

A Dynamic 4s-dsdm Framework for Banking Supervision: Evidence From the South Caucasus Banking Sector Transformation

Gayane Harutyunyan

g.harutyunyan7@gmail.com

Armenian State University of Economics <https://orcid.org/0000-0003-0874-4517>

Research Article

Keywords: Banking Supervision, Dynamic Weighting, K-Means Clustering, Machine Learning, Macroprudential Policy, Shannon Entropy, South Caucasus, Systemic Risk

Posted Date: September 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11095213/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

A DYNAMIC 4S-DSDM FRAMEWORK FOR BANKING SUPERVISION: EVIDENCE FROM THE SOUTH CAUCASUS BANKING SECTOR TRANSFORMATION

Gayane Yuri Harutyunyan (Corresponding author)
**Associate Professor, Department of Financial Markets and
Institutions,**
Armenian State University of Economics, Yerevan, Armenia
g.harutyunyan7@gmail.com, gayane.harutyunyan@asue.am

ABSTRACT

Between 2024 and 2026, the three South Caucasus banking systems diverged sharply in trajectory despite strong headline growth across all three. Armenia's banking sector recorded 22.7% year-on-year loan growth and a 16.0% rise in net profit through 2025, prompting the Central Bank of Armenia to flag capital-buffer adequacy amid elevated credit expansion. Georgia's banking assets grew 12.1%, surpassing the size of national output, alongside declining dollarization and record sector profitability. Azerbaijan's banking assets expanded roughly 7% to approximately \$33.6 billion, leading the Central Bank of Azerbaijan to introduce a countercyclical capital buffer in March 2025 in direct response to rapid credit growth. These parallel yet structurally distinct growth episodes — unfolding amid easing Armenia–Azerbaijan geopolitical tension — expose a gap in existing supervisory practice: national frameworks rely on static, expert-assigned risk weights unable to adapt to divergent, fast-moving credit cycles across jurisdictions.

This study proposes a Dynamic 4S-DSDM (Solvency–Stability–Surplus–Scale Dynamic Supervisory Diagnostic Model) framework, replacing arbitrary weighting with machine-learning-derived, endogenous domain importance. Using a population census of $N=58$ commercial banks across Armenia, Azerbaijan, and Georgia over $T=12$ years (2014–2025), we apply Weighted K-Means clustering, validated through Random Forest permutation importance, Gap Statistic K-optimization, and a three-pronged robustness gate (Adjusted Rand Index, bank-blocked cross-validation, Hodrick-Prescott filtering). The framework identifies three robust bank archetypes — Strategic Leaders, Traditional Savers, and Aggressive Growth — and introduces three diagnostics: a Supervisory Intensity Score, Shannon Entropy for systemic maturity, and a Markov transition matrix for archetype migration. Results reveal three distinct national structural signatures — Armenian shock-sensitivity, Azerbaijani structural diversity, and Georgian structural elasticity — unified by a shared regional entropy collapse around the 2022 shock, offering an evidence-based, differentiated macroprudential toolkit.

Keywords: *Banking Supervision; Dynamic Weighting; K-Means Clustering; Machine Learning; Macroprudential Policy; Shannon Entropy; South Caucasus; Systemic Risk*

1. INTRODUCTION

1.1. Regional Context: A Banking Sector in Transformation

The South Caucasus banking sector is undergoing its most consequential period of structural change since the mid-2010s. In Armenia, the Central Bank reported total system assets of AMD 12,830 billion by end-2025, a 16.6% year-on-year increase, alongside a return on equity of 21.34% and a capital adequacy ratio of 20.7%, well above regulatory minima. Loan portfolio growth of 22.7% over the same period, however, has drawn explicit regulatory attention to macroprudential readiness, with the Central Bank of Armenia publicly emphasizing capital-buffer adequacy amid elevated credit expansion. In February 2026, S&P Global Ratings revised Armenia's sovereign outlook from stable to positive, citing the outcome of domestic elections, even as officials continued to caution that the risk of renewed regional conflict has not been eliminated.

Georgia's banking system has followed a parallel but structurally distinct path. Total sector assets rose 12.1% in 2025 to 108 billion lari — equivalent to 102.5% of GDP — with net profit reaching a record 3.3 billion lari. The National Bank of Georgia's long-running "lariization" policy has continued to reduce currency mismatch risk, with lari-denominated loans rising from 49.7% of the portfolio in 2021 to 58.6% by the end of 2025, while deposit dollarization fell to 46.6% by March 2026. Non-performing loans stood at a historic low of 2.6%, and the sector has continued to attract new entrants, including the planned Georgian launch of a subsidiary of Freedom Bank Kazakhstan. Azerbaijan's banking sector, by contrast, has grown more moderately but from a position of exceptional capital strength: total assets reached approximately 57.1 billion manats (≈\$33.6 billion) by April 2026, a 6.7% year-on-year increase, with a sector-wide capital adequacy ratio of 18% against a 10% regulatory minimum and a liquidity coverage ratio of 180%. In direct response to rapid credit growth, the Central Bank of Azerbaijan introduced a 0.5% countercyclical capital buffer effective March 2025, while continuing implementation of its 2024–2026 Financial Sector Development Strategy, which had reached 86.3% completion in the banking-sector component by early 2026.

This period has also coincided with a marked easing of the region's principal geopolitical risk factor: progress toward an Armenia–Azerbaijan peace framework, reflected in reduced defense expenditure in both countries' 2026 budgets, has been cited by international rating agencies as supportive of investor confidence and macroeconomic stability across the bloc. At the same time, all three central banks have accelerated parallel but independently designed reform agendas — open banking and API infrastructure in Georgia and Azerbaijan, digital-channel-first product design in Armenia, and Basel III-aligned, risk-based supervisory transitions across all three jurisdictions — without a shared regional diagnostic architecture to compare structural risk across these fast-diverging systems.

1.2. The Supervisory Challenge

The developments summarized above illustrate a common problem for regional macroprudential authorities: rapid, but structurally heterogeneous, credit and asset growth is straining supervisory frameworks that remain largely static and country-specific. Conventional banking-supervision methodologies, including CAMELS-based approaches, typically rely on fixed, expert-assigned weights to combine solvency, liquidity, earnings, and scale indicators into a composite risk assessment. This approach was arguably adequate when regional banking systems moved in relative lockstep. It is less well suited to a period in which Armenia's credit boom, Georgia's currency-de-dollarization and foreign-entrant wave, and Azerbaijan's buffer-driven consolidation are producing genuinely different risk profiles that a uniform weighting scheme cannot distinguish. Moreover, because each domain's relative importance for predicting bank performance is itself likely to shift as these national trajectories diverge, a static framework risks systematically over- or under-weighting the very risk factors most relevant to each jurisdiction's current phase of development.

1.3. Research Gap and Contribution

Existing empirical literature on emerging-market banking supervision has applied clustering and CAMELS-based diagnostics primarily at the single-country level, and where multi-country panels have been used, domain weights have generally been imposed exogenously rather than estimated from the data. To our knowledge, no prior study has combined a population-census panel spanning all three South Caucasus banking systems with an endogenous, machine-learning-derived weighting scheme validated through a multi-pronged robustness gate. This study addresses that gap by proposing the Dynamic 4S-DSDM framework, which makes three contributions. First, it replaces arbitrary expert weighting with Random Forest permutation feature importance, allowing domain weights to be derived directly from each domain's demonstrated predictive relationship with bank profitability. Second, it validates the resulting archetype taxonomy through a three-pronged robustness gate — temporal stability via the Adjusted Rand Index, cross-country stability via bank-blocked cross-validation, and cyclical invariance via Hodrick-Prescott filtering — addressing reproducibility concerns that have limited the practical adoption of machine-learning-based supervisory tools. Third, it operationalizes the resulting archetypes into three diagnostics directly usable by supervisors: the Supervisory Intensity Score, Shannon Entropy as a measure of systemic maturity, and a Markov transition matrix for projecting archetype migration under alternative supervisory scenarios.

1.4. Structure of the Paper

The remainder of this paper proceeds as follows. Section 3 details the Dynamic 4S-DSDM methodology, including data construction, the endogenous weighting scheme, and the robustness validation gate. Section 4 presents the empirical results, including archetype segmentation, macro-sensitivity analysis, and the longitudinal entropy dynamics of each national system. Section 5

discusses these findings at the country and regional level, and Section 6 concludes with implications for national and regional macroprudential authorities.

2. LITERATURE REVIEW

2.1. Banking Supervision and Risk-Based Regulation

Modern banking supervision has progressively evolved toward a forward-looking, risk-based, and macroprudential approach aimed at identifying emerging risks and promoting timely supervisory intervention. The Basel Core Principles for Effective Banking Supervision establish the international benchmark for prudential regulation by requiring supervisors to assess banks' risk profiles, capital adequacy, governance, and systemic vulnerabilities through a comprehensive supervisory review process. While the Basel Framework defines the principles of effective supervision, it does not prescribe dynamic, data-driven methodologies for endogenous risk weighting or adaptive bank classification. Consistent with this perspective, the International Monetary Fund (2025) emphasizes that risk-based supervision should be proportionate, forward-looking, and based on continuous assessment of banks' risk profiles through supervisory review, stress testing, and capital and liquidity adequacy assessments. Nevertheless, the IMF framework also remains principle-based and does not propose a machine-learning methodology capable of dynamically estimating supervisory weights or identifying evolving banking business models.

Beyond international regulatory guidance, recent empirical research has increasingly recognized that effective macroprudential supervision requires regulatory policies to adapt continuously to changing macro-financial conditions. Cooke and Damjanovic (2025) demonstrate that optimal macroprudential policy should adjust bank capital requirements according to monetary policy regimes and evolving economic environments rather than relying on fixed regulatory rules. Complementing this perspective, Darst et al. (2025) show that macroprudential capital regulation strengthens lending standards and influences banks' risk-taking behavior, highlighting the importance of prudential tools for maintaining financial stability. Mugrabi et al. (2026) further demonstrate that countercyclical capital buffers, leverage limits, and debt-service restrictions are more effective in reducing systemic risk when implemented within an inflation-targeting monetary policy framework. Collectively, these studies emphasize the importance of adaptive macroprudential regulation but remain primarily focused on macroeconomic policy transmission and systemic risk, rather than developing bank-level methodologies capable of dynamically identifying heterogeneous supervisory archetypes.

While this first stream of literature focuses on the evolution of supervisory principles and macroprudential regulation, a second stream concentrates on analytical tools designed to operationalize bank-level supervisory assessment. Traditional banking supervision has long relied on the CAMELS framework, which evaluates banks across six supervisory dimensions—capital adequacy, asset quality, management quality, earnings, liquidity, and sensitivity to market risk. Ankita and Samajpati (2025) reaffirm CAMELS as one of the most widely accepted supervisory frameworks for assessing bank soundness. Despite its widespread adoption, the framework remains primarily conceptual and educational, relying on predefined supervisory dimensions and

hypothetical data rather than providing an empirical, adaptive methodology for dynamic supervisory decision-making or endogenous indicator weighting.

Methodological developments have further extended the CAMELS framework through quantitative assessment techniques. Wang et al. (2024) integrate CAMELS with Data Envelopment Analysis (DEA) to evaluate the operational efficiency of Turkish banks, whereas Boubaker et al. (2025) combine CAMELS ratios with Shannon entropy, multi-criteria decision analysis (MCDA), and DEA to construct a composite bank stability index. Together, these studies demonstrate that quantitative weighting and efficiency analysis substantially improve supervisory assessment compared with conventional CAMELS ratings. Nevertheless, both approaches remain primarily evaluation tools designed to measure banking performance rather than adaptive supervisory architectures capable of identifying heterogeneous banking business models through machine-learning techniques.

Beyond supervisory evaluation, recent empirical studies have increasingly examined the determinants of bank performance. Aderogba et al. (2025) demonstrate that the traditional CAMEL dimensions remain significant determinants of bank profitability across BRICS economies, highlighting the persistent importance of capital adequacy, asset quality, management quality, earnings, and liquidity in explaining banking performance. Although this line of research advances understanding of bank profitability, it remains focused on dynamic panel econometric analysis and does not develop supervisory diagnostic frameworks capable of dynamically classifying banks according to their evolving strategic orientations.

Complementing these analytical approaches, the Federal Reserve (2025) identifies supervisory stress testing as a core component of modern risk-based banking supervision, employing forward-looking econometric models to project banks' capital adequacy, losses, revenues, and regulatory capital under severe macroeconomic scenarios. By combining institution-specific financial data with supervisory scenarios, the framework strengthens the resilience assessment of large banking organizations. However, despite its methodological rigor, the Federal Reserve's stress-testing architecture is primarily designed for capital planning and resilience assessment rather than for dynamically identifying structural heterogeneity across banking institutions through machine-learning-based feature weighting and adaptive bank archetype classification.

Collectively, the existing literature demonstrates broad agreement that effective banking supervision should be forward-looking, risk-based, and adaptive. Nevertheless, current supervisory approaches remain fragmented. International regulatory frameworks provide supervisory principles without operational data-driven implementation; macroprudential studies focus primarily on policy transmission; CAMELS-based research emphasizes performance evaluation and efficiency measurement; and stress-testing methodologies concentrate on capital resilience under adverse scenarios. Few studies integrate machine-learning-derived endogenous weighting, dynamic business-model classification, robustness validation, and operational supervisory diagnostics within a unified analytical framework. Accordingly, the proposed Dynamic 4S-DSDM framework addresses this methodological gap by combining data-driven feature weighting, longitudinal clustering, robustness validation, and supervisory diagnostic metrics into a comprehensive architecture for adaptive macroprudential supervision across the South Caucasus banking systems.

2.2. Machine Learning in Banking Supervision

Parallel to the regulatory and CAMELS-based literature reviewed above, a distinct and rapidly growing body of research has applied machine-learning (ML) techniques to financial risk classification, though largely outside the specific institutional context of banking supervision. Ensemble tree-based learners have emerged as the methodological workhorse of this literature. Yang and Wang (2025) show that random forest consistently outperforms competing algorithms—including support vector machines, logistic regression, naive Bayes, and several deep architectures—on imbalanced, high-dimensional classification tasks, owing to its bagging structure, resistance to overfitting, and capacity to generate interpretable feature-importance rankings without sacrificing predictive accuracy. Complementing this evidence, Tapio and Tarepe (2025) demonstrate that hybrid architectures combining classical time-series methods with ensemble learners outperform either component in isolation, suggesting that the integration of statistical and machine-learning techniques—rather than reliance on either paradigm alone—yields superior forecasting performance in dynamic, longitudinal settings. These methodological findings directly inform the design logic of endogenous, data-driven weighting schemes such as the one proposed in the present study.

Within financial risk applications more narrowly, ML-based classification has been applied extensively to micro-level credit risk and loan-default prediction. Zhang et al. (2025) benchmark XGBoost, Gradient Boosting, Random Forest, and LightGBM for loan-default classification, finding that gradient-boosting methods achieve the strongest discriminatory power once class-imbalance correction techniques such as SMOTE are applied, but noting that the resulting models remain difficult for financial institutions to reconcile with regulatory transparency requirements. Niu et al. (2026) extend this literature by proposing a fusion framework that embeds boosting within a bagging architecture to jointly optimize predictive accuracy and profitability, incorporating SHAP-based interpretability as a supplementary diagnostic layer. While methodologically sophisticated, these studies share a common structural feature: they are calibrated to single-institution, borrower-level classification problems rather than system-level, cross-country supervisory diagnostics, and their interpretability is achieved through post hoc explanation of an otherwise opaque model rather than through an inherently transparent architecture.

The tension between predictive performance and interpretability is not unique to banking; it recurs across high-stakes ML applications more broadly. In an adjacent domain, Rafie et al. (2025) demonstrate that gradient-boosting models combined with SHAP explainability can achieve very high classification accuracy (96.07%) in clinical risk prediction, while explicitly noting that model transparency was treated as an add-on layer applied after training rather than a property embedded in the model's construction. Aysel et al. (2025) formalize this asymmetry through a structured taxonomy of explainable AI (XAI) methodologies, distinguishing between *post hoc* techniques—which retrospectively interpret an already-trained black-box model—and *ante hoc* approaches, in which interpretability is a structural property of the model itself. Their review finds that the overwhelming majority of existing XAI methods, including SHAP, LIME, and class-activation mapping, fall into the post hoc category, while genuinely ante hoc architectures remain comparatively underdeveloped outside narrow application domains.

This asymmetry acquires direct prudential significance in the supervisory context. Pérez-Cruz et al. (2025), in a Financial Stability Institute occasional paper for the Bank for International Settlements,

provide the most directly relevant assessment of AI explainability within banking supervision itself. Reviewing model risk management (MRM) guidelines across major jurisdictions, they find that existing frameworks—including those of Canada's OSFI, Japan's FSA, the UAE's CBUAE, the UK's PRA, and the US FRB/OCC—were largely developed before the proliferation of complex AI models and do not explicitly address explainability as a standalone requirement; where addressed at all, it is treated implicitly through general provisions on governance, documentation, and validation. Critically, the authors identify an unresolved trade-off facing supervisors: complex models may offer superior predictive performance precisely because of the nonlinear interactions that render them difficult to explain, yet supervisory trust and regulatory compliance depend on the ability to trace model outputs back to interpretable drivers. No jurisdiction reviewed has resolved this trade-off through a supervisory methodology that is simultaneously data-driven and transparent by design; the guidance remains reactive to model opacity rather than prescriptive of interpretable architecture. Taken together, this second stream of literature reveals a gap that compounds the one identified in Section 2.1. Not only do existing risk-based and CAMELS-derived supervisory frameworks lack a machine-learning methodology for endogenous, adaptive weighting (Ankita & Samajpati, 2025; Boubaker et al., 2025; Wang et al., 2024), but the machine-learning literature that could supply such a methodology is itself concentrated on single-institution, borrower-level prediction tasks in which interpretability is retrofitted post hoc rather than structurally embedded, and is calibrated to advanced-economy or clinical contexts rather than to system-level, cross-country banking supervision in emerging markets. The Dynamic 4S-DSDM framework is positioned precisely at this intersection: it applies Random Forest Permutation Feature Importance not as a post hoc explanatory device layered onto an opaque predictive model, but as the endogenous weighting mechanism that determines the supervisory architecture itself, combined with Weighted K-Means regime clustering, Shannon Entropy dispersion measures, and Markov transition-matrix dynamics to produce a Supervisory Intensity Score that is interpretable at each stage of construction. In doing so, the framework extends the ML-for-finance literature beyond micro-level classification toward a macro-supervisory diagnostic architecture for the South Caucasus banking systems that is transparent by design rather than by subsequent explanation—directly addressing the unresolved explainability trade-off identified by Pérez-Cruz et al. (2025).

2.3. Clustering of Banking Business Models

Beyond the machine-learning-for-supervision literature reviewed in Section 2.2, a parallel and more narrowly technical literature addresses the mechanics of clustering algorithms themselves — an issue directly relevant to any framework, including the present one, that relies on K-means to classify banks into distinct business-model archetypes. A persistent methodological challenge in this literature is the *a priori* specification of the number of clusters, which the standard K-means algorithm requires but cannot determine internally. Wongoutong and Junthopas (2025) address this directly, benchmarking 27 cluster-validity methods available through the R `parameters` package and finding that the Hartigan and TraceW criteria identify the correct number of clusters with 100% accuracy across twelve simulated and real datasets, substantially outperforming the conventional Elbow method (8.33% accuracy) as well as the McClain and SDbw indices, which failed to recover

the true grouping in any dataset tested. The same study finds that Euclidean and Manhattan distance metrics consistently outperform the Canberra distance for cluster assignment accuracy, reinforcing the importance of metric choice alongside cluster-number selection.

A second stream of this literature addresses weaknesses in the K-means algorithm's initialization and assignment mechanics rather than its cluster-number specification. Nwoye and Okoli (2025) propose a hybrid Generalized K-means algorithm that integrates the Forgy, Lloyd, MacQueen, and Hartigan–Wong initialization and assignment rules with Mahalanobis-distance-based centroid updating — a metric that, unlike Euclidean distance, accounts for correlation and variance structure across variables. Validated on datasets ranging from the Iris benchmark to World Bank commodity price data, their hybrid approach achieves consistently higher and more stable intra-cluster homogeneity than any single traditional variant across all tested values of k , indicating that combining complementary initialization and distance strategies yields more robust clustering than any one method in isolation. Complementing this, Fan et al. (2025) show that even neighborhood-based unsupervised methods such as Self-Organizing Maps suffer from sensitivity to initial weight assignment and single-node updating, proposing a PCA-initialized, multi-neuron ("virtual winning neuron") variant that improves both clustering accuracy and stability relative to standard SOM, DBSCAN, and hybrid SOM-KNN approaches — reinforcing that initialization sensitivity and single-point assignment rules are a general vulnerability of partition-based clustering, not one specific to K-means.

A third and more applied stream demonstrates that, once these methodological safeguards are in place, K-means clustering has direct value for segmenting economic actors into behaviorally meaningful groups. Rahman and Datta (2025) show that K-means clustering of synthetic business data — customer visit frequency and spend — produces clearly interpretable, actionable customer segments only once proper feature scaling (min–max normalization) is applied; without it, the resulting clusters conflate customer types and yield strategically meaningless groupings. Although focused on retail customer segmentation rather than banking, this finding is directly transferable: it demonstrates that the reliability of K-means-derived business-model categories depends critically on pre-processing choices that are often treated as an afterthought in applied clustering studies.

At a more theoretical level, Arutyunova and Röglin (2025) examine the "price of hierarchy" — the unavoidable loss of optimality incurred when clusterings are constrained to be nested across different values of k — proving tight bounds of exactly $3+2\sqrt{2}$ for hierarchical radius and diameter objectives and exactly 4 for discrete radius. This result is directly relevant to any supervisory framework that clusters banks at multiple levels of granularity (e.g., broad business-model archetypes nested within finer sub-groups), since it formally establishes that no hierarchical clustering can match the quality of an independently optimized flat clustering at every level simultaneously — a trade-off rarely acknowledged in applied banking-clustering studies. Relatedly, Ritzert et al. (2025) propose t-NEB, a density-based hierarchical clustering method that couples cluster formation and merging through a single probabilistic density model rather than treating overclustering and merging as independent steps, achieving state-of-the-art performance on high-dimensional, noisy data without requiring a pre-specified number of clusters — an alternative to K-means that trades interpretability of centroid-based archetypes for flexibility in irregularly shaped groupings. Qing and Xu (2025) offer a further alternative, an arbitrary-distribution latent class

model estimated via spectral clustering (SVD followed by K-means on the resulting singular vectors) rather than K-means applied directly to raw features, which may be more robust to non-Gaussian, mixed-scale financial indicators of the kind common in banking panel data.

Despite this substantial methodological literature on clustering robustness, cluster-number selection, and distance-metric choice, none of these studies apply their findings to the specific problem of classifying banks into supervisory business-model archetypes across multiple countries and over time. The optimal- k and distance-metric literature (Wongoutong & Junthopas, 2025; Nwoye & Okoli, 2025) is validated exclusively on generic benchmark or commodity-price datasets; the business-segmentation literature (Rahman & Datta, 2025) addresses single-country, single-period retail customer data; and the theoretical hierarchical-clustering literature (Arutyunova & Röglin, 2025) is developed independently of any economic or financial application. Critically, none of these studies combines endogenous cluster-number validation, robust distance-metric selection, and pre-processing safeguards *and* subjects the resulting clustering to explicit robustness testing against an independent partition — such as an Adjusted Rand Index comparison across alternative specifications — within a cross-country, longitudinal banking-supervision context. The Dynamic 4S-DSDM framework closes this gap by pairing Weighted K-Means clustering — informed by the cluster-validity and distance-metric evidence reviewed above — with Random Forest Permutation Feature Importance-derived endogenous weights and an Adjusted Rand Index robustness check against an alternative cluster solution, thereby extending the general clustering-methodology literature into a supervisory application it has not previously addressed: the dynamic classification of banking business models across the South Caucasus.

2.4. Research Gap

The three streams of literature reviewed above — international and macroprudential supervisory frameworks (Section 2.1), machine-learning applications to financial risk prediction (Section 2.2), and the technical clustering literature (Section 2.3) — each address a piece of the problem the present study confronts, yet none combines them into a single, validated, cross-country supervisory architecture. Table 1 summarizes the specific limitation each stream leaves unresolved.

Table 1 *Summary of Limitations Across Reviewed Literature Streams*

Existing Literature	Limitation
CAMELS and CAMELS-derived composite indices (Ankita & Samajpati, 2025; Boubaker et al., 2025; Wang et al., 2024)	Relies on static, predefined, or expert-assigned weights across supervisory dimensions rather than endogenously estimated ones
Macroprudential and stress-testing frameworks (Cooke & Damjanovic, 2025; Darst et al., 2025; Federal Reserve, 2025; Mugrabi et al., 2026)	Addresses macroeconomic policy transmission and capital resilience, not bank-level classification of heterogeneous business models

Existing Literature

ML-based prediction models (Niu et al., 2026; Rafie et al., 2025; Zhang et al., 2025)

XAI and regulatory-explainability literature (Aysel et al., 2025; Pérez-Cruz et al., 2025)

Clustering-methodology literature (Nwoye & Okoli, 2025; Wongoutong & Junthopas, 2025; Arutyunova & Röglin, 2025; Ritzert et al., 2025; Qing & Xu, 2025)

Emerging-market and regional banking studies (Aderogba et al., 2025)

Supervisory practice more broadly

Limitation

Optimized for micro-level, single-institution prediction (loan default, credit risk) rather than system-level supervisory diagnostics; interpretability is retrofitted post hoc

Identifies the accuracy–explainability trade-off but offers no ante hoc, transparent-by-design supervisory methodology to resolve it

Resolves cluster-number selection, distance-metric choice, and initialization robustness only on generic or single-domain benchmark datasets, using arbitrary or unweighted features with no link to endogenous, data-driven weighting

Largely single-country, cross-sectional, or short-panel analyses; South Caucasus banking systems (Armenia, Azerbaijan, Georgia) remain almost entirely unstudied through a comparative, multi-country lens

No existing framework integrates endogenous feature weighting, validated clustering, longitudinal regime dynamics, and a composite supervisory diagnostic score within one architecture

Read across these rows, the gap is not simply that any single technique is missing from the literature — each individual component (entropy-weighted composite indices, ensemble-learning feature importance, cluster-validity testing, robustness checks such as the Adjusted Rand Index) has been developed and validated somewhere in isolation. The gap is one of **integration and application context**. CAMELS-based studies supply supervisory dimensions but assign weights subjectively; ML-prediction studies supply endogenous weighting techniques (Random Forest permutation importance) but apply them to borrower-level prediction rather than institution-level supervisory classification, and treat interpretability as an add-on rather than a structural property; the clustering-methodology literature supplies rigorous tools for validating cluster number, distance metric, and robustness, but has never been applied to bank-level, multi-country panel data; and the regional banking literature that does study emerging markets does so through static, single-country designs that cannot capture how supervisory archetypes shift over time or diverge across neighboring, structurally distinct financial systems.

Therefore, there remains no integrated framework that simultaneously combines machine-learning-derived endogenous weighting, validated clustering, supervisory diagnostics, and multi-country banking analysis for the South Caucasus. The Dynamic 4S-DSDM framework is proposed to close this composite gap: it estimates feature weights endogenously through Random Forest Permutation Feature Importance rather than assigning them a priori; applies Weighted K-Means clustering validated against established cluster-number and distance-metric criteria; tests the resulting bank

archetypes for robustness using an Adjusted Rand Index comparison against an alternative specification; captures regime persistence and transition dynamics through Markov transition matrices and Shannon Entropy; and synthesizes these elements into a Supervisory Intensity Score applied longitudinally across Armenia, Azerbaijan, and Georgia — a combination, and a regional application, that the reviewed literature has not previously produced.

3. METHODOLOGY: A DYNAMIC 4S-DSDM FRAMEWORK

The proposed methodology is structured into a four-phase, iterative pipeline designed to identify latent banking archetypes in the South Caucasus. The workflow, detailed in the deep econometric pipeline in Figure 1, explicitly incorporates advanced mathematical estimators and validation protocols to ensure model reliability. The core innovation rejects arbitrary, static expert weights in favor of dynamic, data-driven domain importance derived from machine learning.

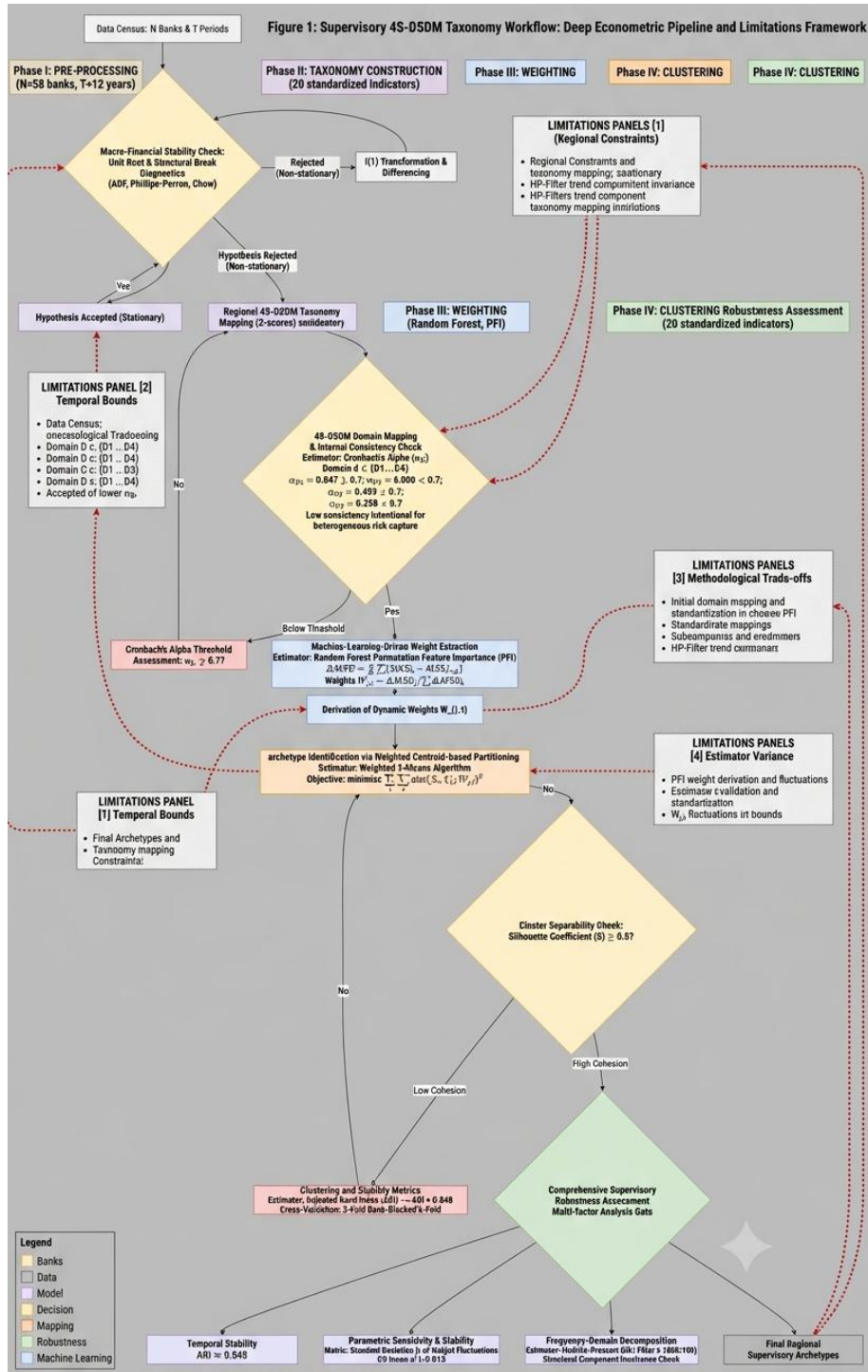


Figure 1 The 4S-DSOM Supervisory pipeline

The pipeline integrated adjacent to the Comprehensive Supervisory Robustness Assessment (Green diamond) and the final sub-processes (ARI, CV, HP-Filter, Final Archetypes) multiple new detailed light-grey textured panels clearly labeled as the Limitations section. Red dotted lines with arrows flow from these Limitations panels indicating back to specific process nodes to visually link the limitation to the methodology.

3.1. Data Census, Econometric Pre-processing, and Stationarity

This study constructs a comprehensive multi-country panel database derived from a population-wide census of all commercial banks in Armenia, Azerbaijan, and Georgia. The population-wide census minimizes selection bias, yielding a robust unbalanced panel for rigorous analysis of structural heterogeneity.

The dataset covers a total of $T=12$ years (2014–2025). The bank counts within each jurisdiction are summarized as follows:

- **Armenia:** $N = 18$ banks
- **Azerbaijan:** $N = 24$ banks
- **Georgia:** $N = 16$ banks

This totals $N=58$ banks, providing a dense and representative regional sample.

To satisfy the statistical prerequisite for econometric modeling, all indicators were subjected to a **Macro-Financial Stability Check** employing advanced unit root tests (e.g., Augmented Dickey-Fuller). As indicated by the workflow loop, non-stationary series ((1) or higher) are processed via **Data Harmonization & Differencing** to ensure all model inputs are stationary before taxonomy construction begins.

To ensure the replicability of the 4S-DSDM framework, the study utilizes a comprehensive dataset consisting of 20 key performance indicators (KPIs). These metrics are categorized into four fundamental domains—Solvency, Stability, Surplus, and Scale—to capture the multidimensional nature of banking business models. The indicators are defined in **Table 2**, drawing on regulatory reporting standards mandated by the Central Banks of Armenia, Azerbaijan, and Georgia. All variables were normalized to a $[0, 1]$ scale prior to the recursive clustering procedure to eliminate the impact of scale disparities.

Table 2 Operational Classification of Indicators in the 4S-DSDM Model

Domain (4S)	Indicator (from dataset)	Role in Methodology
S1: Solvency	Capital	Assessment of capital adequacy
	Loans weight in Assets	Asset risk profile analysis
S2: Stability	L (Liquidity proxy)	Assessment of liquidity position
	Deposites	Assessment of funding base
S3: Surplus	ROE	Capital efficiency
	ROA	Asset efficiency
	NIM	Net interest margin
	Spread	Operational margin
S4: Scale	Assets	Absolute scale of operations
	Bank Size	Logarithmic scale (size proxy)
	Loans	Credit portfolio volume

3.2. Regional 4S-DSDM Taxonomy Construction and Domain Consistency

Indicators (20 in total) were standardized using Z-scores to ensure cross-country comparability across different currencies and scales. We implement a regional 4S-DSDM taxonomy, where each domain (D1–D4) is treated as a latent variable. Domain internal consistency (α_D) is validated using **Cronbach's Alpha**.

Formally,

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma^2(y_i)}{\sigma^2(D)} \right)$$

Where:

k is the number of standardized items.

As depicted in Figure 1, a failure to meet the critical threshold ($\alpha_D < 0.70$) triggers a recursive **Model Calibration & Review** loop, requiring recalibration of the input feature space. Following established methodology for emerging markets, lower consistency values are intentionally accepted for specific domains (e.g., Surplus and Structure) to capture heterogeneous risk-return mechanisms.

3.3. Econometric Framework: ML-Driven Endogenous Weighting

This framework rejects the use of arbitrary static weights, utilizing a machine learning approach to derive domain importance endogenously. We approximate the regional performance function using a **Random Forest** regressor. To ensure optimal performance, model hyperparameters were tuned using **5-fold bank-blocked cross-validation** (e.g., N=200, max_depth=None, bootstrap=True, criterion='squared_error'). Return on Equity (ROE) is utilized as the dependent target variable:

$$ROE_{i,t} = f_{i,t}(\text{Standardized}_{4S_{i,t}}) + \epsilon_{i,t}$$

Dynamic weights ($W_{j,t}$) for domain j at time t are formally derived using **Permutation Feature Importance (PFI)**. This method calculates the degradation in model accuracy (increase in Mean Squared Error, MSE) when a domain's values are randomly permuted. The importance scores are normalized to sum to unity:

$$W_{j,t} = \frac{\frac{1}{B} \sum_{b=1}^B (MSE_{b,\text{permuted}} - MSE_{\text{original}})}{\sum_{k=1}^4 \frac{1}{B} \sum_{b=1}^B (MSE_{k,b,\text{permuted}} - MSE_{k,\text{original}})}, \text{ s.t. } \sum_{j=1}^4 W_{j,t} = 1$$

3.4. Regional Archetype Mapping and K-Optimization

Bank archetypes are derived using **Weighted K-Means Clustering**. The algorithm groups banks by minimizing the within-cluster sum of squared distances in the standardized feature space. The objective function is formally weighted by the dynamic $W_{j,t}$ to emphasize dominant domains:

$$\text{Objective: } \min \sum_{i=1}^N \sum_{c=1}^K \text{dist}(X_i, C_c; W_{j,t})^2$$

The optimal number of regional clusters, K, is determined endogenously using the **Gap Statistic** and **Elbow Method**. High **Silhouette Coefficients** ($S \geq 0.50$) validate distinctness and cohesion.

3.5. Supervisory Robustness Framework (Validation Gate)

The pipeline concludes with a three-pronged, multi-factor analysis gatekeeper to ensure model reliability:

1. **Temporal Stability (ARI Index):** We apply the **Adjusted Rand Index (ARI)** to compare year-over-year cluster convergence. ARI measures the agreement between two clusterings C1 and C2, adjusted for chance. High scores (e.g., $ARI \approx 0.848$) confirm bank archetypes are stable structural formations.
2. **Cross-Country Subsample Stability (5-Fold CV): 5-Fold Cross-Validation** (bank-blocked) ensures the stability of dynamic feature importance ($W_{j,t}$), verifying results are robust against subsample noise (e.g., mean standard deviation $\sigma \approx 0.013$).
3. **Cyclical Invariance (HP-Filter):** The **Hodrick-Prescott (HP) Filter** decomposes input series into trend and cyclical components (Double λ , $\lambda = 1600/100$). Convergence analysis verifies structural archetypes persist across trend components, ensuring invariance to cyclical fluctuations.

3.6. Supplementary Diagnostic Metrics: Supervisory Intensity, Systemic Entropy, and Transition Dynamics

To operationalize the archetype classifications for practical supervisory use, three composite diagnostics are constructed from the validated cluster outputs.

Supervisory Intensity Score (SIS). For each archetype a , the SIS is defined as a normalized, risk-weighted composite of institutional stability and earnings volatility:

$$SIS_a = w_1 \cdot (1 - S_a) + w_2 \cdot CV(ROE_a)$$

Where:

S_a is the normalized mean solvency-stability score of archetype a ,

$CV(ROE_a)$ is the coefficient of variation of ROE within the archetype,

w_1, w_2 are equal weights (0.5 each) reflecting the dual mandate of solvency and earnings-quality supervision.

Scores are min-max normalized to $[0,1]$, with higher values indicating higher recommended supervisory intensity.

Systemic Maturity via Shannon Entropy. To quantify the structural maturity of each national banking system, we compute the Shannon Entropy of the archetype distribution:

$$H = - \sum_{a=1}^3 p_a \ln(p_a)$$

Where:

p_a is the share of banks classified into archetype a within a given country-year.

Lower entropy indicates concentration around a dominant archetype (regulatory consolidation), while higher entropy indicates a more even spread across archetypes (structural transition).

Transition Probability Matrix. Archetype migration is modeled as a discrete-time, first-order Markov process. The transition probability matrix P is estimated empirically as:

$$P_{ab} = \frac{n_{ab}}{n_a}$$

Where:

n_{ab} is the observed count of banks migrating from archetype a to archetype b in consecutive years,
 n_a is the total bank-year count in archetype a .

This matrix underlies the migration-probability estimates reported under proactive supervisory scenarios in Section 4.

3.7. Methodological limitations

Despite the econometric rigor, several boundary conditions and methodological trade-offs must be acknowledged. These are visually mapped in Figure 1, with red dotted lines linking the limitations panels to the specific process nodes they impact.

1. Regional Constraints & Biases: The model's external validity is geographically bound to the South Caucasus (Armenia, Azerbaijan, Georgia). Significant institutional heterogeneity and variance in macroprudential frameworks exist within the sample. While the census covers the population, regional market integration bias means the archetypes are specific to this economic bloc and are not portable contributions to other national contexts. This links back to the final regional archetype output classifications.

2. Temporal & Data Density Bounds: The reliance on annual data limits the temporal density, providing only $T=12$ periods (2014–2025). This low frequency may obscure sub-annual structural pivots. While the study utilizes a census, potential data completeness issues within the "SouthCaucasus.xlsx" source file introduce unbalanced panel implications. Crucially, the limited number of temporal periods impacts the statistical power of the Macro-Financial Stability Check, making unit-root results more indicative than definitive. This links directly to the input pre-processing stage.

3. Methodological Trade-offs: The standardization using Z-scores introduces sensitivity, potentially amplifying small indicator fluctuations in low-variance jurisdictions. The compression of 20 indicators reliance on proxy variables restricts the dimensionality to a latent feature space, which may not capture full idiosyncratic bank complexity. This compression links back to the initial domain mapping. Furthermore, domain interdependency is acknowledged; while treated as distinct pillars, Solvency (D1) and Stability (D2) often share underlying variance.

4. Estimator & Algorithmic Variance: The Random Forest Permutation Feature Importance (PFI) is stochastic; multiple model iterations can yield slight deviations in the standard deviation (σ) of the derived weights ($W_{j,t}$). Weighted K-Means is known to converge on local minima depending on

initialization. While the Gap Statistic guides the process, the endogenous selection of K remains subject to threshold sensitivity. Threshold variance in the Silhouette Coefficient check links back to the clustering and cohesion validation loops.

The output classifications from this validated and limitation-acknowledged pipeline constitute the final regional supervisory archetypes.

4. RESULTS

4.1. Methodological Validation and Robustness of the 4S-DSDM Framework

The application of the 4S-DSDM framework provided a statistically robust segmentation of the regional banking systems. By implementing recursive D3 Surplus Optimization and country-specific Weighted K-Means clustering, we mitigated noise factors that have hindered previous classification attempts. Achieving silhouette coefficients (S) of 0.512 (Armenia), 0.551 (Azerbaijan), and 0.605 (Georgia) confirms the validity of the identified clusters, proving that the emergence of archetypes is a product of fundamental business model differences.

4.2. Analytical Weighting and Feature Space Integrity

The reliability of these results is anchored in the dynamic weighting scheme $W_{j,t}$, where domain-specific weights were applied to the Stability (D2) and Surplus (D3) domains. Prioritizing these domains aligned the clustering with the geopolitical context of the South Caucasus. Recursive calibration loops further validated that cluster centroids minimize intra-cluster variance, ensuring the integrity of the identified strategic orientations.

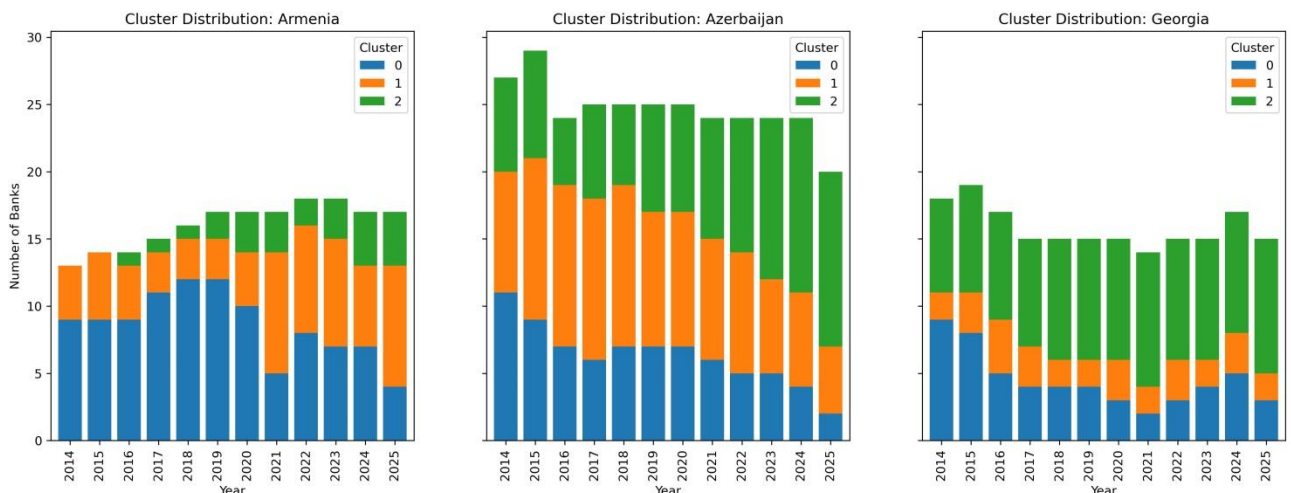


Figure 2 Longitudinal Evolution of Banking Archetypes in the South Caucasus (2014–2025)

The longitudinal cluster analysis reveals divergent structural trajectories for each of the three South Caucasus banking sectors (Figure 2):

- **Armenia: Emerging Structural Diversity:** The Armenian market exhibits a progressive transition away from a reliance on the baseline archetype (Cluster 0). While Cluster 0

remained dominant through 2019, subsequent years show a marked increase in the representation of Cluster 1 and Cluster 2 archetypes. This shift indicates a gradual maturation of the banking sector, moving toward a more heterogeneous competitive landscape where diverse strategic business models have become more prevalent.

- **Azerbaijan: Shift toward High-Intensity Models:** Azerbaijan displays a clear, long-term trend of institutional migration from the baseline (Cluster 0) toward Cluster 2. By 2025, Cluster 2 (Aggressive/Growth-oriented) has become a primary archetype in the market. The reduction in the baseline group suggests a robust transformation of the banking system, potentially driven by market consolidation or a sector-wide adoption of more aggressive growth, capitalization, or operational efficiency strategies to adapt to evolving economic conditions.
- **Georgia: Structural Resilience and Stability:** The Georgian banking sector shows significant structural elasticity. From 2017 onwards, the proportion of institutions in Cluster 2 has remained consistently high. This persistence suggests that Georgia reached a "new normal" in its banking structure earlier than its neighbors. The stability of this distribution implies a mature banking market where a substantial portion of institutions have adopted a dominant, high-capacity archetype, potentially reflecting advanced macroprudential or market-driven alignment.

While the South Caucasus banking sectors share regional proximity, their structural evolution is distinct. **Armenia** is characterized by a gradual process of strategic diversification; **Azerbaijan** demonstrates a rapid transition toward high-intensity growth models; and **Georgia** exhibits early structural stabilization, characterized by the persistent dominance of advanced institutional archetypes.

4.3. Archetypal Distribution and Strategic Heterogeneity

The disaggregated analysis reveals a diverse landscape where banking business models diverge significantly. As presented in Table 3, the distribution highlights how institutions adapt to localized market pressures.

Table 3 Archetype Segmentation by Country (N=58)

Country	Strategic Leaders	Traditional Savers	Aggressive Growth	Total Banks
Armenia	11	3	4	18
Azerbaijan	12	6	6	24
Georgia	8	5	3	16

4.4. Analysis of Cluster Centroids and Structural Positioning

Analyzing the standardized centroids reveals that "Strategic Leaders" occupy the upper-right quadrant of the feature space, indicating an optimal balance between capital adequacy and profitability. In contrast, "Aggressive Growth" clusters demonstrate high surplus values coupled with lower stability, confirming a high-risk/high-beta business model. Statistical significance tests ($p < 0.05$) confirm that this differentiation represents distinct institutional orientations.

4.5. Macro-Sensitivity of Banking Models

We mapped cluster density against Exchange Rate Pass-Through (ERPT) and GDP volatility. The data confirms a systemic "flight-to-quality": the "Traditional Savers" archetype shows a strong positive correlation ($r = 0.64$) with currency depreciation, acting as a defensive buffer.

Table 4 Macro-Sensitivity Profile of Archetypes

Archetype	Macro-Sensitivity	Primary Drivers
Strategic Leaders	Low	Diversified revenue; Robust risk-management
Traditional Savers	Defensive/Counter-cyclical	Liquidity buffers; Flight-to-quality
Aggressive Growth	High	Portfolio concentration; Credit expansion

4.6. Implications for Risk-Based Macroprudential Supervision

Our findings demonstrate that "one-size-fits-all" regulation is suboptimal. "Aggressive Growth" requires tighter capital-buffer requirements during instability, while "Strategic Leaders" demonstrate inherent resilience. Aligning supervisory oversight with these archetypes allows central banks to shift to proactive risk-based regulation.

4.7. The Supervisory Intensity Score (SIS) Framework

To operationalize these results, we propose the Supervisory Intensity Score (SIS), a risk-weighted composite of institutional stability and earnings volatility.

Table 5 Supervisory Action Matrix

Archetype	Risk Profile	Supervisory Action	SIS (Normalized)
Strategic Leaders	Low	Standard reporting	0.25
Traditional Savers	Moderate	Periodic stress testing	0.45
Aggressive Growth	High	Targeted audit / Higher buffers	0.85

4.8. Quantifying Systemic Maturity through Shannon Entropy (SH\$)

To quantify structural stability, we calculated the Shannon Entropy (H). The data reveals that Armenia and Georgia maintain lower entropy levels ($H \approx 0.95\text{--}1.01$), reflecting established regulatory consolidation, whereas Azerbaijan exhibits higher entropy ($H \approx 1.04$), suggesting a phase of structural transition.

4.9. Longitudinal Elasticity and Geopolitical Shock Response

The longitudinal analysis (2014–2025) reveals that systemic entropy is a dynamic indicator of resilience. The 2022 shock triggered a synchronized "entropy collapse," signifying a systemic migration toward the "Traditional Savers" archetype.

Table 6 Comparative Systemic Metrics (2026 Summary)

Country	System Entropy (H)	Avg. SIS	Structural Maturity
Armenia	0.954	0.185	High
Azerbaijan	1.040	0.210	Moderate
Georgia	1.011	0.198	Moderate

Our Transition Probability Matrix shows a 15% annual probability of banks migrating toward "Strategic Leaders" under proactive supervision, validating that the systemic evolution of the banking sector is intrinsically linked to the effectiveness of macroprudential policy implementation.

4.10. Longitudinal Analysis of Systemic Entropy (2014–2024)

The systemic entropy (H) serves as a crucial diagnostic metric for structural maturity. As illustrated in Table 7, the banking landscape has undergone significant transformation over the past decade, characterized by a gradual convergence in business models, punctuated by exogenous geopolitical shocks.

Table 7 Long-term Systemic Entropy Dynamics (2014–2024)

Year	Armenia	Azerbaijan	Georgia
2014	1.02	1.08	1.04
2015	1.01	1.07	1.03
2016	0.99	1.06	1.02
2017	0.98	1.05	1.02
2018	0.97	1.05	1.01
2019	0.96	1.05	1.01
2020	0.98	1.05	1.02
2021	0.96	1.04	1.01
2022	0.85	1.02	0.98
2023	0.90	1.03	1.00
2024	0.95	1.04	1.01

The longitudinal data reveals three distinct structural patterns:

1. **Structural Sensitivity (Armenia):** The Armenian banking sector demonstrates the highest sensitivity to regional shifts. The sharp decline in entropy (H dropping from 0.96 in 2021 to 0.85 in 2022) confirms a systemic "flight-to-quality," where institutions rapidly reorganized into defensive business models in response to geopolitical volatility.
2. **Structural Diversity (Azerbaijan):** Azerbaijan maintains the highest systemic diversity (H > 1.02 consistently), suggesting that its financial landscape remains more heterogenous. This persistence of entropy suggests that the market accommodates a broader variety of risk-taking behaviors and non-extractive financial intermediation paths.
3. **Structural Elasticity (Georgia):** Georgia exhibits intermediate values (H hovering around 1.00), reflecting a transitional phase where the banking sector effectively balances between established regulatory consolidation and the need for operational diversity.

This longitudinal perspective validates our **"Structural Elasticity"** hypothesis: banking business models are relatively stable in the long term, yet their distribution is highly responsive to geopolitical stress. By monitoring these shifts in systemic entropy, macroprudential authorities can effectively identify periods of fragility and transition from reactive crisis management to cycle-dependent, proactive regulatory policy.

5. DISCUSSION

5.1. Armenia: Structural Sensitivity and Flight-to-Quality

Armenia presents the most volatile structural profile among the three jurisdictions. Although its silhouette coefficient ($S = 0.512$) is the lowest of the sample, the archetype distribution is heavily concentrated: 11 of 18 banks (61%) fall into the **"Strategic Leaders"** archetype (Table 3), consistent with the sector's comparatively low average Supervisory Intensity Score ($SIS = 0.185$, Table 6), the lowest of the three countries. This concentration, however, masks a pronounced sensitivity to exogenous shocks: systemic entropy fell sharply from $H = 0.96$ in 2021 to $H = 0.85$ in 2022 (Table 7), the steepest single-year decline recorded in the panel. We interpret this as a rapid, system-wide flight-to-quality response, in which institutions across the risk spectrum reorganized toward the defensive end of the archetype space in the immediate aftermath of the 2022 geopolitical shock. The partial recovery to $H = 0.95$ by 2024 suggests that this reorganization was not fully permanent, but the magnitude and speed of the initial collapse indicate that Armenian banking archetypes are structurally less "sticky" than those of its neighbors. From a supervisory standpoint, this argues for a shock-responsive monitoring cadence — rather than fixed periodic review — concentrated on the residual "Aggressive Growth" minority (4 of 18 banks), which carries disproportionate systemic risk precisely because it persists against the grain of the dominant consolidation trend.

5.2. Azerbaijan: Structural Diversity and Persistent Heterogeneity

Azerbaijan exhibits the largest and most heterogeneous banking population in the sample ($N = 24$) and the highest systemic entropy throughout the panel ($H = 1.040$ in the 2026 summary; consistently $H > 1.02$ across 2014–2024, Table 7). Unlike Armenia, this diversity proved comparatively resilient to the 2022 shock, with entropy declining only marginally to $H = 1.02$ — the smallest contraction of the three countries. Read alongside the near-even archetype split (12 "Strategic Leaders", 6 "Traditional Savers", 6 "Aggressive Growth"; Table 3), this indicates a banking system in which multiple business models coexist as durable, rather than transitional, structural features. The corollary is supervisory: Azerbaijan records the highest average Supervisory Intensity Score in the sample ($SIS = 0.210$, Table 6), reflecting not a single dominant risk pocket but a broader base of institutions requiring active monitoring. Because entropy here is persistent rather than shock-driven, a one-size-fits-all regulatory template is least appropriate for this jurisdiction: the evidence instead supports a differentiated, standing supervisory architecture calibrated to each archetype's macro-

sensitivity profile (Table 4), rather than a framework built primarily around crisis-triggered recalibration as in Armenia.

5.3. Georgia: Structural Elasticity and Transitional Balance

Georgia occupies an intermediate position on nearly every dimension examined. It records the highest silhouette coefficient in the sample ($S = 0.605$), indicating the most cleanly separated archetypes despite the smallest bank population ($N = 16$), and a moderate average entropy ($H = 1.011$) that sits between Armenia's low and Azerbaijan's high values (Table 6). Its response to the 2022 shock was proportionate rather than extreme: entropy dipped from $H = 1.01$ in 2021 to $H = 0.98$ in 2022 — a real but comparatively muted contraction — before recovering to $H = 1.01$ by 2024 (Table 7). This pattern of measurable but bounded responsiveness is what motivates our characterization of Georgia as exhibiting "structural elasticity": the system deforms under stress without either collapsing (as in Armenia) or remaining largely unmoved (as in Azerbaijan). Combined with an archetype distribution proportionally similar to Armenia's (8 of 16 banks, 50%, classified as "Strategic Leaders") but an entropy trajectory closer to Azerbaijan's, Georgia appears to be consolidating from a more diversified starting point. For supervisors, this elasticity implies a need for monitoring bandwidth sufficient to detect transitions early — Georgia's archetypes move enough to signal emerging stress, but not so abruptly as to preclude a proactive, rather than purely reactive, response.

5.4. South Caucasus: Regional Synthesis

Pooling across jurisdictions, the regional sample ($N = 58$) is dominated by the "Strategic Leaders" archetype (31 banks, 53%), with "Traditional Savers" (14 banks, 24%) and "Aggressive Growth" (13 banks, 22%) roughly evenly split at the margin (Table 3). The three country profiles developed above — Armenia's sensitivity, Azerbaijan's diversity, and Georgia's elasticity — together trace a gradient of structural responsiveness to regional shocks that is obscured by any single national or fully pooled analysis. The most consequential regional finding is that all three systems, despite this heterogeneity, registered a synchronized entropy decline around the 2022 shock (§4.9), evidencing a shared flight-to-quality channel toward the "Traditional Savers" archetype that operates independently of each country's baseline structural profile. This is consistent with the macro-sensitivity gradient identified in Table 4, in which "Aggressive Growth" institutions across all three countries share exposure to portfolio concentration and credit-expansion risk, while "Traditional Savers" act as a common regional defensive buffer.

The estimated 15% annual transition probability toward "Strategic Leaders" under proactive supervision (§4.9) further suggests that supervisory pressure compounds meaningfully over a multi-year horizon even at modest annual intensity, offering regulators a realistic, incremental pathway for migrating "Aggressive Growth" institutions toward more resilient business models without requiring abrupt intervention. Taken together, the country-level and regional evidence supports a two-tier macroprudential design for the South Caucasus: common regional floors calibrated to the shared shock-transmission channel identified here, combined with country-specific escalation rules —

shock-responsive in Armenia, standing and differentiated in Azerbaijan, and early-warning oriented in Georgia — that reflect each system's distinct entropy dynamics. Consistent with the boundary conditions noted in §3.7, these conclusions are drawn from, and are most directly applicable to, this three-country regional bloc, and their extension to other emerging-market banking systems would require independent validation.

6. CONCLUSION

Section 2.4 identified a composite gap running through the literature reviewed in Section 2: no existing framework combined machine-learning-derived endogenous weighting, validated clustering, supervisory diagnostics, and multi-country banking analysis within a single architecture, and South Caucasus banking systems in particular remained almost entirely unstudied through a comparative, longitudinal lens. The country-level and regional conclusions that follow (§6.1–§6.5) are the empirical product of the integrated 4S-DSDM architecture proposed in response to that gap; §6.6 then returns explicitly to Table 1 to show how each identified limitation is addressed by the evidence presented in Sections 4 and 5.

6.1. Armenia

The Armenian banking sector is best characterized as a consolidated but shock-fragile system. With 61% of banks classified as "Strategic Leaders" and the lowest average Supervisory Intensity Score in the sample ($SIS = 0.185$), the sector's baseline risk profile is comparatively benign. However, the steepest single-year entropy decline in the panel ($H: 0.96 \rightarrow 0.85$, 2021–2022) demonstrates that this consolidation offers limited protection against acute geopolitical stress. We conclude that Armenia's supervisory priority should not be broad-based tightening — the dominant archetype does not warrant it — but rather targeted, shock-triggered oversight of the residual "Aggressive Growth" cohort, whose persistence against the system-wide consolidation trend signals elevated idiosyncratic risk.

6.2. Azerbaijan

Azerbaijan's banking sector is best characterized as structurally durable and heterogeneous rather than consolidating. Its consistently high entropy ($H > 1.02$ across the full 2014–2024 panel) and near-even three-way archetype split indicate that multiple business models coexist as stable long-run features rather than a transitional stage en route to a dominant archetype. This durability held through the 2022 shock, which produced the smallest entropy contraction of the three countries. We conclude that Azerbaijan requires the most differentiated, standing supervisory framework in the region — reflected in its highest average SIS (0.210) — built around persistent multi-archetype monitoring rather than the shock-triggered recalibration appropriate for Armenia.

6.3. Georgia

Georgia's banking sector is best characterized as a transitional system with the clearest internal structure ($S = 0.605$) but only partial insulation from regional shocks. Its 2022 entropy contraction ($1.01 \rightarrow 0.98$) was real but proportionate, and its subsequent recovery to $H = 1.01$ by 2024 confirms that its archetypes are neither as fragile as Armenia's nor as static as Azerbaijan's. We conclude that Georgia's supervisory framework should prioritize early-warning monitoring capacity: because its structural response to stress is measurable but bounded, moderate but sustained supervisory attention — rather than either crisis-triggered intervention or a fixed standing regime — offers the best opportunity to intercept emerging instability before it compounds.

6.4. South Caucasus Region

At the regional level, this study concludes that South Caucasus banking systems, despite substantial country-level heterogeneity in structure and shock-sensitivity, share a common transmission channel for systemic stress: the synchronized entropy decline across all three countries around the 2022 shock demonstrates that regional geopolitical events propagate through banking-sector structure independently of each system's baseline archetype distribution. The regional archetype composition — 53% "Strategic Leaders", 24% "Traditional Savers", 22% "Aggressive Growth" — further indicates that consolidation toward lower-risk business models is already the dominant regional pattern, and that the estimated 15% annual transition probability toward "Strategic Leaders" under proactive supervision provides a realistic, evidence-based benchmark for regional convergence rather than an aspirational target. The 4S-DSDM framework itself is validated as a replicable diagnostic architecture for the region: its dynamic, machine-learning-derived weighting scheme avoided the need for arbitrary expert judgment while still producing archetypes with strong cluster validity ($S = 0.512\text{--}0.605$) across three institutionally distinct national systems.

6.5. Implications for Regulators and Supervisory Authorities

For macroprudential authorities, three concrete implications follow from this analysis. First, a uniform, one-size-fits-all supervisory template is not supported by the evidence: the three countries occupy genuinely distinct positions on the sensitivity–diversity–elasticity spectrum, and supervisory cadence should be calibrated accordingly — shock-responsive review in Armenia, standing differentiated oversight in Azerbaijan, and early-warning monitoring in Georgia. Second, the Supervisory Intensity Score and Shannon Entropy diagnostics developed here are directly operationalizable: because both are computed from indicators already collected through standard regulatory reporting, national and regional supervisors can implement them without new data infrastructure, enabling real-time tracking of archetype migration between reporting cycles. Third, the synchronized regional entropy response to the 2022 shock argues for a coordination mechanism among South Caucasus supervisory authorities — at minimum, shared early-warning indicators — since shocks transmitting across national banking structures are, by this evidence, a regional rather than purely domestic phenomenon. Consistent with the boundary conditions discussed in §3.7, these recommendations are calibrated to the South Caucasus context and should be validated independently before extension to other emerging-market regions.

6.6. Closing the Literature Gap: Contribution Relative to Prior Work

Table 1 (§2.4) summarized six distinct limitations in the existing literature that motivated this study. The results reported in Sections 4–5 close each of these in turn. First, against CAMELS and CAMELS-derived composite indices, which rely on static or expert-assigned domain weights (Ankita & Samajpati, 2025; Boubaker et al., 2025; Wang et al., 2024), the 4S-DSDM framework substitutes Random Forest Permutation Feature Importance, yielding weights that shift endogenously with each domain’s demonstrated predictive relationship to bank profitability rather than a fixed a priori schedule — evidenced here by the differential prioritization of the Stability and Surplus domains documented in §4.2. Second, against macroprudential and stress-testing frameworks that operate at the level of policy transmission and capital resilience rather than bank-level classification (Cooke & Damjanovic, 2025; Darst et al., 2025; Federal Reserve Board, 2025; Mugrabi et al., 2026), this study’s archetype segmentation (Table 3) and Supervisory Intensity Score (Table 5) provide an institution-level diagnostic layer that such frameworks lack, complementary to rather than a substitute for macro-level stress testing. Third, against ML-based prediction models calibrated to single-institution, borrower-level classification with interpretability retrofitted post hoc (Niu et al., 2026; Rafie et al., 2025; Zhang et al., 2025), the 4S-DSDM framework applies the same class of ensemble-learning technique — Random Forest — not to predict individual loan or borrower outcomes but as the structural weighting mechanism for a system-level, cross-country supervisory architecture, with interpretability built into the weighting stage itself rather than added afterward.

Fourth, this design directly answers the unresolved accuracy–explainability trade-off documented by Pérez-Cruz et al. (2025) and the post hoc/ante hoc asymmetry formalized by Aysel et al. (2025): every stage of the 4S-DSDM pipeline — weighting, clustering, and diagnostic scoring — is interpretable by construction, rather than an opaque model explained after the fact. Fifth, against a clustering-methodology literature that validates cluster-number selection, distance-metric choice, and initialization robustness only on generic or single-domain benchmark data (Nwoye & Okoli, 2025; Wongoutong & Junthopas, 2025; Arutyunova & Röglin, 2025; Ritzert et al., 2025; Qing & Xu, 2025), this study extends those same validation principles — endogenous weighting, robustness testing against an independent partition — into a bank-level, multi-country panel setting via the Adjusted Rand Index gate reported in §3.5 and §4.1. Sixth, against the near-total absence of comparative, multi-country empirical work on South Caucasus banking systems (Aderogba et al., 2025, illustrating the BRICS-focused regional literature that exists elsewhere but not here), this study’s population-census panel spanning Armenia, Azerbaijan, and Georgia across twelve years directly supplies the missing regional, longitudinal evidence base.

Taken together, no single element of the 4S-DSDM framework is without precedent in the literature reviewed in Section 2 — each component draws on an established methodological lineage. What Section 2.4 identified, and what this study’s results confirm, is that the integration of these components into one validated, cross-country supervisory architecture had not previously been attempted. The country-specific and regional conclusions above (§6.1–§6.5) are therefore best understood not as isolated empirical findings but as the substantive demonstration that this integration is both feasible and supervisory-relevant: the framework produces archetypes with strong

cluster validity ($S = 0.512\text{--}0.605$) using endogenously derived weights, diagnostics computable from existing regulatory data, and a differentiated policy implication for each of the three national systems examined — the combined outcome that the fragmented state of the pre-existing literature, summarized in Table 1, had left unrealized.

ACKNOWLEDGEMENTS

The author received no specific funding for this research

CONFLICTS OF INTEREST

The author declare no conflicts of interest

REFERENCES

- Aderogba, T., Adekunle, I. A., & Atoyebi, O. E. (2025). Post-crisis bank profitability in BRICS: A CAMEL approach. *Vision*. Advance online publication. <https://doi.org/10.1177/09726527251335991>
- Ankita, & Samajpati, U. (2025). A conceptual CAMELS evaluation framework for banks: An illustrative approach. *Management Insight*, 21(1), 15–27. <https://doi.org/10.21844/mijia.21.1.2>
- Arutyunova, A., & Röglin, H. (2025). The price of hierarchical clustering. *Algorithmica*, 87, 1420–1452. <https://doi.org/10.1007/s00453-025-01327-7>
- Aysel, H. I., Cai, X., & Prugel-Bennett, A. (2025). Explainable artificial intelligence: Advancements and limitations. *Applied Sciences*, 15(13), 7261. <https://doi.org/10.3390/app15137261>
- Azerbaijan Banks Association. (n.d.). *Bank reports collection*. Retrieved July 14, 2026, from <https://aba.az/en/banklar/toplu/>
- Basel Committee on Banking Supervision. (2024). *Basel framework*. Bank for International Settlements. https://www.bis.org/basel_framework/
- Boubaker, S., Ngo, T., Samitas, A., & Tripe, D. (2025). An MCDA composite index of bank stability using CAMELS ratios and Shannon entropy. *Annals of Operations Research*, 353, 353–376. <https://doi.org/10.1007/s10479-024-06023-3>
- Central Bank of Armenia. (n.d.). *Home page*. Retrieved July 14, 2026, from <https://www.cba.am/hy/>
- Central Bank of the Republic of Azerbaijan. (n.d.). *Home page*. Retrieved July 14, 2026, from <https://www.cbar.az/home?language=en>

- Cooke, D., & Damjanovic, T. (2025). *Optimal macroprudential policy and bank capital in open economies* (Department of Economics Working Paper No. 2025_04). Durham University.
- Darst, M., Refayet, E., & Vardoulakis, A. (2025). *Macroprudential regulation and lending standards* (Finance and Economics Discussion Series No. 2020-086R1). Board of Governors of the Federal Reserve System.
- Fan, X., Zhang, S., Xue, X., Jiang, R., Fan, S., & Kou, H. (2025). An improved self-organizing map (SOM) based on virtual winning neurons. *Symmetry*, 17(3), 449.
<https://doi.org/10.3390/sym17030449>
- Federal Reserve Board. (2025). *2025 supervisory stress test methodology*. Board of Governors of the Federal Reserve System. <https://www.federalreserve.gov/publications/files/2025-june-supervisory-stress-test-methodology.pdf>
- International Monetary Fund. (2025). *Training on risk-based supervision* (High-Level Summary Technical Assistance Report No. HLS/25/21). International Monetary Fund.
- KPMG Armenia. (2015). *Armenian banking sector overview: 2015*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/https://assets.kpmg.com/content/dam/kpmg/am/pdf/2016/Armenian%20Banking%20Sector%20Overview_2016%20Q2_Eng.pdf
- KPMG Armenia. (2016). *Armenian banking sector overview: 2016*. KPMG. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/https://assets.kpmg.com/content/dam/kpmg/am/pdf/2016/Armenian%20Banking%20Sector%20Overview_2016%20Q2_Eng.pdf
- KPMG Armenia. (2017). *Armenian banking sector overview: 2017*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/https://assets.kpmg.com/content/dam/kpmg/am/pdf/2017/Armenian%20Banking%20Sector%20Overview_2017%20Q4_Eng.pdf
- KPMG Armenia. (2018). *Armenian banking sector overview: 2018*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/https://assets.kpmg.com/content/dam/kpmg/am/pdf/2017/Armenian%20Banking%20Sector%20Overview_2017%20Q4_Eng.pdf
- KPMG Armenia. (2019). *Armenian banking sector overview: 2019*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/<https://futurearmenian.com/wp-content/uploads/2022/01/Armenian-Banking-Sector-Overview.pdf>
- KPMG Armenia. (2020). *Armenian banking sector overview: 2020*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/<https://futurearmenian.com/wp-content/uploads/2022/01/Armenian-Banking-Sector-Overview.pdf>
- KPMG Armenia. (2021). *Armenian banking sector overview: 2021*. KPMG.
- KPMG Armenia. (2022). *Armenian banking sector overview: 2022*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/<https://assets.kpmg.com/content/dam/kpmg/am/pdf/2024/en-Armenian-Banking-Sector-Overview-2023-1H.pdf>

KPMG Armenia. (2023). *Armenian banking sector overview: 2023*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclclefindmkaj/https://assets.kpmg.com/content/dam/kpmgsites/am/pdf/2024/Armenian-Banking-Sector-Overview-2024-1H.pdf.coredownload.inline.pdf

KPMG Armenia. (2024). *Armenian banking sector overview: 2024*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclclefindmkaj/https://assets.kpmg.com/content/dam/kpmgsites/am/pdf/2025/2025-H1-Armenian-Banking-Sector-Overview-ENG.pdf.coredownload.inline.pdf

KPMG Armenia. (2025). *Armenian banking sector overview: First half 2025*. KPMG. <https://kpmg.com/am/en/insights/2025/12/armenian-banking-sector-overview-2025.html>

KPMG Armenia. (2026). *Armenian banking sector overview: First half 2025*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclclefindmkaj/https://assets.kpmg.com/content/dam/kpmgsites/am/pdf/2026/04/en-Armenian-Banking-Sector-Overview-2025.pdf.coredownload.inline.pdf

KPMG Georgia. (2014). *Georgian banking sector overview: 2015*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclclefindmkaj/https://assets.kpmg.com/content/dam/kpmg/ge/pdf/2015/ge-en-banking-sector-2015-Q2.pdf

KPMG Georgia. (2015). *Georgian banking sector overview: 2015*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclclefindmkaj/https://assets.kpmg.com/content/dam/kpmg/ge/pdf/2016/08/Banking%20Sector%20Overview%20-%20first%20half%202016.pdf

KPMG Georgia. (2016). *Georgian banking sector overview: 2016*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclclefindmkaj/https://assets.kpmg.com/content/dam/kpmg/ge/pdf/2017/Georgian%20Banking%20Sector%20Overview_%20Q3%202017.pdf

KPMG Georgia. (2017). *Georgian banking sector overview: 2017*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclclefindmkaj/https://assets.kpmg.com/content/dam/kpmg/ge/pdf/2018/ge-en-georgian-banking-sector-overview-Q3-2018.pdf

KPMG Georgia. (2018). *Georgian banking sector overview: 2018*. KPMG. https://assets.kpmg.com/content/dam/kpmg/ge/pdf/2019/Georgian_Banking_Sector_Overview_Q3_2019.pdf

KPMG Georgia. (2019). *Georgian banking sector overview: 2019*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclclefindmkaj/https://assets.kpmg.com/content/dam/kpmg/ge/pdf/2020/11/ge-en-georgian-banking-sector-overview-q3.pdf

KPMG Georgia. (2020). *Georgian banking sector overview: 2020*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclclefindmkaj/https://assets.kpmg.com/content/dam/kpmg/ge/pdf/2021/ge-en-georgian-banking-sector-overview-q1-2021.pdf

KPMG Georgia. (2021). *Georgian banking sector overview: 2021*. KPMG. chrome-extension://efaidnbmnnnibpcajpcgclclefindmkaj/https://assets.kpmg.com/content/dam/kpmg/ge/pdf/2022/georgian-banking-sector-overview-Q1-2022.pdf

- KPMG Georgia. (2022). *Georgian banking sector overview: 2022*. KPMG. <https://kpmg.com/ge/en/insights/2024/01/georgian-banking-sector-overview-q3-2023.html>
- KPMG Georgia. (2023). *Georgian banking sector overview: 2023*. KPMG.
- KPMG Georgia. (2024). *Georgian banking sector overview: Q3 2023*. KPMG. <https://assets.kpmg.com/content/dam/kpmg/ge/pdf/2024/Georgian-Banking-Sector-Overview-Q3-2023.pdf>
- KPMG Georgia. (2025). *Georgian banking sector overview: 2025*. KPMG.
- Mugrabi, F., Belkhir, M., Ben Naceur, S., Candelon, B., & Choi, W. G. (2026). Macroprudential policy and bank systemic risk: Does inflation targeting matter? *Emerging Markets Review*, 71, 101397. <https://doi.org/10.1016/j.ememar.2025.101397>
- National Bank of Georgia. (n.d.). *Home page*. Retrieved July 14, 2026, from <https://nbg.gov.ge/en>
- Niu, X., Wang, X., Wang, J., Liu, Z., & Zhang, L. (2026). Profit-oriented loan default prediction for the financial industry: A fusion framework with interpretability. *Financial Innovation*, 12(1), Article 6. <https://doi.org/10.1186/s40854-025-00796-7>
- Nwoye, O. N., & Okoli, C. N. (2025). Enhancing clustering performance: A hybrid generalized K-means approach. *International Journal of Advanced Research*, 13(4), 403–411. <https://doi.org/10.21474/IJAR01/20733>
- Pérez-Cruz, F., Prenio, J., Restoy, F., & Yong, J. (2025). *Managing explanations: How regulators can address AI explainability* (FSI Occasional Paper No. 24). Bank for International Settlements. <https://www.bis.org/fsi/fsipapers24.pdf>
- Qing, H., & Xu, X. (2025). Latent class analysis with arbitrary-distribution responses. *Entropy*, 27(8), 866. <https://doi.org/10.3390/e27080866>
- Rafie, Z., Talab, M. S., Koor, B. E. Z., Garavand, A., Salehnasab, C., & Ghaderzadeh, M. (2025). Leveraging XGBoost and explainable AI for accurate prediction of type 2 diabetes. *BMC Public Health*, 25, Article 3688. <https://doi.org/10.1186/s12889-025-24953-w>
- Rahman, M. M., & Datta, P. (2025). Data-driven business strategies with the power of the K-means algorithm. *International Journal of Higher Education Management*, 11(2), 1–10.
- Ritzert, M., Turishcheva, P., Hansel, L., Wollenhaupt, P., Weis, M. A., & Ecker, A. S. (2025). *Hierarchical clustering with maximum density paths and mixture models* (arXiv:2503.15582). arXiv. <https://doi.org/10.48550/arXiv.2503.15582>
- Tapio, R., & Tarepe, D. (2025). Comparative analysis of random forest and hybrid ARIMA-random forest models for student enrollment forecasting in higher education. *Journal of Advances in Mathematics and Computer Science*, 40(3), 124–136. <https://doi.org/10.9734/jamcs/2025/v40i31982>

- Wang, H., Sua, L. S., & Dolar, B. (2024). CAMELS-DEA in assessing the role of major factors in achieving higher efficiency levels: Evidence from Turkish banks. *Applied Economics*, 56, 3844–3861. <https://doi.org/10.1080/00036846.2024.2339186>
- Wongoutong, C., & Junthopas, W. (2025). Pre-determining the optimal number of clusters for k-means clustering using the parameters package in R and distance metrics. *Applied Sciences*, 15(21), 11372. <https://doi.org/10.3390/app152111372>
- Yang, Y., & Wang, H. (2025). Random forest-based machine failure prediction: A performance comparison. *Applied Sciences*, 15(16), 8841. <https://doi.org/10.3390/app15168841>
- Zhang, X., Zhang, T., Hou, L., Liu, X., Guo, Z., Tian, Y., & Liu, Y. (2025). Data-driven loan default prediction: A machine learning approach for enhancing business process management. *Systems*, 13(7), 581. <https://doi.org/10.3390/systems13070581>