Journal of Derivatives and Quantitative Studies: 선물연구

From Dot-com to AI: drawdown onsets prediction in AI technology stock market

Maksim Kim and Seungwook Park
Department of Business Administration, Inha University, Incheon, South Korea

Received 13 March 2026
Revised 7 May 2026
4 June 2026
22 June 2026
Accepted 5 July 2026

Abstract

This paper introduces the Iterative Market Understanding and Generation Inference (IMUGI) framework to anticipate the onset of major drawdowns in AI technology stocks at their peak, rather than identifying them once the decline has already unfolded and the hedging window has closed. The study's purpose is to establish whether drawdown-onset prediction with actionable lead time is feasible by combining multiple crisis mechanisms into a unified online probabilistic architecture, rather than relying on the single-mechanism specifications common in the existing literature. IMUGI is a three-layer probabilistic architecture comprising: (1) a Second-Order Hidden Markov Model (SOHMM) with Student-t emissions and Missing-Not-At-Random (MNAR) handling for latent regime inference; (2) a Quadratic Hawkes process for endogenous feedback and extreme-return clustering; and (3) a regularised logistic meta-model that fuses both signals into a daily drawdown-onset probability. Daily price data for 27 NASDAQ-listed AI and technology firms spanning 2003–2025 were retrieved via the yfinance API. The framework is evaluated on a strictly held-out test set of 999 trading days (November 2021 – October 2025), with predictions generated incrementally to prevent look-ahead bias. The framework attained an AUC-ROC of 0.9055 and a Brier score of 0.1175 on the held-out test set. All three drawdown onsets in the test period are detected: two are flagged with advance warning of five and one trading days, respectively, and one is identified within two trading days of onset. The recency of the AI technology sector as a distinct investable field limits the number of observable drawdown onsets; the three events in the test window should be read as statistically constrained but consistent out-of-sample evidence rather than a comprehensive test of generalisation.

Keywords Drawdown onset prediction, Early-warning system, Tail risk management, Hedging strategies, Regime-switching models, Quadratic Hawkes process, Second-order hidden Markov model, AI technology stocks, Probabilistic financial forecasting

Paper type Research article

1. Introduction

Financial markets periodically experience sharp and apparently sudden changes (Ang and Timmermann, 2012), with multiple supposedly rare and extreme events recurring over the past 2 decades. The question at hand is therefore not how often those events occur, but whether they are genuinely unpredictable or possible to anticipate, specifically in the accelerating AI technology sector.

Predicting these shifts remains a complicated endeavour. Most approaches frame it as binary classification, which overlooks distributional shifts (Guan and Tibshirani, 2022) and class imbalance in financial networks (An *et al.*, 2025). A signal in the vicinity of the drawdown starting point is operationally just as valuable for hedging and position management as pinpointing the exact onset, yet standard metrics penalise early warnings equally with misses.

JEL Classification — G17, C58, C53, G01, G13, C41

Funding: This study is supported by Inha University research fund (award id: 2025-2).

Conflict of interest: All authors declare no conflict of interest.

Journal of Derivatives and Quantitative Studies: 선물연구
Emerald Publishing Limited
e-ISSN: 2713-6647
p-ISSN: 1229-988X
DOI 10.1108/JDQS-03-2026-0019

We propose an Iterative Market Understanding and Generation Inference (IMUGI) framework – a multilayered predictive system for assessing the probability of drawdown onset, inspired by regime-switching early-warning models, Hawkes-type jump clustering, and meta-learning architectures that ensemble heterogeneous predictors.

Throughout this paper we distinguish two concepts. A drawdown (DD_t) is the cumulative percentage decline from a historical peak; it is critical when it falls in the worst 20% of its empirical distribution. A crisis event is a temporal episode constructed by clustering consecutive days of extreme drawdowns. Our prediction target is the drawdown start date. We test the framework on a composite AI stock index of 27 US-listed technology companies spanning 2003–2025.

2. Literature review

Understanding what precedes major drawdowns requires building on a set of theoretical foundations, from crash physics and bubble mechanics to regime-switching models and early-warning systems.

Sornette's (2009) Dragon King theory positions extreme crashes as mechanistically distinct outliers arising from identifiable phase transitions, diagnosable ex-ante via LPPL modelling — with predictive validity demonstrated across several historical crashes (Johansen and Sornette, 2010) and confirmed across six independent systems (Sornette and Ouillon, 2012) — in contrast to Taleb's (2007) view of crises as inherently unpredictable. Consistent with both, financial returns exhibit fat tails incompatible with Gaussian assumptions (Mandelbrot, 1963; Mandelbrot and Hudson, 2004; Nolde and Zhou, 2021), motivating reliance on Student-t emissions throughout our framework:

$$P(X_t|\ S_t = k) = \text{Student} - t(\mu_k, \Sigma_k, \nu_k) \tag{1}$$

where each regime k has its own location vector μ_k and scale matrix Σ_k, estimated via EM, while the degrees-of-freedom parameter ν—which governs tail heaviness ($\nu \to \infty$ the distribution converges to the Gaussian)—is shared across regimes and fixed at $\nu = 5$. The use of conditional Student-*t* distributions in hidden Markov models is established in the literature (Bulla, 2011), and the $\nu = 5$ is consistent with recent empirical work showing it minimises the kurtosis gap for equity return distributions without degrading distributional fidelity (Alswaidan and Varner, 2026).

Asset bubbles are the primary phenomena through which severe drawdowns emerge. Cognitive biases and herd behaviour drive speculative buying beyond fundamental value, aligned with Keynesian "animal spirits" theory (Akerlof and Shiller, 2009), while rapid price appreciation, excessive leverage, and credit expansion provide observable warning signals (Jordà *et al.*, 2015). Scheinkman and Xiong (2003) formalised the mechanism: when short-sale constraints relax, accumulated pessimism is gradually incorporated into prices, producing sustained post-restriction declines — the dot-com bubble (1995–2000) being the canonical illustration (Ofek and Richardson, 2003).

Drawdown-based rules serve as risk-reduction techniques in asset selection (Van Hemert *et al.*, 2020) and drawdown-based optimisation (Chekhlov *et al.*, 2005). The Financial Crisis Observatory and the Hyped LPPL model translate bubble diagnostics into trading strategies delivering annualised returns around 34% on equity indices, providing direct empirical evidence of economic gains from crash-onset signals (Cao *et al.*, 2025).

IMUGI integrates five strands established in previous literature: bubble physics and log-periodic power-law (LPPL) modelling (Johansen *et al.*, 2000; Sornette, 2003; Jordà *et al.*, 2015; Sornette *et al.*, 2015), regime-switching and hidden Markov models (Hamilton, 1989; Ang and Bekaert, 2002; Guidolin and Timmermann, 2007), self-exciting Hawkes processes (Hawkes, 1971; Filimonov and Sornette, 2012; Blanc *et al.*, 2017; Chen *et al.*, 2024), empirical crisis prediction and early-warning systems — from leading-indicator models (Kaminsky *et al.*, 1998;

Berg and Pattillo, 1999; Bussière and Fratzscher, 2006; Frankel and Saravelos, 2012; Schularick and Taylor, 2012; Drehmann and Juselius, 2014) to time-to-event/survival specifications (Cox, 1972; Shumway, 2001) — and heavy-tailed, jump, and extremal-event analysis, including extreme-return contagion (Engle and Manganelli, 2004; Embrechts *et al.*, 1997; Mandelbrot, 1963; Bae *et al.*, 2003; Nolde and Zhou, 2021). Prior work within each strand addresses these problems in isolation, typically relying on macroeconomic predictors or single-mechanism specifications; IMUGI combines all five using market microstructure features.

Existing ML approaches (Ghasemieh and Kashef, 2022; Tölö, 2020; Rao and Rojas, 2025; Zhao *et al.*, 2025) share a common failure: optimising for retrospective state identification rather than prospective onset prediction, with detections concentrated near the deepest drawdown point — after the institutional hedging window has closed. The Hyped LPPL model (Cao *et al.*, 2025) suggests that mechanistic priors add predictive value, though this does not constitute proof that Dragon King dynamics are the operative driver. IMUGI addresses these gaps by combining bubble physics, regime memory, and self-excitation within a single online probabilistic architecture (Appendix B, Figure 5).

3. Data and methodology

3.1 AI Tech index

The AI technology index is a market-capitalisation-weighted composite of 27 NASDAQ-listed technology companies with significant AI exposure (see Appendix A, Table 10).

The constituents were selected as the set of NASDAQ-listed firms with identifiable revenue or strategic exposure to AI-relevant activities — semiconductors, cloud infrastructure, enterprise data platforms and AI-native software — as identified at the index construction date. The index is intended to represent the population over which AI-sector overvaluation dynamics are defined, in line with the paper's prediction target. Market-cap weighting downweights recently listed firms in earlier periods, leaving training-period index dynamics dominated by established large-cap technology firms. While this structural weighting mitigates but does not eliminate survivorship and hindsight-driven selection concerns (Brown *et al.*, 1992), the index should be read as representing the AI technology sector as currently identified.

The index spans January 2003 to October 2025 (22.8 years; 5,742 trading days). Daily adjusted closing prices and shares outstanding for all constituents were retrieved via the yfinance API (Aroussi, 2019). Index construction follows the dynamic divisor methodology standard in contemporary index design (S&P Dow Jones Indices, 2022).

First, at the baseline date (January 1st, 2003), the total market capitalisation across all constituent companies is computed. A fixed divisor is calculated as:

$$\text{Divisor} = \frac{\text{Total Market Cap at } t_0}{\text{Base Value}} \tag{2}$$

where Base Value = 1,000. Second, on each trading day t, the index value is calculated as the sum of each stock's price multiplied by shares outstanding, divided by fixed divisor:

$$\text{AI Index}_t = \frac{\sum_{i=1}^{27} P_i(t) \times S_i}{\text{Divisor}} \tag{3}$$

where $P_i(t)$ is the adjusted closing price and S_i is the shares outstanding for company i. This market-cap-weighted approach ensures that larger constituents have proportionally greater influence on index movements, reflecting the concentration characteristic of the AI sector.

3.2 Ground truth: drawdown event detection methodology

Evaluation requires an identification of crisis drawdown events. Following an adaptive distribution-based methodology, major drawdowns are detected via the following procedure:

3.2.1 Detection algorithm. Let P_t denote the price of the AI technology index at time t. Define the drawdown series in accordance with Chekhlov *et al*. (2005) as:

$$\mathrm{DD}_t = \frac{P_t - \max_{s \le t} P_s}{\max_{s \le t} P_s}. \tag{4}$$

Days on which DD_t lies in the worst 20% of its empirical distribution (20th percentile) are initially flagged as crisis days. Consecutive runs of crisis days, allowing for short gaps of up to 10 trading days, are merged into single candidate crisis episodes; episodes shorter than 3 days are discarded as noise. For each remaining episode, the crisis anchor is defined as the date within the episode at which DD_t attains its minimum (maximum peak-to-trough loss).

The 20th percentile threshold is aligned with the results of Kaminsky *et al*. (1998), that were selected via optimisation to minimise the noise-to-signal ratio of crisis signals. The 20th percentile threshold is further grounded in the drawdown literature: Van Hemert *et al.* (2020) demonstrate that in a 10-year window with typical risk characteristics (10% volatility, 0.5 Sharpe ratio), there is a 43% probability of hitting a −20% maximum drawdown, yet only 9.9% probability of exceeding −30%. This positioning of the 20% threshold as a natural boundary between recoverable volatility and systemic distress provides empirical support for the current selection.

For each valid crisis episode, two key dates are attributed: the *start date*, defined as the most recent high-water mark before the anchor,

$$t_{\mathrm{start}} = \sup\{\tau \le t_{\mathrm{anchor}} : P_\tau = \max_{s \le \tau} P_s\} \tag{5}$$

and the anchor date, t_{anchor}, the date of maximum loss during the episode. The primary prediction target is the drawdown start date, i.e. the onset of the correction from its peak. Binary labels for model training and evaluation are constructed by marking an anticipation window of 9 trading days preceding each identified drawdown start as positive ($y_t = 1$), with all other periods marked as negative ($y_t = 0$).

To assess whether the empirical conclusions depend on the event-definition choices, we conducted a local sensitivity analysis around the baseline specification. In addition to the main 20th-percentile drawdown threshold, robustness checks are performed at the 15th and 25th percentiles. Hence, the binary anticipation window is varied across 5, 9, and 15 trading days. This design preserves a common event-detection framework while testing whether predictive performance is stable under variations in drawdown severity.

3.2.2 Historical drawdown events. Analysis of the AI technology index (constructed from 27 NASDAQ constituents, 2003–2025) identified 12 unique drawdown onsets generating 17 distinct crisis events. This multiplicity reflects protracted multi-stage declines where intermediate recoveries create multiple local minima from a single peak. Three onsets exhibit clustering, Table 1 documents all events.

Drawdowns exhibit heterogeneous morphologies: flash corrections (21–29 days, −13.52% to −27.71%), intermediate declines (48–184 days, −14.45% to −26.53%), and protracted crises (311–821 days, −14.31% to −53.86%). Figure 1 illustrates this heterogeneity and reveal multi-stage dynamics with oscillations before final trough depths. The presence of multiple anchors per onset demonstrates that large drawdowns do not follow simple one-stage declines; instead, they exhibit oscillations and recovery attempts before reaching the lowest depths. This complexity validates the SOHMM framework's design for real-time discrimination of distinct crisis regimes.

3.3 Feature engineering

To operationalise the drawdown driving mechanisms identified in the literature – positive feedback loops, log-periodic precursors, jump clustering, volatility amplification, and

Table 1. Complete drawdown events in the US AI and Technology sector (2005–2025)

Event	Start date	Anchor date	Duration	Magnitude
PC Cycle Downturn	2006-01-11	2006-07-14	184	−21.03%
Global Financial Crisis Cluster (Onset: 2007-11-06)				
GFC: Acute Crisis Phase	*2007-11-06*	2008-11-20	380	−53.86%
GFC: Prolonged Recovery Decline	*2007-11-06*	2010-02-04	821	−14.31%
Flash Crash Aftermath Cluster (Onset: 2010-04-22)				
Flash Crash: Short-Term Trough	*2010-04-22*	2010-06-09	48	−14.95%
Flash Crash: Medium-Term Trough	*2010-04-22*	2010-07-02	71	−18.41%
Flash Crash: Extended Decline	*2010-04-22*	2010-08-31	131	−16.87%
US Debt Ceiling and Credit Downgrade	2011-07-26	2011-08-19	24	−15.93%
European Sovereign Debt Crisis	2012-09-19	2012-11-15	57	−14.45%
Oil Collapse and Emerging Market Stress	2015-12-04	2016-02-09	67	−16.89%
Fed Rate Hiking Cycle	2018-10-01	2018-12-24	84	−26.43%
COVID-19 Pandemic Shock	2020-02-19	2020-03-16	26	−27.71%
Profit-Taking After V-Recovery	2020-09-02	2020-09-23	21	−13.52%
Fed Tightening Cycle Cluster (Onset: 2021-12-27)				
Tightening: Sharp Selloff	*2021-12-27*	2022-01-25	29	−15.56%
Tightening: Extended Decline	*2021-12-27*	2022-03-14	77	−21.34%
Tightening: Recovery-Phase Bottom	*2021-12-27*	2022-11-03	311	−40.08%
Yen Carry Trade Unwinding	2024-07-10	2024-08-07	28	−16.42%
AI Bubble and DeepSeek Competition	2024-12-24	2025-04-08	105	−26.53%

Note(s): Magnitudes represent peak-to-trough declines. Events sharing identical start dates represent multiple anchor depths from protracted declines

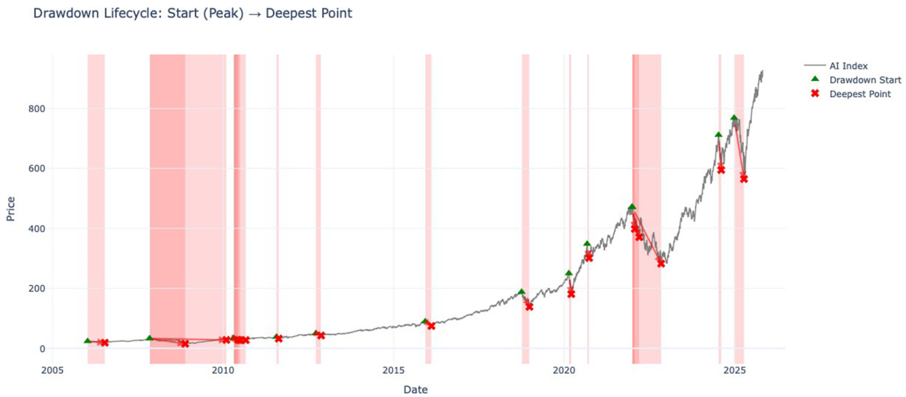

Figure 1. Drawdown lifecycle: onsets and troughs on AI Index (2005–2025). Green triangles denote 12 unique drawdown onsets; red X marks denote 17 crisis anchors. Pink shaded regions indicate duration. Three clusters (2007-11-06, 2010-04-22, 2021-12-27) reveal multi-stage crisis dynamics with regime-dependent oscillations before ultimate trough depths

multivariate regime breakdowns – eight features spanning LPPL bubble physics, microstructure dynamics, and extremal phenomena were computed for daily inference.

3.3.1 LPPL-derived features. Three LPPL-derived indicators are adapted from Sornette *et al.* (2015) with modifications for non-stationary AI stock dynamics and drawdown detection at the 20th-percentile threshold. Rather than Sornette's 126 linearly-spaced windows, we use

15 logarithmically-spaced windows for computational efficiency while retaining coverage across three economically meaningful regimes: long-term hype cycles (600–750 days), medium-term institutional adoption (300–500 days), and short-term retail momentum (125–200 days):

$$W_i = 10^{\log_{10}(750) - i \cdot \frac{\log_{10}(750) - \log_{10}(125)}{14}} \quad \text{for } i = 0, \ldots, 14 \tag{6}$$

The LPPLS model fitted to normalised log-prices is:

$$P(t) = A + B(t_c - t)^m + C(t_c - t)^m \cos(\omega \log(t_c - t) + \phi) \tag{7}$$

where t_c is the critical time (predicted drawdown onset), $m \in (0.1,\ 0.9)$ the power-law exponent, $\omega \in (6,\ 13)$ the angular frequency, and $\phi \in (0,\ 2\pi)$ the phase, derived from the Johansen–Ledoit–Sornette (JLS) rational expectations model (Johansen *et al.*, 2000; Sornette, 2003).

Rather than Sornette's OLS approach, we employ hybrid optimisation: differential evolution (DE, maxiter = 300) optimises the nonlinear parameters (t_c, m, ω, ϕ) under Huber-robust loss, while linear parameters (A, B, C) are solved analytically within each DE iteration. This combination avoids a six-dimensional nonlinear search and provides robustness to the heavy-tailed shocks common in AI stocks (price jumps of ±15%, corporate actions, liquidity events):

$$L_{\text{Huber}}(\mathbf{r}, \delta) = \sum_{i=1}^{n} \left\{ \frac{1}{2} r_i^2 \quad \left| r_i \right| \leq \delta \delta \left(\left| r_i \right| - \frac{1}{2} \delta \right) |r_i| > \delta, \quad \delta = 1.0 \right. \tag{8}$$

Filtering Constraints Valid windows must satisfy all seven constraints simultaneously (Table 2), derived from the JLS framework (Johansen *et al.*, 2000) and operationalised by Shu and Zhu (2019). Unlike crash-focused formulations that maximise crash probability, our drawdown-specific implementation emphasises sustained decline patterns, relaxing strict oscillatory constraints to isolate gradual structural trend exhaustion rather than the finite-time singularities of Dragon King events.

Damping Constraint and Relaxed Bounds for Non-Stationary AI Markets The damping factor:

$$\text{damp} = \frac{m|B|}{\omega|C|} \tag{9}$$

derives from Bothmer and Meister (2003), who established damp ≥1 as the condition for a non-negative crash hazard rate in the JLS framework, subsequently operationalised by Shu

Table 2. Filtering constraints for valid LPPLS window fits

Constraint	Condition	Interpretation
Power-law coefficient	$-10 < B < -10^{-3}$	Ensures negative returns (downward pressure)
Oscillatory amplitude	$0.01 < \|C\| < 1$	Detects oscillatory structure, bounds magnitude
Log-price intercept	$-100 < A < 100$	Numerical stability for normalised log-prices
Minimum oscillations	$n_{\text{osc}} = \frac{\omega}{2\pi} \log \frac{t_c - t_0}{t_c - t_1} \geq 2.0$	Ensures log-periodic signature
Damping factor	$\frac{m\|B\|}{\omega\|C\|} \geq 1$	Non-negative hazard rate (standard constraint)
Maximum relative error	rel_err ≤ 0.2	Model fit quality
Residual stationarity	ADF $p < 0.05$	White-noise residuals

Note(s): All seven conditions must be satisfied simultaneously for a window to contribute to indicator aggregation and Zhu (2019) on Chinese markets and validated by Koistinen (2020) on Finnish equities with 1.4-day average timing accuracy.

For non-stationary AI markets we extend the allowable range to 0 < damp <10: rapid regime shifts, retail-driven herding, and non-stationary intraday microstructure mean that sub-critical damping (0 < damp <1) can coexist with genuine instability signatures.

This relaxation is compensated by the LPPL Trust metric (Section 3.3.1), which detects parameter fragility independently of damping magnitude, retaining the fundamental JLS instability criterion – when market parameters become sensitive to small perturbations the system is approaching a critical transition – without sacrificing sensitivity.

LPPL Probabilistic Proximity Rather than Sornette's theoretical hazard rate $h_t = \alpha(t_c - t)^{m-1}[1 + \cos(\omega \ln(t_c - t) - \phi')]$, we operationalise drawdown imminence via the LPPLS function's instantaneous log-return at normalised window end ($t = 1.0$):

$$h(1.0) = \frac{P'(1.0)}{P(1.0)} \tag{10}$$

where:

$$\begin{aligned} P'(t) = &-mB(t_c - t)^{m-1} - mC(t_c - t)^{m-1}\cos(\omega \log(t_c - t) + \phi) \\ &- \omega C(t_c - t)^{m-1}\sin(\omega \log(t_c - t) + \phi) \end{aligned} \tag{11}$$

The derivative captures both power-law acceleration ($-mB$ term) and oscillatory damping (mC and ωC terms). Sigmoid normalisation maps it to [0, 1]:

$$\text{LPPL\_Probabilistic\_Proximity} = \sigma(h(1.0)) = \frac{1}{1 + \exp(-h(1.0))} \tag{12}$$

Values near 1 indicate accelerating negative returns (drawdown criticality); values near 0 indicate stability or positive momentum.

LPPL Confidence The fraction of windows passing all filtering constraints:

$$\text{LPPL\_Confidence} = \frac{N_{\text{passed}}}{15} \tag{13}$$

High confidence (≥0.7) confirms that LPPLS signatures persist across temporal scales, reducing false positives from transient noise. Values below 0.5 suggest inconsistent fits, indicating stable or transiently non-stationary regimes. Multi-scale consistency is central to the JLS framework: genuine bubbles exhibit log-periodic oscillations at multiple temporal frequencies simultaneously (Johansen *et al.*, 2000; Sornette, 2003).

LPPL Trust Quantifies parameter robustness via bootstrap resampling (Sornette *et al.*, 2015), with $n_{\text{bootstrap}} = 10$ resamples (vs. Sornette's 100) for real-time feasibility:

(1) Compute residuals $\mathbf{r} = \widehat{P} - P$.

(2) Resample with replacement to form synthetic series $P_{\text{syn}} = \widehat{P} + \mathbf{r}^*$.

(3) Refit LPPLS on each replicate and record whether damp ≥1.

$$\text{LPPL\_Trust} = \frac{1}{n_{\text{bootstrap}}} \sum_{b=1}^{n_{\text{bootstrap}}} \mathbb{1}\left[\frac{m_b|B_b|}{\omega|C_b|} \geq 1\right] \tag{14}$$

Low Trust (< 0.3) indicates parameter sensitivity and dynamical fragility near critical transitions – a hallmark of pre-drawdown regimes where small price changes produce dramatically different parameter estimates. Trust is evaluated against the standard constraint (damp $\geq$1) independently of the relaxed operational bounds, so it captures fragility even when the main estimate already satisfies damp $\geq$1.

3.3.2 Microstructure and jump features. Five features capture market microstructure dynamics, jump risk, and distributional anomalies, complementing the LPPL indicators. Log return $r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$ stabilises variance (Campbell *et al.*, 1997); return acceleration $\Delta r_t = r_t - r_{t-1}$ serves as a 2–5-day leading indicator of regime inflection points (Sornette, 2003); and annualised 20-day realised volatility $\sigma_t = \sqrt{252} \cdot \mathrm{std}(r_{t-19:t})$ captures time-varying market stress (Andersen *et al.*, 2001). Lee-Mykland jump statistics (Lee and Mykland, 2008) are aggregated across $\mathcal{W} = \{5-10,\ 10-30,\ 30-60\}$ days as

$$\text{LMJump\_Stat}_i^{(w)} = \frac{r_i}{\sqrt{\text{RV}_{i-w:i}}}, \qquad \text{RV}_{i-w:i} = \frac{1}{w}\sum_{j=i-w}^{i} r_j^2, \tag{15}$$

flagged at $|\text{LMJump\_Stat}| > 2.58$ ($\alpha = 0.01$) and set to NaN otherwise. Mahalanobis distance (Mahalanobis, 1936; Rousseeuw and Van Zomeren, 1990) is computed over windows {20–30, 30–90, 60–180} days on $\mathbf{x}_t = (r_t, \sigma_t, \text{RSI}_t, \text{MACD}_t, \text{momentum}_t, \Delta r_t)^T$ as

$$\text{MD}_t = \sqrt{(\mathbf{x}_t - \mu)^T \Sigma^{-1} (\mathbf{x}_t - \mu)}, \tag{16}$$

$$\text{MD\_Avg\_Outlier} = \text{MD\_Avg} \cdot \mathbb{I}\left[\text{MD}_t^2 > \chi^2_{0.95}(6) \approx 12.59\right], \tag{17}$$

capturing joint deviations across returns, volatility, and technical indicators during regime shifts. All continuous features are z-score normalised over non-NaN observations; binary indicators retain their {0, 1} form without scaling.

3.3.3 Feature architecture. The eight features provide a layered decomposition of crisis morphology: oscillatory bubble accumulation (LPPL indicators), shock self-excitation (LMJump), volatility amplification (rolling volatility), multivariate regime coherence (Mahalanobis distance), and momentum acceleration (return differences). The complete architecture, including feature classes and empirically derived crisis thresholds, is summarised in Table 3.

3.4 Second-Order Hidden Markov Model architecture

The SOHMM provides the probabilistic foundation for regime-switching inference in the IMUGI framework. Unlike conventional first-order HMMs, which condition transitions

Table 3. Eight-feature architecture for crisis detection, organised by feature class with empirically derived crisis thresholds

Feature class	Feature	Crisis threshold
LPPL Physics	Proximity (Hazard Rate)	>0.7 (crash imminence)
	Confidence	>0.7 (11+/15 windows pass)
	Trust	<0.3 (parameter fragility)
Microstructure	Log Return	<-0.04 (heavy-tailed shock)
	Return Acceleration	$\lvert\Delta r_t\rvert > 0.05$ (inflection)
	Volatility (20-day rolling)	>0.40 annualised
Extremal Events	LMJump Detected	$\geq$2 window sizes flagged
	MD Avg Outlier	$\text{MD}^2 > 12.59$

solely on $P(S_t|S_{t-1})$, the SOHMM incorporates trajectory dependence by conditioning on the dyadic state sequence (S_{t-2}, S_{t-1}), capturing momentum and mean-reversion patterns that are empirically significant during crisis episodes characterised by pronounced persistence.

3.4.1 Trajectory-dependent state transitions. The architectural core is the three-dimensional transition tensor $\mathcal{T} \in \mathbb{R}^{K\times K\times K}$, $K = 5$:

$$\mathcal{T}_{ijk} = P(S_t = k| \, S_{t-2} = i, \, S_{t-1} = j) \tag{18}$$

A standard first-order HMM uses a single 5×5 transition matrix (25 parameters). The SOHMM instead instantiates five matrices $A^{(1)}, \ldots, A^{(5)}$, each conditioned on the regime two steps prior, stacked into a $5 \times 5 \times 5$ tensor (125 parameters). This expansion is economically justified: a crisis regime (k) reached via trajectory (i, j) exhibits fundamentally different forward dynamics than crisis arrival via an alternative path, and the rank-3 structure prevents the rapid state flickering endemic to first-order Gaussian HMMs. The model learns these trajectory-dependent distinctions via the EM algorithm.

The empirical motivation is regime persistence. Define the persistence coefficient for regime c as:

$$\rho_c = \frac{1}{K}\sum_{i=1}^{K}\mathcal{T}_{icc} \tag{19}$$

Relative to Nystrup *et al.*'s (2020) jump-penalised first-order HMM, the SOHMM achieves implicit persistence through trajectory-dependent transitions while additionally capturing momentum and mean-reversion dynamics. A third-order extension (625 parameters) would be ill-identified given only 12 training onsets; the rank-3 structure balances expressiveness with estimability.

The graphical dependency structure is shown in Figure 2: blue arrows represent standard Markovian transitions ($t - 1 \rightarrow t$); red dashed arrows capture second-order memory ($t - 2 \rightarrow t$).

3.4.2 Emission parameter initialisation: semi-supervised approach. Initialisation combines K-Means clustering with semantic domain-knowledge scoring in four stages. (1) Global imputation statistics $(\bar{x}_d, \sigma_d)$ are computed from training data to fill missing entries. (2) K-Means partitions the imputed data into $K = 5$ clusters with parameters $\widehat{\mu}_k$, $\widehat{\Sigma}_k$. (3) Each cluster is scored against five target regimes (calm, recovery, euphoria, bubble, crisis) using standardised feature combinations; clusters are then greedily [1] matched to regimes by

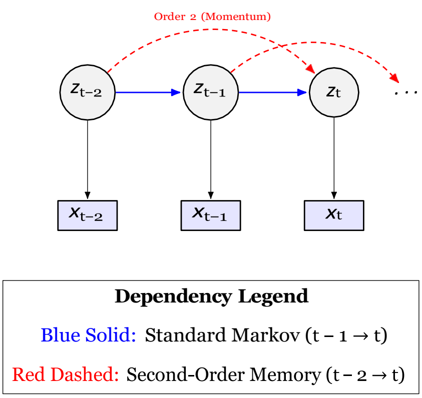

Figure 2. Dependency graph of the second-order HMM (Mari *et al.*, 1997)

iteratively selecting the highest cluster – regime score – conceptually related to the Hungarian algorithm (Kuhn, 1955) but implemented in simplified form for efficiency. (4) Regime-specific missingness probabilities $\Psi_k[d] = P(\text{feature } d \text{ observed}|\text{state } k)$ are learnt from training data, capturing the MNAR structure (Little and Rubin, 2002).

This hybrid approach is a novel contribution to HMM literature (Rabiner, 1989; Murphy, 2012). The semantic scoring layer encodes domain knowledge (LPPL bubble physics, Hamilton (1989) volatility regimes, Rousseeuw and Van Zomeren (1990) outlier patterns) to guide cluster-to-regime assignment without requiring crisis labels at initialisation, rendering the SOHMM *semi-supervised* (Chapelle *et al.*, 2006). Labels are reserved for downstream evaluation; the semantic initialisation acts as a soft prior that accelerates EM convergence and improves interpretability under scarce labelled crises.

3.4.3 Adaptive online learning via recursive Bayesian filtering. Static batch-trained models introduce forward-looking bias by implicitly memorising crash patterns during training and then "predicting" those same patterns out-of-sample. We eliminate this circularity via an online learning framework that updates parameters incrementally as observations arrive.

At each time step, model surprise is quantified via KL divergence:

$$\mathcal{S}_t = D_{KL}\left(\gamma_t \middle\| \pi_{t|t-1}\right) = \sum_{k=1}^{K} \gamma_t(k) \ln \frac{\gamma_t(k)}{\pi_{t|t-1}(k)} \tag{20}$$

where $\gamma_t(k) = P(S_t = k|\ X_1^t)$ is the posterior and $\pi_{t|t-1}(k) = P(S_t = k|\ X_1^{t-1})$ is the one-step-ahead predictive distribution. The learning rate is modulated accordingly:

$$\eta_t = \eta_{\text{base}} + \frac{\eta_{\max} - \eta_{\text{base}}}{1 + \exp(-\beta(\mathcal{S}_t - \tau))} \tag{21}$$

This sigmoid-modulated rate increases under high surprise (enabling rapid adaptation to new regimes) and decreases during stable periods (preserving historical knowledge). The full pipeline—E-Step posterior computation, KL-triggered η_t adjustment, EWMA sufficient-statistic update, and M-Step parameter refinement—is shown in Figure 3, with updated parameters fed back to the E-Step via the state transition tensor $\mathcal{T}$. EM convergence criteria, numerical safeguards, and the online damping coefficients (κ_μ, κ_Ψ, κ_Σ) are specified in Appendix C (Algorithms 1–2).

The SOHMM design synthesises HMM fundamentals (Baum and Petrie, 1966; Baum *et al.*, 1970; Rabiner, 1989), multi-order state dependencies (Juang and Rabiner, 1991; Mari *et al.*, 1997), robust Student-t emissions (Peel and McLachlan, 2000), semi-supervised initialisation (Chapelle *et al.*, 2006), and recursive Bayesian filtering (Cappé *et al.*, 2009) into a unified framework adapted for non-stationary financial regimes, heavy-tailed distributions, and MNAR data structures.

3.5 Quadratic Hawkes process

Self-exciting point processes model temporal event clustering, where past occurrences elevate future event likelihood. Filimonov and Sornette (2012) demonstrated that price changes decompose into exogenous shocks (background intensity μ) and endogenous feedback (memory kernel), with the Hawkes branching process providing a natural mathematical framework for this decomposition. We adopt this semantic logic and extend the linear Hawkes framework with quadratic interactions to capture multiplicative cascading behaviour during market stress.

Let $\mathcal{E} = \{t_1, \ldots, t_N\}$ denote extreme price move timestamps (returns exceeding the 95th or falling below the 5th percentile of training data). The conditional intensity is:

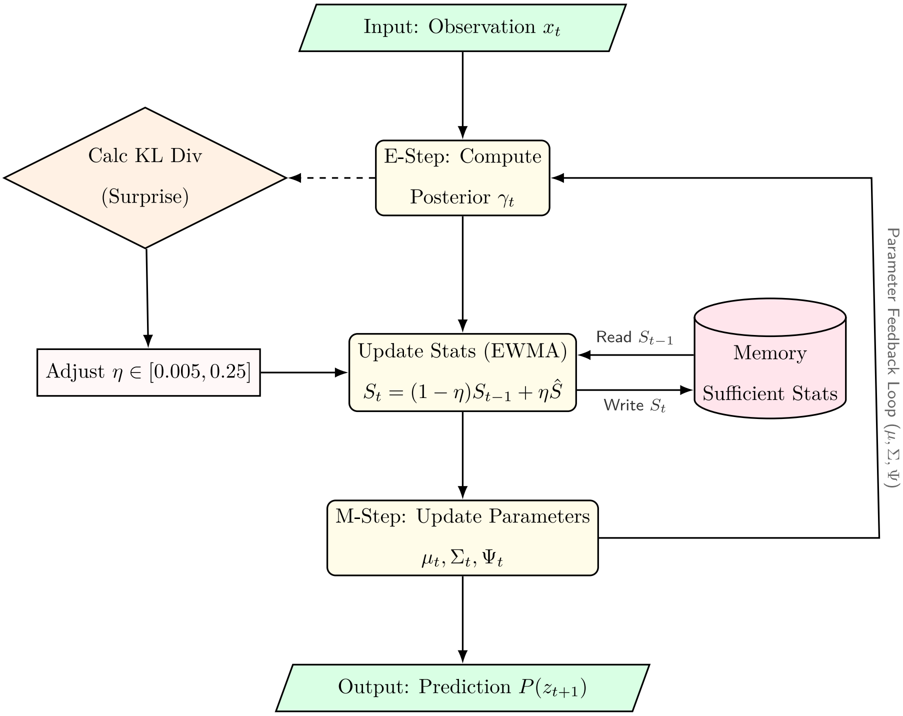

Figure 3. SOHMM online learning architecture

$$\lambda(t) = \mu + \alpha \sum_{t_i \le t} e^{-\theta^{(0)}\tau_i} + \beta^{\text{diag}} \sum_{t_i \le t} e^{-\theta^{(1)}\tau_i} + \beta^{\text{rank1}} \left(\sum_{t_i \le t} e^{-\theta^{(2)}\tau_i} \right)^2 \tag{22}$$

where $\tau_i = t - t_i$. The seven parameters are: μ (baseline exogenous intensity); α (linear self-excitation); β^{diag}, β^{rank1} (quadratic amplitudes); $\theta^{(0)}$, $\theta^{(1)}$, $\theta^{(2)}$ (exponential decay rates).

3.5.1 Three Kernel components. **Linear kernel** (Hawkes, 1971; Bowsher, 2007):

$$\lambda_{\text{linear}}(t) = \alpha \sum_{t_i \le t} e^{-\theta^{(0)}\tau_i} \tag{23}$$

Models standard first-order endogenous feedback, directly analogous to the linear Hawkes term in Filimonov and Sornette (2012).

Diagonal quadratic (Blanc *et al.*, 2017):

$$\lambda_{\text{diag}}(t) = \beta^{\text{diag}} \sum_{t_i \le t} e^{-\theta^{(1)}\tau_i} \tag{24}$$

Captures local event clustering and microstructure effects within individual event epochs.

Rank-1 quadratic (Sornette, 2003):

$$\lambda_{\text{rank1}}(t) = \beta^{\text{rank1}} \left(\sum_{t_i \le t} e^{-\theta^{(2)}\tau_i} \right)^2 \tag{25}$$

Introduces super-exponential scaling: dense event clustering produces intensity spikes through the squared summation, capturing multiplicative cascade effects characteristic of critical market regimes and extending beyond the linear Filimonov–Sornette (2012) framework.

3.5.2 Parameter estimation. The log-likelihood (Hawkes, 1971; Ogata, 1978) is:

$$\ell(\theta) = \sum_{i=1}^{N} \ln \lambda(t_i) - \int_0^T \lambda(u)\, du \tag{26}$$

The integral is approximated via the trapezoidal rule with $N_{\text{quad}} = 1{,}000$ quadrature points. Parameters are estimated by constrained L-BFGS-B optimisation (Byrd *et al.*, 1995) with a stability penalty (Daley and Vere-Jones, 2008):

$$\mathcal{R}_{\text{stab}} = \max\left(0,\ \frac{\beta^{\text{diag}}/\theta^{(1)} + \left(\beta^{\text{rank1}}/\theta^{(2)}\right)^2}{1 - \alpha/\theta^{(0)}} - 0.99\right)^2 \tag{27}$$

This prevents numerical divergence while permitting transient critical bursts. The box constraints and initial values used in estimation are reported in Appendix C (Algorithm 3).

3.5.3 Output: continuous intensity feature. For each calendar date t, the conditional intensity $\lambda(t|\,\mathcal{E}_{\leq t})$ is computed using only events strictly before t. Two sequential transformations are applied:

$$h_{\text{Hawkes}}(t) = \log(1 + \lambda(t)) \tag{28}$$

$$\tilde{h}_{\text{Hawkes}}(t) = \frac{h_{\text{Hawkes}}(t) - \min_{\text{train}}}{\max_{\text{train}} - \min_{\text{train}}} \tag{29}$$

yielding a single normalised feature $\tilde{h}_{\text{Hawkes}}(t) \in [0, 1]$ per date. During testing, the event set $\mathcal{E}$ is updated incrementally using the same training-period percentile thresholds, allowing the intensity to reflect accumulated pressure from recent extreme events without look-ahead bias.

3.6 Framework architecture and design overview

The IMUGI framework is a three-layer architecture:

(1) A **signal layer** comprising: a SOHMM for market regime identification; a Q-Hawkes process for extreme-return clustering.

(2) A **feature layer** constructing: five meta-features from SOHMM and Q-Hawkes outputs.

(3) A **meta-learning layer**: a regularised logistic regression mapping the five-dimensional feature vector to a crisis-onset probability.

The SOHMM is trained on eight features: Log Return, Return Acceleration, Volatility, LMJump Stat Detected, MD Avg Outlier, LPPL Proximity, LPPL Confidence, and LPPL Trust. This bundle integrates price momentum, jump detection, outlier intensity, and bubble-formation indicators, enabling discrimination across five latent market regimes while learning state-dependent missing-data patterns via MNAR-aware parameters.

3.7 Meta-learning: feature engineering and model

3.7.1 Meta-features. The meta-model uses five features derived from Q-Hawkes and SOHMM outputs:

(1) **Hawkes** ($\tilde{h}_t$): normalised Q-Hawkes intensity.

(2) **SOHMM** ($\tilde{s}_t$): normalised crisis-transition probability.

(3) **Hawkes Momentum** ($\Delta_5^H = \tilde{h}_t - \tilde{h}_{t-5}$): five-day change in Hawkes intensity.

(4) **SOHMM Momentum** ($\Delta_5^S = \tilde{s}_t - \tilde{s}_{t-5}$): five-day change in regime probability.

(5) **Interaction** ($I_t = \tilde{h}_t \cdot \tilde{s}_t$): multiplicative coupling of both signals.

These form the meta-feature vector $\mathbf{x}_t = (\tilde{h}_t, \tilde{s}_t, \Delta_5^H, \Delta_5^S, I_t)^\top$.

3.7.2 Logistic regression meta-model. The meta-learner is a regularised logistic regression:

$$\Pr(y_t = 1 | \mathbf{x}_t) = \sigma\left(\beta_0 + \sum_{j=1}^{5} \beta_j x_{t,j}\right), \qquad \sigma(z) = \frac{1}{1 + e^{-z}} \tag{30}$$

with L2 regularisation and balanced class weights to address the $\approx 2.5\%$ positive label rate.

3.7.3 Benchmark models and statistical evaluation. To benchmark IMUGI against alternatives, we additionally evaluate three external models on the same chronological train–test split: a climatology benchmark that assigns the in-sample event rate to every observation, a discrete-time hazard specification following the logic of Shumway (2001), and a time-varying Cox proportional hazards model (Cox, 1972). Raw benchmark and IMUGI probabilities are converted to calibrated probabilities using isotonic regression fitted on out-of-fold training predictions, with a constant fallback when the training predictions are degenerate, following standard probability-calibration practice (Zadrozny and Elkan, 2002). Model-specific classification thresholds are then selected on the calibrated training predictions by maximising balanced accuracy and are held fixed for the test period. For statistical comparison, discrimination differences in AUC and AUPRC are evaluated on raw test probabilities using a paired moving-block bootstrap with 1,000 replications and 21-trading-day blocks, while calibrated probability accuracy is compared using pointwise Brier loss and a one-sided Diebold-Mariano test with Newey-West variance estimation and the Harvey-Leybourne-Newbold small-sample correction ($h = 1$) (Harvey *et al.*, 1997).

4. Results

4.1 Train–test split and causal evaluation protocol

Due to the LPPL module requiring 750 observations before the first valid estimation window, the effective training sample spans 3,992 trading days (2005-12-22 to 2021-10-01), while the strictly held-out test set spans 999 trading days (2021-10-02 to 2025-10-24). To preserve causal ordering and prevent forward-looking bias, all model fitting, feature preprocessing, calibration, and threshold selection procedures were fixed using training data only, and each prediction for day t was generated using information available up to and including that date. In particular, benchmark-model covariates were imputed and standardised using training-sample quantities only, benchmark and IMUGI probabilities were calibrated using isotonic regression fitted on out-of-fold training predictions, and the benchmark-comparison procedures were then applied unchanged over the test period.

4.2 Meta-model coefficients

The estimated coefficients of the regularised logistic meta-model are reported in Table 4. Numerically, the intercept is $\beta_0 = -0.2058$ and the five slope coefficients are $(\beta_1, \ldots, \beta_5) = (-6.33, 0.67, 4.68, 3.98, 5.50)$. The most notable feature is the negative coefficient on Hawkes intensity combined with a large positive interaction term between Hawkes intensity and the SOHMM crisis-transition signal. Substantively, this implies that elevated event-clustering pressure is not interpreted as a crisis warning in isolation, but rather becomes informative when

Table 4. Meta-model coefficients and relative importance (absolute magnitude)

Feature	Coefficient	Importance
Hawkes intensity	−6.33	29.9%
Hawkes × SOHMM	+5.50	26.0%
Hawkes momentum	+4.68	22.1%
SOHMM momentum	+3.98	18.8%
SOHMM regime probability	+0.67	3.2%

it coincides with elevated regime-instability probability from the SOHMM, which is consistent with regime-dependent interaction effects in self-exciting financial systems (Blanc *et al.*, 2017; Chen *et al.*, 2024).

4.3 Held-out performance and benchmark comparison

We first summarise out-of-sample performance on the fixed 999-day test set. Using the pre-specified training-based decision threshold, IMUGI attains an AUC of 0.9055 and a Brier score of 0.1175. Three unique drawdown onsets occur in this window, and Table 5 shows that the framework identifies all three, flagging two in advance and one shortly after onset.

Because these onset-level results rely on only three events, formal inference must be interpreted cautiously. We therefore add two further checks: (1) robustness to alternative crisis-definition thresholds and anticipation windows and (2) comparisons with external benchmark models. Table 6 shows that the 20th-percentile threshold with a 9-day anticipation window is a reasonable baseline. Nearby specifications slightly improve individual metrics but do not dominate across discrimination, calibration, and warning performance. In particular, looser 25th-percentile thresholds worsen Brier loss, while tighter or longer-horizon alternatives reduce either AUC or warning usefulness.

Table 5. Detection performance on the 999-day test set

Onset	Event	Lead time	Probability
2021-12-27	Fed Tightening Cycle (−40.1%)	−5 days (early)	0.5828
2024-07-10	Yen Carry Trade Unwinding (−16.4%)	+2 days (immediate)	0.5806
2024-12-24	AI Bubble and DeepSeek Competition (−26.5%)	−1 day (early)	0.6019

Table 6. Robustness of raw IMUGI under alternative crisis-definition thresholds and anticipation windows

Threshold (%)	Anticipation (days)	AUC	Brier	Advance warning Rate (%)	Immediate capture Rate (%)
15	5	0.8598	0.1171	0.0	0.0
15	9	0.7488	0.1138	50.0	0.0
15	15	0.6885	0.1150	0.0	0.0
20	5	0.9170	0.1156	66.7	33.3
20	9	0.9055	0.1175	66.7	33.3
20	15	0.9142	0.1173	66.7	33.3
25	5	0.9174	0.1333	66.7	0.0
25	9	0.9064	0.1356	66.7	0.0
25	15	0.9140	0.1367	66.7	0.0

We next compare IMUGI with three external benchmarks. Tables 7 and 8 indicate that IMUGI outperforms all three in ranking-based discrimination and achieves significantly lower calibrated Brier loss in every pairwise comparison.

The AUC margin over Climatology is $+0.41$ ($p < 0.01$), confirming that a base-rate model carries virtually no useful timing information for crises. The $+0.14$ margin over Shumway DTH ($p < 0.05$) is more informative: because the Shumway model already incorporates time-varying covariates, this gap isolates the contribution of IMUGI's regime-switching and self-exciting components beyond a standard survival framework. AUPRC gains are positive in all three comparisons but reach conventional significance only against Climatology, a pattern consistent with the known conservatism of precision–recall metrics under severe class imbalance (Williams, 2021). Table 8 shows that absolute Brier differences are small (≤ 0.0007), yet the HLN-corrected DM statistics are large in magnitude (-3.26 to -3.57, all $p < 0.01$), indicating that IMUGI's calibration advantage is systematic rather than driven by a few influential days.

To assess whether the additional layers of IMUGI provide incremental value beyond simpler alternatives, we estimate two logistic-regression benchmarks and six nested IMUGI variants on the test horizon.

Table 9 shows that the main gain comes from combining the SOHMM and Hawkes signals: M3 improves substantially over either single-signal variant. The later layers do not improve monotonically; M4 yields the best Brier score among the nested IMUGI variants, M5 the highest AUC, and the full specification remains close to the strongest variants on both measures.

The logistic benchmarks further clarify the contribution of the layered architecture. Relative to the simplest logistic benchmark, Full IMUGI improves both AUC and Brier. Relative to the raw same-input logistic benchmark, Full IMUGI improves AUC, whereas LR0 attains the lowest Brier score; in an event-sparse sample, however, a lower Brier score can simply reflect more conservative average probability forecasts rather than a better early-warning model. As the paper's objective is drawdown-onset discrimination with actionable lead time, Full IMUGI remains the preferred specification.

Table 7. Bootstrap significance tests for IMUGI relative to external benchmark models

Benchmark	dAUC	dAUPRC	Obs
IMUGI vs Climatology	0.4055***	0.1610*	4,991
IMUGI vs Shumway DTH	0.1355**	0.1032	4,991
IMUGI vs Cox PH	0.3381***	$0.1546^{\dagger}$	4,991

Note(s): Bootstrap test with $B = 1{,}000$ replications and 21-day block resampling. AUC and AUPRC are computed from raw probabilities. dAUC and dAUPRC denote IMUGI minus benchmark performance. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$, $^{\dagger}p < 0.15$

Table 8. HLN-corrected Diebold–Mariano tests for calibrated Brier loss

Comparison	LLA	LLB	Mean d	DMraw	DMHLN	Sig
IMUGI vs Climatology	0.0291	0.0292	−0.0001	−3.261	−3.260	$p < 0.01$
IMUGI vs Shumway DTH	0.0291	0.0299	−0.0007	−3.571	−3.569	$p < 0.01$
IMUGI vs Cox PH	0.0291	0.0298	−0.0007	−3.524	−3.522	$p < 0.01$

Note(s): All models are calibrated and evaluated on the full test set. Loss is pointwise Brier score with HAC Newey–West variance estimation and forecast horizon $h = 1$. The null is H_0: $E[\text{loss}_{\text{IMUGI}} - \text{loss}_{\text{benchmark}}] = 0$ against the one-sided alternative that IMUGI has lower calibrated Brier loss. The reported significance category is based on the one-sided HLN-corrected p-value

Table 9. Ablation analysis and simple benchmark comparison on the test horizon

Model	Specification	AUC	Brier
LR0	Logistic regression	0.8070	*0.1017*
LR1	Logistic regression (no LPPL)	0.6434	0.1700
M1	SOHMM only	0.8579	0.1659
M2	Hawkes only	0.6879	0.1846
M3	M1 + M2 signals	0.9129	0.1257
M4	M3 + momentum terms	0.9134	0.1165
M5	M4 + interaction term	*0.9167*	0.1266
M6	Full IMUGI	0.9055	0.1175

The non-monotonic pattern across M3–M6 is informative rather than necessarily problematic. The largest improvement arises when the SOHMM and Hawkes components are combined, which suggests that these two layers contribute complementary predictive structure. The later additions then redistribute performance across metrics instead of uniformly dominating earlier variants; given a test window containing only three unique drawdown onsets, we do not interpret this as establishing that every additional meta-feature is strictly necessary, and some of the differences among M3–M6 are small relative to the sampling uncertainty.

Taken together, these results indicate that the proposed framework provides better discrimination than standard benchmarks on the full test sample and significantly lower calibrated Brier loss than the external benchmarks. The performance is robust to reasonable variations in the crisis-definition threshold and anticipation window, and the largest single gain arises from combining the SOHMM and Hawkes signals. However performance across the nested variants is not monotonic and the full specification is not uniformly best on every metric; the later layers redistribute performance across discrimination and calibration rather than dominating earlier variants. Within the limits imposed by the small number of significant drawdowns, these results are consistent with the full IMUGI specification extracting incremental predictive value from the same information set.

4.4 Framework performance

The framework detected all three onsets: two with actionable anticipatory signals and one confirmed immediately following the shock. Green markers (circles) indicate days where the model exceeded its decision threshold within the nine-day anticipation window prior to each onset; orange markers (diamonds) indicate threshold crossings within five days after the onset. From an applied risk-management perspective, advance warning of up to two weeks and recognition within one week after onset would leave room for timely adjustments to investment strategies. Although three events cannot establish how reliably such lead times recur across future onsets, these results are best read as out-of-sample evidence whose generalisation remains to be tested, rather than as a fully validated operating regime. This performance is consistent with the framework's design objective of balancing sensitivity to early regime change (SOHMM) with event clustering intensity (Q-Hawkes).

Figure 4 presents the model's output across both its raw and calibrated forms alongside the underlying asset price. Crisis probabilities rise sharply during the early accumulation phases of drawdowns rather than exclusively at their deepest points. The 2021 tightening cluster exemplifies regime coherence: both the calibrated and raw series recognised the shared monetary policy shock across all three subsequent anchors, issuing cohesive anticipatory alerts ahead of the unfolding decline. The isotonic calibration compresses the output range while preserving the signal's temporal structure, with the calibrated series issuing cleaner, more decisive threshold crossings relative to the raw output.

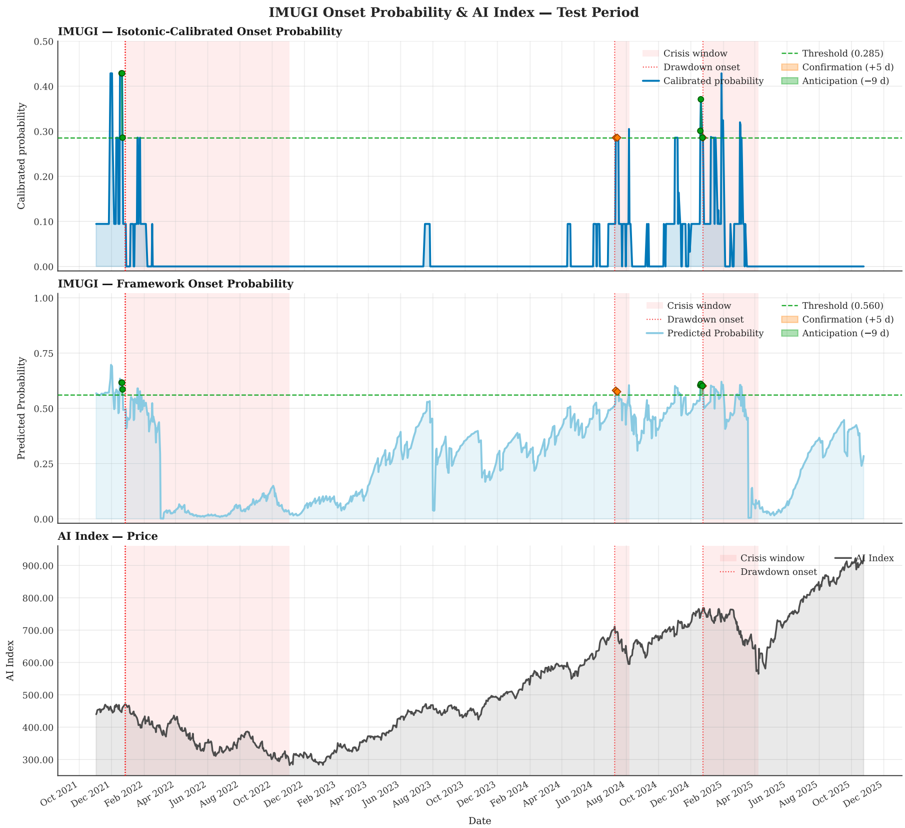

Figure 4. Crisis onset probability and AI Index price over the test period, 2021–10–02 to 2025–10–24. *Top panel*: isotonic-calibrated probability $\widehat{P}_{\mathrm{cal}}(t)$ (left axis, range [0, 0.5]), threshold $\tau^*_{\mathrm{cal}} = 0.285$ (dashed). *Middle panel*: raw model probability $\widehat{P}_{\mathrm{raw}}(t)$, threshold $\tau^*_{\mathrm{raw}} = 0.560$ (dashed). Red-shaded regions: ground-truth crisis windows; dotted vertical lines: drawdown onsets. Green-filled regions with circle markers: anticipatory detections up to nine days prior to onset; orange-filled regions with diamond markers: confirmatory detections within five days after onset. *Bottom panel*: AI Index price (right axis)

5. Discussion

IMUGI's core contribution is anticipating overvaluation-driven drawdown onsets at peak rather than identifying them at trough. Network-topological, Laplacian-energy, and recent machine-learning approaches (Tölö, 2020; Zhao *et al.*, 2025; Rao and Rojas, 2025; Ghasemieh and Kashef, 2022) report strong retrospective classification but locate detections near the deepest drawdown point, by which time the institutional hedging window has closed. Detection of all three test-window onsets at or before materialisation is consistent with onset prediction at actionable lead time being feasible, although three events are too few to establish this as a general operating regime.

The meta-model coefficients reveal an interpretable joint-sufficiency structure. The negative loading on raw Hawkes intensity ($\beta_1 = -6.33$) combined with the large positive SOHMM–Hawkes interaction ($\beta_5 = +5.50$) means jump clustering elevates crisis probability only when concurrent regime instability is detected. Neither signal is sufficient alone, consistent with the regime-dependent jump structure documented by Blanc *et al.* (2017) and Chen *et al.* (2024). The ablation reinforces this: the principal gain arises from combining SOHMM and Hawkes signals, with subsequent layers redistributing rather than uniformly improving performance — a pattern consistent with structural rather than memorised signal extraction.

The framework also exhibits a notable near-miss. Crisis probability rose through late October 2023, approached the alert threshold, then fell on 1–2 November 2023 with no accompanying price decline — coinciding with Federal Reserve Chair Powell's press conference, in which he rejected near-term rate cuts and signalled a "higher for longer" regime (Federal Reserve, 2023, pp. 5–16). This is consistent with the framework's indeterministic character: stabilising policy signals can resolve some of the conditions that elevate drawdown probability, so an alert may subside without a subsequent crash. We read this single episode as illustrative rather than causal evidence, but it shows the model retracting a warning in a way that has an economically coherent interpretation.

6. Limitations

Two constraints bind the strength of any inferential claim. The first is event sparsity: twelve training onsets and three test events do not support strong claims about hit rates in general, and the test events share identifiable monetary-policy components. The second is mechanistic scope: the LPPL layer is built to detect overvaluation relative to fundamental value, so portability is expected primarily to other asset classes where overvaluation-driven drawdowns are the dominant crisis mechanism — high-multiple growth sectors, cryptocurrencies during bubble episodes, real estate during accumulation phases — rather than to liquidity dislocations, geopolitical shocks, or contagion-driven contractions.

The constituent universe was defined with knowledge of the contemporary AI sector; market-capitalisation weighting structurally attenuates the influence of recently-listed firms in earlier periods, but the index should be read as representing the AI technology sector as currently identified rather than a real-time investable portfolio.

7. Conclusion

IMUGI integrates three components — regime memory through a Second-Order Hidden Markov Model, self-excitation through a Quadratic Hawkes process, and bubble physics through LPPL-derived features — into a single online probabilistic architecture targeted at the onset of overvaluation-driven drawdowns in AI technology equities. The framework's empirical performance, robustness across event-definition choices, and outperformance of standard survival and machine-learning benchmarks together support the methodological proposition that drawdown onset prediction benefits from joint rather than single-mechanism modelling.

Three extensions follow naturally. Replication on non-US AI equity indices and cross-asset bubbles — cryptocurrencies, real estate during accumulation — would test whether the mechanistic priors generalise beyond NASDAQ-listed technology firms. Integrating external market signals as additional meta-model features may sharpen onset-versus-continuation discrimination within the early-warning window. Extension to other geographies with rapidly accelerating AI sectors — notably the People's Republic of China and the Republic of Korea — offers a natural cross-market test of the framework's calibration. A formal economic-value analysis, with backtested hedging or position-sizing rules, is the most direct path to translating IMUGI's calibrated onset probabilities into measured institutional benefit.

Authors contributions

Maksim Kim handled aspects of research execution, from conceptualisation through writing. Seungwook Park provided academic supervision and project administration.

Use of AI tools declaration

The English in this paper has been refined using Artificial Intelligence tools.

Acknowledgments
The authors thank 2 reviewers and the editors for helpful comments.

Note

1. A greedy algorithm repeatedly makes a locally optimal choice under the heuristic assumption that this leads to a globally optimal solution (Vince, 2002).

Supplementary material
The supplementary material for this article can be found online.

References

Akerlof, G.A. and Shiller, R.J. (2009), *Animal Spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism*, Princeton University Press, Princeton, NJ.

Alswaidan, A. and Varner, J.D. (2026), "Hybrid hidden Markov model for modeling equity excess growth rate dynamics: a discrete-state approach with jump-diffusion", arXiv preprint arXiv: 2603.10202.

An, S., Gao, X., An, F. and Wu, T. (2025), "Early warning of regime switching in a complex financial system from a spillover network dynamic perspective", *iScience*, Vol. 28 No. 3, 111924, doi: 10.1016/j.isci.2025.111924.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Ebens, H. (2001), "The distribution of realized stock return volatility", *Journal of Financial Economics*, Vol. 61 No. 1, pp. 43-76.

Ang, A. and Bekaert, G. (2002), "International asset allocation with regime shifts", *Review of Financial Studies*, Vol. 15 No. 4, pp. 1137-1187, doi: 10.1093/rfs/15.4.1137.

Ang, A. and Timmermann, A. (2012), "Regime changes and financial markets", *Annual Review of Financial Economics*, Vol. 4 No. 1, pp. 313-337, doi: 10.1146/annurev-financial-110311-101808.

Aroussi, R. (2019), "yfinance: download market data from Yahoo! Finance's API", available at: https://github.com/ranaroussi/yfinance (accessed 19 November 2025).

Bae, K.H., Karolyi, G.A. and Stulz, R.M. (2003), "A new approach to measuring financial contagion", *Review of Financial Studies*, Vol. 16 No. 3, pp. 717-763, doi: 10.1093/rfs/hhg012.

Baum, L.E. and Petrie, T. (1966), "Statistical Inference for probabilistic functions of finite state Markov chains", *Annals of Mathematical Statistics*, Vol. 37 No. 6, pp. 1554-1563, doi: 10.1214/aoms/1177699147.

Baum, L.E., Petrie, T., Soules, G. and Weiss, N. (1970), "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains", *Annals of Mathematical Statistics*, Vol. 41 No. 1, pp. 164-171, doi: 10.1214/aoms/1177697196.

Berg, A. and Pattillo, C. (1999), "Predicting currency crises: the indicators approach and an alternative", *Journal of International Money and Finance*, Vol. 18 No. 4, pp. 561-586.

Blanc, P., Donier, J. and Bouchaud, J.-P. (2017), "Quadratic Hawkes processes for financial prices", *Quantitative Finance*, Vol. 17 No. 2, pp. 171-188, doi: 10.1080/14697688.2016.1193215.

Bothmer, H.C. and Meister, C. (2003), "Predicting critical crashes? A new restriction for the free variables", *Physica A: Statistical Mechanics and its Applications*, Vol. 320, pp. 539-547, doi: 10.1016/s0378-4371(02)01535-2.

Bowsher, C.G. (2007), "Modelling security market events in continuous time: intensity based, multivariate point process models", *Journal of Econometrics*, Vol. 141 No. 2, pp. 876-912, doi: 10.1016/j.jeconom.2006.11.007.

Brown, S.J., Goetzmann, W.N., Ibbotson, R.G. and Ross, S.A. (1992), "Survivorship bias in performance studies", *The Review of Financial Studies*, Vol. 5 No. 4, pp. 553-580, doi: 10.1093/rfs/5.4.553.

Bulla, J. (2011), "Hidden Markov models with *t* components. Increased persistence and other aspects", *Quantitative Finance*, Vol. 11 No. 3, pp. 459-475, doi: 10.1080/14697681003685563.

Bussière, M. and Fratzscher, M. (2006), "Towards a new early warning system of financial crises", *Journal of International Money and Finance*, Vol. 25 No. 6, pp. 953-973, doi: 10.1016/j.jimonfin.2006.07.007.

Byrd, R.H., Lu, P., Nocedal, J. and Zhu, C. (1995), "A limited memory algorithm for bound constrained optimization", *SIAM Journal on Scientific Computing*, Vol. 16 No. 5, pp. 1190-1208, doi: 10.1137/0916069.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.

Cao, Z., Shao, X., Yan, Y. and Geman, H. (2025), "Identifying and quantifying financial bubbles with the hyped log-periodic power law model", available at: https://arxiv.org/abs/2510.10878 (accessed 5 January 2026).

Cappé, O., Moulines, E. and Rydén, T. (2009), *Inference in Hidden Markov Models*, Springer, New York, NY.

Chapelle, O., Schölkopf, B. and Zien, A. (Eds) (2006), *Semi-Supervised Learning*, MIT Press, Cambridge, MA.

Chekhlov, A., Uryasev, S. and Zabarankin, M. (2005), "Drawdown measure in portfolio optimization", *International Journal of Theoretical and Applied Finance*, Vol. 8 No. 1, pp. 13-58, doi: 10.1142/s0219024905002767.

Chen, J., Clements, M.P. and Urquhart, A. (2024), "Modeling price and variance jump clustering using the marked Hawkes process", *The Journal of Financial Econometrics*, Vol. 22 No. 3, pp. 743-772, doi: 10.1093/jjfinec/nbad007.

Cox, D.R. (1972), "Regression models and life-tables", *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 34 No. 2, pp. 187-220, doi: 10.1111/j.2517-6161.1972.tb00899.x.

Daley, D.J. and Vere-Jones, D. (2008), *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*, 2nd ed., Springer, New York, NY.

Drehmann, M. and Juselius, M. (2014), "Evaluating early warning indicators of banking crises: satisfying policy requirements", *International Journal of Forecasting*, Vol. 30 No. 3, pp. 759-780, doi: 10.1016/j.ijforecast.2013.10.002.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

Engle, R.F. and Manganelli, S. (2004), "CAViaR: conditional autoregressive value at risk by regression quantiles", *Journal of Business and Economic Statistics*, Vol. 22 No. 4, pp. 367-381.

Federal Reserve (2023), *Transcript of Chair Powell's Press Conference*, Federal Reserve, Washington, DC.

Filimonov, V. and Sornette, D. (2012), "Quantifying reflexivity in financial markets: towards a prediction of flash crashes", *Physical Review E*, Vol. 85 No. 5, 056108, doi: 10.1103/physreve.85.056108.

Frankel, J. and Saravelos, G. (2012), "Can leading indicators assess country vulnerability? Evidence from the 2008-09 global financial crisis", *Journal of International Economics*, Vol. 87 No. 2, pp. 216-231, doi: 10.1016/j.jinteco.2011.12.009.

Ghasemieh, A. and Kashef, R. (2022), "A robust deep learning model for predicting the trend of stock market prices during market crash periods", *2022 IEEE International Systems Conference (SysCon)*, IEEE, pp. 1-8.

Guan, L. and Tibshirani, R. (2022), "Prediction and outlier detection in classification problems", *Journal of the Royal Statistical Society: Series B*, Vol. 84 No. 2, pp. 524-546, doi: 10.1111/rssb.12443.

Guidolin, M. and Timmermann, A. (2007), "Asset allocation under multivariate regime switching", *Journal of Economic Dynamics and Control*, Vol. 31 No. 11, pp. 3503-3544, doi: 10.1016/j.jedc.2006.12.004.

Hamilton, J.D. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle", *Econometrica*, Vol. 57 No. 2, pp. 357-384, doi: 10.2307/1912559.

Harvey, D., Leybourne, S. and Newbold, P. (1997), "Testing the equality of prediction mean squared errors", *International Journal of Forecasting*, Vol. 13 No. 2, pp. 281-291, doi: 10.1016/s0169-2070(96)00719-4.

Hawkes, A.G. (1971), "Point spectra of some mutually exciting point processes", *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 33 No. 3, pp. 438-443, doi: 10.1111/j.2517-6161.1971.tb01530.x.

Johansen, A. and Sornette, D. (2010), "Shocks, crashes and bubbles in financial markets", *Brussels Economic Review – Cahiers Economiques de Bruxelles*, Vol. 53 No. 2, pp. 201-230.

Johansen, A., Ledoit, O. and Sornette, D. (2000), "Crashes as critical points", *International Journal of Theoretical and Applied Finance*, Vol. 3 No. 2, pp. 219-255, doi: 10.1142/s0219024900000115.

Jordà, O., Schularick, M. and Taylor, A.M. (2015), "Leveraged bubbles", *Journal of Monetary Economics*, Vol. 76, pp. S1-S20.

Juang, B.-H. and Rabiner, L.R. (1991), "Hidden Markov models for speech recognition", *Technometrics*, Vol. 33 No. 3, pp. 251-272, doi: 10.1080/00401706.1991.10484833.

Kaminsky, G.L., Lizondo, S. and Reinhart, C.M. (1998), "Leading indicators of currency crises", *IMF Staff Papers*, Vol. 45 No. 1, pp. 1-48, doi: 10.2307/3867328.

Koistinen, J. (2020), "The log-periodic power law: evidence from the Finnish stock market", Unpublished Master's thesis, Aalto University School of Business, Espoo.

Kuhn, H.W. (1955), "The Hungarian method for the assignment problem", *Naval Research Logistics Quarterly*, Vol. 2 Nos 1-2, pp. 83-97, doi: 10.1002/nav.3800020109.

Lee, S.S. and Mykland, P.A. (2008), "Jumps in financial markets: a new nonparametric test and jump dynamics", *The Review of Financial Studies*, Vol. 21 No. 6, pp. 2535-2563, doi: 10.1093/rfs/hhm056.

Little, R.J.A. and Rubin, D.B. (2002), *Statistical Analysis with Missing Data*, 2nd ed., Wiley, Hoboken, NJ.

Mahalanobis, P.C. (1936), "On the generalized distance in statistics", *Proceedings of the National Institute of Sciences of India*, Vol. 2 No. 1, pp. 49-55.

Mandelbrot, B. (1963), "The variation of certain speculative prices", *The Journal of Business*, Vol. 36 No. 4, pp. 394-419, doi: 10.1086/294632.

Mandelbrot, B. and Hudson, R. (2004), *The (Mis)Behavior of Markets: A Fractal View of Risk, Ruin, and Reward*, Basic Books, New York, NY.

Mari, J.-F., Haton, J.-P. and Kriouile, A. (1997), "Automatic word recognition based on second-order hidden Markov models", *IEEE Transactions on Speech and Audio Processing*, Vol. 5 No. 1, pp. 22-25, doi: 10.1109/89.554265.

Murphy, K.P. (2012), *Machine Learning: A Probabilistic Perspective*, MIT Press, Cambridge, MA.

Nolde, N. and Zhou, C. (2021), "Extreme value analysis for financial risk management", *Annual Review of Statistics and its Application*, Vol. 8 No. 1, pp. 217-240, doi: 10.1146/annurev-statistics-042720-015705.

Nystrup, P., Lindström, E. and Madsen, H. (2020), "Learning hidden Markov models with persistent states by penalizing jumps", *Expert Systems with Applications*, Vol. 150, 113307, doi: 10.1016/j.eswa.2020.113307.

Ofek, E. and Richardson, M. (2003), "DotCom mania: the rise and fall of internet stock prices", *The Journal of Finance*, Vol. 58 No. 3, pp. 1113-1137, doi: 10.1111/1540-6261.00560.

Ogata, Y. (1978), "The asymptotic behaviour of maximum likelihood estimators for stationary point processes", *Annals of the Institute of Statistical Mathematics*, Vol. 30 No. 1, pp. 243-261, doi: 10.1007/bf02480216.

Peel, D. and McLachlan, G.J. (2000), "Robust mixture modelling using the *t* distribution", *Statistics and Computing*, Vol. 10 No. 4, pp. 339-348, doi: 10.1023/a:1008981510081.

Rabiner, L.R. (1989), "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE*, Vol. 77 No. 2, pp. 257-286, doi: 10.1109/5.18626.

Rao, P. and Rojas, R.R. (2025), "Predicting market troughs: a machine learning approach with causal interpretation", available at: https://arxiv.org/abs/2509.05922 (accessed 8 January 2026).

Rousseeuw, P.J. and Van Zomeren, B.C. (1990), "Unmasking multivariate outliers and leverage points", *Journal of the American Statistical Association*, Vol. 85 No. 411, pp. 633-639, doi: 10.1080/01621459.1990.10474920.

S&P Dow Jones Indices (2022), *S&P U.S. Indices Methodology*, S&P Dow Jones Indices LLC, New York, NY, available at: https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-us-indices.pdf

Scheinkman, J.A. and Xiong, W. (2003), "Overconfidence and speculative bubbles", *Journal of Political Economy*, Vol. 111 No. 6, pp. 1183-1220, doi: 10.1086/378531.

Schularick, M. and Taylor, A.M. (2012), "Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870-2008", *American Economic Review*, Vol. 102 No. 2, pp. 1029-1061, doi: 10.1257/aer.102.2.1029.

Shu, M. and Zhu, W. (2019), "Detection of Chinese stock market bubbles with LPPLS confidence indicator", available at: https://arxiv.org/abs/1905.09640 (accessed 20 January 2026).

Shumway, T. (2001), "Forecasting bankruptcy more accurately: a simple hazard model", *The Journal of Business*, Vol. 74 No. 1, pp. 101-124, doi: 10.1086/209665.

Sornette, D. (2003), *Why Stock Markets Crash: Critical Events in Complex Financial Systems*, Princeton University Press, Princeton, NJ.

Sornette, D. (2009), "Dragon-kings, black swans and the prediction of crises", *International Journal of Terraspace Science and Engineering*, Vol. 2, pp. 1-18.

Sornette, D. and Ouillon, G. (2012), "Dragon-kings: mechanisms, statistical methods and empirical evidence", *European Physical Journal Special Topics*, Vol. 205, pp. 1-26, doi: 10.1140/epjst/e2012-01559-5.

Sornette, D., Demos, G., Zhang, Q., Cauwels, P., Filimonov, V. and Zhang, Q.Z. (2015), "Real-time prediction and post-mortem analysis of the Shanghai 2015 stock market bubble and crash", working paper No. 15-31, Swiss Finance Institute, Geneva.

Taleb, N.N. (2007), *The Black Swan: The Impact of the Highly Improbable*, Random House, New York, NY.

Tölö, E. (2020), "Predicting systemic financial crises with recurrent neural networks", *Journal of Financial Stability*, Vol. 49, 100746, doi: 10.1016/j.jfs.2020.100746.

van Hemert, O., Ganz, M., Harvey, C.R., Rattray, S., Martin, E.S. and Yawitch, D. (2020), "Drawdowns", *Journal of Portfolio Management*, Vol. 46 No. 8, pp. 34-50, doi: 10.3905/jpm.2020.1.170.

Vince, A. (2002), "A framework for the greedy algorithm", *Discrete Applied Mathematics*, Vol. 121 Nos 1-3, pp. 247-260, doi: 10.1016/s0166-218x(01)00362-6.

Williams, C.K.I. (2021), "The effect of class imbalance on precision-recall curves", *Neural Computation*, Vol. 33 No. 4, pp. 853-857, doi: 10.1162/neco_a_01362.

Zadrozny, B. and Elkan, C. (2002), "Transforming classifier scores into accurate multiclass probability estimates", *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 694-699, doi: 10.1145/775047.775151.

Zhao, X., Huang, C., Yang, X., Cao, J. and Yang, X. (2025), "Can we better predict financial crisis? The role of Laplacian-energy-like measure", *International Review of Economics and Finance*, Vol. 103, 104396, doi: 10.1016/j.iref.2025.104396.

Corresponding author
Seungwook Park can be contacted at: separk6112@inha.ac.kr