Artificial Intelligence in Financial Risk Management: A Proposed Machine Learning Framework for Predicting Financial Market Stress in an Emerging-Market Context

Ioan-Anton SĂVULESCU, I.A.S.*
Petroleum-Gas University of Ploiesti, Ploiesti, Romania; Romanian Academy, School of Advanced Studies of the Romanian Academy, Doctoral School of Economic Sciences, National Institute for Economic Research "Costin C. Kirițescu", Institute of National Economy, Bucharest, Romania
isavulescu@upg-ploiesti.ro

Acknowledgments: This research was supported by grant from the Petroleum-Gas University of Ploiești, Romania, project number GO-GICS-14819/26.06.2026 within the Internal Grants for Scientific Research.

ABSTRACT

Artificial intelligence (AI) and machine learning (ML) are used in financial risk management, and Gartner (2024) predicted that 90% of finance functions would deploy at least one AI-enabled solution by 2026. Traditional autoregressive models, however, remain widely used to forecast financial market stress. Though they struggle to capture nonlinear and regime-dependent dynamics. Recent central bank evidence shows that tree-based ML models predict stress in US Treasury, foreign exchange and money markets more accurately than traditional time-series approaches. This paper does not report empirical results. It proposes a research framework for testing whether ML models outperform autoregressive benchmarks in forecasting stress in the equity, foreign exchange, money and government bond markets of an emerging economy, using Sri Lanka as the reference case. The framework fixes the stress measure, data, forecast horizons, chronological validation scheme, evaluation metrics and hypotheses before estimation. It also embeds model risk management and explainability in the design, drawing on supervisory model-risk guidance and explainable-AI research. The expected contribution is a replicable protocol and a governance-oriented evaluation approach for early-warning systems in data-limited emerging markets.

Keywords: *Artificial Intelligence, Machine Learning, Financial Market Stress, Early-Warning Systems, Model Risk Management, Emerging Markets*

JEL Classification: C53, G17, G32, E44

1. INTRODUCTION

Financial markets are closely interconnected through cross-border capital flows, financial innovation and technology. A shock in one segment, such as equities, can quickly spread to exchange rates, interest rates, credit spreads and liquidity. Early identification of market stress is therefore essential for financial institutions, central banks and supervisors. Financial market stress denotes conditions in which markets function abnormally or with unusual uncertainty, visible in sharp price movements, higher volatility, wider spreads, reduced liquidity and weaker investor confidence. Because several indicators typically move together, stress is multidimensional, which is why composite indicators such as the Composite Indicator of Systemic Stress (CISS) aggregate stress across market segments (Holló et al., 2012).

Autoregressive (AR) models are transparent and capture the persistence of stress, but they impose a fixed functional form and may deteriorate under nonlinearities, structural breaks and interactions among many predictors. ML methods can learn such patterns directly from data (Mullainathan & Spiess, 2017; Gu et al., 2020). Random forests (Breiman, 2001) average many decision trees and handle interactions without prespecification. Aldasoro et al. (2025) show that tree-based models significantly outperform traditional time-series approaches in predicting the distribution of market conditions in key US markets. Related central bank work finds that random forests also outperform logit models in early warning of fiscal stress, correctly predicting 70–80% of stress events and tranquil periods (Jarmulska, 2020).

Evidence concentrates on advanced economies, while emerging markets have shorter and noisier data, more frequent structural change and thinner liquidity, so transferability of ML advantages is uncertain. Second, adoption in finance has advanced faster than governance: a model that forecasts well but cannot be explained, validated and monitored is unsuitable for supervisory use (Board of Governors of the Federal Reserve System, 2011; Rudin, 2019). This is a conceptual and methodological paper presenting a pre-specified research protocol. It contains no estimated results, and statements about ML performance describe hypotheses, not findings.

2. AIMS/OBJECTIVES

The primary aim is to specify and justify a protocol for testing whether ML models forecast financial market stress in an emerging economy more accurately than autoregressive benchmarks.

1. Do ML models forecast stress more accurately than AR benchmarks at horizons of 1, 3, 6 and 12 months?
2. Do ML models provide earlier and more reliable warnings of high-stress episodes, at acceptable false-alarm rates?
3. Which predictors drive forecasts, and are they stable across regimes?

4. What model-risk and explainability controls are needed to deploy such models responsibly?

3. HYPOTHESES

H1: Random forests achieve lower out-of-sample forecast loss than AR benchmarks, with larger gains in the upper tail of stress and at longer horizons.

H2: Random forests give longer warning lead times than AR and logit benchmarks at comparable false-alarm rates.

H3: Predictive gains are larger in high-stress regimes than in calm periods.

4. THEORY

The financial instability hypothesis holds that stability breeds risk-taking and leverage that eventually produce abrupt, nonlinear adjustments (Minsky, 1986). This implies threshold effects and interactions, for example funding-liquidity strains combined with currency depreciation, which linear AR specifications represent poorly and tree-based models, such as random forests (Breiman, 2001), can approximate. Portfolio-theoretic stress indices add that systemic stress rises when stress is correlated across market segments (Holló et al., 2012), which motivates multi-market predictors. Empirically, Gu et al. (2020) show that tree-based and neural methods capture nonlinear predictor interactions in asset-pricing forecasts. Aldasoro et al. (2025) use Shapley value analysis to identify funding liquidity, investor overextension and the global financial cycle as important predictors of future tail realisations of market conditions. Evidence for frontier and emerging markets remains scarce.

For model risk, supervisory guidance defines model risk as the potential for adverse consequences from decisions based on incorrect or misused models and calls for sound development, independent validation and ongoing monitoring (Board of Governors of the Federal Reserve System, 2011). Explainable-AI research offers post hoc attribution through Shapley values (Lundberg & Lee, 2017) while cautioning that interpretable models should be preferred in high-stakes settings where they perform comparably (Rudin, 2019). Regulation of high-risk AI also emphasises documentation, human oversight and robustness (European Parliament & Council, 2024).

5. METHODS

The reference case is Sri Lanka, an emerging economy that has recently experienced pronounced external, sovereign and currency stress. The proposed sample is monthly observations from January 2010 to December 2025, subject to data-availability checks. If Sri Lankan series prove too short or discontinuous, the protocol allows a robustness extension to a panel of comparable emerging markets.

Stress is measured by a composite index built in the CISS tradition (Holló et al., 2012) from four segments: equity, foreign exchange, money market and government bonds. Each segment subindex averages three raw stress measures (Table 1), each transformed with an empirical cumulative distribution function computed recursively on an expanding window, so no future information enters. Subindices are aggregated using time-varying cross-correlations. The continuous index is the target for regression forecasts. The binary target equals 1 when the index exceeds its recursively estimated 80th percentile (sensitivity: 75th and 90th percentiles). Forecasts are made h months ahead, with h in {1, 3, 6, 12}.

Table 1. Proposed Stress Measures and Predictors

Segment	Stress measures (target index)	Additional predictors
Equity	Negative return of the All-Share Price Index, realized volatility, maximum drawdown	Trading volume, market capitalisation change
Foreign exchange	LKR/USD depreciation, realised exchange-rate volatility, reserve change	Current account, external debt service, VIX, US policy rate
Money market	Overnight/policy-rate spread, short-term rate volatility, interbank spread	Liquidity injections, reserve money growth
Government bonds	Yield change, term spread, yield volatility	Sovereign spreads, credit growth, inflation, output proxy

Note. Variable lists are proposed and will be finalized after data-availability checks.

Sources: Central Bank of Sri Lanka, Colombo Stock Exchange, and international databases for global factors.

All predictors are lagged so that only information available at the forecast origin is used. Series are transformed to stationarity, standardized using training-window moments only, and checked for missing values (forward fill limited to one month; otherwise, exclusion). Outliers are retained and assessed through minorized robustness checks, because extremes are the object of interest. Benchmarks are (i) AR(p) with p chosen by the Bayesian information criterion up to 12 lags, (ii) AR with exogenous macro-financial variables, and (iii) for the binary target, penalised logistic regression. ML models are random forest (primary; regression and quantile versions), gradient-boosted trees and an elastic net as a low-complexity control. Including several algorithms tests whether any gain is specific to random forests. ML models use 1, 3 and 6 monthly lags of each predictor plus the lagged target.

Data are split chronologically: initial training 2010–2017, expanding-window monthly re-estimation, and an out-of-sample test period 2018–2025 that includes major stress episodes. Hyperparameters are tuned by time-series cross-validation

inside the training window, with purging and an embargo of h months between training and validation folds to prevent leakage from overlapping horizons (López de Prado, 2018). Random shuffling is not used, and grids and seeds are reported.

Point forecasts are evaluated with mean absolute error and root mean squared error, and tail forecasts with quantile loss at the 50th, 90th and 95th quantiles. Accuracy differences against the AR benchmark are tested with the Diebold–Mariano test (Diebold & Mariano, 1995). Binary forecasts are evaluated with ROC-AUC, precision, recall, F1 and precision–recall curves. Early-warning value is measured by warning lead time, false-alarm rate and missed-event rate, reported separately for stress episodes and calm periods. Variable importance uses permutation importance and Shapley values (Lundberg & Lee, 2017), interpreted as predictive, not causal.

Governance is operationalized in four testable elements: (1) documentation of purpose, data lineage, assumptions and limitations in a model card; (2) independent validation through replication by a second analyst, back-testing and benchmark comparison; (3) explainability through global and local Shapley explanations, regime-stability tests of rankings and an interpretable challenger model; and (4) monitoring of forecast errors, data drift and predictor-importance shifts, with predefined thresholds triggering review, human override and recalibration. Costs are made explicit through a loss function that weights missed stress events more heavily than false alarms.

6. RESULTS

No empirical results are reported, because the protocol has not yet been executed. Table 2 shows the output that the completed study will report, so that reviewers can see how hypotheses H1–H3 will be judged. No value in this paper should be read as a finding.

Table 2. Illustrative Household Sorting Outcomes

Output	Content	Hypothesis
Forecast accuracy	MAE, RMSE and quantile loss by model, market segment and horizon; Diebold–Mariano p-values vs AR	H1
Classification	ROC-AUC, precision, recall and F1 for the binary stress target; ROC and precision–recall curves	H1, H2
Early-warning value	Warning lead time, false-alarm and missed-event rates for each stress episode	H2
Regime analysis	Metrics in high stress versus calm periods	H3
Explainability	Shapley and permutation importance rankings; stability across regimes	RQ3

7. DISCUSSION

Because no estimation has been performed, this section describes how outcomes would be interpreted. If H1–H3 hold, they would extend evidence from advanced-economy markets (Aldasoro et al., 2025) to an emerging market. If they fail, the finding would be equally informative: it would indicate that short samples and structural breaks limit ML advantages and that simple benchmarks remain adequate.

Some challenges are foreseeable. Complex models can overfit when observations are few relatives to predictors, so inference rests on out-of-sample results and on comparison with the elastic net and boosted trees. Foreign exchange stress in small open economies is strongly driven by external shocks, so global factors are included. Money-market stress may reflect policy intervention, which can weaken historical relationships. Segments are therefore evaluated separately before generalizing.

8. IMPLICATIONS AND LIMITATIONS

The protocol allows a central bank or supervisor to compare ML and AR early-warning tools under identical, leakage-free conditions and to judge them on lead time, false alarms and explainability, not accuracy alone. For practice, human oversight, local calibration and sensitivity analysis should complement algorithmic forecasts where data are short or structural change is frequent. For theory, the design provides a test of whether nonlinear, interaction-based mechanisms of instability are detectable in an emerging market.

The study is not yet implemented and presents no evidence of ML superiority. It depends on the availability and consistency of Sri Lankan data for 2010–2025, relies on one country, and uses a constructed stress index whose thresholds involve judgement. Predictor importance does not imply causation. Future work should execute the protocol, report full empirical results, extend it to other emerging markets and compare additional explainable-AI methods.

REFERENCES

Aldasoro, I., Hördahl, P., Schrimpf, A., & Zhu, S. X. (2025). *Predicting financial market stress with machine learning* (BIS Working Paper No. 1250). Bank for International Settlements.

Board of Governors of the Federal Reserve System. (2011). *Supervisory guidance on model risk management* (SR 11-7). Federal Reserve.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*.

Gartner. (2024, September 12). *Gartner predicts that 90% of finance functions will deploy at least one AI-enabled technology solution by 2026*. https://www.gartner.com/en/newsroom/press-releases/2024-09-12-gartner-predicts-that-90-percent-of-finance-functions-will-deploy-at-least-one-ai-enabled-tech-solution-by-2026

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Holló, D., Kremer, M., & Lo Duca, M. (2012). *CISS – A composite indicator of systemic stress in the financial system* (ECB Working Paper No. 1426). European Central Bank.

Jarmulska, B. (2020). *Random forest versus logit models: Which offers better early warning of fiscal stress?* (ECB Working Paper No. 2408). European Central Bank.

López de Prado, M. (2018). *Advances in financial machine learning*. Wiley.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Minsky, H. P. (1986). *Stabilizing an unstable economy*. Yale University Press.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215.