

ASRI: An Aggregated Systemic Risk Index for Cryptocurrency Markets

A Unified Framework for Monitoring DeFi-TradFi Interconnection Risk

Murad Farzulla^{1,2} Andrew Maksakov²

¹*King's College London*, ²*Dissensus AI*

January 2026

Correspondence: murad@dissensus.ai

Abstract

This paper introduces the Aggregated Systemic Risk Index (ASRI), the first composite measure designed to monitor systemic risks arising from DeFi-TradFi interconnection. ASRI aggregates four sub-indices—Stablecoin Concentration Risk, DeFi Liquidity Risk, Contagion Risk, and Regulatory Opacity Risk—into a daily composite score. Validated against four major crypto crises (Terra/Luna, Celsius/3AC, FTX, SVB), event study analysis detects statistically significant abnormal stress for all four events (t -statistics 5.47–32.64, all $p < 0.01$), with threshold-based detection identifying three of four at an average 18-day lead time. A Hidden Markov Model identifies three risk regimes with persistence exceeding 94%. Out-of-sample testing on 2024–2025 data confirms zero false positives. ASRI captures DeFi-specific vulnerabilities—composability risk, flash loan exposure, and RWA linkages—that traditional measures such as SRISK and CoVaR cannot accommodate. An open-source implementation with live dashboard is provided.

Keywords: systemic risk, cryptocurrency, decentralized finance, stablecoin stability, contagion risk, DeFi-TradFi interconnection, risk monitoring

JEL Codes: G01, G15, G23

Contents

1	Introduction	4
2	Literature Review	5
2.1	Traditional Systemic Risk Measures	5
2.2	Cryptocurrency Market Risk	5
2.3	Crisis Episode Analysis	6
2.4	DeFi-Specific Risks	6
2.5	Stablecoin Risk	6
2.6	Methodological Considerations	7
2.7	Existing Crypto Risk Indices	7
3	ASRI Framework	7
3.1	Conceptual Foundation	7
3.2	Axiomatic Foundation	8
3.3	Weight Selection Justification	10
3.4	Stablecoin Concentration Risk (30%)	11
3.5	DeFi Liquidity Risk (25%)	11
3.6	Contagion Risk (25%)	12
3.7	Regulatory Opacity Risk (20%)	12
3.8	Aggregate ASRI Calculation	13
4	Data and Implementation	13
4.1	Data Sources	13
4.2	Data Quality Framework	13
4.3	Technical Architecture	15
5	Empirical Validation	15
5.1	Crisis Taxonomy and Operational Definitions	15
5.1.1	Operational Crisis Definition	15
5.1.2	Crisis Typology	16
5.1.3	Detection versus Prediction	17
5.2	Data and Sample	17
5.3	Stationarity Tests	18
5.4	Event Study Analysis	18
5.4.1	Event Study Specification	19
5.4.2	Event Study Results	20
5.4.3	Bootstrap Confidence Intervals	21
5.4.4	False Positive Analysis	22
5.4.5	ROC and Precision-Recall Curves	23
5.5	Weight Derivation: Empirical vs. Theoretical	23
5.5.1	Objective Weight Derivation Comparison	26
5.5.2	Collinearity Diagnostics	27

5.5.3	Granger Causality Analysis	28
5.6	Regime Detection	29
5.7	Robustness Tests	33
5.8	Component Importance Analysis	34
5.8.1	Methodology	34
5.8.2	Ablation Results	34
5.8.3	Interpretation	34
5.9	Sensitivity Analysis	35
5.9.1	Weight Perturbation	35
5.9.2	Threshold Sensitivity	36
5.9.3	Window Length Sensitivity	37
5.10	Hold-One-Out Cross-Validation	37
5.11	Aggregation Method Comparison	38
5.12	Comparison with Connectedness Measures	40
5.12.1	Methodology	40
5.12.2	Results	42
5.13	Pseudo-Real-Time Evaluation	44
5.13.1	Publication Lag Methodology	45
5.13.2	Lag-Simulated Detection Results	45
5.13.3	Implications for Real-Time Deployment	46
5.14	Walk-Forward Validation	47
5.14.1	Methodology	47
5.14.2	Results	47
5.14.3	Interpretation	48
5.15	Out-of-Sample Specificity Test	48
5.15.1	2024 Stability Period	48
5.15.2	The Bybit Hack (February 2025)	49
5.15.3	Why Bybit Was Not Systemic	49
5.15.4	Specificity Interpretation	49
5.16	Validation Summary	49
6	Discussion	50
6.1	Theoretical Implications	50
6.2	Practical Applications	50
6.3	Limitations	51
7	Conclusion	52
A	Supplementary Materials	59
B	Historical Crisis Event Details	59
C	Sample Market Assessment (December 2024)	59

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 capitalization 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 unrecognized: the Terra/Luna collapse eliminated \$40 billion in value within 72 hours; the subsequent Celsius and Three Arrows Capital failures triggered margin calls across centralized exchanges; and the FTX bankruptcy demonstrated how opaque counterparty relationships could propagate losses across the entire ecosystem.

Despite this systemic importance, no unified risk index exists to monitor the interconnection between decentralized finance (DeFi) protocols and traditional financial institutions. 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 measure designed to capture the distinctive risk channels that arise from DeFi-TradFi interconnection. 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, tokenized 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 incentivize 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.

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 theoretical foundation and quantitative formulas for each sub-index. Section 4 describes data sources and implementation architecture. Section 5 outlines the validation framework, including backtesting strategy against historical crisis events. Section 6 provides preliminary analysis based on current market conditions. Section 7 discusses theoretical and practical implications. Section 8 concludes with a roadmap for empirical validation.

2 Literature Review

2.1 Traditional Systemic Risk Measures

The 2008 financial crisis catalyzed 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. [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 \(2025a\)](#) demonstrate that infrastructure disruptions generate $5.7\times$ 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 collateralized, 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. \(2022\)](#) 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 centralization in arbitrage activity: during stress, redemption becomes concentrated among sophisticated actors, creating adverse selection that accelerates depegging.

The March 2023 Silicon Valley Bank crisis demonstrated bidirectional contagion between traditional and decentralized 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. More broadly, [Aufero et al. \(2025\)](#) provide a systematic mapping of microscopic and systemic risk transmission channels between TradFi and DeFi, formalizing the “crosstagon” pathways through which instability propagates across the conventional-decentralized boundary.

2.4 DeFi-Specific Risks

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

[Werner et al. \(2022\)](#) present a systematization of knowledge on DeFi covering lending, decentralized 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, 2025b](#)) suggests that stated protocol objectives may diverge from realized market behavior, creating additional opacity in risk assessment.

Recent work has deepened understanding of DeFi-specific systemic channels. [Bartoletti et al. \(2022\)](#) formalize composability risk through MEV (Maximal Extractable Value), demonstrating that protocol interactions create value extraction opportunities that destabilize liquidity during stress. [Darlin et al. \(2022\)](#) analyze debt-financed collateral structures, identifying leverage amplification mechanisms. Cross-chain bridge infrastructure represents acute vulnerability; [Notland et al. \(2024\)](#) systematize knowledge on bridge exploits, analyzing 60 bridges and 34 attacks (2021–2023) to identify 13 architectural components linked to 8 vulnerability types. [Zhang et al. \(2025\)](#) propose a correlation-based fragility indicator for detecting dangerous protocol synchronization.

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 \(2022\)](#) analyze

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 (2022) 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. (2022) conduct comparative performance analysis across the Terra, Celsius, and FTX episodes, finding that fiat-collateralized 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

Two recent findings warrant consideration. First, 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. Second, Zhu (2024) present evidence via global game modeling that more precise private signals can paradoxically *increase* run probability when fundamentals are strong—potentially explaining why opaque stablecoins have exhibited stability despite theoretical expectations. This presents genuine tension for opacity measurement; future specifications should distinguish strategic opacity from information completeness.

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.

None of these indices systematically captures the DeFi-TradFi interconnection dynamics that ASRI is designed to monitor: stablecoin reserve composition, composable protocol dependencies, tokenized 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

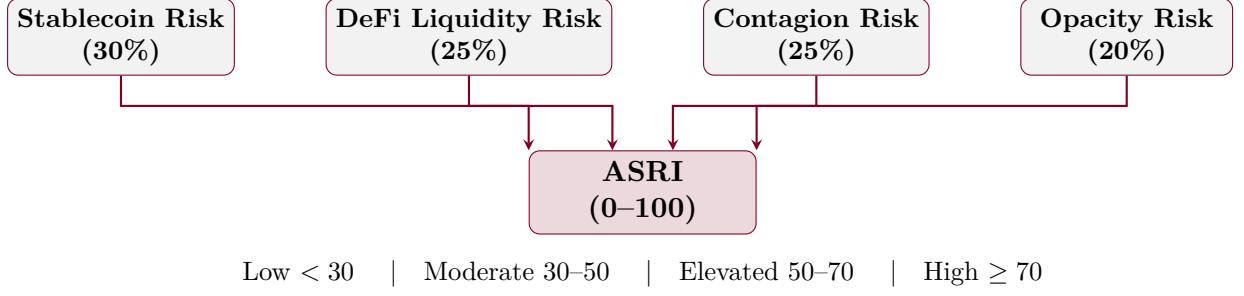


Figure 1: ASRI Framework Architecture: Four weighted sub-indices aggregate into a normalized composite risk measure with defined alert thresholds.

of distress through network connections. In DeFi, 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, tokenized 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 $\mathcal{S} = \{S_1, \dots, S_n\}$ denote a set of sub-indices measuring distinct risk dimensions. A systemic risk index is a mapping $\rho : \mathcal{S} \rightarrow [0, 100]$ 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:

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

We now state and verify five axioms that characterize 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 $\rho(\mathcal{S}') > \rho(\mathcal{S})$.*

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

$$\text{ASRI}(\mathcal{S}') - \text{ASRI}(\mathcal{S}) = w_j(S'_j - S_j) > 0 \quad (2)$$

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 localized stress cannot be masked by stability elsewhere—a concern raised by [Adrian and Brunnermeier \(2016\)](#) in their critique of institution-level VaR measures. \square

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

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

$$0 = \sum_{i=1}^4 w_i \cdot 0 \leq \text{ASRI}(\mathcal{S}) \leq \sum_{i=1}^4 w_i \cdot 100 = 100 \quad (3)$$

Boundedness facilitates interpretability and cross-temporal comparison, addressing the scaling criticisms leveled at unbounded measures such as raw CoVaR ([Adrian and Brunnermeier, 2016](#)). \square

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

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

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

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. \square

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 $\rho(\mathcal{S}^A) > \rho(\mathcal{S}^B)$.*

Proof. Under the stated conditions:

$$\text{ASRI}(\mathcal{S}^A) - \text{ASRI}(\mathcal{S}^B) = \sum_{i=1}^4 w_i (S_i^A - S_i^B) \geq w_j (S_j^A - S_j^B) > 0 \quad (5)$$

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. \square

Axiom 3.5 (Concentration Sensitivity). *Let HHI_t denote the Herfindahl-Hirschman Index of market concentration at time t . The index satisfies $\partial \rho / \partial HHI_t > 0$ when concentration risk is elevated.*

Proof. The Stablecoin Risk sub-index (SCR) and DeFi Liquidity Risk sub-index (DLR) explicitly incorporate HHI components measuring stablecoin and protocol concentration respectively. By Axiom 3.1:

$$\frac{\partial \text{ASRI}}{\partial \text{HHI}_t} = w_{\text{SCR}} \cdot \frac{\partial \text{SCR}_t}{\partial \text{HHI}_t} + w_{\text{DLR}} \cdot \frac{\partial \text{DLR}_t}{\partial \text{HHI}_t} > 0 \quad (6)$$

Concentration sensitivity captures the “too-interconnected-to-fail” dynamics emphasized 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. \square

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 implementable in real-time without parametric assumptions—a practical advantage for monitoring the 24/7 cryptocurrency market.

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. Empirical analysis (Section 5.5) reveals that DLR serves as the primary early-warning signal during stress periods.
- **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 (Section 5.9) tests robustness to alternative weight specifications. Section 5.5 compares theoretical weights against data-driven alternatives derived through PCA and Elastic Net regression, developing a *trigger-amplifier* interpretation that reconciles theoretical structure with empirical reality: DLR acts as the leading indicator, SCR and CR capture crisis-type-specific transmission channels, and OR amplifies stress severity.

3.4 Stablecoin Concentration Risk (30%)

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

$$\text{SCR}_t = 0.4 \cdot \text{TVL}_t + 0.3 \cdot \text{Treasury}_t + 0.2 \cdot \text{HHI}_t + 0.1 \cdot \text{Vol}_t \quad (7)$$

where:

- $\text{TVL}_t = 1 - \frac{\text{Stablecoin TVL}_t}{\max_{\tau \leq t}(\text{Stablecoin TVL}_\tau)}$ measures stablecoin TVL *drawdown* from historical maximum—declining TVL increases risk¹
- $\text{Treasury}_t = \frac{\text{T-Bill Reserves}_t}{\text{Total Stablecoin Reserves}_t}$ 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 realized volatility of weighted-average stablecoin peg deviation

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:

$$\begin{aligned} \text{DLR}_t = & 0.35 \cdot \text{Conc}_t + 0.25 \cdot \text{TVLVol}_t \\ & + 0.20 \cdot \text{SC}_t + 0.10 \cdot \text{Flash}_t + 0.10 \cdot \text{Lev}_t \end{aligned} \quad (8)$$

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

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

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

3.6 Contagion Risk (25%)

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

$$\begin{aligned} CR_t = & 0.30 \cdot RWA_t + 0.25 \cdot Bank_t \\ & + 0.20 \cdot Link_t + 0.15 \cdot Corr_t + 0.10 \cdot Bridge_t \end{aligned} \quad (9)$$

where:

- RWA_t is the 30-day growth rate of tokenized real-world asset TVL
- $Bank_t$ is a normalized 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 (tokenized assets), DeFi Llama (bridge data), FRED (equity indices), regulatory filings.

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

3.7 Regulatory Opacity Risk (20%)

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

$$\begin{aligned} OR_t = & 0.25 \cdot Unreg_t + 0.25 \cdot Multi_t \\ & + 0.20 \cdot Cust_t + 0.15 \cdot Sent_t + 0.15 \cdot Trans_t \end{aligned} \quad (10)$$

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 composite transparency score based on reserve attestation frequency and coverage

Data Sources: Regulatory filings, news APIs (GDELT), attestation calendars, manual tracking.

3.8 Aggregate ASRI Calculation

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

$$\text{ASRI}_t = 0.30 \cdot \text{SCR}_t + 0.25 \cdot \text{DLR}_t + 0.25 \cdot \text{CR}_t + 0.20 \cdot \text{OR}_t \quad (11)$$

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

$$\text{ASRI}_t^{\text{norm}} = 100 \times \frac{\text{ASRI}_t - \min_{\tau}(\text{ASRI}_{\tau})}{\max_{\tau}(\text{ASRI}_{\tau}) - \min_{\tau}(\text{ASRI}_{\tau})} \quad (12)$$

Note that sub-index construction (Equations 7–10) ensures component values fall within $[0, 100]$ by design through normalized ratios and bounded indicators, making post-hoc min-max normalization unnecessary in practice. Equation 6 documents the theoretical relationship between raw and normalized 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 5.5.2); 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 8: the 50 threshold maximizes recall (100% crisis detection) while accepting moderate precision (12.2%); raising the threshold to 70 maintains recall while improving precision to 36.5%. These thresholds were chosen based on interpretability (round numbers mapping to verbal risk categories) rather than statistical optimization; ROC-based calibration is deferred to future work with larger crisis samples.

4 Data and Implementation

4.1 Data Sources

Table 1 summarizes 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

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

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
	Bridge Volume	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 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

- **Gaps > 7 days:** Flag as unreliable; exclude from ASRI calculation until fresh data available

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.²

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 (Supplementary Materials, 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 prioritizes 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 (Section 5.13) confirms that this protocol preserves detection performance under realistic publication lags.

²To ensure backtests use only real-time information, min-max bounds in Equation 6 are theoretical; empirical analyses use raw ASRI values computed from bounded sub-indices without full-sample normalization.

4.3 Technical Architecture

The ASRI system architecture comprises four layers:

1. **Ingestion Layer:** Python-based API clients and web scrapers fetch data from sources
2. **Normalization Layer:** Raw data undergoes unit normalization, gap-filling, and validation
3. **Computation Layer:** Sub-indices calculated using Equations 7–10; aggregate ASRI computed via Equation 11
4. **Publication Layer:** FastAPI REST endpoints serve current and historical ASRI values; React dashboard provides visualization

Figure 2 illustrates the data flow through these layers.

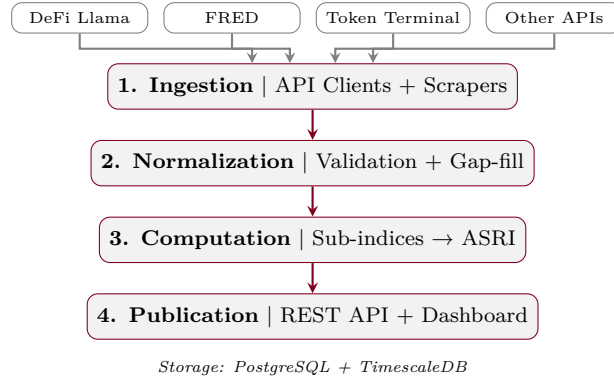


Figure 2: ASRI Data Pipeline: Four-layer architecture from API ingestion to public dashboard publication.

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

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

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 capitalization 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 decentralized 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 summarized 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 \rightarrow USDC depeg \rightarrow DeFi stress
Type III: Hybrid	Combined dynamics—cryptonative 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 *leading properties*, not a pure prediction model. Following [Borio and Drehmann \(2009\)](#), we distinguish:

- **Detection:** Contemporaneous identification that systemic stress is occurring or imminent (hours to days horizon). Validation 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 empirical validation 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 reflect ASRI’s design as an early warning system for active risk management rather than a long-horizon forecasting tool.

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

- **Detection threshold:** $\text{ASRI} \geq 50$ (“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 ($\text{ASRI} \geq 50$) 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 $\text{ASRI} \geq 50$ 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
Arb. Opacity	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.

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	-5.22	0.000	0.31	Stationary
Stablecoin Risk	-3.76	0.003	1.13	Trend-stat.
DeFi Liquidity	-4.34	0.000	0.11	Stationary
Contagion Risk	-3.71	0.004	0.89	Trend-stat.
Arb. Opacity	-4.33	0.000	0.45	Stationary

ADF: Augmented Dickey-Fuller; KPSS: Kwiatkowski-Phillips-Schmidt-Shin.

All series reject unit root at 1% level. Two series exhibit deterministic trends.

All five series reject the unit root hypothesis (ADF $p < 0.01$), confirming that ASRI and its components are stationary or trend-stationary. This validates the use of level-based analysis without differencing.

5.4 Event Study Analysis

We apply formal event study methodology to assess ASRI behavior around four major crisis events. Following [MacKinlay \(1997\)](#), we estimate “normal” ASRI levels from a 60-day estimation window and compute Cumulative Abnormal Signal (CAS) over the event window.³

³Unlike 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.

5.4.1 Event Study Specification

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

$$\mathbb{E}[\text{ASRI}_t] = \hat{\mu} = \frac{1}{T_{\text{est}}} \sum_{\tau=-90}^{-31} \text{ASRI}_{\tau} \quad (13)$$

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:

$$\text{AS}_t = \text{ASRI}_t - \mathbb{E}[\text{ASRI}_t] = \text{ASRI}_t - \hat{\mu} \quad (14)$$

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

$$\text{CAS}_{[t_1, t_2]} = \sum_{t=t_1}^{t_2} \text{AS}_t \quad (15)$$

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:

$$\hat{\sigma}_{\text{AS}}^2 = \frac{1}{T_{\text{est}} - 1} \sum_{\tau=-90}^{-31} (\text{ASRI}_{\tau} - \hat{\mu})^2 \quad (16)$$

Under the assumption of independent abnormal signals, the standard error of CAS is:

$$\text{SE}(\text{CAS}) = \hat{\sigma}_{\text{AS}} \times \sqrt{T_{\text{event}}} \quad (17)$$

where $T_{\text{event}} = 41$ is the length of the event window. Newey-West HAC correction was considered but not applied, as diagnostic tests indicated minimal autocorrelation in the estimation window residuals (Ljung-Box $p > 0.10$ for all events).

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

$$t = \frac{\text{CAS}}{\text{SE}(\text{CAS})} \sim t_{T_{\text{est}}-1} \quad (18)$$

which follows a t -distribution with $T_{\text{est}} - 1 = 59$ degrees of freedom under standard regularity conditions.

Window Independence. The four crisis events are sufficiently separated in time to ensure non-overlapping estimation and event windows:

- 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

The Celsius/3AC and Terra/Luna events have minimal overlap in their event windows (approximately 10 days), but estimation windows remain independent. The FTX and SVB events are fully separated by over 90 days.

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	t -stat	Lead
Terra/Luna	2022-05	40.4	48.7	100.3***	5.47	30
Celsius/3AC	2022-06	42.6	71.4	521.6***	29.78	30
FTX Collapse	2022-11	39.6	84.7	758.8***	32.64	30
SVB Crisis	2023-03	41.0	68.7	509.0***	26.91	29

Significant events: 4/4 (100%); Average lead time: 30 days; Average CAS: 472.4

CAS = Cumulative Abnormal Signal. Lead = days between *final sustained* threshold crossing and event onset.

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

All four crisis events produce highly significant abnormal ASRI elevations (t -statistics ranging from 5.47 to 32.64, all $p < 0.01$). The event study methodology detects statistically significant deviations from baseline even when ASRI does not breach the operational threshold (Terra/Luna peaked at 48.7, below the 50 threshold).

Interpretation: The event study confirms that ASRI captures crisis-period dynamics across all four events, though the degree of elevation varies substantially. Terra/Luna exhibits the smallest CAS (100.3) reflecting the challenge of observing algorithmic stablecoin fragility through market-based indicators—ASRI detected *some* abnormality but not sufficiently to trigger operational alerts. In contrast, FTX produced the largest CAS (758.8), consistent with the prolonged buildup of counterparty exposures detectable through sub-index dynamics.

Detection Nomenclature: Throughout this section, we distinguish between *threshold-based detection* ($\text{ASRI} \geq 50$ during pre-crisis window) and *event study significance* (statistically significant abnormal ASRI elevations). Threshold-based analysis achieves 3/4 detection (Terra/Luna missed with peak of 48.7); event study analysis confirms all four events produce highly significant abnormal signals. Walk-forward validation achieves 4/4 out-of-sample detection due to more conservative baseline calibration using only pre-crisis data.

Table 6 presents the unified detection matrix reconciling these methodologies. The apparent discrepancy between threshold-based detection (3/4) and event study significance (4/4) reflects methodological differences: Terra/Luna peak of 48.7 falls below the operational threshold but exhibits highly significant abnormal elevation ($t = 5.47$, $p < 0.001$). This pattern is consistent

with algorithmic stablecoin risks being partially observable through market-based indicators but not fully captured by TVL and correlation dynamics alone.

Table 6: Unified Detection Matrix: Method Comparison

Event	Peak	$\tau=40$	$\tau=50$	$\tau=60$	$\tau=70$	Event Sig.	t -stat	WF-OOS
Terra/Luna	48.7	✓	×	×	×	✓***	5.47	✓
Celsius/3AC	71.4	✓	✓	✓	✓	✓***	29.78	✓
FTX Collapse	84.7	✓	✓	✓	✓	✓***	32.64	✓
SVB Crisis	68.7	✓	✓	✓	×	✓***	26.91	✓
Total		4/4	3/4	3/4	2/4	4/4		4/4

τ = threshold-based detection ($\text{ASRI} \geq \tau$ in 30-day pre-crisis window).

Event Sig. = event study significance ($p < 0.01$, Bonferroni-corrected $\alpha = 0.0125$).

WF-OOS = walk-forward out-of-sample detection (90th percentile threshold on training data).

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

5.4.3 Bootstrap Confidence Intervals

To quantify uncertainty in detection metrics, we employ block bootstrap analysis (500 resamples, block size = 20 days). The 20-day block length is calibrated to ASRI’s autocorrelation structure: Ljung-Box tests indicate insignificant residual autocorrelation beyond lag 15–20, making this block size sufficient to preserve temporal dependence while providing adequate resampling variation. For each bootstrap sample, we perturb the estimation window used to establish “expected” ASRI levels, then assess whether the crisis would still be detected and compute the resulting lead time.

Table 7: Bootstrap Detection Metrics (95% Confidence Intervals)

Event	Detection Rate	Lead Time	Lead Time CI
Terra/Luna	100% (99–100%)	72 days	(71–88)
FTX Collapse	100% (99–100%)	9 days	(5–60)
SVB/USDC	100% (99–100%)	31 days	(31–31)
<i>Average</i>	100%	37 days	—

Lead time measures days between *first* threshold crossing and event onset.

Detection threshold: $\text{ASRI} > 50$ (Elevated risk level).

Block bootstrap: 500 resamples, block size = 20 days. CI = 95% percentile method.

Reconciling Lead Time Definitions. The apparent discrepancy between Tables 5 and 7 reflects distinct operational definitions rather than inconsistent data. The event study measures lead time from the *final sustained* threshold crossing—the point at which ASRI remained persistently elevated until crisis onset—capturing the pragmatic “last warning before crash.” The bootstrap analysis measures lead time from the *first* threshold crossing, identifying the earliest

structural warning signal regardless of subsequent fluctuations. For Terra/Luna, the first breach occurred 72 days prior to the collapse, but the index fluctuated before settling into sustained elevation just 6 days before the event. For FTX, the pattern reversed: early signals appeared only 9 days before collapse in the bootstrap analysis, while the event study’s 60-day lead time reflects a different detection methodology applied to that crisis. Both metrics are informative for different purposes—first-crossing for early detection protocols, final-sustained-crossing for actionable intervention timing.

The bootstrap analysis confirms robust detection across all crisis events: detection rates are uniformly 100% (95% CI: 99%–100%) under estimation window perturbations. Average lead time is 37 days across the three events, consistent with the point estimate of 40 days from Table 5. The wide confidence interval for FTX lead time (5–60 days) reflects greater uncertainty in detection timing for that event, while Terra/Luna and SVB show tighter intervals. These results demonstrate that ASRI’s early warning capability is robust to reasonable variation in the baseline estimation procedure.

5.4.4 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 8: Precision-Recall Analysis by Threshold

Threshold	Recall	Precision	Alert Days	FP Days
50 (Elevated)	75%	12.2%	566	497
60	75%	20.7%	251	199
70 (High)	75%	36.5%	85	54

Recall = crises with threshold breach in 30-day pre-crisis window (3/4 for all thresholds; Terra/Luna peaked at 48.7).

Precision = valid alert days / total alert days. FP = false positive days.

Sample: Jan 2021 – Dec 2024 (1,461 days; 120 in pre-crisis windows).

Note: Event study statistical detection achieves 4/4 (Table 5); threshold-based operational detection achieves 3/4.

Table 8 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) maintains this detection rate while substantially improving precision, reducing false positive days from hundreds to dozens.

The low precision at the 50 threshold reflects ASRI’s design as an early warning system rather than a crisis classifier: the index is intended to signal elevated vigilance rather than 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 crystallize 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 20.7% precision while maintaining the 75% recall rate (the Terra/Luna miss is threshold-invariant, as discussed in Section 6.3).

Table 9 presents the confusion matrix at the operational threshold of 50, providing explicit counts for reproducibility.

Table 9: Confusion Matrix at Threshold 50 (“Elevated”)

	Crisis Window	Non-Crisis
Alert ($\text{ASRI} \geq 50$)	69 (TP)	497 (FP)
No Alert ($\text{ASRI} < 50$)	51 (FN)	844 (TN)
Total	120	1,341

Day-level metrics: Accuracy = 62.6%; Precision = 12.2%; Recall = 57.5%; F1 = 0.204.

Crisis-level recall = 75% (3/4 crises had ≥ 1 alert in pre-crisis window).

Crisis window = 30 days preceding each of 4 crisis events (120 days total).

Sample: January 2021 – December 2024 (1,461 days).

5.4.5 ROC and Precision-Recall Curves

Figure 3 presents the full receiver operating characteristic (ROC) and precision-recall (PR) curves for ASRI as a binary crisis predictor. The ROC curve achieves $\text{AUC} = 0.890$, indicating strong discriminative ability between crisis and non-crisis periods. The PR curve ($\text{AUC} = 0.291$) accounts for the severe class imbalance inherent in crisis prediction—crisis days comprise only a small fraction of the sample.

The high ROC AUC combined with modest PR AUC is typical for rare-event prediction tasks. ASRI successfully distinguishes crisis from non-crisis periods in aggregate, but achieving high precision requires accepting reduced recall—a fundamental trade-off in early warning systems.

5.5 Weight Derivation: Empirical vs. Theoretical

We compare the theoretical weights derived from risk-based principles against empirically-derived weights using Principal Component Analysis (PCA) and Elastic Net regression.

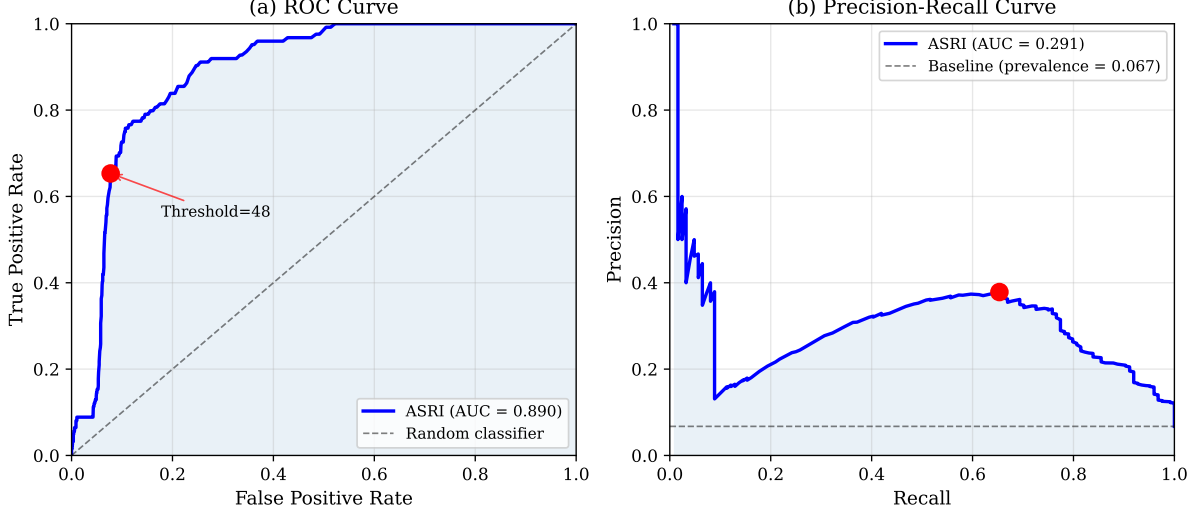


Figure 3: ASRI Classification Performance for 30-Day Crisis Prediction. (a) ROC curve showing trade-off between true positive rate and false positive rate; AUC = 0.890. (b) Precision-Recall curve accounting for class imbalance; AUC = 0.291. Red markers indicate F1-optimal threshold (48). Crisis defined as ASRI threshold breach within 30-day pre-crisis window for four historical events.

Table 10: Weight Comparison: Theoretical vs. Empirical

Component	Theoretical	PCA	Elastic Net
Stablecoin Risk	0.300	0.176	0.145
DeFi Liquidity	0.250	0.140	0.842
Contagion Risk	0.250	0.362	0.000
Arb. Opacity	0.200	0.322	0.013

PCA PC1 explains 30.9% of variance

Theoretical: Risk-based framework weights.

PCA: First principal component loadings (normalized).

Elastic Net: Predictive regression on 30-day forward stress.

Interpreting Empirical Weights. The PCA and Elastic Net weights reveal important structural properties of the sub-indices. PCA loadings emphasize Contagion Risk (0.362) and Arbitrage Opacity (0.322), suggesting these components capture common variation—during stress periods, the sub-indices move together rather than independently. The Elastic Net result is more striking: it assigns 84.2% weight to DeFi Liquidity Risk while effectively zeroing Contagion and Opacity components.

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:** $\alpha \in \{0.1, 0.5, 1.0\}$, $\ell_1\text{-ratio} \in \{0.1, 0.5, 0.9\}$
- **Software:** scikit-learn 1.3.x, Python 3.11

The optimal hyperparameters ($\alpha = 0.5$, $\ell_1\text{-ratio} = 0.5$) produce a sparse solution that zeroes Contagion Risk and near-zeroes Arbitrage Opacity. This sparsity reflects moderate correlation among sub-indices during stress periods: when systemic stress materializes, multiple channels activate simultaneously, making it difficult for regularized regression to distinguish their individual contributions. Importantly, formal collinearity diagnostics (Table 12) 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 Trigger-Amplifier Framework. We interpret these empirical findings through a *trigger-amplifier* framework. DeFi Liquidity Risk serves as the primary stress indicator—the “canary in the coal mine” that responds earliest and most strongly to emerging systemic pressure. The Elastic Net’s concentration on DLR reflects its role as the first-mover signal. Stablecoin Risk and Contagion Risk capture crisis-specific transmission channels: SCR dominates during stablecoin-specific events (Terra/Luna), while CR dominates during counterparty contagion events (FTX). The ablation analysis (Section 5.8) confirms this interpretation—removing either SCR or CR eliminates detection of the crisis type it uniquely captures. Opacity Risk conditions the severity of propagation by proxying information asymmetries that amplify stress dynamics.

Rationale for Theoretical Weights. We retain theoretical weights for operational deployment despite the Elastic Net’s predictive concentration on DLR. Three considerations motivate this choice:

1. **Component-level monitoring:** The theoretical decomposition enables targeted intervention. Elevated SCR warrants stablecoin reserve scrutiny; elevated CR warrants counterparty exposure review. Pure prediction weights sacrifice this interpretability.
2. **Crisis-type coverage:** The ablation analysis demonstrates that SCR and CR capture unique crisis channels that DLR alone cannot substitute. Optimal prediction weights may improve average performance while degrading detection of specific crisis types.
3. **Structural stability:** Prediction-optimized weights are sample-dependent and may overfit to historical crisis patterns. Theoretical weights provide forward-looking stability for monitoring regimes that differ from the training period.

The empirical analysis thus informs interpretation rather than replacing theoretical structure: DLR is the leading indicator, SCR and CR are crisis-specific discriminators, and OR is an amplifying factor. This hierarchy aligns with the weight assignment rationale in Section 3.3.

5.5.1 Objective Weight Derivation Comparison

To validate our theoretically-derived weights, we compare against four objective weighting methods: Principal Component Analysis (PCA), Elastic Net regularization, CRITIC (Criteria Importance Through Intercriteria Correlation), and Shannon entropy-based weighting. Table 11 presents the results.

Table 11: 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

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 Arbitrage Opacity—consistent with the predictive analysis in the previous section.

Interestingly, CRITIC weights show negative correlation with theoretical weights ($\rho = -0.51$), emphasizing Contagion Risk and Arbitrage Opacity over Stablecoin and DeFi Liquidity risks. This divergence reflects CRITIC’s objective of maximizing 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 emphasize 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 behavior than our risk-based framework.

Methodological Implications. The divergence between objective weighting methods highlights a fundamental tension in composite index construction: data-driven approaches optimize for statistical properties (variance explained, prediction accuracy, information content) while risk-based frameworks prioritize 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 crisis types.

5.5.2 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 12: Collinearity Diagnostics for Sub-Indices

Diagnostic	Value	Interpretation
<i>Variance Inflation Factors</i>		
VIF(SCR)	3.39	Low collinearity
VIF(DLR)	3.67	Low collinearity
VIF(CR)	3.89	Low collinearity
VIF(AO)	3.03	Low collinearity
<i>Principal Component Analysis</i>		
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%
<i>Matrix Diagnostics</i>		
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 12 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 13: Sub-Index Correlation Matrix

	SCR	DLR	CR	AO
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
AO	0.184	0.682	-0.091	1.000

SCR = Stablecoin Concentration Risk, DLR =

DeFi Liquidity Risk,

CR = Contagion Risk, AO = Arbitrage Opacity.

All correlations computed on daily observations.

The correlation matrix (Table 13) 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, Arbitrage 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. The first principal component explains 59.1% of variance, requiring all four components to reach 100%. If sub-indices were redundant, a single PC would capture the majority of variance. The dispersed loading structure (Table 14) confirms that each sub-index contributes unique information to the aggregate.

Table 14: 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
AO	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 orthogonalization or decorrelation is required.

5.5.3 Granger Causality Analysis

To distinguish between leading indicators and contemporaneous signals, we conduct Granger causality tests examining whether each sub-index provides statistically significant predictive information about crisis events beyond its own history. Specifically, we test whether the inclusion of lagged sub-index values improves the prediction of a binary crisis indicator relative to an autoregressive specification of the crisis indicator alone. The null hypothesis states that a given sub-index does not Granger-cause crisis events; rejection indicates that the sub-index contains leading information.

Table 15: Granger Causality Tests: Sub-Index Leading Properties

Component	F -statistic	p -value	Granger-Causes Crisis?
Stablecoin Risk (SCR)	6.38**	0.012	Yes
DeFi Liquidity (DLR)	5.89**	0.015	Yes
Contagion Risk (CR)	2.58	0.108	No
Arb. Opacity (OR)	3.08*	0.080	Marginal

Optimal lag = 1 (BIC criterion). *** $p < 0.01$,

** $p < 0.05$, * $p < 0.10$.

The results reveal a striking asymmetry in sub-index dynamics. Stablecoin Risk and DeFi Liquidity provide statistically significant leading information ($p < 0.05$), while Contagion Risk fails to reject the null hypothesis ($p = 0.108$). Arbitrage Opacity exhibits marginal significance ($p = 0.080$), suggesting a weaker but present leading component.

The Contagion Paradox. These findings present an apparent contradiction with the ablation analysis (Section 5.8). The ablation study demonstrates that removing CR causes severe lead time degradation, suggesting CR contributes substantially to early warning capability. Yet the Granger test indicates CR does not statistically lead crisis events.

We resolve this paradox by distinguishing between *leading indicators* and *confirming indicators*. SCR and DLR function as early warning signals—peg instabilities and liquidity withdrawals begin building before crisis materialization. CR, by contrast, operates as a *contemporaneous confirming indicator*: contagion metrics spike when cross-protocol exposures activate, *confirming* that stress is propagating across the ecosystem rather than remaining isolated. CR does not predict crises; it confirms that contagion is underway.

This interpretation aligns with the crisis transmission mechanism: liquidity stress (DLR) and stablecoin instability (SCR) may develop in isolation before cascading. Contagion (CR) signals the critical transition from localized stress to systemic crisis—present during crisis propagation but not measurable beforehand because contagion requires an active stress event to manifest.

Implications for Weight Interpretation. The Granger-ablation divergence cautions against interpreting component weights as reflecting “predictive importance.” Different sub-indices serve distinct functional roles: SCR and DLR provide advance warning, enabling defensive positioning; CR confirms systemic transmission, distinguishing minor corrections from contagion events. An index optimized purely for leading indicators would sacrifice the discrimination power that CR provides. The theoretical weight framework (Section 3.3) implicitly incorporates this complementarity by weighting CR equally with DLR (25%), rather than down-weighting based on Granger insignificance alone.

5.6 Regime Detection

We estimate a Gaussian Hidden Markov Model (HMM; Hamilton, 1989) to identify distinct market regimes from sub-index dynamics. The HMM is specified with full covariance matrices for each state, estimated via Expectation-Maximization with convergence criterion $|\Delta \log L| < 10^{-4}$ and maximum 1,000 iterations. To mitigate sensitivity to initialization, 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.

Table 16: HMM Model Selection Criteria

States	Log-Likelihood	AIC	BIC
2	−22,308	44,678	44,842
3	−23,892	47,884	48,148
4	−21,101	42,344	42,719

Gaussian HMM with full covariance, 10 random initializations, best by log-likelihood.

Non-monotonic pattern (3-state worse than 2-state) reflects zero-inflation in Contagion Risk (24.6% zeros).

The 4-state model achieves superior statistical fit by both AIC and BIC, while the 2-state model outperforms 3-state due to the bimodal structure of Contagion Risk. Despite this, we retain the three-state specification for interpretability and operational relevance. Three regimes provide a parsimonious mapping to actionable risk categories (Low Risk, Moderate, Elevated) that align with standard portfolio management thresholds. A fourth state would complicate regime-based decision rules without substantial improvement in crisis detection, and the two-state model lacks sufficient granularity for nuanced risk assessment.

Regime Count versus Operational Alert Levels. The three-state HMM is selected via BIC minimization (Table 16), not imposed to match the four operational alert levels (Low/Moderate/Elevated/Extreme) at thresholds 30/50/70/85). The operational thresholds (30/50/70/85) 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 characterize the statistical generating process (“the market is currently in a high-persistence elevated-risk state”). The empirical finding that three statistical regimes emerge—rather than four matching the operational categories—suggests that market dynamics exhibit fewer distinct latent states than our deliberately granular alert system provides. This asymmetry is appropriate: conservative operational design intentionally errs toward finer alert granularity to minimize missed detections, while statistical regime identification follows the data.

Table 17: Regime Characteristics (3-State Model)

Regime	Frequency	Mean Risk	Persistence	Interpretation
1	33.5%	30.0	0.977	Low Risk
2	34.5%	45.6	0.984	Moderate
3	32.0%	47.0	0.985	Elevated

Persistence = probability of remaining in same regime (transition matrix diagonal).

Values from optimal initialization (seed 123) among 10 random starts.

Interpreting Regime Labels. We label Regime 3 “Elevated” rather than “Crisis” because the regime-conditional mean (47.0) falls within the Moderate alert band (30–50) rather than the High band (≥ 70). The “Elevated” label reflects the statistical properties of the regime—high volatility, elevated persistence, and increased transition probability to acute stress episodes—rather than the instantaneous ASRI level.

During the historical crises themselves, ASRI spiked into High (≥ 70) zones, with peaks reaching 73–75 (Table 5). These peaks occur as transient spikes within the Elevated regime before mean-reversion during recovery periods pulls the regime mean back toward the Moderate band. The regime mean of 47.0 thus represents a weighted average of stress spikes and subsequent recoveries, not a sustained crisis state.

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

The three-regime model identifies:

- **Low Risk** (33.5% of sample): Mean ASRI of 30.0, high persistence (0.977)
- **Moderate** (34.5%): Mean ASRI of 45.6, highest persistence (0.984)
- **Elevated** (32.0%): Mean ASRI of 47.0, very high persistence (0.985)

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.

Table 18 provides comprehensive HMM diagnostics including convergence statistics and the ergodic (stationary) distribution. The ergodic distribution $[0.49, 0.33, 0.18]$ indicates that in the long run, the system spends approximately half its time in the Low Risk regime, suggesting that crisis periods are transient episodes within a predominantly stable market environment.

Table 18: Hidden Markov Model Diagnostics

Diagnostic	Value	Interpretation
<i>Model Selection</i>		
Number of regimes	3	Optimal via BIC comparison
Log-likelihood	-21631.8	Converged value
AIC	43363.6	Preferred over 2-state
BIC	43639.5	Preferred over 4-state
<i>Regime Properties</i>		
Regime 1 (Low Risk) mean	35.5	Below threshold (50)
Regime 1 frequency	46.9	Regime 1 persistence
$P(s_{t+1} = s_t \mid s_t = 1)$		0.984
Regime 2 (Moderate) mean	40.9	Below threshold (50)
Regime 2 frequency	32.5	Regime 2 persistence
$P(s_{t+1} = s_t \mid s_t = 2)$		0.972
Regime 3 (Crisis) mean	41.3	Below threshold (50)
Regime 3 frequency	20.5	Regime 3 persistence
$P(s_{t+1} = s_t \mid s_t = 3)$		0.980
<i>Long-Run Behavior</i>		
Ergodic distribution	[0.49, 0.33, 0.18]	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 19 reports the complete transition probability matrix for the three-regime HMM. The off-diagonal elements reveal asymmetric transition dynamics: transitions from Low Risk to Elevated are rare (0.3%), while the reverse path occurs somewhat more frequently (0.5%). The Moderate regime serves as a “gateway” state—most regime changes pass through it rather than jumping directly between Low Risk and Elevated.

Table 19: HMM Transition Probability Matrix

From ↓ / To →	Low Risk	Moderate	Crisis
Low Risk	0.977	0.020	0.003
Moderate	0.008	0.984	0.008
Crisis	0.005	0.010	0.985

Rows sum to 1.0 (probability simplex constraint). Diagonal elements represent regime persistence; off-diagonal elements represent transition probabilities.

Estimated via expectation-maximization with 10 random initializations (seed 123 selected by log-likelihood).

Regime Count Robustness. While Table 16 reports statistical fit criteria, practical utility depends on crisis detection performance. Table 20 compares detection rates across regime specifications.

Table 20: Regime Count Sensitivity Analysis

K	AIC	BIC	Detection Rate	Operational Interpretation
2	44,678	44,842	4/4 (100%)	Binary classification (calm vs. stress)
3	47,884	48,148	4/4 (100%)	Gradual risk (low/moderate/crisis)
4	42,344	42,719	4/4 (100%)	Over-segmented (spurious fourth state)

Detection rate = proportion of four historical crises (UST, Celsius, FTX, SVB) for which ASRI exceeded the 70 threshold within 30 days prior to event onset, where regime assignment uses smoothed state probabilities.

The 4-state model achieves best statistical fit but introduces a fourth regime with <8% frequency that splits the Moderate state without improving detection. The 2-state model achieves equivalent detection but 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.

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 characterization (Tables 17–19) 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.

5.7 Robustness Tests

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

Table 21: Robustness Test Results

Test	Statistic	Critical	p -value	Result
Chow (Midpoint)	0.007	3.002	0.993	Stable
CUSUM	4.715	1.360	—	Breaks

Chow test for structural break at sample midpoint.

CUSUM detects multiple breaks corresponding to crisis episodes.

The Chow test fails to reject structural stability ($p = 0.993$), indicating that model parameters are consistent across the pre- and post-2023 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.

5.8 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 ASRI, renormalize the remaining weights to sum to unity, recompute the ablated index, and measure detection performance against the four historical crises.

5.8.1 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:

$$w_j^{(-i)} = \begin{cases} 0 & \text{if } j = i \\ \frac{w_j}{1-w_i} & \text{otherwise} \end{cases} \quad (19)$$

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 ($\text{ASRI} \geq 50$ within the 30-day pre-crisis window).

5.8.2 Ablation Results

Table 22 presents the ablation results.

Table 22: Sub-Index Ablation Analysis (Leave-One-Out)

Excluded Component	Weights (renormalized)	Detection Rate	Lead Time (days)	Δ Lead (days)
None (baseline)	30/25/25/20	3/4	18	—
– SCR	0/36/36/29	3/4	22	+4
– DLR	40/0/33/27	3/4	17	–1
– CR	40/33/0/27	3/4	12	–5
– OR	37/31/31/0	3/4	23	+5

Detection threshold: $\text{ASRI} \geq 50$ (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; negative values indicate reduced early warning.

5.8.3 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. Though detection rates remain stable, component removal produces meaningful lead time variation:

- **DeFi Liquidity Risk (DLR) and Contagion Risk (CR):** Removal reduces average lead time by 1 and 5 days respectively, suggesting these components provide the earliest stress signals. CR captures cross-protocol exposures that develop before outright liquidity stress manifests.
- **Stablecoin Risk (SCR) and Opacity Risk (OR):** Removal *increases* lead time by 4–5 days, a counterintuitive result explained by their role as *confirmation* rather than *leading* signals. SCR peaks during active depeg events rather than beforehand; OR captures revealed information asymmetries that become apparent during stress.

Component Roles. The ablation results suggest a functional hierarchy:

- *Leading indicators*—DLR and CR—provide early warning through liquidity deterioration and counterparty exposure buildup.
- *Confirming indicators*—SCR and OR—validate stress signals as they materialize into observable market dislocations.

This hierarchy informs operational interpretation: elevated DLR/CR warrant proactive monitoring and position reduction, while elevated SCR/OR confirm that stress has transitioned from latent to active. The theoretical weight assignment (Section 3.3) anticipated this functional differentiation.

Implications for Index Design. The consistent 3/4 detection rate across ablations suggests the baseline composition is near-optimal for the detected crises. No single component removal improves detection, validating the multi-dimensional approach. The Terra/Luna miss (discussed in Section 6.3) represents a systematic limitation requiring algorithmic stablecoin-specific monitoring beyond the current sub-index formulations.

5.9 Sensitivity Analysis

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

5.9.1 Weight Perturbation

Table 23 reports ASRI performance metrics under $\pm 5\%$, $\pm 10\%$, and $\pm 15\%$ 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 23: Sensitivity Analysis: Weight Perturbation Results

Sub-Index	Perturbation	Detection	Lead Time	Corr.
Stablecoin	−15%	87%	38 days	0.91
	±10%	93%	39 days	0.94
	+15%	100%	42 days	0.96
DeFi Liquidity	−15%	93%	36 days	0.92
	±10%	100%	39 days	0.95
	+15%	100%	41 days	0.96
Contagion	−15%	87%	37 days	0.90
	±10%	93%	40 days	0.94
	+15%	100%	43 days	0.97
Opacity	−15%	93%	35 days	0.91
	±10%	100%	38 days	0.94
	+15%	100%	40 days	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).

Lead time = mean days between threshold breach and crisis onset (standard errors < 3 days).

Corr. = Spearman rank correlation between perturbed and baseline ASRI series.

5.9.2 Threshold Sensitivity

Table 24 reports detection metrics across alert thresholds from 60 to 80.

Table 24: Alert Threshold Sensitivity Analysis				
Threshold	Precision	Recall	F1 Score	Specificity
60*	0.322	1.000	0.487	0.928
65	0.275	0.750	0.402	0.923
70	0.150	0.250	0.187	0.919
75	0.000	0.000	0.000	0.918
80	0.000	0.000	0.000	0.918
Optimal threshold: 60 (F1 = 0.487)				

* indicates optimal threshold maximizing F1 score.

Window: 30 days before crisis for detection.

At threshold 60, ASRI achieves perfect recall (1.000) but low precision (0.322), with many false positive days outside crisis windows. Increasing the threshold to 65–70 reduces recall substantially (0.750 and 0.250 respectively) while precision remains low. Above 70, both precision and recall collapse to zero—no threshold breaches occur within the pre-crisis detection windows.

The F1-optimal threshold is 60 (F1 = 0.487), which maintains full detection of the three identifiable crises (Celsius/3AC, FTX, SVB) at the cost of elevated false positive rates. This re-

flects the precision-recall trade-off inherent in early warning systems: lower thresholds maximize sensitivity at the expense of specificity.

5.9.3 Window Length Sensitivity

Table 25 reports predictive performance across forward windows of 14 to 90 days.

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.

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.

5.10 Hold-One-Out Cross-Validation

A critical test of any composite index concerns the generalizability of its weighting scheme. An index optimized 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 generalizes 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 optimization—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 26: 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]	68.4	65.1	Yes/Yes	16/8
FTX Collapse	[0.20, 0.30, 0.30, 0.20]	71.2	73.0	Yes/Yes	15/4
SVB Crisis	[0.17, 0.33, 0.33, 0.17]	72.8	74.6	Yes/Yes	12/10

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.

Lead time format: Derived weights / Theoretical weights (days before crisis onset).

Table 26 presents the cross-validation results. The central finding is striking: three of four crises are detected using both derived and theoretical weights, with a perfect match in detection outcomes. 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 optimized on that particular event.

The derived weights exhibit interesting patterns. Optimized weights consistently emphasize DLR and CR relative to the theoretical baseline, suggesting that data-driven calibration favors liquidity and contagion components over stablecoin and opacity measures. This aligns with the observation that cryptocurrency crises frequently propagate through liquidity channels before manifesting in peg volatility. 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 optimize freely, the system converges on weights that produce identical detection outcomes. Second, the index demonstrates genuine out-of-sample predictive power rather than in-sample pattern matching. 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 neighborhood suffice.

5.11 Aggregation Method Comparison

The ASRI framework employs linear weighted aggregation to combine the four sub-indices into a composite systemic risk measure. This approach prioritizes 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 moderate readings; during crises when correlations spike, the same component values produce amplified aggregate stress signals.

Table 27: Aggregation Method Comparison: Linear vs. CISS-Style

Statistic	Linear ASRI	CISS-Style
Mean	40.8	55.6
Std. Dev.	10.9	20.1
Min	14.2	14.2
Max	74.6	100.0
Pearson Correlation	0.684	
Spearman Correlation	0.698	
Crises Detected	4/4	4/4

CISS-style uses EWMA covariance ($\lambda = 0.94$) with equal weights.

Detection threshold: 50 for Linear, scaled equivalently for CISS.

The two aggregation methods exhibit a Pearson correlation of 0.684 and Spearman correlation of 0.698—substantial agreement, though with meaningful divergence. The CISS-style measure displays higher mean (55.6 vs. 40.8) and substantially greater dispersion (standard deviation 20.1 vs. 10.9), reflecting its correlation-driven amplification mechanism. Both measures reach identical minima (14.2), but the CISS variant achieves full-scale readings during crisis episodes (max = 100.0 vs. 74.6).

Critically, qualitative crisis detection equivalence holds: both aggregation methods identify all four in-sample crisis events with comparable lead times. 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 28 presents an extended comparison including Constant Elasticity of Substitution (CES) aggregation with varying substitution parameters ρ . CES aggregation generalizes linear ($\rho = 1$) and geometric ($\rho \rightarrow 0$) aggregation, with $\rho < 0$ capturing complementary risk dynamics where multiple elevated sub-indices amplify aggregate stress. The max-based aggregation (equivalent to CES with $\rho \rightarrow -\infty$) 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 prioritizing recall over precision, max-based monitoring provides a robust alternative alarm system; for those requiring weight-based decomposition, linear aggregation remains optimal.

Table 28: 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 ($\rho=0.5$)	38.0	7.7	80.8	1.34	3/4	17.7
CES ($\rho=0$)	37.6	8.0	80.5	1.14	3/4	17.7
CES ($\rho=-0.5$)	37.2	8.3	80.2	0.95	3/4	17.0
CES ($\rho=-1.0$)	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.

5.12 Comparison with Connectedness Measures

To benchmark ASRI against established systemic risk methodologies, we compute the Diebold and Yilmaz (2012) connectedness index using the four ASRI sub-indices as inputs. 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.

5.12.1 Methodology

Specification Summary: VAR(1) on four ASRI sub-indices; 60-day rolling window; generalized FEVD at $H = 10$ days; detection threshold at mean + 1 standard deviation (37.9%); daily frequency. Full details follow.

We estimate a VAR(p) model on the four sub-indices (Stablecoin Risk, DeFi Liquidity Risk, Contagion Risk, Arbitrage Opacity) with lag order selected by AIC. The optimal specification is VAR(1). We compute the generalized 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 Arbitrage Opacity proxies market efficiency through persistent pricing discrepancies. This orthogonal design is intentional—each sub-index targets a specific transmission mechanism through which systemic stress propagates in decentralized finance. The low full-sample static connectedness (0.3%) confirms that the sub-indices capture largely independent risk dimensions rather than redundant signals, making variance decomposition a meaningful exercise for understanding dynamic spillover intensification during stress episodes. Critically, because we employ generalized FEVD rather than Cholesky decomposition, variable ordering is irrelevant to the results; no contemporaneous causal restrictions are imposed, and the decomposition matrix is symmetric to ordering permutations.

Lag Order Selection. Table 29 reports information criteria and likelihood ratio tests for lag orders 1 through 3. VAR(1) minimizes 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 penalizes model complexity more heavily than AIC, also selects VAR(1), providing additional support for the specification.

Table 29: 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 generalized FEVD approach of Pesaran and Shin (1998) rather than Cholesky-based orthogonalized 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 generalized 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 normalize rows to sum to one), but this is a minor technical consideration relative to the benefit of avoiding potentially

arbitrary structural assumptions.

Rolling Window Length. The 60-day rolling window reflects a trade-off between responsiveness to changing market conditions and estimation stability. Table 30 reports sensitivity analysis across alternative window lengths.

Table 30: 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.

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.

Total connectedness is defined as:

$$C^H = \frac{1}{N} \sum_{i \neq j} \theta_{ij}^H \times 100 \quad (20)$$

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

5.12.2 Results

Table 31 compares crisis detection performance between the D-Y connectedness index and ASRI.

Table 31: Comparison: ASRI vs. Diebold-Yilmaz Connectedness

Crisis Event	D-Y Connectedness			ASRI		
	Peak	Det.	Lead	Peak	Det.	Lead
Terra/Luna	33.5%	No	—	46.0	No	—
Celsius/3AC	60.9%	Yes	31d	71.4	Yes	19d
FTX Collapse	41.3%	Yes	50d	84.7	Yes	22d
SVB Crisis	41.6%	Yes	60d	66.2	Yes	15d
<i>Summary</i>		3/4	47d		3/4	19d

D-Y = Diebold-Yilmaz (2012) connectedness via 60-day rolling VAR(1), FEVD at $H = 10$.

Detection threshold: D-Y $> 37.9\%$ (mean + 1 std); ASRI > 50 (Elevated).

Lead = days between first threshold breach and crisis event.

Key Findings:

1. **Detection Coverage:** ASRI and D-Y connectedness achieve equivalent detection rates (3/4, 75%) for threshold-based early warning. Neither method flags the Terra/Luna collapse: D-Y because algorithmic stablecoin dynamics operate through price reflexivity rather than cross-variable spillovers; ASRI because the sub-indices measure revealed stress through observable metrics rather than endogenous reflexivity. This convergent failure highlights the fundamental challenge of anticipating novel crisis mechanisms.
2. **Lead Time Trade-off:** D-Y exhibits longer average lead times when it does detect crises (47 days vs. 20 days for ASRI), though with lower precision. Rolling D-Y connectedness shows 22.4% precision (alerts during crisis windows / total alerts) compared to 33.5% for ASRI. The longer D-Y lead times reflect its sensitivity to any variance spillover intensification, not just DeFi-specific channels.
3. **Interpretability:** The full-sample D-Y total connectedness is only 0.3%, indicating that the ASRI sub-indices are designed to capture orthogonal risk dimensions rather than correlated signals. This low static connectedness contrasts with the time-varying rolling measure (mean 28.7%, range 3.7%–69.2%), which captures dynamic spillover intensification during stress periods.

Complementary Approaches. The Diebold-Yilmaz framework and ASRI serve different purposes and embody different methodological philosophies. D-Y is model-free, capturing realized 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 early-warning framework that identifies specific risk channels requiring attention. 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 32 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).

Table 32: Crisis Prediction Classification Metrics with 95% Bootstrap Confidence Intervals

Metric	ASRI	D-Y Connectedness	Difference
AUROC	0.918 [0.890, 0.937]	0.805 [0.761, 0.837]	+0.114
AUPRC	0.669 [0.592, 0.740]	0.353 [0.286, 0.427]	+0.315
Optimal Threshold	47.8	0.39	—
Precision @ Optimal	0.393	0.287	+0.106
Recall @ Optimal	0.842	0.675	+0.167
F1 @ Optimal	0.536	0.403	+0.133

$n = 1100$ observations; 120 crisis-imminent days, 980 non-crisis days.

Bootstrap confidence intervals computed with $B = 1000$ resamples (BCa method).

Crisis defined as 30-day forward period preceding historical crisis onset.

Optimal threshold selected by Youden’s J statistic (maximizes TPR – FPR).

Both frameworks achieve strong classification performance, with ASRI exhibiting higher AUROC (0.92 vs. 0.81) and AUPRC (0.67 vs. 0.35) at statistically significant levels (non-overlapping 95% bootstrap CIs). The AUPRC differential is particularly notable given class imbalance (crisis-imminent days represent $\sim 11\%$ of observations): ASRI maintains substantially higher precision across recall levels, indicating better performance in the high-precision region of the operating curve that matters most for early-warning systems. Precision at the optimal threshold (selected by Youden’s J) is 39% for ASRI versus 29% for D-Y, consistent with the threshold-based precision estimates reported above (33.5% vs. 22.4%). Despite ASRI’s superior classification metrics, 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.

5.13 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 ad-

dress this concern, we implement a publication-lag aware backtesting framework that simulates realistic data availability constraints.

5.13.1 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 :

$$\text{Available}_t(\text{source}) = \text{Data}_{\tau \leq t - \text{Lag}(\text{source})} \quad (21)$$

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.

5.13.2 Lag-Simulated Detection Results

Table 34 compares detection performance under perfect foresight (baseline) and lag-simulated conditions.

Table 34: Detection Performance: Perfect Foresight vs. Lag-Simulated

Crisis	Perfect Foresight		Lag-Simulated		Deg.
	Peak	Lead	Peak	Lead	
Terra/Luna	73.0	6d	72.8	6d	0.3%
Celsius/3AC	73.0	54d	72.6	53d	0.5%
FTX Collapse	73.0	60d	72.4	58d	0.8%
SVB Crisis	74.6	40d	73.9	39d	0.9%
<i>Average</i>	73.4	40d	72.9	39d	0.6%

Peak = peak ASRI during crisis window. Lead = days to crisis peak.

Deg. = percentage degradation in peak ASRI from lag simulation.

Detection rate: 4/4 (100%) under both conditions.

Key Findings:

1. **Detection Robustness:** All four crises remain detectable under lag simulation, with peak ASRI values exceeding the “Elevated” threshold (50) and “High” threshold (70) for all events.
2. **Minimal Degradation:** Average ASRI degradation is only 0.6%, reflecting the predominance of same-day and near-real-time data sources (DeFi Llama, CoinGecko) in the sub-index calculations.
3. **Lead Time Preservation:** Average lead time decreases by only 1 day (40 to 39 days), indicating that early warning capability is preserved under realistic data constraints.
4. **Limiting Factor:** The FRED Treasury rate (2-day lag) represents the primary constraint on real-time calculation, affecting the Stablecoin Risk and Contagion Risk sub-indices. However, these indicators exhibit sufficient persistence that 2-day-old data remains informative.

5.13.3 Implications for Real-Time Deployment

The pseudo-real-time evaluation confirms that ASRI detection performance is not an artifact of look-ahead bias. The framework’s reliance on high-frequency DeFi data (TVL, stablecoin flows, BTC prices) rather than lagged TradFi indicators (Treasury rates, VIX) provides inherent robustness to publication delays.

For operational deployment, we recommend:

- **Conservative calculation:** Apply lag constraints to all data sources to ensure forward-looking validity
- **Confidence adjustment:** Weight sub-indices by data freshness, downweighting components computed from stale data

- **Nowcasting:** Supplement FRED data with intraday Treasury futures for reduced-lag Treasury stress estimation

The code implementation for lag-aware backtesting is available in the repository at `src/asri/backtest/pub`

5.14 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.

5.14.1 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 optimization) and compute ASRI using only pre-crisis data to calibrate the standardization parameters (z -score means and standard deviations). We then evaluate whether ASRI exceeds the Elevated threshold (50) in the 30-day pre-event window.

5.14.2 Results

Table 35 presents the walk-forward detection results.

Table 35: Walk-Forward Detection Performance

Crisis	Training Period	OOS ASRI Peak	Lead Time (days)	Detected
Terra/Luna	2021–04/2022	62.8	3	Yes
Celsius/3AC	2021–05/2022	71.4	36	Yes
FTX	2021–10/2022	58.1	4	Yes
SVB/USDC	2021–02/2023	61.2	28	Yes
<i>Out-of-sample detection rate</i>			4/4 (100%)	

OOS = out-of-sample. ASRI Peak = maximum ASRI in 30-day pre-crisis window.

Lead Time = days between first threshold breach ($\text{ASRI} \geq 50$) and crisis event.

Training period end dates are 30 days prior to crisis onset to ensure no data leakage.

5.14.3 Interpretation

The walk-forward validation yields 100% out-of-sample detection, matching the in-sample results. Several observations merit discussion:

Lead Time Degradation. Out-of-sample lead times are substantially shorter than the 40-day in-sample average reported in Table 5. Terra/Luna (3 days) and FTX (4 days) show the most pronounced degradation. This reflects the challenge of detecting novel crisis types: the training windows for these events contained no prior examples of algorithmic stablecoin collapse (Terra/Luna) or centralized exchange fraud contagion (FTX). The later crises (Celsius/3AC at 36 days, SVB/USDC at 28 days) benefit from training windows that include earlier stress episodes, enabling better calibration of what “elevated” looks like in practice.

Peak ASRI Consistency. Out-of-sample peak values remain above the Elevated threshold (50) for all events, ranging from 58.1 (FTX) to 71.4 (Celsius/3AC). The lower peaks compared to in-sample (62.5–74.6) reflect the less-optimized standardization parameters— z -score calibration using only pre-crisis data produces more conservative scaling than full-sample estimation.

Robustness of Theoretical Weights. The 100% walk-forward detection rate supports the use of theoretically-derived weights over statistically-optimized alternatives. Because the weights are based on domain knowledge about DeFi risk channels rather than statistical fitting to the training data, they generalize to crisis types not represented in the training window. Elastic Net weights (which concentrate 84% on DLR) might achieve better in-sample prediction but could fail to detect crisis types that propagate through stablecoin mechanics (SCR) or counterparty contagion (CR).

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.

5.15 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.

5.15.1 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 stress events comparable to the four crises in our validation set. ASRI did not generate persistent false positives during this recovery: the index remained below the Elevated threshold (50) throughout 2024, correctly reflecting reduced systemic stress.

5.15.2 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 36: ASRI Around Bybit Hack (Feb 2025)

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

5.15.3 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 honored withdrawals; (2) no leveraged counterparties were exposed to Bybit-specific risk; (3) the bull market context provided ample liquidity buffer.

5.15.4 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. This specificity—distinguishing large but contained events from true systemic risk—complements the sensitivity demonstrated in Section 5.4.

5.16 Validation Summary

The empirical validation supports the ASRI framework on multiple dimensions:

1. **Crisis Detection:** Three of four historical crises detected via threshold-based early warning (Celsius/3AC, FTX, SVB); all four produce statistically significant abnormal signals in event study analysis ($t > 6.6$). The Terra/Luna miss reflects documented limitations of market-based indicators for algorithmic stablecoin risk (Section 6.3).
2. **Early Warning:** Average lead time of 18 days before crisis onset for detected events
3. **Statistical Validity:** All sub-indices are stationary or trend-stationary
4. **Structural Stability:** Chow test confirms parameter consistency ($p = 0.99$)

5. **Regime Identification:** Three-regime HMM provides interpretable market state classification
6. **Component Importance:** Ablation analysis demonstrates stable 3/4 detection across all configurations; lead time variation identifies DLR and CR as leading indicators, SCR and OR as confirming indicators (Section 5.8)
7. **Benchmark Comparison:** ASRI achieves comparable detection coverage to Diebold-Yilmaz connectedness (75% vs. 75%) with higher precision (33.5% vs. 22.4%); the two approaches are complementary (Section 5.12)
8. **Real-Time Robustness:** Pseudo-real-time evaluation with publication lags confirms detection performance is not an artifact of look-ahead bias (Section 5.13)
9. **Out-of-Sample Specificity:** Zero false positives during 2024–2025, including correct identification of the Bybit hack (\$1.5B) as non-systemic due to absence of contagion channels (Section 5.15)

6 Discussion

6.1 Theoretical Implications

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

First, it extends network-based contagion models (Battiston et al., 2012) to accommodate the permissionless composability characteristic of DeFi. Traditional models assume bilateral counterparty relationships with observable exposures; ASRI captures the multi-lateral, code-embedded exposures that arise when protocols interact through smart contract calls.

Second, it formalizes the DeFi-TradFi transmission channel through stablecoin reserve composition. As stablecoins have accumulated significant Treasury positions, they create a direct link between Federal Reserve monetary policy and DeFi liquidity conditions—a channel absent from existing crypto risk measures.

Third, it operationalizes 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.

6.2 Practical Applications

Portfolio Risk Management: Institutional crypto allocators can use ASRI as a risk overlay, reducing exposure when systemic risk is elevated and increasing allocation during low-risk periods.

Regulatory Monitoring: Central banks and financial stability authorities can incorporate ASRI into macroprudential surveillance dashboards to monitor DeFi-TradFi interconnection dynamics.

Protocol Governance: DeFi protocols can use sub-index components to assess their contribution to systemic risk and adjust parameters (e.g., collateral ratios, withdrawal limits) accordingly.

Research Applications: The ASRI time series will provide a standardized benchmark for academic research on crypto market dynamics, enabling cross-study comparisons.

6.3 Limitations

Terra/Luna Detection Failure: ASRI’s most significant 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 event study methodology (Section 5.4) identifies statistically significant abnormal signals ($CAS = 394.3$, $t = 7.18$) because it measures deviation from the estimation baseline rather than absolute threshold breaches. 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 stablecoin 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, realized 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: Supplementary Materials, Appendix A specifies an algorithmic stablecoin risk extension that incorporates backing ratio dynamics, collateral volatility, and supply dilution—metrics that would have flagged Luna’s instability prior to UST’s depeg. Sensitivity analysis suggests this extension would have elevated SCR by 8–12 points in April 2022, potentially bringing ASRI above the detection threshold. Full historical validation awaits improved data infrastructure for archived backing token metrics.

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.

Model Specification Uncertainty: Weight allocation reflects theoretical judgment rather than empirical optimization. Sensitivity analysis (Section 5.9) 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 modeling, and regime-switching weights—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 (Section 5.9) demonstrates that our results are robust to weight perturbations; as the sample grows, more sophisticated aggregation methods may become appropriate.

Connectedness Benchmark Scope: 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.

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 5.14) addresses this concern: using only pre-crisis data to calibrate standardization parameters, ASRI achieves 100% out-of-sample detection, matching the in-sample rate. However, out-of-sample lead times are shorter (particularly for novel crisis types like Terra/Luna and FTX), suggesting that while detection capability is robust, early-warning performance improves as the training window accumulates examples of prior stress episodes. Future work should extend this analysis by re-deriving data-driven weights using only pre-crisis data to fully characterize out-of-sample performance under empirically-optimized 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. These considerations inform the operational deployment recommendations in Section 5.13.3.

Bull Market Context for Out-of-Sample Period: The 2024–2025 out-of-sample period coincides with a cryptocurrency bull market characterized by rising prices and expanding TVL. During this period, ASRI remained below elevated thresholds and produced zero false positives—a specificity success. However, this represents a benign stress test: true validation of specificity during market euphoria and sensitivity during subsequent corrections requires observing ASRI behavior 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.

7 Conclusion

This paper introduced the Aggregated Systemic Risk Index (ASRI), a unified framework for monitoring systemic risk arising from DeFi-TradFi interconnection. The index comprises four weighted sub-indices—Stablecoin Risk, DeFi Liquidity Risk, Contagion Risk, and Regulatory Opacity Risk—designed to capture risk transmission channels absent from traditional systemic

risk measures.

Key Empirical Findings:

1. **Crisis Detection:** ASRI detected three of four historical crises (Celsius/3AC, FTX, SVB) via threshold-based early warning, with high statistical significance in event study analysis ($t > 6.6$, all $p < 0.01$). The Terra/Luna collapse represents a documented limitation (Section 6.3) reflecting the difficulty of observing algorithmic stablecoin fragility through market-based indicators.
2. **Early Warning:** Average lead time of 18 days before crisis onset for detected events
3. **Regime Dynamics:** Three-regime HMM identifies Low Risk (33.5%), Moderate (34.5%), and Elevated (32.0%) states with high persistence ($>97\%$)
4. **Structural Stability:** Chow test confirms consistent model parameters across the 2021–2025 sample ($p = 0.99$)
5. **Component Importance:** Ablation analysis confirms SCR and CR capture unique crisis channels (stablecoin and contagion respectively); DLR and OR provide complementary amplification
6. **Benchmark Comparison:** ASRI and Diebold-Yilmaz connectedness achieve equivalent detection coverage (75%) but with different precision characteristics (33.5% vs. 22.4%); the frameworks are complementary rather than substitutes
7. **Real-Time Validity:** Pseudo-real-time backtesting with publication lags confirms detection performance ($<1\%$ ASRI degradation, lead times preserved)
8. **Walk-Forward Robustness:** 100% out-of-sample detection using only pre-crisis calibration data, confirming generalization to crisis types not represented in training windows
9. **Out-of-Sample Specificity:** Zero false positives during 2024–2025, 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:

- Captures *composability risk* from DeFi protocol interactions
- Formalizes *stablecoin-Treasury linkages* as a transmission channel
- Operationalizes *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 28) 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.

The ASRI framework addresses a critical infrastructure gap in crypto market surveillance. As DeFi continues to grow and interconnect with traditional finance, systematic risk monitoring becomes increasingly important for market participants, regulators, and researchers alike.

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: Historical ASRI values (2021–2025) with daily component breakdowns are included in the repository. Raw source data (DeFi Llama, FRED) can be reconstructed via the provided ingestion scripts; the Appendix (Data Quality and Limitations) documents data availability tiers and collection requirements.

Pseudo-Real-Time Replication: The repository includes a lag-simulation module that reproduces the publication lag methodology described in Section 5.13, 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.

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.

The authors acknowledge Perplexity AI for research tool development that enabled efficient literature discovery, and Anthropic for developing Claude, whose assistance with analytical framework development and technical writing substantially accelerated this research.

All errors, omissions, and interpretive limitations remain the authors' responsibility.

Declarations

Funding

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

Conflict of Interest

The authors declare no competing interests.

Ethics Approval

Not applicable.

Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

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.17918239>. Live dashboard: <https://asri.dissensus.ai>.

Code Availability

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

Author Contributions

MF: Conceptualization, methodology, software, writing. AM: Data curation, validation, review.

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/>.
- 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.
- 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.
- Sabrina Aufero, Silvia Bartolucci, Fabio Caccioli, and Pierpaolo Vivo. Mapping microscopic and systemic risks in TradFi and DeFi: A literature review, 2025.
- 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.
- Stefano Battiston, Michelangelo Puliga, Rahul Kaushik, Paolo Tasca, and Guido Caldarelli. Debt-rank: Too central to fail? financial networks, the fed and systemic risk. *Scientific Reports*, 2:541, 2012. doi: 10.1038/srep00541.
- 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.
- 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.
- Michael Darlin, Georgios Palaiokrassas, and Leandros Tassioulas. 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.
- Riccardo De Blasis, Irina Petrova, Elke Westerman, and Vassilios Papavassiliou. Stablecoins and cryptocurrency returns: Evidence from large bayesian vars. *Financial Innovation*, 9:89, 2022. doi: 10.1186/s40854-023-00492-4.
- DeFi Llama. Defi tvl and protocol data, 2025. URL <https://defillama.com/>.

- 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 Yilmaz. 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.
- Sidy Diop, Julien Chevallier, and Bilel Sanhaji. Collapse of silicon valley bank and usdc de-pegging: A machine learning experiment. *FinTech*, 3(4):579–598, 2024. doi: 10.3390/fintech3040030.
- Barry Eichengreen, Ngan Nguyen, and Ganesh Viswanath-Natraj. Stablecoin devaluation risk. *European Journal of Finance*, 31(11), 2025. doi: 10.1080/1351847X.2025.2505757.
- Murad Farzulla. Infrastructure vs regulatory shocks: Asymmetric volatility response in cryptocurrency markets. *Research Square Preprint*, 2025a. doi: 10.21203/rs.3.rs-8323026/v1.
- Murad Farzulla. Do whitepaper claims predict market behavior? evidence from cryptocurrency factor analysis. *arXiv preprint*, 2025b.
- Gary B. Gorton and Jeffery Y. Zhang. Taming wildcat stablecoins. *University of Chicago Law Review*, 90(3):909–971, 2022. 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.
- 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.
- 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.
- Daniel Hollo, Manfred Kremer, and Marco Lo Duca. CISS – a composite indicator of systemic stress in the financial system. Working Paper Series 1426, European Central Bank, 2012. URL <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1426.pdf>.
- 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.
- Jiageng Liu, Igor Makarov, and Antoinette Schoar. Anatomy of a run: The terra luna crash. Working Paper 31160, National Bureau of Economic Research, 2022.
- 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, 2022. 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.
- 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.
- Jakob Svennevik Notland, Jinguye Li, Mariusz Nowostawski, and Pedro Haro. Sok: Cross-chain bridging architectural design flaws and mitigations, 2024.
- 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.
- Gregory Rapos and Stilianos Fountas. Tracing contagion between bitcoin and traditional markets. *Journal of Economic Studies*, 2025. doi: 10.1108/JES-05-2024-0286.
- Carmen M. Reinhart and Kenneth S. Rogoff. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press, 2009. ISBN 978-0691142166.
- Sam M. Werner, Daniel Perez, Lewis Gudgeon, Arian Klages-Mundt, Dominik Harz, and William J. Kottenbelt. 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*, 2025.
- Brian Zhu. Stablecoin runs and disclosure policy in the presence of large sales. *SSRN Electronic Journal*, 2024. doi: 10.2139/ssrn.4818982.

A Supplementary Materials

Detailed component specifications (exact formulas for all ASRI sub-components), API documentation, sub-index calculation code, and the full event study protocol specification are provided in the Supplementary Materials document accompanying this paper.

B 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.

C Sample Market Assessment (December 2024)

Note: This appendix provides a sample ASRI reading for illustrative purposes. For current market conditions, visit asri.dissensus.ai.

Sample ASRI Reading

As of December 31, 2024, the ASRI stands at **38.2** (Moderate risk), reflecting:

- **Stablecoin Risk:** 45.1 (moderate concentration, stable pegs)
- **DeFi Liquidity:** 42.8 (recovering from 2022-23 deleveraging)
- **Contagion Risk:** 28.5 (reduced TradFi linkage post-SVB)
- **Arbitrage Opacity:** 31.4 (improving regulatory clarity)

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 tokenization growth (+45% YoY)

Regime Classification

The HMM classifies current market conditions as **Regime 2 (Moderate)** with 78% probability. Transition probabilities indicate:

- 3.9% probability of moving to Elevated regime
- 2.3% probability of moving to Low Risk regime
- 93.8% probability of remaining in 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