IOCMA 2026 Conference
The 2nd International Online Conference on Mathematics and Applications
10–12 June 2026 | Online
MDPI

Regional and Sectoral Heterogeneity in Corporate Insolvency: Evidence for Spain Using Fixed Effects Models and Machine Learning

Yu Zhang
IUMPA. Universitat Politècnica de València

INTRODUCTION & AIM

Corporate insolvency dynamics reflect the interaction between firm-level fragility and macro-financial conditions, shaped by strong regional and sectoral heterogeneity. While classical models highlight the predictive power of financial ratios (e.g., liquidity, leverage, profitability), they largely abstract from systemic and spatial factors that significantly influence failure patterns.
This study adopts a regional–sectoral panel framework for Spain (2008–2024) to capture these dimensions. The empirical strategy integrates three complementary approaches:
Two-Way Fixed Effects (TWFE): identifies structural relationships while controlling for time-invariant regional and sectoral heterogeneity, as well as common aggregate shocks.
Machine Learning (ML): enhances predictive performance and captures nonlinear interactions in complex, imbalanced environments.
Clustering (Fuzzy C-Means): uncovers latent group structures by classifying regions and sectors according to their response to macro-financial shocks.
This integrated framework combines interpretability (econometrics), predictive accuracy (ML), and structural segmentation (clustering), providing a unified applied mathematics approach to studying insolvency as a spatially and sectorally heterogeneous phenomenon.

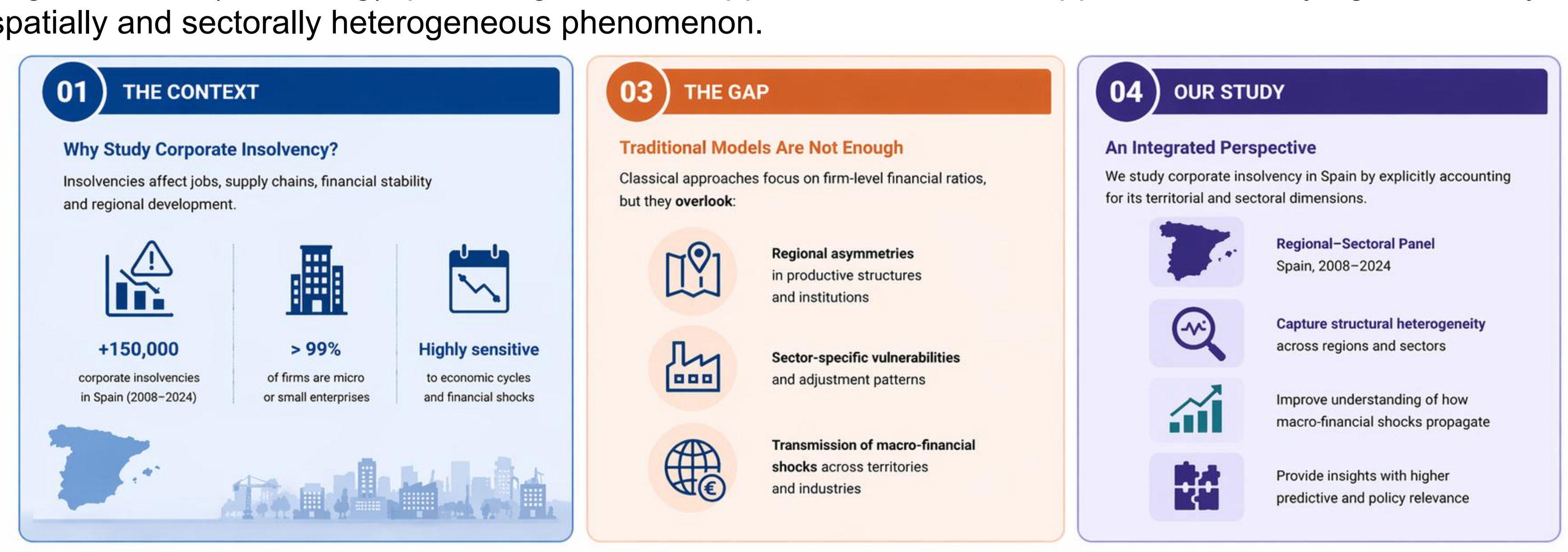


Figure 1. Conceptual Structure of Corporate Insolvency Dynamics

METHOD

The analysis is based on a balanced regional–sectoral panel for Spain (2008–2024) at the level (r, s, t). Financial data (returns and volatility of major indices and Spanish banks) are obtained from Yahoo Finance, while macroeconomic variables (CPI, unemployment, NPL) and insolvency data come from INE and the Bank of Spain. All series are harmonised to annual frequency. Regions are standardised to 18 units and sectors to 12 CNAE groups. For 2021–2024, missing sectoral detail is reconstructed using historical weights from pre-2020 data, ensuring consistency while preserving the underlying sectoral structure.
The dependent variable is the insolvency rate, transformed as log(1+IR). Aggregate financial conditions are summarised by a Market Risk Index (MRI), constructed as a standardised combination of returns and volatility (r×σ), capturing both the magnitude and direction of market stress. This formulation allows the index to distinguish between contractionary and expansionary regimes, which is crucial for interpreting its effect on insolvency dynamics. The resulting annual index is expressed in standard deviation units and used as a common macro-financial shock.
The structural analysis relies on a two-way fixed effects model, decomposing insolvency dynamics into region–sector heterogeneity and common time effects. This framework identifies differential sensitivities to aggregate shocks across units while controlling for persistent structural differences. Estimation is performed using weighted least squares, with Driscoll–Kraay standard errors to account for cross-sectional dependence, serial correlation and heteroskedasticity.
A predictive layer based on gradient boosting models (CatBoost, XGBoost, LightGBM) is implemented using an expanding-window scheme, capturing nonlinearities and improving out-of-sample risk detection. This setup preserves the temporal structure of the data and allows model performance to be evaluated under changing economic conditions.
Fuzzy C-Means clustering is used to identify latent vulnerability profiles across regions and sectors, revealing three stable groups. The fuzzy framework allows partial membership, reflecting gradual differences across units rather than imposing sharp boundaries.

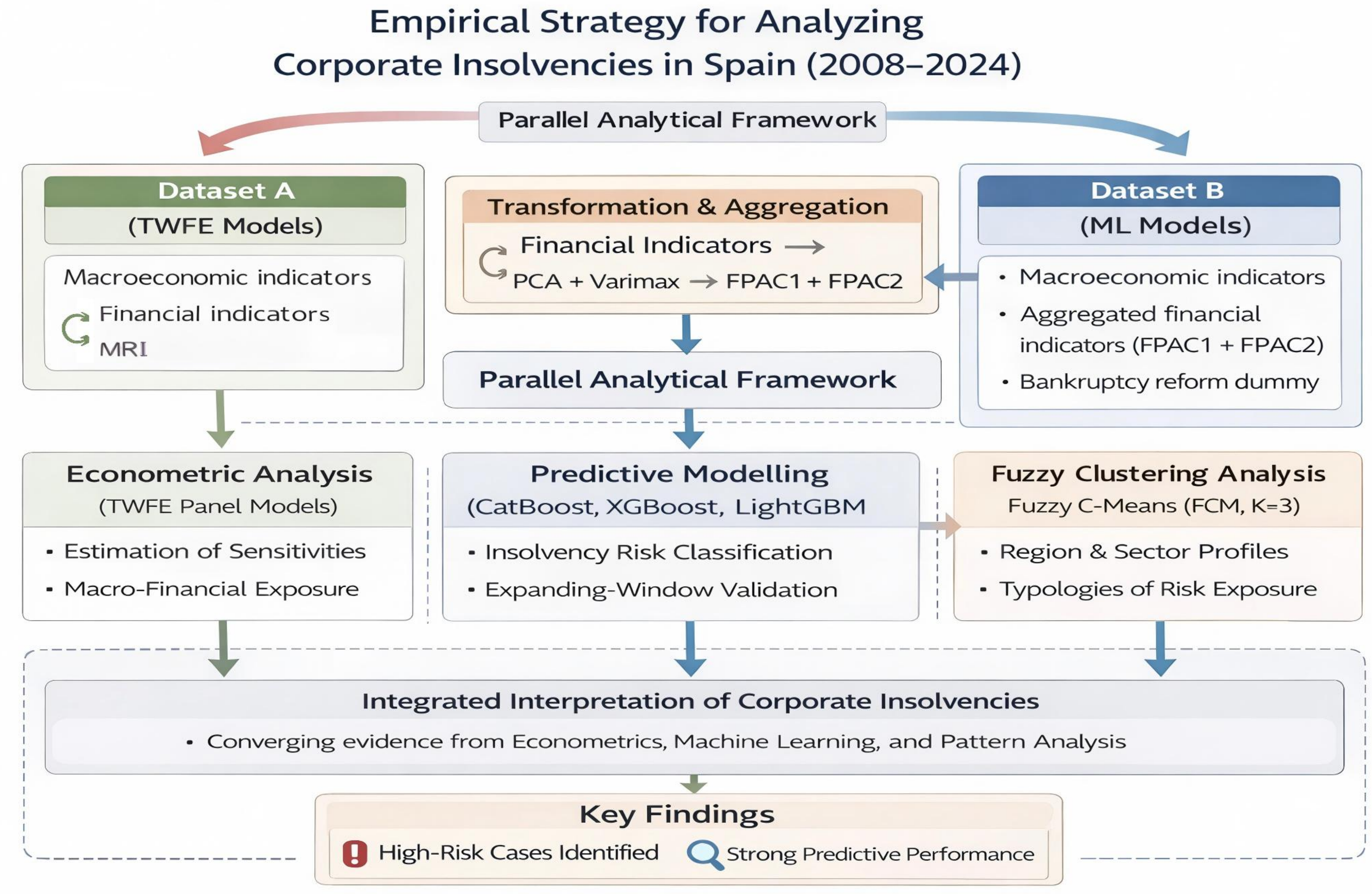


Figure 2. Integrated Modelling Framework

RESULTS & DISCUSSION

Under normal conditions, higher financial stress (MRI) is associated with higher insolvency rates, but this relationship is not stable over time. During the pandemic, insolvency remained unusually low despite elevated stress, reflecting policy interventions that delayed firm exits. The subsequent increase suggests a postponed adjustment effect rather than an immediate response.
Panel estimates reveal strong heterogeneity in the transmission of financial shocks. Sectoral effects are generally weaker, while regional differences are more pronounced, indicating uneven exposure across territories. Moving beyond a single aggregate indicator, decomposing financial conditions into orthogonalised shocks significantly improves explanatory power (within (R^2): 0.01→0.17)., highlighting the role of multiple transmission channels.

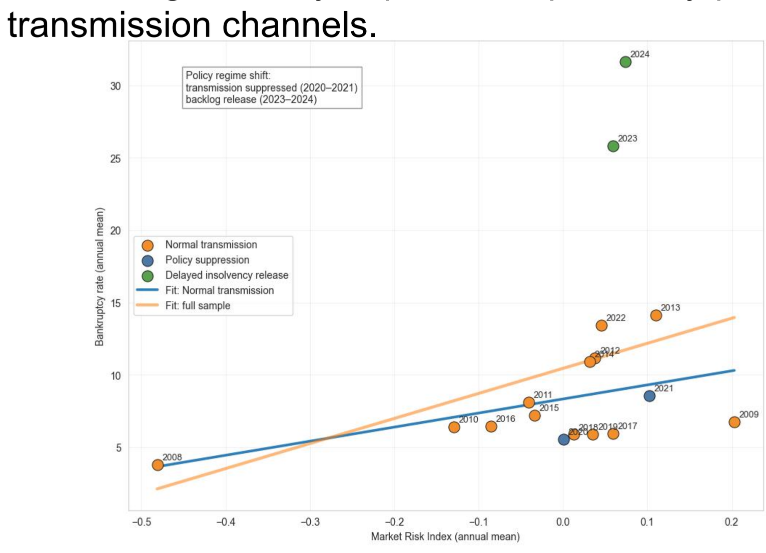


Figure 3. Financial Stress–Insolvency Relationship with Policy-Induced Regime Shifts

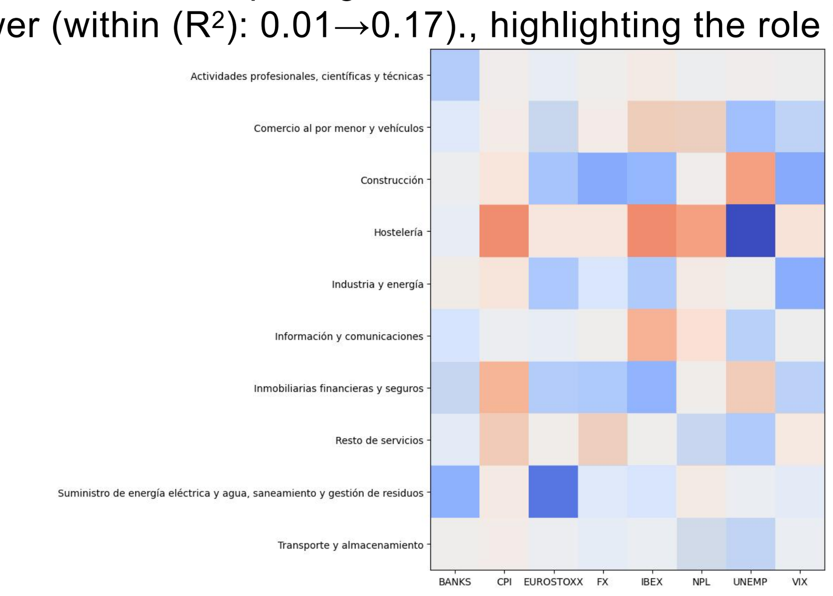
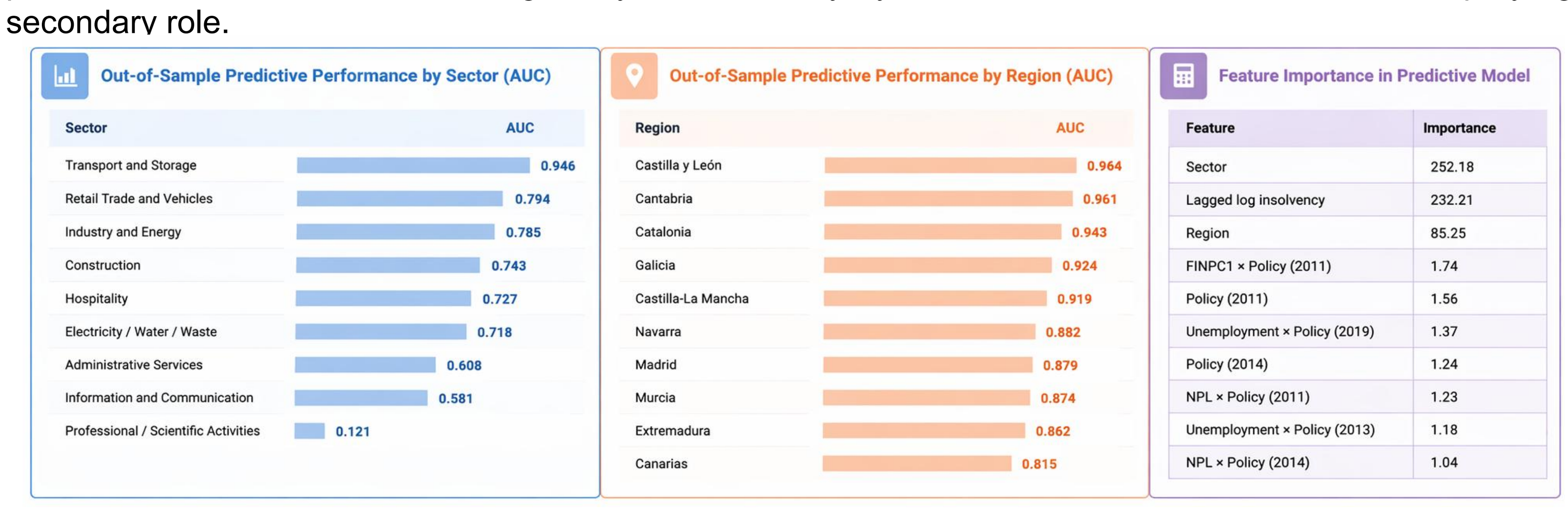


Figure 4. Sectoral sensitivities to orthogonalised shocks

A principal component approach summarises financial conditions into two factors explaining around 88% of variance. FINPC1 (equity and banking stress) drives positive responses in most sectors—especially construction, retail and services—whereas FINPC2 (global risk and exchange rate) plays a more limited role under normal conditions but becomes relevant at the regional level and under high stress. This suggests sectoral dominance in normal times and regional amplification under stress.
Predictive analysis confirms these patterns. Gradient boosting models show stable performance overall, with differences emerging after 2022 due to structural shifts. Accuracy is higher in financially exposed sectors and lower in knowledge-intensive activities, while regional differences are less pronounced. Feature importance shows that structural variables dominate prediction (sector, lagged insolvency, region), reflecting strong persistence and structural heterogeneity in insolvency dynamics, with macro-financial factors playing a secondary role.



Out-of-Sample Predictive Performance by Sector (AUC)

Sector	AUC
Transport and Storage	0.946
Retail Trade and Vehicles	0.794
Industry and Energy	0.785
Construction	0.743
Hospitality	0.727
Electricity / Water / Waste	0.718
Administrative Services	0.608
Information and Communication	0.581
Professional / Scientific Activities	0.121

Out-of-Sample Predictive Performance by Region (AUC)

Region	AUC
Castilla y León	0.964
Cantabria	0.961
Catalonia	0.943
Galicia	0.924
Castilla-La Mancha	0.919
Navarra	0.882
Madrid	0.879
Murcia	0.874
Extremadura	0.862
Canarias	0.815

Feature Importance in Predictive Model

Feature	Importance
Sector	252.18
Lagged log insolvency	232.21
Region	85.25
FINPC1 × Policy (2011)	1.74
Policy (2011)	1.56
Unemployment × Policy (2019)	1.37
Policy (2014)	1.24
NPL × Policy (2011)	1.23
Unemployment × Policy (2013)	1.18
NPL × Policy (2014)	1.04

Figure 5. Predictive Performance and Feature Importance

Clustering further identifies three stable groups for both sectors and regions. Differences arise from the interaction between structural characteristics and financial exposure, separating financially sensitive activities from service sectors, and more resilient regions from those more exposed to shocks.
Insolvency emerges as a structured and uneven response to financial stress, where underlying economic configurations shape both the intensity and timing of firm failures across sectors and regions.

Panel A. Sectors		Panel B. Regions	
Cluster	Sectors	Cluster	Regions
0	Construction; Industry and Energy; Information and Communications	0	Cantabria; Castile and León; Castile–La Mancha; Galicia; Navarra; Basque Country
1	Administrative and Support Services; Professional, Scientific and Technical Activities; Real Estate, Financial and Insurance Activities; Electricity, Water Supply, Sanitation and Waste Management; Transport and Storage	1	Andalusia; Canary Islands; Ceuta and Melilla; Extremadura; Balearic Islands; La Rioja
2	Retail Trade and Vehicles; Hospitaility; Other Services	2	Aragon; Asturias; Catalonia; Valencian Community; Madrid; Murcia

Table 1. Cluster composition by sector and region

CONCLUSION

Corporate insolvency is a heterogeneous, system-driven process shaped by macro-financial shocks and structural characteristics.The relationship between financial stress and insolvency is time-varying, with policy interventions generating delayed adjustments. Transmission is uneven: sectoral exposure dominates in normal periods, while regional heterogeneity intensifies under stress.Multiple financial channels drive insolvency dynamics, while structural factors (sector, persistence, region) dominate predictive performance. Clustering identifies distinct vulnerability profiles across sectors and regions.
Overall, insolvency risk is nonlinear and evolving, requiring disaggregated approaches beyond aggregate indicators.

FUTURE WORK / REFERENCES

Future research could extend this framework by improving the identification of causal mechanisms and incorporating richer data structures. In particular, combining macro-financial indicators with firm-level information and exploring dynamic or network-based approaches would help better capture the propagation of shocks and enhance the predictive and interpretative power of the analysis.