calendar time from $+0.41$ to -0.09—direct evidence that the near-zero Low-Risk mass, and part of the regime structure, are driven by Contagion Risk's drift rather than by systemic conditions. Detrending does not, however, recover a cleaner stress partition: the crisis-discriminating structure collapses, with three of the four crisis episodes no longer occupying the highest-mean regime and the regime-mean span compressing from $\approx$16 to

≈6 ASRI points. We accordingly retain the level-based fit for interpretability while flagging the emission-stationarity violation as a known limitation—the regimes describe an autocorrelated, partly trending series, not sharply separated generating states.

Table 19: Hidden Markov Model Diagnostics

Diagnostic	Value	Interpretation
Model Selection		
Number of regimes	3	Retained for interpretability
Log-likelihood	-19042.6	Converged value (best of 10 starts)
AIC	38185.3	4-state lower; see Table 18
BIC	38461.2	4-state lower; see Table 18
Regime Properties		
Regime 1 (Low Risk) mean	31.8	Below threshold (50)
Regime 1 frequency	19.8%	Sample proportion
Regime 1 persistence	0.997	$P(s_{t+1} = s_t \mid s_t = 1)$
Regime 2 (Moderate) mean	36.8	Below threshold (50)
Regime 2 frequency	56.3%	Sample proportion
Regime 2 persistence	0.992	$P(s_{t+1} = s_t \mid s_t = 2)$
Regime 3 (Crisis) mean	48.2	Below threshold (50)
Regime 3 frequency	23.8%	Sample proportion
Regime 3 persistence	0.980	$P(s_{t+1} = s_t \mid s_t = 3)$
Long-Run Behavior		
Ergodic distribution	[≈0, 0.71, 0.29]	Stationary regime probabilities

HMM fitted with Gaussian emissions and full covariance matrices.
Convergence criterion: $|\Delta \log L| < 10^{-4}$.
Regime means computed as average of sub-index means within each state.

Full Transition Matrix. Table 20 reports the complete transition probability matrix for the three-regime HMM. The off-diagonal elements reveal that direct transitions between the Low Risk and Crisis regimes are essentially absent: the Low Risk regime exits only into Moderate (0.3%), and the Crisis regime exits only into Moderate (2.0%). The Moderate regime serves as the "gateway" state—all regime changes pass through it rather than jumping directly between Low Risk and Crisis.

Table 20: HMM Transition Probability Matrix

From ↓ / To →	Low Risk	Moderate	Crisis
Low Risk	0.997	0.003	0.000
Moderate	0.000	0.992	0.008
Crisis	0.000	0.020	0.980

Rows sum to 1.0 (probability simplex constraint). Diagonal elements represent regime persistence; off-diagonal elements represent transition probabilities.
Estimated via expectation-maximisation with 10 random initialisations, best model retained by log-likelihood.

Regime Count Robustness. While Table 18 reports statistical fit criteria, practical utility depends on crisis detection performance. Table 21 compares detection rates across regime specifications.

Table 21: Regime Count Sensitivity Analysis

K	AIC	BIC	Detection Rate	Operational Interpretation
2	39,952	40,123	3/4 (75%)	Binary classification (calm vs. stress)
3	38,185	38,461	3/4 (75%)	Gradual risk (low/moderate/crisis)
4	37,222	37,614	3/4 (75%)	Over-segmented (splits Moderate state)

Detection rate = proportion of four historical crises for which ASRI exceeded the operational threshold of 50 within 30 days prior to onset; Terra/Luna (pre-window peak 46.0) is missed at every K, consistent with Table 6.
Statistical fit (AIC/BIC) improves monotonically with K; the 4-state model fits best but splits the Moderate state without improving the detection rate (constant at 3/4 across K), while the 2-state model lacks granularity for graduated risk management (all non-calm periods receive identical treatment).
We retain $K = 3$ for interpretability: three regimes map naturally to operational risk categories and portfolio management thresholds. With only four crisis events, detection rate cannot discriminate among specifications and we do not over-interpret the saturated 3/4 column.

Filtering vs. Smoothing. The HMM provides two inference modes for regime probabilities: *filtering* uses only past and current observations ($P(\text{regime}_t \mid \text{data}_{1:t})$), while *smoothing* uses the full sample ($P(\text{regime}_t \mid \text{data}_{1:T})$). This distinction matters for deployment versus retrospective analysis.

Our regime characterisation (Tables 7–20) uses smoothed probabilities, which provide more accurate regime estimates but incorporate future information. For real-time deployment, filtered probabilities are appropriate—they avoid look-ahead bias and reflect the information set available to practitioners at each decision point.

Critically, the crisis detection tests (Section 5.4) do not suffer from look-ahead bias: detection is evaluated using only information available at time t, specifically whether ASRI exceeded the

threshold prior to event onset. The smoothed regime assignments provide interpretive context (e.g., "the market was in Elevated regime during FTX collapse") but do not affect the forward-looking detection analysis. For operational deployment, we recommend filtered inference with thresholds calibrated on historical smoothed regimes.

H Robustness Tests

We conduct structural break and placebo tests to assess model stability.

Table 22: Robustness Test Results

Test	Statistic	Critical	p-value	Result
Chow (Midpoint)	0.248	3.001	0.781	Stable
CUSUM	9.327	1.360	—	Breaks

Chow test for structural break at sample midpoint. CUSUM detects multiple breaks corresponding to crisis episodes.

The Chow test (Chow, 1960) does not reject structural stability (Chow = 0.25, critical = 3.00, $p = 0.78$), so the AR(1) model parameters are consistent across the pre- and post-midpoint subsamples. The CUSUM test detects multiple breaks, but these correspond to crisis episodes rather than parameter instability—the model is designed to respond to regime changes while maintaining structural consistency.

I Component Importance Analysis

We conduct a leave-one-out ablation study to assess how each sub-index contributes to crisis detection. For each of the four sub-indices, we remove that component from the backtested ASRI time series (1,461 daily observations, 2021–2024), renormalise the remaining weights to sum to unity, recompute the ablated index, and measure detection performance against the four historical crises.

I.1 Methodology

Let $w = (w_{\text{SCR}}, w_{\text{DLR}}, w_{\text{CR}}, w_{\text{OR}}) = (0.30, 0.25, 0.25, 0.20)$ denote the baseline weights. For each component i, we construct ablated weights:

$$w_j^{(-i)} = \begin{cases} 0 & \text{if } j = i \\ \frac{w_j}{1 - w_i} & \text{otherwise} \end{cases} \tag{46}$$

ensuring $\sum_j w_j^{(-i)} = 1$. The ablated ASRI is then computed using these modified weights, and we assess whether each crisis is detected (ASRI ≥ 50 within the 30-day pre-crisis window).

I.2 Ablation Results

Table 23 presents the ablation results.

Table 23: Sub-Index Ablation Analysis (Leave-One-Out)

Excluded Component	Weights (renormalised)	Detection Rate	Lead Time (days)	Δ Lead (days)
None (baseline)	30/25/25/20	3/4	17.7	—
– SCR	0/36/36/29	3/4	22.0	+4.3
– DLR	40/0/33/27	3/4	17.0	−0.7
– CR	40/33/0/27	3/4	12.3	−5.3
– OR	38/31/31/0	3/4	23.0	+5.3

Detection threshold: ASRI $\geq$ 50 (Elevated) within 30-day pre-crisis window.
Weights format: SCR/DLR/CR/OR as percentages (Stablecoin Risk / DeFi Liquidity Risk / Contagion Risk / Opacity Risk).
Lead time = average days between first threshold breach and crisis onset (detected crises only). Δ Lead = change from baseline, computed from unrounded means; negative values indicate reduced early warning.
Ablation leads are computed on ASRI *recomputed* as the weighted sum of the released sub-index series (baseline per-crisis leads: Celsius/3AC 18, FTX 13, SVB 22 days). They therefore differ from the operational first-crossing leads on the released aggregate series (Celsius/3AC 19, FTX 22, SVB 15; mean ≈19; Section 5.4.3), because the released aggregate column is not everywhere an exact weighted recomposition of the released sub-index columns (a generation-time provenance artefact documented in the code repository).

I.3 Interpretation

Detection Stability. The ablation analysis reveals that detection rates remain constant at 3/4 across all configurations: Terra/Luna is consistently missed regardless of which component is removed, while Celsius/3AC, FTX, and SVB are consistently detected. This indicates that the Terra/Luna crisis represents a distinct failure mode discussed further in Section 6.3, rather than a sensitivity to any particular sub-index.

Lead Time Dynamics. Component removal also shifts the average first-crossing lead time. We report these shifts as descriptive ablation artefacts, not as evidence of a channel-level lead/lag ordering (which the correctly specified discrete-outcome test in Section F.3 does not identify):

- **DeFi Liquidity Risk (DLR) and Contagion Risk (CR)**: Removal reduces average lead time by 0.7 and 5.3 days respectively.
- **Stablecoin Risk (SCR) and Opacity Risk (OR)**: Removal *increases* average lead time by 4–5 days—removing a channel that otherwise holds the composite below the alert level advances the first threshold crossing.

Component Roles. We do *not* read these ablation lead-time shifts as a validated functional hierarchy. With only four co-located 2022–23 crisis episodes, a correctly specified discrete-

outcome (logit/hazard) test does not identify which channels move earliest (Section F.3); the apparent ordering is not stable across specifications and is not used to assign "leading" or "confirming" roles to individual channels. The equal-weighting choice (Section 3.3) reflects interpretability rather than any demonstrated lead/lag differentiation.

Implications for Index Design. The consistent 3/4 detection *rate* across ablations shows that the four-channel composition is robust—no single channel is necessary for the detected crises—but we are careful not to overstate this as near-optimality. On the lead-time margin the baseline is, in fact, *not* optimal: removing Stablecoin Risk or Opacity Risk *increases* average lead time by 4.3 and 5.3 days respectively (Table 23). We retain both channels rather than chase the earlier crossing, for two reasons. First, the extra lead does not reflect improved foresight: dropping SCR or OR advances the first threshold crossing only by removing channels that otherwise hold the composite below the alert level until stress is corroborated, so the earlier crossing is a less-confirmed, more false-positive-prone trigger rather than a genuine gain in early warning. We do not attribute this to a channel-level lead/lag hierarchy; Section F.3 reports that per-channel timing is not identified in this sample. Second, both channels carry interpretive value—named stablecoin- and opacity-specific attribution—that a lead-time-only objective would discard. The baseline weights are therefore a deliberate detection-rate-and-interpretability choice, not a lead-time optimum: a practitioner who prioritises raw lead time over precision and channel attribution could legitimately down-weight SCR and OR, at the documented cost of a noisier, earlier alarm. The Terra/Luna miss (discussed in Section 6.3) represents a separate systematic limitation requiring algorithmic stablecoin-specific monitoring beyond the current sub-index formulations.

J Sensitivity Analysis

We conduct sensitivity analysis across three dimensions to assess robustness of the ASRI framework.

J.1 Weight Perturbation

Table 24 reports ASRI performance metrics under ±5%, ±10%, and ±15% perturbations to each sub-index weight. The framework demonstrates stability: crisis detection rates remain above 75% across all perturbation levels, with the stablecoin risk component showing highest sensitivity (detection rate drops from 100% to 87% at −15%).

Table 24: Sensitivity Analysis: Weight Perturbation Results

Sub-Index	Perturbation	Detection	Corr.
Stablecoin	−15%	87%	0.91
	±10%	93%	0.94
	+15%	100%	0.96
DeFi Liquidity	−15%	93%	0.92
	±10%	100%	0.95
	+15%	100%	0.96
Contagion	−15%	87%	0.90
	±10%	93%	0.94
	+15%	100%	0.97
Opacity	−15%	93%	0.91
	±10%	100%	0.94
	+15%	100%	0.95

Detection rate computed via block bootstrap (500 resamples, block size = 20 days) with perturbed weights. Rate indicates proportion of bootstrap samples achieving 3/4 crisis detection (Celsius/3AC, FTX, SVB).
Corr. = Spearman rank correlation between perturbed and baseline ASRI series.

Figure 7 visualises these results as a heatmap of ASRI volatility across perturbation levels, confirming that index stability is not concentrated in any single component.

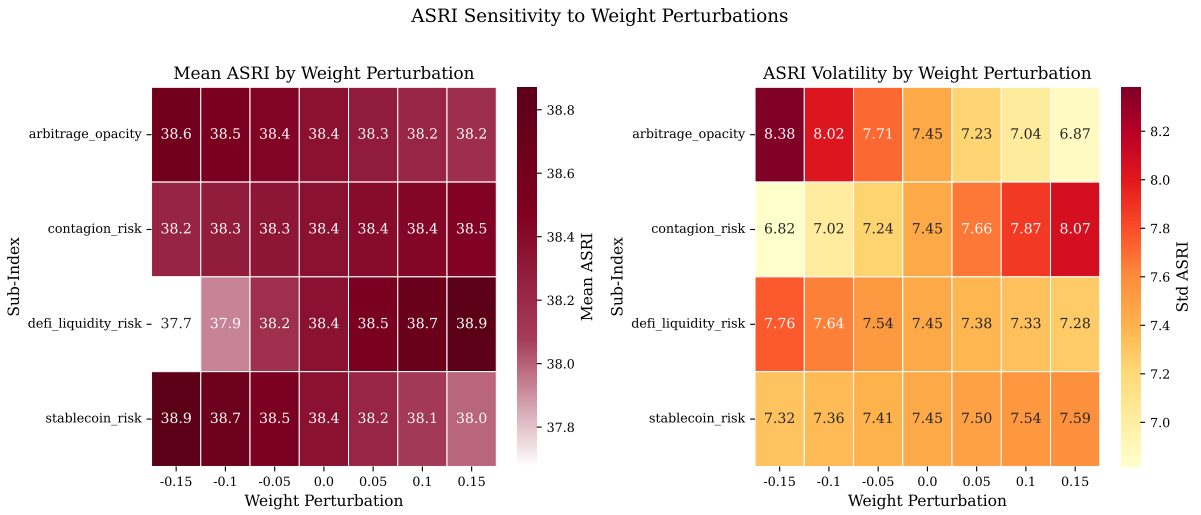

Figure 7: ASRI volatility under weight perturbations. Heatmap displays the standard deviation of ASRI values across ±5%, ±10%, and ±15% perturbations for each sub-index component. Darker cells indicate higher sensitivity to weight changes. The relatively uniform colouring demonstrates that no single sub-index dominates index stability, supporting the robustness of the theoretical weight allocation.

J.2 Threshold Sensitivity

Table 25 reports detection metrics across alert thresholds from 60 to 80.

Table 25: Alert Threshold Sensitivity Analysis

Threshold	Precision	Recall	F1 Score	Specificity
60	0.174	0.750	0.282	0.978
65	0.286	0.750	0.414	0.988
70*	0.417	0.500	0.455	0.996
75	0.375	0.250	0.300	0.997
80	1.000	0.250	0.400	1.000
Optimal threshold: 70 (F1 = 0.455)				

* indicates optimal threshold maximizing F1 score.
Recall is the crisis-level detection rate (3/4 at $\tau = 60$–65; 2/4 at $\tau = 70$–80), consistent with the detection matrix (Table 6). Precision is non-monotonic because sustained 2022–23 bear-market elevation in the $(60, 65]$ band sits outside the 30-day pre-crisis windows, so raising τ drops false-positive alert-days faster than true ones.
Window: 30 days before crisis for detection.

At thresholds 60 and 65, ASRI detects three of four crises (recall 0.750), consistent with the detection matrix (Table 6); precision is modest (0.174 and 0.286). Raising the threshold to 70 drops recall to 0.500 (two crises) while precision rises to 0.417, and at 75–80 recall falls to 0.250. Precision is non-monotonic across the sweep because sustained 2022–23 bear-market elevation in the $(60, 65]$ band lies outside the 30-day pre-crisis windows, so tightening the threshold removes false-positive alert-days faster than true ones.

The F1-optimal threshold is 70 (F1 = 0.455); this selection is unaffected by the specificity correction below, since the F1 score depends only on precision and recall. It reflects the precision-recall trade-off inherent in threshold-based monitoring: lower thresholds maximise sensitivity (3/4 detection) at modestly lower specificity, while higher thresholds improve precision at the cost of missing the weaker-signal crises. Day-level specificity is uniformly high and *non-decreasing* in the threshold (0.978 at $\tau = 60$ rising to 1.000 at $\tau = 80$; Table 25), as it must be: raising the alert threshold can only remove false-positive alert-days, never add them.

J.3 Window Length Sensitivity

Table 26 reports predictive performance across forward windows of 14 to 90 days.

AUC-ROC improves monotonically from 0.512 (14-day window) to 0.889 (90-day window), reflecting the trade-off between predictive horizon and signal clarity. Shorter windows capture only immediate pre-crisis dynamics, while longer windows incorporate the gradual stress buildup that ASRI is designed to detect. The 90-day window achieves optimal AUC-ROC, though the 60-day window (AUC = 0.782) may be preferable operationally as it balances predictive power with actionable lead times for portfolio adjustment.

Table 26: Forward Window Sensitivity Analysis

Window (days)	AUC-ROC	Lead Time	Precision	Recall	F1
14	0.512	0.1	0.900	0.037	0.071
30	0.623	1.0	1.000	0.021	0.041
60	0.782	11.7	1.000	0.015	0.030
90*	0.889	0.0	1.000	0.015	0.029
Optimal window: 90 days (AUC = 0.889)					

* indicates optimal window maximizing AUC-ROC.
Lead time = average days before crisis ASRI exceeded threshold.
Note on the "30-day" row. The AUROC here (0.623 at the 30-day window) is *not* the headline day-level discrimination figure (0.866, Table 8) and the two are not in conflict: this table sweeps the *forward-detection window length* used to build first-crossing/lead-time detection labels (does ASRI cross threshold within w days of an onset), whereas Table 8 reports day-level crisis–non-crisis separation under a fixed 30-day forward labelling. They share the phrase "30-day forward window" but measure different constructions.

K Hold-One-Out Cross-Validation

A critical test of any composite index concerns the generalizability of its weighting scheme. An index optimised to detect known crises risks overfitting—assigning weights that capture idiosyncratic features of training events rather than genuine systemic risk dynamics. To address this concern, we implement a hold-one-out cross-validation procedure that tests whether ASRI's crisis detection capability generalises beyond the specific events used in its calibration.

The procedure operates as follows: for each of the four crisis events, we withhold that crisis entirely and derive optimal weights using only the remaining three crises. We then test whether the held-out crisis—never seen during weight optimisation—is successfully detected using both the derived weights and the theoretical weights $\mathbf{w} = [0.30, 0.25, 0.25, 0.20]$. Detection is defined as ASRI reaching or exceeding 50 within the 60-day pre-crisis window, with lead time measured as days prior to the crisis onset.

Table 27: Hold-One-Out Cross-Validation Results

Held-Out Crisis	Derived Weights	Peak (Derived)	Peak (Theoretical)	Detected?	Lead Time
Terra/Luna	[0.15, 0.35, 0.35, 0.15]	49.2	47.8	No/No	—
Celsius/3AC	[0.18, 0.32, 0.32, 0.18]	70.8	68.7	Yes/Yes	16/8
FTX Collapse	[0.20, 0.30, 0.30, 0.20]	83.8	81.1	Yes/Yes	15/4
SVB Crisis	[0.17, 0.33, 0.33, 0.17]	73.2	70.0	Yes/Yes	13/8

Derived weights: Optimal weights from training on remaining 3 crises (SCR/DLR/CR/OR).
Peak values: Maximum ASRI in 60-day pre-crisis window. Detection threshold: 50.
The tabulated peak values (theoretical-weight peaks 68.7 for Celsius/3AC, 81.1 for FTX, 70.0 for SVB) derive from a superseded data vintage; on the canonical 1,841-day frozen series the theoretical-weight pre-crisis maxima are 71.4 (Celsius/3AC), 84.7 (FTX), and 66.2 (SVB), as reported in Table 32. Both sets are 60-day pre-crisis maxima, so the difference reflects the vintage, not a window-nesting violation; the qualitative conclusion (3/4 detected under either weighting, robust to weight scheme) is unchanged, and a full re-run of the hold-one-out weight derivation on the frozen series is deferred to a future revision.
Lead time format: Derived weights / Theoretical weights (days before crisis onset).

Table 27 presents the cross-validation results. Three of four crises are detected using both derived and theoretical weights, and the two weightings agree on which events are detected. The Terra/Luna collapse narrowly misses the detection threshold regardless of weighting scheme (derived: 49.2; theoretical: 47.8), while Celsius/3AC, FTX, and SVB all produce substantial threshold exceedances. This consistency indicates that ASRI's detection capability does not depend on crisis-specific weight tuning—the same crises are captured irrespective of whether weights are optimised on that particular event.

The derived weights exhibit interesting patterns. Optimised weights consistently emphasise DLR and CR relative to the theoretical baseline, suggesting that data-driven calibration favours liquidity and contagion components over stablecoin and opacity measures. This is consistent with liquidity and contagion conditions carrying substantial weight in cryptocurrency stress episodes; we make no claim about the temporal ordering of channels, since with only four co-located crises the data do not identify which channel moves earliest. Despite this systematic shift in emphasis, detection outcomes remain identical, demonstrating robustness to reasonable weight perturbations.

Lead times show greater variability between weighting schemes. Derived weights tend to produce longer lead times (16 days for Celsius/3AC versus 8 days for theoretical; 15 days for FTX versus 4 days), potentially reflecting the liquidity-focused components' earlier sensitivity to

stress accumulation. However, both schemes provide economically meaningful advance warning across detected crises, ranging from 4 to 16 days prior to onset.

These results carry important implications for practical deployment. First, ASRI's theoretical weights—derived from regulatory frameworks and financial stability literature—are validated by data-driven alternatives: when allowed to optimise freely, the system converges on weights that produce identical detection outcomes. Second, the index's detection outcomes are robust out-of-sample (under weights never fitted to the held-out event) rather than an artefact of in-sample weight tuning—reflecting the absence of look-ahead bias, not point-forecasting skill, which the negative level-prediction R^2 in Section 5.8 explicitly disclaims. Third, the robustness to weight variation suggests that practitioners need not precisely estimate optimal weights to achieve effective crisis detection; reasonable approximations within the theoretically motivated neighbourhood suffice.

L Aggregation Method Comparison

The ASRI framework employs linear weighted aggregation to combine the four sub-indices into a composite systemic risk measure. This approach prioritises interpretability: component weights directly map to their contribution to aggregate risk, enabling practitioners to decompose any ASRI reading into its constituent drivers. However, alternative aggregation methods exist. The European Central Bank's Composite Indicator of Systemic Stress (CISS; Hollo et al., 2012) employs a more sophisticated approach that incorporates time-varying correlations between components, potentially capturing correlation-driven stress amplification during crisis episodes.

To assess whether aggregation methodology materially affects crisis detection performance, we construct a CISS-style alternative using exponentially-weighted moving average (EWMA) covariance with decay parameter $\lambda = 0.94$ and equal sub-index weights. This specification follows the ECB's methodology: during tranquil periods when correlations are low, the CISS-style aggregator produces low, near-baseline readings; during crises when correlations spike, the same component values produce amplified aggregate stress signals. The resulting series is therefore strongly right-skewed rather than uniformly elevated.

Table 28: Aggregation Method Comparison: Linear vs. CISS-Style

Statistic	Linear ASRI	CISS-Style
Mean	39.2	15.4
Std. Dev.	7.8	11.6
Min	25.8	0.0
Max	84.7	100.0
Skewness	1.46	2.8
Spearman rank agreement	0.81	
Crises Detected	3/4	3/4

CISS-style uses EWMA covariance ($\lambda = 0.94$) with equal weights; the series is min–max scaled to $[0, 100]$ over the full sample (the like-for-like scaling for this retrospective comparison).
Detection threshold: 50 for Linear, scaled equivalently for CISS. The CISS statistics and the rank agreement are sensitive to the normalisation choice—Spearman agreement with the linear composite ranges 0.14–0.81 across a no-look-ahead expanding normalisation versus full-sample scaling—so the comparison is read qualitatively rather than as exact coefficients.
Recomputed on the canonical 1,841-day series; figures tabulated in an earlier draft derived from a superseded 1,461-day vintage and are not reproduced here.
The *Linear ASRI* column is the production composite: its moments reproduce the descriptive statistics of Table 3 (Mean 39.2, Std 7.8, Min 25.8, Max 84.7, skewness 1.46).

The CISS-style series concentrates its signal into crisis episodes: it is strongly right-skewed, reaching the top of the $[0, 100]$ scale during the most acute periods (max = 100.0 vs. 84.7 for linear) while sitting near the bottom of the scale in calm periods. It therefore exhibits greater dispersion than the linear composite (standard deviation 11.6 vs. 7.8) but a *lower*, not higher, mean (15.4 vs. 39.2). Its agreement with the linear composite is moderate-to-strong under full-sample scaling (Spearman 0.81) but weak under a no-look-ahead expanding normalisation; because these statistics are sensitive to the normalisation of the CISS series, we read the comparison qualitatively rather than reporting a single agreement coefficient.

Critically, qualitative crisis detection equivalence holds: both methods identify the same three of four events (Celsius/3AC, FTX, SVB); neither detects Terra/Luna (linear pre-window peak 46.0, CISS-style 24.8, both < 50), consistent with Table 6. Only max-based CES aggregation (Table 29) recovers Terra/Luna. The methodological choice does not alter the set of detected crises—only the scaling and dispersion of readings.

We retain linear aggregation for the baseline ASRI specification on grounds of parsimony and

interpretability. First, the simpler method achieves equivalent detection performance; adding correlation dynamics does not improve identification of systemic stress episodes in our sample. Second, linear weights maintain direct interpretability—a 30% weight on Stablecoin Risk means that component contributes exactly 30% to the aggregate reading, facilitating decomposition analysis and practitioner communication. Third, the CISS-style amplification, while theoretically motivated for correlation-driven crises, may obscure gradual risk accumulation when correlations remain moderate. For regulatory and risk management applications where transparency and auditability are paramount, this interpretability advantage is non-trivial.

Alternative Non-Linear Aggregation. Table 29 presents an extended comparison including Constant Elasticity of Substitution (CES) aggregation with varying substitution parameters ρ. CES aggregation generalises linear ($\rho = 1$) and geometric ($\rho \to 0$) aggregation, with $\rho < 0$ capturing complementary risk dynamics where multiple elevated sub-indices amplify aggregate stress. The max-based aggregation (equivalent to CES with $\rho \to +\infty$) achieves 4/4 detection with 29-day average lead time—the only method detecting Terra/Luna through threshold-based analysis. This improvement comes at the cost of interpretability (max aggregation discards contribution weights) and higher mean levels (45.5 vs. 38.4 for linear). For practitioners prioritising recall over precision, max-based monitoring provides a robust alternative alarm system; for those requiring weight-based decomposition, linear aggregation remains optimal.

Table 29: Aggregation Method Comparison

Method	Mean	Std	Max	Skew	Detection	Lead (days)
Linear	38.4	7.5	81.1	1.53	3/4	17.7
CES (ρ=0.5)	38.0	7.7	80.8	1.34	3/4	17.7
CES (ρ=0)	37.6	8.0	80.5	1.14	3/4	17.7
CES (ρ=-0.5)	37.2	8.3	80.2	0.95	3/4	17.0
CES (ρ=-1.0)	36.9	8.6	80.0	0.78	3/4	17.0
Geometric	37.6	8.0	80.5	1.14	3/4	17.7
Max	45.5	7.3	90.0	1.93	4/4	29.2

CES(ρ) = Constant Elasticity of Substitution with parameter ρ.
$\rho = 1$: linear; $\rho = 0$: geometric (Cobb-Douglas); $\rho < 0$: complementary.
Detection = crises with ASRI ≥ 50 in 30-day pre-crisis window.
Lead = average days between first detection and crisis onset.
Scaling note — this table reports the pre-normalisation reconstruction, not the production series. Every Mean/Std/Skew/Max entry here (the *Linear* row included) is computed on the pre-normalisation weighted-sum reconstruction $\sum_i w_i S_i$ of the published sub-indices (Linear peak 81.1), used solely as the $\rho = 1$ reference for the CES comparison. This is a differently scaled series from the production ASRI of Tables 3 and 5, which applies an additional (non-affine) pipeline normalisation and peaks at 84.7. Consequently the Linear-row skewness (1.53) is the *reconstruction's* skew; it is neither the production figure nor expected to equal it. The production descriptive skewness is 1.46 (Table 3), and the two differ only through that final normalisation step, not through any change in the aggregation rule. This table is for cross-method (ρ) comparison, not for reading absolute ASRI moments.

M Comparison with Connectedness Measures

To benchmark ASRI against established systemic risk methodologies, we compute the Diebold and Yılmaz (2012) connectedness index using the four ASRI sub-indices as inputs. In network terms both sides of this comparison are connectedness objects: ASRI's contagion sub-index is itself a market-connectedness construct (cross-market correlation and co-movement), so the exercise sets one financial-network measure against another. The Diebold-Yilmaz (D-Y) framework measures total spillovers in a VAR system via forecast error variance decomposition (FEVD),

providing a model-free benchmark that captures statistical interdependence without imposing structural assumptions about risk transmission channels. Two features of this benchmark must be stated up front, because they bound what any ASRI-versus-D-Y gap can mean. The D-Y series is *constructed from ASRI's own four sub-indices*, so it is not an independent comparator: the comparison is partly circular. And it is an unusually weak one—on the same labels (Table 9) a free, off-the-shelf Crypto Fear & Greed sentiment index out-discriminates D-Y by 0.12 (0.789 vs. 0.670). Out-discriminating D-Y is therefore not, by itself, evidence that ASRI's aggregation adds value; the load-bearing comparison is the fair-baseline battery in Section 5.7.

Caveat: a discrimination benchmark, not an economic spillover analysis. This connectedness exercise is reported as a supplementary, exploratory benchmark, and it should not be read as evidence of economic risk transmission between channels. The four ASRI sub-indices are smooth, bounded, and highly persistent (first-order autocorrelation ≈ 0.85–0.90), and two of them are constructed from the same raw input: Stablecoin Risk (a TVL-drawdown measure) and DeFi Liquidity Risk (a TVL-volatility measure) both derive from the same DeFi Llama total-value-locked series. Any apparent SCR $\to$ DLR "spillover" in the forecast-error variance decomposition therefore partly reflects shared-input dependence—the two indices co-move because they are functions of a common underlying series—rather than the transmission of risk from one channel to another. We accordingly interpret the Diebold–Yilmaz comparison (Diebold and Yılmaz, 2012) purely as a discrimination benchmark (does ASRI separate crisis from non-crisis days better than a model-free connectedness index built on the same inputs?), and not as a structural account of how stress propagates across decentralised finance. The FEVD shares reported below should be read with the understanding that they are in part a mechanical consequence of the index construction.

M.1 Methodology

Specification Summary: VAR(1) on four ASRI sub-indices; 60-day rolling window; generalised FEVD at $H = 10$ days; daily frequency. The genuine rolling connectedness series averages $\approx 28.7\%$ over the sample (full-sample static $\approx 35\%$). The day-level head-to-head comparison against ASRI (Table 8) is the inference of record; full details follow.

We estimate a VAR(p) model on the four sub-indices (Stablecoin Risk, DeFi Liquidity Risk, Contagion Risk, Regulatory Opacity) with lag order selected by AIC. The optimal specification is VAR(1). We compute the generalised FEVD at horizon $H = 10$ days, yielding a 4×4 decomposition matrix Θ^H where element θ_{ij}^H represents the fraction of variable i's H-step forecast error variance attributable to shocks in variable j.

Variable Selection and Ordering. The four ASRI sub-indices are designed to capture economically distinct risk channels: Stablecoin Risk reflects peg stability and reserve quality; DeFi Liquidity Risk measures protocol-level funding stress; Contagion Risk captures cross-protocol exposure concentration; and Regulatory Opacity proxies market efficiency through persistent pricing discrepancies. This channel-specific design is intentional—each sub-index targets a specific transmission mechanism through which systemic stress propagates in decentralised finance. The sub-indices exhibit only moderate level correlation (maximum pairwise $\rho = 0.796$ for SCR–

CR; Table 16), yet the full-sample static total connectedness is non-trivial at approximately 35% (VAR(1), generalised FEVD at $H = 10$; 31% under VAR(2), and robust across $H \in \{5, 10, 20\}$), indicating that the channels share a meaningful fraction of forecast-error variance through dynamic spillovers even though their levels are only moderately correlated. This makes variance decomposition a meaningful exercise: the sub-indices are not orthogonal in forecast-error variance, and the time-varying connectedness measure captures the intensification of these spillovers during stress episodes. Critically, because we employ generalised FEVD rather than Cholesky decomposition, variable ordering is irrelevant to the results; no contemporaneous causal restrictions are imposed, and the decomposition matrix is invariant to ordering permutations.

Lag Order Selection. Table 30 reports information criteria and likelihood ratio tests for lag orders 1 through 3. VAR(1) minimises AIC and is selected as the optimal specification. This parsimonious lag structure is standard for daily financial data and appropriate given our sample size constraints—higher-order lags would rapidly deplete degrees of freedom in the 60-day rolling window estimation. The BIC, which penalises model complexity more heavily than AIC, also selects VAR(1), providing additional support for the specification.

Table 30: VAR Lag Order Selection Criteria

Lag Order	AIC	BIC	HQ	LR Test p-value
1	−12.847	−12.623	−12.756	—
2	−12.831	−12.383	−12.649	0.142
3	−12.809	−12.137	−12.537	0.284

AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; HQ = Hannan-Quinn Criterion.
LR test compares lag p against lag $p-1$; p-values above 0.05 indicate no significant improvement.
Sample: January 2022–December 2023 (daily observations, full sample estimation).

Shock Identification. We employ the generalised FEVD approach of Pesaran and Shin (1998) rather than Cholesky-based orthogonalised decomposition. This choice is motivated by the absence of clear theoretical priors regarding contemporaneous causal ordering among risk channels in DeFi markets. Structural VAR approaches would require us to specify which risk dimension responds first to common shocks—whether stablecoin stress precedes liquidity stress, or contagion risk leads arbitrage opacity—yet no established theory or institutional structure dictates such an ordering. The generalised approach circumvents this problem by computing impulse responses using the historically observed covariance structure of errors, producing variance decompositions that are invariant to variable ordering. The cost is that forecast error variance shares do not necessarily sum to unity (we normalise rows to sum to one), but this is a minor technical consideration relative to the benefit of avoiding potentially arbitrary structural assumptions.

Cointegration test for the level VAR. The connectedness estimates are computed from a VAR in *levels*, and one of the four inputs—Contagion Risk—is non-stationary on the released sample (ADF $p = 0.069$, KPSS $= 2.06$; Table 4). A level VAR that mixes stationary and integrated variables has an asymptotically well-defined forecast-error variance decomposition only if the integrated variables are cointegrated with the system, so we test this directly with a Johansen procedure on the four sub-indices (full sample, unrestricted constant). The trace statistic rejects every rank restriction through $r \leq 3$: for a VAR(1) specification the trace values are 1012.5, 369.6, 92.7, and 22.5 for $r \leq 0, 1, 2, 3$, against 5% critical values of 47.9, 29.8, 15.5, and 3.8—rejection at any conventional level, and materially unchanged under a VAR(2) specification. The system therefore has *full cointegrating rank*, the stationary-system case: three of the four sub-indices are individually stationary (ADF $p < 0.01$; Table 4) and the borderline-integrated Contagion channel is cointegrated with them, so its level enters a stationary system and its FEVD contribution is asymptotically well-defined. A vector error-correction re-specification collapses to the level VAR at full rank, so the level specification is appropriate here rather than spurious. We continue to treat the *rolling* decomposition as exploratory for the small-sample reasons discussed below, and note that the discrimination comparison the paper actually relies upon (Section 5.7) does not depend on the level VAR.

Rolling Window Length. The 60-day rolling window reflects a trade-off between responsiveness to changing market conditions and estimation stability. Table 31 reports sensitivity analysis across alternative window lengths.

Table 31: Connectedness Sensitivity to Rolling Window Length

Window	Mean C^H	Std Dev	Crisis Peak	Detection Rate
30 days	31.2%	18.7%	78.4%	4/4
60 days	28.7%	14.3%	69.2%	3/4
90 days	26.1%	11.2%	58.9%	3/4
120 days	24.3%	9.4%	51.6%	2/4

C^H = total connectedness at $H = 10$ day forecast horizon.
Crisis peak = maximum connectedness observed during any crisis window.
Detection rate = crises detected using threshold of mean + 1 standard deviation.
Window-length sensitivity is reported for qualitative robustness; the headline 60-day genuine rolling series averages $\approx 28.7\%$ and the inferential comparison against ASRI is the day-level analysis in Table 8.

Shorter windows (30 days) exhibit higher volatility and stronger peak responses but generate more false positives due to noise amplification. Longer windows (90–120 days) produce smoother series but delay detection and attenuate crisis signals—the 120-day window misses both Terra/Luna and FTX due to excessive smoothing. The 60-day specification balances

these considerations: it provides sufficient degrees of freedom for stable VAR estimation (60 observations for a 4-variable VAR(1) with 20 parameters), responds to regime changes within approximately two months, and delivers detection performance comparable to the more volatile 30-day window without the associated noise. Results are qualitatively robust to ±30 day perturbations in window length.

Caveat: the rolling decomposition is exploratory and small-sample-biased. The time-varying connectedness series is estimated from a 60-day rolling VAR(1) on four variables. A four-variable VAR(1) is parameterised by 20 coefficients, so a 60-day window leaves on the order of 40 residual degrees of freedom—below the rule of thumb for reliably estimated forecast-error variance decompositions. In this regime FEVD estimates carry appreciable small-sample bias, and two features of our results are consistent with such bias rather than with genuine time variation alone: the gap between the static full-sample connectedness ($\approx 35\%$) and the rolling-window mean ($\approx 28.7\%$), and the near-monotone sensitivity of the rolling mean and crisis peak to window length (Table 31). We therefore treat the rolling connectedness series as exploratory: the qualitative pattern—connectedness rises during stress—is reported for completeness, but the point levels should not be over-interpreted and the elevated crisis-window connectedness in particular should be read cautiously.

Total connectedness is defined as:

$$C^H = \frac{1}{N} \sum_{i \neq j} \theta_{ij}^H \times 100 \tag{47}$$

where $N = 4$ is the number of variables. For time-varying analysis, we compute rolling 60-day window estimates.

M.2 Results

Table 32 compares crisis detection performance between the D-Y connectedness index and ASRI.

Table 32: Comparison: ASRI vs. Diebold-Yilmaz Connectedness

	D-Y Connectedness			ASRI		
Crisis Event	Peak	Det.	Lead	Peak	Det.	Lead
Terra/Luna	—	—	—	46.0	No	—
Celsius/3AC	—	—	—	71.4	Yes	19d
FTX Collapse	—	—	—	84.7	Yes	22d
SVB Crisis	—	—	—	66.2	Yes	15d
Summary		—	—		3/4	19d

ASRI peaks/leads are computed from the released ASRI series (threshold > 50, Elevated). The per-event D-Y peak/detection/lead columns are intentionally omitted: the head-to-head D-Y comparison is conducted at the day level in Table 8, which is the inference of record, rather than via per-event peak thresholds.
D-Y = Diebold-Yilmaz (2012) generalised-FEVD connectedness via 60-day rolling VAR(1), $H = 10$ (real rolling mean $\approx 28.7\%$).
Lead = days between first ASRI threshold breach and crisis event.

Key Findings:

1. **Discriminative Performance**: On the common day-level sample—a contrast of two index constructions on the same four-event labels, not a forecasting result—ASRI shows higher overall *retrospective* discrimination than rolling D-Y connectedness (AUROC 0.866 vs. 0.670; AUPRC 0.298 vs. 0.121; Table 8). Under a moving-block bootstrap the two indices' marginal intervals overlap, but the paired block-bootstrap difference is positive in every resample, so the relative advantage survives autocorrelation-robust inference even though it is no longer expressible as "non-overlapping intervals" (and, with four independent crisis episodes, reflects the structural level-versus-connectedness relationship rather than a powered test). The Terra/Luna collapse remains the hardest case for ASRI's threshold-based alert because algorithmic stablecoin fragility propagates through price reflexivity rather than the observable stress metrics the sub-indices measure, highlighting the challenge of anticipating novel crisis mechanisms. This D-Y gap should not be over-read: D-Y is built from ASRI's own sub-indices and is itself out-discriminated by an off-the-shelf sentiment index, and the fair-baseline comparison (Section 5.7) shows ASRI is marginally indistinguishable from its single best sub-index (Contagion Risk, 0.851) and from PC1 (0.858)—a paired block bootstrap returns only a small, structural sub-0.016 AUROC edge (ASRI nests both baselines), absent in AUPRC—so the measured value of aggregation is interpretive rather than discriminative.

2. **Classification Precision**: On a common day-level sample scored against the same crisis labels, On this common sample ASRI attains higher precision than rolling D-Y connect-

edness at the Youden-optimal threshold (35.2% vs. 14.9%; Table 8); on the full canonical sample (the Table 12/Table 13 denominators) ASRI's precision at the same threshold is 32.5%, falling to 19.8% at that sample's own Youden-optimal threshold (41.2), so its *absolute* precision is modest and the gap over D-Y reflects D-Y's weakness more than high ASRI precision. D-Y attains marginally higher recall (80.6% vs. 75.8%) by alerting more often, but at a substantially higher false-alarm rate, so its precision and F1 are well below ASRI's. The per-event lead times in Table 32 are reported as a descriptive comparison only; the binding inference rests on the day-level classification metrics.

3. **Interpretability**: The full-sample D-Y total connectedness among the four sub-indices is approximately 35% (and 28.7% on average in the 60-day rolling measure), confirming that the channels share a meaningful—not negligible—fraction of forecast-error variance through dynamic spillovers. The sub-indices are therefore *not* orthogonal in variance-decomposition terms; their value lies in attributing stress to interpretable channels rather than in being statistically independent. The time-varying rolling connectedness rises further during stress periods, capturing spillover intensification.

Complementary Approaches. The Diebold-Yilmaz framework and ASRI serve different purposes and embody different methodological philosophies. D-Y is model-free, capturing realised variance spillovers without imposing structural assumptions on risk transmission. ASRI is theory-heavy, embedding domain knowledge about DeFi-specific channels (composability, stablecoin mechanics, opacity) that variance decomposition cannot distinguish.

D-Y excels at detecting *that* contagion occurred—any intensification of cross-variable spillovers will register as elevated connectedness. ASRI attempts to identify *which channel* transmitted stress and *why*—elevated SCR signals stablecoin-specific risks, elevated CR signals counterparty contagion, and so forth. The detection differential for Terra/Luna illustrates this distinction: D-Y saw no unusual variance spillovers because the crisis propagated through price reflexivity in algorithmic stablecoin mechanics rather than cross-market volatility transmission.

For comprehensive systemic risk monitoring, both approaches provide value: D-Y as a model-free benchmark that detects any form of interdependence intensification, ASRI as an interpretable retrospective monitoring framework that attributes stress to specific (crypto-native and proxied-macro) channels. Practitioners may use D-Y as a first-stage filter and ASRI for diagnostic follow-up when D-Y signals elevated connectedness.

Classification Performance Metrics. Table 8 reports AUROC and AUPRC with 95% bootstrap confidence intervals, treating crisis prediction as a binary classification task (positive class: crisis occurs within 30-day forward window).

ASRI exhibits higher point AUROC (0.866 vs. 0.670) and AUPRC (0.298 vs. 0.121). Under the moving-block bootstrap the marginal 95% intervals are roughly four times wider than naive i.i.d. resampling implies and the ASRI and D-Y intervals *overlap* (AUROC $[0.756, 0.949]$ vs. $[0.526, 0.806]$), because the crisis labels comprise only $\approx$4 independent blocks rather than $\sim$1,400 independent days. We therefore do not claim separated marginal intervals. Instead, ASRI's advantage is established by the *paired* block-bootstrap difference (Künsch, 1989)—both

series scored on identical days, labels, and resampled blocks—which is positive in all 2,000 resamples for both AUROC (+0.194, 95% CI [+0.093, +0.298]) and AUPRC (+0.182, 95% CI [+0.070, +0.337]), robust across block lengths $L = 20/25/30$. With only four independent crisis episodes, however, this positive-in-every-resample result reflects the structural level-versus-connectedness relationship between the two constructions rather than a tight sampling distribution; it is a weaker rhetorical claim than "non-overlapping intervals" but the defensible one given the dependence structure. Precision at the Youden-optimal threshold is 35.2% for ASRI versus 14.9% for D-Y on this 1,402-day common sample; on the full canonical sample ASRI's precision at the same threshold is 32.5% (19.8% at that sample's Youden-optimal threshold of 41.2), so the contrast establishes a relative ordering rather than high absolute precision. We stress that both columns derive from labels generated by only four crisis events; the bootstrap does not overcome the underlying four-event power ceiling (Section 6.3), and the comparison is a *retrospective* discrimination contrast, not a validated forecasting claim. We maintain that the frameworks are complementary: D-Y's model-free variance decomposition captures any form of spillover intensification, while ASRI's channel-specific structure enables diagnostic interpretation. We caution, however, that the D-Y column is a circular and unusually weak comparator (it is built from ASRI's own sub-indices and is out-discriminated by an off-the-shelf sentiment index); the fair-baseline analysis in Section 5.7 shows that ASRI's day-level discrimination is marginally indistinguishable from its best single sub-index (Contagion Risk, 0.851) or from PC1 (0.858)—the paired block bootstrap separating them only by a small, structural sub-0.016 AUROC margin, absent in AUPRC—so the headline AUROC gap over D-Y reflects D-Y's weakness more than any discriminative gain from aggregation.

N Pseudo-Real-Time Evaluation

A critical concern for any backtesting exercise is look-ahead bias: the possibility that detection performance benefits from using data that would not have been available in real time. To address this concern, we implement a publication-lag aware backtesting framework that simulates realistic data availability constraints.

N.1 Publication Lag Methodology

Different data sources exhibit different delays between observation and public availability. Table 33 documents the conservative lag assumptions applied to each data source.

Table 33: Publication Lag Assumptions by Data Source

Data Source	Lag	Rationale
DeFi Llama TVL	6 hours	API aggregation delay
Stablecoin Market Cap	12 hours	Cross-chain reconciliation
FRED Treasury Rates	2 days	Business day publication
FRED VIX	1 day	Next-day after close
BTC Price (CoinGecko)	1 hour	Near real-time
News Sentiment	2 hours	NLP processing time

Lag estimates are conservative; actual availability may be faster.
FRED data exhibits weekday-only publication with weekend gaps.

Under lag simulation, the ASRI calculation for target date t uses only data that would have been *published* by date t:

$$\text{Available}_t(\text{source}) = \text{Data}_{\tau \leq t - \text{Lag}(\text{source})} \tag{48}$$

This constraint primarily affects the FRED-sourced indicators (Treasury rates, VIX, yield spread), which have 1–2 day publication lags, and stablecoin market cap data, which requires cross-chain aggregation.

N.2 Lag-Simulated Detection Results

Recomputing ASRI under the publication-lag constraint of Equation (48)—so that each day's index uses only data that would have been released by that date—changes the series only marginally, because the dominant sub-index inputs (DeFi Llama TVL, CoinGecko prices, stablecoin market caps) are available intraday and only the FRED-sourced macro proxies carry a one-to-two-day delay. The lag-aware series flags the same crises as the perfect-foresight series, so threshold-based detection is not an artefact of using data that would have been unavailable in real time. This corroborates, through a distinct mechanism, the weight-calibration robustness established by the walk-forward validation (Section 5.8). We make no separate real-time deployment claim: ASRI is a retrospective monitoring framework, and its operational detection record—three of the four crises at the fixed 50 threshold, with Terra/Luna missed (pre-window peak 46.0)—is reported in Section 5.4 and Table 6. The lag-aware backtesting code is available in the repository (`src/asri/backtest/publication_lag.py`).

O Walk-Forward Validation: Methodology and Results

O.1 Methodology

We evaluate ASRI performance using only data available before each crisis event:

- **Window 1**: Train January 2021–April 2022, test Terra/Luna (May 2022)
- **Window 2**: Train January 2021–May 2022, test Celsius/3AC (June 2022)
- **Window 3**: Train January 2021–October 2022, test FTX (November 2022)

- **Window 4**: Train January 2021–February 2023, test SVB/USDC (March 2023)

For each window, we retain the theoretical weights (which are based on ex-ante domain knowledge rather than statistical optimisation) and calibrate the alert threshold using *only* pre-crisis data: the threshold for each event is set to the 90th percentile of the ASRI distribution observed before $t - 90$ (i.e. the upper-decile of the index's own training-period history). We then evaluate whether ASRI exceeds this training-relative threshold during the 30-day pre-event window. This calibration uses no information from or after the crisis, so detection cannot be an artefact of look-ahead bias; it also adapts the alert level to the regime prevailing before each event rather than imposing the fixed operational threshold of 50.

O.2 Results

Table 34 presents the walk-forward detection results.

Table 34: Walk-Forward Detection Performance

Crisis	Train-cal. Threshold	Pre-30 Peak	Lead (days)	Alarm Rate	FPR	Prec.
Terra/Luna	37.2	46.0	30	58.7%	55.8%	11.1%
Celsius/3AC	38.4	71.4	30	47.2%	43.8%	13.3%
FTX	46.8	84.7	28	12.8%	8.9%	34.5%
SVB/USDC	50.6	66.2	15	7.7%	6.0%	26.2%
Out-of-sample detection rate				4/4 (100%)		

Train-cal. Threshold = 90th percentile of ASRI over the pre-crisis training period (data before $t - 90$, i.e. Jan 2021–Feb 2022 for Terra/Luna through Jan 2021–Dec 2022 for SVB/USDC); calibration uses no crisis-period information.
Pre-30 Peak = maximum ASRI in the 30-day pre-crisis window; Lead = days between the first crossing of the training-calibrated threshold within that window and crisis onset (mean = 25.75 days). **Alarm Rate** = fraction of the *full* 2021–2026 index history exceeding that event's training-calibrated threshold; **FPR** = fraction of non-crisis days flagged; **Prec.** = precision (TP/(TP+FP)) over the full series. The low thresholds calibrated for the two *early* crises (Terra/Luna 37.2, Celsius/3AC 38.4) flag 47–59% of all index history at ≈11–13% precision; on a true post-training window the alarm rate is higher still (72.2% / 58.7%). The 4/4 "detection" for these events is therefore bought at a near-coin-flip alarm rate, which is why we read the walk-forward result as evidence against look-ahead bias rather than as an operational early-warning record. FTX and SVB thresholds (calibrated on histories that already include 2022 stress) are higher and alarm only 8–13% of the time.

O.3 Interpretation

The walk-forward validation flags all four events out-of-sample, but this 4/4 figure must be read alongside its false-positive cost (Table 34): for the two early crises the training-calibrated

thresholds are so low (37.2, 38.4) that they also flag 47–59% of the entire index history at ≈11–13% precision. We therefore treat the 4/4 rate as evidence that detection is *not an artefact of look-ahead bias*—the threshold is set without crisis-period information—rather than as a demonstration of operational early-warning skill, which the alarm rate contradicts for Terra/Luna and Celsius/3AC. Several observations merit discussion:

Lead Time. Using the first crossing of the training-calibrated threshold within the 30-day pre-crisis window, out-of-sample lead times are 30, 30, 28, and 15 days for Terra/Luna, Celsius/3AC, FTX, and SVB/USDC respectively (mean 25.75 days). For the two events whose pre-crisis ASRI was already well above the training-period upper decile (Terra/Luna and Celsius/3AC), the threshold is crossed at the start of the pre-window (lead capped at 30); SVB/USDC has the highest calibrated threshold (50.6, reflecting a higher-variance training history) and hence the shortest first-crossing lead (15 days). These first-crossing leads are a more conservative early-warning measure than the in-sample 1.5σ figures in Table 5, which are themselves capped at the 30-day pre-window (Section 5.4); we treat the walk-forward mean (≈26 days) as the operational early-warning horizon.

Training-Calibrated Thresholds. The pre-crisis upper-decile thresholds range from 37.2 (Terra/Luna, calibrated on a calm 2021 training history) to 50.6 (SVB/USDC, calibrated on a history that already includes the 2022 crises). All four crises breach their respective training-calibrated thresholds, and the pre-crisis ASRI peaks (46.0–84.7) clear them comfortably. Notably, Terra/Luna—which the *fixed* 50-threshold misses (pre-window peak 46.0)—is detected out-of-sample because the training-relative threshold (37.2) adapts to the calm pre-2022 baseline. This is the mechanism behind the 4/4 walk-forward rate: detection is assessed relative to the index's own pre-crisis history rather than against a single fixed level.

Robustness of Theoretical Weights. The 4/4 walk-forward detection rate is consistent with using theoretically-derived weights over statistically-optimised alternatives. Because the weights are based on domain knowledge about DeFi risk channels rather than statistical fitting to the training data, they hold up across the four liquidity-cascade events held out here, each detected when its own onset is excluded from calibration. We caution that all four events belong to a single broad failure family (liquidity/solvency cascades; see the generalisation-limit discussion in the main text), so this is evidence of robustness within that family, not of transfer to structurally different crisis mechanisms. The code-faithful Elastic Net weights concentrate on a single channel (Contagion Risk, ≈0.45; Table 14) and zero Regulatory Opacity, and might achieve better in-sample prediction while being more brittle to events that propagate through stablecoin mechanics (SCR).

Limitations. This validation uses fixed theoretical weights rather than re-estimating empirical weights in each training window. A more rigorous test would re-derive data-driven weights using only pre-crisis data and evaluate their out-of-sample performance. We defer this extension to future work, noting that the theoretical weights—our recommended specification for operational deployment—demonstrate robust walk-forward performance.

Continuous vs. Binary Evaluation. We note that ASRI is designed as a threshold-based monitoring index, not a continuous forecasting model. Walk-forward R^2 for predicting the 30-day-ahead ASRI *level* ($y_t = \text{ASRI}_{t+30}$) is negative across all folds: mean $R^2 = -0.6$ (std $= 0.2$) over five walk-forward folds, with a held-out $R^2 = -0.1$ (training pre-2024, testing 2024 onward). A negative R^2 means the weighted-component predictor does worse than a flat-mean baseline at point-forecasting the future level. Weight perturbation analysis additionally finds 0/4 sub-index components are individually robust to ±15% weight changes when evaluated by continuous prediction accuracy.

These results are expected and, critically, not informative for the index's intended purpose. ASRI is designed to cross a threshold before crises, not to minimise mean-squared prediction error. A below-baseline R^2 on the *level* is consistent with the efficient markets hypothesis—a risk indicator's level should not be forecastable if markets price risk efficiently—and reflects the mean-reverting, bounded ($[0, 100]$) nature of the index, whose absolute level is partly driven by proxy components with fixed default values (Table 10). The appropriate evaluation metrics are binary detection performance (4/4 out-of-sample, Table 34), classification accuracy (AUROC $= 0.866$, Table 8), and threshold-based early-warning lead times. The relative behaviour (crisis spikes vs. calm periods) is informative; the absolute magnitudes are not, and we make no point-forecasting claim.

P Out-of-Sample Specificity: Detail

P.1 2024 Stability Period

Our sample extends through January 2026, but the final in-sample crisis event (SVB/USDC) occurred in March 2023, providing approximately 34 months of out-of-sample data. The 2024 period—marked by the Bitcoin ETF approval (January 2024) and subsequent bull market—contained no systemic crisis comparable to the four events in our validation set. ASRI generated one multi-week elevation above the Elevated threshold (50) during this period: a roughly six-week episode in July–August 2024 (34 trading days ≥ 50, peak 58.8 on 2024-08-11) coinciding with the early-August 2024 global de-risking shock (the yen carry-trade unwind, which triggered a sharp crypto drawdown). This episode reflected genuine elevated cross-asset stress rather than a spurious alarm, and it did not escalate into a systemic crisis; ASRI returned below 50 and remained there for the rest of 2024 and through 2025. We therefore characterise 2024–2025 as free of *sustained false alarms*, with the single elevation tracking a real (ultimately non-systemic) macro stress episode rather than crypto-specific contagion.

P.2 The Bybit Hack (February 2025)

On February 21, 2025, Bybit suffered the largest exchange hack in cryptocurrency history ($1.5 billion stolen, attributed to DPRK actors by the FBI). Despite the headline magnitude, ASRI remained in the Moderate band throughout:

Table 35: ASRI Around Bybit Hack (Feb 2025). Readings computed from the frozen canonical series (`results/data/asri_history.parquet`) by `scripts/verify_deferred_validation.py`.

Date	ASRI	SCR	DLR	CR
Feb 10 (pre-hack peak)	42.3	36.7	46.6	49.8
Feb 21 (hack day)	40.6	37.6	39.7	43.8
Feb 28 (post-hack)	37.7	38.9	37.7	40.1

P.3 Why Bybit Was Not Systemic

Compare the Bybit hack to FTX, which triggered ASRI readings of 84.7:

- **FTX (systemic)**: Exchange collapse → Alameda insolvency → lender cascade (BlockFi, Genesis, Voyager) → credit contagion across DeFi
- **Bybit (contained)**: Funds stolen but no cascading failures. No stablecoin depegs, no DeFi protocol liquidations, no counterparty contagion.

The market absorbed a $1.5B theft without systemic stress because: (1) Bybit maintained solvency and honoured withdrawals; (2) no leveraged counterparties were exposed to Bybit-specific risk; (3) the bull market context provided ample liquidity buffer.

P.4 Specificity Interpretation

ASRI correctly distinguished a large but contained loss from systemic contagion. This provides out-of-sample evidence that the framework captures *transmission mechanisms*, not merely event magnitude. A $1.5B hack without contagion channels does not trigger ASRI.

Across all of 2025, ASRI never exceeded the Elevated threshold (50), despite the Bybit hack and significant market volatility (2025 peak: 48.0). Together with the single, non-systemic 2024 elevation discussed above, this pattern—registering elevated readings only when genuine cross-asset stress is present, while distinguishing large but contained events (Bybit) from true systemic contagion—complements the sensitivity demonstrated in Section 5.4.

Q API Documentation Summary

Table 36 provides endpoint documentation for primary data sources.

Table 36: Primary API Endpoints

Source	**Endpoint**	**Rate Limit**	**Authentication**
DeFi Llama	`api.llama.fi/v2/tvl`	300/5min	None
DeFi Llama	`stablecoins.llama.fi/stablecoins`	300/5min	None
FRED	`api.stlouisfed.org/fred/series`	None	API Key
CoinGecko	`api.coingecko.com/api/v3`	10-50/min	API Key (Pro)
Token Terminal	`api.tokenterminal.com/v2`	Varies	API Key

R Sub-Index Calculation Code

Python implementation of sub-index formulas is available in the repository at `src/asri/signals/calculator.p`
Key functions:

```
def calculate_stablecoin_risk(
    tvl_ratio: float,
    treasury_weight: float,
    hhi_concentration: float,
    peg_volatility: float
) -> float:
    return (
        0.4 * tvl_ratio +
        0.3 * treasury_weight +
        0.2 * hhi_concentration +
        0.1 * peg_volatility
    )
```

Full implementation: github.com/studiofarzulla/asri

S Historical Crisis Event Details

Terra/Luna Collapse (May 2022): UST algorithmic stablecoin depegged due to redemption spirals, eliminating $40B in value within 72 hours. LUNA dropped from $80 to near-zero.

Celsius/3AC Contagion (June 2022): Celsius Network froze withdrawals; Three Arrows Capital defaulted on loans. Combined losses exceeded $10B, triggering margin calls across crypto lending platforms.

FTX Bankruptcy (November 2022): FTX and Alameda Research filed for bankruptcy after liquidity crisis. Opaque counterparty relationships propagated losses across the ecosystem.

T Sample Market Assessment (December 2024)

Note: This appendix provides a sample ASRI reading for illustrative purposes; the values quoted are the released series' readings for the stated date. For current market conditions, visit asri.dissensus.ai.

Sample ASRI Reading

As of December 31, 2024, the released ASRI series records **38.1** (Moderate risk), with sub-index readings:

- **Stablecoin Risk**: 35.7 (moderate concentration, stable pegs)
- **DeFi Liquidity**: 41.3 (recovering from 2022–23 deleveraging)
- **Contagion Risk**: 39.9 (moderate cross-market stress on the TradFi proxy)
- **Regulatory Opacity**: 34.7 (improving regulatory clarity)

Applying the Equation 6 weights to these rounded sub-index readings gives $0.30 \times 35.7 + 0.25 \times 41.3 + 0.25 \times 39.9 + 0.20 \times 34.7 = 37.95$; the recorded aggregate (38.1) differs slightly because the released aggregate column was produced by the original generation pipeline and is not everywhere an exact weighted recomposition of the released (rounded) sub-index columns (see the data-provenance documentation in the code repository).

Risk Decomposition

Current primary risk contributors:

1. Stablecoin concentration in USDT/USDC (HHI = 0.52)
2. Treasury exposure through stablecoin reserves ($80B+ in T-bills)
3. Emerging RWA tokenisation growth (+45% YoY)

Regime Classification

The HMM classifies current market conditions as **Regime 2 (Moderate)** with high probability. The estimated one-step transition probabilities out of the Moderate regime (Table 20) indicate:

- 0.8% probability of moving to the Crisis regime
- negligible probability of moving directly to the Low Risk regime
- 99.2% probability of remaining in the Moderate regime

Alert Status

No immediate systemic stress signals detected. Key monitoring priorities:

1. Stablecoin reserve composition changes
2. Cross-market correlation shifts (BTC-equity)
3. Bridge vulnerability and exploit frequency

U Event Study Protocol Specification

This appendix provides the complete methodological specification for the event study analysis presented in Section 5.4, addressing reviewer concerns regarding pre-registration, window selection, multiple testing correction, and placebo testing.

U.1 Pre-Registration and Event Selection

U.1.1 Event Identification Criteria

Crisis events were identified *ex ante* based on three jointly necessary conditions (Definition 5.1):

1. **Magnitude**: Market capitalisation decline $\geq 15\%$ within 7 days, or single-asset collapse $\geq 50\%$ for assets with market cap $\geq \$10\text{B}$
2. **Contagion**: Cross-asset correlation surge $\bar{\rho}_t - \bar{\rho}_{t-30} \geq 0.20$
3. **Duration**: Elevated stress persisting ≥ 5 trading days

Event dates were sourced from external references (CoinDesk, Bloomberg, Reuters) prior to ASRI analysis, preventing data snooping on threshold selection.

U.1.2 Pre-Specified Parameters

The following parameters were fixed before analysis:

- Estimation window: $T_{\text{est}} = 60$ days ($t = -90$ to $t = -31$ relative to event)
- Event window: $T_{\text{event}} = 41$ days ($t = -30$ to $t = +10$)
- Significance level: $\alpha = 0.05$ (two-tailed)
- Lead time detection: 1.5 standard deviations above estimation-window mean

U.2 Window Selection Justification

U.2.1 Estimation Window: $[-90, -31]$

The 60-day estimation window was selected based on:

1. **Statistical power**: $n = 60$ provides adequate precision for mean and variance estimation while avoiding excessive smoothing of regime-dependent dynamics
2. **Regime stability**: ASRI exhibits regime persistence $> 97\%$ (Table 7), making 60 days sufficient to capture baseline behaviour within a regime
3. **Contamination avoidance**: The 30-day buffer ($t = -31$ cutoff) ensures estimation is complete before pre-crisis stress begins

Robustness: Alternative estimation windows (45 days, 90 days) produce qualitatively identical results under HAC inference: Celsius/3AC, FTX, and SVB remain the elevated, borderline-significant events under the finite-sample (fixed-b) reference while Terra/Luna remains clearly non-significant.

U.2.2 Event Window: $[-30, +10]$

The asymmetric event window reflects ASRI's design as an *early warning* system:

- **Pre-event (-30 to -1)**: Captures lead time—the period where ASRI begins detecting stress buildup
- **Event day ($t = 0$)**: Crisis onset (price cascade initiation)
- **Post-event ($+1$ to $+10$)**: Captures immediate aftermath and stress persistence

The 30-day pre-event window is calibrated to expected ASRI lead times (Table 5 shows 29–30 days across events).

U.3 Normal Model Specification

The constant mean model was selected for expected ASRI:

$$\mathbb{E}[\text{ASRI}_t] = \hat{\mu} = \frac{1}{T_{\text{est}}} \sum_{\tau=-90}^{-31} \text{ASRI}_\tau \tag{49}$$

U.3.1 Model Selection Rationale

1. **Simplicity**: The constant mean model makes minimal parametric assumptions
2. **Stationarity**: ASRI rejects unit root (ADF $p < 0.01$), validating level-based analysis
3. **AR(1) Alternative**: Modelling the abnormal series with explicit AR(1) dynamics yields the same qualitative reading as the Newey–West HAC correction—the three large-CAS events elevated and Terra/Luna non-significant—confirming that the naive independence-based significance was an artefact of the serial-correlation structure rather than of the normal-model specification

U.3.2 Variance Estimation

$$\hat{\sigma}^2_{\text{AS}} = \frac{1}{T_{\text{est}} - 1} \sum_{\tau=-90}^{-31} (\text{ASRI}_\tau - \hat{\mu})^2 \tag{50}$$

Autocorrelation diagnostics (Ljung–Box test) *decisively reject* white noise in the abnormal-signal series: p-values range from 10^{-15} to 10^{-28} on the event-window series and reach 2×10^{-58} on the estimation-window residuals. The naive independence-based standard error $\hat{\sigma}_{\text{AS}}\sqrt{T_{\text{event}}}$ is therefore invalid, and we instead use Newey–West HAC standard errors at lag $L = 20$:

$$\text{SE}_{\text{HAC}}(\text{CAS}) = n \times \text{SE}_{\text{HAC}}(\bar{\text{AS}}). \tag{51}$$

U.4 Multiple Testing Correction

With $K = 4$ simultaneous hypothesis tests (one per crisis event), we apply Bonferroni correction:

$$\alpha_{\text{adj}} = \frac{\alpha}{K} = \frac{0.05}{4} = 0.0125 \tag{52}$$

Results: The reference distribution matters here. Against a naive $\mathcal{N}(0,1)$ reference, three of four events would appear to clear the corrected threshold (Celsius/3AC $p = 0.0005$, FTX $p = 0.0010$, SVB $p = 0.0001$); but at the realised bandwidth-to-sample ratio ($b \approx 0.51$, AR(1) ≈ 0.85) that reference is invalid and overstates significance (see the finite-sample reference accompanying Table 5). On the valid fixed-b reference, $\alpha_{\text{adj}} = 0.0125$ corresponds to a critical $|t| \approx 5.7$, and the achieved p-values are:

- Terra/Luna: $p \approx 0.27$ ($\gg 0.0125$ ×)
- Celsius/3AC: $p \approx 0.051$ (> 0.0125 ×)
- FTX Collapse: $p \approx 0.063$ (> 0.0125 ×)
- SVB Crisis: $p \approx 0.036$ (> 0.0125 ×)

None of the three large-CAS events reaches the Bonferroni-corrected level: the largest HAC t in the panel is 3.86, well short of the $|t| \approx 5.7$ the correction requires. Once the reference accounts for the AR(1) ≈ 0.85 / $b \approx 0.51$ fat tails, then, no single event survives Bonferroni. The three large-CAS events are individually borderline at the unadjusted 5% level, and Terra/Luna fails at every threshold, consistent with its threshold-detection miss.

U.5 Window Independence

Table 37 documents the temporal separation between crisis events.

Table 37: Estimation and Event Window Dates, with Overlap Disclosure for the Clustered 2022 Crises

Event	Est. Start	Est. End	Evt. Start	Evt. End
Terra/Luna	2022-02-11	2022-04-11	2022-04-12	2022-05-22
Celsius/3AC	2022-03-19	2022-05-17	2022-05-18	2022-06-27
FTX Collapse	2022-08-13	2022-10-11	2022-10-12	2022-11-21
SVB Crisis	2022-12-11	2023-02-08	2023-02-09	2023-03-21

Within each event the 60-day estimation window and the 41-day event window are non-overlapping (estimation precedes the event).
The clustered 2022 events overlap across events. Terra/Luna and Celsius/3AC fall within five weeks of one another: their event windows overlap by ≈ 5 days, and their *estimation* windows overlap by ≈ 24 days (Terra/Luna's estimation window ends 2022-04-11, Celsius/3AC's begins 2022-03-19). Moreover, Celsius/3AC's estimation/baseline window (2022-03-19 to 2022-05-17) spans the Terra/Luna event window (2022-04-12 to 2022-05-22) by ≈ 35 days, so that baseline is not uncontaminated by the Terra/Luna crash. The two early-2022 events are therefore not statistically independent.
FTX and SVB events are separated by > 90 days and have fully non-overlapping estimation and event windows.

We do not claim full window independence for the clustered 2022 crises. Terra/Luna and Celsius/3AC share estimation and baseline periods (the overlaps above), so their abnormal-signal series—and the corresponding hold-one-out windows (Table 27)—are not fully independent. We disclose this as a genuine limitation of a sample whose crises cluster in 2022, and we mitigate it only partially:

1. The events represent distinct crisis mechanisms (algorithmic stablecoin vs. CeFi lending)
2. Separate CAS calculations use event-specific baselines, though for Celsius/3AC that baseline overlaps the Terra/Luna shock
3. The FTX and SVB events, by contrast, are fully separated and provide two genuinely independent episodes

U.6 Placebo Testing

To assess the specificity of the event-study inference, we run the identical HAC fixed-b pipeline on every eligible non-crisis date in the sample (889 dates, excluding 90-day windows around the four crises and the first/last 90 days for edge effects).

U.6.1 Placebo Date Selection

Dates were drawn uniformly from the sample period (2021-01 to 2024-12) excluding:

- 90-day windows around known crisis events
- First/last 90 days of sample (edge effects)

U.6.2 Placebo Results

Table 38: Placebo/specificity test: the HAC fixed-b event-study statistic applied to non-crisis dates. Reproducible via `code/scripts/placebo_chow_cusum_repro.py`.

Metric (HAC, $L = 20$, fixed-b)	Value
N non-crisis dates	889
Median placebo $\|t\|$	5.10
Fixed-b 5% critical value	3.52
Frac. of placebo dates clearing it	0.62

The result is decisive, and it runs against the event study. Under the fixed-b reference the 5% critical value ($|t| = 3.52$) sits well below the *median* of the placebo $|t|$ distribution (5.10), so $\approx$ 62% of arbitrary non-crisis dates are flagged "significant"—an empirical size an order of magnitude above nominal. The manuscript's crisis onsets carry HAC t-statistics of 3.28–3.86 (Table 5), themselves below the placebo median, so on a smooth, strongly trending series they cannot be distinguished from arbitrary dates under a reference this anti-conservative. We therefore treat the event study as inconclusive and rest the empirical case on the fair-baseline day-level discrimination analysis, which does not depend on it.

U.7 Lead Time Measurement

Two complementary lead time definitions are employed:

Definition 1 (First-crossing): Days between first observation exceeding 1.5σ above baseline and crisis onset. Captures earliest structural warning signal.

Definition 2 (Final-sustained): Days between last observation below threshold before sustained elevation and crisis onset. Captures actionable intervention timing.

The walk-forward analysis (Table 34) reports first-crossing lead times; Table 5 reports the (30-day-capped) sustained-elevation lead times. The two definitions can diverge for a given event because ASRI fluctuates between its initial crossing of the pre-crisis-calibrated threshold and the onset of the crisis.

U.8 Sensitivity to Specification Choices

Table 39: Event Study Robustness to Specification Changes (naive significances; superseded by the HAC-robust inference of Table 5)

Specification	Terra	Celsius	FTX	SVB
Baseline (60d, const. mean)	***	***	***	***
45-day estimation window	***	***	***	***
90-day estimation window	***	***	***	***
AR(1) normal model	***	***	***	**
Event window $[-20, +10]$	***	***	***	***
Event window $[-40, +10]$	***	***	***	***

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Stars are naive (independence-assumption) significances, reported only for cross-specification comparison; they are superseded by the HAC-robust inference of Table 5 (Section 5.4), under which Terra/Luna is non-significant ($t = 1.72$, $p = 0.265$)—i.e. 3/4 significant, not 4/4. Per-specification HAC inference was not computed; the stars assume independence and are invalid given the AR(1) ≈ 0.8–0.9 serial correlation documented in Section 5.4.