INTERNATIONAL MONETARY FUND

Hungary's Corporate Sector Risk

A Machine Learning Approach

Jakree Koosakul and Xuege Zhang

SIP/2024/038

IMF Selected Issues Papers are prepared by IMF staff as background documentation for periodic consultations with member countries. It is based on the information available at the time it was completed on July 18, 2024. This paper is also published separately as IMF Country Report No 24/269.

2024
AUG

SELECTED ISSUES PAPER

 SIP/2024/038

IMF Selected Issues Paper
European Department

Hungary's Corporate Sector Risk: A Machine Learning Approach
Prepared by Jakree Koosakul and Xuege Zhang

Authorized for distribution by Anke Weber
August 2024

***IMF Selected Issues Papers* are prepared by IMF staff as background documentation for periodic consultations with member countries.** It is based on the information available at the time it was completed on July 18, 2024. This paper is also published separately as IMF Country Report No 24/269.

ABSTRACT: In recent years, Hungary's non-financial corporations were confronted with multiple shocks, ranging from the pandemic and rising geopolitical tensions to the historic tightening of domestic monetary policy. Employing machine learning techniques, this paper examines the determinants of Hungarian listed firms' credit risk evolution over this period. Our analysis shows that both firm-specific and macroeconomic factors played a role in explaining the observed rise in firms' default probability at onset of the pandemic, although Hungarian corporates proved broadly resilient, with risk indicators quickly improving a year after. Firms' credit risk rose again in 2022, however, as both long-term interest rates and sovereign risk premia sharply increased, despite continued improvements in firms' financial ratios. This development merits continued monitoring, particularly since a significant portion of corporate loans are set to mature within the next few years and could be repriced at higher interest rates.

RECOMMENDED CITATION: Jakree Koosakul and Xuege Zhang. 2024. "Hungary's Corporate Sector Risk: A Machine Learning Approach" IMF Selected Issues Paper (SIP/2024/038). Washington, D.C.: Inernational Monetary Fund.

JEL Classification Numbers:	E43, E44, E52, G32
Keywords:	Hungary, corporate sector, credit risk, sovereign risk, probability of default, machine learning, COVID-19 pandemic, geopolitical tension, monetary policy
Author's E-Mail Address:	jkoosakul@imf.org, xzhang2@imf.org

SELECTED ISSUES PAPERS

Hungary's Corporate Sector Risk: A Machine Learning Approach

Hungary

Prepared by Jakree Koosakul and Xuege Zhang[1]

[1] The authors are grateful to Anke Weber for providing overall guidance on this project, and to Kevin Wiseman for sharing the code previously used to examine corporate sector risks in the context of Mexico 2023 Article IV Staff Report. They are also thankful to the Hungarian authorities, including seminar participants at the Magyar Nemzeti Bank, for helpful comments and feedback, and to Estefania Cohn Bech for excellent research assistance and Ninfa Gonzales for technical support.

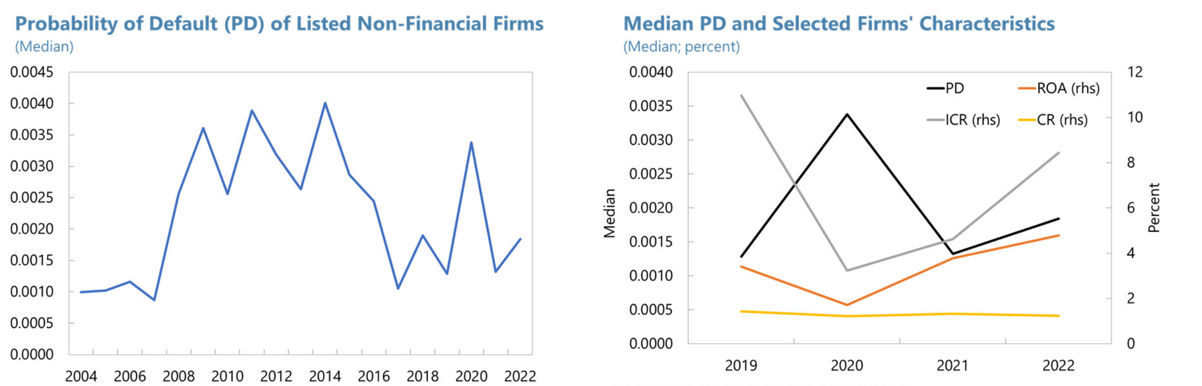
HUNGARY'S CORPORATE SECTOR RISK: A MACHINE LEARNING APROACH[1]

A. Introduction

1. **In recent years, Hungary's non-financial corporations were confronted with multiple shocks.** Global and domestic supply and demand disruptions caused by the pandemic resulted in a significant contraction of Hungary's export-oriented economy, ending a period of steady growth over 2016-2019. Russia' war in Ukraine and the resulting pressures on energy prices further reduced households' disposable income and subjected firms to significant cost pressures. The inflationary pressures stemming from these shocks necessitated a historic tightening of monetary policy. With market interest rates peaking at around 18 percent, firms' borrowing costs rose markedly, while the post-pandemic recovery in demand became subdued as monetary policy transmission worked its way through the economy.[2]

2. **Despite these shocks, the Hungarian corporate sector proved quite resilient.** The probability of default (PD) estimates for median Hungarian listed firms increased markedly at the onset of the pandemic, almost to the levels observed during the global financial and European crises (top left panel). This occurred as firms' financial ratios deteriorated (top right panel) and growth plummeted (bottom left panel). Such risks eased significantly in 2021, however, as the economy began to recover and firms' profitability and debt-servicing capacity improved. The median PD picked up again in 2022, although at a much more moderate pace, on the back of continued forint depreciation and rising interest rates and sovereign risk premia (bottom right panel).

Figure 1. Hungary's Corporate Sector Risk and Selected Firms' Characteristics and Macroeconomic Variables



Sources: NUS-CRI; and IMF staff calculations.

Sources: NUS-CRI; Orbis; and IMF staff calculations.
Note: ROA=return on assets, ICR=interest coverage ratio, CR=current ratio.

---

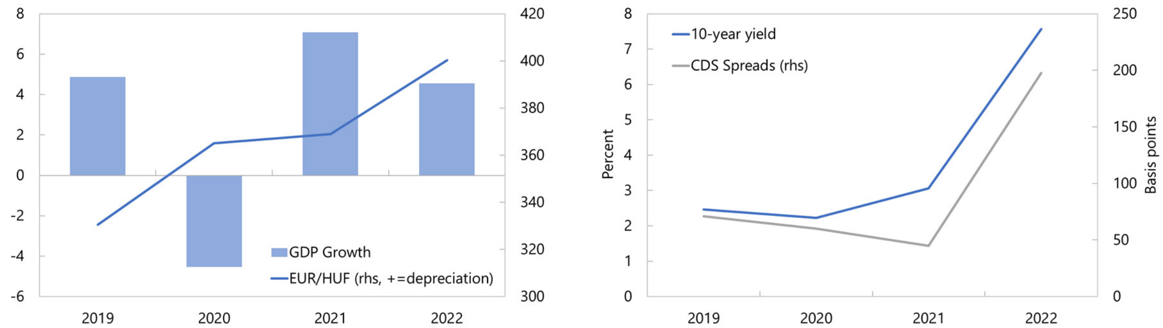
[1] Prepared by Jakree Koosakul (EUR) and Xuege Zhang (MCM). The authors are grateful to Kevin Wiseman for sharing the code previously used to examine corporate sector risks in the context of Mexico 2023 Article IV Staff Report.

[2] It is worth noting that despite the historic increase in the monetary policy rate, a range of subsidized programs were introduced to help shield companies since the breakout of the pandemic. In addition, a significant portion of market-based loans appear to depend on long-term yields rather than the policy rate. These factors would imply that the effective loan rates for companies did not increase as much as the policy rate.

Figure 1. Hungary's Corporate Sector Risk and Selected Firms' Characteristics and Macroeconomic Variables (Concluded)

Real GDP Growth and EUR/HUF Exchange Rate
(Percent)

10-Year Government Bond Yields and CDS Spreads
(Percent; basis points)



Sources: National authorities; Haver Analytics; and IMF staff calculations.

Sources: Haver Analytics; and IMF staff calculations.

3. **Against this backdrop, this annex utilizes machine learning techniques to better understand the evolution of risks in Hungary's corporate sector.** The objectives are twofold: i) explore the general determinants of firms' PDs over the past two decades, based on both firm-specific and macroeconomic factors, and ii) examine more specifically the influence of these determinants on firms' PDs during and after the pandemic. As is standard in the literature, our focus is on non-financial corporations (NFCs). Due to data availability, the analysis is based on NFCs listed in Hungary's stock market (and extended to include other CESEE countries as a robustness check).

B. Data and Methodology

4. **The analyses are based on data from several sources over the span of two decades.** The estimates on firms' PDs are obtained from the Credit Research Initiative, National University of Singapore (NUS-CRI).[3] Such PDs are transformed through a logistic (logit) function.[4] Firms' financial ratios are calculated based on balance sheet data from the Orbis database.[5] To account for the impact of macroeconomic factors, data on Hungary's GDP growth, 10-year government bond yields, CDS spreads, and the EUR/HUF exchange rate are obtained from Haver.[6] The sample period is 2000-2022. The matched sample of firms contains 527 firm-year observations covering 55 unique firms, with 24 firms present on average each year. Inspection into the characteristics of these firms indicates that a range of industries are well represented, although given the listed nature of these firms, the sample focuses on large corporations in the Hungarian economy. Due to the rather small sample, a robustness exercise conducted in the final section of this annex relies on an expanded sample of listed firms not only in Hungary, but also in seven other CESEE countries, namely Bulgaria,

---

[3] For more details on the measure, which is based on the forward intensity model of Duan et al. (2012), see CRI-NUS (2022)

[4] The choice of the logit transformation ensures that the predicted PD falls within the [0,1] interval.

[5] To clean firms' balance sheet data, we apply the procedure proposed by Kalemli-Ozcan et al. (2015).

[6] The choice of macroeconomic variables roughly follows a similar study done on Mexican corporates in the Mexico 2023 Article IV Staff Report.

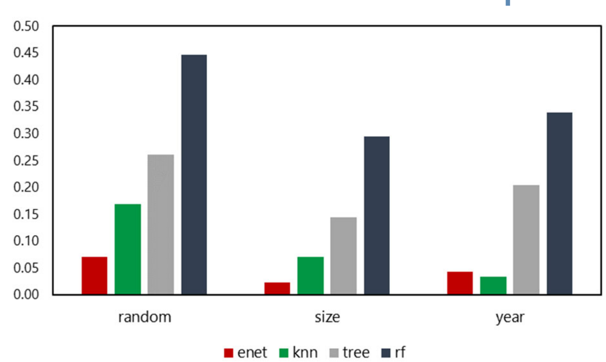
Croatia, Czech Republic, Poland, Romania, Slovakia, Slovenia. This brings the number of firms and firm-year observations to 1,182 and 10,086, respectively.

5. Machine learning techniques are utilized. Relative to more traditional regression techniques, machine learning methods offer several advantages, including their ability to handle non-linear relationships, non-standard distributions, data gaps, and outliers that are prevalent in firms' financial ratios, and interaction effects between explanatory variables. Without a strong prior on which specific machine learning models may be appropriate, a variety of candidate models with different levels of flexibility and complexity are considered. These include elastic net, decision tree, K-nearest neighbors, and random forest, which are the more commonly used algorithms to model continuous dependent variables in economics.[7]

6. Candidate models are evaluated based on out-of-sample performance. To avoid overfitting, the selection of the final machine learning method for modelling Hungarian firms' PDs is based on evaluating their out-of-sample predictive performance through cross-validation. The estimation sample is divided into subgroups or "folds", with one fold removed to be used as a validation set and the remaining folds used in the estimation of the model. The model's performance is then evaluated based on its predictive power over the validation fold. This is repeated for each fold and the results across all folds are averaged to arrive at an overall metric of performance. This cross-validation exercise is conducted by dividing folds based on random allocation, firm size, and time.

7. The random forest model has the best predictive performance. Through the horse race, we find the random forest model to produce the highest value of the average coefficient of determination (akin to R^2 in the OLS regression context) among the four candidate models (Figure 2). This is true under all combinations of cross-validation sampling method.

Figure 2. R^2 for Candidate Models Under Different Cross-Validation Groups



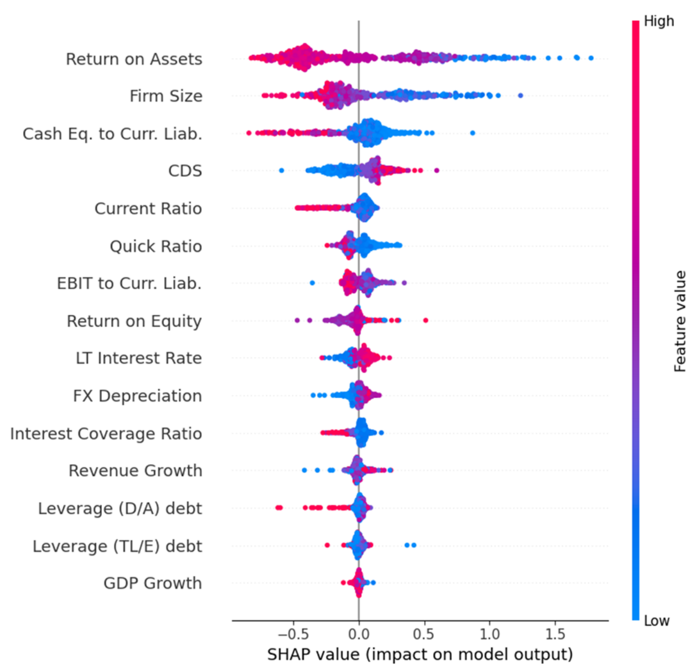
Source: IMF staff calculations.
Note: enet = elastic net, knn = K-nearest neighbors, tree = decision tree, and rf = random forest.

[7] Athey and Imbens (2019) provide a useful reference on the description and use of these models in economics.

C. Determinants of Corporate Sector Risk

8. Both firm-specific and macroeconomic variables are found to be important determinants of firms' PDs. As with other non-linear machine learning models, results from the random forest model can be summarized via Shapley values, which represent additive contributions of each factor to the predicted values of the dependent variable. A "beeswarm" plot provides a useful visual representation of each factor's contributions based on such values, where the variables are ranked in order of their average absolute contribution to risk assessment and where the color of each dot represents high (in red) or low (in blue) values for the variable in question. From Figure 3, return on assets (ROA), firm size, and cash equivalents to current liabilities (CE/CL) are the top three most important variables in the model, highlighting the importance of firm profitability, size, and liquidity buffers. The direction of the effects is as expected, with high risks associated with low values of these variables. On the macro side, sovereign risk premia, captured by the CDS, appears as the fourth most important factor, with high risks correctly associated with high values. Long-term interest rates and the value of the currency are found to play a relatively minor role, although their variations appear to explain some of the variations in the predicted PDs. By contrast, the role of GDP growth in predicting corporate risk seems very limited, potentially reflecting the positive correlation between growth and firms' profitability ratios that are also included in the model.

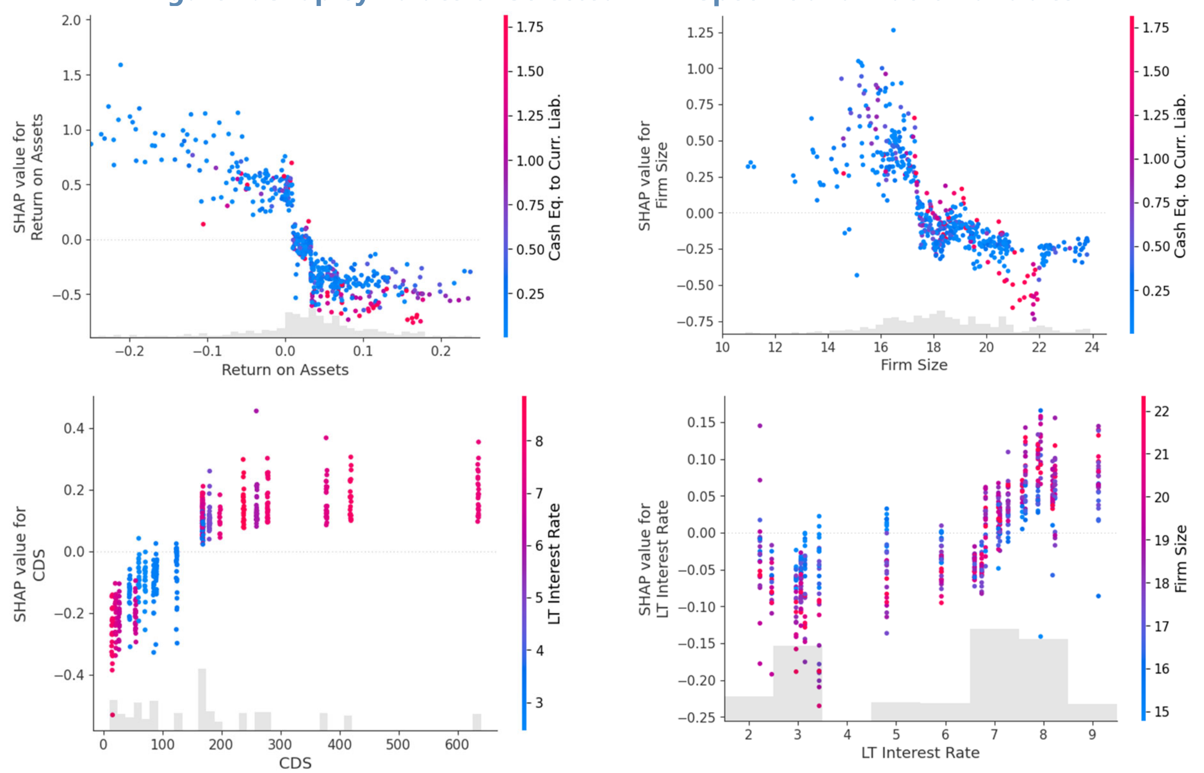
Figure 3. Contributions to Risk Assessment



Source: IMF staff calculations.

9. Plots of Shapley values for individual variables also highlight some non-linearity and interaction effects. Figure 4 displays plots of Shapley values for the top two firm-specific and macroeconomic variables. There appears to be significant non-linearity in the effects on risk for all four variables. More specifically, the contributions to predicted PDs from ROA and firm size fall sharply when returns become positive or firms are sufficiently large. Similarly, the contributions from sovereign risk premia and long-term interest rates increase substantially when CDS spreads rise above around 170 and when bond yields are higher than 7 percent. In addition, some interaction effects are evident. In particular, the reduction in the contributions of ROA and firm size to predicted PDs is more pronounced for firms with higher CE/CL (red dots). Put differently, the beneficial effects of higher profitability and larger size appear to be greater for firms with higher liquidity buffers. Similarly, larger firms (red dots) also appear to benefit more in terms of the risk reduction from a decline in long-term interest rates than smaller firms.

Figure 4. Shapley Values of Selected Firm-Specific and Macro Variables



Source: IMF staff calculations.

D. Risk Evolution Through the Pandemic and High Interest Rates

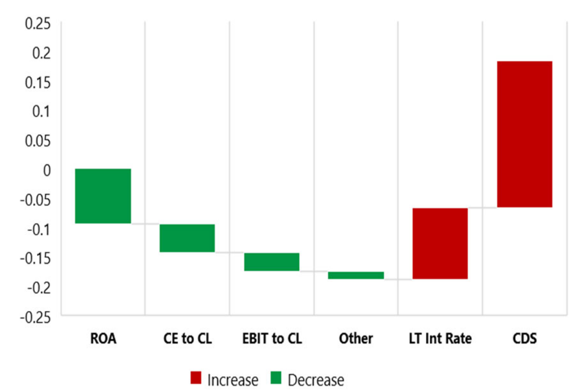
10. Risk indicators deteriorated across the board at the onset of the pandemic before improving one year after the pandemic. Since Shapley values are additive, they can be averaged across firms for each variable for a given year. The difference in the average Shapley values across two years for a given variable hence allows us to capture the contribution of such a variable to the change in the predicted PD between the two years. The waterfall charts below provide information on the main contributors to the change in the predicted PDs during and after the pandemic. At the onset of the pandemic in 2020, the left-hand panel of Figure 5 shows a deterioration of key risk factors across the board. In particular, deteriorating profitability contributed the most to the observed rise in Hungarian corporate sector risk in 2020. Other firm-specific and macro factors, namely firms' liquidity buffers and size, GDP growth, and sovereign risk premia also played a role, although to a limited extent. The right-hand panel then shows reversals in the contributions of risk factors in 2021, with an improvement in firms' profitability playing the most significant role in reducing corporate sector risk, followed by a decline in CDS spreads and improvements in revenue growth and GDP growth.

Figure 5. Contributions to Change in Average Predicted PD During the Pandemic

Source: IMF staff calculations.
Note: in the left-hand panel, the average value of the logit PDs across all firms in 2021 is normalized to 0, such that the sum of the individual bars captures the change in the average logit PD across firms between 2021 and 2022. The same treatment applies to the average value of the logit PDs across all firms in 2022, in the right-hand panel.

11. **While firm-specific risk factors continued to improve in 2022, their beneficial effects were more than offset by higher interest rates and rising sovereign risk premia.** While successful at taming inflationary pressures, our analysis highlights the effects of the significant monetary tightening on corporate sector risk in 2022, especially since this occurred at a time of rising sovereign risk premia. Going forward, it would be important to continue monitoring interest rate and maturity risks carefully, since a significant portion of corporate loans are set to mature within the next few years and could be repriced at higher interest rates.

Figure 6. Contributions to Change in Predicted Logit PD Between 2021-2022



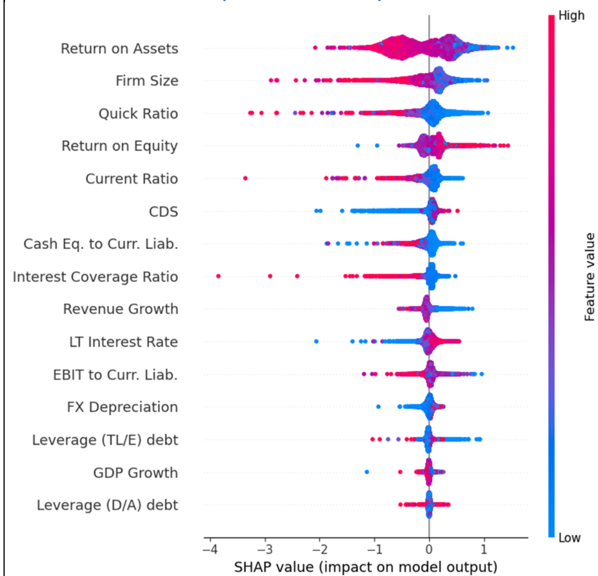
Source: IMF staff calculations.
Note: the average value of the logit PDs across all firms in 2021 is normalized to 0, such that the sum of the individual bars captures the change in the average logit PD across firms between 2021 and 2022.

E. Robustness Check

12. Our baseline results are robust to including an expanded sample of countries. Two potential issues with respect to the baseline sample are worth highlighting. First, the sample contains a relatively limited number of firms. This potentially limits the variations of both the dependent and explanatory firm-specific variables that would otherwise result in a more precise identification of the effects of different factors. Second, the focus on only one country also implies relatively limited variations in the values of macroeconomic variables, with each variable only having just over 20 likely non-independent observations. To address these issues, the baseline exercises are replicated based on an expanded sample of countries. As shown in the charts below, the baseline results are robust with respect to both the general determinants of credit risk as well as the assessment on the evolution of Hungary's corporate sector risk during and after the pandemic.

Figure 7. Contributions to Risk Assessment

(Expanded Sample)



Source: IMF staff calculations.

Figure 8. Shapley Values of Selected Firm-Specific and Macro Variables

(Expanded Sample)

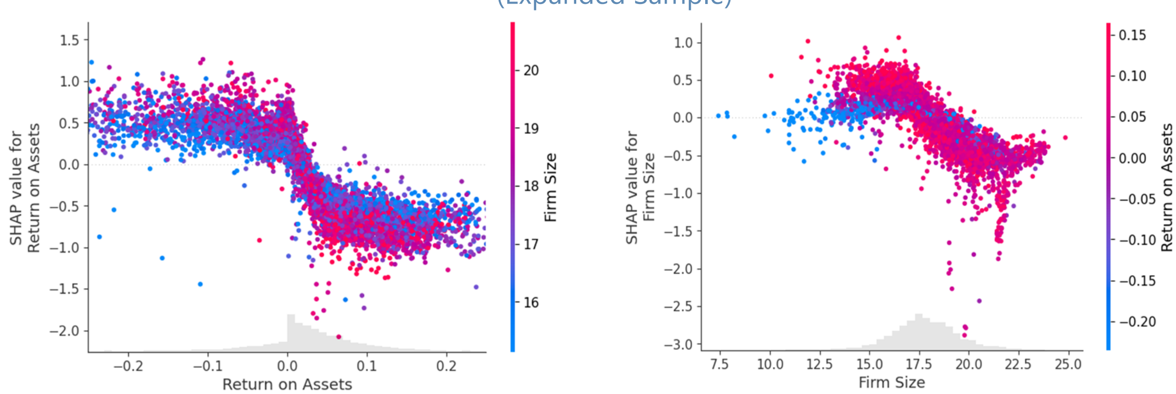
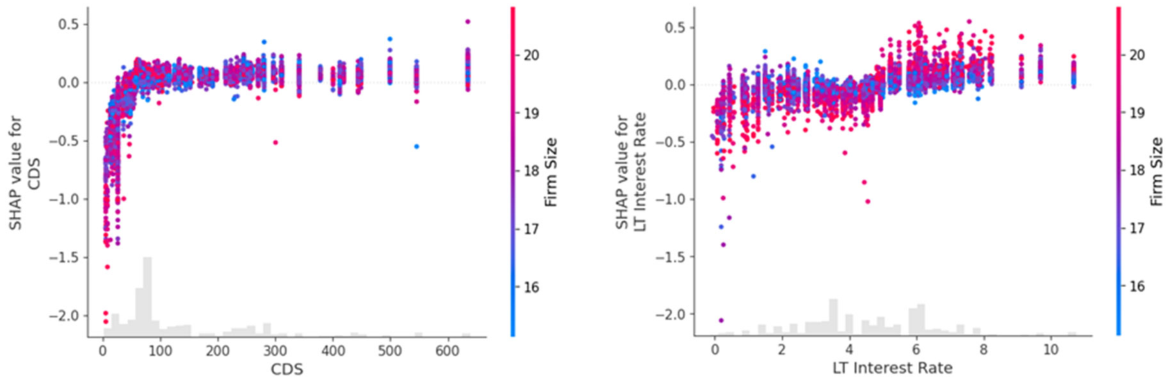


Figure 8. Shapley Values of Selected Firm-Specific and Macro Variables (Concluded)

(Expanded Sample)



Source: IMF staff calculations.

Figure 9. Top Three Contributors to Hungary's Corporate Sector Risk Evolution

(Expanded Sample)

Source: IMF staff calculations.

Note: While the model is estimated based on the expanded country sample, the results in this figure are computed specifically based on the Shapley values for Hungarian firms in the sample, and are hence Hungary-specific.

References

Athey, Susan, and Guido W. Imbens. 2019. "Machine Learning Methods That Economists Should Know About", *Annual Review of Economics,* 11: 685-725.

Duan, Jin-Chuan, Jie Sun, and Tao Wang. 2012. “Multiperiod Corporate Default Prediction – A Forward Intensity Approach”, *Journal of Econometrics*, 179:191-209.

International Monetary Fund. 2023. "Mexico: 2023 Article IV Consultation-Press Release and Staff Report." Country Report No. 2023/356. November 1.

Kalemli-Ozcan, Sebnem, Bent Sørensen, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcan Yesiltas. 2015. "How to Construct Nationally Representative Firm Level Data from the Orbis Global Database: New Facts and Aggregate Implications." Working Paper 21558, National Bureau of Economic Research. Revised January 2022.

The Credit Research Initiative of the National University of Singapore. 2022. “Probability of Default: White Paper.” Available at https://nuscri.org/en/white_paper/