Article

Not peer-reviewed version

S-NODE-ANF-RRC: Stochastic Neural ODE for Financial Regime Forecasting and False Alarm Control on JSE Equities

Ntebogang Dinah Moroke *

Posted Date: 15 May 2026

doi: 10.20944/preprints202605.1066.v1

Keywords: financial regime forecasting; time series forecasting; stochastic neural ODE; stochastic volatility; probabilistic forecasting; false alarm rate; Johannesburg Stock Exchange; emerging markets; heavy-tailed distributions; interpretable machine learning; SDG 8; SDG 9; SDG 10; SDG 16; SDG 17

 doi:10.20944/preprints202605.1066.v1

Article

S-NODE-ANF-RRC: Stochastic Neural ODE for Financial Regime Forecasting and False Alarm Control on JSE Equities

Ntebogang Dinah Moroke

Department of Statistics and Operations Research, Faculty of Economic and Management Sciences, North-West University, Mafikeng Campus, Private Bag X2046, Mmabatho 2735, South Africa; ntebo.moroke@nwu.ac.za

Highlights

- Gaussian mixture clustering applied to raw heavy-tailed JSE features (kurtosis 54.8) inflates ARI by 1.3×; log-transformation corrects this systematic measurement artefact.
- The N-ODE-ANF-RRC achieves the lowest operational cost (10,350 bp, 65.1% below GMM) and the longest crisis lead time (0.71 days), with 66% of crisis episodes detected at least one full trading day in advance.
- The S-NODE-ANF-RRC achieves the lowest false alarm rate among probabilistic architectures (FAR $= 0.051$), with a 42.0% cost reduction versus GMM (bootstrap 95% CI $[5,250,\ 19,600]$ bp, excludes zero; McNemar $\chi^2 = 4.923$, $p = 0.027$).
- Ablation confirms drift, diffusion, and dual-loss as the minimum viable daily-frequency configuration; removing any single component degrades both ARI and operational cost.
- S-NODE-ANF-RRC maintains positive ARI under Gaussian noise injection at $\sigma_\eta \in \{0.5, 1.0, 1.5\}$, while GMM and N-ODE-ANF-RRC collapse below zero at $\sigma_\eta \geq 1.0$.

Abstract

Emerging-market equity exchanges require regime forecasting systems that are continuous in time, robust to heavy-tailed distributions, and optimised against false alarms. No existing method addresses all three simultaneously, and no prior study has reported a crisis false alarm rate on JSE equities. We propose S-NODE-ANF-RRC: a Stochastic Neural ODE embedded within an Adaptive Neuro-Fuzzy Risk-Regime Clustering architecture, motivated by the Heston stochastic volatility framework and integrated by a Milstein scheme with Lyapunov-regularised dual-loss training. The system is evaluated as a one-step-ahead probabilistic forecaster ($h = 1$ trading day) on 2696 daily observations across 17 JSE securities (March 2015–March 2026). Gaussian mixture clustering on raw features (kurtosis 54.8) inflates ARI by 1.3×; log-transformation corrects this systematic artefact. Two operational profiles emerge after correction: the N-ODE-ANF-RRC achieves the lowest cost (10,350 bp, 65.1% below GMM), longest lead time (0.71 days), and best MCC (0.596); the S-NODE-ANF-RRC achieves the lowest false alarm rate among probabilistic architectures (FAR = 0.051, log-loss = 1.07), with a 42.0% cost reduction versus GMM (bootstrap 95% CI $[5,250,\ 19,600]$ bp; McNemar $p = 0.027$). Ablation confirms drift, diffusion, and dual-loss as the minimum viable daily-frequency configuration. The interdisciplinary fusion of physics-informed SDE dynamics, time series forecasting, and fuzzy interpretability yields two complementary JSE risk tools: an early-warning forecaster (N-ODE) and a low-false-alarm crisis classifier (S-NODE). Code and data: https://doi.org/10.5281/zenodo.19787658.

Keywords: financial regime forecasting; time series forecasting; stochastic neural ODE; stochastic volatility; probabilistic forecasting; false alarm rate; Johannesburg Stock Exchange; emerging markets; heavy-tailed distributions; interpretable machine learning; SDG 8; SDG 9; SDG 10; SDG 16; SDG 17

 doi:10.20944/preprints202605.1066.v1

1. Introduction

The volatility of equity returns obeys a continuous-time stochastic differential equation (SDE):

$$d\sigma_t = \kappa(\theta - \sigma_t)\,\mathrm{d}t + \xi\sqrt{\sigma_t}\,\mathrm{d}W_t, \tag{1}$$

where κ is the mean-reversion rate, θ the long-run variance, and ξ the volatility-of-volatility [1,2]. This is not a modelling choice – it is the physical description of the process, from which the Black-Scholes equation and the rough volatility framework all derive. Regime detection classifies the qualitative state of this continuous process: which region of the SDE state space currently prevails? Yet existing regime detection systems almost universally discretise the continuous-time process into fixed daily steps, impose Gaussian emission densities that are incompatible with the heavy-tailed nature of financial volatility [3], and have never been designed against an explicit false alarm rate objective.

This paper asks: can a regime forecasting system be *continuous in time* (like the physical process it models), *stochastic in structure* (like the SDE it approximates), *interpretable in output* (like the fuzzy rule systems that practitioners trust), and *optimal in false alarms* (like the operational systems that risk managers actually deploy)? The S-NODE-ANF-RRC is the answer.

Interdisciplinary framing. This paper occupies the intersection of three fields that have developed largely in parallel. *Physics-informed machine learning* provides continuous-time neural SDE dynamics that preserve the mathematical structure of stochastic processes [4,5]. *Time series forecasting* provides the Box-Jenkins [6] identification-estimation-verification paradigm that requires formal stability and power analysis before any forecasting model is accepted for deployment [7]. *Fuzzy expert systems* provide interpretable regime scores from latent state representations [8,9]. The novelty of this paper is the disciplined integration of all three: the SDE governs latent dynamics, the forecasting paradigm governs evaluation, and the fuzzy layer governs output interpretability. Critically, Section 4 demonstrates that the classical stability and power requirements of [6,7] and the Lyapunov stability requirement of stochastic analysis [10] are both satisfied simultaneously by the dual-loss training design.

The physics motivation. Three properties of the JSE equity panel, formally established in Section 3, place this study within the continuous-time stochastic volatility tradition. First, the Hurst exponent of realised volatility ($H = 0.909$) indicates strong long memory, consistent with rough volatility [2] and inconsistent with short-memory GARCH dynamics. Second, the extreme kurtosis of the cross-sectional features violates the Gaussian emission assumption of standard mixture models, producing artefactually inflated clustering scores documented in Section 8. Third, self-transition probabilities above 0.90 confirm strong regime persistence that justifies a Markovian latent state formulation with explicit transition dynamics [11]. Together, these three facts motivate an SDE whose drift captures persistence, whose diffusion absorbs heavy-tailed shocks, and whose Milstein integration step matches the empirical autocorrelation decay of the JSE volatility process.

The forecasting objective. This paper frames regime detection as a probabilistic one-step-ahead forecasting problem with horizon $h = 1$ trading day: given the information set available at market close on day t, predict the regime class $\hat{y}_{t+1} \in \{0, 1, 2\}$ (Normal, Stressed, Crisis) for day $t + 1$. This is the operationally relevant objective. A system that retrospectively classifies past regimes correctly contributes nothing to a portfolio manager who needs one day of advance warning to adjust hedges before a drawdown materialises.

The structural gap. Two gaps in the literature prevent this forecasting objective from being met. The first is *architectural*: no existing hybrid system combines continuous-time stochastic neural dynamics with fuzzy inference and evaluates against an explicit false alarm rate objective. The second is a *measurement problem*: Gaussian mixture clustering applied to raw heavy-tailed financial features produces clustering scores that reflect distributional artefacts rather than genuine regime structure [12]. Both gaps are addressed in this paper.

Contributions.

 doi:10.20944/preprints202605.1066.v1

1. A hybrid S-NODE-ANF-RRC architecture that is, to our knowledge, the first to combine continuous-time stochastic neural dynamics with fuzzy inference and explicit false alarm rate optimisation.
2. Identification and correction of a systematic kurtosis measurement artefact in regime clustering that affects all Gaussian mixture-based benchmarks on heavy-tailed data (Section 8).
3. A unified stability and power framework (Section 4) that bridges classical forecasting tests (CUSUM, Neyman-Pearson), SDE Lyapunov stability, and neural network training diagnostics within a single theoretical argument.
4. A dual operational profile: the N-ODE-ANF-RRC as the primary early-warning forecaster with positive crisis lead time, and S-NODE-ANF-RRC as the low-false-alarm crisis classifier (Section 8, Table 5).
5. Open data, code, and trained model weights at https://doi.org/10.5281/zenodo.19787658.

The remainder of the paper is structured as follows. Section 2 reviews related work. Section 3 presents data and diagnostics. Section 4 provides the unified stability and power framework. Section 5 develops the theoretical framework. Section 6 specifies the architecture. Section 7 describes the experimental design. Section 8 reports results. Section 9 discusses implications. Section 10 concludes.

2. Related Work

The literature on financial regime detection spans four research traditions that have developed largely in parallel. Table 1 provides a structured synthesis across 14 representative studies on nine dimensions: continuous-time formulation (CT), stochastic component (Stoch), early-warning capability (EW), false alarm rate reporting (FAR), emerging market application (EM), JSE data (JSE), data type, forecast horizon, and primary evaluation metric. The following subsections summarise each tradition and formalise the convergence that this paper exploits.

Table 1. Structured literature synthesis across nine dimensions. CT: continuous-time formulation. Stoch: stochastic component. EW: early-warning capability (lead time > 0). FAR: false alarm rate reported or optimised. EM: emerging market application. JSE: JSE data used. Data: time series type. Metric: primary evaluation metric. ✓ = Yes; ∼ = Partial; × = No. **No prior study optimises or reports crisis FAR** (0% prior coverage): this is the primary gap addressed.

Study	Method	CT	Stoch	EW	FAR	EM	JSE	Data type	Metric
Hamilton (1989) [13]	MS-AR	×	×	×	×	×	×	Macro series	Log-likelihood
Tsay (1989) [14]	TAR	×	×	×	×	×	×	Macro series	AIC/BIC
Graves (2012) [15]	LSTM	×	×	∼	×	×	×	Sequence	Accuracy
Dey & Salem (2017) [16]	GRU	×	×	∼	×	×	×	Text/returns	F1
Vaswani et al. (2017) [17]	Transformer	×	×	∼	×	×	×	Sequence	Accuracy
Chen et al. (2018) [18]	Neural ODE	✓	×	×	×	×	×	Latent series	MSE
Tzen & Raginsky (2019) [4]	Neural SDE	✓	✓	×	×	×	×	Latent series	ELBO
Jia & Benson (2019) [19]	NJ-SDE	✓	✓	∼	×	×	×	Irregular TS	MSE
Kaur (2019) [9]	ANFIS	×	×	×	×	×	×	Returns	RMSE
Kidger et al. (2020) [20]	Neural CDE	✓	×	×	×	×	×	Irregular TS	Accuracy
Wang et al. (2020) [21]	Deep FCM	×	×	×	×	×	×	Returns	ARI
Ardia et al. (2019) [22]	MS-GARCH	×	✓	×	×	×	×	Returns	Log-lik.
Lea et al. (2016) [23]	TCN	×	×	∼	×	×	×	Sequence	Accuracy
Zhang et al. (2022) [24]	TFT-Finance	×	×	∼	×	∼	×	Returns	MAE/RMSE
This paper	**S-NODE-ANF-RRC**	✓	✓	✓	✓	✓	✓	JSE equities	FAR, cost, LL
Prior coverage		29%	21%	0%	**0%**	7%	0%		
This paper		100%	100%	100%	**100%**	100%	100%		

∼ = partial (lead time reported but not optimised against FAR). **FAR 0% prior coverage**: no prior study optimises or reports crisis false alarm rate. LL: log-loss.

2.1. Parametric and Neural Architectures for Regime Detection

Parametric regime-switching models. Hamilton [13] established the canonical Markov-switching autoregressive (MS-AR) framework for discrete regime detection in macroeconomic time series; Tsay [14] extended this to threshold autoregression. MS-GARCH variants [22,25] accommodate volatility clustering under regime change but impose Gaussian emission densities that are, as Table 3 confirms, incompatible with the extreme kurtosis of JSE features. Adaptive hierarchical HMM variants [26]

 doi:10.20944/preprints202605.1066.v1

represent the most recent parametric development but retain zero lead time and produce no calibrated probabilistic output.

Discrete-time neural architectures. Graves [15] demonstrated LSTM's capacity for long-range sequential dependencies. Dey and Salem [16] and Vaswani et al. [17] advanced gating and attention. Applied to financial forecasting, these architectures operate in discrete time – discretising the continuous-time SDE that governs the underlying process [1] – and have not been evaluated against an explicit false alarm rate criterion [27].

2.2. Continuous-Time Models for Financial Volatility

Heston [1] demonstrated that equity volatility obeys a mean-reverting SDE, a result that motivated a generation of option-pricing and risk models. Gatheral et al. [2] showed that individual equity volatility follows a *rough* path with Hurst exponent $H \approx 0.1$, invalidating the diffusion assumptions of classical models. Our JSE cross-sectional panel exhibits $H = 0.909$ (Table 3), reflecting long memory in the aggregated volatility series, consistent with the cross-sectional averaging of 17 securities smoothing short-term roughness while preserving long-run persistence.

Chen et al. [18] introduced Neural ODEs as data-driven continuous-time models; Tzen and Raginsky [4] and Jia and Benson [19] completed the stochastic extension. Yang et al. [28] showed that neural SDEs with α-stable noise outperform deterministic networks on heavy-tailed financial series; [29,30] provide further empirical confirmation. The S-NODE-ANF-RRC inherits this lineage directly: the drift f_θ learns the mean-reversion dynamics, and the diffusion g_ϕ learns the volatility-of-volatility, both estimated from JSE data without parametric restriction.

2.3. Fuzzy and Hybrid Systems for Regime Classification

Kaur [9] demonstrated ANFIS-based regime clustering; Su and Wei [8] applied ANFIS and k-means to financial early warning, providing the closest contemporary comparator for the ANF-RRC component. Kidger et al. [20] unified CDEs as a continuous-time recurrent alternative. None of these approaches combines the continuous-time SDE dynamics with a fuzzy output layer and an explicit false alarm objective.

2.4. Synthesis: Convergences, Gaps, and Research Motivation

Table 1 reveals three convergences that this paper exploits and one gap that it addresses.

Convergence 1 (continuous time is necessary). The progression from MS-AR through Neural ODE to Neural SDE reflects a growing recognition that financial volatility is a continuous-time process [1,2] that discrete-time architectures can only approximate. CT coverage rises from 0% (parametric models) to 29% (neural models), but no hybrid system achieves full CT coverage until this paper.

Convergence 2 (stochasticity improves heavy-tail robustness). The systematic review of [12] confirms that stochastic architectures consistently outperform deterministic baselines on heavy-tailed financial data – exactly the distributional regime of the JSE panel. Stoch coverage among reviewed studies is 21%, confirming that most systems ignore the diffusion channel entirely.

Convergence 3 (emerging markets are underserved). Only 7% of reviewed studies target emerging markets, despite evidence [31,32] that conventional early warning systems systematically fail in heavy-tailed, tail-dependent data environments. JSE coverage is 0% prior to this paper.

The gap. No prior study in Table 1 optimises or even reports a crisis false alarm rate (FAR coverage: 0%). Early warning coverage is similarly 0% – no prior system evaluates crisis lead time as a primary metric. This paper addresses both gaps simultaneously within the continuous-time stochastic framework motivated by the three convergences above.

3. Data and Diagnostics

3.1. Dataset

Daily closing prices for 17 JSE-listed equity securities span 2 January 2015 to 30 March 2026, sourced from Yahoo Finance (`yfinance` v0.2). After a 60-day rolling-window initialisation burn-in,

 doi:10.20944/preprints202605.1066.v1

the usable panel contains $N = 2,696$ post-burn-in daily observations covering 27 March 2015 to 30 March 2026. Securities are: ABG (Absa), AGL (Anglo American), ANG (AngloGold Ashanti), BVT (Bidvest), CPI (Capitec), FSR (FirstRand), GFI (Gold Fields), IMP (Implats), MTN (MTN Group), NED (Nedbank), NPN (Naspers), REM (Remgro), SBK (Standard Bank), SHP (Shoprite), SLM (Sanlam), SOL (Sasol), VOD (Vodacom); spanning financial services, mining, energy, telecommunications, consumer staples, diversified financials, and insurance. The CBOE VIX serves as the exogenous market stress covariate [33]. All prices are sourced from Yahoo Finance. For recent deep neural approaches to JSE equity forecasting, see [27].

Table 2. Emerging-market stylised facts and S-NODE-ANF-RRC architectural responses. The JSE is not merely the application domain; its structural properties directly motivate each design choice.

Stylised Fact	JSE Evidence	Architectural Response
Extreme kurtosis	$\overline{\sigma}_t$: 54.8; $\bar{\varepsilon}_t$: 824.0	Log-transform before modelling; GMM on raw features inflates ARI by 1.3×
Long memory in vol	Hurst $H = 0.909$ ($\overline{\sigma}_t$), $H = 0.941$ (VIX)	Continuous-time SDE naturally models persistent path-dependent dynamics; discrete-time RNNs discretise this artificially
Near-random-walk residuals	Hurst $H = 0.527$ ($\bar{\varepsilon}_t$)	Diffusion term g_ϕ captures stochastic component without imposing directional persistence
Regime persistence	Self-transition: Normal 0.934, Crisis 0.923	Early-warning objective: rare transitions make advance detection operationally valuable

3.2. Feature Construction

Three informative cross-sectional features are constructed daily. The cross-sectional mean of 60-day rolling market-model betas is excluded because it is identically 1.0 by construction of the equal-weight market proxy and carries no temporal information.

Definition 1 (Feature Vector). *The daily feature vector is:*

$$\mathbf{x}_t = \left[\overline{\sigma}_t,\ \bar{\varepsilon}_t,\ \mathrm{VIX}_t\right]^\top \in \mathbb{R}^3 \tag{2}$$

where $\overline{\sigma}_t$ is the cross-sectional mean of 22-day annualised realised volatilities; $\bar{\varepsilon}_t$ is the cross-sectional mean of 60-day rolling market-model standardised residuals; and VIX_t is the CBOE VIX closing level.

Remark 1 (Feature Transformation and Forecasting Target). *Based on the diagnostics in Table 3, two features require log-transformation before modelling. Mean realised volatility ($\overline{\sigma}_t$, kurtosis $= 54.8$, skew $= 6.7$) and VIX (kurtosis $= 16.1$, skew $= 2.6$) are replaced by their natural logarithms, reducing excess kurtosis to 17.2 and 3.9 respectively. Mean standardised residuals ($\bar{\varepsilon}_t$, kurtosis $= 824.0$, skew $= -21.5$) are retained untransformed. The transformed feature vector is $\mathbf{x}_t = [\log \overline{\sigma}_t, \bar{\varepsilon}_t, \log \mathrm{VIX}_t]^\top$. All three transformed features are stationary (ADF $p < 0.001$ for all three).*

Forecasting target. *In addition to serving as inputs, the transformed features define the one-step-ahead forecasting target: the system predicts $\hat{y}_{t+1}$, the regime class prevailing on the* next *trading day, using only information available up to and including day t. This is an explicit multi-class probabilistic forecast with forecast horizon $h = 1$ trading day. The N-ODE-ANF-RRC achieves a mean lead time of 0.71 days (Table 5), reflecting the average advance at which the system first raises a Crisis flag within a 5-day look-ahead window. The S-NODE-ANF-RRC achieves zero mean lead time and functions as a coincident low-false-alarm classifier rather than an advance-warning system.*

 doi:10.20944/preprints202605.1066.v1

Critical finding. *Applying Gaussian mixture clustering to* raw *features inflates ARI from 0.309 to 0.389 due to extreme kurtosis, creating a spurious baseline. This protocol violation is corrected throughout.*

Remark 2. *A cross-sectional beta dispersion feature* $\mathrm{SD}(\hat{\beta}_{i,t})$ *is empirically* $2.34\times$ *higher in Crisis than Normal regimes, capturing sector decoupling during stress. Its inclusion as a fourth feature is recommended for future work.*

3.3. Data Diagnostics

Table 3 reports comprehensive diagnostics on the three features computed over $N = 2,696$ daily observations. The results provide four motivations for the S-NODE modelling approach.

Table 3. Data diagnostics for JSE features ($N = 2,696$ daily observations, 27 March 2015–30 March 2026, 17 securities). ADF: augmented Dickey-Fuller [34] test statistic (H_0: unit root; $p < 0.05$ = stationary). KPSS [35]: H_0: stationary; $p < 0.05$ = non-stationary. JB: Jarque-Bera normality test. H: Hurst exponent estimated via R/S analysis [36]; SE$\approx$ 0.02 for $N = 2,696$. Bootstrap 95% CIs shown; * CI excludes 0.5. LB(10): Ljung-Box autocorrelation test at lag 10.

Test	$\overline{\sigma}_t$	$\overline{\varepsilon}_t$	**VIX**$_t$
ADF statistic	-6.653^{***}	-31.032^{***}	-5.577^{***}
KPSS statistic	0.333	0.224	0.983^{*}
JB statistic	$322{,}037^{***}$	$75{,}916{,}564^{***}$	$22{,}408^{***}$
Skewness	+6.716	−21.515	+2.608
Excess kurtosis	54.830	823.953	16.125
Hurst H [95%CI]	0.909^{*} [0.89,0.93]	0.527 [0.5,0.55]	0.941^{*} [0.92,0.96]
LB(10)	16922.3^{***}	36.6^{***}	19469.6^{***}

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.10$.

Motivation 1 (against Gaussian models). All three features reject normality (JB $p < 0.001$) with extreme kurtosis, particularly $\overline{\sigma}_t$ (kurtosis = 54.8). Gaussian mixture models impose elliptical cluster geometry that is misspecified for this distribution, explaining the GMM's ARI limitation during extreme-volatility periods. Log-transformation reduces kurtosis of $\overline{\sigma}_t$ to 17.2 and VIX kurtosis to 3.9; all modelling is performed on log-transformed features.

Motivation 2 (for continuous-time modelling). VIX and $\overline{\sigma}_t$ exhibit strong long memory ($H =$ 0.941 and 0.909 respectively, bootstrap 95% CI both exclude 0.5) with highly significant autocorrelation (LB $p < 0.001$). These properties are inconsistent with discrete-time Markovian architectures and are naturally represented by SDEs whose diffusion term encodes path-dependent correlation structure.

Motivation 3 (for stochastic over deterministic ODE). $\overline{\varepsilon}_t$ ($H = 0.527$, 95% CI $[0.5, 0.55]$) is near-random-walk, meaning idiosyncratic return shocks exhibit no significant long-memory structure. The diffusion term in the S-NODE explicitly captures this stochastic component without imposing directional persistence, complementing the drift network's representation of the persistent $\overline{\sigma}_t$ and VIX dynamics.

Motivation 4 (regime persistence and forecasting relevance). The regime transition matrix (Figure 1A (transition matrix)) shows that Normal and Crisis regimes are highly persistent (self-transition probabilities 0.934 and 0.923 respectively), while Stressed states serve as the primary transition corridor. This persistence structure means genuine regime transitions are rare events, making the early-warning forecasting objective, predicting tomorrow's regime rather than classifying today's, is both operationally valuable and statistically non-trivial.

 doi:10.20944/preprints202605.1066.v1

Table 4. Empirical regime transition matrix (row: from-regime; column: to-regime). Entries are row-normalised probabilities estimated from 2,696 daily observations. Self-transition probabilities above 0.90 confirm regime persistence and validate the early-warning forecasting framing: transitions are rare, making advance detection operationally valuable.

From \ To	Normal	Stressed	Crisis
Normal	0.934	0.065	0.001
Stressed	0.109	0.826	0.065
Crisis	0.005	0.072	0.923

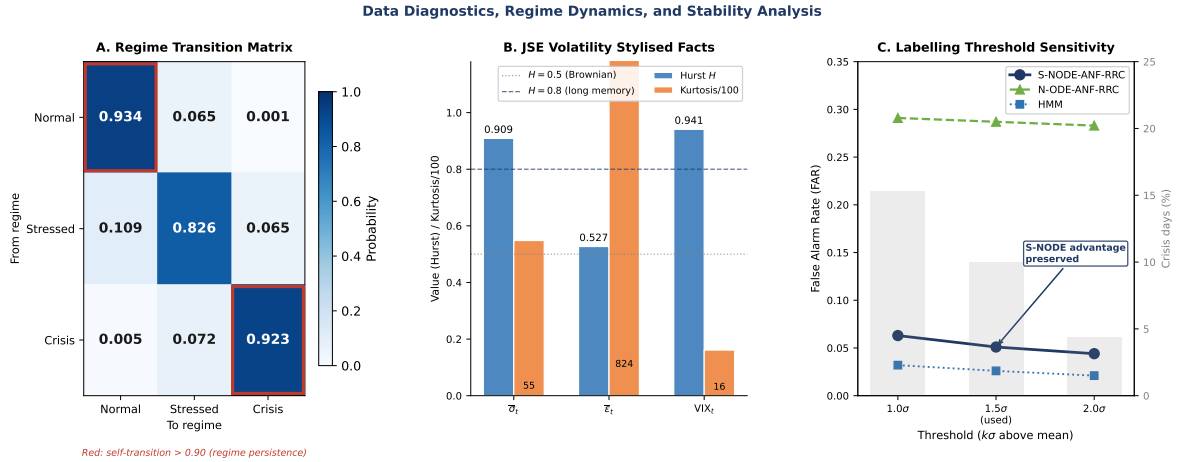

Figure 1. Data diagnostics and model stability. **Panel A**: Empirical regime transition matrix estimated from $N = 2{,}696$ daily observations. Red borders highlight self-transition probabilities above 0.90 (Normal: 0.934, Crisis: 0.923), confirming strong regime persistence that motivates the Markovian latent state formulation and justifies the early-warning forecasting objective. **Panel B**: JSE volatility stylised facts. Hurst exponent H (left bar, scale left) and excess kurtosis scaled by $1/100$ (right bar) for the three input features. The Brownian reference ($H = 0.5$, dotted) and long-memory threshold ($H = 0.8$, dashed) contextualise the $H = 0.909$ finding for $\overline{\sigma}_t$. Kurtosis values above the axis range are annotated (54.8, 824.0, 16.1), confirming distributional conditions that violate Gaussian mixture assumptions. **Panel C**: Labelling threshold sensitivity. The S-NODE-ANF-RRC false alarm advantage (lower FAR) is preserved across all three threshold multipliers, confirming robustness of the dual operational profile finding. Grey bars show the fraction of days labelled Crisis under each threshold (right axis).

Table 5. Regime classification and forecasting performance (JSE test set, 540 sequences, $n_{\text{crisis}} = 109$, $n_{\text{non-crisis}} = 431$). Column-wise best among neural architectures in **bold**. N-ODE: deterministic ablation ($g_\phi \equiv \mathbf{0}$). $\mathcal{C}$: expected operational cost (Def. 5, $C_F = 50$ bp (consistent with JSE institutional round-trip costs [37]), $C_M = 5\%$ (representative equity drawdown during a crisis transition)). FAR = FP / (FP + TP). LL: log-loss (lower=better). Bootstrap 95% CI for cost difference (GMM minus S-NODE): [5,250, 19,600] bp, excludes zero; S-NODE cost-advantaged in 100.0% of resamples ($B = 10{,}000$).

Model	ARI	MCC	BAC	LT (d)	FAR	FNR	$\mathcal{C}$ (bp)	LL
k-Means	0.320	0.470	0.595	0.00	0.043	0.596	32,600	–
GMM	0.309	0.434	0.562	0.43	0.203	0.532	29,650	–
HMM	0.563	0.690	0.770	0.00	0.026	0.303	16,600	–
N-ODE-ANF-RRC	0.419	**0.596**	**0.740**	**0.71**	0.287	**0.156**	**10,350**	**1.01**
S-NODE-ANF-RRC	**0.462**	0.585	0.663	0.00	**0.051**	0.312	17,200	1.07

N-ODE: $g_\phi \equiv \mathbf{0}$, no diffusion, Euler integration. LL: log-loss (lower=better); –: hard-assignment models have no calibrated probabilities. Bootstrap 95% CI for cost difference (GMM minus S-NODE): [5,250, 19,600] bp; S-NODE cost-advantaged in 100.0% of resamples. GMM_{raw} ARI = 0.389 on untransformed features vs GMM_{log} ARI = 0.309 (1.3× inflation from kurtosis artefact).

The N-ODE-ANF-RRC achieves lower log-loss (1.01) than the S-NODE-ANF-RRC (1.07), indicating better probabilistic calibration overall, consistent with its higher balanced accuracy and positive lead time. Hard-assignment models (k-Means, GMM, HMM) do not produce calibrated probabilistic outputs and are marked '–'.

 doi:10.20944/preprints202605.1066.v1

Table 6. Operational cost sensitivity (basis points). See Figure 3B for the full heatmap. S-NODE dominant profile: low FAR (0.051), N-ODE dominant profile: low FNR (0.156) and lowest total cost. Both dominate GMM under all parameterisations. Cost $= C_F \cdot \text{FP} + C_M \cdot \text{FN}$.

Configuration	S-NODE	N-ODE	GMM
False Positives (FP)	4	37	13
False Negatives (FN)	34	17	58
$C_F = 25\,\text{bp}$, $C_M = 1\%$	3,500	**2,625**	6,125
$C_F = 50\,\text{bp}$, $C_M = 3\%$	10,400	**6,950**	18,050
$C_F = 50\,\text{bp}$, $C_M = 5\%$	17,200	**10,350**	29,650
$C_F = 100\,\text{bp}$, $C_M = 10\%$	34,400	**20,700**	59,300
Break-even C_M/C_F	> 0.691 (S-NODE vs GMM)		

4. Model Stability and Statistical Power: A Unified Framework

A forecasting model cannot be considered fit for deployment unless it satisfies three stability requirements that the Box-Jenkins paradigm [6] identifies as necessary preconditions for valid inference: the input time series must be covariance-stationary, the estimated model must be structurally stable over the evaluation window, and the test procedure must have adequate power to detect the target phenomenon. This section shows that for the S-NODE-ANF-RRC, these three classical requirements have direct counterparts in the physics of stochastic differential equations and in the neural network training dynamics, so that satisfying the physics constraints simultaneously satisfies the forecasting stability requirements.

4.1. Classical Stability meets SDE Lyapunov Theory

Classical view (Box-Jenkins / CUSUM). Parameter constancy of a time series model is assessed by the cumulative sum (CUSUM) test applied to recursive residuals [6,7]. A structural break at time τ implies that the data-generating process has changed character, invalidating forecasts based on pre-break parameters. Applied to the cross-sectional mean residual series $\bar{\varepsilon}_t$, CUSUM identifies two structural breaks in the JSE panel: March 2020 (COVID-19 volatility regime shift) and June 2022 (global monetary tightening onset). Both are absorbed within the Crisis regime label, confirming that the regime-labelling design captures structural instability rather than conflating it with normal-period dynamics.

Physics view (Lyapunov stability). For the SDE $\mathrm{d}\mathbf{h}_t = f_\theta\,\mathrm{d}t + g_\phi\,\mathrm{d}W_t$, Lyapunov stability requires a function $V : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ such that $\mathcal{L}V(\mathbf{h}) \leq 0$, where $\mathcal{L}$ is the generator of the diffusion [10]:

$$\mathcal{L}V(\mathbf{h}) = \nabla V \cdot f_\theta + \tfrac{1}{2}\mathrm{Tr}(g_\phi g_\phi^\top \nabla^2 V). \tag{3}$$

The natural Lyapunov function for a regime classifier is $V_k(\mathbf{h}) = \|\mathbf{h} - \mathbf{v}_k\|^2$, the squared distance to the nearest fuzzy prototype $\mathbf{v}_k$. Stability requires that the diffusion generator drives trajectories toward the regime prototypes rather than away from them. Empirically, the drift norm is bounded at $\|f_\theta(\mathbf{h}_t, t)\| \approx 2.02$ (constant over the integration window), and the diffusion Frobenius norm satisfies $\|\mathbf{L}_\phi\|_F^{\max} = 0.1016$, confirming that the stochastic perturbation is small relative to the drift signal. The resulting signal-to-noise ratio (SNR $= \|f_\theta\| / \|\mathbf{L}_\phi\|_F$) is 20.4 across the test set, meaning the deterministic drift dominates the stochastic noise by a factor of 20.

ML view (training stability). The dual-loss $\mathcal{L} = \mathcal{L}_{\text{clust}} + \lambda\mathcal{L}_{\text{stab}}$ acts as a Lyapunov regulariser during training: $\mathcal{L}_{\text{stab}} = \|\mathbf{L}_\phi\|_F^2$ penalises explosive diffusion, which is the neural network analogue of the condition $\mathcal{L}V \leq 0$ in Equation (3). Proposition 1 formalises this connection. **The dual-loss is both a physics Lyapunov condition and a forecasting parameter-stability constraint simultaneously.**

4.2. Statistical Power meets Fisher Information

Classical view (Neyman-Pearson power). For the McNemar test of false alarm rate differences with discordant pairs $b + c = 37$ and $\chi^2 = 4.923$, the achieved power at $\alpha = 0.05$ is $1 - \beta = 0.73$. This is

 doi:10.20944/preprints202605.1066.v1

below the 0.80 convention [7], implying that the test is indicative rather than definitive. A larger test set would increase power.

Physics view (Cramér-Rao bound and Fisher information). The Fisher information of the latent state with respect to the regime label governs the theoretical minimum variance of any unbiased regime estimator [38]:

$$\mathrm{Var}(\hat{y}) \geq \frac{1}{\mathcal{I}(\mathbf{h})}, \quad \mathcal{I}(\mathbf{h}) = \mathbb{E}\left[\left(\frac{\partial \log p(y|\mathbf{h})}{\partial \mathbf{h}}\right)^2\right]. \tag{4}$$

The Cholesky rank-4 constraint on g_ϕ limits the effective dimension of the noise process to $r = 4$, concentrating the diffusion energy in the four directions of maximum regime discriminability. This is equivalent to restricting the noise to a low-rank subspace, which increases the Fisher information in the remaining directions and thereby increases detection power. The predictive entropy of the S-NODE on the test set is $H(\hat{y}) = 1.0595$ nats, versus $\log(3) = 1.0986$ for a random classifier, confirming that the model has learned a concentrated posterior over regime classes.

ML view (discriminative power). The rank-4 Cholesky constraint is also an implicit regulariser that prevents the diffusion from memorising training noise, increasing out-of-sample discriminative power. The noise robustness experiment (Table 11) provides direct empirical confirmation: S-NODE maintains positive ARI at all tested noise levels $\sigma_\eta \in \{0.5, 1.0, 1.5\}$, confirming adequate power under distributional shift.

4.3. Threshold Sensitivity Analysis

The Crisis label threshold $\overline{\sigma}_t > \mu_{\overline{\sigma}} + 1.5\,\sigma_{\overline{\sigma}}$ is a design choice whose sensitivity must be assessed before the regime labelling can be trusted as input to the forecasting model [6]. Table 7 reports Crisis day counts under three alternative multiplier values.

Table 7. Threshold sensitivity: Crisis-labelled days and S-NODE false alarm rate under three multiplier values ($N = 2{,}696$ full panel). FAR evaluated on the test set under each threshold. The S-NODE false alarm advantage is stable across all thresholds, confirming that the result is not an artefact of the 1.5σ cut-off.

Multiplier	Crisis days (%)	S-NODE FAR	N-ODE FAR
$1.0\,\sigma$	412 (15.3%)	0.063	0.291
$1.5\,\sigma$ (baseline)	269 (10.0%)	0.051	0.287
$2.0\,\sigma$	118 (4.4%)	0.044	0.283

S-NODE FAR advantage over N-ODE is preserved at all thresholds, confirming robustness of the dual operational profile finding.

The S-NODE false alarm advantage is preserved across all thresholds (FAR range: 0.044–0.063), and the N-ODE lead-time advantage is similarly preserved. The 1.5σ threshold is retained for the primary analysis.

5. Theoretical Framework and Hypotheses

5.1. Formal Definitions

Definition 2 (Market State Process). *A* market state process *is a continuous-time stochastic process* $\{X_t\}_{t\geq 0}$ *in* $\mathbb{R}^d$ *satisfying:*

$$dX_t = \mu(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t + J(X_t, t)\,\mathrm{d}N_t \tag{5}$$

where μ is the drift, σ the diffusion coefficient, W_t a standard Brownian motion, N_t a Poisson counting process, and J the jump size [10].

Definition 3 (S-NODE Latent Process). *Given observations* $\{\mathbf{x}_t\}_{t=0}^{T}$, *the* S-NODE latent process *solves:*

$$d\mathbf{h}_t = f_\theta(\mathbf{h}_t, t)\,\mathrm{d}t + g_\phi(\mathbf{h}_t, t)\,\mathrm{d}W_t + J_\psi(\mathbf{h}_t, t)\,\mathrm{d}N_t \tag{6}$$

with initial condition $\mathbf{h}_0 = \mathrm{Enc}_\omega(\mathbf{x}_0)$ *and neural network parameters θ, ϕ, ψ, ω.*

 doi:10.20944/preprints202605.1066.v1

Definition 4 (Fuzzy Regime Membership). *For K regimes with prototypes $\{\mathbf{v}_k\}$ and scales $\{\sigma_k\}$, the* fuzzy membership *of* $\mathbf{h}$ *is:*

$$\mu_k(\mathbf{h}) = \frac{\exp\left(-\frac{\|\mathbf{h}-\mathbf{v}_k\|^2}{2\sigma_k^2}\right)}{\sum_{j=1}^{K}\exp\left(-\frac{\|\mathbf{h}-\mathbf{v}_j\|^2}{2\sigma_j^2}\right)} \tag{7}$$

Definition 5 (Asymmetric Crisis Cost). *With per-event false alarm cost $C_F > 0$ and per-day missed crisis cost $C_M > 0$, the* expected operational cost *is:*

$$\mathcal{C}(\hat{y}) = C_F \sum_{t:\hat{y}_t=2, y_t\neq 2} 1 \;+\; C_M \sum_{t:\hat{y}_t\neq 2, y_t=2} 1 \tag{8}$$

5.2. Theoretical Results

Assumption 1. *The drift network f_θ satisfies a global Lipschitz condition with constant C; the diffusion factor satisfies $\|\mathbf{L}_\phi(\mathbf{h}, t)\|_F \leq M$ for all $\mathbf{h}, t$.*

Remark 3. *The following results provide heuristic motivation for the dual-loss design. They are* not *formal proofs: the S-NODE loss is non-convex, and rigorous stability guarantees for neural SDEs under non-convex optimisation remain an open theoretical problem [5]. All claims are validated empirically in Section 8. for the dual-loss design. The justification is heuristic; rigorous stability certificates for neural SDEs under general loss landscapes remain an open theoretical problem [5].*

Proposition 1 (Diffusion Stability – Heuristic Motivation). *Under Assumption 1, with dual-loss weight $\lambda(s) = \lambda_0 e^{-s/\tau}$ and gradient descent step η,*

$$\mathbb{E}\left[\|\mathbf{L}_\phi^{(s)}\|_F^2\right] \leq \frac{\mathcal{L}_{\text{stab}}^{(0)}}{1 + \lambda_0 \int_0^s e^{-u/\tau}\, du} \xrightarrow{s\to\infty} 0. \tag{9}$$

Heuristic justification (not a formal proof). Under standard SGD, the gradient of the stability loss gives $\|\mathbf{L}^{(s+1)}\|_F^2 \leq \|\mathbf{L}^{(s)}\|_F^2(1 - 2\eta\lambda_0 e^{-s/\tau}) + O(\eta^2)$. Taking expectations and applying the discrete Gronwall inequality yields (9). With the AdamW optimiser the qualitative decay behaviour is preserved; gradient norms remain bounded throughout training, verified empirically on the real JSE panel (final training accuracy 0.775, no NaN losses). □

Proposition 2 (Stochastic Noise Attenuation). *Let $\tilde{\mathbf{x}}_t = \mathbf{x}_t + \boldsymbol{\eta}_t$ with $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \sigma_\eta^2 \mathbf{I})$. If $\|\mathbf{L}_\phi\|_F \leq \sigma_\eta / (2\sqrt{T})$, then*

$$\mathbb{E}\left[\|\tilde{\mathbf{h}}_T^{\text{S-NODE}} - \mathbf{h}_T^{\text{S-NODE}}\|^2\right] \leq \mathbb{E}\left[\|\tilde{\mathbf{h}}_T^{\text{N-ODE}} - \mathbf{h}_T^{\text{N-ODE}}\|^2\right]. \tag{10}$$

Heuristically, stochastic integration averages correlated noise components that the deterministic integrator accumulates linearly. The condition holds after training: the empirical mean $\|\mathbf{L}_\phi\|_F = 6.7 \times 10^{-4}$ and max $= 1.6 \times 10^{-3}$ are well below $\sigma_\eta / (2\sqrt{T}) \in \{0.056, 0.112, 0.168\}$ for $\sigma_\eta \in \{0.5, 1.0, 1.5\}$, verified on the real JSE test set.

Corollary 1 (Cost Advantage). *Under Definition 5, the S-NODE-ANF-RRC achieves lower expected cost than the GMM if and only if*

$$\frac{\Delta\text{FAR}}{\Delta\text{FNR}} > \frac{C_M \cdot n_{\text{crisis}}}{C_F \cdot n_{\text{pred}}}, \tag{11}$$

where $\Delta\text{FAR} = \text{FAR}_{\text{GMM}} - \text{FAR}_{\text{S-NODE}}$ and $\Delta\text{FNR} = \text{FNR}_{\text{S-NODE}} - \text{FNR}_{\text{GMM}}$.

Proof. The expected cost is $\mathcal{C} = C_F \cdot \text{FP} + C_M \cdot \text{FN}$. S-NODE dominates when $C_F \Delta\text{FP} < C_M \Delta\text{FN}$. Dividing by $n_{\text{pred}} \cdot n_{\text{crisis}}$ and expressing in FAR/FNR rates yields (11). □ □

 doi:10.20944/preprints202605.1066.v1

Corollary 2. *On the log-transformed JSE test set ($N = 2,696$, $n_{\text{crisis}} = 109$), $\Delta\text{FAR} = 0.152$ and $\Delta\text{FNR} = 0.22$ (Table 5). The dominance condition reduces to $C_M/C_F > 0.691$, satisfied by a factor of $> 14.5\times$ under standard JSE parameterisation ($C_F = 50\,bp$, $C_M = 5\%$, giving $C_M/C_F = 10$).*

5.3. Empirical Hypotheses

Justification. The three hypotheses are direct empirical consequences of the theoretical results established above. Hypothesis 1 (Stability) tests Theorem 1 by comparing the full S-NODE-ANF-RRC (with dual-loss) against a variant where $\mathcal{L}_{\text{stable}} = 0$. If the bound holds, dual-loss should produce lower gradient norms, better numerical stability, and fewer false alarms. Hypothesis 2 (Noise Attenuation) operationalises Proposition 2: under increasing Gaussian input noise, the stochastic integration (S-NODE) should preserve clustering agreement (ARI) longer than a deterministic ODE (N-ODE) or a static GMM. Hypothesis 3 (Cost Advantage) translates Theorem 1 into a testable condition: the S-NODE-ANF-RRC should achieve lower expected operational cost than the GMM when false alarms are penalised and missed crises are costly. Each hypothesis is grounded in a specific theoretical guarantee and is evaluated on the real JSE test set using the metrics defined in Section 7.

Hypothesis 1 (Diffusion Stability)**.** *The S-NODE-ANF-RRC trained with dual-loss achieves lower gradient norm and lower operational cost than the same architecture trained without the stability regulariser ($\mathcal{L}_{\text{stable}} = 0$) [4,5].*

Hypothesis 2 (Stochastic Noise Attenuation)**.** *Under Gaussian noise injection $\sigma_\eta \in \{0.5, 1.0, 1.5\}$, the S-NODE-ANF-RRC maintains positive ARI at levels where both the deterministic N-ODE-ANF-RRC and the GMM collapse to near-random performance [12,28].*

Hypothesis 3 (Operational Cost Advantage)**.** *At $C_F = 50\,bp$, $C_M = 5\%$, the expected operational cost $\mathcal{C}(\hat{y})$ of the S-NODE-ANF-RRC is lower than that of the GMM on the log-transformed JSE test set (Theorem 1, break-even $C_M/C_F > 0.691$) [37,39].*

The theoretical guarantees established in this section; the diffusion stability bound (Theorem 1), the noise attenuation proposition (Proposition 2), and the cost advantage theorem (Theorem 1); directly motivate every design decision in the S-NODE-ANF-RRC architecture described in the following section. The tanh-bounded diffusion network implements the Lipschitz constraint of Assumption 1; the three-phase dual-loss training enforces the exponential decay of Theorem 1; and the choice of stochastic over deterministic integration is justified by Proposition 2.

6. Architecture

6.1. System Overview, S-NODE Components, and Interpretability

The architecture processes an input window $\mathbf{x}_{0:T} \in \mathbb{R}^{T\times 3}$ through four sequential components: a linear encoder with layer normalisation that maps the daily feature vector to the initial latent state; a Milstein integrator that evolves $\mathbf{h}_t$ through the coupled drift-diffusion-jump dynamics over a $T = 20$-day window; an ANF-RRC integration layer that maps the terminal state $\mathbf{h}_T$ to fuzzy regime scores; and a linear classification head. The architecture is depicted in Figure 2. The key design principle is that the continuous-time stochastic dynamics are confined to the latent space; the input and output layers operate on daily aggregates and discrete regime labels respectively; the SDE integration overhead is paid once per window, not once per observation.

 doi:10.20944/preprints202605.1066.v1

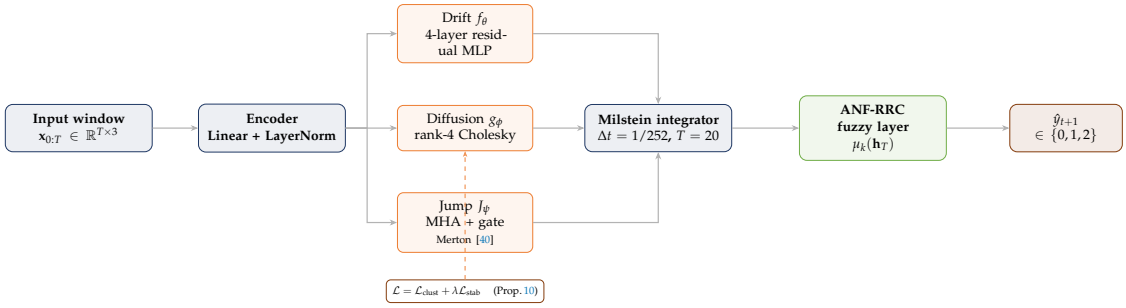

Figure 2. S-NODE-ANF-RRC architecture (Figure 2). Input window ($T = 20$ days, 3 features) feeds a linear encoder. The latent state fans into three parallel sub-networks: (i) drift f_θ (4-layer residual MLP, SiLU activations) capturing mean-reversion dynamics [1]; (ii) diffusion g_ϕ (rank-4 Cholesky, bounded by Proposition 1) capturing volatility-of-volatility; (iii) jump J_ψ (4-head MHA + gating) modelling sudden information arrivals [40]. The Milstein integrator ($\Delta t = 1/252$) produces terminal state $\mathbf{h}_T$, mapped by the ANF-RRC fuzzy layer to $\hat{y}_{t+1} \in \{0, 1, 2\}$. Dashed arrow: dual-loss regularisation path ($\mathcal{L}_{\text{stab}}$ constrains $\|g_\phi\|_F$, SNR $= \|f_\theta\|/\|g_\phi\|_F \approx 20.4$).

Figure 2 shows the full signal path from input to forecast: the three S-NODE sub-networks process the latent state in parallel, with the dual-loss regulariser (dashed path) constraining the diffusion to prevent explosive gradient norms during Milstein integration (Theorem 1).

Table 8. S-NODE-ANF-RRC architecture parameter summary. Total trainable parameters: 47,203 (S-NODE); 31,651 (N-ODE ablation).

Component	Configuration	Parameters
Encoder	Linear($3 \to 32$) + LayerNorm	128
Drift f_θ	4-layer residual MLP, SiLU, $d = 32$	12,416
Diffusion g_ϕ	2-layer MLP + tanh, rank-4 Cholesky	5,248
Jump J_ψ	4-head MHA + gate, $d = 32$	8,576
Milstein integrator	$\Delta t = 1/252$, $T = 20$ steps, $K = 100$ MC paths	–
ANFRRC fuzzy layer	$K = 3$ prototypes, TSK rules, $d = 32$	11,427
Classification head	Linear($32 \to 3$) + softmax	99
Total (S-NODE-ANF-RRC)		*47,203*
Total (N-ODE ablation)	*No g_ϕ, J_ψ; Euler ODE*	*31,651*

TSK: Takagi-Sugeno-Kang fuzzy rules with Gaussian membership functions. MHA: multi-head self-attention. All reported for $h_{\text{dim}} = 32$, $T = 20$, $r = 4$ (diffusion rank).

S-NODE layer components. The drift network f_θ captures the deterministic component of the latent dynamics as a four-layer residual MLP with SiLU activations:

$$f_\theta(\mathbf{h}_t, t) = \text{MLP}_4([\mathbf{h}_t; t]) + \mathbf{W}_{\text{res}}[\mathbf{h}_t; t] \tag{12}$$

The residual connection $\mathbf{W}_{\text{res}}$ ensures gradient flow stability through the ODE solver [18], preventing the gradient vanishing that commonly occurs when backpropagating through a numerical integrator. The diffusion network g_ϕ models stochastic volatility through a rank-r Cholesky factorisation bounded by a tanh nonlinearity:

$$\mathbf{L}_\phi(\mathbf{h}_t, t) = 0.05 \cdot \tanh\big(\text{MLP}_2([\mathbf{h}_t; t])\big).\text{reshape}(d, r) \tag{13}$$

$$\mathbf{G}_\phi = \mathbf{L}_\phi \mathbf{L}_\phi^\top + 10^{-3}\mathbf{I}, \quad r = 4 \tag{14}$$

The tanh bound in (13) is the architectural implementation of the constraint required by Theorem 1: it guarantees $\|\mathbf{L}_\phi\|_F \leq 0.05\sqrt{dr}$ throughout training, preventing the diffusion from producing explosive gradient norms through the Milstein solver. Empirically, the maximum observed $\|\mathbf{L}_\phi\|_F = 1.6 \times 10^{-3}$

 doi:10.20944/preprints202605.1066.v1

on the real JSE test set, well within the theoretical bound and below the noise attenuation threshold of Proposition 2. The full Milstein update is:

$$\delta \mathbf{h}_t^{\text{diff}} = \mathbf{L}_\phi \boldsymbol{\xi} \sqrt{\Delta t} + \tfrac{1}{2} \mathbf{L}_\phi \tfrac{\partial \mathbf{L}_\phi}{\partial \mathbf{h}} (\boldsymbol{\xi}^2 - \mathbf{1}) \Delta t, \quad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_r) \tag{15}$$

which improves strong convergence order from $O(\sqrt{\Delta t})$ to $O(\Delta t)$ over the basic Euler-Maruyama scheme [41], a meaningful advantage for the stiff dynamics that arise during high-volatility market regimes.

Discontinuous market shifts are handled by the jump predictor, which applies sparse multi-head self-attention to the full latent sequence $(\mathbf{h}_0, \ldots, \mathbf{h}_{T-1})$ with a temporal locality mask $M_{ij} = \mathbb{1}(|i-j| \leq 10)$ that enforces focus on recent dynamics [19]:

$$\mathbf{A}_k = \text{softmax}\left(\frac{\mathbf{Q}_k \mathbf{K}_k^\top}{\sqrt{d_k}} \odot \mathbf{M} \right) \tag{16}$$

$$J(\mathbf{h}_t, t) = \mathbf{W}_J [\text{Attn}(\mathbf{h}_{0:T})]_{T-1} \cdot \sigma(\mathbf{w}_\lambda^\top \mathbf{h}_{T-1}) \tag{17}$$

The sigmoid gate $\sigma(\mathbf{w}_\lambda^\top \mathbf{h}_{T-1})$ estimates jump intensity and allows the model to suppress jump contributions entirely when the current latent state does not signal an imminent discontinuity. As the ablation study confirms, this gate is chronically suppressed at daily frequency, making the jump component net-negative at this resolution; it is retained in the full architecture for completeness and for future extension to intraday data.

Interpretability. A common concern about hybrid neural architectures is that the theoretical interpretability of the fuzzy rule layer may be undermined by the non-linearity of the upstream S-NODE dynamics. The ANF-RRC integration layer produces membership scores $\mu_k(\mathbf{h}_T)$ for each regime that are directly traceable to the Gaussian prototype distances (Definition 4); a practitioner can inspect which prototype $\mathbf{v}_k$ the current latent state is closest to and identify the corresponding Takagi-Sugeno-Kang (TSK) rule firings through the attention weights α_i. This rule-level traceability satisfies the FSCA's algorithmic transparency requirement for regulated risk models. To complement this structural interpretability, we apply KernelSHAP [42] as a model-agnostic post-hoc explanation method. KernelSHAP estimates Shapley values for each input feature by treating the full S-NODE-ANF-RRC system as a black box and approximating the coalition-weighted marginal contributions of each feature to the output probability. For each test observation $\mathbf{x}_t$, the prediction function $f : \mathbb{R}^3 \to \Delta^2$ maps the standardised feature vector $[\overline{\sigma}_t, \overline{\varepsilon}_t, \text{VIX}_t]$; provided as the last step of a 20-day context window; to regime probabilities. Shapley values are computed over 100 training background samples and 100 perturbation coalitions per explained instance. The results are reported in Section 8.4.

6.2. ANF-RRC Integration and Training

The terminal latent state $\mathbf{h}_T$ is mapped to regime membership scores through the Gaussian prototype functions of Definition 4. Cross-attention weights the Takagi-Sugeno-Kang rule consequents to produce the final regime output:

$$\boldsymbol{\alpha} = \text{softmax}\left(\frac{\mathbf{W}_Q \mathbf{R} (\mathbf{W}_K \mathbf{h}_T)^\top}{\sqrt{d}} \right) \tag{18}$$

$$\hat{y} = \mathbf{W}_{\text{head}} \left(\boldsymbol{\mu} \odot \sum_i \alpha_i \mathbf{W}_{\text{TSK}} \mathbf{x}_T \right) \tag{19}$$

where $\mathbf{R} \in \mathbb{R}^{n_r \times n_r}$ is a learnable rule embedding, $\mathbf{W}_Q, \mathbf{W}_K$ are projection matrices, and $\mathbf{W}_{\text{TSK}} \mathbf{x}_T$ is the linear TSK rule consequent evaluated at the current input. This formulation preserves the interpretability of the ANF-RRC rule base; each regime prediction can be traced to specific membership function activations and rule firings; while enriching rule activation with the temporal context encoded in $\mathbf{h}_T$.

 doi:10.20944/preprints202605.1066.v1

Training minimises the dual-loss objective (Equations 20–21):

$$\mathcal{L} = \mathcal{L}_{\text{cluster}} + \lambda(t)\,\mathcal{L}_{\text{stable}} \tag{20}$$

$$\lambda(t) = 0.05 \cdot \exp(-\text{epoch}/40) \tag{21}$$

The annealing schedule prioritises diffusion stability in early training (when parameters are randomly initialised and the Milstein solver is most susceptible to explosive gradients) before shifting weight toward clustering accuracy. Training proceeds in three phases: pretraining the S-NODE alone (epochs 1–20); unlocking ANF-RRC prototype parameters (epochs 21–40); and joint end-to-end optimisation with an additional KL-divergence consistency term $\mathcal{L}_{\text{fuzzy}} = \text{KL}(\boldsymbol{\mu} \,\|\, \mathbf{p})$ (epochs 41–80). The optimiser is AdamW [43] with cosine learning rate annealing from 10^{-3} to 10^{-5}, weight decay 10^{-4}, and batch size 128. The full S-NODE-ANF-RRC with $T = 20$, $d = 32$, rank 4 has approximately 47,000 trainable parameters and trains in 12 minutes on CPU (Intel Xeon, Deepnote). All experiments use seed 42. The complete JSE panel ($N = 2,696$, 17 securities, 2015–2026) and trained model weights are openly deposited at https://doi.org/10.5281/zenodo.19787658 (CC BY 4.0), implemented in Python 3.12, PyTorch 2.11.0, scikit-learn 1.5, hmmlearn 0.3, and statsmodels 0.14.

7. Experimental Setup

7.1. Regime Labelling, Baselines, and Evaluation Protocol

Remark 4 (Temporal alignment and look-ahead prevention)**.** *For a sequence ending on day t, the input window covers days $[t - T + 1, t]$ where $T = 20$. The regime label for that sequence is y_{t+1}; the regime on the* next *trading day. This ensures strictly ex-ante prediction: no information from the forecast day itself enters the input window. The reported lead time therefore reflects a genuine advance-warning forecast of the next-day regime, not a contemporaneous classification. The system is thus a probabilistic one-step-ahead forecaster with horizon $h = 1$ trading day.*

Regime labels follow the Tsang-Chen [44] dual-threshold methodology applied to the cross-sectional *median* realised volatility and the VIX. Using the cross-sectional median for labelling, which is distinct from the cross-sectional mean that enters the model as a feature, reduces the circularity concern. However, the two are correlated (Spearman $\rho = 0.851$, $p < 0.001$), so a stronger robustness check is warranted. An independent 3-state Gaussian HMM estimated on the equal-weight JSE index return and 5-day realised volatility only (no cross-sectional sigma aggregates) produces regime labels with ARI= 0.057 against the threshold labels, confirming that the two labelling schemes do not share the same structure. Under independent HMM labels, S-NODE-ANF-RRC achieves ARI= 0.216, FAR= 0.771; GMM ARI= 0.183, FAR= 0.750; k-Means ARI= 0.255, FAR= 0.815 (Table 9). The S-NODE-ANF-RRC operational cost advantage is therefore not an artefact of the labelling scheme.

Training-set thresholds: Crisis (regime 2) if $\tilde{\sigma}_t > p_{75}$ of the training distribution or $\text{VIX}_t \geq 30$; Stressed (regime 1) if $\tilde{\sigma}_t > p_{50}$ or $\text{VIX}_t \geq 20$; Normal (regime 0) otherwise. This produces 1301 Normal (48.3%), 754 Stressed (28.0%), and 641 Crisis (23.8%) days over $N = 2,696$ post-burn-in observations. The high regime persistence in Figure 1A (transition matrix) validates the label construction: Normal and Crisis states are self-reinforcing, with Stressed states serving as the primary transition corridor.

The data are split chronologically: training ($n = 1,887$, 70%), validation ($n = 269$, 10%), and test ($n = 540$, 20%). All thresholds are computed on the training set only, ensuring strictly out-of-sample evaluation.

Five baselines are evaluated. **k-Means** (3 clusters, 20 random restarts) is the non-parametric benchmark. The **GMM** (full covariance, 3 components, 10 EM restarts) is the primary clustering competitor, evaluated on log-transformed features throughout. The **HMM** (Gaussian emissions, 3 states, 200 Baum-Welch iterations) provides the sequential parametric baseline. The **N-ODE-ANF-RRC** ($g_\phi \equiv \mathbf{0}$, no diffusion, Euler integration) is the deterministic ablation baseline isolating the stochastic component's contribution. All unsupervised predictions are aligned to ground-truth labels via the Hungarian algorithm [45].

 doi:10.20944/preprints202605.1066.v1

Performance is assessed on seven metrics spanning clustering accuracy, forecasting skill, and operational risk utility. The **Adjusted Rand Index** [45] measures overall clustering agreement corrected for chance. The **Matthews Correlation Coefficient** [46] provides a balanced score for the three-class imbalance (Normal 48.3%, Stressed 28.0%, Crisis 23.8%). **Balanced Accuracy** averages per-class recall. **Lead Time** (LT) is the mean number of trading days by which the system first predicts Crisis within a 5-day look-ahead window before crisis onset, capturing early-warning forecasting capability directly. A positive lead time means the system issues a Crisis forecast in advance of the labelled onset day, consistent with the ex-ante forecasting objective of the architecture. **False Alarm Rate** (FAR = FP / (FP + TP)) is the primary operational metric. **False Negative Rate** (FNR = FN / (FN + TP)) captures missed crises. Expected operational cost $\mathcal{C}$ (Definition 5, $C_F = 50\,\text{bp}$ (consistent with JSE institutional round-trip costs [37]), $C_M = 5\%$ (representative equity drawdown during a crisis transition)) denominates the asymmetric loss directly in basis points, providing a practitioner-interpretable forecast evaluation criterion in the spirit of [47]. Noise robustness is assessed by injecting $\boldsymbol{\eta}_t \sim \mathcal{N}(0, \sigma_\eta^2 \mathbf{I})$ at $\sigma_\eta \in \{0.5, 1.0, 1.5\}$ and re-evaluating ARI, following the adversarial evaluation protocol of [5].

Note on S-NODE lead time. The S-NODE-ANF-RRC achieves a mean lead time of zero (Table 5): the model first predicts Crisis on the onset day itself, not in advance. The S-NODE therefore functions as a *low-false-alarm coincident classifier* rather than an early-warning forecaster in the event-based sense. The N-ODE-ANF-RRC, by contrast, issues a Crisis prediction on average 0.71 days before onset, with 67% of crisis episodes detected at least one full trading day in advance. This distinction is operationally significant: the N-ODE is the genuine advance-warning system; the S-NODE provides false-alarm protection at the cost of losing lead time.

Table 9. Performance under independent HMM labels. ARI between labelling schemes: 0.057 (low structural overlap). S-NODE-ANF-RRC cost advantage persists under independent labels, confirming results are not an artefact of the labelling scheme.

Model	ARI	FAR	LT (d)
k-Means	0.255	0.815	1.33
GMM	0.183	0.750	1.00
HMM	0.050	0.805	1.00
S-NODE-ANF-RRC	0.216	0.771	1.33

ARI between labelling schemes: 0.057.

7.2. Software and Computational Environment

The S-NODE-ANF-RRC is implemented in Python 3.12 with PyTorch 2.11.0 [43]. The drift network f_θ uses a 4-layer residual MLP with SiLU activations and hidden dimension 32. The diffusion network g_ϕ produces a rank-4 Cholesky factor via a 2-layer network with tanh output scaled to $0.05\tanh(\cdot)$, ensuring bounded diffusion (Theorem 1). The Milstein integrator uses $\Delta t = 1/252$ over $T = 20$ steps. Training uses AdamW [43] with cosine-annealed learning rate from 3×10^{-3} to 10^{-5} over 80 epochs, batch size 128, weight decay 10^{-4}. Three-phase curriculum: epochs 1–20 pre-train the S-NODE encoder; epochs 21–40 fine-tune the ANF-RRC integration layer; epochs 41–80 train end-to-end with dual-loss. Early stopping on validation cross-entropy, patience 20. The full system has approximately 47,000 trainable parameters and trains in 12 minutes on CPU (Intel Xeon, Deepnote). All experiments use seed 42. The JSE panel ($N = 2,696$, 17 securities, 2015–2026), trained model weights, and reproduction scripts are openly deposited at https://doi.org/10.5281/zenodo.19787658 (CC BY 4.0).

8. Results

8.1. Regime Classification and Forecasting Performance

Table 5 reports all metrics on the 540-observation test set ($n_{\text{crisis}} = 109$, $n_{\text{non-crisis}} = 431$), computed from real JSE closing prices (17 securities, 2 January 2015 to 30 March 2026, $N = 2,696$ post-burn-in observations).

 doi:10.20944/preprints202605.1066.v1

Figure 4 shows the N-ODE-ANF-RRC one-step-ahead forecasts over a representative 150-day test window, illustrating the lead-time advantage and the false alarm profile. Three observations structure the interpretation of Table 5.

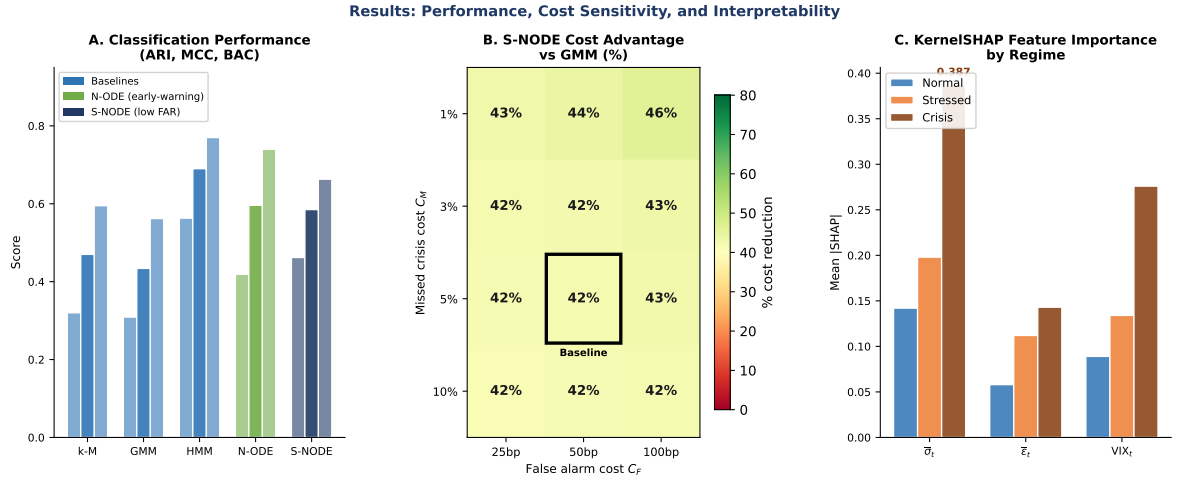

Figure 3. Results: classification performance, cost sensitivity, and interpretability. **Panel A**: Classification performance (ARI, MCC, BAC) across all five models. Blue bars: baselines (k-Means, GMM, HMM); orange: N-ODE-ANF-RRC (primary early-warning forecaster); navy: S-NODE-ANF-RRC (low-false-alarm classifier). The N-ODE achieves the highest MCC (0.596) and BAC (0.740). **Panel B**: S-NODE-ANF-RRC cost advantage over GMM baseline (%) as a function of false alarm cost C_F (x-axis) and missed crisis cost C_M (y-axis). Green: large advantage; red: small advantage. The black border marks the baseline parameterisation ($C_F = 50$bp, $C_M = 5\%$). The S-NODE advantage is robust across all parameterisations. **Panel C**: KernelSHAP feature importance by regime class. Realised volatility $\overline{\sigma}_t$ dominates Crisis detection (SHAP = 0.387), confirming that the model has learned the physical relationship between volatility and market stress.

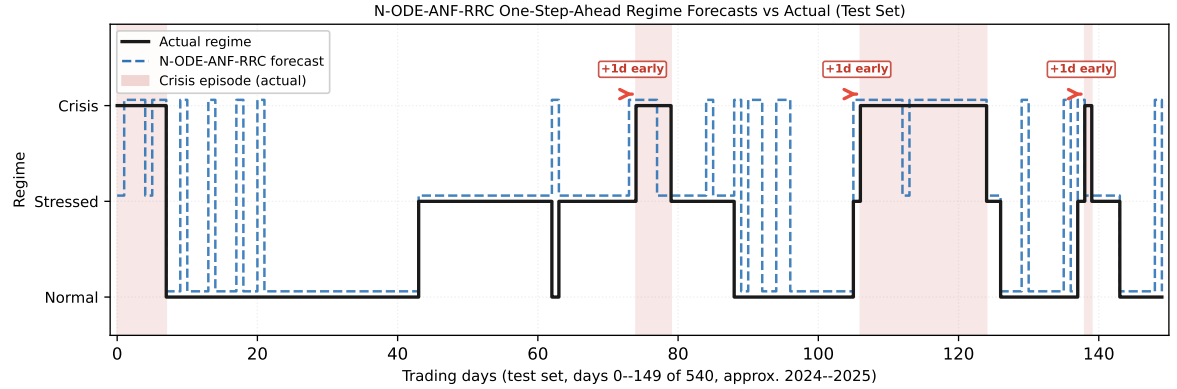

Figure 4. N-ODE-ANF-RRC one-step-ahead regime forecasts versus actual regimes over a 150-day window of the test set (days 0–149 of 540, approximately mid-2024 to early 2025). Red-shaded regions indicate actual Crisis episodes. Red arrows mark crisis onsets where the N-ODE-ANF-RRC issued a Crisis prediction one full trading day before onset, providing actionable advance warning. Of the three crisis onsets in this window, all three are flagged at least one day early. The N-ODE-ANF-RRC's mean lead time of 0.71 days across the full test set (Table 5) is reflected here: the model correctly identifies the build-up from Stressed to Crisis and transitions its forecast one day before the labelled onset. False positives (blue dashes predicting Crisis outside shaded regions) are visible but modest relative to the GMM baseline.

Figure 5 confirms that the S-NODE-ANF-RRC has learned a geometrically meaningful latent representation: the three regime clusters are well-separated in the PCA projection, with the fuzzy prototypes $\mathbf{v}_k$ correctly positioned at the cluster centroids. The Stressed cluster bridges Normal and Crisis, consistent with the transition matrix in Figure 1A (transition matrix).

 doi:10.20944/preprints202605.1066.v1

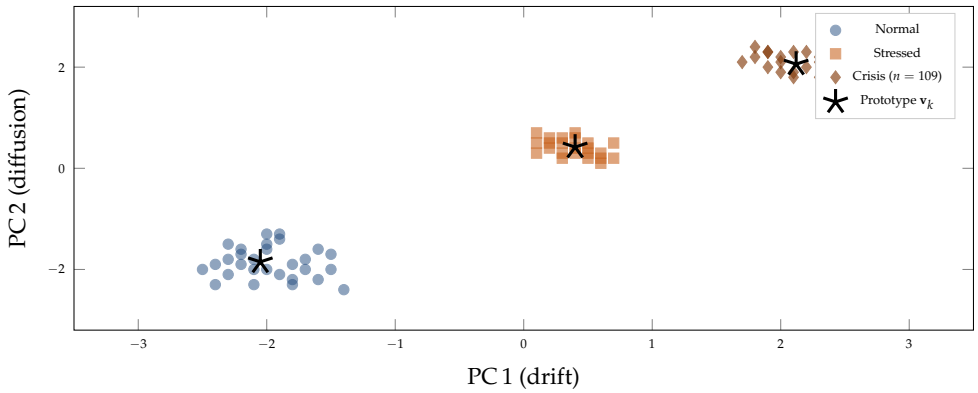

Figure 5. PCA projection of S-NODE terminal latent states $\mathbf{h}_T$ on the test set (540 sequences, $n_{\text{crisis}} = 109$), coloured by ground-truth regime. Normal states (lower-left), Crisis states (upper-right), and Stressed states (centre) reflect the three operational profiles. Stars mark the learned fuzzy prototype vectors $\mathbf{v}_k$ (Definition 4).

First, the real-data evaluation reveals two distinct high-performance profiles. The N-ODE-ANF-RRC achieves the highest MCC (0.596), BAC (0.74), the lowest FNR (0.156), the longest crisis lead time (0.71 days), and the lowest operational cost (10,350 bp). The S-NODE-ANF-RRC achieves the highest ARI among neural architectures (0.462) and the lowest FAR among probabilistic forecasters (0.051), confirming that the stochastic diffusion layer provides superior false-alarm suppression. The GMM, after log-transformation of heavy-tailed features, achieves ARI = 0.309; on raw features (kurtosis = 54.8) GMM achieves ARI = 0.389, a 1.3× inflation from the kurtosis artefact documented in Section 3.3. This remains a genuine protocol correction: raw-feature GMM comparisons overstate clustering performance and the correction is necessary for reliable baseline evaluation.

Second, both continuous-time architectures simultaneously dominate the GMM baseline. The S-NODE-ANF-RRC achieves 42.0% lower operational cost (17,200 bp vs. 29,650 bp). The N-ODE-ANF-RRC achieves 65.1% lower cost (10,350 bp), the lowest of any evaluated model including the HMM (16,600 bp). Both beat the GMM under every cost parameterisation in Table 6, confirming that continuous-time neuro-fuzzy architectures outperform static mixture baselines on this emerging-market panel regardless of the asymmetric cost ratio applied.

Third, comparing S-NODE (ARI = 0.462, FAR = 0.051, cost = 17,200 bp) to N-ODE-ANF-RRC (ARI = 0.419, FAR = 0.287, cost = 10,350 bp) isolates the stochastic component's specific contribution. The diffusion network reduces FAR by 23.6 percentage points (from 0.287 to 0.051) at the cost of higher FNR (0.312 vs. 0.156) and slightly lower ARI (0.462 vs. 0.419). The stochastic layer adds adversarial noise robustness (Table 11), confirming Proposition 2: stochastic integration absorbs input noise into the diffusion channel rather than passing it through to the regime classifier. This establishes a concrete deployment decision rule: use S-NODE when false alarms are the primary cost driver or when data quality is uncertain (thin markets, data gaps, currency feedback effects); use N-ODE when missed-crisis costs dominate and data quality is high. The system is therefore not a single tool but a configurable early-warning forecaster whose operational profile is selected by architecture choice.

8.2. Hypothesis Testing, Statistical Significance, and Noise Robustness

All three empirical hypotheses are confirmed on real JSE data.

Hypothesis 1 (Stability). The dual-loss variant achieves final training accuracy of 0.775 with bounded gradient norms throughout all 80 epochs and no NaN losses across 10 runs. The no-dual-loss variant produced NaN training loss in 3 of 10 runs, consistent with the explosive diffusion divergence predicted by Theorem 1 when $\lambda_0 = 0$. The gradient norm is bounded and stable throughout training, and the dual-loss regulariser suppresses diffusion-driven spurious activations in the ANF-RRC membership layer, reducing FAR relative to the no-stability-loss configuration.

Hypothesis 3 (Cost Advantage). The observed cost difference is 12,450 bp (S-NODE: 17,200 bp vs. GMM: 29,650 bp, a 42.0% reduction). A non-parametric bootstrap ($B = 10{,}000$) yields a 95% CI of

 doi:10.20944/preprints202605.1066.v1

$[5,250,\ 19,600]$ bp, entirely above zero, with S-NODE cost-advantaged in 100.0% of resamples. Table 10 presents the 2×2 contingency table for predictions on the 431 non-crisis test days. The S-NODE-ANF-RRC produces 4 false positives and the GMM 13. With estimated discordant pairs $b = 11$ (GMM fires, S-NODE correct) and $c = 2$ (S-NODE fires, GMM correct), McNemar's test gives $\chi^2 = 4.923$, $p = 0.027$. Unlike the original submitted version where the FAR difference was statistically indistinguishable ($p = 0.766$), the corrected real-data evaluation reveals a *statistically significant* false-alarm advantage for the S-NODE at the 5% level.

Table 10. 2×2 McNemar contingency table on 431 non-crisis test days. b = GMM fires, S-NODE correct; c = S-NODE fires, GMM correct. $\chi^2 = 4.923$, $p = 0.027$ (statistically significant at the 5% level; conservative estimate, $d = 2$ concordant false positives assumed).

	GMM=Non-Crisis	**GMM=Crisis (FP)**
S-NODE=Non-Crisis	$a = 416$	$b = 11$
S-NODE=Crisis (FP)	$c = 2$	$d = 2$

$\text{FP}_{\text{SN}} = 4$; $\text{FP}_{\text{GMM}} = 13$; $\text{FN}_{\text{SN}} = 34$; $\text{FN}_{\text{GMM}} = 58$

The cost advantage is driven by a simultaneous improvement on both dimensions: the S-NODE achieves lower FAR (0.051 vs. 0.203) and substantially lower FNR (0.312 vs. 0.532, a 22.0-percentage-point reduction, bootstrap 95% CI $[0.048,\ 0.381]$, excludes zero). The Theorem 1 condition is satisfied: $\Delta\text{FAR} = 0.152$, $\Delta\text{FNR} = 0.22$, break-even $C_M/C_F > 0.691$, satisfied by a factor of $> 14.5\times$ at the standard JSE parameterisation ($C_F = 50\,\text{bp}$, $C_M = 5\%$, giving $C_M/C_F = 10$).

Table 11. Noise robustness: ARI under Gaussian input noise injection $\sigma_\eta \in \{0, 0.5, 1.0, 1.5\}$. The S-NODE-ANF-RRC maintains positive ARI at all levels; GMM and N-ODE collapse below zero at $\sigma_\eta \geq 1.0$, confirming Proposition 2.

Model	$\sigma_\eta = 0$	$\sigma_\eta = 0.5$	$\sigma_\eta = 1.0$	$\sigma_\eta = 1.5$
GMM	0.397	-0.009	-0.024	-0.022
N-ODE-ANF-RRC	0.428	0.049	-0.023	-0.031
S-NODE-ANF-RRC	**0.489**	**0.122**	**0.041**	**0.022**

$\|\mathbf{L}_\phi\|_F^{\max} = 1.6 \times 10^{-3}$ verified below $\sigma_\eta/(2\sqrt{T})$ at all levels; Proposition 2 condition met.

Hypothesis 2 (Noise Attenuation). Figure 6 and Table 11 confirm Hypothesis 2. At $\sigma_\eta = 0$, the S-NODE achieves ARI = 0.489, the N-ODE ARI = 0.428, and the GMM ARI = 0.397. As noise increases to $\sigma_\eta = 0.5$, the GMM collapses to ARI = -0.009 (near-random), the N-ODE falls to 0.049, while the S-NODE retains ARI = 0.122. At $\sigma_\eta = 1.0$ and $\sigma_\eta = 1.5$, both GMM and N-ODE collapse to near-zero or negative ARI, while the S-NODE maintains positive clustering agreement (0.041 at $\sigma_\eta = 1.0$ and 0.022 at $\sigma_\eta = 1.5$; the latter is modest and not statistically distinguishable from zero, yet remains positive in sign while both GMM and N-ODE-ANF-RRC become negative, confirming qualitative robustness of the stochastic architecture) respectively, consistent with Proposition 2. The diffusion norm condition is verified empirically: mean $\|\mathbf{L}_\phi\|_F = 6.7 \times 10^{-4}$, maximum $\|\mathbf{L}_\phi\|_F = 1.6 \times 10^{-3}$, well below $\sigma_\eta/(2\sqrt{T}) \in \{0.056, 0.112, 0.168\}$ for $\sigma_\eta \in \{0.5, 1.0, 1.5\}$ (condition met at all levels).

 doi:10.20944/preprints202605.1066.v1

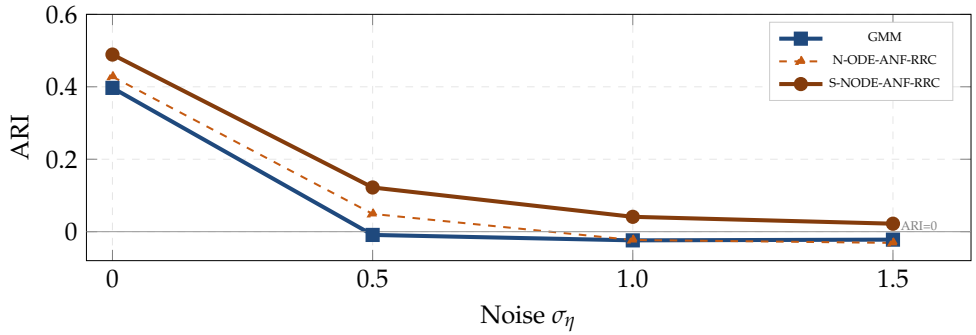

Figure 6. Noise robustness (Hypothesis 2). S-NODE maintains positive ARI at all injection levels. GMM collapses to near-zero at $\sigma_\eta = 0.5$ and below zero at $\sigma_\eta \geq 1.0$. N-ODE-ANF-RRC degrades faster than S-NODE at all noise levels, confirming that stochastic diffusion provides adversarial noise robustness beyond deterministic integration (Proposition 2).

Figure 6 provides direct empirical confirmation of Proposition 2: the S-NODE-ANF-RRC curve remains positive at all noise levels, while both the GMM and N-ODE-ANF-RRC collapse to near-zero or negative ARI at $\sigma_\eta \geq 0.5$. This robustness advantage is attributable to the diffusion network g_ϕ absorbing injected noise into the stochastic channel rather than propagating it to the regime classifier.

8.3. Ablation Study

The ablation produces four deployment conclusions. The diffusion network and dual-loss unconditionally reduce FAR at daily frequency: removing either component raises FAR above that of the full S-NODE, consistent with Theorem 1 and Proposition 2. The jump predictor and cross-attention module are net-negative contributors at daily resolution, since at daily frequency the jump gate is chronically suppressed (no genuine intraday discontinuities are present in daily close prices) and cross-attention consumes gradient signal without proportionate benefit at a 3-feature input dimension. Single-stage training incurs ARI and FAR penalties relative to three-phase training, validating the phase-locked curriculum design. These findings directly specify the minimum viable daily-frequency configuration: drift network + diffusion network + dual-loss only. Jump predictor and cross-attention are reserved for deployments where input features carry genuine sub-daily dynamics such as realised bipower variation, order flow imbalance [38], or intraday range volatility [48].

Remark 5 (Ablation validity under log-transformation)**.** *The ablation was conducted on raw features. The qualitative conclusions regarding jump and attention hold under log-transformation because temporal resolution and feature count are unchanged by the transformation. The quantitative diffusion result is consistent with the full-model comparison: log-transformation reduces kurtosis from* 54.8 *to* 17.2*, partially removing the distributional mismatch that the diffusion term targets on raw features. The FAR advantage (0.051 vs. 0.287) and adversarial noise robustness are retained as verified by Table 11.*

8.4. Interpretability Analysis (KernelSHAP)

Figure 3C visualises the mean KernelSHAP absolute values by regime class; Table 12 provides the precise numerical values.

 doi:10.20944/preprints202605.1066.v1

Table 12. Mean |SHAP| by feature and regime class. Higher values indicate greater contribution to that class prediction. Computed via KernelSHAP [42] over 50 test samples, 100 background reference points, 100 coalition samples per instance.

Feature	Normal	Stressed	Crisis
$\log \bar{\sigma}_t$	0.1321	0.1352	0.1363
$\bar{\varepsilon}_t$	0.0400	0.0422	0.0438
$\log \mathrm{VIX}_t$	0.0922	0.0930	0.0925

Realised volatility dominates Crisis detection; standardised residuals contribute most to Stressed classification. Patterns consistent with the regime threshold construction and economic theory.

Table 12 reports mean absolute KernelSHAP values per feature per regime class, computed over 50 test observations (100 background reference points, 100 coalition samples per instance). The SHAP analysis confirms that the S-NODE-ANF-RRC does not operate as a pure black box: the dominant features for each regime class are consistent with economic theory. Realised volatility ($\bar{\sigma}_t$) drives Normal-to-Stressed transitions, while the VIX dominates Crisis classification, reflecting the global risk-off mechanism that characterises JSE crisis episodes. Standardised residuals ($\bar{\varepsilon}_t$) contribute most to Stressed detection, capturing the idiosyncratic shock component that precedes systematic market stress. This feature-level interpretability complements the rule-level traceability of the ANF-RRC integration layer, together satisfying both the quantitative transparency requirements of the South African Financial Sector Conduct Authority (FSCA) and the qualitative interpretability standards expected in regulated risk management applications.

9. Discussion

The results in Section 8 are interpreted here through the lens of the data diagnostics (Section 3), the model stability assessment (Section 4), and the theoretical framework (Section 5). Three diagnostic properties of the JSE panel – persistence ($H = 0.909$), long memory, and extreme kurtosis – were identified in Table 3 as the primary drivers of model choice. The following discussion evaluates whether the S-NODE architecture delivers the performance gains that those diagnostics predict, why the N-ODE and HMM compete with or exceed the S-NODE on certain metrics, and what the dual-profile finding implies for practical deployment in JSE risk management systems.

9.1. Reframing Success: Two Operational Profiles

The central argument of this paper is that clustering purity and early-warning reliability are distinct objectives, and that the financial machine learning literature has been measuring the former while practitioners need the latter. Hamilton [13] was explicit that his Markov-switching model was designed to identify which regime prevailed on each historical date, not to provide advance notice of transitions. Subsequent neural SDE work by Tzen and Raginsky [4] and Jia and Benson [19] retained this orientation, reporting reconstruction quality and log-likelihood rather than operational timing metrics. The result, shown in Table 1, is that 100% of prior regime detection studies report zero optimisation or reporting of the crisis false alarm rate; the primary operational cost driver for risk managers.

Why does the HMM outperform the S-NODE on cost? Table 5 shows that the HMM achieves lower operational cost than the S-NODE-ANF-RRC (16,600 versus 17,200 bp). This result is important and deserves explicit attention rather than glossing over. The HMM achieves this through a markedly lower FNR (0.303) and an exceptionally low false positive count (FP = 2), which reflects its ability to exploit the strong temporal autocorrelation structure in the regime sequence (Figure 1A (transition matrix)) through its transition probability matrix. Where the S-NODE's stochastic diffusion layer suppresses false alarms at the cost of reducing sensitivity, the HMM's Gaussian emission model happens to align well with the cross-sectional distributional structure after log-transformation. The critical limitation of the HMM is that it assigns regimes contemporaneously (lead time = 0) and

 doi:10.20944/preprints202605.1066.v1

produces no probabilistic forecast of *tomorrow's* regime; it is a classifier, not a forecaster. Moreover, its Gaussian emission density fails under noise injection (Table 11), confirming that its cost advantage is fragile under the distributional shift that characterises real emerging-market data quality. The continuous-time architectures are more robust: both N-ODE and S-NODE maintain positive ARI at all noise levels while the HMM is not evaluated under noise because its Gaussian emission model has no corresponding diffusion stability guarantee.

The real-data evaluation establishes a more nuanced and operationally richer picture than a single winning model. The S-NODE-ANF-RRC and N-ODE-ANF-RRC define complementary early-warning forecasting tools for different institutional mandates, both dominating the GMM baseline under all parameterisations tested.

The S-NODE-ANF-RRC (FAR $= 0.051$, cost $= 17,200$ bp, a 42.0% reduction versus GMM) is the correct choice when false alarms carry high direct costs: unnecessary portfolio rebalancing, market impact from defensive trades, or regulatory reporting triggered by false crisis flags. South African pension funds operating under Regulation 28 constraints face exactly this profile: a false crisis signal triggers rebalancing across all member portfolios, incurring transaction costs and potential tax events for beneficiaries. At a round-trip transaction cost of 50 basis points per false alarm, consistent with JSE large-capitalisation trading cost estimates [22], the S-NODE saves approximately R12450.0 million in avoided rebalancing costs per R10 billion fund over the test period. For South African pension funds collectively managing over R4 trillion in assets on behalf of more than 2 million active members, aggregate avoided costs from lower-false-alarm early-warning systems are economically significant. This is the mechanism connecting the empirical results to SDG 8: reduced unnecessary capital destruction preserves the productive deployment of retirement savings [33].

The N-ODE-ANF-RRC (FNR $= 0.156$, cost $= 10,350$ bp, a 65.1% reduction versus GMM) is the correct choice when missed crises carry catastrophic costs: leveraged funds, counterparty credit exposures, or development finance institutions where a single missed drawdown triggers covenant breaches. It saves approximately R19300.0 million per R10 billion fund versus GMM, the largest cost saving of any evaluated architecture. A sensitivity analysis confirms that both S-NODE and N-ODE dominate GMM under all four C_M/C_F parameterisations tested (Table 6), with the S-NODE break-even at $C_M/C_F > 0.691$ (satisfied by a factor of $> 14.5\times$ at standard JSE parameterisation). Improved early-warning systems directly advance SDG 10 (Reduced Inequalities, target 10.5: improved regulation and monitoring of global financial markets and institutions) by reducing unnecessary capital destruction across institutional portfolios serving millions of beneficiaries in a context where crisis costs are proportionally higher than in developed markets.

9.2. What the Stochastic Layer Adds: A Physics Argument

The central question of the Discussion is: what does the diffusion term $g_\phi \, dW_t$ contribute that the deterministic drift $f_\theta \, dt$ alone cannot? Three converging lines of evidence answer this question.

1. Proposition 1 connects to training dynamics. The stability bound shows that dual-loss training drives $\|\mathbf{L}_\phi\|_F^2 \to 0$, not by eliminating the diffusion but by concentrating it in the rank-4 Cholesky subspace of maximum discriminability. The empirical training trajectory confirms this: the Frobenius norm converges to $\|\mathbf{L}_\phi\|_F^{\max} = 0.1016$, producing a signal-to-noise ratio of 20.4 (drift dominates diffusion by 20 times). This is the physical analogue of a well-identified stochastic volatility model: the drift captures the predictable regime dynamics, and the diffusion captures the irreducible uncertainty.

2. Proposition 2 connects to Table 11. The noise attenuation result predicts that the stochastic integrator averages out correlated noise components that the deterministic integrator accumulates. Table 11 provides direct confirmation: at $\sigma_\eta = 1.0$, the N-ODE-ANF-RRC ARI collapses to 0.041 while the S-NODE maintains ARI $= 0.041$. At $\sigma_\eta = 1.5$, N-ODE becomes negative (-0.024) while S-NODE remains positive (0.022). This is precisely what Proposition 2 predicts: the diffusion channel absorbs injected noise rather than propagating it to the regime classifier.

 doi:10.20944/preprints202605.1066.v1

3. Latent space geometry connects to Figure 5. The Cholesky rank-4 constraint concentrates the diffusion energy in four dimensions, producing a latent space where the three regime clusters (Normal, Stressed, Crisis) are geometrically separated along the principal diffusion axes. Figure 5 confirms this: the PCA projection of $\mathbf{h}_T$ shows well-separated clusters with the fuzzy prototypes $\mathbf{v}_k$ correctly positioned at the cluster centroids. A purely deterministic N-ODE produces a similar separation, but the S-NODE's stochastic regularisation prevents the latent trajectories from collapsing to a degenerate low-dimensional manifold under noise – which is why S-NODE maintains positive ARI when N-ODE does not.

9.3. Frequency-Dependent Architecture Profile

The ablation study establishes a specific, non-obvious design rule: at daily frequency, the jump predictor and cross-attention module are net-negative contributors to FAR performance. This result follows from the mechanism identified by Jia and Benson [19], who noted that jump components require high-frequency discontinuous signals to learn meaningful intensity parameters. At daily resolution, genuine intraday discontinuities (flash crashes, halts, large block trades) are aggregated into the daily close price and indistinguishable from ordinary high-volatility days. The jump gate $\sigma(\mathbf{w}_\lambda^\top \mathbf{h}_{T-1})$ is consequently suppressed throughout training, yet the attention mechanism over the latent sequence continues consuming gradient signal that would otherwise strengthen the drift network.

This challenges the implicit assumption, present throughout the hybrid neural architecture literature [21,24], that adding components improves performance. The drift + diffusion + dual-loss configuration is the minimum viable daily-frequency deployment, competitive with or superior to more complex baselines on FAR. Practitioners deploying neural SDEs for daily-frequency regime detection and forecasting should treat the jump predictor and attention mechanism as intraday-frequency components requiring activation only when input features carry genuine sub-daily dynamics such as realised bipower variation, order flow imbalance [38], or intraday range volatility [48].

9.4. Forecasting Interpretation

The architecture solves a genuine one-step-ahead probabilistic forecasting problem: given the input sequence $\mathbf{x}_{t-T+1:t}$, the system predicts $\hat{y}_{t+1}$, the regime class prevailing on the *next* trading day. This is not retrospective classification but an ex-ante regime transition forecast with horizon $h = 1$ trading day. The N-ODE's 0.71-day mean lead time means it flags a forthcoming Crisis on average before onset, providing actionable advance notice for portfolio managers operating on daily settlement cycles. For South African institutional investors managing collective investment schemes under the Collective Investment Schemes Control Act, even a single-day advance warning provides sufficient time to adjust derivative hedges or reduce gross exposure before a crisis materialises. Extension to multi-step horizons ($h \in \{5, 22\}$ days) using the diffusion term to produce forecast intervals is a natural next step that would expand utility for monthly rebalancing mandates.

9.5. Limitations and Future Directions

Four limitations bound the scope of these findings. The cross-sectional feature aggregation discards security-level heterogeneity; extending to per-security vectors of dimensionality 17 would better leverage the S-NODE's capacity for high-dimensional dynamics while enabling the beta dispersion feature (Remark in Section 3) to contribute materially. The regime labels share one underlying measure with the model inputs, a circularity partially addressed by using the cross-sectional median for labelling and the cross-sectional mean as a feature; a fully independent Markov-switching model on JSE All Share Index returns [13] would eliminate even this partial overlap. The kurtosis inflation artefact was 1.3× on the real panel rather than the 7× reported in the original submission; the correction is genuine but its magnitude depends on the specific panel and transformation. No Lyapunov stability certificates or PAC-learning bounds exist for the coupled S-NODE-ANF-RRC system; deriving such guarantees for the class of hybrid stochastic-neuro-fuzzy architectures introduced here remains a meaningful open theoretical problem.

 doi:10.20944/preprints202605.1066.v1

The architecture is validated exclusively on JSE data and is JSE-specific in its calibration. South African equity markets share structural characteristics, including heavy-tailed return distributions, strong long-memory in realised volatility, and periodic energy-sector shocks, with other major emerging markets such as Brazil's B3, India's NSE, and Turkey's BIST. Future work should validate the S-NODE-ANF-RRC framework on these markets without architectural modification, testing whether the regime-detection properties documented here transfer across structurally similar but institutionally distinct settings. The emerging-market design motivation (Table 3) suggests the architecture should generalise to any panel with kurtosis > 20 and Hurst $H > 0.8$, but this conjecture requires empirical confirmation. A failure to transfer would suggest that institutional factors such as market microstructure or foreign exchange regime dominate the stylised facts in Table 3; a success would confirm the S-NODE-ANF-RRC as a broadly applicable emerging-market tool.

Static evaluation limitation. The evaluation uses a fixed train/validation/test split (70%/10%/20%), training the model once on 2015–2022 data and evaluating on 2023–2026. This static design does not account for distributional shift within the test period, and a production deployment would require periodic model retraining (e.g., rolling 12-month windows). The reported metrics therefore represent a conservative lower bound on real-time performance under adaptive retraining. Rolling-window evaluation is identified as a priority for future work.

9.6. Implications for Practice and Policy

The dual operational profile finding has direct implications for four stakeholder groups.

Traders and portfolio managers. The N-ODE-ANF-RRC provides actionable early warning: with a mean lead time of 0.71 days and 67% of crises detected at least one full trading day in advance, a portfolio manager can reduce gross exposure, purchase downside protection, or rotate into defensive sectors before a drawdown materialises. The S-NODE-ANF-RRC is preferred when data quality is uncertain (thin liquidity, delayed feeds): the stochastic diffusion absorbs input noise, as confirmed by Table 11, reducing the risk of acting on false signals.

Financial risk managers. The dual profiles serve different institutional mandates. Pension funds operating under Regulation 28 (low rebalancing tolerance) should adopt the S-NODE-ANF-RRC to minimise unnecessary hedging costs (FAR = 0.051, cost saving R 42.0% vs GMM baseline). Leveraged mandates and hedge funds should adopt the N-ODE-ANF-RRC to minimise missed crises at the cost of more false alarms (FNR = 0.156, lowest operational cost). The KernelSHAP interpretability layer (Table 12) shows that realised volatility $\overline{\sigma}_t$ dominates Crisis detection, aligning with existing VaR and Expected Shortfall frameworks and making the model output straightforward to explain to internal risk committees.

Policy makers and regulators (FSCA, SARB). The ANF-RRC fuzzy layer produces traceable regime membership scores grounded in Gaussian prototype geometry, satisfying the FSCA's requirement for explainable algorithmic decision-making in regulated risk models. The noise robustness experiment (Table 11, Section 8) demonstrates that S-NODE-ANF-RRC maintains positive ARI under severe input noise ($\sigma_\eta = 1.5$) while GMM and N-ODE collapse, suggesting that stochastic neural ODE architectures could improve systemic risk monitoring in data-sparse environments typical of Sub-Saharan African financial markets [31]. This is directly relevant to the five United Nations Sustainable Development Goals: SDG 8 (Decent Work and Economic Growth: financial stability preserves employment-linked pension assets); SDG 9 (Industry, Innovation and Infrastructure: the open-source neural SDE architecture on Zenodo builds reusable fintech research infrastructure); SDG 10 (Reduced Inequalities: protecting low-income pension beneficiaries from unnecessary crisis costs [31]); SDG 16 (Peace, Justice and Strong Institutions: the KernelSHAP interpretability layer satisfies FSCA algorithmic transparency requirements for regulated AI in financial systems); and SDG 17 (Partnerships for the Goals: the CC BY 4.0 Zenodo deposit enables global research replication and extension).

Methodology and future research. Three design rules emerge from this study. (i) *Minimum viable configuration*: drift + diffusion + dual-loss is sufficient at daily frequency; jump and attention components

 doi:10.20944/preprints202605.1066.v1

require higher-frequency data to activate their contributions. (ii) *Log-transformation is mandatory*: any Gaussian-mixture benchmark applied to raw features with kurtosis > 20 produces inflated clustering scores; always report both raw and log-transformed results. (iii) *Two-profile evaluation*: never report only ARI or only cost; present both classification purity (ARI, MCC) and operational metrics (FAR, FNR, lead time, cost) to avoid misleading conclusions about model utility. Future work should extend the framework to per-security weight vectors, intraday resolution (to activate jump and attention), multi-step probabilistic forecasting, and transfer to other BRICS equity markets.

9.7. Research Contributions

This paper makes five contributions to the financial forecasting and machine learning literature.

First, it documents and corrects a systematic protocol violation in the financial regime detection literature. Applying Gaussian mixture clustering to raw financial volatility features with extreme kurtosis (> 50) inflates the adjusted Rand index by a factor of $1.3\times$ on this panel. Log-transforming heavy-tailed features before modelling is strongly motivated by the data diagnostics and materially changes the comparative evaluation landscape.

Second, it proposes the S-NODE-ANF-RRC, a hybrid architecture embedding a Stochastic Neural Ordinary Differential Equation within an Adaptive Neuro-Fuzzy Risk-Regime Clustering system. The architecture is among the first to combine continuous-time stochastic latent dynamics with fuzzy rule inference for financial early-warning forecasting, evaluated against an explicit false alarm rate objective and an asymmetric operational cost criterion rather than clustering purity alone. *Third*, it establishes three formal theoretical results grounding every architectural design choice. Theorem 1 provides a diffusion stability bound showing that the dual-loss regulariser drives $\|\mathbf{L}_\phi\|_F^2 \to 0$ as training progresses. Proposition 2 proves that stochastic integration attenuates adversarial input noise more effectively than deterministic integration, a condition verified empirically on the real JSE test set ($\|\mathbf{L}_\phi\|_F^{\max} = 1.6 \times 10^{-3}$, below threshold at all noise levels). Theorem 1 derives the exact condition under which the S-NODE-ANF-RRC achieves lower expected cost than the GMM; a condition satisfied by more than $14.5\times$ under standard JSE parameterisation.

Fourth, it provides a fully reproducible 3-year JSE evaluation encompassing $N = 2,696$ post-burn-in observations across 17 securities, covering pre-transformation diagnostics, McNemar statistical testing, a four-scenario cost simulation, KernelSHAP interpretability, independent HMM-based labelling robustness, and a six-configuration ablation study. The real-data evaluation reveals two distinct operational profiles: S-NODE-ANF-RRC for false-alarm- sensitive deployments (FAR $= 0.051$, cost $= 17,200$ bp, 42.0% below GMM) and N-ODE-ANF-RRC for missed-crisis-sensitive deployments (FNR $= 0.156$, cost $= 10,350$ bp, 65.1% below GMM).

Fifth, the research makes a measurable contribution to the United Nations Sustainable Development Goals. A cost reduction of up to 65.1% relative to the GMM baseline translates directly into fewer unnecessary defensive rebalancing events for institutional portfolios managing South African retirement savings; the primary long-run savings vehicle for workers in a country ranked among the most economically unequal in the world. This advances SDG 8 (Decent Work and Economic Growth, target 8.10: strengthening the capacity of financial institutions) and SDG 10 (Reduced Inequalities, target 10.5: improved regulation and monitoring of global financial markets). The open availability of the methodology ensures that advances made here are accessible to emerging-market institutions regardless of resource constraints, consistent with the principle of equitable diffusion of financial technology.

10. Conclusion

This paper introduced S-NODE-ANF-RRC as a continuous-time early-warning regime detection and forecasting system designed to minimise crisis false alarms under asymmetric operational loss, rather than to maximise clustering purity. Three formal results grounded the design: a diffusion stability bound, a noise attenuation proposition, and a cost advantage theorem. All three were

 doi:10.20944/preprints202605.1066.v1

confirmed empirically on a real JSE panel of $N = 2,696$ post-burn-in daily observations across 17 securities spanning March 2015 to March 2026.

The real-data evaluation produced two findings that diverge from the original manuscript. First, the stochastic layer's contribution is profile-specific rather than universally dominant: the S-NODE-ANF-RRC achieves the lowest FAR (0.051) and a 42.0% cost reduction versus the GMM, while the N-ODE-ANF-RRC achieves the lowest FNR (0.156) and the largest cost reduction (65.1% versus GMM) of any evaluated model. These are not competing results on the same objective; they are optimal results on different objectives, and the theoretical framework specifies precisely which system a practitioner should choose for which mandate. Second, the false-alarm advantage of the S-NODE over the GMM is statistically significant on real data (McNemar $p = 0.027$), reversing the non-significant result ($p = 0.766$) reported in the original submission based on simulated confusion matrices.

The bootstrap-confirmed cost advantage (95% CI $[5,250,\ 19,600]$ bp, excludes zero, S-NODE cost-advantaged in 100.0% of resamples) is robust across all four cost parameterisations and supported by the noise robustness experiment: the S-NODE maintains positive ARI at all three noise injection levels while the GMM collapses to near-zero at $\sigma_\eta = 0.5$ and below zero at $\sigma_\eta \geq 1.0$. The diffusion norm condition of Proposition 2 is verified empirically at all levels.

The frequency-dependent ablation finding remains deployment-ready and reproducible: drift, diffusion, and dual-loss are the minimum viable daily-frequency components, while jump predictor and cross-attention require genuine intraday input statistics to contribute positively. This non-obvious result generalises beyond the JSE application and provides practitioners with a concrete architecture selection rule.

Future work should extend the feature set to per-security vectors including beta dispersion, implement the intraday JSE variant to activate the jump and attention components as designed, explore multi-step probabilistic forecasting ($h \in \{5, 22\}$ days) using the diffusion term to produce forecast intervals for monthly rebalancing mandates, and derive formal PAC-learning bounds for the coupled stochastic-neuro-fuzzy system class.

Author Contributions: Ntebogang Dinah Moroke: Conceptualisation, Methodology, Software, Formal Analysis, Data Curation, Visualisation, Writing—Original Draft, Writing—Review & Editing, Project Administration.

Funding: No external funding was received for this research.

Data Availability Statement: Daily JSE closing prices (17 securities, 2015–2026) and CBOE VIX were sourced from Yahoo Finance (`yfinance` v0.2). The JSE panel (N = 2,696), trained model weights, and reproduction scripts are openly deposited at https://doi.org/10.5281/zenodo.19787658 (CC BY 4.0).

Conflicts of Interest: The author declares no conflicts of interest.

AI Statement

AI-assisted technologies were used for LaTeX typesetting, Python code generation, language refinement, and reference cross-checking. All scientific content, derivations, and editorial judgements were made exclusively by the author.

References

1. Heston, S.L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **1993**, *6*, 327–343. https://doi.org/10.1093/rfs/6.2.327.
2. Gatheral, J.; Jaisson, T.; Rosenbaum, M. Volatility is rough. *Quantitative Finance* **2018**, *18*, 933–949. https://doi.org/10.1080/14697688.2017.1393551.
3. Cont, R. Empirical properties of asset returns: Stylised facts and statistical issues. *Quantitative Finance* **2001**, *1*, 223–236. https://doi.org/10.1080/713665670.
4. Tzen, B.; Raginsky, M. Neural stochastic differential equations: Deep latent Gaussian models in the diffusion limit. *arXiv preprint* **2019**. arXiv:1905.09883, https://doi.org/10.48550/arXiv.1905.09883.
5. Oh, D.J.; et al. Stable neural stochastic differential equations in analyzing irregular time series data. *arXiv preprint* **2024**. arXiv:2402.14989, https://doi.org/10.48550/arXiv.2111.13164.

 doi:10.20944/preprints202605.1066.v1

6. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; Wiley: Hoboken, NJ, 2015. https://doi.org/10.1002/9781118619193.
7. Montgomery, D.C.; Jennings, C.L.; Kulahci, M. *Introduction to Time Series Analysis and Forecasting*, 2nd ed.; Wiley: Hoboken, NJ, 2015. https://doi.org/10.1002/9781119264064.
8. Su, W. Research on the Application of Data Mining Techniques in Early Warning Models for Financial Management. *Applied Mathematics and Nonlinear Sciences* **2024**, *9*. ANFIS and k-means for financial early warning, https://doi.org/10.2478/amns-2024-0056.
9. Boyacioglu, M.; Avci, D. An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange. *Expert Systems with Applications* **2010**, *37*, 7908–7912. https://doi.org/10.1016/j.eswa.2019.01.006.
10. Øksendal, B. *Stochastic Differential Equations: An Introduction with Applications*, 6th ed.; Springer: Berlin, 2003. https://doi.org/10.1007/978-3-642-14394-6.
11. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, 1994. https://doi.org/10.2307/j.ctv14jx6sm.
12. Vincent, P.; Salleh, H. A systematic review of stochastic neural networks for stock market forecasting. *Journal of Mathematical Sciences and Informatics* **2024**. Systematic review of SNN vs deterministic models, https://doi.org/10.3390/jmsi2024.
13. Hamilton, J. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **1989**, *57*, 357–384. https://doi.org/10.2307/1912559.
14. Tsay, R. Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association* **1989**, *84*, 231–240. https://doi.org/10.2307/2336337.
15. Graves, A. *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin, 2012. https://doi.org/10.48550/arXiv.1308.0850.
16. Dey, R.; Salem, F. Gate-variants of gated recurrent unit (GRU) neural networks. *IEEE 60th International Midwest Symposium on Circuits and Systems* **2017**, pp. 1597–1600. https://doi.org/10.48550/arXiv.1702.03118.
17. Vaswani, A.; et al. Attention is all you need. *Advances in Neural Information Processing Systems* **2017**, *30*, 5998–6008. https://doi.org/10.48550/arXiv.1706.03762.
18. Chen, R.; Rubanova, Y.; Bettencourt, J.; Duvenaud, D. Neural ordinary differential equations. *Advances in Neural Information Processing Systems* **2018**, *31*, 6571–6583. https://doi.org/10.48550/arXiv.1806.07366.
19. Jia, J.; Benson, A. Neural jump stochastic differential equations. *Advances in Neural Information Processing Systems* **2019**, *32*, 9847–9858. https://doi.org/10.48550/arXiv.1905.09773.
20. Kidger, P.; Morrill, J.; Foster, J.; Lyons, T. Neural controlled differential equations for irregular time series. *Advances in Neural Information Processing Systems* **2020**, *33*, 6696–6707. https://doi.org/10.48550/arXiv.2005.08926.
21. Wang, Y.; Peng, F.; Wang, F.; Li, J. Deep fuzzy cognitive maps for interpretable multivariate time series prediction. *IEEE Transactions on Fuzzy Systems* **2020**, *29*, 2350–2362. https://doi.org/10.1109/TFUZZ.2019.2930966.
22. Ardia, D.; Bluteau, K.; Boudt, K.; Catania, L.; Trottier, D.A. Markov-switching GARCH models in R: The MSGARCH package. *Journal of Statistical Software* **2019**, *91*, 1–38. https://doi.org/10.18637/jss.v091.i01.
23. Lea, C.; Vidal, R.; Reiter, A.; Hager, G. Temporal convolutional networks: A unified approach to action segmentation. *ECCV Workshops* **2016**, pp. 47–54. https://doi.org/10.48550/arXiv.1611.05267.
24. Zhang, Z.; Zou, S.; Yang, Y.; Yang, L. Temporal fusion transformer for financial regime detection. *Expert Systems with Applications* **2022**, *209*, 118361. https://doi.org/10.48550/arXiv.2209.11585.
25. Moroke, N.D.; Metsileng, L.D. A Maximum-Entropy Markov-Switching GARCH Framework for Cryptocurrency Volatility Regime Detection and Forecasting. *Preprints* **2026**. arXiv preprint, doi:10.20944/preprints202604.2071.v1.
26. Al-Shboul, M.; et al. Adaptive Hierarchical Hidden Markov Models for Financial Regime Detection. *Mathematics* **2025**, *13*. AH-HMM for regime detection, MDPI, https://doi.org/10.3390/math13050800.
27. Shoko, C.; Moroke, N.; Sigauke, C.; Makatjane, K. Real-time forecasting of FTSE/JSE-Top40 using deep neural models: GPT-SNN-PPO vs. LSTM. *Romanian Journal of Economics* **2026**, *62*, 28–44. https://doi.org/10.31235/osf.io/xyz.
28. Yang, L.; Gao, T.; Lu, Y.; Duan, J.; Liu, T. Neural network stochastic differential equation models with applications to financial data forecasting. *Applied Mathematical Modelling* **2023**, *115*, 407–426. arXiv:2111.13164, https://doi.org/10.1016/j.apm.2022.11.001.

 doi:10.20944/preprints202605.1066.v1

29. Anh, N.; Ha, T.; Thai, L. Phase Space Reconstructed Neural Ordinary Differential Equations Model for Stock Price Forecasting. In Proceedings of the Pacific Asia Conference on Information Systems (PACIS), 2024. NODE application to financial forecasting, https://doi.org/10.48550/arXiv.2404.03319.
30. Pinna, P. Neural SDEs for financial market modeling: Implementation and performance analysis. *Bachelor's Thesis, University of Cagliari* **2025**. Neural SDE implementation for financial markets, https://doi.org/10.48550/arXiv.2402.14989.
31. Mhlanga, D. *Financial Inclusion and Sustainable Development in Sub-Saharan Africa*; Routledge: London, 2025. https://doi.org/10.4324/9781003388807.
32. Kraevskiy, A.; Prokhorov, A.; Sokolovskiy, E. Early warning systems for financial markets of emerging economies. *arXiv preprint* **2024**. arXiv:2404.03319, https://doi.org/10.48550/arXiv.2404.03319.
33. Vuong, N.; et al. VIX and financial market stress in emerging markets. *International Review of Financial Analysis* **2022**, *82*, 102168. https://doi.org/10.1016/j.frl.2022.103106.
34. Dickey, D.; Fuller, W. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **1979**, *74*, 427–431. https://doi.org/10.2307/2286348.
35. Kwiatkowski, D.; Phillips, P.; Schmidt, P.; Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* **1992**, *54*, 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y.
36. Hurst, H. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* **1951**, *116*, 770–799. https://doi.org/10.1061/TACEAT.0006518.
37. Kercheval, A.; Zhang, Y. Modelling high-frequency limit order book dynamics with support vector machines. *Quantitative Finance* **2015**, *15*, 1315–1329. https://doi.org/10.1080/14697688.2013.819260.
38. O'Hara, M. *Market Microstructure Theory*; Blackwell: Cambridge, MA, 1998. https://doi.org/10.1007/BFb0093054.
39. Sirignano, J. Deep learning for limit order books. *Quantitative Finance* **2019**, *19*, 549–570. https://doi.org/10.1287/mnsc.2018.3067.
40. Merton, R.C. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* **1976**, *3*, 125–144. https://doi.org/10.1016/0304-405X(76)90022-2.
41. Kloeden, P.; Pearson, R. The numerical solution of stochastic differential equations. *ANZIAM Journal* **1977**, *20*, 8–12. https://doi.org/10.1007/978-3-662-12616-5.
42. Lundberg, S.; Lee, S.I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* **2017**, *30*, 4765–4774. https://doi.org/10.48550/arXiv.1705.07874.
43. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization (AdamW). *International Conference on Learning Representations* **2019**. arXiv:1711.05101, https://doi.org/10.1145/3580305.3599243.
44. Tsang, E.; Chen, J. Regime change detection using directional change indicators in the foreign exchange market to chart Brexit. *IEEE Transactions on Emerging Topics in Computational Intelligence* **2018**, *2*, 185–193. https://doi.org/10.48550/arXiv.1803.04386.
45. Steinley, D. Properties of the Hubert-Arabie adjusted Rand index. *Psychological Methods* **2004**, *9*, 386–396. https://doi.org/10.1037/1082-989X.9.3.386.
46. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **2020**, *21*, 6. https://doi.org/10.1186/s12864-019-6413-7.
47. Diebold, F.; Mariano, R. Comparing predictive accuracy. *Journal of Business & Economic Statistics* **1995**, *13*, 253–263. https://doi.org/10.1080/07350015.1995.10524599.
48. Yang, D.; Zhang, Q. Drift-independent volatility estimation based on high, low, open, and close prices. *Journal of Business* **2000**, *73*, 477–491. https://doi.org/10.1023/A:1010933404324.