

HERA-XAI

**Explaining One-Year-Ahead Corporate
Rating Forecasts Through
Hierarchical Artificial Intelligence**



6/2026

Greta Falavigna

<i>Direttore</i>	Elena Ragazzi
<i>Direzione</i>	CNR-IRCrES <i>Istituto di Ricerca sulla Crescita Economica Sostenibile</i> Strada delle Cacce 73, 10135 Torino, Italy Tel. +39 011 3977612 segreteria.ircres@cnr.it www.ircres.cnr.it
<i>Sede di Roma</i>	Via dei Taurini 19, 00185 Roma, Italy Tel. +39 06 49937809 / Fax +39 06 49937808
<i>Sede di Milano</i>	Via Corti 12, 20121 Milano, Italy Tel. +39 02 23699501 / Fax +39 02 23699530
<i>Sede di Genova</i>	Corso Ferdinando Maria Perrone 24, 16152 Genova, Italy Tel. +39 010 6598798

Comitato di Redazione

Grazia Biorci, Igor Benati, Antonella Emina, Serena Fabrizio, Lucio Morettini, Susanna Paleari, Anna Perin, Emanuela Reale, Secondo Rolfo, Andrea Orazio Spinello, Isabella Maria Zoppi.



<https://www.ircres.cnr.it/produzione-scientifica/>

I Working Paper del CNR-IRCrES rappresentano il punto di vista degli autori e non riflettono necessariamente quello dell'Istituto nel suo complesso.

The Working Papers published by CNR-IRCrES represent the views of the respective author(s) and not of the Institute as a whole.

CNR-IRCrES Working Paper 6/2026



ottobre 2026 by CNR-IRCrES

HERA-XAI: EXPLAINING ONE-YEAR-AHEAD CORPORATE RATING FORECASTS THROUGH HIERARCHICAL ARTIFICIAL INTELLIGENCE

GRETA FALAVIGNA

CNR-IRCrES, Consiglio Nazionale delle Ricerche, Istituto di Ricerca sulla Crescita Economica Sostenibile, Economica Sostenibile, Strada delle Cacce 73, 00135 Torino, Italia

corresponding author: greta.falavigna@cnr.it

ABSTRACT

Corporate rating forecasting models are often characterized by a trade-off between predictive performance and interpretability. While advanced machine learning algorithms can generate accurate forecasts, understanding the financial mechanisms underlying their predictions remains a significant challenge. This paper presents HERA-XAI, an explainable extension of the Hierarchical Artificial Intelligence framework for one-year-ahead corporate rating forecasting. Building upon the sector-specific hierarchical architecture introduced in HERA, the proposed framework combines Random Forest classifiers with a comprehensive explainability layer based on variable importance measures, Partial Dependence Plots (PDPs), Shapley Additive Explanations (SHAP), and simulation-based analyses.

The analysis is conducted on a large dataset of Italian firms and considers three macro-sectors: Production, Trade, and Services. First, alternative probability threshold calibration criteria are investigated within the hierarchical classification process. Second, predictor importance rankings are extracted across the hierarchy to identify the financial indicators most relevant to rating assignment. Third, PDPs are employed to characterize the direction, magnitude, and non-linear effects of key financial variables on predicted outcomes. Fourth, local and global SHAP analyses are used to quantify the contribution of individual predictors to both firm-level classifications and overall model behaviour. Finally, scenario simulations are performed to evaluate how changes in firms' financial conditions propagate through the hierarchical architecture and affect future rating assignments.

The results show that the integration of hierarchical classification and explainable artificial intelligence provides transparent and economically interpretable rating forecasts while preserving predictive effectiveness. By combining forecasting, explainability, and simulation capabilities within a unified framework, HERA-XAI allows researchers, financial institutions, and policymakers to identify the financial drivers of rating transitions, understand the mechanisms underlying machine learning predictions, and explore the potential consequences of alternative financial scenarios.

KEYWORDS: Corporate Ratings, Explainable Artificial Intelligence, SHAP, Partial Dependence Plots, Hierarchical Classification.

DOI: 10.23760/2421-7158.2026.006

HOW TO CITE

Falavigna, G. (2026). *HERA-XAI: Explaining One-Year-Ahead Corporate Rating Forecasts Through Hierarchical Artificial Intelligence* (CNR-IRCrES Working Paper 6/2026). Istituto di Ricerca sulla Crescita Economica Sostenibile. <http://dx.doi.org/10.23760/2421-7158.2026.006>

SUMMARY

1.	INTRODUCTION	5
2.	LITERATURE REVIEW	6
2.1.	Corporate Rating Forecasting and Machine Learning	6
2.2.	Explainable Artificial Intelligence in Finance	6
2.3.	Partial Dependence Plots and SHAP Explanations	7
2.4.	Research Gap and Contribution	8
3.	DATA AND METHODS	8
3.1.	Dataset and Financial Indicators	8
3.2.	Methodological framework	9
3.2.1.	The HERA Framework	9
3.2.2.	Preliminary Assessment of Random Forest Size	11
3.2.3.	Node-Level Threshold Optimization	12
3.2.4.	End-to-End HERA Simulation	15
4.	HERA 2.0 EXPLAINABILITY ANALYSIS	17
4.1.	Predictor Importance Analysis	17
4.1.1.	Methodology	17
4.1.2.	Results	18
4.1.3.	Interpreting Predictor Importance within the HERA Hierarchy	21
4.2.	Partial Dependence Plots (PDP)	22
4.2.1.	Methodology	22
4.2.2.	Results	23
4.3.	SHAP Explanations	26
4.3.1.	Local SHAP	26
4.3.2.	Methodology	27
4.3.3.	Results	27
4.3.4.	Global SHAP	28
4.3.5.	Methodology	28
4.3.6.	Results	29
5.	ILLUSTRATIVE POLICY SIMULATION	35
5.1.	Methodology	35
5.2.	Results	36
6.	INTERPRETING EXPLAINABILITY RESULTS: SCOPE AND LIMITATIONS	41
7.	CONCLUSION AND FUTURE DEVELOPMENT	42
8.	SOFTWARE STATEMENT	43
9.	REFERENCES	43
10.	APPENDIX A	44
11.	APPENDIX B	50

1. INTRODUCTION

The growing availability of corporate financial data and the increasing adoption of artificial intelligence techniques have significantly expanded the range of tools available for corporate risk assessment and rating forecasting. In this context, forecasting models are increasingly expected not only to provide accurate predictions but also to offer transparent and interpretable explanations capable of supporting managerial, financial, and policy decisions.

This study builds upon the HERA framework proposed by Falavigna (2026) and develops a comprehensive interpretability procedure for hierarchical corporate rating forecasting models. HERA (Hierarchical Engine for Rating Assessment) demonstrated that one-year-ahead corporate ratings can be effectively predicted through a sector-specific hierarchical structure composed of consecutive Random Forest classifiers. The hierarchical architecture showed that rating transitions can be represented as a sequence of binary classification problems and highlighted the existence of significant differences across Production, Trade, and Services sectors.

The HERA framework itself originated from earlier research by Falavigna (2012), where Artificial Neural Networks were employed to assign financial ratings to firms using accounting and balance-sheet information. Consequently, HERA inherited the same set of financial indicators originally identified in that study, allowing the comparison of alternative artificial intelligence approaches while preserving a consistent economic interpretation of the explanatory variables.

The original HERA framework was primarily developed as a forecasting architecture. Nevertheless, it also introduced preliminary elements of model interpretability through predictor importance measures and illustrative Partial Dependence Plot (PDP) analyses. These initial investigations provided useful evidence regarding the financial variables associated with rating transitions and demonstrated the potential of explainability techniques within a hierarchical forecasting environment. However, interpretability was not yet developed as a systematic component of the framework and several important questions remained open. In particular, limited attention was devoted to threshold calibration, the comprehensive analysis of rating drivers across all hierarchical nodes, and the decomposition of model predictions into variable-specific contributions.

The purpose of the present study is therefore to extend the original framework and introduce HERA-XAI, an explainable evolution of the HERA architecture. Rather than replacing the original forecasting structure, HERA-XAI builds upon the already validated hierarchical framework and complements it with a comprehensive explainability layer. The proposed extension pursues four main objectives. First, it investigates the calibration of the hierarchical forecasting process through the identification of optimal classification thresholds. Second, it evaluates the relative importance of financial indicators within each node of the hierarchy. Third, it examines the functional relationship between key predictors and forecasted rating outcomes through an extensive application of Partial Dependence Plots (PDPs). Finally, it introduces both local and global Shapley Additive Explanations (SHAP) to quantify the contribution of individual predictors to model predictions and to identify the most influential financial drivers underlying rating transitions.

In this perspective, HERA-XAI transforms HERA from a forecasting architecture with preliminary interpretability features into a fully explainable rating forecasting framework. By integrating forecasting accuracy with explainable machine learning techniques, the proposed architecture provides useful insights for firms seeking to improve their financial profiles, financial institutions interested in understanding rating dynamics, and policymakers evaluating the potential effects of economic and financial interventions on corporate performance.

More generally, HERA-XAI contributes to the growing literature on Explainable Artificial Intelligence (XAI) by demonstrating how threshold calibration procedures, predictor importance analyses, Partial Dependence Plots, and SHAP explanations can be systematically integrated into a hierarchical forecasting architecture without compromising predictive effectiveness. The

resulting framework represents a bridge between predictive analytics, financial decision-support systems, and explainable machine learning, thereby providing a richer understanding of both corporate rating forecasts and their underlying economic determinants.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature on corporate rating forecasting, hierarchical classification models, and explainable artificial intelligence. Section 3 presents the dataset and the methodological framework. Section 4 discusses the explainability analyses, including predictor importance rankings, Partial Dependence Plots, and SHAP explanations. Section 5 introduces an illustrative policy simulation exercise. Finally, Section 6 summarizes the main findings and outlines future developments of the HERA framework, with particular emphasis on scenario analysis and decision-support applications. Two appendices provide additional empirical results and supporting analyses.

2. LITERATURE REVIEW

2.1. Corporate Rating Forecasting and Machine Learning

Corporate credit ratings represent one of the most widely used indicators of firms' creditworthiness and play a fundamental role in financing decisions, investment allocation, risk management, and regulatory supervision. Consequently, forecasting future rating changes has attracted increasing attention within both academic and professional communities.

Traditional approaches to rating assessment were primarily based on statistical techniques such as discriminant analysis, logistic regression, and other econometric models. Although these methods provided relatively transparent interpretations of financial risk, their ability to capture nonlinear relationships and complex interactions among accounting indicators was often limited. The growing availability of large financial datasets and increasing computational power have therefore stimulated the adoption of machine learning techniques for corporate rating forecasting.

Recent studies show that algorithms such as Random Forests, Support Vector Machines, Gradient Boosting Machines, and Artificial Neural Networks frequently outperform conventional statistical approaches in terms of predictive accuracy. However, these predictive gains are often accompanied by a substantial reduction in interpretability, giving rise to the so-called black-box problem. In credit risk applications, this lack of transparency may limit the practical usefulness of machine learning models, particularly when predictions affect financial decisions involving investors, lenders, regulators, and policymakers. Recent research therefore emphasizes the need to reconcile predictive performance with model transparency and explainability (Ghosh et al., 2024; Oliveira and Basso, 2025).

Within this context, Falavigna (2026) introduced the HERA framework (Hierarchical Engine for Rating Assessment), a sector-specific hierarchical architecture based on consecutive Random Forest classifiers for one-year-ahead corporate rating forecasting. The framework demonstrated that complex multiclass rating assignments can be decomposed into a sequence of binary classification tasks while preserving satisfactory forecasting performance and highlighting significant differences across Production, Trade, and Services sectors¹. Nevertheless, the original framework was mainly oriented toward prediction, leaving open the issue of understanding the economic mechanisms underlying model decisions.

2.2. Explainable Artificial Intelligence in Finance

¹ The Production sector includes firms with 2-digits NACE codes from 01 to 43, covering primary activities, mining and quarrying, manufacturing industries, utilities, and construction activities. The Trade sector includes firms with 2-digits NACE codes from 45 to 47, corresponding to wholesale and retail trade and the repair of motor vehicles. The Services sector includes firms with 2 digits NACE codes from 49 to 99.

The increasing diffusion of artificial intelligence and machine learning models within the financial industry has generated a growing demand for explainability. Financial decisions often involve substantial economic consequences, and stakeholders are rarely willing to rely on models whose internal logic cannot be understood or justified. As a consequence, Explainable Artificial Intelligence (XAI) has emerged as one of the most rapidly expanding research areas at the intersection between finance and data science. Recent literature reviews highlight the growing importance of XAI techniques in financial decision-making contexts characterized by high accountability requirements, regulatory constraints, and operational risk (Černevičienė and Kabašinskas, 2024; Mohsin and Nasim, 2025).

Explainable Artificial Intelligence can be broadly defined as a collection of methodologies aimed at making machine learning models understandable to human users. Existing approaches can be divided into intrinsically interpretable models and post-hoc explanation methods. The first category includes models whose decision rules are directly interpretable, whereas the second seeks to explain complex predictive algorithms after model estimation. Since highly performing methods such as Random Forests and Gradient Boosting Machines are inherently difficult to interpret, post-hoc explanation techniques have become the dominant approach in many financial applications. Current evidence suggests that feature importance measures, Partial Dependence Plots (PDPs), LIME, and SHAP are among the most frequently adopted explainability tools in finance (Mohsin and Nasim, 2025; Marathe, 2024).

The relevance of explainability extends beyond purely technical considerations. Transparent models enhance trust, facilitate validation processes, improve communication between analysts and decision-makers, and contribute to regulatory compliance. More importantly, explainability techniques can transform predictive algorithms from simple forecasting tools into decision-support systems capable of generating actionable economic insights. This aspect is particularly important in corporate rating forecasting, where understanding the determinants of future rating transitions is often as valuable as the prediction itself (Černevičienė and Kabašinskas, 2024; Marathe, 2024).

2.3. Partial Dependence Plots and SHAP Explanations

Among the numerous explainability techniques proposed in recent years, Partial Dependence Plots (PDPs) and Shapley Additive Explanations (SHAP) have become particularly influential.

Partial Dependence Plots were originally introduced as a tool for interpreting nonlinear machine learning models by describing the marginal effect of individual predictors on the model output (Friedman, 2001). PDPs provide a global perspective on model behavior by estimating how predicted probabilities change as a selected variable varies across its empirical distribution while all remaining predictors maintain their observed values. This approach allows researchers to identify nonlinear patterns, threshold effects, saturation phenomena, and sector-specific sensitivities that cannot be inferred from variable importance measures alone.

A more comprehensive explanation can be obtained through SHAP values. Introduced by Lundberg and Lee (2017), the SHAP methodology applies concepts from cooperative game theory to machine learning interpretation. In this framework, predictors are viewed as players contributing to the final prediction, while the model output represents the collective payoff. The contribution of each predictor is quantified through its Shapley value, calculated by averaging its marginal contribution across all possible predictor combinations (Lundberg et al., 2020; Rudin, 2019).

A major advantage of SHAP lies in its ability to provide both local and global explanations. Local SHAP values explain individual predictions by decomposing the forecast associated with a specific observation into additive contributions attributable to each predictor. Global SHAP measures are instead obtained by aggregating local explanations across multiple observations, thereby identifying the variables that exert the greatest overall influence on model behavior. Because SHAP satisfies desirable theoretical properties such as local accuracy, consistency, and

additivity, it has become one of the most widely adopted explainability frameworks in contemporary machine learning research (Černevičienė and Kabašinskas, 2024; Marathe, 2024).

Recent applications to credit risk assessment and corporate rating prediction demonstrate that SHAP can reveal heterogeneous drivers of rating transitions across sectors and risk categories while preserving the predictive advantages of machine learning algorithms. By simultaneously identifying the magnitude and direction of predictor contributions, SHAP analyses provide a natural bridge between predictive performance and economic interpretation (Ghosh et al., 2024; Oliveira and Basso, 2025).

2.4. Research Gap and Contribution

Although substantial progress has been achieved in both corporate rating forecasting and Explainable Artificial Intelligence (XAI), the two streams of literature have largely evolved independently. Most studies on rating forecasting primarily focus on predictive performance, whereas explainability research has generally concentrated on generic credit-risk applications or non-hierarchical machine learning architectures. As a result, limited attention has been devoted to the development of integrated frameworks capable of combining forecasting accuracy, model interpretability, and scenario analysis within a unified rating forecasting environment.

While the original HERA framework introduced a sector-specific hierarchical architecture and provided preliminary insights into the financial drivers of rating transitions through predictor importance measures and illustrative PDP analyses, its primary focus remained on forecasting performance. Consequently, a comprehensive explainability framework capable of systematically interpreting model predictions and supporting forward-looking simulations was still lacking.

HERA-XAI addresses this gap by extending the original HERA architecture from a forecasting framework into an explainable and simulation-oriented decision-support system. The proposed approach integrates hierarchical Random Forest classification, threshold calibration procedures, permutation-based predictor importance measures, Partial Dependence Plots, local and global SHAP analyses, and scenario-based simulations. By jointly exploiting these complementary tools, HERA-XAI enables the identification of the financial drivers of rating transitions, explains how such drivers influence model predictions, and assesses the potential impact of alternative financial scenarios on future rating assignments.

The contribution of the paper is therefore twofold. From a methodological perspective, it proposes one of the first integrated explainability frameworks specifically developed for hierarchical corporate rating forecasting systems. From an applied perspective, it provides a practical environment in which researchers, financial institutions, firms, and policymakers can not only forecast future ratings but also understand the mechanisms underlying model predictions and explore the consequences of hypothetical changes in firms' financial conditions.

3. DATA AND METHODS

3.1. Dataset and Financial Indicators

HERA-XAI uses the same database employed in the original HERA framework (Falavigna, 2026). The dataset consists of Italian firms operating in the Production, Trade, and Services sectors and includes balance-sheet information used to construct firm-specific financial ratings and one-year-ahead rating transitions.

To ensure comparability with previous research, the analysis relies on the same set of financial indicators originally adopted in Falavigna (2012) and subsequently incorporated into HERA. These variables capture different dimensions of firms' financial structure, profitability, indebtedness, and asset composition. Table 1 summarizes the financial indicators inherited from

HERA (Falavigna, 2026) and originally introduced in Falavigna (2012). The table reports the variable name, the corresponding financial ratio, and its economic interpretation.

Table 1. Financial indicators used in the HERA Framework

Indicator	Financial Ratio	Economic Interpretation
DEBTA	Total Debt / Total Assets	Measures the overall level of indebtedness relative to total assets. Higher values indicate greater financial leverage and a stronger reliance on external financing.
DEBPROD	Total Debt / Production Value	Measures the amount of debt supported by the firm's productive activity. Higher values may indicate a heavier debt burden relative to operating capacity.
ROS	Operating Income / Production Value	Return on Sales. Measures operating profitability and the firm's ability to generate earnings from sales activities.
ROA	Operating Income / Total Assets	Return on Assets. Evaluates the profitability generated by the firm's asset base and the efficiency with which assets are employed.
DEBCOV	Operating Income / Total Debt	Debt Coverage Ratio. Measures the firm's ability to generate operating earnings relative to total debt. Higher values indicate stronger debt-servicing capacity.
FAPROD	Fixed Assets / Production Value	Fixed Assets to Production Ratio. Captures the intensity of fixed capital employed in productive activities. Higher values generally characterize more capital-intensive firms.
FCCOV	Financial Charges / Operating Income	Financial Charges Coverage Ratio. Measures the burden of financial expenses relative to operating earnings. Higher values may signal greater financial pressure.
FAEQ	Fixed Assets / Net Equity	Fixed Assets to Equity Ratio. Evaluates the extent to which fixed investments are financed by long-term equity sources. Lower values generally indicate a more balanced financial structure.

3.2. Methodological framework

3.2.1. The HERA Framework

The HERA-XAI framework builds upon the Hierarchical Engine for Rating Assessment (HERA) introduced in Falavigna (2026). The original framework was designed to forecast one-year-ahead corporate ratings through a hierarchical architecture composed of sector-specific Random Forest classifiers. Rather than treating rating assignment as a single multiclass classification problem, HERA decomposes the forecasting process into a sequence of binary classification tasks organized within a decision tree structure.

The hierarchical architecture consists of nine interconnected classification nodes. At each node, a Random Forest model is estimated to discriminate between two groups of rating classes. The resulting classification determines the path followed by the observation through the hierarchy until a terminal node is reached and a final rating category is assigned. This approach reduces the complexity of the original forecasting problem, allowing each classifier to focus on a more homogeneous subset of firms while preserving the overall structure of the rating system.

To facilitate the interpretation of the hierarchical architecture, Figure 1 provides a simplified representation of the HERA framework and the sequence of binary classification tasks performed at each node. The figure is adapted from Falavigna (2026) and highlights how the original multiclass rating problem is decomposed into a set of consecutive binary decisions. At each stage, firms are assigned either to Class 0 or Class 1 and subsequently routed to the corresponding branch of the hierarchy until a final rating category is reached. Throughout the paper, the terms *Node* and *Random Forest classifier* (RF) are used interchangeably. Each node of the HERA hierarchical architecture is implemented through a dedicated Random Forest model; consequently, Node 1 corresponds to RF1, Node 2 to RF2, and so forth until Node 9 (RF9). For the sake of exposition, the term *node* is generally adopted when referring to the hierarchical

structure of the framework, whereas *RF* is used when discussing model estimation, predictive performance, and explainability analyses. Since the two terms refer to the same component of the architecture, they should be considered equivalent throughout the paper.

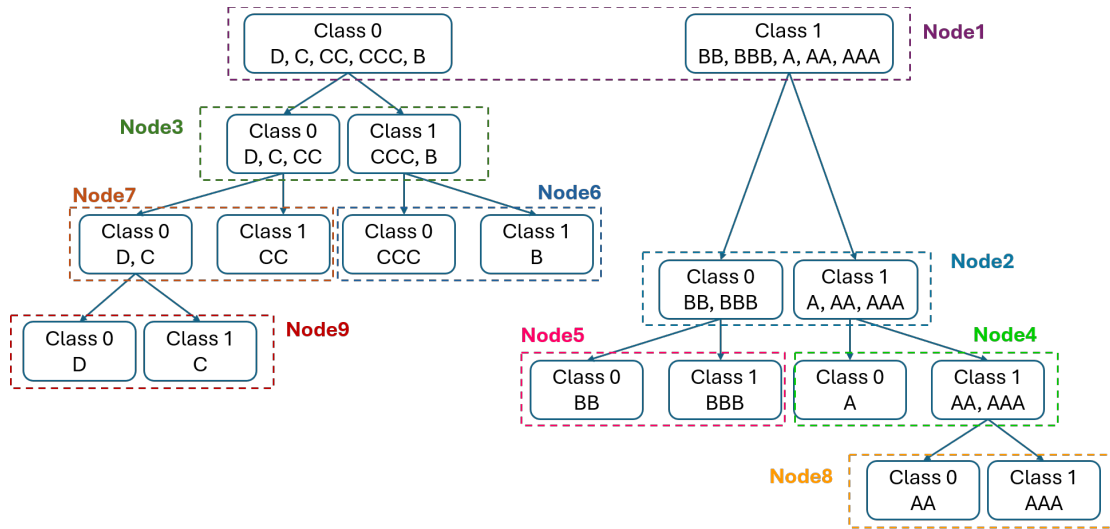
The figure also clarifies the economic meaning of the terminal classifiers (Nodes 4-9), which are directly associated with specific rating boundaries. This representation is particularly useful for interpreting the results of the variable importance, PDP, and SHAP analyses presented in the following sections, as each predictor should be interpreted with respect to the specific rating discrimination performed by the corresponding node.

A distinctive characteristic of HERA is its sector-specific implementation. Separate hierarchical models are estimated for Production, Trade, and Services firms, acknowledging the existence of structural differences across sectors in terms of financial conditions, risk profiles, and rating dynamics. The framework therefore combines hierarchical classification with sector specialization, providing a more flexible representation of corporate creditworthiness than a single pooled model.

While HERA demonstrated satisfactory forecasting performance, its original formulation primarily focused on predictive accuracy. Consequently, the interpretation of model decisions remained largely unexplored. In particular, the framework did not provide detailed information regarding the optimal calibration of classification thresholds, the relative contribution of financial indicators, or the mechanisms through which individual predictors influence rating forecasts.

HERA-XAI extends the original framework by introducing a comprehensive explainability layer. Building upon the same hierarchical structure, the proposed approach integrates threshold calibration procedures, permutation-based variable importance measures, Partial Dependence Plots (PDPs), and Shapley Additive Explanations (SHAP). The resulting framework allows researchers to move beyond prediction and investigate the financial drivers underlying rating transitions, thereby transforming HERA from a forecasting model into a fully explainable decision-support system.

Figure 1. Hierarchical Structure of the HERA Framework and Associated Rating Classifications



Note: adapted from Falavigna, 2026.

Although the original HERA framework focused primarily on forecasting performance, it also included preliminary analyses of rating drivers based on predictor importance measures and illustrative Partial Dependence Plots. These analyses provided an initial understanding of the financial variables underlying rating forecasts but did not constitute a complete explainability framework. HERA-XAI extends this original approach by systematically integrating threshold calibration, variable importance rankings, PDP analyses, and SHAP explanations within a unified explainability architecture. The evolution from HERA to HERA-XAI is summarized in Figure 2.

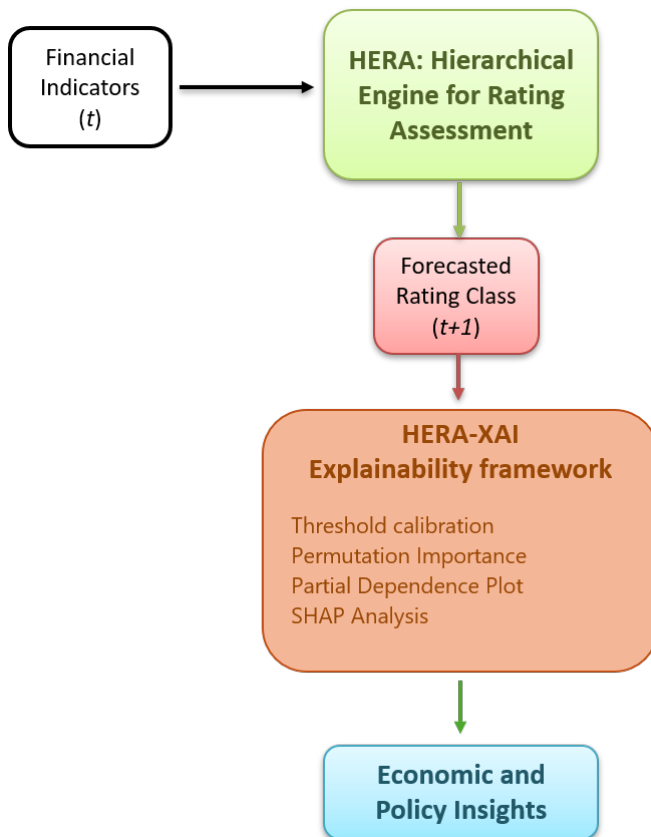
3.2.2. Preliminary Assessment of Random Forest Size

The original HERA framework was based on Random Forest classifiers composed of 200 trees. While this configuration provided satisfactory predictive performance, the development of HERA-XAI introduced additional computationally intensive procedures, including threshold calibration, Partial Dependence Plot (PDP) generation, and SHAP-based explainability analyses. Consequently, a preliminary assessment was conducted to evaluate whether a smaller forest size could preserve the predictive performance of the original specification while reducing computational requirements.

The analysis was performed on the Trade sector, which represents the smallest sector-specific dataset available within the HERA framework. Random Forest models composed of 50 and 200 trees were estimated and compared using the Area Under the Receiver Operating Characteristic Curve (AUC). The AUC measures the discriminatory capability of a classifier independently of any specific classification threshold and is therefore particularly suitable for assessing the intrinsic predictive quality of binary classification models. Values close to 0.5 indicate no discriminatory power, whereas values approaching unity indicate near-perfect class separation.

The comparison presented in Table 2 revealed only limited differences between the two specifications, suggesting that a forest composed of 50 trees is capable of preserving most of the predictive information contained in the larger ensemble. Given the substantial reduction in computational burden, the 50-tree specification was retained for the subsequent calibration and explainability analyses developed within HERA-XAI.

Figure 2. Evolution of the HERA Framework from Forecasting to Explainable Forecasting



Given this favorable trade-off between predictive performance and computational efficiency, the 50-tree configuration was adopted for the subsequent calibration and explainability analyses developed within HERA-XAI.

Table 2. Comparison of Predictive Performance between 200-Tree and 50-Tree Random Forest Classifiers (Trade sector)

	AUC RF (200)	AUC RF (50)
Node1	0.8649	0.8607
Node2	0.8969	0.8933
Node3	0.7723	0.7661
Node4	0.8197	0.8085
Node5	0.7826	0.7768
Node6	0.7474	0.7407
Node7	0.8435	0.8238
Node8	0.7765	0.7415
Node9	0.7029	0.6945

3.2.3. Node-Level Threshold Optimization

The second stage of the analysis focuses on the calibration of the terminal and intermediate classifiers composing the HERA framework. Following the results presented in Section X, all subsequent analyses are based on Random Forest models composed of 50 trees, since the reduction in forest size generates negligible losses in predictive performance while substantially improving computational efficiency.

For each classifier, predictive performance was evaluated using the Out-of-Bag (OOB) observations generated during the bootstrap estimation procedure. In Random Forest models, OOB observations correspond to the units not selected in the bootstrap sample used to estimate a given tree. Since these observations are not used during model estimation, they provide an internal and unbiased validation sample, allowing predictive performance to be evaluated without the need for an additional hold-out dataset.

For each node/RF and sector, several threshold selection criteria were evaluated. Following the approach adopted in HERA (Falavigna, 2026), the analysis considers the maximum F1-score, the maximum Youden Index, and the threshold minimizing total classification error. The corresponding thresholds, classification accuracies, and error rates were computed using OOB predictions.

The definitions and theoretical properties of the F1-score, Youden Index, and classification error metrics are extensively discussed in Falavigna (2026) and are therefore not repeated here. The objective of the present analysis is not to introduce new performance metrics but rather to compare the empirical implications of alternative threshold calibration strategies within the hierarchical architecture.

Table 3 (panel a: Production; panel b: Trade; panel c: Services) reports the calibration results for all nodes and sectors. The evidence reveals that the thresholds maximizing F1-score, Youden Index, and total classification accuracy often differ substantially. In some cases, these differences are relatively small, whereas in others they become particularly pronounced. For example, in RF 3 of the Trade sector the optimal threshold varies from 0.35 under the F1 criterion to 0.70 under the Youden criterion, while the minimum-error threshold is equal to 0.44. Such discrepancies suggest that different optimization criteria may generate substantially different classification outcomes even when applied to the same underlying classifier.

These findings raise an important methodological question. While threshold optimization can be performed at the level of individual nodes, the final objective of HERA is the assignment of a corporate rating through a sequence of hierarchical decisions. Consequently, the threshold that appears optimal for a single classifier may not necessarily produce the best overall performance once classification errors propagate throughout the hierarchy. To address this issue, an additional end-to-end simulation exercise was conducted.

Table 3. Node-Level Threshold Calibration Results Based on Out-of-Bag Predictions

Panel a - Production											
	AUC	Max F1	Th F1	Accuracy F1	Error F1	Max Youden	Th Youden	Accuracy Youden	Error Youden	Minimum Total Error	Th_ Error
Node1	0.860	0.851	0.420	0.800	0.200	0.578	0.620	0.792	0.208	0.197	0.500
Node2	0.867	0.730	0.370	0.795	0.205	0.586	0.370	0.795	0.205	0.193	0.510
Node3	0.801	0.877	0.410	0.803	0.197	0.475	0.710	0.760	0.240	0.196	0.450
Node4	0.801	0.458	0.280	0.845	0.155	0.465	0.160	0.766	0.234	0.115	0.580
Node5	0.746	0.695	0.350	0.661	0.339	0.377	0.490	0.689	0.311	0.311	0.520
Node6	0.743	0.816	0.330	0.720	0.280	0.358	0.620	0.695	0.305	0.274	0.420
Node7	0.782	0.971	0.270	0.945	0.055	0.456	0.930	0.778	0.222	0.055	0.330
Node8	0.732	0.276	0.220	0.918	0.082	0.373	0.060	0.742	0.258	0.049	0.770
Node9	0.703	0.822	0.310	0.712	0.288	0.306	0.680	0.660	0.340	0.283	0.440

Panel b - Trade											
	AUC	Max F1	Th F1	Accuracy F1	Error F1	Max Youden	Th Youden	Accuracy Youden	Error Youden	Minimum Total Error	Th_ Error
Node1	0.853	0.850	0.420	0.795	0.205	0.565	0.620	0.786	0.214	0.202	0.480
Node2	0.884	0.747	0.390	0.818	0.182	0.624	0.360	0.813	0.187	0.174	0.510
Node3	0.750	0.881	0.350	0.797	0.203	0.388	0.730	0.733	0.267	0.200	0.440
Node4	0.792	0.421	0.310	0.877	0.123	0.450	0.140	0.767	0.233	0.096	0.590
Node5	0.765	0.698	0.340	0.680	0.320	0.410	0.450	0.704	0.296	0.293	0.480
Node6	0.735	0.832	0.290	0.735	0.265	0.345	0.640	0.697	0.303	0.261	0.420
Node7	0.806	0.977	0.330	0.955	0.045	0.504	0.940	0.803	0.197	0.045	0.330
Node8	0.732	0.240	0.210	0.935	0.065	0.391	0.080	0.846	0.154	0.036	0.620
Node9	0.679	0.804	0.280	0.688	0.312	0.268	0.630	0.655	0.345	0.303	0.400

Panel c - Services											
	AUC	Max F1	Th F1	Accuracy F1	Error F1	Max Youden	Th Youden	Accuracy Youden	Error Youden	Minimum Total Error	Th_ Error
Node1	0.830	0.834	0.400	0.774	0.226	0.521	0.610	0.766	0.234	0.222	0.490
Node2	0.843	0.740	0.380	0.761	0.239	0.538	0.430	0.769	0.231	0.227	0.520
Node3	0.772	0.851	0.360	0.760	0.240	0.411	0.710	0.707	0.293	0.235	0.460
Node4	0.786	0.363	0.260	0.951	0.049	0.475	0.060	0.819	0.181	0.037	0.550
Node5	0.708	0.675	0.310	0.613	0.387	0.312	0.490	0.657	0.343	0.342	0.520
Node6	0.737	0.781	0.280	0.681	0.319	0.341	0.580	0.677	0.323	0.306	0.440
Node7	0.777	0.973	0.320	0.948	0.052	0.444	0.930	0.789	0.211	0.052	0.320
Node8	0.671	0.277	0.200	0.789	0.211	0.259	0.120	0.664	0.336	0.101	0.600
Node9	0.703	0.824	0.260	0.709	0.291	0.305	0.680	0.660	0.340	0.282	0.410

Furthermore, the choice of the optimal threshold depends on the objectives of the classification process itself. In HERA (Falavigna, 2026), the F1-score was adopted on theoretical grounds because of its ability to balance precision and recall, particularly in the presence of class imbalance. However, the hierarchical nature of HERA requires not only the correct identification of positive cases but also an adequate separation between positive and negative classes. In contexts where the number of true positives is limited, correctly identifying these observations is certainly important, but not at the expense of a substantial increase in false positives. From this perspective, the Youden criterion may be particularly relevant because it explicitly seeks a balance between sensitivity and specificity, thereby emphasizing the overall discriminatory power of the classifier rather than the identification of positive cases alone. The presence of substantial differences between F1-based and Youden-based thresholds in several nodes further motivates the empirical comparison conducted in the next step of the analysis.

3.2.4. End-to-End HERA Simulation

To evaluate the impact of threshold selection on the final rating assignment, a simulation exercise was conducted using a random sample of 300 firms for each of the three sectors (Production, Trade, and Services). The objective was to assess how different threshold calibration criteria affect the behavior of the hierarchical classification system and, ultimately, the accuracy of the predicted ratings.

Three alternative threshold-selection strategies were considered:

1. the threshold maximizing the F1-score (MaxF1);
2. the threshold maximizing the Youden Index (Youden);
3. the threshold minimizing the overall classification error (MinError).

For each strategy, the optimal threshold was estimated separately for every sector and every node of the HERA hierarchy. Consequently, the routing decisions taken within the hierarchical structure depend on both the sector to which the firm belongs and the calibration criterion being evaluated.

The simulation can be illustrated through the following example. Consider Firm A, operating in the Production sector, for which a one-year-ahead rating forecast is required. The firm first enters the Random Forest classifier (RF1), which produces a predicted probability score. This score is then compared with the node-specific threshold associated with the selected calibration strategy. If the score is greater than or equal to the threshold, the prediction is converted into class 1; otherwise, it is converted into class 0. Depending on this binary outcome, the firm is routed to the next classifier in the hierarchy (e.g., RF2 or RF3). The same procedure is repeated sequentially until a terminal node is reached and a final rating class is assigned.

The entire routing process was fully automated and applied to all sampled firms. The simulation was repeated three times, once for each threshold-selection criterion (MaxF1, Youden, and MinError), thereby generating three alternative sets of predicted ratings.

To evaluate the resulting classification performance, the predicted ratings were compared with the observed ratings through confusion matrices (Tables 4-6). In each matrix, rows correspond to the rating classes assigned by HERA, while columns report the actual observed rating classes. Each cell therefore contains the number of firms classified in the corresponding predicted-observed combination. The confusion matrices provide a direct assessment of the quality of the rating forecasts and allow the identification of the threshold-selection criterion yielding the most accurate overall classification. Notice that since the simulation sample was randomly drawn from the database, no firms belonging to the AAA rating class were selected. This outcome reflects the limited representation of the highest rating category in the population rather than any intentional exclusion from the analysis.

The numbers in bold indicate the correct classifications. While the three calibration strategies generate different classification patterns, the lowest overall error is achieved using Youden-based thresholds. The total number of misclassifications falls from 35 under the F1 criterion to 27 under

the Youden criterion, whereas the minimum-error thresholds lead to 57 misclassifications. This evidence suggests that a threshold optimized at the node level does not necessarily translate into the best performance of the hierarchical architecture as a whole.

Table 4: Confusion Matrix of HERA Rating Assignments Using the MaxF1 Threshold Criterion. Rows: HERA score; Columns: Real score

	AAA	AA	A	BBB	BB	B	CCC	CC	C	D	Total
AAA											
AA		7	1								8
A			60		3						63
BBB				55	2		3		1		61
BB					60			2			62
B				1	1	50	7	4		1	64
CCC			1				20		1	2	24
CC								13	4	1	18
C											
D											
Total		7	62	56	66	50	30	19	6	4	300

Table 5. Confusion Matrix of HERA Rating Assignments Using the Youden Threshold Criterion. Columns: Real score

	AAA	AA	A	BBB	BB	B	CCC	CC	C	D	Total
AAA		1									1
AA		6	7								13
A			53								53
BBB				53							53
BB				1	63						64
B			1	1	1	49	1	1			54
CCC				1		1	28				30
CC			1		2			17	5		25
C							1	1			2
D									1	4	5
Total		7	62	56	66	50	30	19	6	4	300

Table 6 Confusion Matrix of HERA Rating Assignments Using the Minimum Error Threshold Criterion. Rows: HERA score; Columns: Real score

	AAA	AA	A	BBB	BB	B	CCC	CC	C	D	Total
AAA											
AA		7									7
A			60								60
BBB				53	2		1				56
BB			1	2	63		1				67
B			1	1	1	34	17	4	1	1	60
CCC						16	11			2	29
CC								15	5	1	21
C											
D											
Total		7	62	56	66	50	30	19	6	4	300

4. HERA 2.0 EXPLAINABILITY ANALYSIS

4.1. Predictor Importance Analysis

4.1.1. Methodology

Variable importance measures constitute the first layer of the explainability framework implemented in HERA-XAI. While forecasting models are primarily evaluated in terms of predictive performance, understanding which financial indicators drive classification decisions is equally important from both an economic and a managerial perspective. Variable importance analysis provides an initial assessment of the contribution of each predictor to the prediction process and represents the starting point for the subsequent explainability analyses based on Partial Dependence Plots (PDPs) and SHAP values.

Following the methodology adopted in HERA (Falavigna, 2026), predictor importance is evaluated through the Out-of-Bag (OOB) permutation importance measure generated by the Random Forest algorithm. The basic idea underlying this approach is straightforward. For each predictor, its values are randomly permuted among the observations belonging to the Out-of-Bag sample, while all remaining variables are kept unchanged. The resulting decrease in predictive performance is then measured and compared with the original performance obtained before permutation. Variables whose random permutation produces a large increase in classification error are considered more important because the classifier relies heavily on the information they contain. Conversely, variables producing negligible changes in predictive performance are interpreted as having a limited contribution to the classification process. As already anticipated in HERA (Falavigna, 2026), predictor importance measures provide a first indication of the financial variables driving rating assignments. However, the original analysis was limited to selected nodes and was primarily intended to support the interpretation of forecasting results.

The use of OOB permutation importance offers several advantages within the HERA framework. First, importance measures are obtained from observations that are not used during the construction of individual trees, thereby providing an internally validated assessment of predictor relevance. Second, the procedure naturally accommodates nonlinear relationships and interaction effects among financial indicators. Finally, variable importance can be computed separately for each node of the hierarchical architecture, allowing the identification of sector-specific and node-specific rating drivers.

For each classifier, predictors were ranked according to their OOB importance values. The ranking was subsequently used to identify the most influential financial indicators associated with each classification problem. Particular attention was devoted to the top-ranked predictors, as these variables constitute the candidates for the PDP and SHAP analyses presented in the following sections. While variable importance measures identify which variables matter most, they do not provide information regarding the direction, magnitude, or functional form of their effects. Consequently, additional explainability tools are required to understand how individual financial indicators influence the probability of belonging to a specific rating class.

The variable importance analysis therefore represents the first step of the HERA-XAI explainability framework. By identifying the most relevant predictors across nodes and sectors, it provides a bridge between predictive performance and economic interpretation and establishes the foundation for the more detailed analyses based on PDPs and SHAP explanations.

4.1.2. Results

Tables 7 reports the most important predictor identified for each node of the HERA architecture across the Production, Trade, and Services sectors. Predictors were ranked according to their Out-of-Bag (OOB) permutation importance values, and only the highest-ranked variable was retained for each node-sector combination.

The reported importance score measures the increase in classification error generated when the values of a predictor are randomly permuted across the Out-of-Bag observations while all remaining variables are left unchanged. Consequently, the score quantifies the contribution of a variable to the predictive performance of the classifier. Higher values indicate that the model relies more heavily on the information contained in that predictor, whereas lower values suggest a weaker contribution to the classification process.

From an interpretative perspective, the importance score does not provide information regarding the sign or functional form of the relationship between a predictor and the forecasted rating outcome. Rather, it measures the relative contribution of each variable to classification accuracy. The analysis therefore represents a first step in identifying the most influential financial drivers within the hierarchical architecture. For instance, in Node 1 of the Production sector, DEBTA is associated with an importance score of 15.887. This means that the model's predictive accuracy decreases considerably when the information provided by DEBTA is no longer available in its original form. Therefore, DEBTA represents one of the most influential variables driving classification decisions at this node.

The interpretation of predictor importance varies according to the position of the node within the hierarchy. In the upper levels of the architecture, particularly Nodes 1, 2, and 3, the identified predictors contribute to broad rating separations that affect all subsequent classification paths. For example, the most important predictor in Node 1 captures the variable that plays the greatest role in the initial discrimination performed by the framework and therefore influences the routing of firms toward all downstream branches of the hierarchy.

Greater attention, however, should be devoted to Nodes 4 through 9. These terminal classifiers are directly responsible for the assignment of the final rating classes and therefore provide the most detailed information regarding the financial drivers underlying rating outcomes. The most important predictors identified in these nodes represent the variables that most strongly influence the final stages of the rating assignment process and constitute the primary candidates for the explainability analyses developed in the following sections.

Overall, Table 7 highlights the existence of both common and sector-specific rating drivers. Some financial indicators repeatedly emerge across multiple nodes and sectors, suggesting the presence of structural determinants of creditworthiness, whereas others appear only in specific branches of the hierarchy, reflecting the heterogeneous financial characteristics of Production, Trade, and Services firms.

Table 7. Most Important Predictor by Node and Sector Based on OOB Permutation Importance

Node	Var (Production)	Importance	Var (Trade)	Importance	Var (Services)	Importance
1	DEBTA	15.887	FCCOV	16.405	FAEQ	22.460
2	FCCOV	29.680	FCCOV	19.533	FCCOV	35.626
3	DEBTA	15.617	DEBTA	9.693	FAEQ	17.806
4	FAPROD	10.318	DEBPROD	7.245	DEBTA	4.919
5	FCCOV	18.994	FCCOV	24.023	FAEQ	20.924
6	FAEQ	19.205	FAEQ	11.457	FAEQ	22.315
7	DEBTA	4.698	ROA	2.866	DEBCOV	5.207
8	DEBPROD	2.909	DEBPROD	2.116	DEBPROD	3.581
9	DEBTA	8.831	DEBPROD	5.558	FAEQ	9.128

Note: DEBTA=Debt/Assets; DEBPROD=Debt/Production; ROS=Return on Sales; ROA=Return on Assets; DEBCOV=Debt Coverage; FAPROD=Fixed Assets/Production; FCCOV=Financial Charges Coverage; FAEQ=Fixed Assets/Equity'

A general decline in permutation importance values can be observed moving from the upper to the lower levels of the hierarchy. This pattern is likely attributable to the progressive reduction in sample size within terminal nodes, as well as to the increasing homogeneity of firms reaching the final classification stages. Consequently, the predictive contribution of individual variables tends to become more evenly distributed across the available financial indicators.

Tables 8-10 present the three most important predictors identified for each node and sector according to the OOB permutation importance ranking. These variables constitute the candidate predictors considered for the subsequent Partial Dependence Plot (PDP) analyses.

The inclusion of the three highest-ranked predictors provides a more complete representation of the variable importance structure underlying each classifier. While the top-ranked predictor captures the largest contribution to predictive performance, the second and third predictors often represent complementary dimensions of firms' financial conditions and may contribute substantially to the classification process.

The selection of candidate variables based on permutation importance ensures that the explainability analysis remains closely linked to the predictive behavior of the Random Forest models. Rather than examining financial indicators arbitrarily, the PDP analysis focuses on variables that the classifiers have identified as the most relevant sources of information.

To maintain a manageable number of graphical outputs and facilitate interpretation, Partial Dependence Plots were ultimately computed only for the highest-ranked predictor associated with each node. This choice reflects the objective of investigating the marginal effect of the single most influential financial driver while limiting redundancy across highly correlated predictors.

Consequently, the variables reported in the first position of Tables 8-10 constitute the predictors analyzed through PDPs in the following section. The resulting plots provide information that cannot be inferred from importance measures alone, namely the direction, magnitude, and potential non-linearities characterizing the relationship between key financial indicators and the predicted rating outcomes.

Table 8. Top Three Predictors Selected for PDP Analysis: Production Sector

Node	Top1	Top2	Top3
1	DEBTA	FCCOV	FAEQ
2	FCCOV	FAPROD	DEBPROD
3	DEBTA	FAEQ	DEBPROD
4	FAPROD	FAEQ	DEBTA
5	FCCOV	DEBTA	FAPROD
6	FAEQ	DEBTA	FAPROD
7	DEBTA	DEBCOV	DEBPROD
8	DEBPROD	DEBTA	ROS
9	DEBTA	FAEQ	DEBPROD

Note: DEBTA=Debt/Assets; DEBPROD=Debt/Production; ROS=Return on Sales; ROA=Return on Assets; DEBCOV=Debt Coverage; FAPROD=Fixed Assets/Production; FCCOV=Financial Charges Coverage; FAEQ=Fixed Assets/Equity'

Table 9. Top Three Predictors Selected for PDP Analysis: Trade Sector

Node	Top1	Top2	Top3
1	FCCOV	DEBTA	DEBPROD
2	FCCOV	DEBTA	FAEQ
3	DEBTA	FAEQ	FCCOV
4	DEBPROD	FAPROD	DEBTA
5	FCCOV	DEBTA	DEBPROD
6	FAEQ	DEBTA	FAPROD
7	ROA	DEBTA	DEBCOV
8	DEBPROD	FAPROD	ROS
9	DEBPROD	DEBTA	FAEQ

Note: DEBTA=Debt/Assets; DEBPROD=Debt/Production; ROS=Return on Sales; ROA=Return on Assets; DEBCOV=Debt Coverage; FAPROD=Fixed Assets/Production; FCCOV=Financial Charges Coverage; FAEQ=Fixed Assets/Equity'

Table 10. Top Three Predictors Selected for PDP Analysis: Services Sector

Node	Top1	Top2	Top3
1	FAEQ	DEBTA	FAPROD
2	FCCOV	FAEQ	DEBTA
3	FAEQ	DEBTA	DEBPROD
4	DEBTA	DEBPROD	ROS
5	FAEQ	FCCOV	DEBTA
6	FAEQ	DEBTA	FAPROD
7	DEBCOV	FAPROD	ROA
8	DEBPROD	DEBTA	ROS
9	FAEQ	DEBTA	DEBPROD

Note: DEBTA=Debt/Assets; DEBPROD=Debt/Production; ROS=Return on Sales; ROA=Return on Assets; DEBCOV=Debt Coverage; FAPROD=Fixed Assets/Production; FCCOV=Financial Charges Coverage; FAEQ=Fixed Assets/Equity'

4.1.3. Interpreting Predictor Importance within the HERA Hierarchy

The interpretation of predictor importance should always be linked to the specific classification task performed by each node of the HERA architecture. Unlike traditional multiclass classifiers, each HERA node is responsible for discriminating between two groups of rating categories. Consequently, the most important predictor identified within a node should not be interpreted as a general driver of creditworthiness, but rather as the variable that contributes most to a specific rating transition or rating boundary.

For example, in Table 10, the variable FAEQ emerges as the most important predictor in Node 5 within the Services sector. This result indicates that FAEQ is the financial indicator that contributes most to the classifier responsible for distinguishing firms located around the BBB-BB rating boundary. In other words, the importance of FAEQ should be interpreted in relation to the model's ability to separate firms belonging to the investment-grade segment from those exhibiting a more speculative risk profile.

More generally, the economic interpretation of predictor importance becomes increasingly meaningful in the lower levels of the hierarchy. While Nodes 1, 2, and 3 perform broader rating separations affecting all downstream classifications, Nodes 4 through 9 are directly associated with the final stages of the rating assignment process. These terminal classifiers operate close to specific rating boundaries and therefore provide more detailed information regarding the financial factors driving the final rating outcome. To facilitate the interpretation of predictor importance measures, Table 11 summarizes the economic meaning of each node within the HERA architecture. Since every classifier is associated with a specific rating boundary, the most important predictor identified in a given node should be interpreted as the variable contributing most to that particular classification task.

Table 11. Economic Interpretation of the HERA Hierarchical Nodes

Node	Main Classification Task	Economic Interpretation
1	Upper vs Lower Rating Groups (D+C+CC+CCC+B) vs (BB+BBB+A+AA+AAA)	Initial separation of firms into broad credit-quality groups. The decision affects all subsequent paths within the hierarchy.
2	Upper-Branch Classification (BB+BBB) vs (A+AA+AAA)	Intermediate classification within the upper segment of the rating hierarchy.
3	Lower-Branch Classification (D+C+CC) vs (CCC+B)	Intermediate classification within the lower segment of the rating hierarchy.
8	AA vs AAA	Identification of the financial factors that distinguish firms belonging to the two highest rating classes.
4	A vs (AA+AAA)	Analysis of the financial characteristics associated with firms positioned between the highest credit-quality segment and the A rating category.
5	BB vs BBB	Identification of the determinants underlying the investment-grade versus speculative-grade boundary.
6	CCC vs B	Analysis of the drivers distinguishing firms characterized by elevated speculative risk from those experiencing severe financial weakness.
7	(D+C) vs CC	Identification of the factors associated with extreme credit deterioration and firms approaching default conditions.
9	D vs C	Analysis of the financial indicators driving the distinction between default firms and firms experiencing severe but non-default financial distress.

For this reason, the PDP and SHAP analyses presented in the following sections mainly focus on Nodes 4 through 9. The objective is not merely to identify which variables are important, but also to understand how the most influential financial indicators affect the probability of crossing specific rating thresholds. This approach allows the interpretation of the HERA architecture in

terms of economically meaningful rating transitions rather than generic measures of variable relevance.

4.2. Partial Dependence Plots (PDP)

4.2.1. Methodology

In order to support the interpretability stage of the HERA-XAI framework, predictor importance rankings were first extracted from all Random Forest classifiers composing the hierarchical architecture. For each node and sector, predictors were ranked according to the Out-of-Bag (OOB) Permutation Importance measure. Two summary datasets were subsequently generated: the first reports the most influential predictor associated with each node-sector model, while the second reports the three most relevant predictors. These rankings were used to identify the candidate variables for the PDP analysis.

Because the objective of HERA-XAI is to explain the final rating assignment process, the PDP analysis was restricted to the terminal nodes of the hierarchy (RF4-RF9), which are directly responsible for assigning the final rating classes. Focusing on terminal classifiers allows the investigation of the economic and financial drivers underlying the ultimate rating decision rather than the intermediate routing decisions occurring within the hierarchy.

For each sector (Production, Trade and Services), a representative random sample of firms was extracted from the original database. Given the size of the complete dataset, which exceeds four million firm-year observations, the use of random subsamples substantially reduced computational costs while preserving the overall distributional characteristics of the data. The resulting samples were used exclusively for explainability analyses and did not affect model estimation.

Prior to the computation of PDPs, all explanatory variables were pre-processed to ensure numerical consistency. Infinite values generated by financial ratios were replaced with missing values, while missing observations were imputed using the variable-specific median calculated within the corresponding sample. This procedure was adopted to guarantee that all observations could be evaluated by the Random Forest classifiers.

For each terminal node and sector, only the most influential predictor (Top Predictor) identified through the permutation importance analysis was retained. This choice was motivated by interpretability considerations. While PDPs can be generated for all predictors, restricting the analysis to the leading variable substantially reduces the number of graphs while focusing attention on the financial indicator most strongly associated with classification performance.

The PDPs were computed using a manual implementation procedure. For a selected predictor, a grid of equally spaced values covering the central portion of its empirical distribution was generated. The chosen predictor was then successively fixed at each grid value for all observations, while the remaining predictors retained their observed values. The modified dataset was subsequently submitted to the corresponding Random Forest classifier and predicted probabilities were computed. The average probability across all observations was then calculated for each grid point. Repeating this procedure over the entire grid generated an estimate of the partial dependence function.

Formally, the partial dependence function of predictor x_j can be expressed as

$$PD(x_j) = \frac{1}{N} \sum_{i=1}^N f(x_j, x_{i,-j})$$

where $f(\cdot)$ denotes the prediction function of the Random Forest model, N is the number of observations, and $x_{i,-j}$ represents the observed values of all predictors other than x_j .

Unlike local explainability methods that focus on individual firms, PDPs provide a global interpretation of model behaviour. The resulting curves describe the average relationship between a financial indicator and the predicted probability generated by the classifier. Consequently, PDPs allow the identification of monotonic trends, threshold effects, turning points, and non-linear relationships that cannot be inferred from variable importance measures alone. Within HERA-XAI, PDPs therefore complement permutation importance rankings by explaining not only which variables are important, but also how changes in their values influence rating forecasts on average.

Since the original database contains approximately four million firm-year observations, a representative random sample of 50,000 observations was extracted for each sector. The resulting samples were subsequently used to compute Partial Dependence Plots while maintaining computational feasibility and preserving the distributional properties of the original data.

4.2.2. Results

To illustrate the interpretative potential of Partial Dependence Plots, the analysis initially focuses on Node 9, the terminal classifier responsible for distinguishing firms belonging to the D and C rating classes. This node is particularly relevant from an economic perspective, as it separates firms experiencing severe financial distress from firms already classified in default conditions.

Node 9 also provides an interesting comparison across sectors because the variable importance analysis identified different top predictors for Production, Trade, and Services firms. Consequently, the PDPs presented below allow us to investigate how the most influential financial indicator affects the probability of belonging to the D class within each sector and to assess whether different financial mechanisms underlie the final stage of credit deterioration. The variable importance analysis reveals substantial sectoral heterogeneity within RF9. Specifically, DEBTA (Total Debt to Total Assets Ratio) emerges as the most influential predictor in the Production sector, DEBPROD (Debt to Production Ratio) represents the top predictor in the Trade sector, whereas FAEQ (Fixed Assets to Net Equity Ratio) is identified as the most important variable in the Services sector. This evidence suggests that different financial dimensions drive the distinction between firms rated D and C across sectors, highlighting the relevance of sector-specific rating determinants.

Unlike variable importance measures, which only identify the relevance of a predictor, PDPs provide information regarding the direction, magnitude, and potential non-linearities of the relationship between a financial indicator and the predicted rating outcome.

Figure 3 presents Partial dependence of the predicted probability associated with the C rating class on the top-ranked predictor identified through OOB permutation importance within the three sectors (panel a: Production; panel b: Trade; panel c: Services). In other words, Figure 3 reports the Partial Dependence Plots associated with RF9, the terminal classifier responsible for distinguishing firms belonging to rating classes D and C. The analysis highlights the existence of sector-specific financial drivers characterizing the final stage of financial distress.

In details, it should be noted that Node 9 discriminates between firms belonging to rating classes D and C, where class C represents the positive outcome (class = 1) and class D represents the negative outcome (class = 0). Consequently, the values reported on the vertical axis should be interpreted as the predicted probability of assignment to class C. A decreasing PDP therefore implies an increasing likelihood of assignment to class D, whereas an increasing PDP indicates a higher probability of belonging to class C.

In the Production sector (Figure 3, panel a), the top-ranked predictor is DEBTA (Total Debt to Total Assets Ratio). The PDP shows a predominantly negative and almost monotonic relationship, indicating that higher levels of indebtedness relative to total assets are associated with a lower probability of being classified in rating class C and, consequently, a greater likelihood of assignment to class D. The strongest variation occurs at relatively low values of DEBTA, whereas beyond approximately 25-30 units the curve becomes substantially flatter. This behaviour suggests the presence of a threshold effect followed by a clear saturation point, beyond

which additional increases in indebtedness generate only limited changes in the predicted rating outcome.

In the Trade sector (Figure 3, panel b), the most important predictor is DEBPROD (Debt to Production Ratio). The relationship is again largely monotonic decreasing, confirming the relevance of debt-related indicators in the final stages of rating deterioration. The probability of assignment to class C declines rapidly for low and intermediate values of DEBPROD and then progressively stabilizes. A noticeable change in slope can be observed around values between 150 and 170, after which the curve reaches a relatively stable plateau. This pattern points to the existence of a non-linear threshold effect and a subsequent saturation phenomenon, suggesting that extremely high debt burdens relative to productive activity provide little additional discriminatory information once severe financial distress has already been reached.

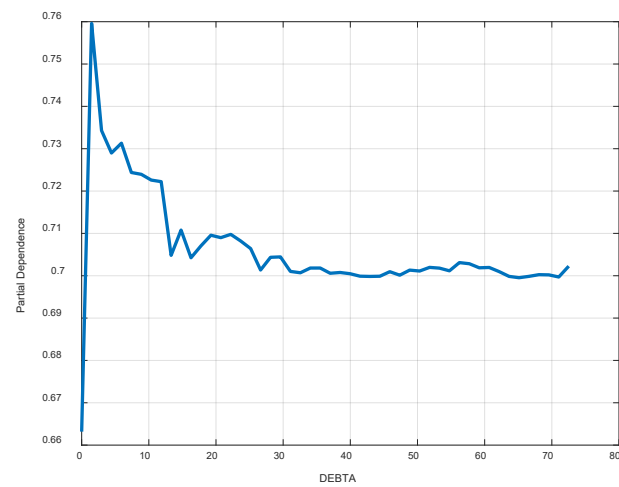
The pattern observed in the Services sector (Figure 3, panel c) differs substantially from those identified in Production and Trade. Here the top predictor is FAEQ (Fixed Assets to Net Equity Ratio). After some instability in the extreme lower tail of the distribution, the PDP exhibits a generally positive relationship, indicating that higher FAEQ values are associated with a higher probability of assignment to class C rather than class D. The relationship appears less regular and more complex than those observed for debt-related indicators. Several local fluctuations emerge across the distribution, suggesting the presence of non-linear effects, while the curve stabilizes for values exceeding approximately 450-500 units, indicating a saturation point. The evidence suggests that, within the Services sector, the distinction between C and D firms is influenced more by the structure of long-term financing than by indebtedness measures alone. One possible interpretation is that firms located in the D category may have already experienced significant reductions in their stock of fixed assets, possibly as a consequence of restructuring or asset disposal processes typically associated with severe financial distress.

Overall, the PDP analysis reveals important sectoral differences in the financial drivers underlying the final stage of credit deterioration. While indicators related to indebtedness dominate the discrimination between classes C and D in the Production and Trade sectors, the Services sector appears to be more strongly influenced by the relationship between fixed investments and equity financing. These findings provide additional support for the sector-specific architecture underlying the HERA framework and highlight the heterogeneous nature of corporate financial distress across economic sectors.

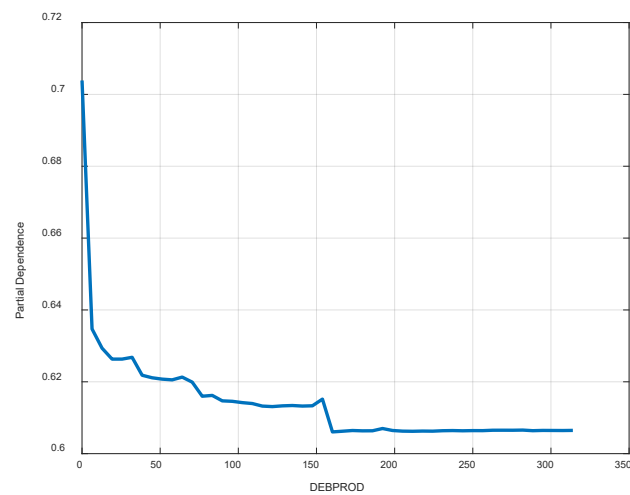
The PDPs reported in Figure 3 are intended to provide illustrative examples of the interpretability potential offered by HERA-XAI. To avoid excessive redundancy, the remaining PDPs (from Node4 to Node8) are presented in Appendix A. Figures A1-A5 report the partial dependence plots associated with the remaining nodes of the HERA hierarchy. Each figure is organized into three panels, corresponding to the Production, Trade, and Services sectors, respectively. This format facilitates direct comparison across sectors and confirms the presence of heterogeneous, frequently non-linear, and sector-specific relationships between financial indicators and rating outcomes.

Figure 3. Partial Dependence Plot for Node 9 (D vs C): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)

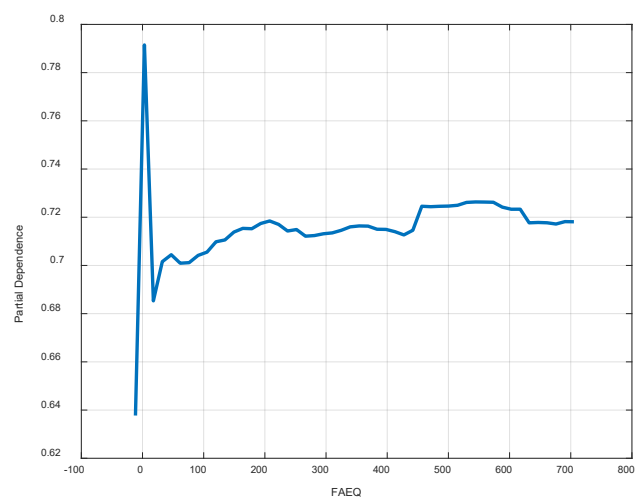
Production – panel a



Trade – panel b



Services – panel c



4.3. SHAP Explanations

While predictor importance measures identify the financial indicators that contribute most to predictive performance and Partial Dependence Plots describe the average effect of these variables on model predictions, neither approach is able to fully explain how individual predictors contribute to specific classification outcomes. To address this limitation, HERA-XAI incorporates Shapley Additive Explanations (SHAP), an explainable artificial intelligence technique derived from cooperative game theory.

The SHAP methodology interprets each predictor as a "player" contributing to the final prediction generated by the classifier. The contribution of every variable is quantified through a Shapley value, which measures how much the predicted probability changes when the information provided by that predictor is considered in combination with all other variables. As a result, SHAP provides an additive decomposition of model predictions and allows complex machine learning algorithms to be interpreted in a transparent and theoretically consistent manner.

A distinctive feature of SHAP is its ability to support interpretation at two complementary levels: local and global.

Local SHAP explanations focus on individual observations and are designed to answer the question:

Why has the model assigned this specific prediction to this particular firm?

For a given company, the predicted probability can be decomposed into the contributions generated by each financial indicator, allowing the identification of the variables that increase or decrease the likelihood of belonging to a specific rating class. Local SHAP therefore provides firm-specific explanations and reveals the financial drivers underlying individual rating forecasts.

Global SHAP explanations, in contrast, are obtained by aggregating local SHAP values across a sample of observations. Their objective is to answer a different question:

Which variables drive the overall behaviour of the model?

By averaging the absolute contributions of predictors across multiple firms, Global SHAP measures identify the variables that exert the greatest influence on the classification process as a whole. Consequently, Global SHAP provides a ranking of predictor importance that can be directly compared with the permutation importance measures previously discussed.

The two perspectives are therefore complementary. Local SHAP explains individual predictions, whereas Global SHAP explains the behaviour of the classifier. Together, they provide a detailed understanding of both firm-level decisions and the overall mechanisms through which the HERA framework generates rating forecasts.

To ensure consistency with the PDP analysis, SHAP explanations were derived from the same random samples of 50,000 observations extracted from the original sector-specific datasets. Subsequently, random subsets of these samples were selected for the computation of Global SHAP measures in order to reduce the computational burden associated with Shapley value estimation.

4.3.1. Local SHAP

In the local interpretation framework, SHAP values are computed for a single observation. For a given firm, the predicted probability generated by the Random Forest classifier is decomposed into the sum of the baseline prediction and the contribution of each predictor. Positive SHAP values indicate variables that increase the probability of belonging to the positive class, whereas negative values indicate variables that decrease that probability. Local SHAP explanations

therefore provide a transparent and firm-specific interpretation of rating assignments, revealing the financial indicators that drive individual predictions.

4.3.2. Methodology

Mathematically, the prediction for observation i can be expressed as:

$$\hat{y}_i = \phi_0 + \sum_{j=1}^p \phi_{ij}$$

where ϕ_0 represents the baseline prediction and ϕ_{ij} denotes the Shapley contribution of predictor j for observation i .

Local SHAP explanations were obtained by selecting an individual observation from the SHAP sample and decomposing the predicted probability into the additive contribution of each financial indicator.

4.3.3. Results

As an illustrative example, Local SHAP values were computed for a randomly selected firm belonging to the SHAP sample. Table 12 reports the Local SHAP values computed for the randomly selected firm belonging to the RF8 sample. RF8 is responsible for the final discrimination between the AA and AAA rating classes. Positive SHAP values increase the probability of being assigned to the AAA class, whereas negative values increase the probability of being classified as AA.

Among the positive contributions, ROS (0.0179) and DEBPROD (0.0132) emerge as the most influential predictors, followed by DEBTA (0.0079) and FAEQ (0.0060). These variables contribute to moving the prediction toward the highest rating category.

Conversely, DEBCOV (-0.0320) and FAPROD (-0.0313) generate the largest negative contributions, while ROA (-0.0131) also exerts a negative effect. These variables therefore shift the prediction toward the AA category.

Considering the absolute magnitude of the SHAP values, DEBCOV and FAPROD exert the strongest influence on the prediction for this particular firm. The example illustrates how Local SHAP explanations allow the final prediction to be decomposed into the additive contributions of individual financial indicators, thereby revealing the specific variables that support or weaken the probability of achieving the highest rating class.

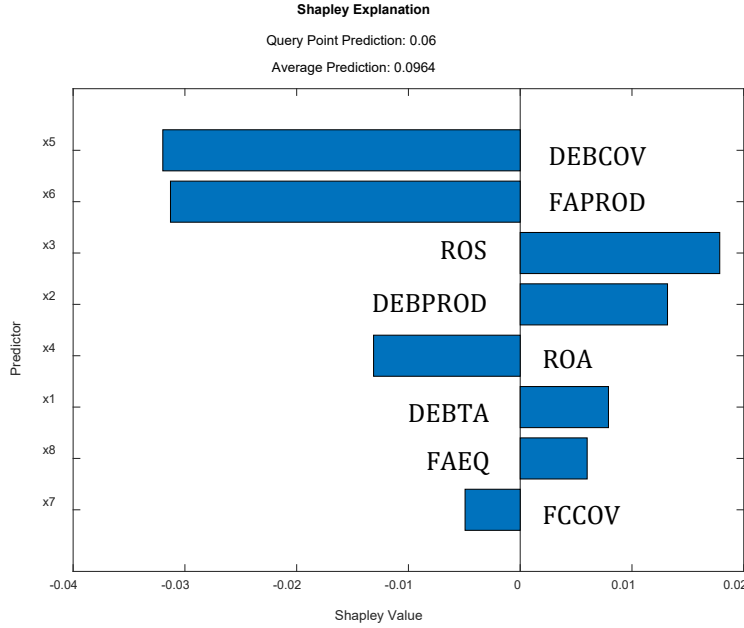
Table 12. Local SHAP Contributions to the AA versus AAA Classification for a Randomly Selected Firm

Predictor	ShapleyValue
DEBTA	0.0078972
DEBPROD	0.013179
ROS	0.017852
ROA	-0.01312
DEBCOV	-0.031977
FAPROD	-0.031289
FCCOV	-0.0049333
FAEQ	0.005991

Note: DEBTA=Debt/Assets; DEBPROD=Debt/Production; ROS=Return on Sales; ROA=Return on Assets; DEBCOV=Debt Coverage; FAPROD=Fixed Assets/Production; FCCOV=Financial Charges Coverage; FAEQ=Fixed Assets/Equity'

Figure 4 reports the Local SHAP explanation for a randomly selected firm belonging to the RF8 sample. The classifier associated with RF8 is responsible for the final distinction between the AA and AAA rating classes. The graph decomposes the predicted probability into the contribution of each financial indicator, allowing the identification of the variables that increase or decrease the likelihood of obtaining the highest rating category. The prediction for the selected firm is compared with the average prediction generated by the classifier, highlighting how individual financial characteristics influence the final rating assignment.

Figure 4. Local SHAP Decomposition of the Predicted Probability in RF8 (AA vs AAA)



The average probability of belonging to the AAA class is equal to 9.64%, whereas the selected firm's predicted probability is 6%. The gap between these two values is explained by the combined contribution of the financial indicators represented in the SHAP decomposition. In particular, DEBCOV and FAPROD exert the strongest negative effects, pushing the prediction toward the AA class, while ROS and DEBPROD provide the largest positive contributions toward AAA classification.

4.3.4. Global SHAP

4.3.5. Methodology

While Local SHAP values explain individual predictions, the primary objective of HERA-XAI is to understand the overall behavior of the hierarchical rating forecasting framework. For this reason, Global SHAP importance is obtained by aggregating Local SHAP values across multiple observations. Following the standard SHAP approach, predictor importance is measured through the mean absolute Shapley value:

$$\text{MeanAbsSHAP}_j = \frac{1}{N} \sum_{i=1}^N |\phi_{ij}|$$

where ϕ_{ij} denotes the Shapley value associated with predictor j for observation i , and N represents the number of evaluated observations.

The resulting measure quantifies the average contribution of each predictor to the model predictions across the analysed sample. Variables associated with larger mean absolute SHAP

values exert a stronger influence on the classification process. Unlike traditional permutation importance measures, Global SHAP not only identifies whether a predictor is important, but also quantifies the magnitude of its contribution to the final prediction.

The original database contains approximately four million firm-year observations. The exact computation of Shapley values is computationally demanding because the procedure requires the evaluation of a large number of predictor combinations for each observation. Performing SHAP analyses on the full dataset would therefore imply excessive computational costs and execution times without generating substantial additional information.

To ensure consistency with the PDP analysis, SHAP values were computed using the same random samples of 50,000 observations previously extracted from the original sector-specific datasets. Subsequently, Global SHAP measures were estimated on smaller random subsamples drawn from these datasets. Specifically, 40 observations were randomly selected for each node and sector. The use of a common sample size across all models facilitates comparison among sectors and classification nodes while maintaining computational feasibility.

The objective of the analysis was not the explanation of individual firms but the estimation of average predictor contributions and the identification of the most influential variables driving rating forecasts. Given the limited number of predictors considered in the HERA framework and the focus on global importance rankings, a sample size of 40 observations per model was deemed sufficient to obtain stable estimates while substantially reducing computational requirements. Preliminary analyses suggested that predictor rankings remained highly stable across alternative sample sizes, indicating that the reduction in the number of evaluated observations did not materially affect the interpretation of the results.

4.3.6. Results

To complement the PDP analysis, Global SHAP values were computed for the terminal classifiers of the HERA hierarchy (Node4-Node9). Since these nodes are directly responsible for the assignment of the final rating classes, they represent the most relevant stage for investigating the financial drivers underlying rating forecasts. Tables 13-15 report the three most influential predictors identified by the SHAP analysis for each terminal node and sector. The rankings are based on the mean absolute Shapley values and therefore reflect the average contribution of each predictor to the classification process. By comparing the results across sectors, it is possible to assess both the stability of the main financial drivers and the presence of sector-specific patterns in the determinants of future rating assignments.

Table 13. Top Three Predictors Selected by SHAP Analysis: Production Sector

Node	Top1	Top2	Top3
4	FCCOV	ROS	DEBPROD
5	DEBCOV	DEBTA	FCCOV
6	DEBTA	FAEQ	DEBCOV
7	FAPROD	ROA	FAEQ
8	ROS	FAPROD	DEBCOV
9	FAEQ	ROA	DEBTA

Note: DEBTA=Debt/Assets; DEBPROD=Debt/Production; ROS=Return on Sales; ROA=Return on Assets; DEBCOV=Debt Coverage; FAPROD=Fixed Assets/Production; FCCOV=Financial Charges Coverage; FAEQ=Fixed Assets/Equity'

Table 14. Top Three Predictors Selected by SHAP Analysis: Trade Sector

Node	Top1	Top2	Top3
4	DEBTA	DEBCOV	ROA
5	FCCOV	DEBTA	DEBCOV
6	FAEQ	DEBTA	FCCOV
7	FAEQ	ROS	DEBTA
8	DEBCOV	ROS	ROA
9	FAEQ	DEBTA	DEBPROD

Note: DEBTA=Debt/Assets; DEBPROD=Debt/Production; ROS=Return on Sales; ROA=Return on Assets; DEBCOV=Debt Coverage; FAPROD=Fixed Assets/Production; FCCOV=Financial Charges Coverage; FAEQ=Fixed Assets/Equity'

Table 15. Top Three Predictors Selected by SHAP Analysis: Services Sector

Node	Top1	Top2	Top3
4	FCCOV	FAPROD	DEBPROD
5	DEBCOV	FAEQ	DEBTA
6	FAEQ	DEBTA	ROA
7	FAPROD	ROA	FAEQ
8	FAPROD	ROA	DEBCOV
9	FAEQ	FAPROD	ROA

Note: DEBTA=Debt/Assets; DEBPROD=Debt/Production; ROS=Return on Sales; ROA=Return on Assets; DEBCOV=Debt Coverage; FAPROD=Fixed Assets/Production; FCCOV=Financial Charges Coverage; FAEQ=Fixed Assets/Equity'

To complement the information provided by the predictor rankings, Global SHAP results are presented through a graphical representation combining overall predictor importance and the distribution of SHAP contributions across observations. Figure 5 illustrates the results obtained for Node 9, the terminal classifier responsible for distinguishing firms rated D from those rated C. Similar figures are reported for the remaining nodes and sectors in Appendix B.

Each panel consists of two complementary components. The left-hand side reports the mean absolute SHAP value associated with each predictor. These values provide a ranking of variable importance by measuring the average magnitude of the contribution generated by each financial indicator, regardless of its sign. Consequently, predictors located at the top of the ranking exert the strongest influence on the classification process.

The right-hand side displays the distribution of SHAP values for the same predictors across all observations included in the analysis. The boxplots provide information that cannot be obtained from importance rankings alone, namely the variability, dispersion, and direction of predictor contributions. Positive SHAP values indicate observations for which a given variable increases the probability of belonging to the positive class of the node, whereas negative values indicate observations for which the corresponding variable decreases that probability. The width of the distributions and the presence of outliers further highlight the heterogeneous role that the same financial indicator may play across different firms.

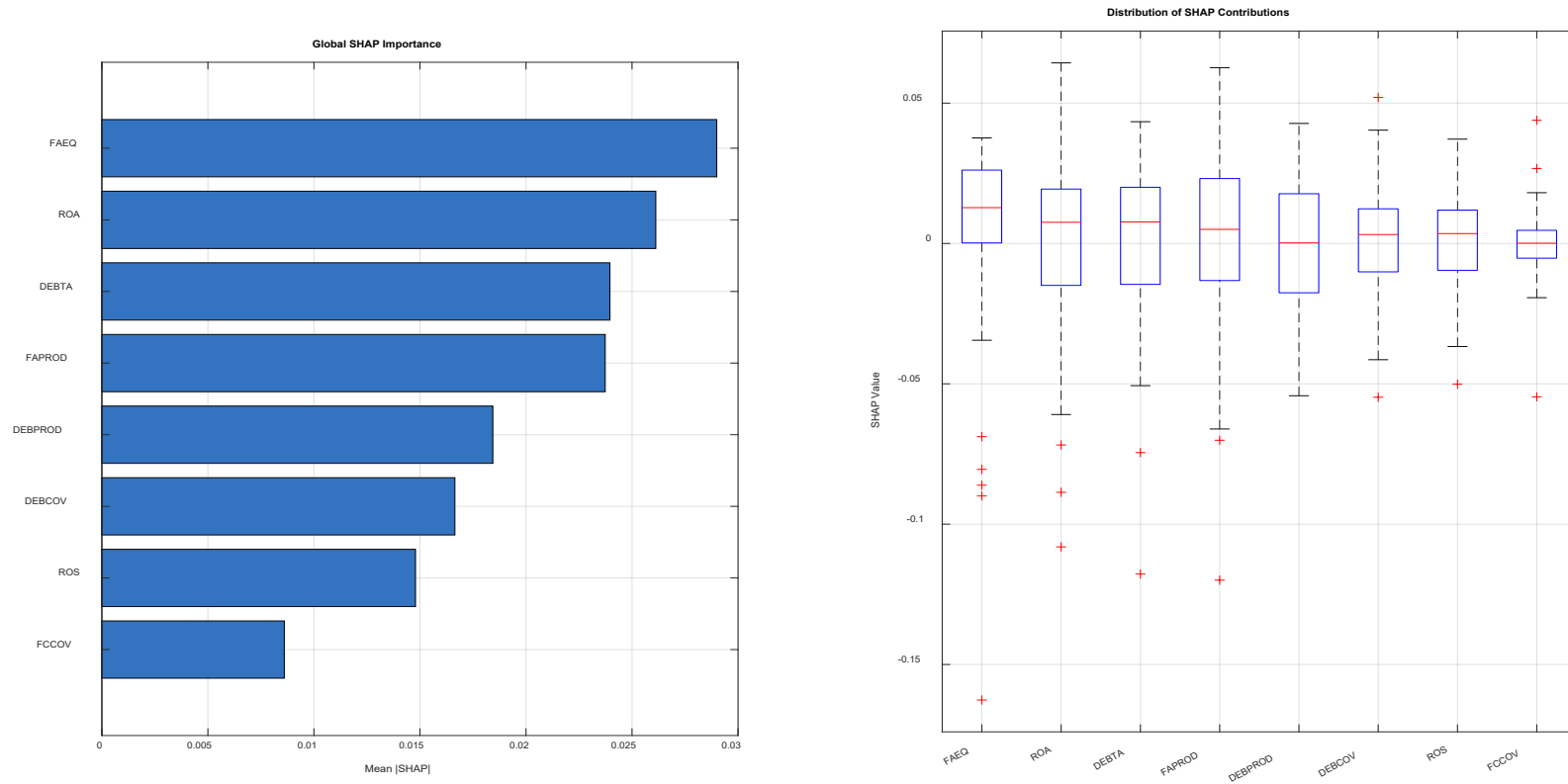
Taken together, the two panels provide a more complete interpretation of the model than either approach alone. While the mean absolute SHAP values identify the variables that matter most in the classification process, the boxplots reveal how their contributions are distributed among firms, offering additional insights into the mechanisms through which the HERA framework generates rating forecasts.

Although several predictors identified through Global SHAP analysis coincide with those emerging from permutation importance rankings, important differences also arise. This finding is not unexpected, as the two approaches quantify different dimensions of predictor relevance. Permutation importance measures the deterioration in predictive performance caused by

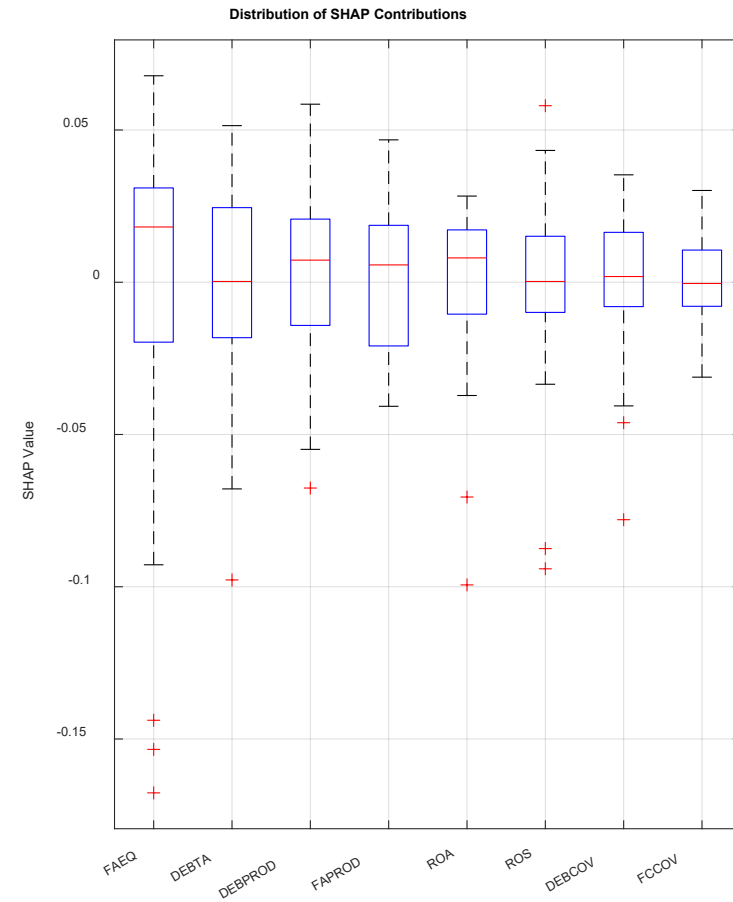
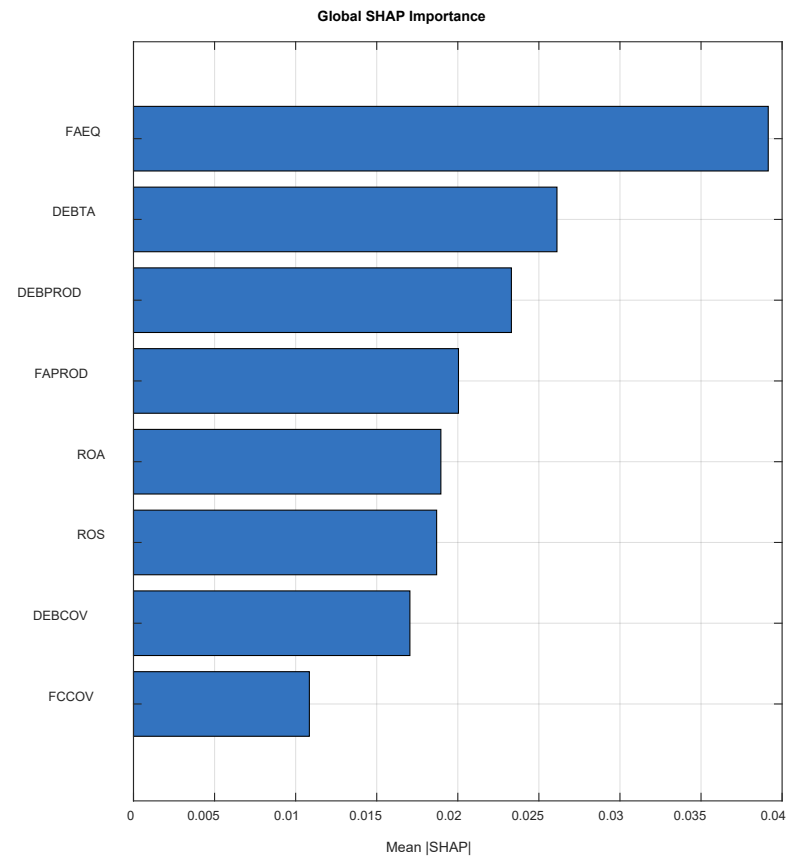
randomly destroying the information contained in a predictor, thereby assessing its contribution to model accuracy. In contrast, Global SHAP evaluates the average contribution of a predictor to model predictions, independently of its impact on classification performance. For instance, let us consider the distinction Node 9. According to permutation importance, the most influential predictor is DEBTA in the Production sector, DEBPROD in the Trade sector, and FAEQ in the Services sector. Conversely, the Global SHAP analysis identifies FAEQ as the leading predictor in all three sectors. This result suggests that FAEQ consistently contributes to the prediction process across observations, while the debt-related indicators appear more important in terms of predictive accuracy within their respective sector-specific models. The comparison therefore provides complementary perspectives on the financial drivers underlying the HERA framework: permutation importance highlights the variables most critical for classification accuracy, whereas SHAP reveals those exerting the greatest influence on model predictions.

Figure 5: Global SHAP Distribution for Node 9 (D vs C): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)

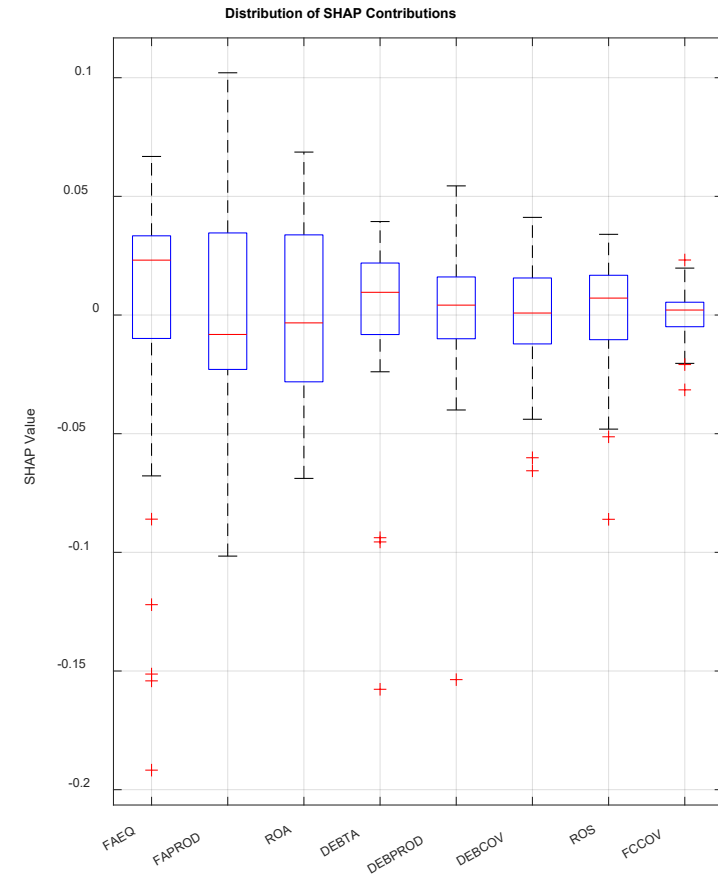
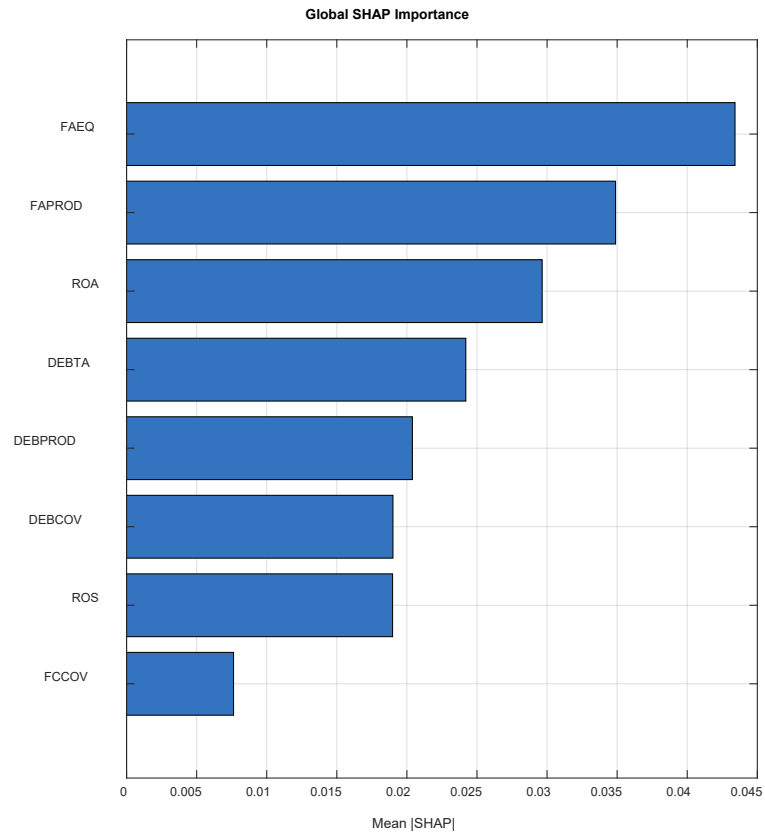
Production – panel a



Trade – panel b



Services – panel c



5. ILLUSTRATIVE POLICY SIMULATION

One of the most promising developments of the HERA-XAI framework concerns its potential use as a decision-support system. Once the main financial drivers of rating classifications have been identified through predictor importance measures, PDP analyses, and SHAP explanations, the framework can be used to explore how changes in firms' financial conditions affect predicted rating probabilities.

The objective of this section is not to develop a fully-fledged policy simulation framework, which is left to future research, but rather to provide an illustrative example of the potential applications of HERA-XAI. Specifically, the analysis investigates how hypothetical changes in selected financial indicators modify the probabilities generated by the hierarchical rating model. The exercise can therefore be interpreted as a simple what-if analysis designed to illustrate the practical implications of the explainability results.

The previous analyses highlighted that corporate ratings are determined by the combined effect of multiple financial indicators rather than by individual variables considered in isolation. Nevertheless, to better understand the sensitivity of the model to changes in firms' financial conditions, the simulation initially examines the effect of varying one indicator at a time while keeping all remaining variables unchanged. This approach makes it possible to identify which financial dimensions are most effective in altering predicted rating probabilities and, potentially, the final rating assignment.

5.1. Methodology

The explainability analyses presented in the previous sections identify the financial indicators that contribute most to rating forecasts. However, predictor importance measures, PDPs, and SHAP values remain essentially descriptive tools. To illustrate the potential use of HERA-XAI as a decision-support framework, an additional policy simulation exercise was conducted.

The simulation focuses on RF6 (Node 6), the classifier responsible for distinguishing firms belonging to the CCC rating category (Class 0) from those belonging to the B category (Class 1). This node was selected because the CCC class represents firms characterized by substantial financial vulnerability, whereas firms rated B, although still speculative, are not generally considered to face the same degree of financial distress. Consequently, the transition from CCC to B may be interpreted as a meaningful improvement in credit quality.

The analysis was performed on a randomly selected Production-sector firm belonging to the RF6 node. The firm was initially classified using the Youden threshold identified during the threshold calibration phase. The predicted probability generated by RF6 correctly assigned the firm to the CCC category, providing the baseline scenario for the simulation exercise.

Subsequently, a series of what-if simulations was conducted by modifying one financial indicator at a time while keeping all remaining variables unchanged. For each indicator, values were progressively varied within a range from -30% to $+30\%$, using increments of 5 percentage points. After each modification, the probability generated by RF6 was recalculated and compared with the Youden threshold in order to determine whether the predicted rating class remained unchanged or shifted from CCC to B.

The simulation was initially performed for all financial indicators included in the HERA framework. Interestingly, the variables producing the most relevant changes in classification outcomes were ROA (Return on Assets), DEBPROD (Debt-to-Production Ratio), and ROS (Return on Sales). These variables were not among the top-ranked predictors identified by either the permutation importance analysis, which highlighted FAEQ, DEBTA, and FAPROD, or the Global SHAP analysis, which identified FAEQ, DEBTA, and DEBCOV as the most influential predictors.

5.2. Results

This finding suggests that the variables exerting the largest influence on model predictions are not necessarily those generating the strongest response under hypothetical policy interventions. While permutation importance and SHAP measure predictor relevance within the fitted model, the policy simulation evaluates how rating probabilities react to explicit changes in firms' financial conditions. The two approaches therefore provide complementary perspectives on the behaviour of the HERA framework.

Table 16 reports the numerical results of the simulation for ROA, DEBPROD, and ROS, whereas Figure 6 provides a graphical representation of the corresponding probability trajectories.

Table 16 reports, for each simulated shock, the predicted probability that the selected firm belongs to Class 1, which in the RF6 classifier corresponds to the B rating category. The table also indicates whether a change in the predicted rating class occurs by comparing the estimated probability with the Youden threshold. Consequently, it is possible to identify the scenarios under which the firm transitions from the CCC category to the B category.

Results presented in Figure 6 show that the response of the rating system is highly non-linear.

The baseline value reported in the SHAP analysis represents the average prediction generated by the Random Forest classifier over the reference sample. It can be interpreted as the expected probability obtained in the absence of firm-specific information. SHAP values then quantify how each financial indicator contributes to increasing or decreasing this baseline prediction. Consequently, the final predicted probability can be decomposed into the baseline value plus the sum of the individual Shapley contributions associated with the firm's financial characteristics.

While substantial variations in individual indicators often produce limited changes in the predicted rating, the simultaneous modification of multiple variables may lead to more pronounced rating transitions. This finding is consistent with the structure of Random Forest classifiers, which base their predictions on the joint interaction of several financial dimensions rather than on isolated indicators.

The analysis also reveals that the relevance of a predictor identified through PDP or SHAP analysis does not necessarily imply that a marginal change in that variable alone will immediately generate a rating upgrade or downgrade. Instead, predictor importance reflects the contribution of a variable within the overall decision process, whereas rating transitions emerge from the combined effect of multiple financial characteristics evaluated simultaneously by the hierarchical classification framework.

To further investigate the simulation capabilities of the HERA-XAI framework, an additional scenario analysis was performed by simultaneously modifying all financial indicators included in the model. Unlike the previous simulations, which evaluated the effect of individual predictors, this experiment was designed to assess the overall sensitivity of rating forecasts to coordinated changes in a firm's financial profile.

For each simulation step, a common shock ranging from -30% to $+30\%$ was applied to all financial indicators. The direction of the adjustment was selected to represent either a deterioration or an improvement in the firm's overall financial condition. Specifically, profitability indicators (ROS and ROA), debt coverage capacity (DEBCOV), and equity support (FAEQ) were increased under positive shocks and decreased under negative shocks. Conversely, leverage indicators (DEBTA and DEBPROD), financial pressure (FCCOV), and capital intensity (FAPROD) were reduced under positive shocks and increased under negative shocks, consistently reflecting an improvement or deterioration in financial sustainability.

For each shock level, the modified financial profile was submitted to the RF6 classifier (Node 6), and the resulting predicted probability was compared with the node-specific Youden threshold. The corresponding rating class was then assigned according to the same decision rule adopted within the HERA simulation framework (Table 17).

Figure 7 illustrates the evolution of the predicted probability as the overall financial shock varies from -30% to $+30\%$. The horizontal dashed line represents the Youden threshold used by

the classifier, while the dotted line indicates the baseline probability associated with the original firm profile.

Several interesting patterns emerge. First, the relationship between financial improvements and predicted probability is clearly non-linear, confirming that the Random Forest classifier does not react proportionally to changes in financial indicators. Second, the trajectory exhibits local increases and decreases even under progressively larger positive shocks. This behavior reflects the complex interaction mechanisms embedded in the Random Forest architecture, where predictions result from the combined effect of multiple variables and multiple decision trees rather than from a simple linear aggregation of financial ratios.

Most importantly, the combined simulation highlights that rating changes are not necessarily driven by a single financial indicator. While the previous PDP and SHAP analyses identified the variables most relevant to the classification process, the simultaneous modification of all indicators produces effects that may differ from those observed when variables are analyzed individually. Consequently, the combined scenario should not be interpreted as the effect of a specific financial ratio, but rather as the response of the rating system to a coordinated improvement or deterioration of the firm's overall financial condition.

Table 16. Policy Simulation Results for RF6 (CCC vs B): Effects of ROA, DEBPROD, and ROS Variations on Predicted Probabilities and Rating Classifications

ROA			DEBPROD			ROS		
Shock %	Prob. to be assigned in B class	Predicted Class	Shock %	Prob. to be assigned in B class	Predicted Class	Shock %	Prob. to be assigned in B class	Predicted Class
-30	0.6	CCC	-30	0.72	B	-30	0.6	CCC
-25	0.64	B	-25	0.76	B	-25	0.6	CCC
-20	0.62	CCC	-20	0.68	B	-20	0.6	CCC
-15	0.62	CCC	-15	0.6	CCC	-15	0.62	CCC
-10	0.54	CCC	-10	0.64	B	-10	0.66	B
-5	0.6	CCC	-5	0.6	CCC	-5	0.66	B
0	0.6	CCC	0	0.6	CCC	0	0.6	CCC
5	0.66	B	5	0.66	B	5	0.66	B
10	0.66	B	10	0.72	B	10	0.68	B
15	0.7	B	15	0.74	B	15	0.62	B
20	0.68	B	20	0.76	B	20	0.72	B
25	0.7	B	25	0.72	B	25	0.76	B
30	0.72	B	30	0.68	B	30	0.72	B

Figure 6. Policy Simulation for RF6 (CCC vs B): Effects of ROA (panel a), DEBPROD (panel b), and ROS (panel c) Variations on the Predicted Probability of Class B

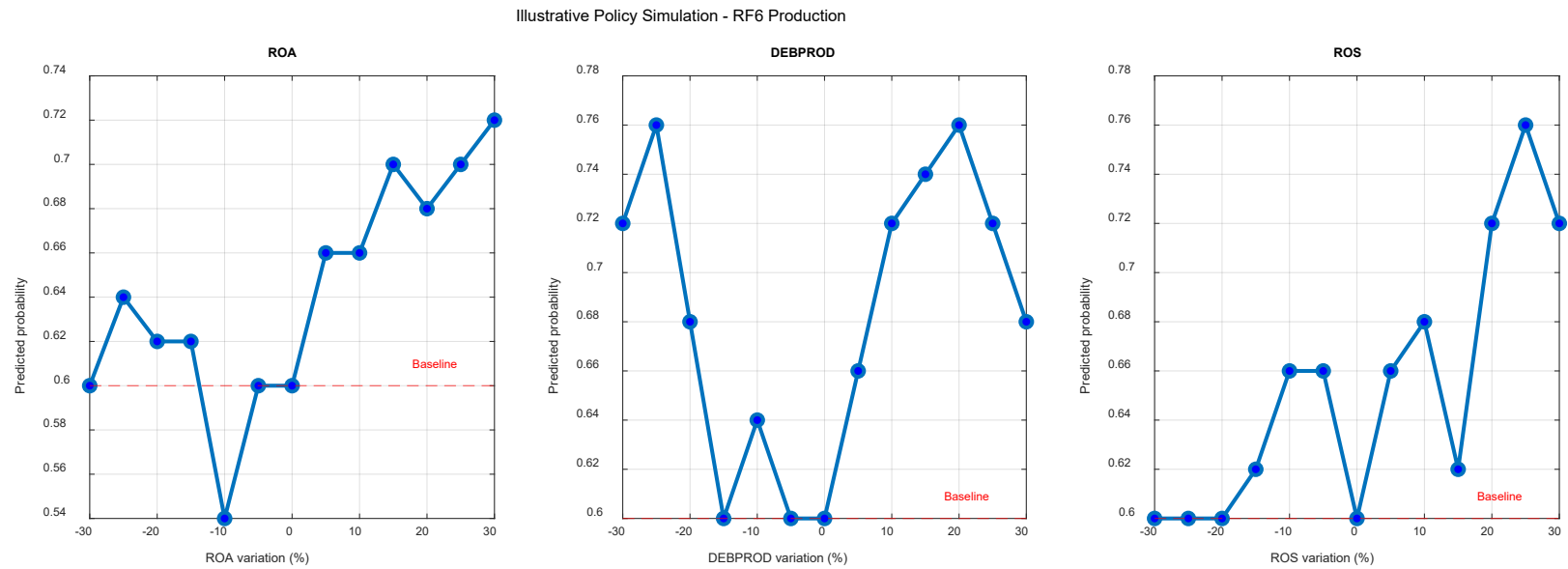
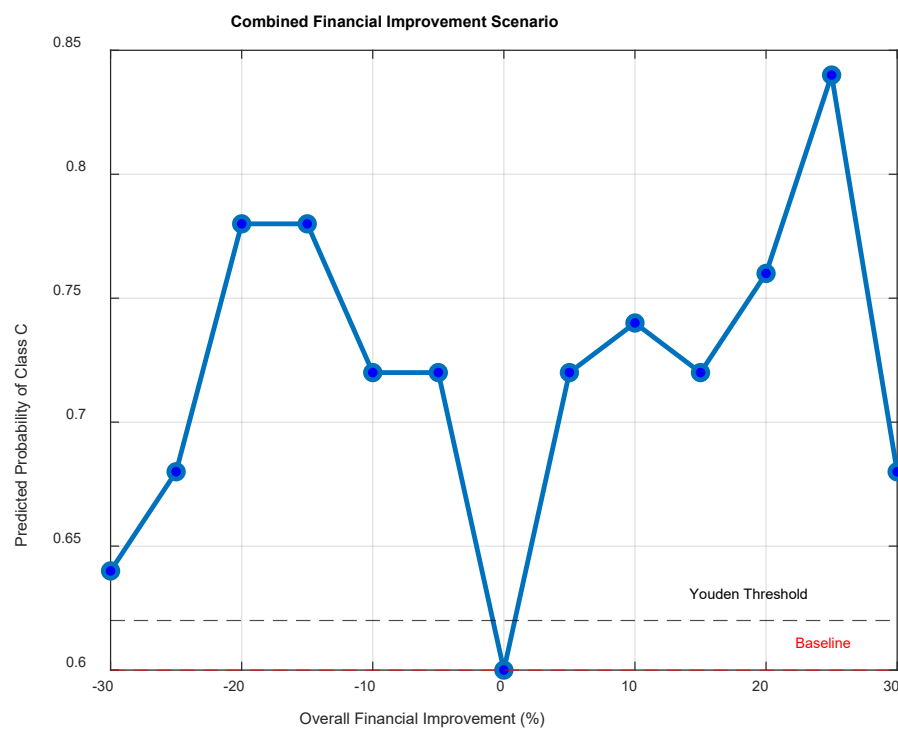


Table 17. Rating Assignments Associated with Progressive Improvements and Deteriorations in Firms' Financial Profiles

Shock%	Prob. to be assigned in B class	Predicted Class
-30	0.64	B
-25	0.68	B
-20	0.78	B
-15	0.78	B
-10	0.72	B
-5	0.72	B
0	0.6	CCC
5	0.72	B
10	0.74	B
15	0.72	B
20	0.76	B
25	0.84	B
30	0.68	B

Figure 7. Response of the RF6 Classifier to Progressive Changes in Overall Financial Conditions



The simulation exercise should be interpreted as a complementary layer of analysis within the HERA-XAI framework rather than as a substitute for PDP and SHAP investigations. While PDPs describe the average effect of individual predictors and SHAP values quantify their contribution

to model decisions, scenario simulations evaluate how changes in firms' financial profiles propagate through the hierarchical architecture and ultimately influence rating assignments. Consequently, the differences observed between variable importance rankings, PDP patterns, SHAP results, and simulated rating transitions should not be regarded as inconsistencies. Rather, they reflect the distinct but complementary perspectives offered by each interpretability tool.

Taken together, these analyses provide a richer understanding of the forecasting process. Variable importance identifies the predictors that are most relevant for classification performance, PDPs reveal the average relationship between predictors and predicted outcomes, SHAP values explain the contribution of individual variables to model decisions, and scenario simulations assess the response of the entire rating system to hypothetical changes in firms' financial conditions. The integration of these components confirms the internal coherence of HERA-XAI and extends the framework beyond forecasting toward an explainable and simulation-oriented decision-support environment.

Although the current simulation exercise remains deliberately illustrative, it demonstrates the potential of the HERA architecture to support increasingly sophisticated scenario analyses. Future developments may allow firms, financial institutions, and policymakers to evaluate the possible consequences of alternative financial and economic strategies before their implementation, thereby strengthening the practical value of the framework for forward-looking decision making.

6. INTERPRETING EXPLAINABILITY RESULTS: SCOPE AND LIMITATIONS

A potential concern is whether the simultaneous use of Permutation Importance, PDPs, and SHAP values is necessary. The three techniques are not intended to provide redundant information but rather to address different dimensions of model interpretability.

Permutation Importance identifies which predictors are most influential for predictive performance and therefore answers the question "Which variables matter most?". PDPs focus on the shape of the relationship between individual predictors and rating probabilities, thereby addressing the question "How does a variable affect the prediction on average?". Finally, SHAP values provide both global and local explanations, quantifying the contribution of each predictor to model outcomes and explaining why a specific prediction was generated.

Consequently, the three approaches should be regarded as complementary rather than alternative tools. Together, they provide a more comprehensive understanding of the decision-making process embedded in the Random Forest classifiers than any single explainability technique considered in isolation.

Another issue concerns the relationship between global and local explainability results. Global methods such as Permutation Importance and Global SHAP summarize the behaviour of the model across the entire sample, whereas Local SHAP explanations focus on the determinants of a specific prediction.

Differences between global and local explanations are therefore not contradictory. A predictor that is highly influential on average may have a limited role for a particular firm, while a relatively less important predictor at the global level may strongly influence an individual prediction.

In this respect, global explanations help identify the general drivers of rating forecasts, whereas local explanations provide firm-specific insights. The two perspectives should therefore be interpreted as complementary levels of analysis.

A further methodological clarification concerns the interpretation of explainability outputs. Neither PDPs nor SHAP values should be interpreted as evidence of causal relationships. These techniques explain how the Random Forest model uses information to generate predictions, but they do not identify the causal mechanisms underlying rating dynamics.

Consequently, the relationships identified through PDPs and SHAP values should be interpreted as predictive associations rather than causal effects. A positive SHAP contribution or a monotonic PDP relationship does not imply that changing a given indicator will necessarily cause an improvement in the firm's rating.

Establishing causal relationships would require additional methodological assumptions and dedicated econometric or causal inference frameworks, which fall outside the scope of the present study.

7. CONCLUSION AND FUTURE DEVELOPMENT

This paper represents a natural extension of the HERA framework originally introduced in Falavigna (2026). While the original study focused on the development of a hierarchical machine learning architecture capable of forecasting one-year-ahead corporate ratings, HERA-XAI expands the framework by introducing a comprehensive explainability layer. In particular, threshold calibration procedures, variable importance measures, Partial Dependence Plots (PDPs), SHAP explanations, and illustrative policy simulations have been integrated within the hierarchical architecture with the objective of understanding the economic and financial mechanisms underlying rating classifications.

The main contribution of HERA-XAI is therefore not merely an improvement in predictive performance, but the transformation of HERA from a forecasting model into an interpretable analytical framework. From a methodological perspective, the study demonstrates that the explainability of hierarchical rating systems cannot rely on a single interpretative technique. Permutation importance measures identify the variables that contribute most to predictive accuracy, PDPs reveal the shape and direction of the relationship between financial indicators and classification probabilities, whereas SHAP values provide both local and global explanations of model behaviour. These methods should therefore be regarded as complementary rather than alternative explainability tools. Permutation importance identifies which variables matter most, PDPs explain how they affect predictions on average, and SHAP values quantify their contribution to both individual and aggregate forecasting outcomes. Taken together, these techniques provide a substantially richer understanding of model behaviour than any single explainability method considered in isolation.

From a practical perspective, the added value of HERA-XAI relative to the original HERA framework lies in its ability to transform rating forecasting into an interpretable decision-support tool. While HERA provides accurate predictions regarding future rating outcomes, it does not explain the mechanisms underlying such forecasts. HERA-XAI addresses this limitation by identifying the financial variables driving rating assignments, quantifying their contribution to model predictions, and illustrating how variations in those variables affect predicted rating probabilities. As a result, managers can move beyond simply observing a forecast and gain a deeper understanding of the financial dimensions that most strongly influence their expected rating trajectory.

A particularly interesting result concerns the differences observed among predictor importance rankings, PDP analyses, SHAP explanations, and policy simulations. Variables identified as highly relevant by permutation importance or SHAP analyses do not necessarily coincide with those generating the largest changes in predicted probabilities during scenario analyses. This finding highlights the importance of combining multiple explainability tools and suggests that predictor relevance, predictor contribution, and policy sensitivity represent distinct dimensions of model interpretation. More generally, the results indicate that understanding why a model assigns a rating and understanding how rating probabilities react to changes in financial indicators are related but conceptually different objectives.

The illustrative policy simulations further support this interpretation. By modifying individual financial indicators and observing the resulting changes in predicted rating probabilities, the analysis provides a first example of how HERA-XAI can be employed in a decision-support context. Although preliminary in nature, these exercises demonstrate that explainability analyses can be used not only to interpret model outputs but also to explore the consequences of hypothetical changes in firms' financial conditions. In this respect, HERA-XAI represents an intermediate step between pure prediction and prescriptive decision-support systems.

At the same time, several limitations should be acknowledged. First, the computation of PDPs and SHAP values remains computationally intensive, particularly within a hierarchical architecture consisting of multiple Random Forest classifiers. For this reason, explainability analyses were performed on representative subsamples of the original database containing approximately four million firm-year observations. Although robustness checks suggested that the results remain stable, future studies could explore more efficient computational strategies for large-scale explainable artificial intelligence applications. Second, the policy simulations presented in this paper are intended solely as illustrative examples and do not yet constitute a fully developed prescriptive framework. Third, the relationships identified through PDPs and SHAP values should be interpreted as predictive associations generated by the Random Forest models rather than as causal effects. While the proposed explainability methods clarify how the model uses information to generate predictions, they do not identify the causal mechanisms underlying rating dynamics.

Several avenues for future research emerge from the present work. A first extension concerns the development of a formal policy simulation environment capable of jointly modifying multiple financial indicators and evaluating their impact on rating trajectories over alternative time horizons. A second direction involves the integration of counterfactual optimization procedures capable of identifying the minimum financial adjustments required to achieve a desired rating outcome. A third extension concerns the comparison of alternative explainability techniques within hierarchical machine learning frameworks. Finally, future versions of the framework could incorporate dynamic scenario generation and prescriptive decision-support modules, transforming HERA into a comprehensive platform for forecasting, explanation, and managerial recommendation.

Overall, HERA-XAI demonstrates that hierarchical artificial intelligence can move beyond the traditional objective of predictive accuracy and become a transparent and interpretable framework for understanding the drivers of corporate creditworthiness. In doing so, the framework answers not only the question “*What rating is expected next year?*” but also “*Why is this rating expected?*” and “*Which financial factors are most responsible for the forecast?*”. Consequently, HERA-XAI represents a crucial intermediate step in the evolution of the HERA family and lays the methodological foundations for HERA-PRO, where explainability is extended towards counterfactual reasoning and prescriptive analytics capable of answering a further question: “*What should change to obtain a better rating outcome?*”

8. SOFTWARE STATEMENT

All empirical analyses, model estimations, threshold calibrations, PDP computations, SHAP explanations, and policy simulations were developed and executed using MATLAB R2024a.

9. REFERENCES

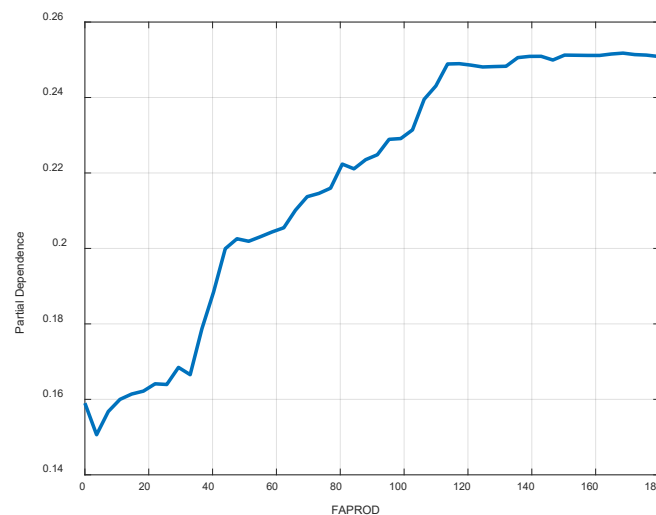
- Falavigna, G. (2012). Financial ratings with scarce information: A neural network approach. *Expert Systems with Applications*, 39(2), pp. 1784-1792.
- Falavigna, G. (2026). *HERA: Hierarchical Engine for Rating Assessment. A Sector-Specific Hierarchical Artificial Intelligence Framework for One-Year-Ahead Corporate Rating Forecasting* (CNR-IRCrES Working Paper 4/2026). Istituto di Ricerca sulla Crescita Economica Sostenibile. <http://dx.doi.org/10.23760/2421-7158.2026.004>
- Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: a systematic literature review: J. Černevičienė, A. Kabašinskas. *Artificial Intelligence Review*, 57(8), 216.
- Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), pp. 1189-1232. <https://doi.org/10.1214/aos/1013203451>

- Ghosh, A., Chakraborty, M., Khatua, S., & Das, S. (2024). Corporate credit rating prediction using explainable AI. *European Economic Letters*, 14(3), pp. 1862-1885. <https://doi.org/10.52783/eel.v14i3.1956>
- Mohsin, M.T., & Nasim, N.B. (2025) Explaining the unexplainable: A systematic review of explainable AI in finance. *International Journal of Science and Research Archive*, 16(03), pp. 476-497. Article DOI: <https://doi.org/10.30574/ijrsra.2025.16.3.2581>.
- Marathe, S.D. (2024). Demystifying interpretable AI in finance: A review of SHAP and LIME. *International Journal of Innovative Research in Technology*, 11(5), pp. 1633-1641.
- Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. In I. Guyon et al. (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), pp. 56-67. <https://doi.org/10.1038/s42256-019-0138-9>
- Oliveira, N.A., & Basso, L.F.C. (2025). Explaining corporate ratings transitions and defaults through machine learning. *Algorithms*, 18(10), 608. <https://doi.org/10.3390/a18100608>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), pp. 206-215. <https://doi.org/10.1038/s42256-019-0048-x>

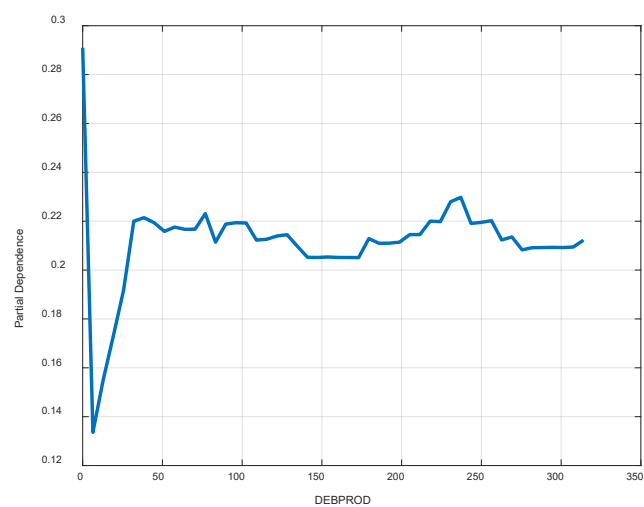
10. APPENDIX A

Figure A1. Partial Dependence Plot for Node 4 (A vs AAA-AA): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)

Production – panel a



Trade – panel b



Services – panel c

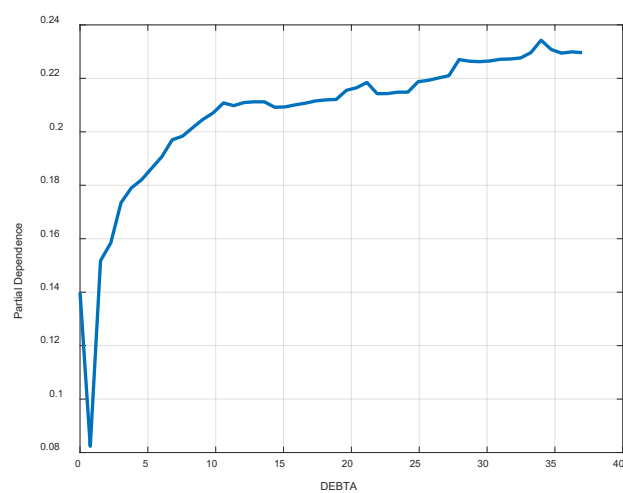
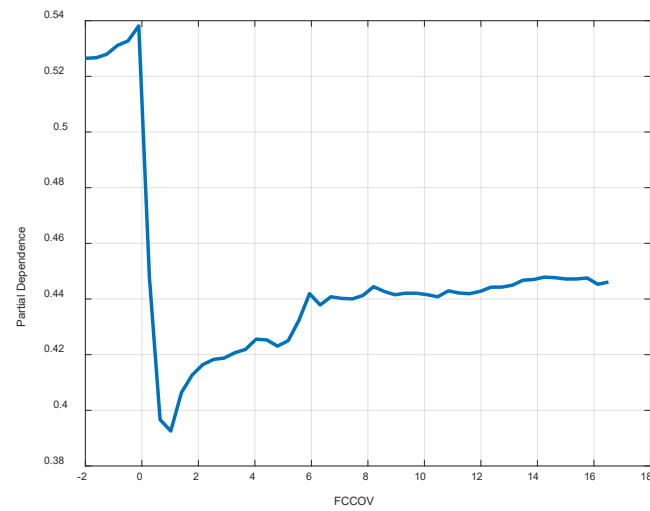
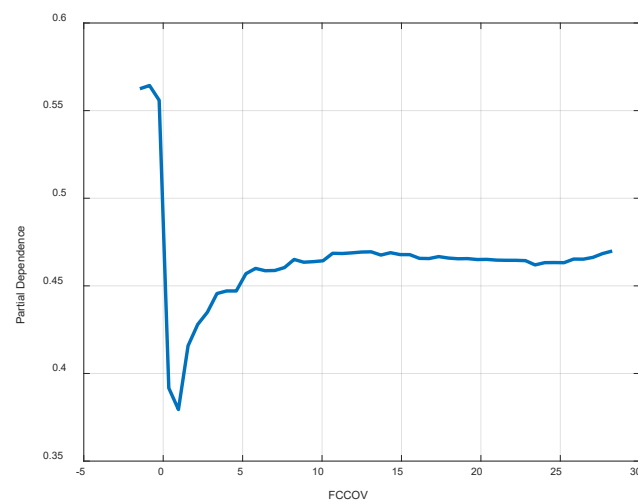


Figure A2. Partial Dependence Plot for Node 5 (BB vs BBB): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)

Production – panel a



Trade – panel b



Services – panel c

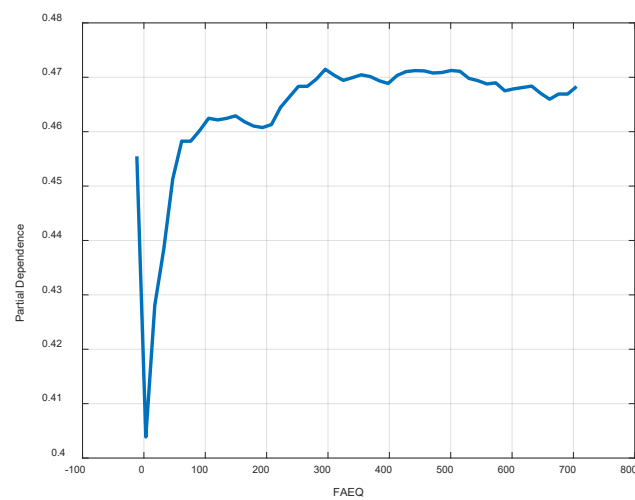
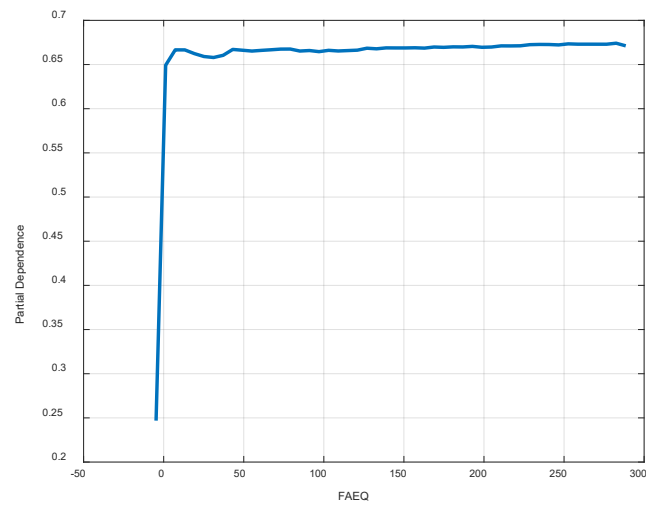
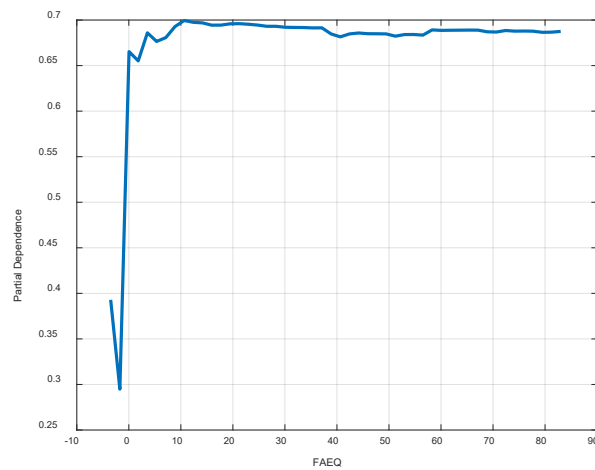


Figure A3. Partial Dependence Plot for Node 6 (CCC vs B): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)

Production – panel a



Trade – panel b



Services – panel c

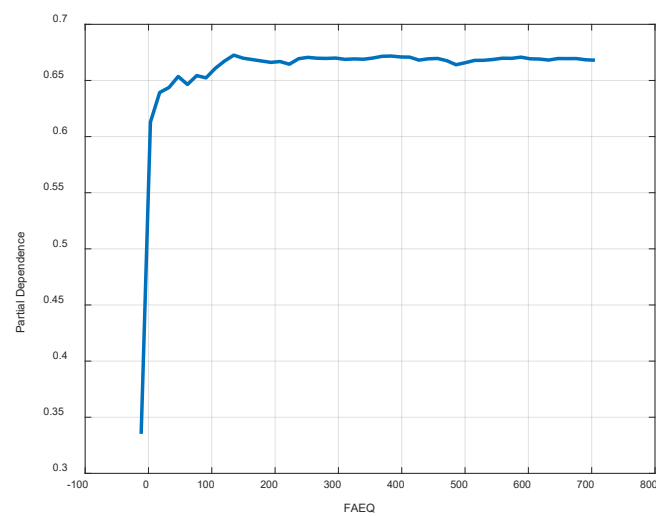
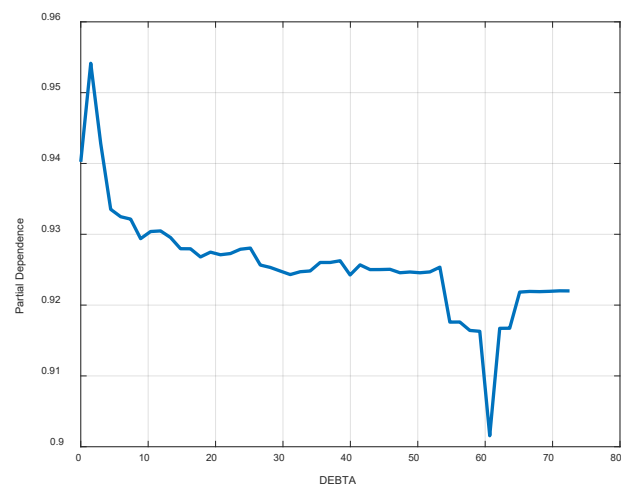
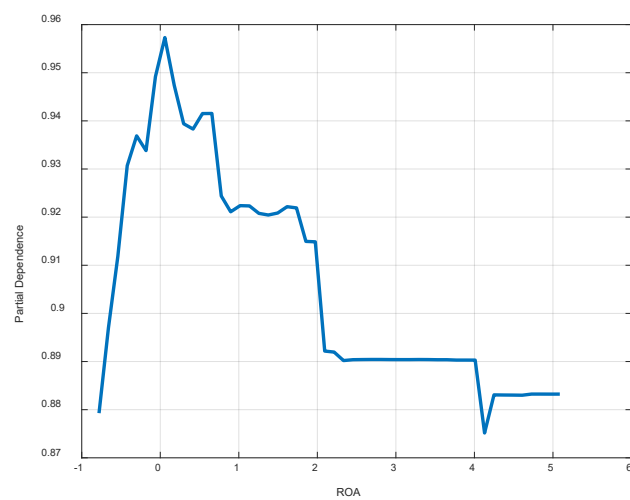


Figure A4. Partial Dependence Plot for Node 7 (D-C vs CC): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)

Production – panel a



Trade – panel b



Services – panel c

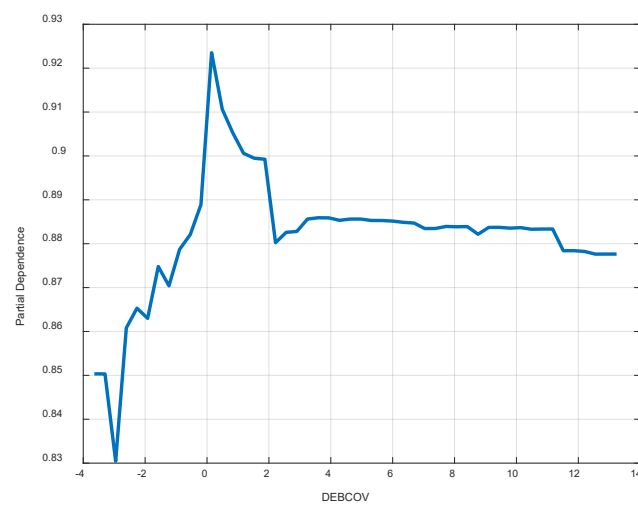
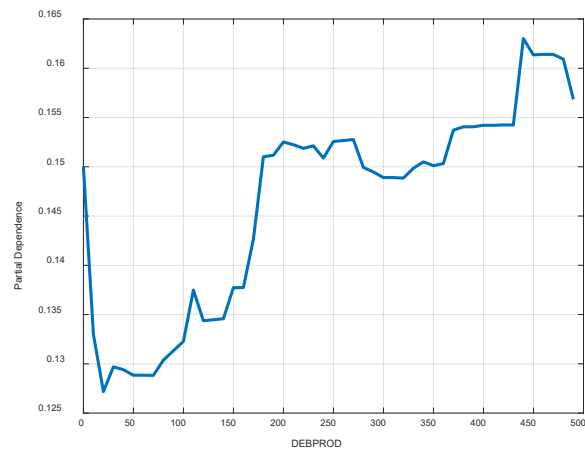
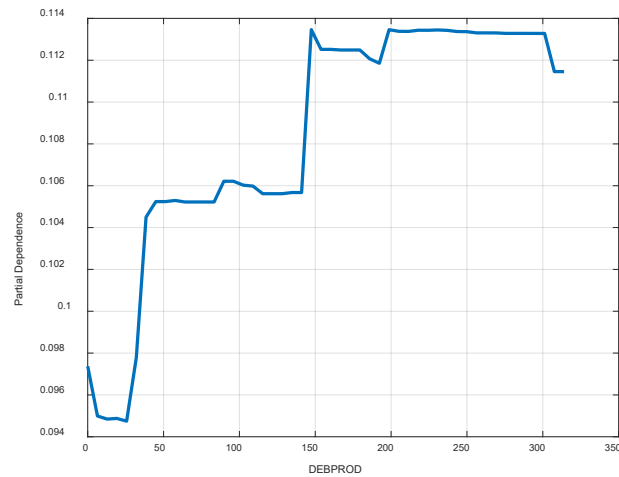


Figure A5. Partial Dependence Plot for Node 8 (AA vs AAA): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)

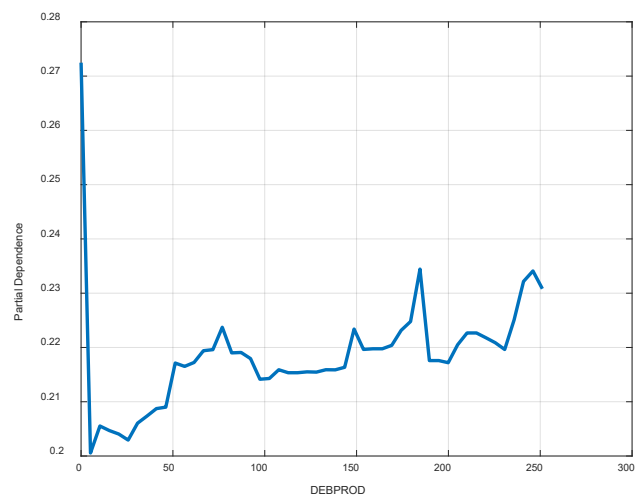
Production – panel a



Trade – panel b



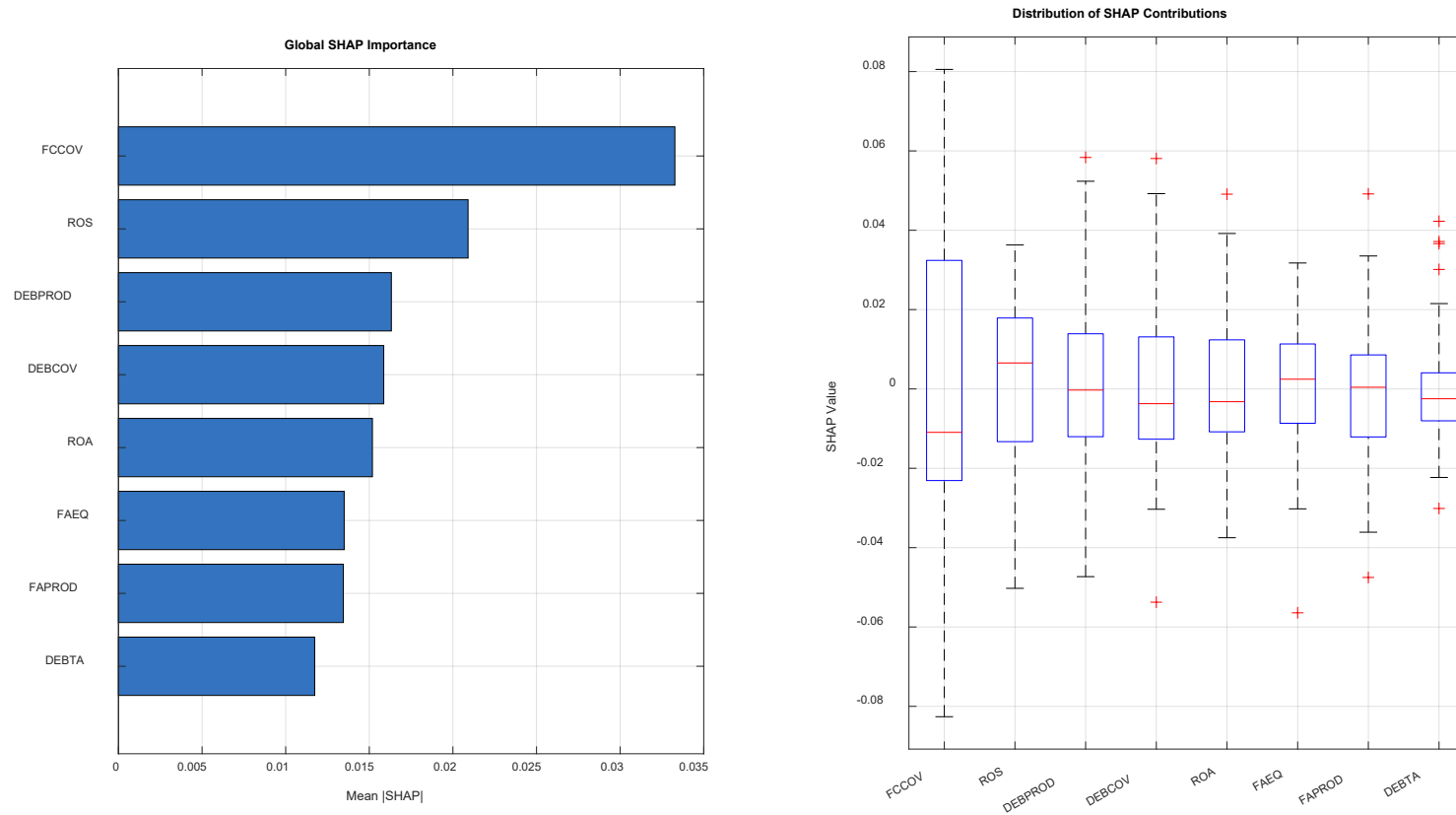
Services – panel c



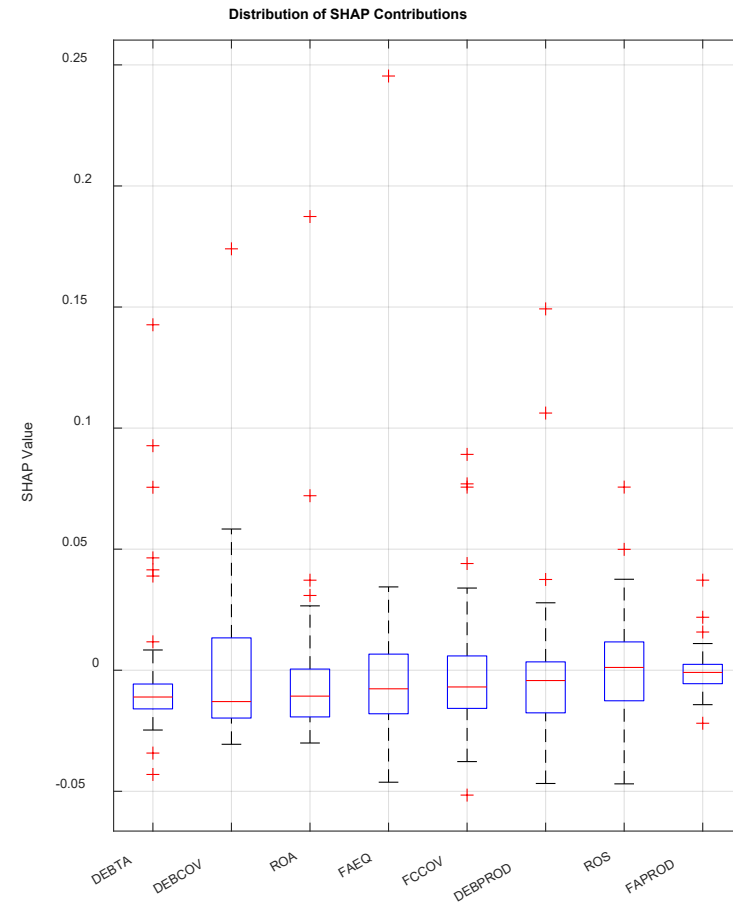
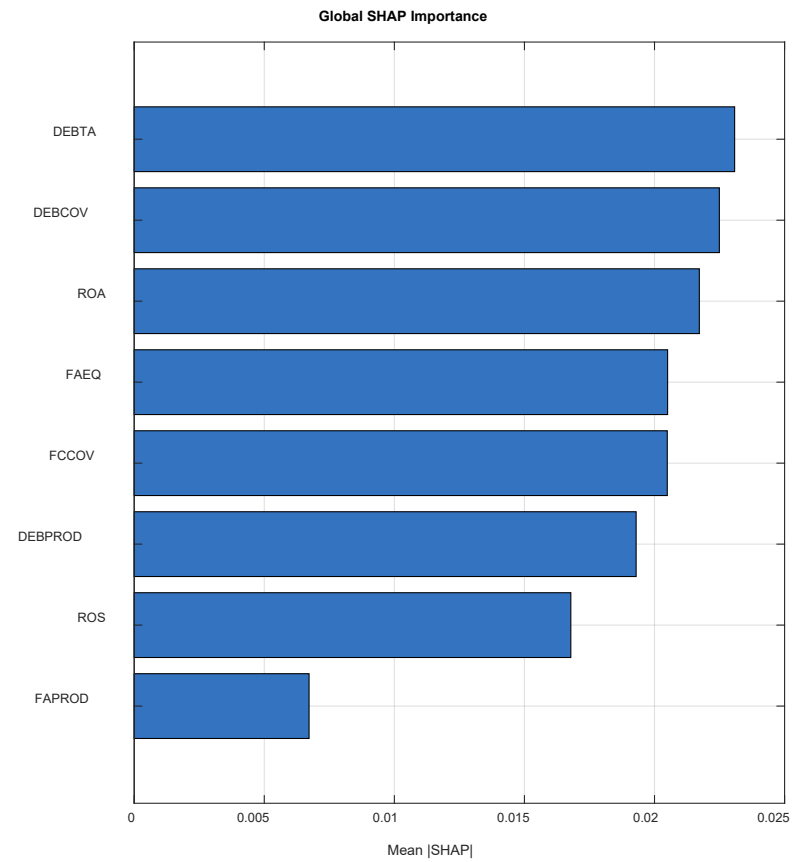
11. APPENDIX B

For the sake of readability, the main financial indicators employed in the HERA-XAI framework are briefly summarized as follows: DEBTA (Debt-to-Assets Ratio), DEBPROD (Debt-to-Production Ratio), ROS (Return on Sales), ROA (Return on Assets), DEBCOV (Debt Coverage Ratio), FAPROD (Fixed Assets-to-Production Ratio), FCCOV (Financial Charges Coverage Ratio), and FAEQ (Fixed Assets-to-Equity Ratio).

Figure B1. Global SHAP Distribution for Node 4 (A vs AA-AAA): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)
Production – panel a



Trade – panel b



Services – panel c

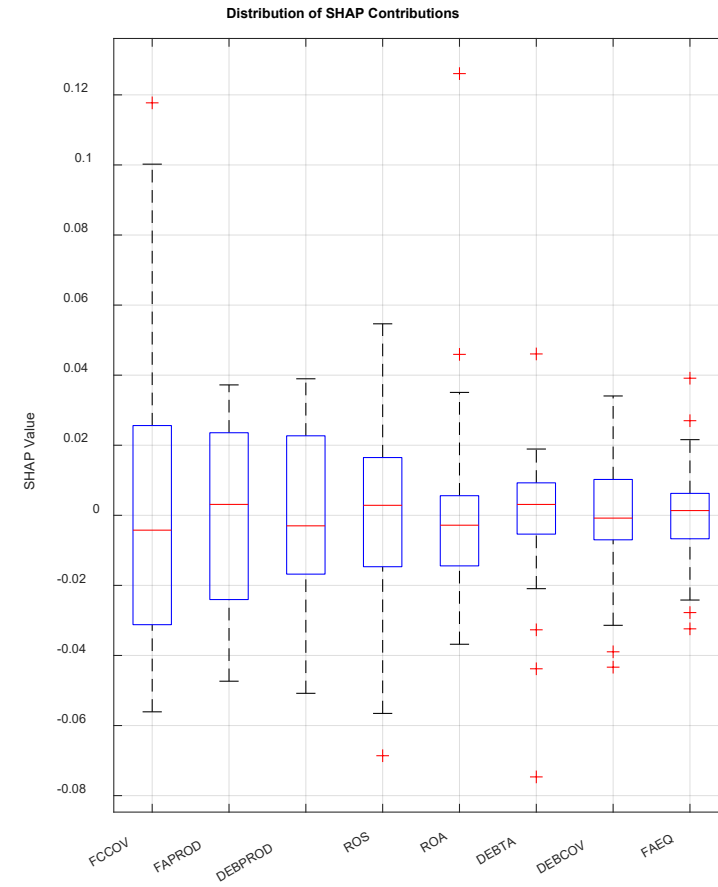
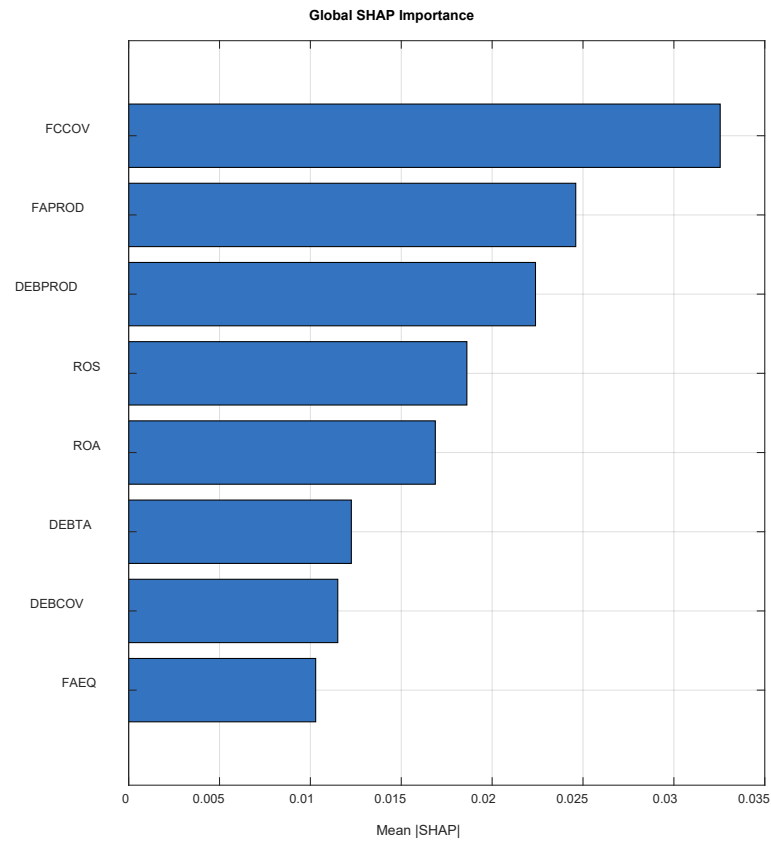
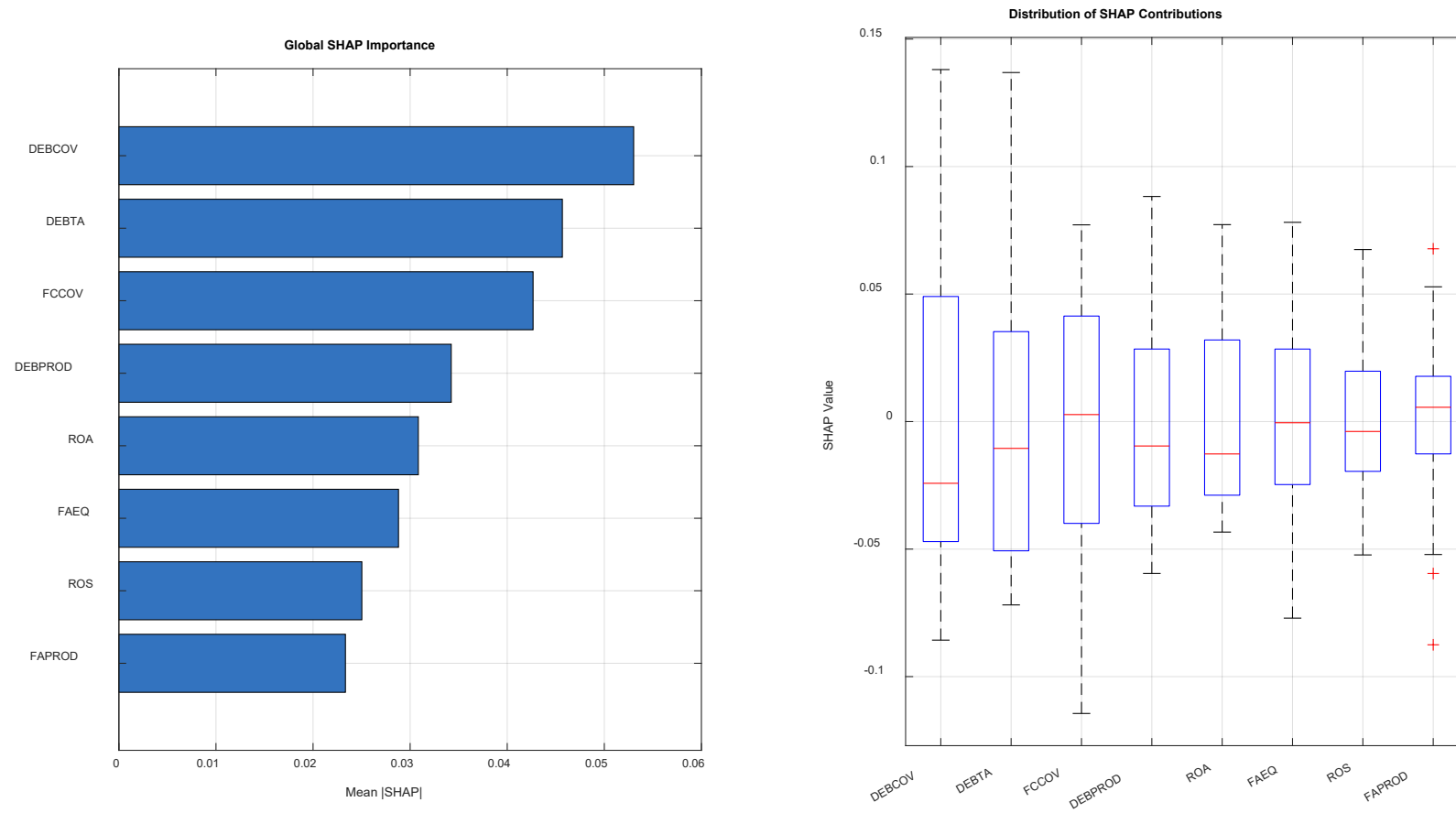
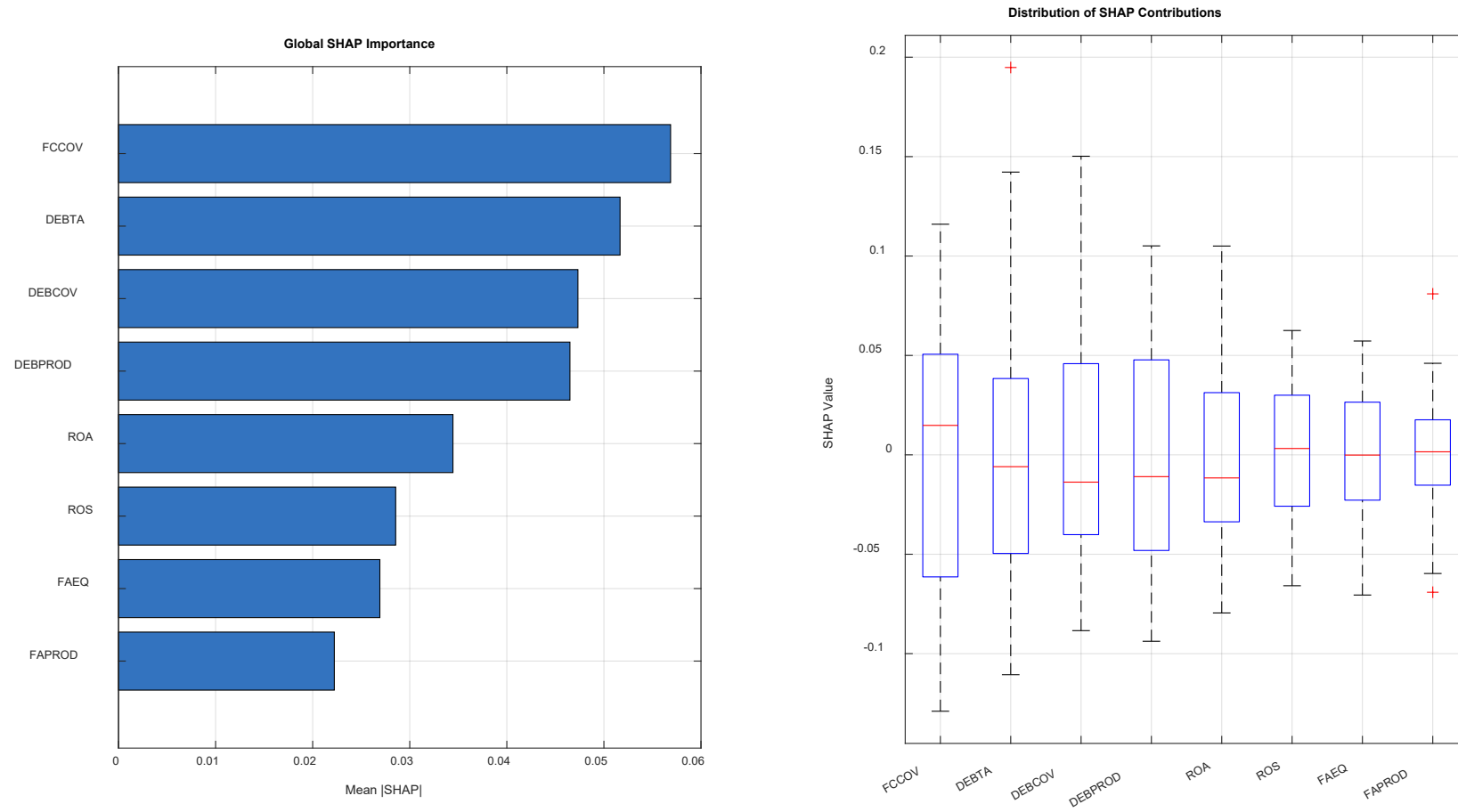


Figure B2. Global SHAP Distribution for Node 5 (BB vs BBB): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)

Production – panel a



Trade – panel b



Services – panel c

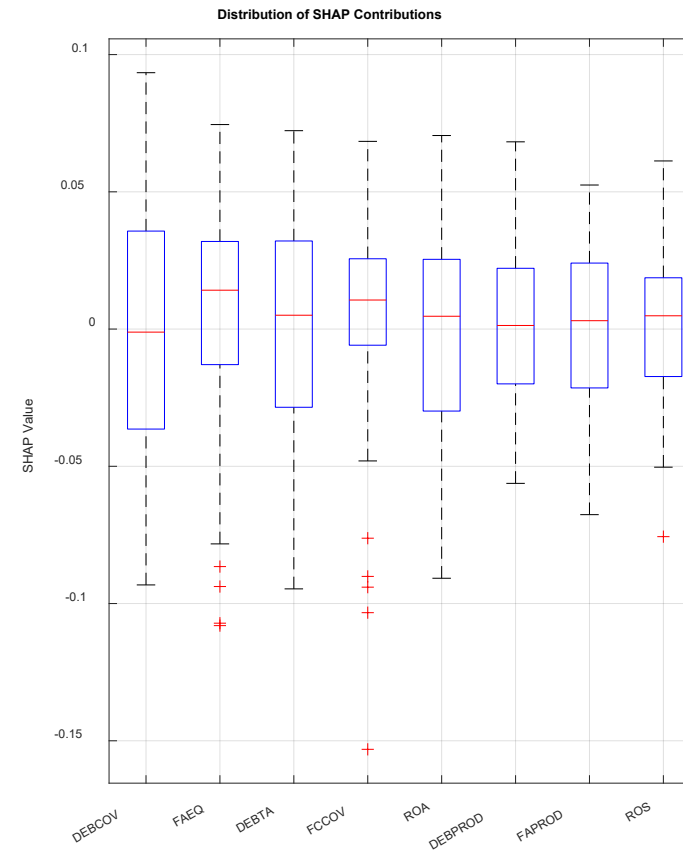
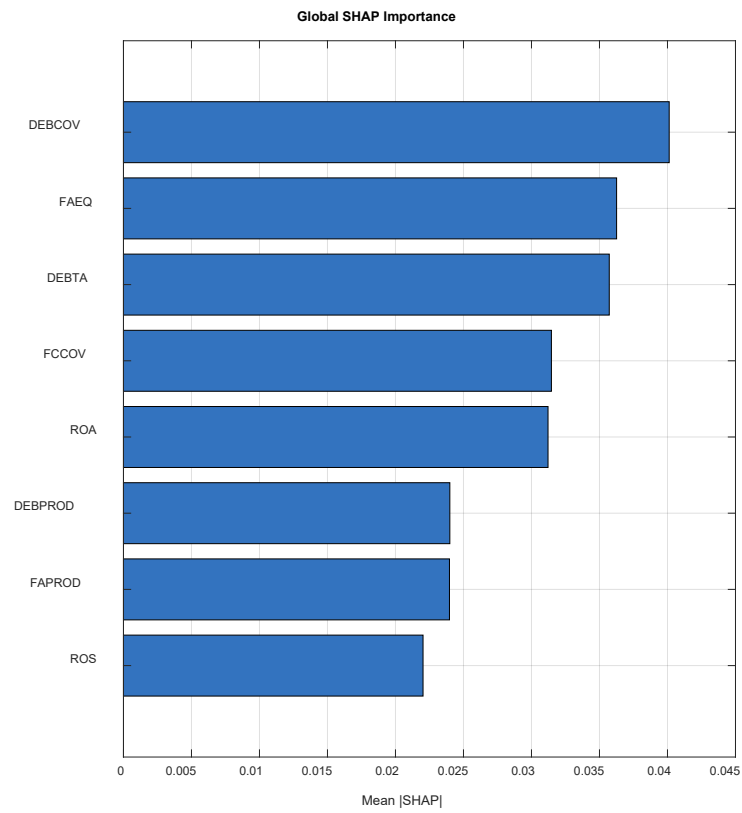
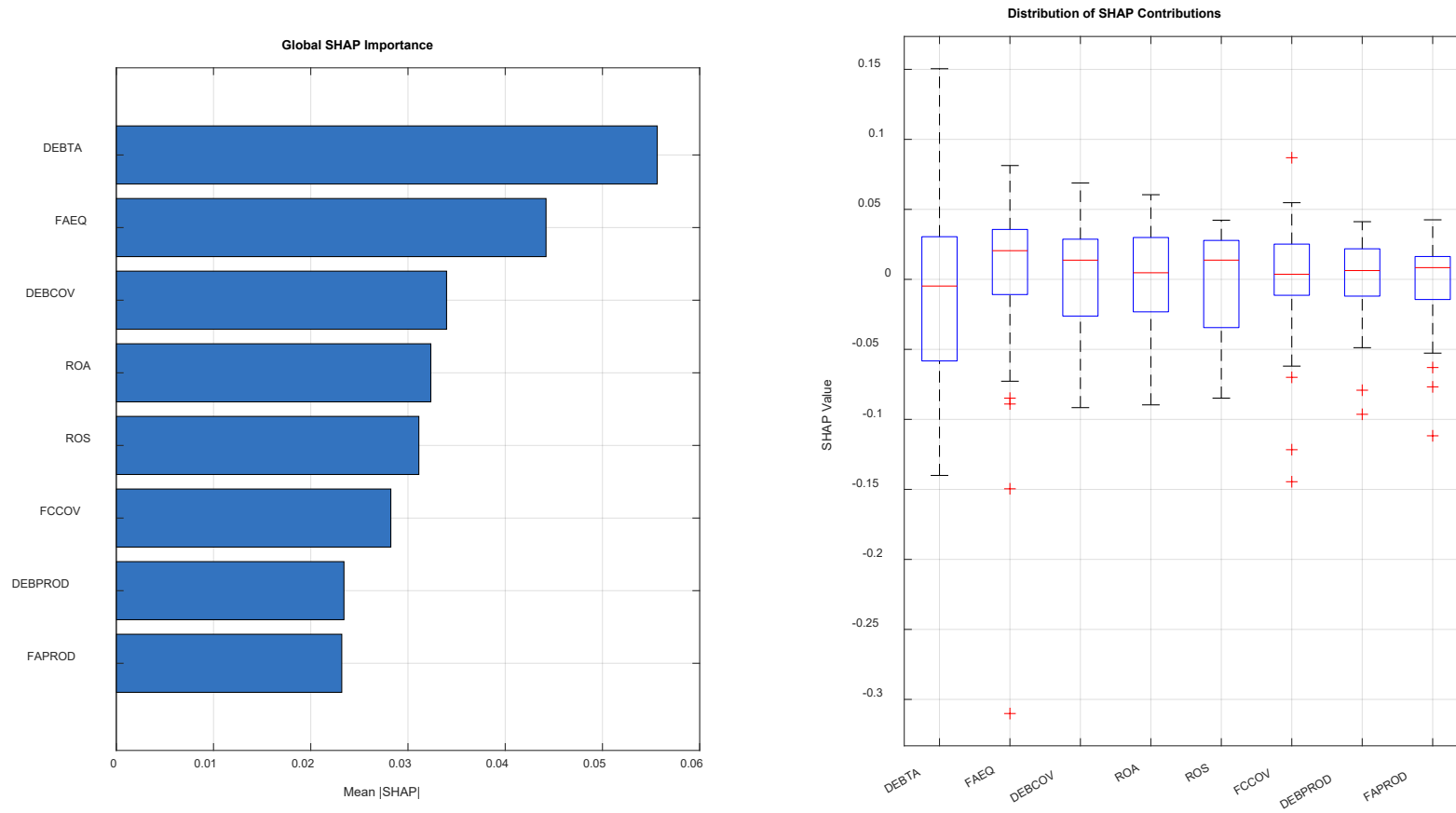
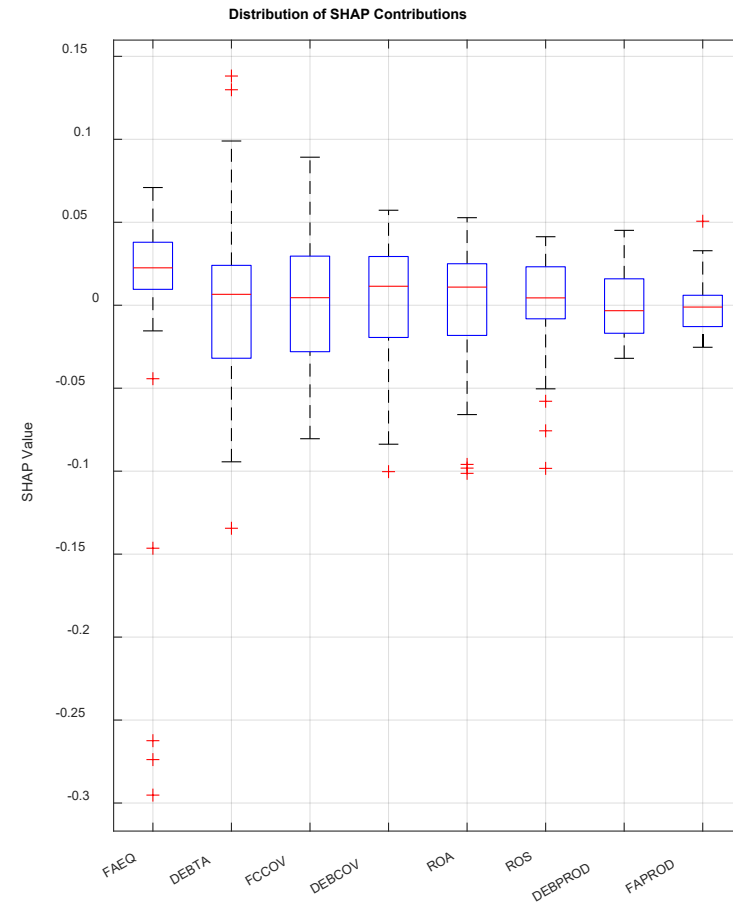
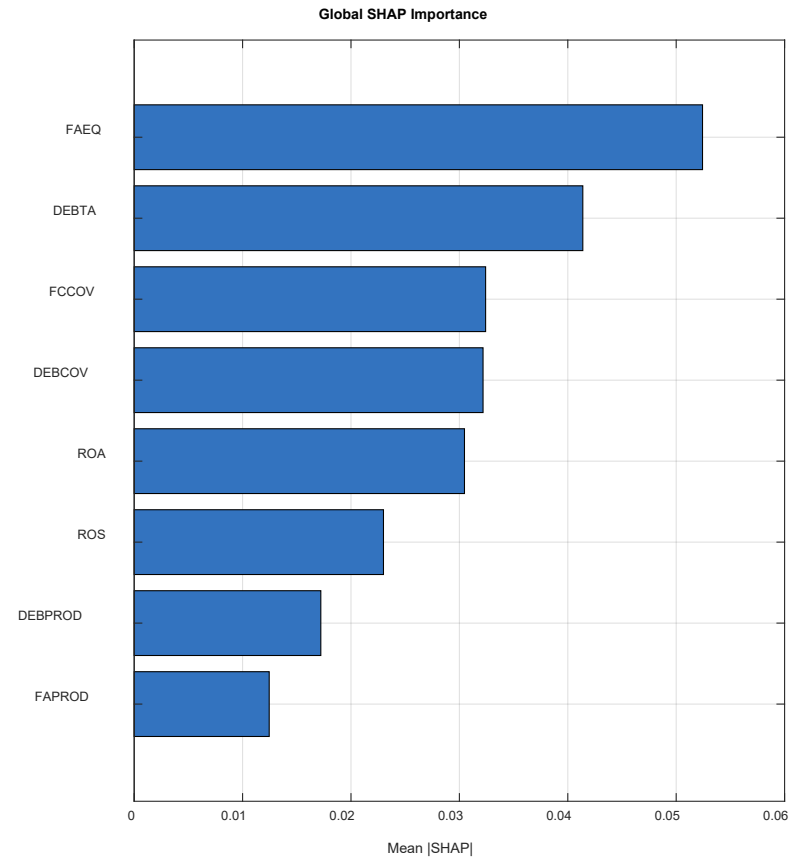


Figure B3. Global SHAP Distribution for Node 6 (CCC vs B): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)
Production – panel a



Trade – panel b



Services – panel c

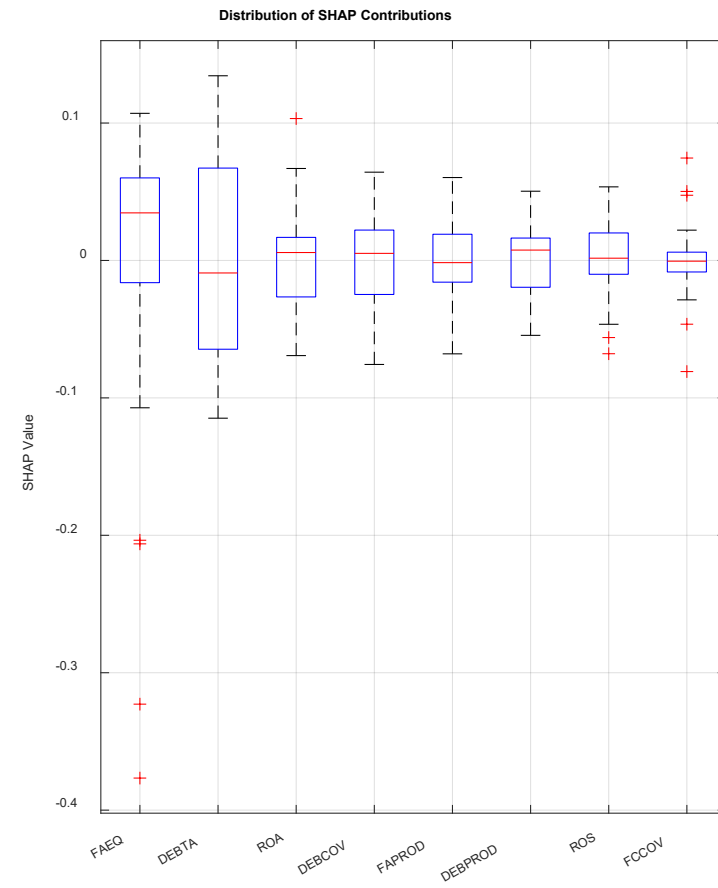
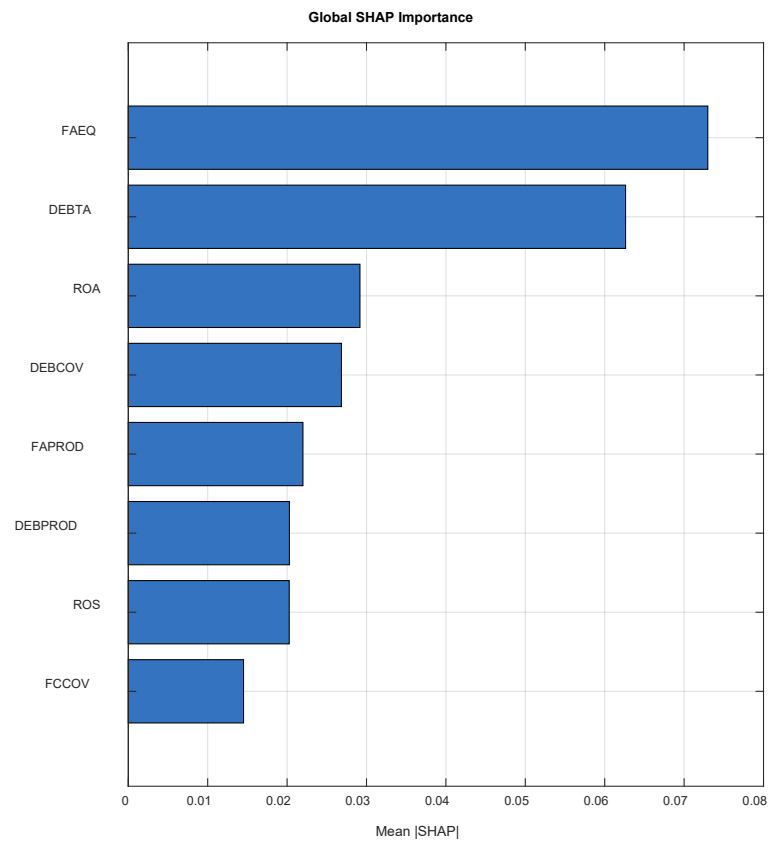
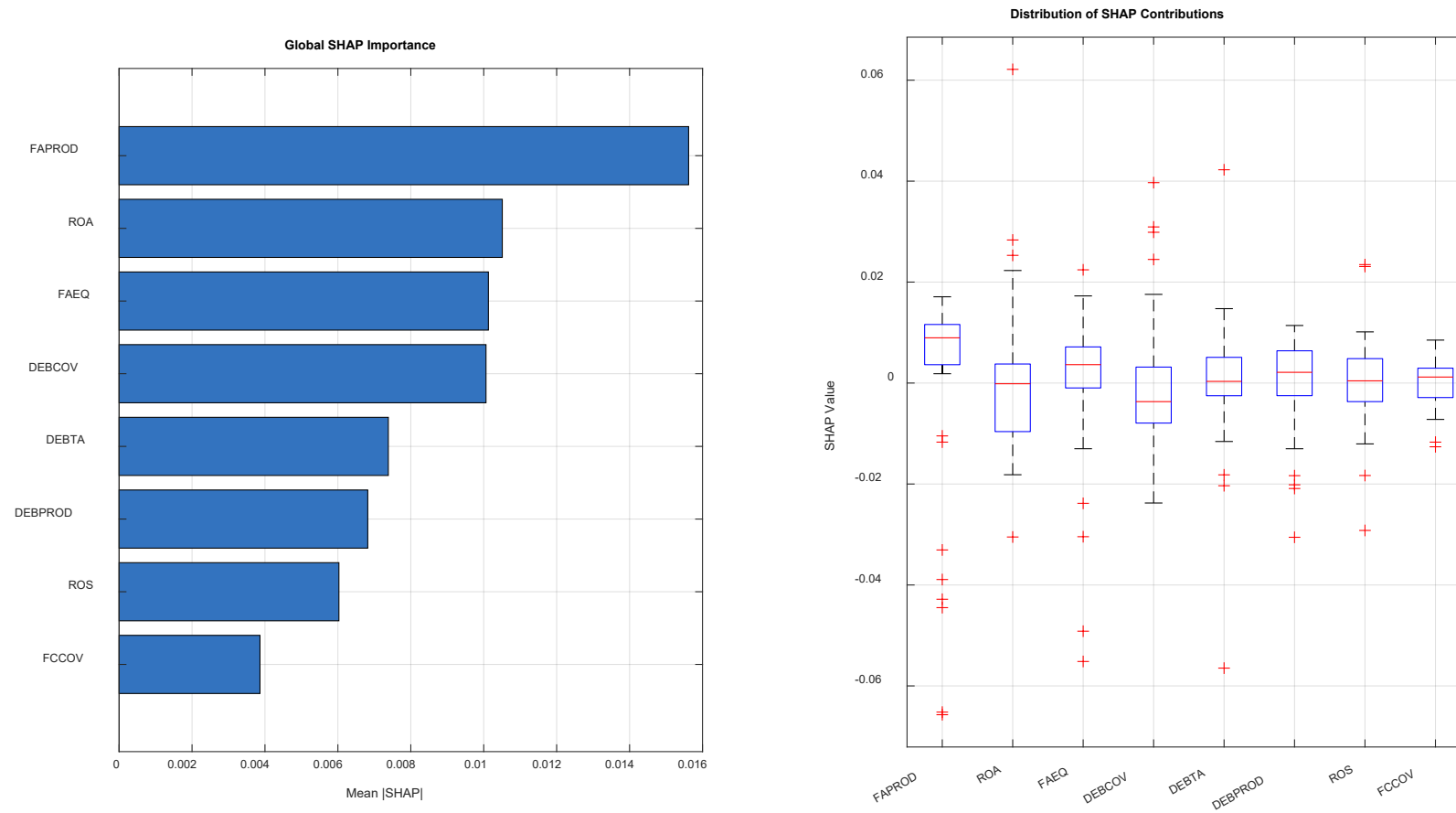
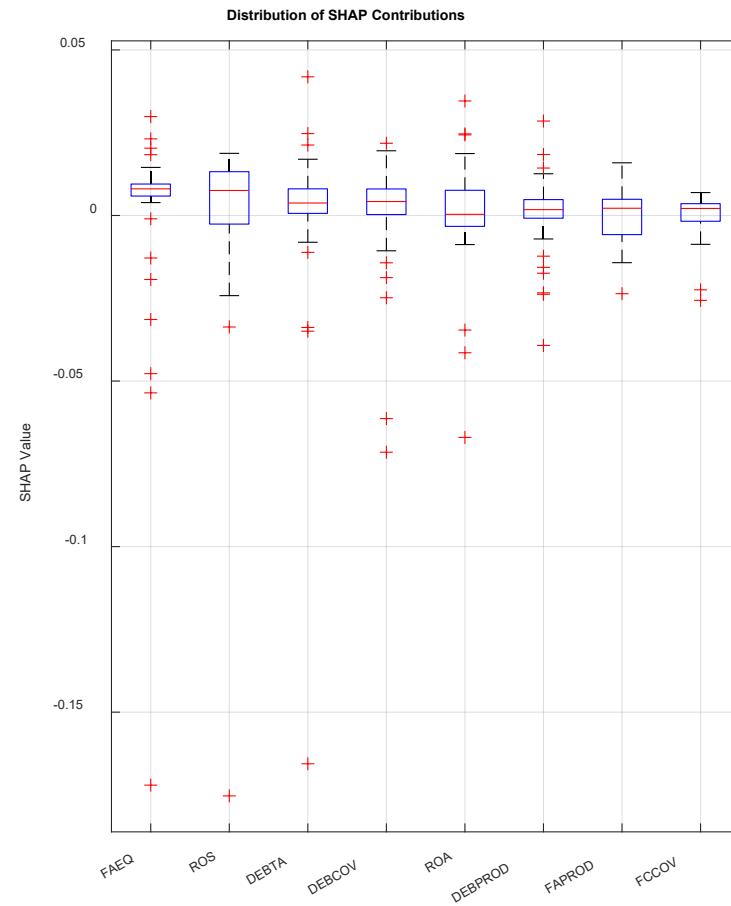
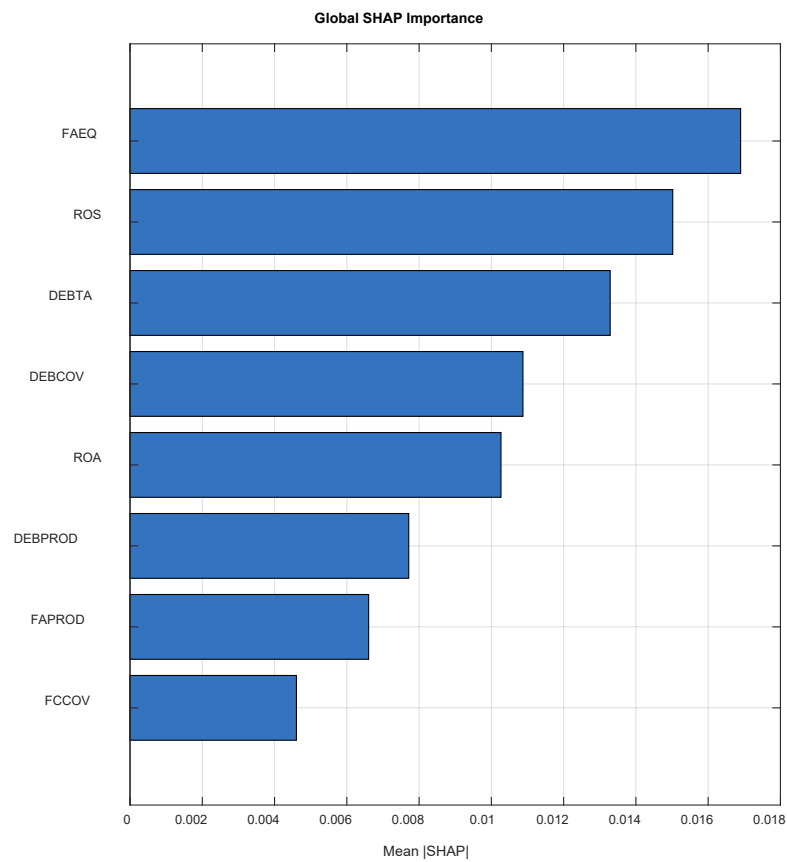


Figure B4. Global SHAP Distribution for Node 7 (D-C vs CC): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)

Production – panel a



Trade – panel b



Services – panel c

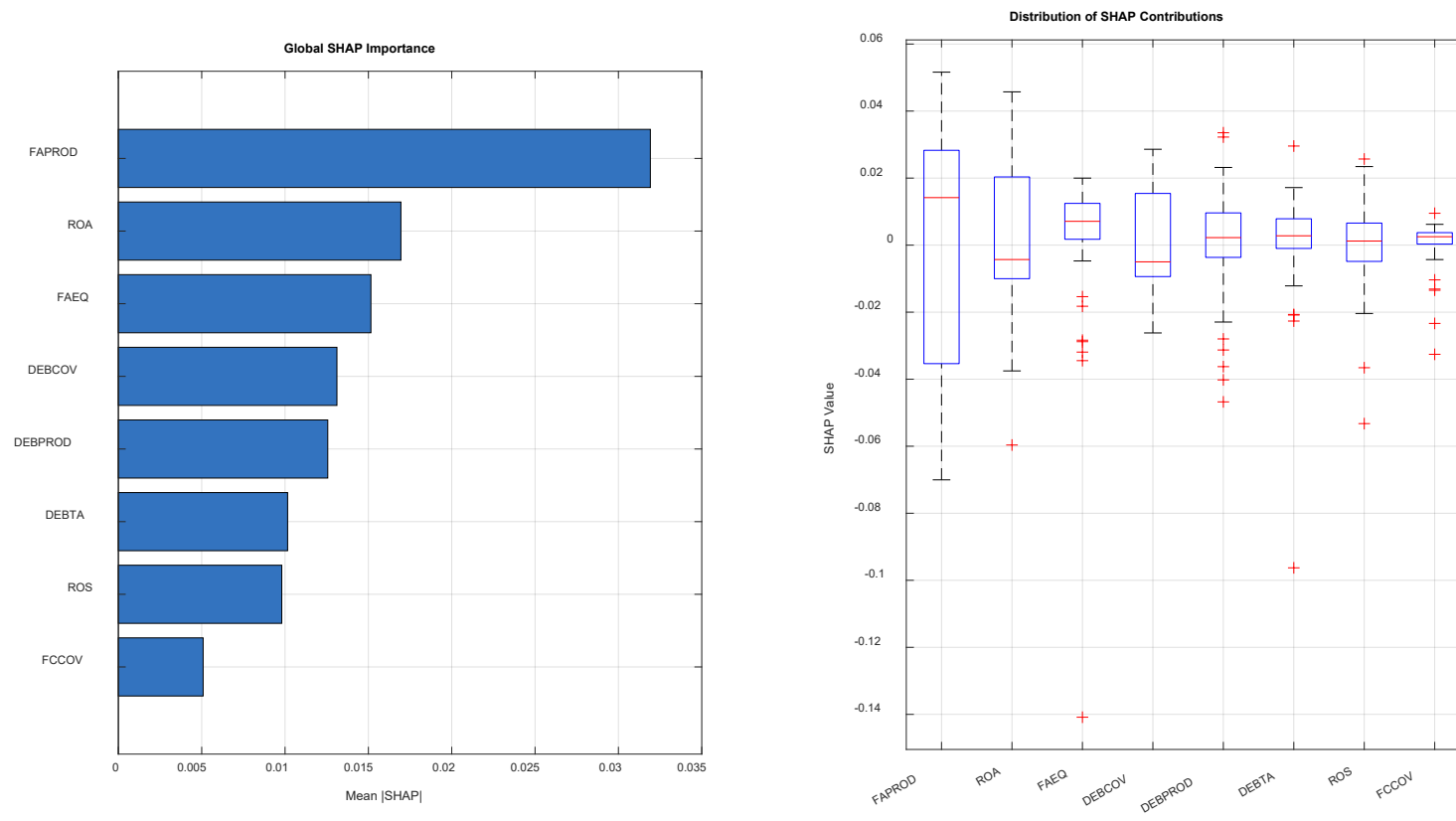
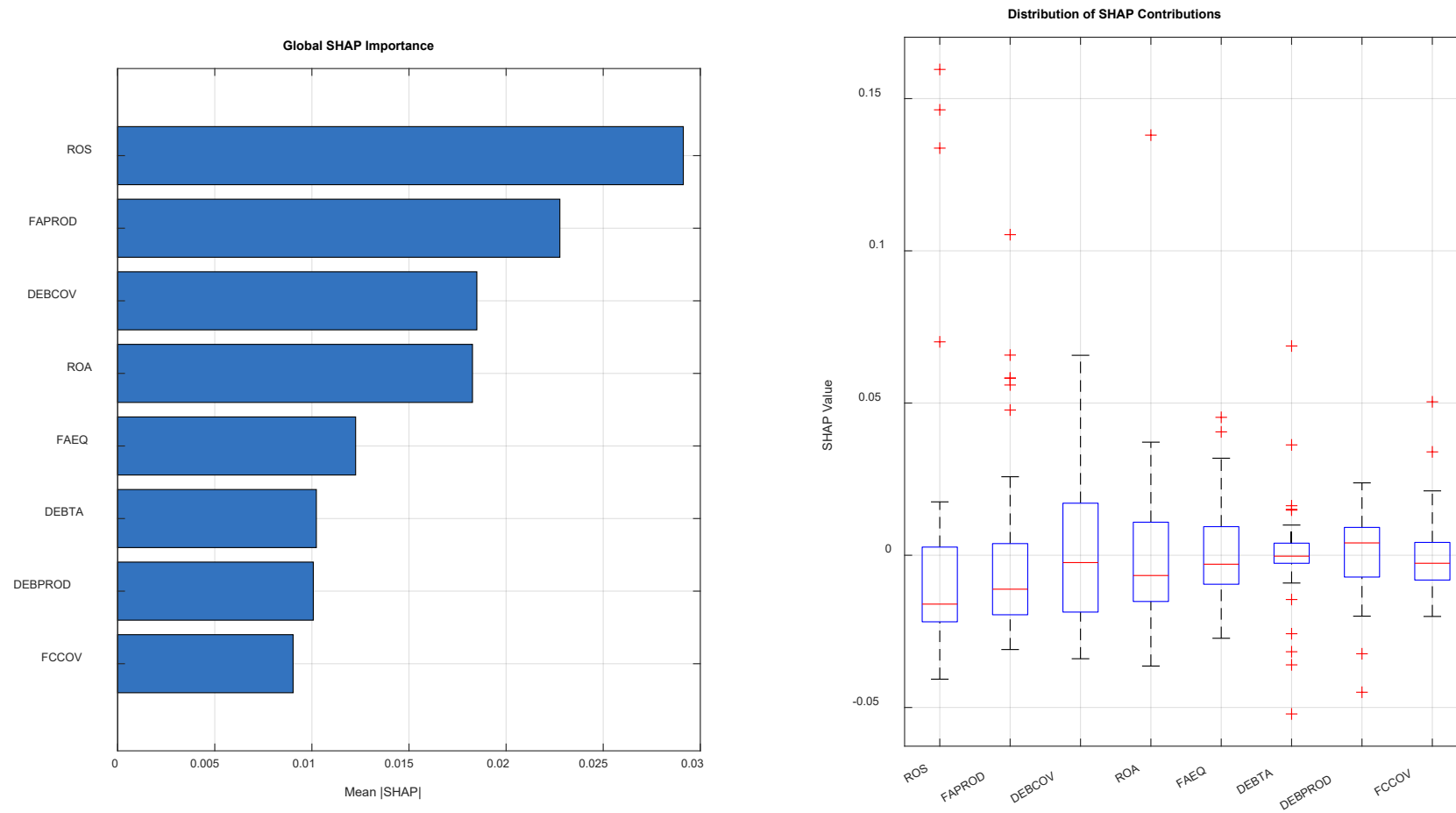
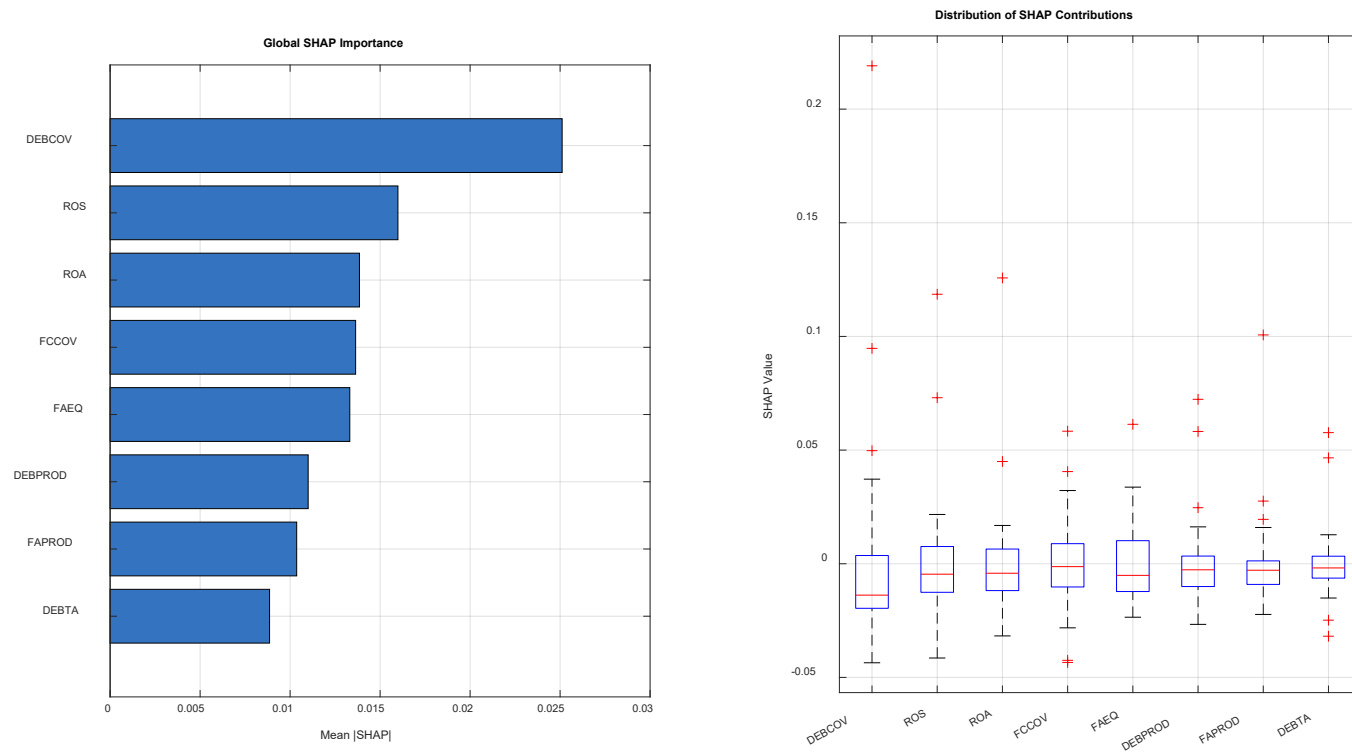


Figure B5. Global SHAP Distribution for Node 8 (AA vs AAA): Production Sector (panel a); Trade Sector (panel b); Services Sector (panel c)

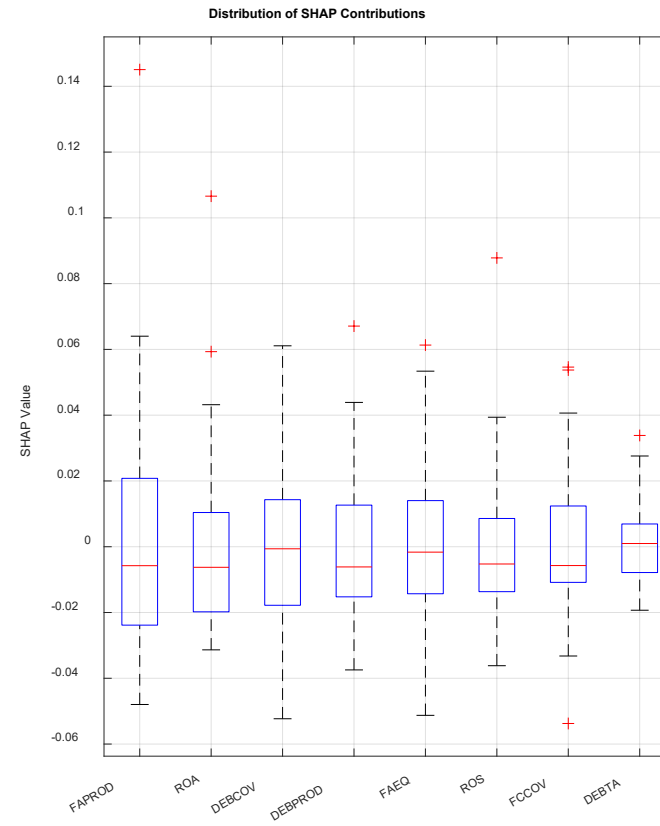
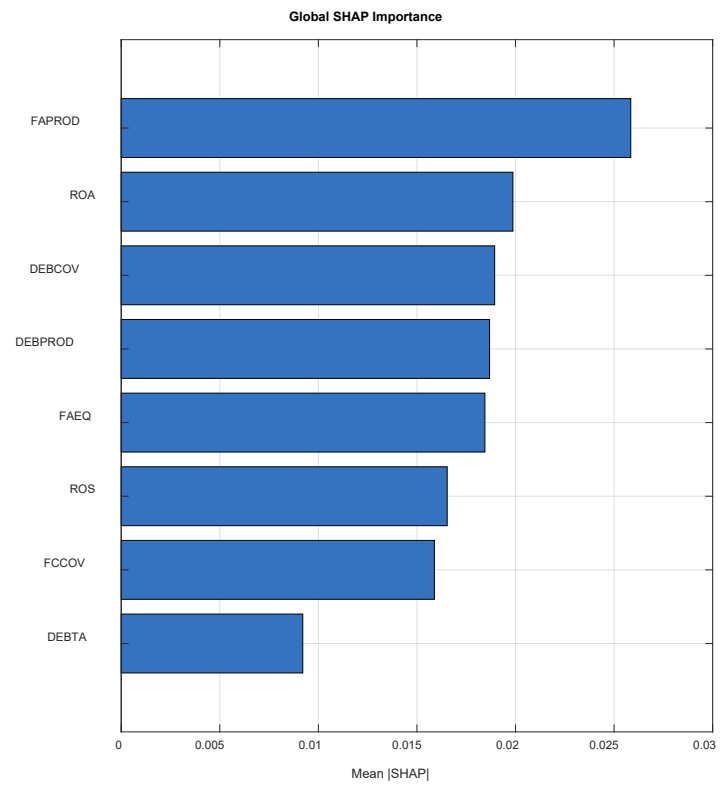
Production – panel a



Trade – panel b



Services – panel c



CNR-IRCrES Working Papers

2026

N.5/2026 [CAN AI ENHANCE INCLUSION? A Participatory Approach to Defining and Evaluating New Technological Innovations](#). Maddalena Bartolini, Elena Ragazzi, Elia Moscoso Thompson, Caterina Rizzo, Alice De Luca, Monica Gori, Katia Lupinetti, Mauro Migliavacca, Lisa Sella, Michela Spagnuolo.

N.4/2026 [HERA: Hierarchical Engine for Rating Assessment. A sector-specific hierarchical artificial intelligence framework for one-year-ahead corporate rating forecasting](#). Greta Falavigna.

N.3/2026 [Il detective sociale: laboratorio di divulgazione delle scienze sociali per la scuola primaria](#). Antonella Emina, Anna Perin, Lisa Sella, Isabella Maria Zoppi.

N. 2/2026 [Il calcio come infrastruttura sociale? Riflessioni ed esempi dal contesto italiano](#). Chiara D'Amato, Luca Tricarico.

N. 1/2026 [Regolamentazione ambientale europea e impatti economici sulle imprese](#). Greta Falavigna, Giampaolo Vitali.

2025

N. 9/2025 [Finanziamento delle comunità energetiche: attori, strumenti e aspettative dell'ecosistema ESG nel contesto italiano](#). Maksym Koltunov. Luca Tricarico, Adriano Bisello, Chiara D'Amato.

N. 8/2025 [Importazioni, esportazioni e strategie green: il caso di studio della chimica in Italia](#). Greta Falavigna, Giampaolo Vitali.

N. 7/2025 [Mapping international collaboration in Nanosciences and Nanotechnologies: a bibliometric and econometric study of South African research](#). Ugo Finardi, Jeanne C.M. Vallette d'Osia.

N. 6/2025 [Dispersione salariale e performance nel calcio in Italia: un'analisi degli incentivi tra sforzo e risultato nella serie A \(2010-2022\)](#). Mario Nosvelli, Matteo Colzato.

N. 5/2025 [Designing a multiple case study approach for studying universities' community engagement](#). Andrea Orazio Spinello, Valentina Carazzolo, Ugo Finardi, Stefania Frongia, Valentina Ghibellini, Emanuela Reale, Andrea Vargiu.

N. 4/2025 [Tools and Methods for Assessing Active Ageing and Inclusion: The Rima Experimental Project](#). Lisa Sella, Nicola Pollo, Valentina Lamonica.

N. 3/2025 [Sinfonie urbane: il ruolo della ritmanalisi nella rigenerazione degli spazi. Il caso di Piazza San Venceslao a Praga](#). Vittoria Virgili, Marco Marucci.

N. 2/2025 [From Single to Multi-Project Participation: SME Absorptive Capacity in Horizon 2020](#). Loreta Isaraj.

N. 1/2025 [Invecchiamento e welfare responsabile: la "condizione anziana" nelle periferie di Roma Capitale](#). Cristiana Di Pietro.

2024

N. 5/2024 [National frameworks for universities' community engagement: perspectives from Italy, France and the UK](#). Valentina Carazzolo, Ugo Finardi, Emanuela Reale, Andrea Orazio Spinello.

N. 4/2024 DAPIL – [Digital archive of pilgrimage literature. Un archivio digitale per la letteratura medioevale di pellegrinaggio](#). Giuseppe Cuscunà.

N. 3/2024 [Thesaurus e glossario ragionato della lingua della IA e della robotica. Thesaurus and reasoned glossary of AI and robotic language](#). Grazia Biorci.

N. 2/2024 [Comprendere le trasformazioni di San Lorenzo, Roma. Un tentativo di ordine concettuale tra gentrificazione e politica urbana](#). Antonia Astore, Luca Tricarico.

N. 1/2024 [How to measure the social acceptability of alternative environmental management solutions in wetlands and other ecosystems](#). Lisa Sella, Francesca Silvia Rota, Nicola Pollo.

2023

N. 5/2023 [Laboratori creativi sperimentali a distanza per la valutazione](#). Isabella Maria Zoppi, Antonella Emina.

N. 4/2023 [Eco-innovations and labor in the European automotive industry: an econometric study](#). Anna Novaresio.

N. 3/2023 [Climate change adaptation planning: tools and methods for effective and sustainable decisions](#). Filippo Frascini, Marco Pregolato.

N. 2/2023 [The effect of propensity to savings on rate of profit](#). Mario De Marchi.

N. 1/2023 [UNESCO o non UNESCO? Quando il brand incentiva il turismo: il caso dei paesaggi vitivinicoli delle Langhe e del Monferrato](#). Orsola Borsani, Greta Falavigna.

2022

N. 5/2022 [Logit statico, Logit dinamico e modelli hazard](#). Franco Varetto.

- N. 4/2022 [Evolution of Deep Learning from Turing machine to Deep Learning next generation](#). Greta Falavigna.
- N. 3/2022 [A Simulation Model of Technology Innovation of a Territory](#). Angelo Bonomi.
- N. 2/2022 [Technology and Environmental Policies](#). Angelo Bonomi.
- N. 1/2022 [Le donne marittime: fra stereotipi di genere, discriminazioni e scarse opportunità occupazionali](#). Barbara Bonciani, Silvia Peveri.
- 2021
- N. 9/2021 [Management of open access research infrastructures in large EU projects: the "CultureLabs" case](#). Andrea Orazio Spinello, Danilo Giglito, Eleanor Lockley.
- N. 8/2021 [Francia-Italia: parole in campo. Intorno alla narrazione del Campionato del Mondo di calcio donne](#). Antonella Emina.
- N. 7/2021 [Covid-19 e rischio di insolvenza: il punto di vista del mercato azionario](#). Franco Varetto.
- N. 6/2021 [Institutional efficiency and budget constraints: a Directional Distance Function approach to lead a key policy reform](#). Greta Falavigna, Roberto Ippoliti.
- N. 5/2021 [Different waves and different policy interventions in 2020 Covid-19 in Italy: did they bring different results?](#). Mario Nosvelli.
- N. 4/2021 [On Search of a General Model of Technology Innovation](#). Angelo Bonomi.
- N. 3/2021 [Design and implementation of a web survey on the effects of evaluation on academic research](#). Andrea Orazio Spinello, Emanuela Reale, Antonio Zinilli.
- N. 2/2021 [An online survey on the effects of agile working in Italian Public Research Organisations](#). Serena Fabrizio, Valentina Lamonica, Andrea Orazio Spinello.
- N. 1/2021 [Technology Transfer Activities in Universities and Public Research Organizations: A Literature Overview](#). Ugo Finardi, Rolfo Secondo, Isabella Bianco.
- 2020
- N. 12/2020 [Unexpected loss multiperiodale e pricing del rischio di credito](#). Franco Varetto.
- N. 11/2020 [La ricerca in Nanotecnologie e Nanoscienze in Italia: spesa del settore pubblico e aree tematiche prevalenti](#). Ugo Finardi, Andrea Orazio Spinello.
- N. 10/2020 [Persistent fast growth and profitability](#). Lucio Morettini, Bianca Poti, Roberto Gabriele.
- N. 9/2020 [Binomio Burnout e Mindfulness nelle organizzazioni. Alcuni studi e scenari di applicazione](#). Oriana Ippoliti, Riccardo Briotti, Bianca Crocamo, Antonio Minopoli.
- N. 8/2020 [Innovation and communication of companies on Twitter before and during COVID-19 crisis](#). José N. Franco-Riquelme, Antonio Zinilli, Joaquín B. Ordieres-Meré and Emanuela Reale.
- N. 7/2020 [The proposal of a new hybrid methodology for the impact assessment of energy efficiency interventions. An exploratory study](#). Monica Cariola, Greta Falavigna.
- N. 6/2020 [The technology innovative system of the Silicon Valley](#). Angelo Bonomi.
- N. 5/2020 [Storia dell'industria delle macchine utensili in Piemonte dalle origini alla seconda guerra mondiale](#). Secondo Rolfo.
- N. 4/2020 [Blockchain e Internet of Things per la logistica Un caso di collaborazione tra ricerca e impresa](#). Edoardo Lorenzetti, Lucio Morettini, Franco Mazzenga, Alessandro Vizzarri, Romeo Giuliano, Paolo Peruzzi, Cristiano Di Giovanni.
- N. 3/2020 [L'impatto economico e fiscale di un evento culturale: misure e scala territoriale](#). Giovanna Segre, Andrea Morelli.
- N. 2/2020 [Mapping the tangible and intangible elements of the historical buildings and spaces](#). Edoardo Lorenzetti, Nicola Maiellaro.
- N. 1/2020 [Il lavoro agile negli enti pubblici di ricerca](#). Emanuela Reale, Serena Fabrizio, Andrea Orazio Spinello.
- 2019
- N. 6/2019 [Women's candidatures in local elections: does the context matter? Empirical evidence from Italian municipalities](#). Igor Benati, Greta Falavigna, Lisa Sella.
- N. 5/2019 [Research activities in Nanotechnologies and Nanosciences: an analysis of Piedmont's nanotech research system](#). Ugo Finardi.

[Numeri precedenti/Previous issues](#)

How can artificial intelligence improve corporate rating forecasts without becoming a “black box”?

In corporate credit risk assessment, predictive accuracy is essential, but understanding the financial factors behind model decisions is equally important. This volume introduces HERA-XAI, an innovative framework that combines advanced machine learning techniques with state-of-the-art explainable artificial intelligence methods to make rating forecasts both accurate and transparent.

Drawing on a large dataset of Italian firms across the Production, Trade, and Services sectors, the book explores how financial indicators influence rating assignments, how their effects vary across firms and industries, and how changes in financial conditions can shape future rating trajectories. By integrating forecasting, explainability, and simulation within a unified hierarchical architecture, HERA-XAI provides new tools for interpreting machine learning predictions and assessing alternative financial scenarios.

Bridging the gap between predictive performance and interpretability, this work offers valuable insights for researchers, financial analysts, risk managers, banking professionals, and policymakers interested in the responsible application of artificial intelligence to credit risk evaluation and corporate finance.