Journal of Applied Economic Sciences
Volume XX, Summer, Issue 2(88), 2025

Multidimensional Surveillance of the Indian Banking System: A Cluster Approach

Tasneem CHHERAWALA
https://orcid.org/0009-0002-9396-1670
National Institute of Bank Management, Pune, India
tasneem@nibmindia.org

Alka VAIDYA[1]
https://orcid.org/0009-0005-9039-4596
National Institute of Bank Management, Pune, India
alka@nibmindia.org

Sanjay BASU
https://orcid.org/0009-0008-1311-9349
National Institute of Bank Management, Pune, India
sbasu@nibmindia.org

Article's history:

Received 1st of May, 2025; *Received in revised* form 31st of May, 2025; Accepted 15th of June, 2025; Available online: 25th of June, 2025. Published as article in the Volume XX, Spring, Issue 2(88), 2025.

Suggested citation:

Chherawala, T., Vaidya, A., & Basu, S. (2025). Multidimensional surveillance of the Indian banking system: A cluster approach. *Journal of Applied Economic Sciences,* Volume XX, Summer, Issue 2(88), 255–272. https://doi.org/10.57017/jaes.v20.2(88).07

Abstract:

This paper develops a multidimensional risk ranking model for bank supervision, using the k-means cluster approach. It employs a combination of size, balance sheet, and market-based indicators for predicting idiosyncratic and systemic risk. The risk rankings are benchmarked to a long-run threshold, which regulators may wish to target for the resolution of financial crises. When the tool is applied to data on Indian banks between 2005 and 2023, several important results emerge. The effectiveness of different signals depends on the nature of the impending financial stress event. Market-based indicators are better at predicting external shocks, while markers related to asset size forecast credit booms and busts with greater efficiency. Moreover, the model is superior to heuristic regulatory measures of bank-specific distress and its resolution.

The framework is also able to distinguish between the risk performance of public and private sector banks, during a period that spans the global financial crisis (GFC), the non-performing asset (NPA) crisis in India, and the COVID pandemic. Private banks exhibit better risk profiles during the GFC and NPA crises. This study emphasizes a multifaceted approach to bank supervision, in order to capture the heterogeneity of financial crises.

Keywords: K-means clustering, systemic risk, idiosyncratic risk, bank supervision.

JEL Classification: C88; G21; G28.

[1] Corresponding Author

Introduction

Continuous oversight of banking sector vulnerabilities is a critical task for authorities worldwide. Micro-prudential risk-based supervision requires the detection of bank-wide stress for early intervention and pre-emptive actions. Recognition of potential system-level financial stress is also essential for appropriate macro-prudential policy measures. These objectives are paramount in emerging economies like India, where the banking system plays a stellar role, against the backdrop of financial market underdevelopment.

The surveillance of banking systems is a complex task, because it is based on multiple parameters, derived from both accounting ratios and market prices. The interactions between balance sheet and market-based indicators can have significant implications for bank safety and soundness as well as financial stability. Traditionally, supervisors have used the CAMELS model (comprising indicators of capital adequacy, asset quality, management, earnings, liquidity, and sensitivity to market risk) to rank and identify weak banks (Faulk et al., 2018; Gaul & Jones, 2021). Academic research on bank insolvency risk has often employed the Z-Score (Boyd & Runkle, 1993). Both these measures are based on balance sheet data. On the other hand, the systemic stability of banks has been captured through a range of metrics that rely on market data. Some prominent examples are the marginal expected shortfall (Acharya et al., 2017) and the capital shortfall measure called SRISK (Brownlees & Engle, 2017).

Given the plethora of established indicators, it is essential to create an optimal combination of early warning signals of financial distress (Acharya et al., 2025). This set of indicators must be benchmarked to the long-run risk and performance thresholds, which are unique to the banking system of every country. Hence, the objective of this study is to develop a tractable analytical framework for optimal aggregation of multiple idiosyncratic and systemic risk indicators, to group and rank banks over time, using the unsupervised machine learning technique of k-means clustering. In this method, 'k' distinct groups or clusters are created. Each observation is assigned to the nearest cluster average. The variance within each cluster is minimized, while the variance between different clusters is maximized, in this approach.

In this article, cluster analysis is applied to quarterly panel data of Indian commercial banks for the period between March, 2005 and December, 2023. The end date has been chosen because quarterly data was available, only till December, 2023[2]. As explained in Section 2, the year 2004 marked the onset of rapid credit growth and asset overvaluation in India (IMF 2006). In order to capture the impact of such early warning signals on bank-specific and systemic risk, the start date is chosen as March, 2005. The variables comprise 21 balance sheet-based indicators and 7 market-based indicators of standalone and systemic vulnerability of banks, which are reduced to a limited set of factors through principal component analysis. Four clustering models are run, taking different combinations of risk factors, to identify the models that provide the greatest sensitivity to known idiosyncratic and systemic distress scenarios. For each model, an optimal set of 5 clusters is ranked based on the aggregate risk scores of the factor centroids.

The results show that the share of banks in the riskiest cluster spikes before the occurrence of the Global Financial Crisis (GFC; 2007 – 2009) and the Non-Performing Assets (NPA) crisis (2015 onwards), implying that the metric is able to predict these shocks well before they materialize. The sharpness of the spikes also attests to the severity of the episodes. In contrast, the spike for the COVID-19 pandemic, if it appears, happens around the same time as the event itself. This implies that financial system indicators did not anticipate this crisis. The smaller uptick also suggests that the GFC and NPA crises had a stronger impact on the financial sector.

Furthermore, when the size factor is included, the metric captures periods of rapid asset growth, both before the NPA crisis and after the Covid pandemic. This has often been associated with the build-up of systemic risk. It is also shown that cluster-based risk ranking of banks is able to identify the emergence and resolution of the NPA crisis ahead of the regulatory Prompt Corrective Action (PCA) framework. Finally, the results depict that, on

[2] The extension of our analysis with data for additional two quarters (March 2024 and June 2024), which became available with a lag, is discussed in Section 5.

average, public sector banks were riskier than private sector banks, during both the GFC and NPA crisis. The study demonstrates that cluster analysis provides enhanced predictive power of standalone distress and systemic risk.

The article makes several vital contributions to the literature. First, the multidimensional surveillance system is shown to perform better than the regulatory PCA framework in the prediction and management of bank fragility in India. A small body of research studies the effectiveness of government policies and regulations across many countries, in the management of bank distress, through cluster analysis. The argument is that the multiple indicator approach, employed in this paper, can be a more powerful tool for risk assessment than regulatory models for bank supervision. Second, the model compares financial fragility between state-owned and private banks. This aspect has not been studied so far, through cluster analysis. Eichengreen & Gupta (2013) and Acharya and Kulkarni (2019) compare the performance of public and private sector banks only during the GFC, to highlight the role of government guarantees. However, this study also compares their risk profiles during the NPA crisis and the Covid pandemic. The results have significant implications for government intervention. Finally, the centroid for each cluster, in this model, is a static one. This means that, unlike many other papers, we assess bank fragility in each cluster vis-à-vis a long-run benchmark level of risk, enabling easier identification of outlier banks. This gives regulators a target level of risk, to which they may wish to revert through their policies, after crisis episodes. The analogy here is with the inflation targeting framework, at central banks, in which the long-run optimal rate of inflation is often constant.

A recent paper by Mercadier et al. (2025) comes the closest to the spirit of this paper. The authors also develop a multidimensional model, based on the k-means clustering algorithm, for monitoring bank-specific and systemic fragility worldwide. They also infer that the GFC had a stronger impact than the COVID-19 pandemic on the cluster framework. However, the present study differs from their article in important respects. First, four cluster models are employed, to offer a holistic perspective on the sources of bank distress and highlight the marginal impact of size and market factors on the risk measures. Unlike Mercadier et al. (2025), bank size is considered a positive indicator of risk. At a system level, this treatment aligns better with regulatory concerns relating to rapid asset growth and its signal of a brewing crisis. Second, a key conclusion of the present study is that it forecasts the idiosyncratic risk of Indian banks during the domestic NPA crisis and its ultimate resolution better than the regulatory PCA framework. Since the Mercadier metric makes a global comparison of bank fragility, country-specific problems may get obscured. Third, this article compares risk profiles between ownership types to show that public sector banks have been riskier than the private sector counterparts during different crises.

The rest of the paper is organized as follows. Section 1 contains the literature review. Section 2 describes the data and variable selection. Section 3 focuses on the methodology. Section 4 discusses the results and their implications. Section 5 extends the main results, with new data. Section 6 concludes.

1. Literature Review

Cluster analysis is a well-established unsupervised learning technique (Jain et al., 1999) that has been used in many research studies on banking and finance. For instance, Leon et al. (2017) show that clustering algorithms create stock portfolios with lower volatilities than those designed by the mean-variance optimization process. The risk-adjusted performance of these portfolios, measured by the Omega Ratio, is also better. Sass and Thos (2024) also find that cluster analysis improves risk-adjusted performance, measured with Sharpe, Sortino & Omega Ratios, of equally weighted equity portfolios. Hence, these models are useful tools for market risk management. Likewise, cluster algorithms can be used to enhance the recovery performance of Non-Performing Loans when credit portfolio heterogeneity is high (Carleo, et al., 2024). Hence, these methods add value to credit risk management as well.

However, the literature on the application of this methodology to the classification of banks, based on overall risk profiles, is limited. In one of the earliest studies, the European Central Bank (FSR, 2006) proposes cluster analysis for identifying large and complex banking groups in the euro area. It includes features like asset size, degree of interconnectedness, and nature of off-balance sheet activities, to distinguish banks that pose high systemic risks from the others. Dardac & Boitan (2009) apply cluster analysis to sixteen Romanian universal banks based on their risk profile, profitability, and intermediation cost. They use a variety of annual financial indicators for the period between 2004 – 2006. They employ the single linkage agglomerative method and discover that the clusters remain relatively stable over time in terms of exposure to risk and profitability. The study is unable to differentiate between individual banks or clusters in terms of degrees of risk.

In a cross-country study, Reverchuck et al. (2013) conduct cluster analysis on summary data of banking systems across 48 countries over the period 2004 – 2009, to measure their efficiency levels during crises. This approach helps them assess the effectiveness of government policies and regulations for monitoring and regulating their banking sectors, on various risk and profitability indicators, relative to other countries. Using cluster analysis on a sample of more than 3000 banks from 32 European countries, between 2010 and 2017, Ayadi et al (2021) identify and track the evolution of bank business models. They demonstrate that higher risk and lower profitability trigger changes in bank business models. Through a propensity score matching approach, they show that migration to target business models improves bank profitability, stability, and cost efficiency. João et al. (2023) develop a dynamic clustering model to study time-varying bank group structures. Their framework, applied to quarterly data of 299 European banks between 2008 and 2018, demonstrates that bank business models have become more homogenous over time, in terms of key attributes like size, leverage, share of trading book, and funding sources. The model implies that the convergence of business models may aggravate systemic concentration risk and interbank spillover effects.

Mercadier et al. (2025) rank five clusters of 256 banks from 43 countries, based on 72 bank-specific and systemic risk indicators, reduced to 10 main factors through Principal Components Analysis, for the period between 2004 and 2024. It not only classifies individual banks but also segregates countries and regions in terms of financial fragility. The model, based on K-means cluster algorithms, indicates that the global financial crisis had a more severe effect on the banking industry than the COVID-19 pandemic. It also shows that large banks need not always create higher systemic risk. This study is an important contribution to the literature on bank supervision and pre-emptive action, using machine learning approaches.

In the Indian context, Vashisht & Sarva (2024) employ k-means clustering to analyze the risk profiles of 30 Indian commercial banks over the period from 2009 to 2020. In their approach, banks are categorized into two distinct risk clusters based on a limited set of financial variables. They identify profitability and non-performing loans ratios as critical drivers of bank performance. However, their study does not consider market indicators, nor does it address systemic risk implications. To the best of our knowledge, the present application is the first to utilize cluster analysis for evaluating risks in the Indian banking sector, based on a comprehensive set of accounting and market-based signals.

2. Data and Variable Selection

The study is conducted using quarterly data of domestic public and private sector scheduled commercial banks (SCBs) in India, for the period between March 2005 to December 2023. The start date is chosen as end-March 2005, for several reasons related to developments in the Indian banking sector. First, the Benchmark Prime Lending Rate (BPLR) Regime was introduced by the Reserve Bank of India (RBI) in April 2003. This system was expected to make bank loan rates more sensitive to market interest rates and credit allocation more rational (Mohanty, 2010). Secondly, after a string of heavy losses in bond markets, banks were allowed by RBI to shift a larger fraction of their investments in government securities from the trading book to the banking book, in September 2004 (RBI 2004). This measure reduced their incentives for investment in low-yield bonds and made credit disbursement more lucrative. Finally, real estate and stock prices – two assets preferred as collateral - had begun to surge from early 2004 onwards (IMF 2006). All these factors resulted in year-on-year bank credit growth of more than 30% in 2024-25, with the expectation that rapid credit growth would persist over the next few years. It is well known that credit booms are strong leading indicators of financial crises. Selecting March, 2005 as the start date allows us to examine the linkage between growth in asset size and future banking crises in India. The end-date – that is, December, 2023 – was the last quarter for which data were available for the original cluster models. An extension of the analysis, with data till June 2024 has been shown in Section 5. The total sample period selected covers three financial crises in India.

Without loss of generality, very small and highly specialized banks like Small Finance Banks and Payment Banks are excluded. Foreign Banks are also omitted since they are highly diversified in terms of the nature of business activities that they conduct in India. The number of banks in the sample changes over time, because both new banks have been established during the period and some banks have been merged. Also, for the clustering models that use equity market data, banks are included from the period when they were first listed on the stock exchange. The sample of banks represents 85.69 % of the total assets of Indian SCBs as on March 2005 and 94.13% as on December 2023.

Table 1 describes the variables used in the analysis. The rationale for the inclusion of each of the variables, along with their directional relationship to bank risk, is supported by well-established literature, reference to which is provided in Table 1. 21 indicators have been derived from financial statements and regulatory disclosures of banks. These indicators describe different dimensions of risk and performance of individual banks. The indicators are grouped into eight broad factors representing Size, Credit Concentration, Interconnectedness, Illiquidity, Loss Buffers, Asset Quality, Cost Inefficiency, and Profitability. To create a unidirectional positive relationship with banks' overall riskiness, some of the original variables have been appropriately transformed.

Many studies suggest that traditional balance sheet data are only partially able to detect ex-ante, financial institutions' risk of failure, and market-based indicators are often able to give some advanced signals of risk (IMF, 2009). As such, another 7 indicators, derived from the stock price data of individual banks, are included. Of these, the variable for market capitalization (LN_MCAP) is included in the Size factor, and Price to Book Value (P_BV) is included in the Profitability factor. The remaining 5 indicators (BETA, EQ_VOL, 5PCVAR, LRMES, and MZSCORE) are grouped into a new factor: Sensitivity to Market Risk.

Bank-wise quarterly financial ratios and other balance sheet indicators are derived from published financial statements and regulatory disclosures available from the Reserve Bank of India database. Market parameters have been constructed from equity market data sourced from CMIE Prowess.

Table 1: Variable description

No.	Variable	Description	Factor	Relation with Bank Risk*	Reference
1	SHARE_TA	Bank's Total Assets as a percentage of the Total Assets of SCBs	Size	+	Bos et al. (2024),
2	LN_TA	Natural Logarithm of Bank's Total Assets		+	
3	LN_MCAP	Natural Logarithm of Bank's Market Capitalization		+	
4	GA_TA	Gross Loans as a % of Total Assets	Credit Concent-ration	+	BIS (2006),
5	NRA_GA	Non-Retail Advances as a % of Gross Loans		+	
6	EQ_TA	Equity as a Proportion of Total Assets	Loss Buffer	-	Hugonnier and Morellec (2017), Conlon et al. (2020), Velliscig et al. (2023)
7	T1CAR	Tier 1 Capital as a Proportion of Risk-Weighted Assets		-	
8	CRAR	Total Regulatory Capital as a Proportion of Risk-Weighted Assets		-	
9	PCR	Provisions Coverage Ratio (adjusted for Write Offs)		-	
10	CD	Credit to Deposit Ratio	Liquidity	+	Boďa and Zimková (2021), Hugonnier and Morellec (2017)
11	SLR_TA	Investments in Statutory Liquidity Ratio Securities to Total Assets		-	
12	GIB	Gross sum of on balance sheet Interbank assets and liabilities to total assets	Inter-connected ness	+	Dinger and von Hagen (2009)

No.	Variable	Description	Factor	Relation with Bank Risk*	Reference
13	NNPAR	Ratio of Non-Performing Loans Net of Specific Provisions to Net Loans	Asset Quality	+	Velliscig et al. (2023)
14	SAR	Ratio of Stressed Advances (Non-Performing + Stressed Restructured Loans) to Gross Loans		+	
15	RWAD	Ratio of Regulatory Risk-Weighted Assets to Total Assets		+	
16	CIR	Cost to Income Ratio	Cost Efficiency	+	Fiordelisi et al. (2011)
17	OER	Overhead Efficiency Ratio		-	
18	CASA_TD	Ratio of low-cost current and savings deposits to total deposits		-	
19	NIM	Net Interest Margin	Profita-bility	-	Martynova et al. (2015)
20	ROE	Return on Equity = Net Profit to Total Book Value of Equity		-	
21	NONINTINC_TI	Non-Interest Income to Total Income		-	
22	RP_NII	Risk Provisions to Net Interest Income		+	Velliscig et al. (2023)
23	P_BV	Equity Price to Book Value Ratio		-	
24	BETA	Bank's stock beta	Sensiti-vity to Market Risk	+	
25	EQ_VOL	Annualised volatility of the Bank's stock returns		+	
26	5PCVAR	Annualised Value at Risk of Bank's stock at 5% level of confidence		+	
27	MZSCORE	Market Z Score		+	
28	LRMES	Long Run Marginal Expected Shortfall		+	

Note: *"+" denotes positive relationship, "- "denotes negative relationship with bank risk.* Source: Authors

3. Methodology

3.1. Cluster Analysis

We use the k-means clustering technique to derive risk clusters of Indian commercial banks from 2005 to 2023. Clustering is an exploratory data analysis technique used to study the patterns within a complex and unlabelled data set, which cannot be described only through the traditional statistical measures such as mean or covariance. (Sami & Tommi, 2006). Observations in the same group are similar to each other in certain ways, while being dissimilar in different groups. Since we do not associate any prior categorization in terms of risk performance on the banks in our sample, our approach to risk-ranking of banks uses the unsupervised learning method of cluster analysis.

Clustering algorithms are broadly classified into two categories - partition-based and hierarchical. In partitioning methods, distinctive groups are created by dividing the datasets, whereas hierarchical clustering aims at forming a tree linkage or relationships for the data objects, which are then grouped. (Oyewole & Thopil, 2023). Hierarchical clustering is more suitable when multi-tiered structures are expected within the datasets, such as organizational structures in corporate businesses. With large datasets, however, it can become difficult to determine the optimal number of clusters.

The k-means formulation is the most popular partitioning-based clustering algorithm (Oti et al., 2021). It first creates an initial set of 'k' partitions using a centre point or "centroid" for each cluster, which corresponds to the mean of observations assigned to the cluster. It uses an iterative relocation technique that moves objects from one cluster to another by maximising the expected similarity between data points and their associated cluster centroids. The quality of k-means clustering is measured through the within-cluster squared error criterion (Yuan and Yang, 2019). Since the variables in our dataset do not have a pre-defined hierarchy, and we wish to rank the clusters based on the relative risk of cluster centroids, we use k-means clustering. The number of clusters, k, can be defined optimally using methods such as the Elbow Curve or WSS (within cluster sum of squares), silhouette, and gap-statistics. The present work uses the WSS method, which is a graphical representation technique where WSS are plotted against different values of k and an "elbow" point is identified where the rate of decrease in WSS slows down significantly. We obtain k = 5 based on an arm-shaped line chart where the breakpoint happens and use the 'k-means' algorithm available in 'R' version 4.3.2.

João et al. (2023) describe how alternative cluster analysis techniques have developed in the recent past in terms of their applications to panel data. They classify panel data clustering based on two criteria. First, whether the cluster location in terms of the centroids is static or changes over time. Second, whether the cluster membership is static (that is, each time series is allocated to one cluster over the entire sample period) or dynamic (that is, cluster assignment of each cross-sectional element can change over time). In our analysis, we create static k-means clusters with dynamic cluster composition. The static k-means allows us to interpret cluster centroids as stable, long-run average risk benchmarks. The dynamic cluster composition implies that each bank can be assigned to different clusters in different periods, depending on how its point-in-time risk profile compares to the through-the-cycle benchmarks. The time-variant cluster membership, which captures cluster switches, is preferable from the perspective of supervisory monitoring of changes in banks' risk profile.

We run 4 models of cluster analysis based on different combinations of risk indicators as depicted in Table 2. All 4 models have a common set of financial statement-based variables. Models M1 and M2 exclude market variables, which are included in Models M3 and M4. Size-related variables are included in M1 and M3 but excluded in M2 and M4. These combinations are created for two reasons. First, given that the academic literature is divided over the impact of size and concentration on bank risk, we wish to examine the clusters with and without the Size factor. Second, applications of cluster analysis for bank performance evaluation have traditionally considered only financial and regulatory ratios derived from balance sheet and income statements. We wish to assess whether the introduction of market-based parameters (Mercadier et al., 2025) enhances the surveillance of bank risks.

Table 2: Cluster analysis models

Model	Financial Statement-based Variables	Size-related Variables	Market data-based Variables
M1: Without Market Data; With Size	✓	✓	X
M2: Without Market Data; Without Size	✓	X	X
M3: With Market Data; With Size	✓	✓	✓
M4: With Market Data; Without Size	✓	X	✓

Source: Authors

3.2. Dimensionality Reduction with Principal Components Analysis

A critical issue with big datasets with high-dimensional data, such as ours, is "The Curse of Dimensionality" (Velliangiria et al., 2019). Our dataset includes 28 variables, which can lead to an inelegant and complex cluster structure that is difficult to interpret. Thus, dimensionality reduction is an important data pre-processing step that improves computational efficiency and enhances the interpretability of our models. The technique not only identifies the most informative features, but it also constructs new features based on existing ones that capture the essence of the original data. (Mutinda & Langat, 2024). We use the well-established Principal Component Analysis method for dimensionality reduction within each risk factor.

Principal Component Analysis is a commonly used multivariate technique in which several related variables are transformed to a set of uncorrelated variables. (Jackson, 1991). The method reduces the number of variables by transforming them to essential features, called principal components, which are linear combinations of the original variables that maximally explain the variance of all the variables while preserving as much information as possible. In this study, we select the principal components that explain a minimum of 75% cumulative variance for each risk factor. PCA is applied on the datasets of each of the four cluster analysis models to obtain a reduced number of features as described below:

M1: 21 variables reduced to 14 principal components

M2: 19 variables reduced to 13 principal components

M3: 28 features reduced to 18 principal components

M4: 26 features reduced to 17 principal components

PCA is applied to the standardised values of all variables to transform them to a similar scale for better performance of machine learning models in terms of accuracy and speed. We use 'Standardisation' method for the same (Trebuňa et al., 2014). It may be noted that due to the variable transformation and the incidence of high positive correlation across variables within each factor, the values of each of the principal components are also positively related to bank risk. Based on the positive relationship of transformed variables to bank risk, the cluster centroids are scored on a linear scale of 1-5, where 1 is assigned to the lowest value and represents the lowest risk, and 5 is assigned to the highest value to represent the highest risk. The cluster's risk rank is based on the aggregate score across all its centroids. The cluster with the highest aggregate score is the riskiest cluster, assigned Risk1, and the cluster with the lowest aggregate score is the least risky cluster, assigned Risk5. Each panel observation (bank x period) is given the same risk rank as the cluster to which it belongs.

4. Research Results

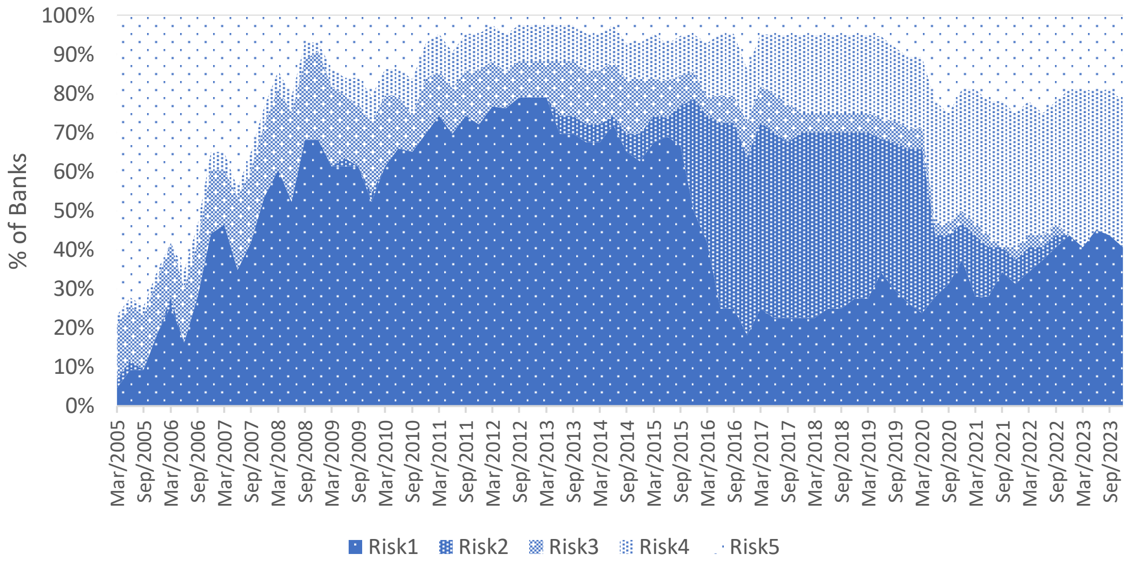
The results of the 4 models of cluster analysis are presented in this section. First, the dynamics of the cluster compositions for each of the models are demonstrated and the implications for systemic risk are discussed (Section 4.1). Next, the early warning signals for micro-prudential supervision that the risk clustering method provides are analyzed through comparison with the regulatory PCA classification of weak banks (Section 4.2). Finally, the cluster-based risk performance of public and private banks is statistically compared (Section 4.3).

4.1. Dynamics of System Risk

In this section, the risk clusters of Indian commercial banks over the period 2005 to 2023 using each of the cluster models, are graphically presented. During this period, there were three distinct financial crises that occurred – the GFC (2007-2008), the NPA crisis (2014-2015) and the Covid pandemic (2020-2021). The trend of the riskiest cluster (Risk1) is evaluated in the context of the crises.

Figure 1 depicts the trend of cluster compositions for model M1. As can be seen, when clusters include asset size as a driver of risk, there is a sharp increase in the proportion of banks in the riskiest cluster (Risk1) between 2005 to 2012-13, peaking at 79% by Mar-2013. This corresponded with the period when bank asset growth was the fastest and overall risk profiles worsened. The riskiest cluster continued to capture the largest proportion of banks till end-2015, the period over which the NPA crisis occurred for Indian banks. The share of Risk1 falls during the period between 2015 and 2019. This may be explained by the slowdown in credit growth and the improvement in balance sheet quality after the NPA crisis. This was the phase when Reserve Bank of India- RBI's asset quality reviews (AQRs) prompted banks to reduce their risky loans. It is interesting that, after the Covid-19 pandemic, there is once again an uptick in the share of the riskiest cluster due to credit growth rebound, despite stronger performance by banks.

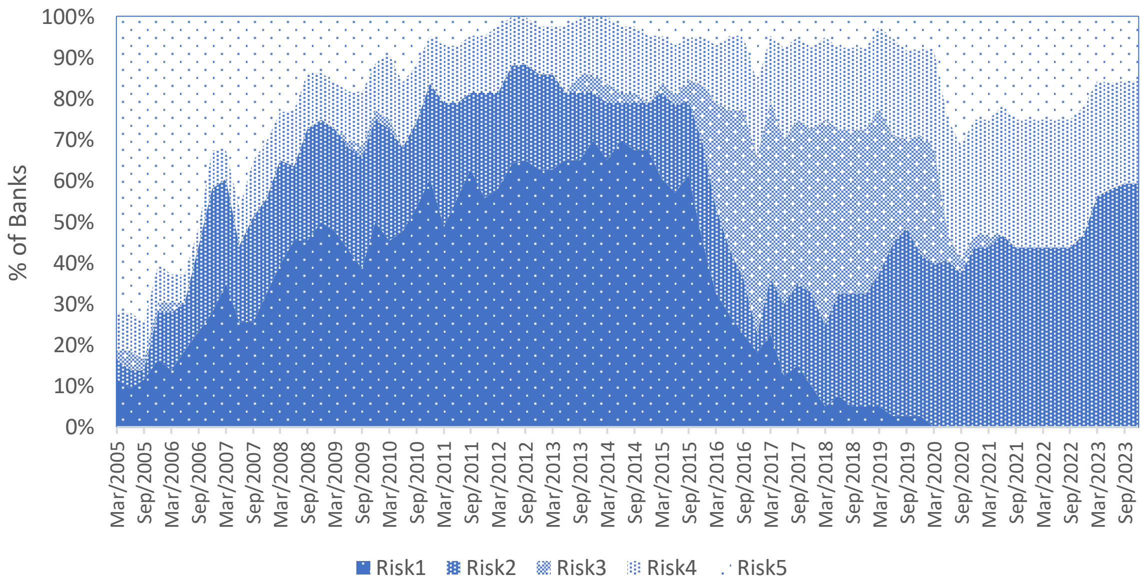
Figure 1: Model M1- with Size, without Market Indicators



Source: Authors

In Figure 2, which captures the long run bank risk clusters without size and market factors (model M2), we continue to see a rapid rise in the proportion of banks in Risk1, from around 10% in 2005 to a maximum of 67.44% in Dec 2014. Thus, even without size consideration, the risk profile of Indian banks continues to worsen in terms of pure accounting ratios. It is reflected in the increasing share of the riskiest cluster, during the period of the GFC and NPA crises. However, the spike is lower during these crises, than for model M1, where the size factor is included. Thus, weaker accounting ratios alone do not explain the severity of the episodes. Rapid credit growth or size also plays an important role in the buildup of bank distress. After the NPA crisis, the share of banks in Risk1 plummeted, and was replaced by an increasing proportion of banks in the next risk cluster (Risk2), thereby capturing the post-pandemic stress in the banking sector, albeit at a lower level of severity.

Figure 2: Model M2- without Size, without Market Indicators

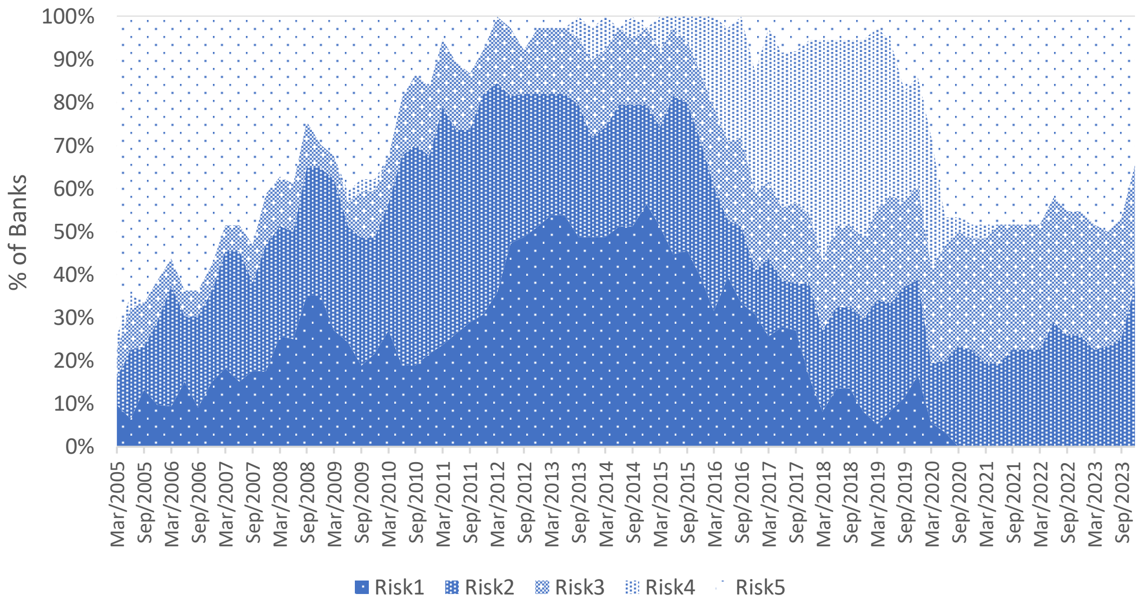


Source: Authors

When long run risk clusters are defined using both market and size factors (as depicted in Figure 3 for model M3), a number of interesting results emerge. First, throughout the period of analysis, the overall proportion of banks in Risk1 is lower for clusters defined with market factors (Figure 3), compared to clusters defined without market factors (Figure 1 and Figure 2). Thus, equity markets do not seem to penalize balance sheet size and financial quality as significant idiosyncratic risks. In other words, the impact of poorer accounting ratios and fast credit growth may have been masked by (that is, negatively correlated with) equity market performance.

Second, the domestic financial market stress during GFC (2008-09) is more prominent, compared to both models M1 and M2, defined without market factors (Figure 1 and Figure 2). This is seen in the distinct peak of Risk1 share at 35.14% in Sep-Dec 2008. Thus, inclusion of market factors in the clustering model enhances predictive ability of external stress factors. Third, the period of the NPA crisis remains distinctly highlighted with the rapid increase of banks in the Risk1 cluster during 2012-2015. However, the proportion of riskiest banks is lower when market factors are considered (peak of 56.41% of banks in Risk1 in Dec-2014) versus when market factors are excluded. Once again, addition of market factors appears to dampen the adverse impact of size and accounting ratios. Finally, the Covid pandemic does not show any increase in the riskiest cluster.

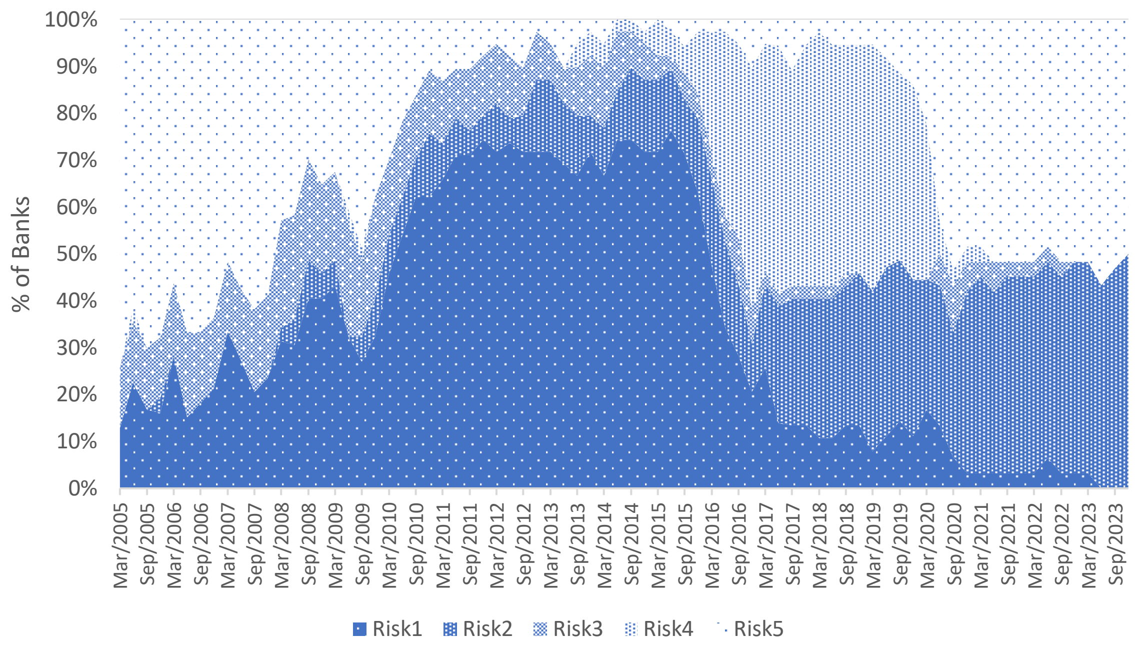
Figure 3: Model M3- with Size, with Market Indicators



Source: Authors

Figure 4 shows the trend in cluster compositions when these are defined including Market factors but excluding Size (model M4). Without Size factors, there is a much higher share of Risk1 during various crisis periods, since the negative correlation between size and market performance is absent in this model. The inference is that larger banks may be associated with better market indicators. Simultaneous inclusion of both signals cancels out the effects of each other. When one factor is omitted, the impact of the other one is stronger.

Figure 4: Model M4- without Size, with Market Indicators



Source: Authors

4.2. Prompt Corrective Action (PCA) for Banks

The Prompt Corrective Action (PCA) scheme was introduced by the RBI in December 2002, for the purpose of monitoring and resolving weak banks in India. At the time of its inception, it took into consideration three trigger points – Capital to Risk Weighted Asset Ratio (CRAR), Net Non-Performing Assets (Net NPA) and Return on Assets (ROA). Banks which breached one or more of these ratios would be subject to regulatory actions and constraints on business growth and risk-return management.

The PCA framework was revised in April, 2017 and became effective from FY 2017-18. It enhanced the indicators to additionally include the CET1 capital ratio and Leverage (Tier 1 Leverage Ratio as defined under Basel III norms). Three risk thresholds were heuristically specified for each of the indicators. The menu of regulatory actions was expanded and breach of the riskiest threshold of CET1 ratio by a bank would qualify it for a potential amalgamation, restructuring or even winding up.

Between 2015 to 2019, 12 commercial banks were put under the PCA classification, remaining under business constraints for an average of 12 consecutive quarters before their improved performance warranted a removal from PCA. 11 of the PCA banks were government-owned and 5 of these were ultimately merged with other public sector banks between 2019-2020.

The bank-wise risk cluster assignments are back-tested with the PCA classification of Indian banks. Table 3 below summarises the number of quarters before their PCA classification that these banks were assigned for the first time to the riskiest cluster (Risk1) under each of the four clustering models. The average lead warning time of increasing risk as per the cluster approach, was the highest at 31 quarters for the model M4. The minimum number of quarters of early warning was also highest (18 quarters) for the same model. However, if size was included along with market indicators (M3), both the average and minimum number of quarters of early warning reduced to 24 and 10 quarters, respectively. Cluster models M1 and M2, also underperformed where the minimum number of early warning quarters to PCA was marginally lower than M4.

Table 3: No. of Quarters Prior to First PCA categorization when Banks First Moved to Risk1

	Clusters With Size, Without Market Indicators (M1)	Clusters Without Size or Market Indicators (M2)	Clusters With Size and Market Indicators (M3)	Clusters Without Size, With Market Indicators (M4)
Average	30	29	24	31
Min	0	9	10	18
Max	48	46	38	48

Source: Authors

Table 4 below summarises the number of quarters before moving out of PCA that these banks got upgraded from Risky1 to a lower risk cluster under each clustering model. The higher the number of quarters in this assessment, the earlier the model captures signals of improved / stronger risk profile of the banks vis-à-vis the regulatory approach. However, faster or earlier removal of banks from the riskiest cluster can also indicate that the model is less conservative. We see that the cluster model with size, without market indicators (M1) transfers banks out of the Risk1 cluster 20 quarters on average before the banks have been removed from PCA. This model thus provides the earliest signals of risk-based performance improvement compared to the other models. However, the models M2 (Cluster without size or market indicator) and M4 (cluster without size with market indicator) both have a slightly more conservative early warning signal of improving bank performance, which may be more optimal.

Table 4: No. of Quarters Before Last PCA Categorization when Banks First Moved out of Risk1

	Clusters With Size, Without Market Indicators (M1)	Clusters Without Size or Market Indicators (M2)	Clusters With Size and Market Indicators (M3)	Clusters Without Size, With Market Indicators (M4)
Average	20	18	17	18
Min	12	10	6	10
Max	38	28	31	28

Source: Authors

Thus, overall, several important points emerge from the results. First, cluster-based risk ranking of banks provides better early warning signals of worsening risk as well as advanced indication of improving risk for individual banks, as compared to heuristic regulatory approaches like the PCA framework. Second, defining clusters with accounting and market indicators, but without size indicators, provides the most optimal combination for making timely predictions of changing standalone bank risk.

4.3. Risk Profiles of Public and Private Sector Banks

This segment presents the cluster-based risk profile comparison of government-owned and private banks during three crisis periods – the GFC (2007-2008), the NPA crisis (2014-2015) and the Covid pandemic (2020-2021). The Z-test (given the sample size) is applied to compare the difference in the proportion of riskiest banks (that is, in Risk1 cluster) in both ownership categories, for each clustering model. The results are shown in Table 5

During the GFC, three out of four models show a statistically significant higher proportion of Risk1 for PSBs, compared to PVBs. The NPA Crisis is demarcated by a higher proportion of Risk1 cluster for PSBs, compared to PVBs across all models, at 1% level of significance. The inference is that PSBs are worse performers in terms of size, accounting ratios and market signals, during both episodes. The Covid Pandemic, which was more of a real sector crisis than a financial crisis, has a much lower proportion of riskiest banks as compared to the previous crises. However, in cluster models M1, which considers size but not market data, and M4, which considers market data but not size, the proportion of Risk1 is higher for private sector banks than for PSBs, albeit at a higher significance level.

Table 5: Z-Test for Difference of Proportions in Risk1 Clusters for Public and Private Sector Banks

Crisis / Clustering Model	Private Banks Proportion of Risky 1	Public Banks Proportion of Risky 1	Difference of Proportion	Standard Error	Z-Statistic	P-Value
GFC (2007 - 2008)						
M1	0.3162	0.6730	0.3568	0.0549	6.5038	0.0000
M2	0.2868	0.4313	0.1445	0.0532	2.7151	0.0033
M3	0.2727	0.2189	-0.0538	0.0523	-1.0279	0.1520
M4	0.2455	0.3550	0.1096	0.0568	1.9308	0.0268
NPA Crisis (2014-2015)						
M1	0.4656	0.7681	0.3025	0.0532	5.6834	0.0000
M2	0.2748	0.8357	0.5609	0.0542	10.3425	0.0000
M3	0.0935	0.6985	0.6050	0.0599	10.0976	0.0000
M4	0.4112	0.8744	0.4632	0.0543	8.5357	0.0000
Covid Pandemic (2020-2021)						
M1	0.336	0.255	-0.081	0.057	-1.410	0.079
M2	0	0	Not Applicable			
M3	0.0150	0.0085	-0.0066	0.0137	-0.4776	0.3165
M4	0.0902	0.0424	-0.0479	0.0318	-1.5058	0.0661

Source: Authors

To the best of our knowledge, this study is the first one to use cluster analysis to distinguish between the risk profiles of public and private sector banks, over a phase that spans three crises. The results may encourage policymakers to consolidate public sector banks or merge them with safer private banks. The government of India has already taken concrete steps along both these directions.

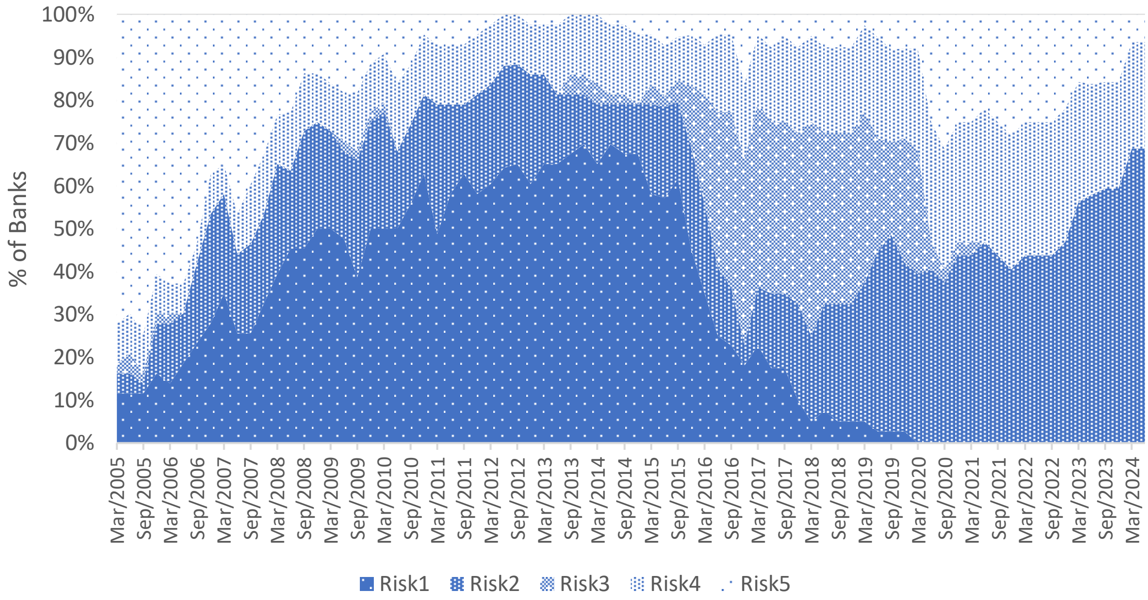
5. Extensions of Cluster Analysis for Incremental Period[3]

In this section, some new results are presented with a bigger sample. It has been extended to include two more quarters until June 2024, which is the latest available RBI data. For the sake of brevity, only two graphs are discussed. The first one (Figure 5) captures the evolution of risk profile based only on accounting ratios of banks, that is, without inclusion of size or market-based indicators. This analysis portrays the inherent strength of banks and bank systems based only on balance sheet quality.

There's a marginal increase in the number of banks in the riskiest cluster (Risk1), compared to Model M2 (Figure 2), after the addition of two quarters of data. The change is expected to be slight, since only two quarters of data should not change the general pattern since 2005. Since FY 2022-23, Indian banks have registered an impressive performance in terms of accounting ratios. When more such data is added, till June 2024, the resultant improvement in risk profiles will lower the mean in each cluster and shift the centroid further away from extreme values. Therefore, stress periods capture a larger number of banks when benchmarked to normal phases. This is captured by the marginal rise in the share of riskiest banks, during the GFC and NPA crises.

[3] This section has benefitted from the comments and suggestions of the editorial team.

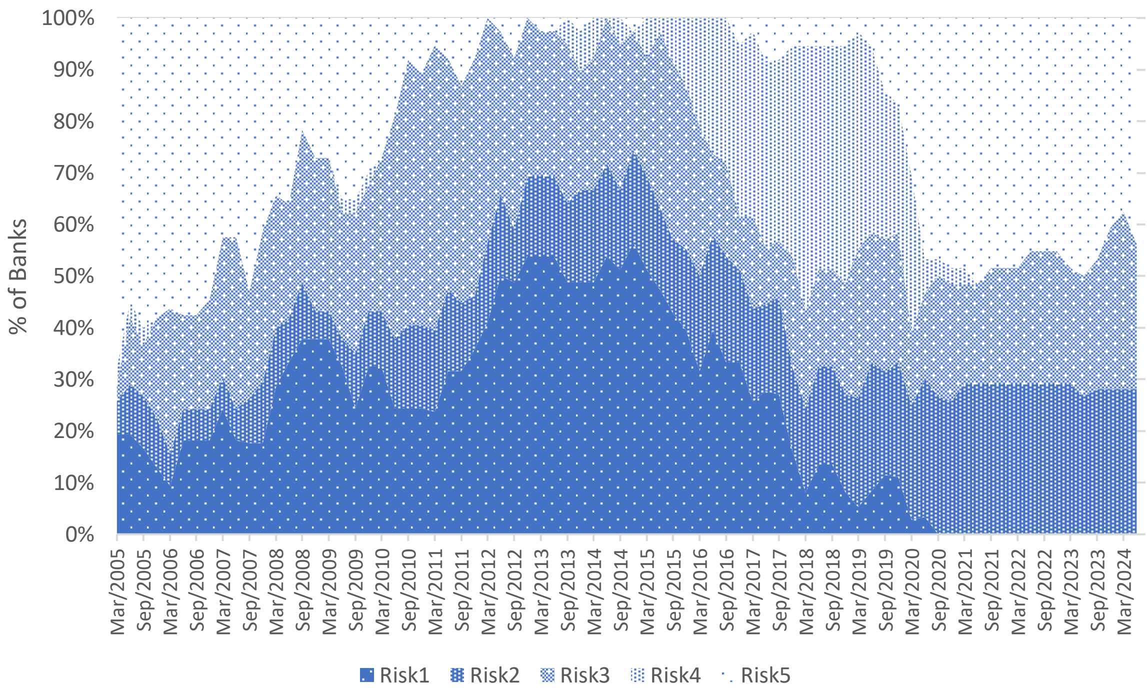
Figure 5: Model M5 Without Size, Without Market Indicators (Up to Jun-2024)



Source: Authors

The next diagram (Figure 6) depicts the bank clusters derived from the interaction of all three factors – accounting, size and market-based indicators – with the larger sample. Once again, a comparison with Figure 3 shows a slight increase in the share of Risk1 banks (the riskiest cluster), particularly during the crisis's periods. The logic is the same as in the previous case.

Figure 6: Model M6 With Size With Market Indicators (Up to Jun-2024)



Source: Authors

It is noteworthy that k-means clustering is known to be sensitive to initial conditions. This is known as the initialization problem (Ikotun et.al., 2023), due to which, addition of new data may result in substantial changes in the outcomes. To avoid this problem, incremental k-means clustering is often employed (Chakraborty & Nagwani, 2011). Therefore, this analysis may be refined with the application of the incremental k-means clustering approach, to account for continuous addition of new data of banks in future periods. This is an important goal for future research, in this direction.

Conclusions

Employing a k-means clustering model, this study analyses the idiosyncratic and systemic risk profiles of Indian banks between 2005 and 2023. The framework demonstrates the extent to which bank size, accounting ratios, and market factors predict three severe shocks, that is, the GFC, NPA crisis, and the Covid pandemic. It is shown that optimal combinations of these variables can forecast the GFC and NPA crisis. However, the pandemic, which was a real sector event, shows up as a concurrent stress. We also show how these drivers interact with each other as they assess the severity of these episodes. The implication is that a multidimensional perspective is more informative than the effect of individual signals. Furthermore, the performance of cluster-based risk ranking is superior to the regulatory monitoring framework in terms of anticipation of both the onset of standalone bank distress and subsequent recovery. The model also suggests that public sector banks were riskier than private sector ones during the GFC and NPA crisis.

The policy implications of this paper are significant. Governments and regulators can use the model to monitor the financial fragility of banks and bank systems, to initiate remedial strategies well before crises occur. In particular, they can enrich the set of indicators used to predict the imposition of the countercyclical capital buffer (CCCB). Under Basel III, CCCB is contingent only on the divergence of the Credit-GDP ratio, from a historical trend. This paper presents a multidimensional tool, which uses a combination of size, market and balance sheet signals, to forecast systemic risk better than the Credit-GDP gap alone. Banks can also utilize the framework to design Pillar II (internal) capital and liquidity buffers for material risks. Appropriate regulatory measures may be undertaken, in light of our results, to improve the risk profiles of public sector banks and their resilience to large shocks.

There is significant scope of future research in this domain. First, in light of the performance gap between public and private sector banks, regulators may contemplate mergers or acquisitions in the Indian banking sector. They may apply scenario analysis and simulation models on this multidimensional framework to forecast the risk performance of the merged entities and the impact of the mergers on systemic risk. In other words, a large number of hypothetical and historical scenarios on accounting ratios, market shocks and size indicators may be generated by regulators and policymakers, to capture the risk implications of possible mergers and acquisitions. Second, the analysis presented in this paper may be replicated for non-bank financial companies (NBFCs). The shadow banking system has emerged as an important source of systemic risk worldwide. A multidimensional surveillance system will enable regulators to unravel the various shades of NBFC vulnerability. It will help them introduce appropriate guidelines, to harmonize risk management standards of both banks and NBFCs.

The most important message of this paper is that financial sector risk, both idiosyncratic and systemic, is kaleidoscopic. Hence, the monitoring system that anticipates financial sector distress should also be multifaceted. For instance, as the results show, market-based signals may capture external stress better, while size-based indicators may predict credit booms and busts with greater efficiency. As the global financial sector becomes more interconnected and interactive, a one size fits all approach to bank supervision may be suboptimal. The present study offers a multidimensional surveillance framework that allows banks, regulators, and policymakers to gauge the diverse sources of bank-specific and systemic risk so that the treatment will be tailored to the nature of the ailment.

Credit Authorship Contribution Statement

Chherawala, T. contributed to the conceptualization of the study, conducted the data analysis, and led the writing of the original draft. Vaidya, A. was involved in the methodology design, validation of results, and contributed to the review and editing of the manuscript. Basu, S. provided supervision, contributed to the theoretical framework, and assisted in the interpretation of results and final manuscript refinement. All authors reviewed and approved the final version of the manuscript.

Conflict of Interest Statement

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Acharya, V., Bhadury, S., Surti, J. (2025). Financial Vulnerability and Risks to Growth in Emerging Markets. *The Indian Economic Journal*, 73(1), 18-56. https://doi.org/10.1177/00194662241292664

Acharya, V., & Kulkarni, N. (2019). Government Guarantees and Bank Vulnerability during a Crisis: Evidence from an Emerging Market. *National Bureau of Economic Research*, Working Paper. http://www.nber.org/papers/w26564

Acharya, V., Pedersen, L., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. *Review of Financial Studies*, 30(1), 2-47. https://doi.org/10.1093/rfs/hhw060

Ayadi, R., Bongini, P., Casu, B., & Cucinelli, D. (2021). Bank Business Model Migrations in Europe: Determinants and Effects. *British Journal of Management*, Volume 32, Issue 4, 1007-1024. https://doi.org/10.1111/1467-8551.12437

BIS (2006). Studies on credit risk concentration, Working Paper No. 15. https://www.bis.org/publ/bcbs_wp15.pdf

Boďa, M., & Zimková, E. (2021). Overcoming the loan-to-deposit ratio by a financial intermediation measure - A perspective instrument of financial stability policy. *Journal of Policy Modelling,* Volume 43, Issue 5, 1051-1069. https://doi.org/10.1016/j.jpolmod.2021.03.012

Bos, J., Lamers, M., Peyrache, A., & Purice, V. (2024). Carrying the (paper) burden: A portfolio view of systemic risk and optimal bank size. *Journal of the Operational Research Society*, 76(4), 607-616. https://doi.org/10.1080/01605682.2024.2383372

Boyd, J., & Runkle, D. (1993). Size and performance of banking firms : Testing the predictions of theory. *Journal of Monetary Economics*, Volume 31(1), 47-67. https://doi.org/10.1016/0304-3932(93)90016-9

Brownlees C., & Engle R. (2017). SRISK: a conditional capital shortfall measure of systemic risk. *The Review of Financial Studies*, Volume 30, Issue 1, 48–79. https://doi.org/10.1093/rfs/hhw088

Carleo, A., & Rocci, R. (2024). Functional clustering of NPLs recovery curves. *Socio-Economic Planning Sciences*, Volume 95(C). https://doi.org/10.1016/j.seps.2024.102018

Chakraborty, S., & Nagwani, N. K. (2011). Analysis and Study of Incremental K-Means Clustering Algorithm. In: Mantri, A., Nandi, S., Kumar, G., Kumar, S. (eds) High Performance Architecture and Grid Computing. HPAGC 2011. *Communications in Computer and Information Science*, Volume 169. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22577-2_46

Conlon T., Cottera J., & Molyneuxb P. (2020). Beyond common equity: The influence of secondary capital on bank insolvency risk. *Journal of Financial Stability*, Volume 47, 100732. https://doi.org/10.1016/j.jfs.2020.100732

Dardac N., Boitan I., (2009), A Cluster Analysis Approach for Bank's Risk Profile: The Romanian Evidence. *European Research Studies Journal, XII(1):* 109-118. https://doi.org/10.35808/ersj/213

Dinger, V., & Von Hagen, J. (2009). Does Interbank Borrowing Reduce Bank Risk? *Journal of Money, Credit and Banking*, 41, 491-506. https://doi.org/10.1111/j.1538-4616.2009.00217.x

Eichengreen, B., & Gupta, P. (2013). The two waves of service-sector growth. *Oxford Economic Papers*, Volume 65, Issue 1, 96–123. https://doi.org/10.1093/oep/gpr059

Faulk, B., Faulk, W., & Lutton, T. (2018) Risk and the probability of insolvency: a regulatory perspective, chapter in "The Most Important Concepts in Finance, Edited by Benton Gup. https://www.federalreserve.gov/SECRS/2019/December/20191210/OP-1681/OP-1681_102219_136090_473235180759_1.pdf

Fiordelisi, F., Ibanez, D., & Molyneux, P. (2011). Efficiency and risk in European banking. European Central Bank Working Paper Series. https://doi.org/10.1016/j.jbankfin.2010.10.005

FSR (2006). European Central Bank. https://www.ecb.europa.eu/pub/pdf/fsr/financialstabilityreview200612en.pdf

Gaul, L., & Jones, J. (2021). CAMELS Ratings and their information content. Office of The Comptroller of the Currency the US Department of the Treasury, First Version, 1(6), 1-37. https://www.occ.treas.gov/publications-and-resources/publications/economics/working-papers-banking-perf-reg/pub-econ-working-paper-camels-ratings.pdf

Hugonnier, J., & Morellec, E. (2017). Bank capital, liquid reserves, and insolvency risk. *Journal of Financial Economics*, Volume 125, Issue 2, 266-285. https://doi.org/10.1016/j.jfineco.2017.05.006

IMF (2006). India: Selected Issues, Staff Country Report No. 06/56, February. https://www.imf.org/external/pubs/ft/scr/2006/cr0656.pdf

IMF (2009). Detecting Systemic Risk. *Global Financial Stability Report, April 2009: Responding to the Financial Crisis and Measuring Systemic Risk*, 111-147. https://www.imf.org/en/Publications/GFSR/Issues/2016/12/31/Global-Financial-Stability-Report-April-2009-Responding-to-the-Financial-Crisis-and-22583

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B. & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178 – 210. https://doi.org/10.1016/j.ins.2022.11.139

Jackson, E. (1991). *A Use's Guide to Principal Components*. John Wiley & Sons, Inc. https://doi.org/10.1002/0471725331

Jain, A., Murty, M., & Flynn, P. (1999). Data clustering: a review. *ACM Computer Surveys (CSUR*), 31, 264-323. https://doi.org/10.1145/331499.331504

João C., Lucas, A., Schaumburg J., & Schwaab, B. (2023). Dynamic clustering of multivariate panel data. *Journal of Econometrics*, Volume 237, Issue 2, Part B. https://doi.org/10.1016/j.jeconom.2022.03.003

Leon, D., Aragón, A., Sandoval, J., Hernández, G., Arévalo, A., & Ni Jño (2017). Clustering algorithms for Risk-Adjusted Portfolio Construction. *Procedia Computer Science*, Volume 108, 1334-1343. https://doi.org/10.1016/j.procs.2017.05.185

Martynova, N., Ratnovski, L., & Vlahu, R. (2015). Bank Profitability and Risk-Taking, IMF Working Paper, WP/15/249.

Mercadier, M., Tarazi, A. Armand, P., & Lardy, J. (2025), Monitoring bank risk around the world using unsupervised learning, *European Journal of Operational Research*, Volume 324, Issue 2, 590-615. https://doi.org/10.1016/j.ejor.2025.01.036

Mohanty, D. (2010). Perspectives on Lending Rates in India, RBI Bulletin, July, 1363 – 1371. https://www.bis.org/review/r100617d.pdf

Mutinda, J., & Langat, A. (2024). Exploring the Role of Dimensionality Reduction in Enhancing Machine Learning Algorithm Performance. *Asian Journal of Research in Computer Science*, Volume 17, Issue 5, 57-166. https://doi.org/10.9734/ajrcos/2024/v17i5445.

Oti, U., Olusula, M., Eze, F., & Enogwe, S. (2021). Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering,* 07(08), 64-69. https://doi.org/10.31695/IJASRE.2021.34050

Oyewole, G. J., & Thopil, G. A. (2023). Data clustering: application and trends. *Artificial Intelligence Review,* Volume 56, 6439–6475. https://doi.org/10.1007/s10462-022-10325-y

RBI (2004). Prudential norms on classification of investment portfolio of banks. RBI/2004-05/157/DBOD.No.BP.BC.37/21.04.141/2004-05, September 2. https://www.corporatelawreporter.com/wp-content/uploads/2013/05/Prudential-norms-on-classification-of-Investment-Portfolio-of-Banks.pdf

Reverchuk, S., Lobozynska, S., & Megits, N. (2013). Evaluating the Efficiency of Banking Systems During a Pre-Crisis and Crisis Period by Using Cluster Analysis (2004-2009). *Sage Open,* 3(4), 1-19. https://doi.org/10.1177/2158244013504575

Sami A., & Tommi K., (2006), Introduction to partitioning-based clustering methods with a robust example. *Reports of the Department of Mathematical Information Technology Series C. Software and Computational Engineering*, No. C. 1/2006. https://doi.org/10.13140/RG.2.2.32908.33927

Sass J., & Thös A., (2024). Risk reduction and portfolio optimization using clustering methods. *Econometrics and Statistics*, Volume 32, 1-16. https://doi.org/10.1016/j.ecosta.2021.11.010

Trebuňa P., Halčinová, J. Fil'o, M., & Markovič, J. (2014). The importance of normalization and standardization in the process of clustering. IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl'any, Slovakia, 381-385. https://doi.org/10.1109/SAMI.2014.6822444

Vashisht, S., Sarva, M. (2024). Risk profiling of Indian commercial banks - a clustering approach. *Afro-Asian Journal of Finance and Accounting,* Volume 14, Issue 6, 854 - 872. https://doi.org/10.1504/AAJFA.2024.142112

Velliangiri, S, Alagumuthukrishnan, S, & Joseph, S., (2019). A Review of Dimensionality Reduction Techniques for Efficient Computation. *Procedia Computer Science*, Volume 165, 104–111. https://doi.org/10.1016/j.procs.2020.01.079

Velliscig, G., Floreani, J., & Polato, M. (2023). Capital and asset quality implications for bank resilience and performance in the light of NPLs' regulation: a focus on the Texas ratio. *Journal of Banking Regulation*, 24(1), 66-88. https://doi.org/10.1057/s41261-021-00184-y

Yuan, C., & Yang, H., (2019), Research on K-Value Selection Method of K-Means Clustering Algorithm. *J*, 2(2), 226-235. https://doi.org/10.3390/j2020016