Enhancing Predictive Framework for Fiscal Crises Using Machine Learning

Akhil Baghla

August 2023

1 Introduction

Financial crises are a recurring phenomenon in economic history, imposing substantial economic and societal burdens. Although complete prevention might be challenging for policymakers, the ability to identify early indicators could empower macroprudential authorities to enact measures that reduce the likelihood and impact of these crises. Predicting fiscal crises presents a significant challenge due to the limited observations of these disruptive events and their enduring economic impact (Medas et al., 2018). In response, economists have focused on developing early warning systems to detect and prevent such crises—an endeavor ongoing since the 1970s (Moreno Badia et al., 2020). These systems hold immense policy implications, emphasizing the critical importance of predictive accuracy. Research predominantly delves into econometric and machine learning methods, evaluating their effectiveness in forecasting fiscal crises.

However, forecasting fiscal crises is uniquely complex. Even established early warning systems struggled to anticipate major crises like the Asian currency crises of the 1990s and the Global Financial Crisis (Berg and Pattillo, 1999; Christofides et al., 2016). Exploring machine learning methods offers a promising avenue for capturing the intricate patterns underpinning fiscal crises. By adapting well-established machine learning techniques and optimizing the balance between overfitting and underfitting, this research aims to enhance predictive capabilities and deepen the understanding of the economic vulnerabilities driving fiscal crises.

2 Objective

The primary objective is to predict financial crises with a lead time of 2-3 years. To achieve this, leverage will be drawn from extensive databases containing macroeconomic and financial indicators spanning over 50 years for various economies. The central research goal is to develop an advanced early warning system for predicting financial crises through machine learning. By comparing the predictive performance of machine learning models with traditional econometric methods, this study aims to illuminate potential enhancements achievable through this approach.

The research strategy entails a rigorous comparison between a conventional logistic regression model and a diverse array of advanced machine learning models. The focus will be on methods like decision trees, random forests, extremely randomized trees, support vector machines (SVM), and gradient boosting. Through out-of-sample evaluations, the predictive efficacy of each model variant will be systematically assessed. Performance and accuracy metrics will guide the selection of the most suitable models under specific conditions. Furthermore, the contributions of individual variables will be ranked, offering insights into their significance. This meticulous analysis aims to unveil the substantial potential of machine learning techniques in improving the ability to anticipate financial crises within a specified temporal horizon.

3 Methodology

To fulfill this objective, this study will build upon the Macrohistory Database curated by Jordà et al. (2017), a comprehensive source of macroeconomic and financial data spanning 140 years across 17 advanced economies (tentative). A comparative analysis will be conducted, pitting traditional models against a diverse range of machine learning models. Rigorous out-of-sample evaluations will gauge the performance of these models.Different datasets will be combined by leveraging diverse sources yet to be explored.

4 Contributions

The research aims to make the following contributions:

Model Performance: Systematic comparisons between machine learning models and conventional logistic regression will quantify the predictive power of each model variant. Identifying superior models for forecasting financial crises will provide insights into the potential benefits of advanced techniques.

Interpretability: Addressing the challenge posed by complex machine learning models, this study aims to uncover the most critical variables, enhancing the interpretability of these models.

Insights and Implications: The research will offer insights into the pivotal factors driving financial crisis predictions. This understanding will empower policymakers to make informed decisions regarding macroprudential policies, especially during periods of heightened vulnerability.

5 Expected Outcomes

Anticipated outcomes include the identification of machine learning models with superior predictive capabilities, insights into critical economic variables influencing crisis predictions, and an enriched understanding of their nonlinear relationships and interactions. These findings will contribute to the development of more effective early warning systems and the refinement of macroprudential policy strategies.

6 Literature Review

In the world of predicting financial crises, researchers have faced a common issue: there aren't many instances of crises to learn from. To tackle this, they've been exploring new ways to improve predictions. One interesting approach is using machine learning models, which seem to do better than the usual logistic regression in forecasting financial crises over longer periods.

Machine learning models have a special knack for finding complex, non-linear patterns and connections among variables that traditional methods might miss. This is really useful for understanding the hidden factors behind financial crises, which are rare events with intricate and hard-to-predict behaviors before they happen (Alessi and Detken, 2018). This capacity positions machine learning as a potent tool in constructing robust early warning systems. Machine learning's appeal lies in its ability to tackle persistent challenges in macroeconomic early warning systems. The complex dynamics preceding fiscal crises often elude simple linear or threshold models, which struggle to capture intricate complexities. Furthermore, the challenge of small sample sizes in macroeconomic panels poses the risk of identifying spurious patterns that lack relevance in predicting future crises. This risk of overfitting—fitting data too closely to the estimation sample—plagues traditional methods, underscoring the need for approaches that optimally balance complexity and predictive power. Machine learning algorithms are equipped to address this trade-off, making them a preferred choice for various prediction tasks (Kleinberg et al., 2015).

A significant step forward in this area of research involves understanding how black-box machine learning models predict financial crises. Researchers have achieved this by breaking down the models' predictions into the contributions of individual variables using the innovative Shapley value framework (Strumbelj and Kononenko, 2010; Joseph, 2020). This method not only helps grasping the important economic factors driving these models but also gives a way to test their predictions. It's especially crucial in making these forecasts more credible. What's more, this technique addresses a common issue faced by policymakers when using machine learning models—their lack of transparency. By offering explanations that support policy decisions based on these models, this framework connects the dots between machine learning results and policy choices. This fosters transparency and accountability.

To test the effectiveness of machine learning models in predicting financial crises, researchers have explored an extensive dataset—the Macrohistory Database by Jordà et al. (2017). This dataset encapsulates macroeconomic and financial variables from a span of more than 140 years across 17 advanced economies. Through a rigorous comparison, various machine learning models, including decision trees, random forests, extremely randomized trees, support vector machines (SVM), and artificial neural networks, have been evaluated against the traditional logistic regression model. The findings underscore a clear pattern—machine learning models, except individual decision trees, consistently showcase robust predictive capabilities, surpassing the performance of the logistic regression.

Digging deeper into the drivers of these predictive models, the research reveals that credit growth and the shape of the yield curve emerge as central predictors across a diverse set of modeling approaches. While the significance of domestic credit growth in predicting financial crises is well-documented (Borio and Lowe, 2002; Drehmann et al., 2011; Schularick and Taylor, 2012; Aikman et al., 2013; Jordà et al., 2013, 2015b; Giese et al., 2014), the role of the yield curve has received less attention, often limited to predicting recessions rather than crises. The study's findings emphasize the crucial linkage between a flatter or more inverted domestic yield curve and an elevated likelihood of a crisis, even when controlling for recessions. This phenomenon may be indicative of compressed net interest margins or, intriguingly, might reflect the amplification of the search for yield and increased risk-taking tendencies preceding financial crises. These findings also hold relevance on a global scale, where interactions between credit growth and the yield curve slope play a pivotal role in forecasting crises, contingent upon the chosen time period for global credit growth and global yield curve slope vis-à-vis recessions. In addition, the research explores the predictive prowess of other variables, such as stock prices, money, the current account, and house prices. While the predictive capacity of these variables varies, house prices appear to marginally enhance model performance post-1945, albeit without robustness, indicating its situational relevance across specific countries and periods.

In summary, the literature on predicting financial crises has witnessed an evolution towards leveraging machine learning models for enhanced predictive accuracy. These models, characterized by their flexibility and capacity to unearth nonlinear relationships, present a powerful approach in forecasting rare and intricate financial crises. The utilization of the Shapley value framework further enhances the interpretability of these models, fostering a connection between machine learning outputs and policy decisions. The evaluation of machine learning models against traditional logistic regression through a comprehensive dataset underscores their superior predictive capabilities. Moreover, the study's findings regarding the role of credit growth, the yield curve, and other key variables contribute to a nuanced understanding of the drivers behind financial crises, shedding light on previously unexplored dynamics and interactions.