Knowledge & Data Engineering Systems

"A Distributed Framework for Financial Market Trend Prediction Using Hybrid Fuzzy Clustering and Hidden Markov Models"

Fatima Ghanduri, Christos Anagnostopoulos

School of Computer Science, University of Glasgow

University of Glasgow

IEEE SMC Systems, Man, and Cybernetics Society

About

School of Computer Science-University of Glasgow, UK

FATIMA GHANDURI
PhD Researcher - UNIVERSITY OF GLASGOW
Data & AI Technical Sales- Energy Economics– SLB

DR CHRISTOS ANAGNOSTOPOULOS
Associate Professor in Distributed Computing & Data Engineering
UNIVERSITY OF GLASGOW

Agenda

Knowledge & Data
Engineering Systems

- Finance Problem Statement
- Fault-Monitoring lens for markets
- Solution mapping & scope
- Results & Ablation study
- Deployment & Scalability
- Limitations & Takeaways

Why is it hard in finance?

signals **decay fast**, **regimes are overlapping** not crisp, microstructure adds **noise**, and non-stationarity makes centralized models brittle!

there is no single model that can universally predict prices across all market conditions. A model that performs well in one regime will break in another.

Knowledge & Data Engineering Systems

Why is it hard in finance?

there is no single model that can universally predict prices across all market conditions. A model that performs well in one regime will break in another.

signals **decay fast**, **regimes are overlapping** not crisp, microstructure adds **noise**, and non-stationarity makes centralized models brittle!

(1) Alpha decay & Drift → Online Adaption

(2) Overlapping regimes → Fuzzy Membership

(3) Fragmented liquidity/venues → Distributed nodes

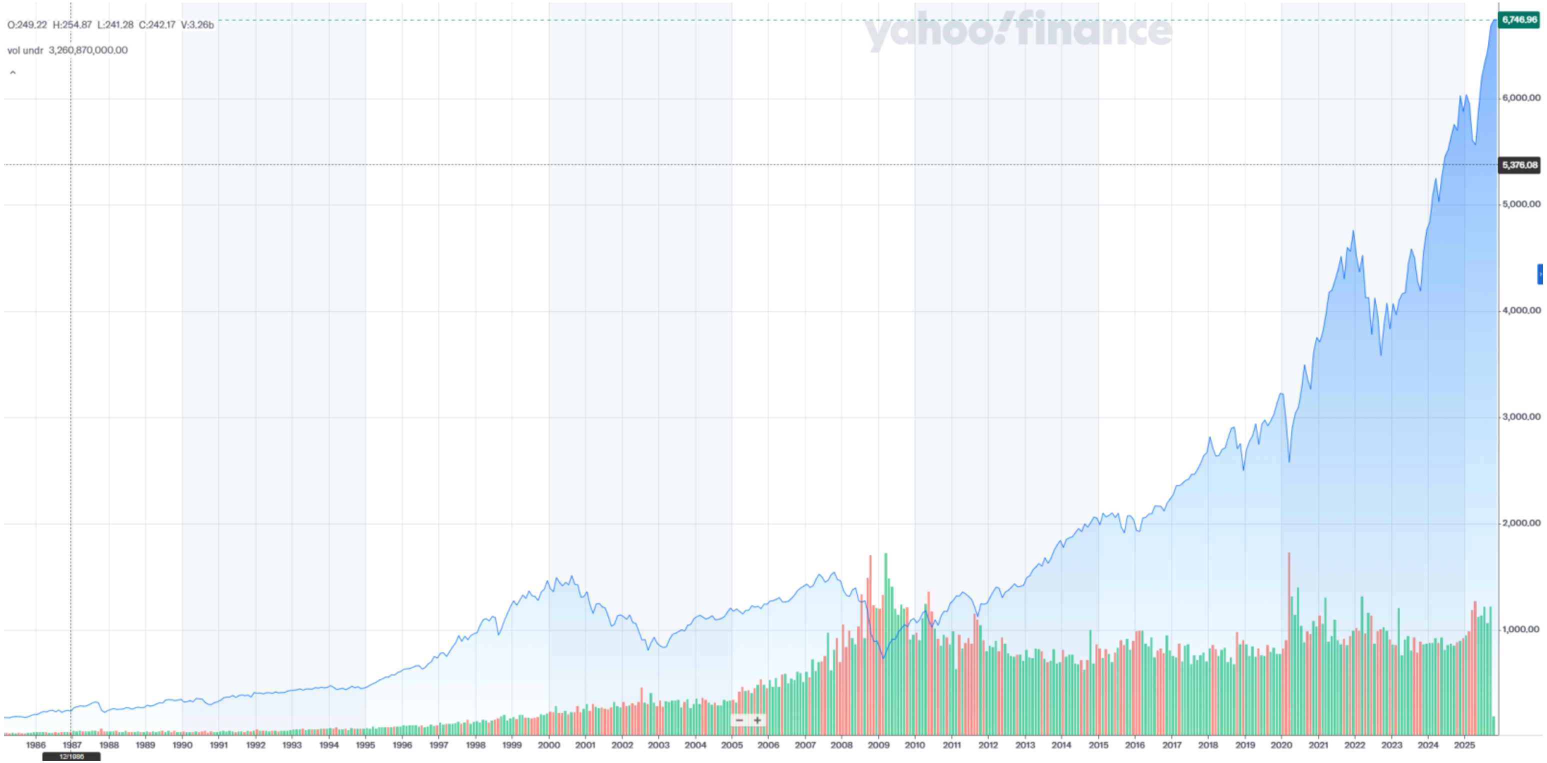
(4) Cost/slippage → Decision-relevant metrics

Not accuracy alone!

Knowledge & Data Engineering Systems

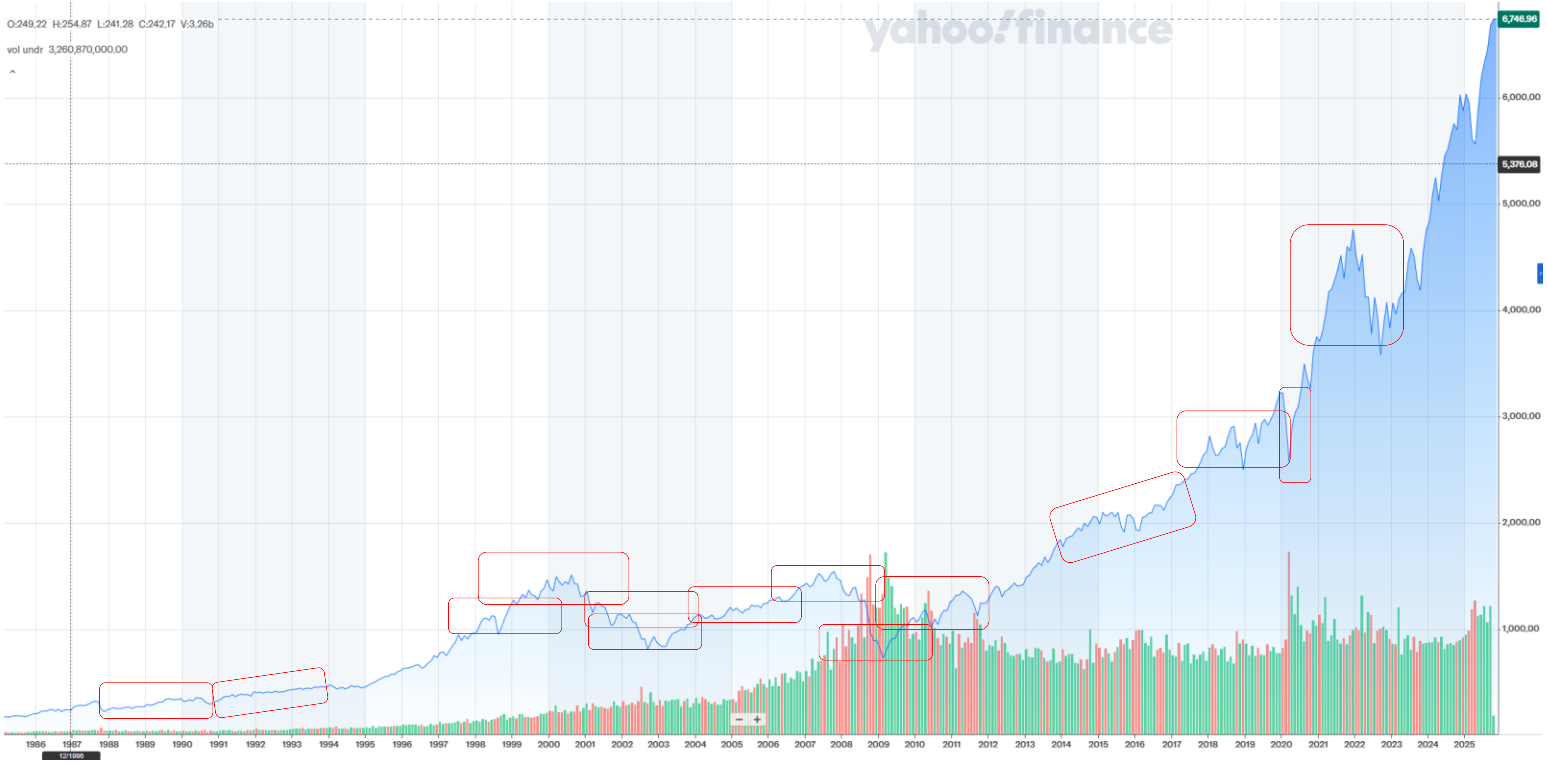
Fault-Monitoring lens for markets

Knowledge & Data Engineering Systems



Fault-Monitoring lens for markets

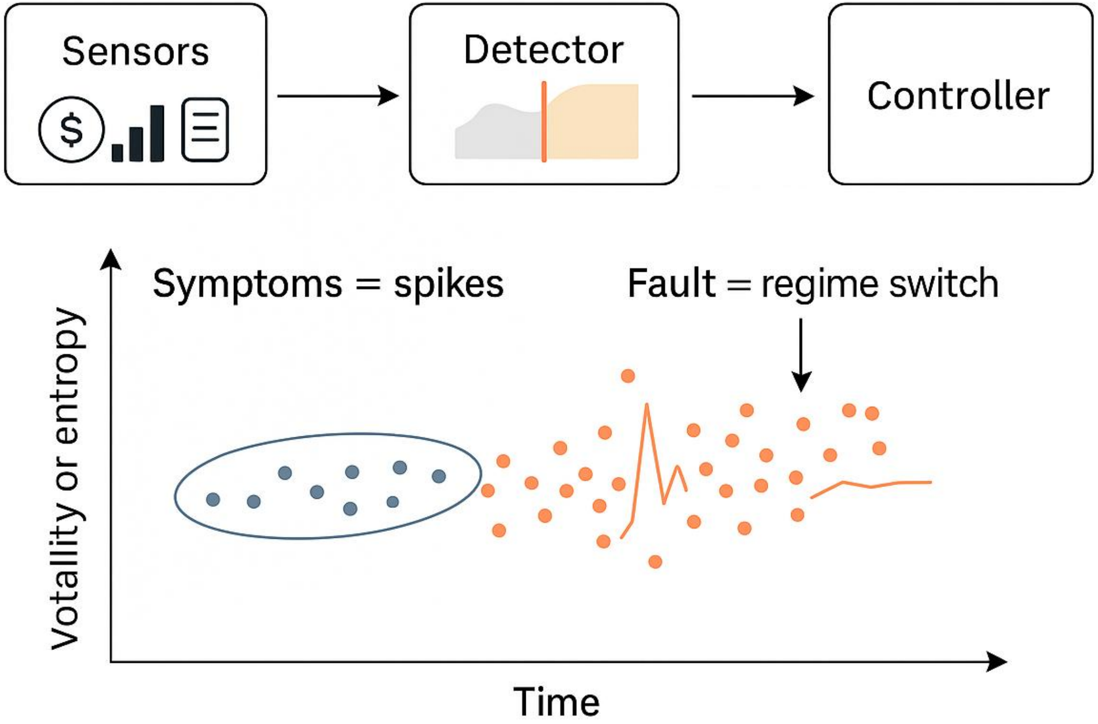
Knowledge & Data Engineering Systems



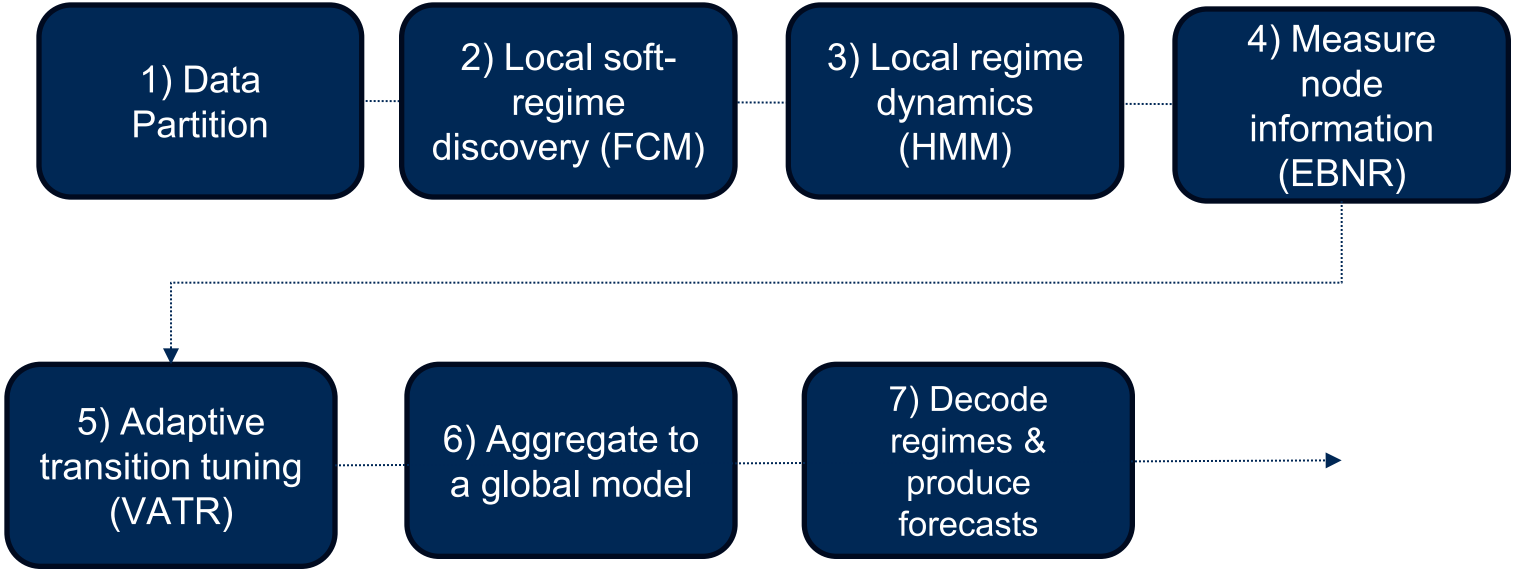
Yahoo Finance – S&P500 Prices

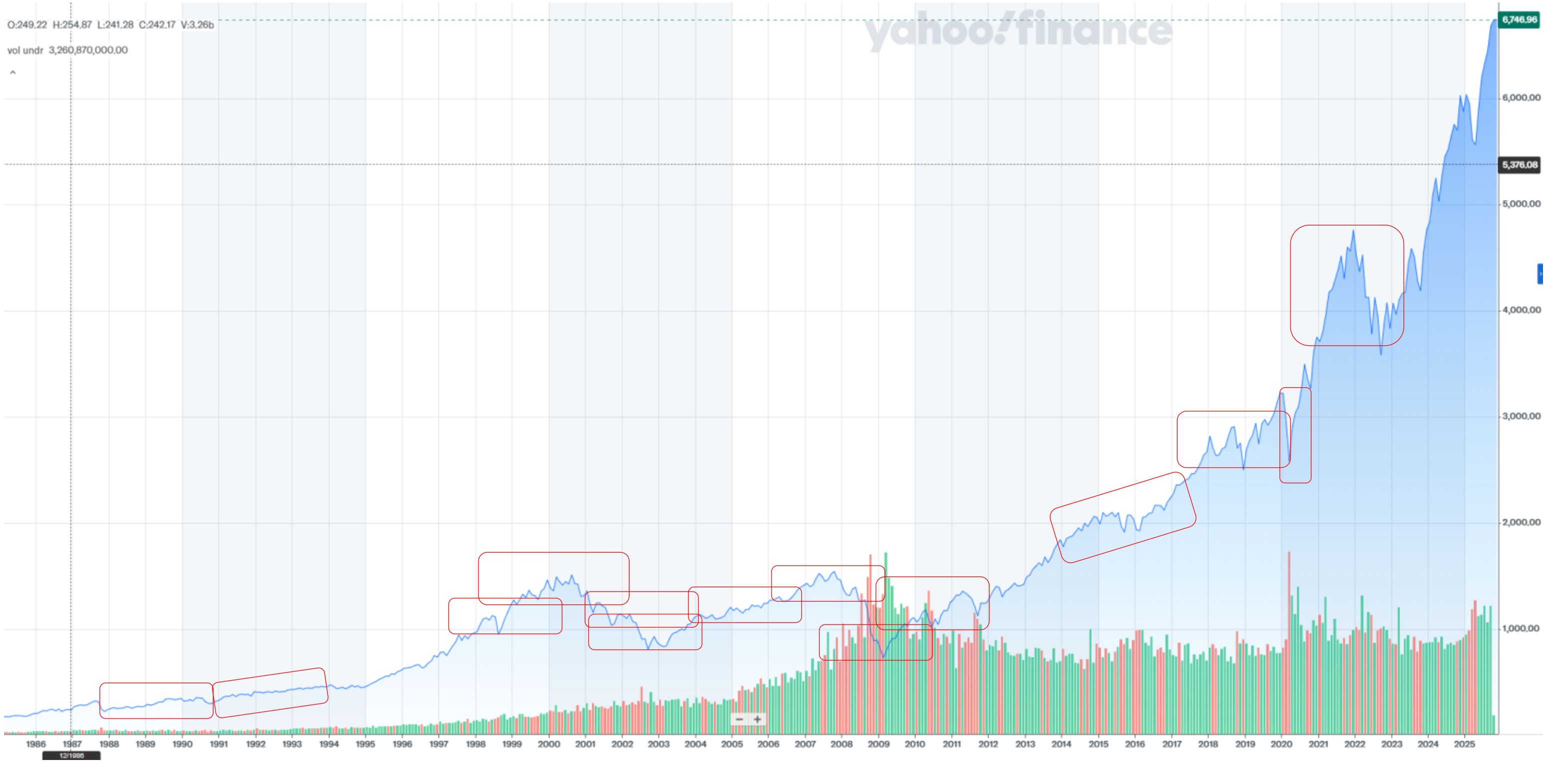
A **regime switch** is a **sudden change in the underlying statistical behavior of the market**.

Fault-Monitoring lens for markets



Solution Ideology: Proposition

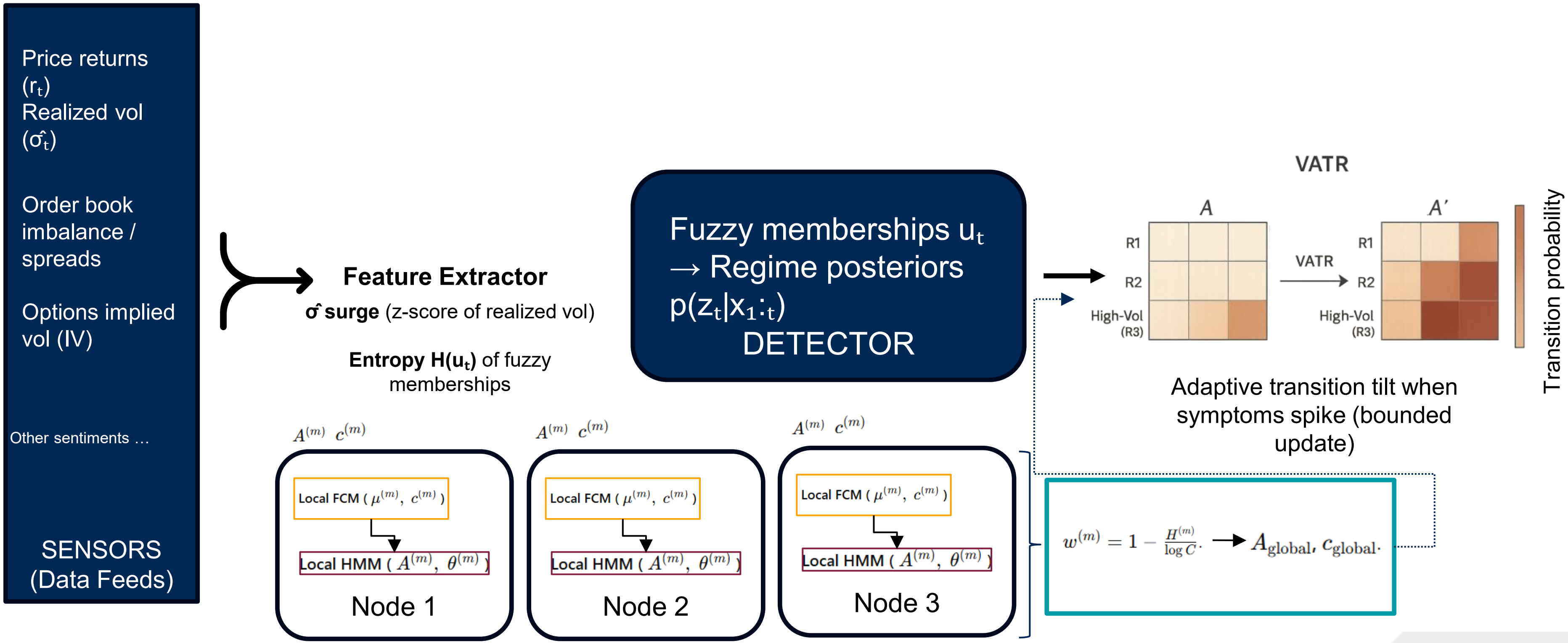




Knowledge & Data Engineering Systems

SLB-Private

Solution Ideology



Results

Knowledge & Data Engineering Systems

Detection quality, contributions, efficiency, robustness

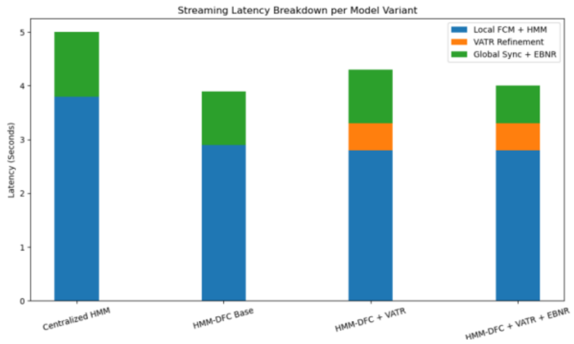
Model	FPC	Silhouette	Precision	Recall	F1
HMM-DFC (ours)	**0.7769**	**0.61**	**0.83**	**0.88**	**0.84**
Single HMM	—	—	0.71	0.76	0.73
GARCH	—	—	0.65	0.68	0.66
Distributed FCM	0.7200	0.54	0.70	0.73	0.71

+11 pts F1 vs single HMM; +18 pts vs GARCH; better clustering (FPC/Silhouette).

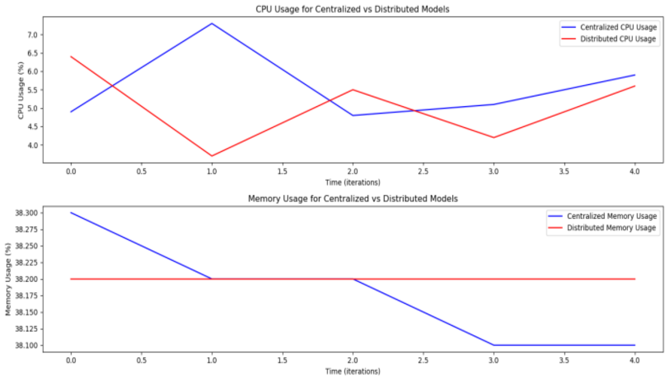
Base (no VATR/EBNR): F1 **0.77**, FPC **0.728**, Sil **0.53**
+ VATR: F1 **0.82**, FPC **0.749**, Sil **0.57** *(~+5 F1 from faster switching)*
+ VATR + EBNR (Full): F1 **0.84**, FPC **0.7769**, Sil **0.61** *(extra gain from entropy-weighted sync)*

Results

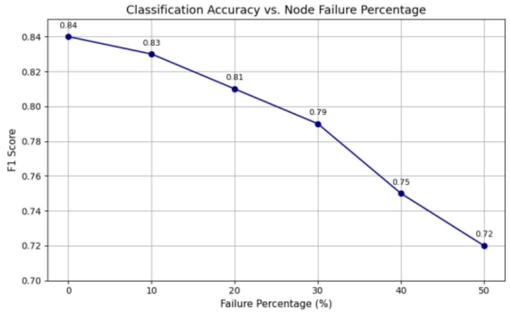
Detection quality, contributions, efficiency, robustness



Latency: Streaming breakdown shows **lower response time** than centralized; **VATR/Sync overhead is small**



Compute footprint: Smoother CPU and **lower memory** over iterations vs centralized



Robust to node loss: F1 degrades **gently up to ~30%** node failures (**≈0.84** → **0.79**), still **≈0.72** at **50%** failure

Ablation: why VATR & EBNR matter

Knowledge & Data Engineering Systems

Scalability & multi-node processing

*Each node learns local structure—close to the feed—and ships only A^m and c^m for **entropy-weighted sync**; the **VATR** controller then adapts the global transitions when risk spikes.*

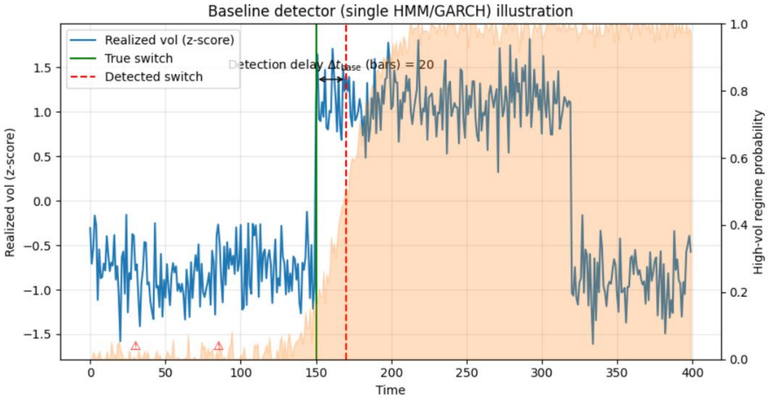
VATR (Volatility-Adaptive Transition Refinement)

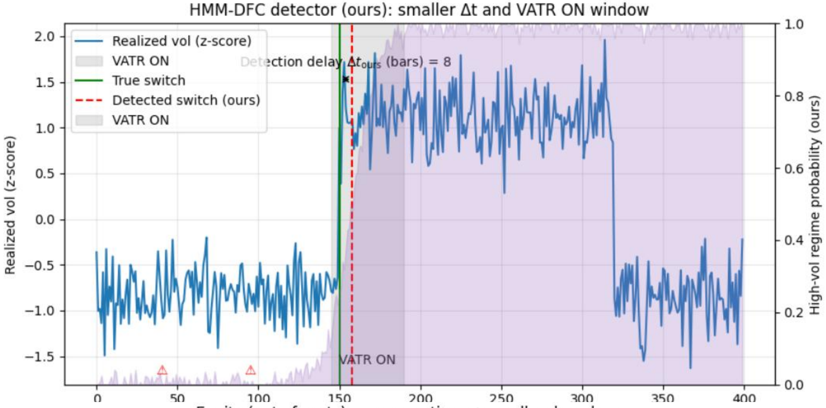
*Static HMM transitions **lag** during risk spikes → **late regime switches**, **whipsaw**, **turnover**.*

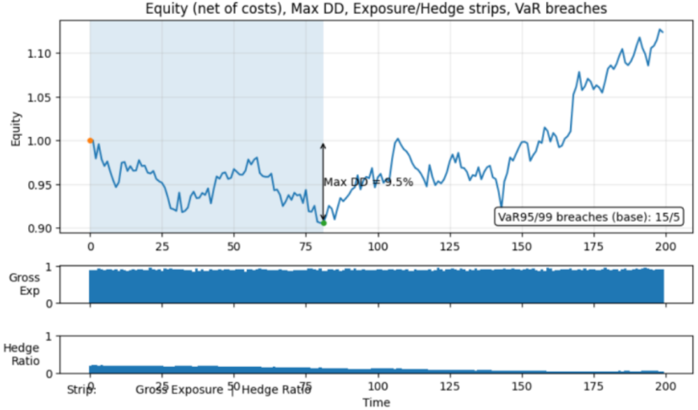
EBNR (Entropy-Based Node Reweighting)

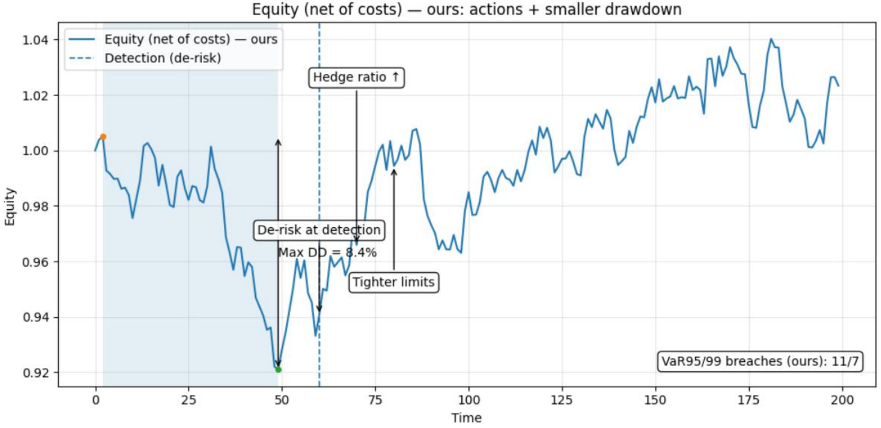
*In distributed data, some windows/nodes are **noisy** or **uninformative**; naive averaging **washes out** boundary signals.*

Stress case: Abrupt volatility regime (macro-shock)









Deployment & scalability

Knowledge & Data Engineering Systems

Why it scales:

- Add nodes, add coverage — no central bottleneck.
- Data stays local → less shuffling, lower latency.
- Resilient: if a node/coordinator blips, the system keeps the last good settings and degrades gracefully.
- In tests: faster response than a centralized setup and almost linear growth

Takeaways & Limitations

Knowledge & Data Engineering Systems

Takeaways

- **Earlier, safer switches:** VATR cuts detection delay and whipsaw.
- **Cleaner aggregation:** EBNR down-weights noisy windows → fewer false alarms.
- **Economic impact:** smaller drawdowns & fewer VaR/ES breaches; better net P&L after costs.
- **Deployable:** low-latency edge nodes, data stays local; fault-tolerant (graceful under 30–50% node loss).
- **Scalable:** add nodes to add assets; near-linear to ~7 clusters; VATR/Sync overhead is small.

Limitations

- **Labels & drift:** ex-post, feature-dependent → need rolling relabeling & stability monitors.
- **VATR tuning:** bad τ/β\tau/\betaτ/β can flip-flop → use bounded tilt, hysteresis, decay.
- **Entropy ≠ informativeness:** augment EBNR with liquidity/data-quality weights.
- **Sync staleness:** coordinator lag can stale → TTL + fallback to local.
- **Backtest realism/capacity:** cost/impact are proxies → validate with live/shadow & stress tests.

Knowledge & Data Engineering Systems

Thank you!

fghanduri20@gmail.com
f.ghanduri.1@research.gla.ac.uk

SCAN ME

Add me on linkedin!