PREDICTING EXTERNAL MACROECONOMIC CRISES: MACHINE
LEARNING PERSPECTIVE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SONGUL SIVA CELIK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF ECONOMICS

MARCH 2024

Approval of the thesis:

**PREDICTING EXTERNAL MACROECONOMIC CRISES: MACHINE
LEARNING PERSPECTIVE**

submitted by **SONGUL SİVA CELİK** in partial fulfillment of the requirements
for the degree of **Doctor of Philosophy in Economics, the Graduate School of
Social Sciences of Middle East Technical University** by,

Prof. Dr. Sadettin KİRAZCI
Dean
Graduate School of Social Sciences   _____

Prof. Dr. Onur YILDIRIM
Head of Department
Department of Economics   _____

Assist. Prof. Dr. Kağan   PARMAKSIZ
Supervisor
Department of Economics   _____

Examining Committee Members:

Prof. Dr. Erol TAYMAZ (Head of the Examining Committee)
Middle East Technical University
Department of Economics   _____

Assist. Prof. Dr. Kağan PARMAKSIZ (Supervisor)
Middle East Technical University
Department of Economics   _____

Assoc. Prof. Dr. İlhan Can ÖZEN
Middle East Technical University
Department of Economics   _____

Assoc. Prof. Dr. Seçil Aysed BAHÇE
Ankara University
Department of Economics   _____

Assist. Prof. Dr. Umut ÖNEŞ
Ankara University
Department of Economics   _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: SONGUL SIVA CELIK

Signature:

ABSTRACT

PREDICTING EXTERNAL MACROECONOMIC CRISES: MACHINE
LEARNING PERSPECTIVE

CELIK, SONGUL SIVA
Ph.D., The Department of Economics
Supervisor: Dr. Kağan Parmaksız

March 2024, 221 pages

In this study, our primary objective is to predict Sudden Stops using Machine Learning (ML) methods and evaluate their out-of-sample prediction power. Conducted in two phases, the first involves establishing a baseline with Forbes and Warnock (2021) as our model, replicating and assessing its out-of-sample predictions. We then introduce various ML methods, including Elastic Net, Random Forests, Support Vector Machines, kNN, AdaBoost, XGBoost, and Multi-Layer Perceptron, for a comprehensive comparison against the baseline, utilizing metrics like accuracy, precision, recall, F1-score, and ROC curve.

In the second phase, we expand the dataset from the IMF, prioritizing data availability, and employ ML methods for feature selection. Selected features are used for ML estimation, involving traditional methods like Elastic Net, Random Forests, Support Vector Machines, XGBoost, Logistic Regression, and the modern deep learning method, a Long Short-Term Memory (LSTM). The analysis aims to explore potential improvements in out-of-sample performances facilitated by ML algorithms and feature selection.

The results indicate that XGBoost and Random Forests exhibit high prediction performance in both data sets. However, considering the temporal dependencies, the

Long Short-Term Memory (LSTM) method outperforms traditional tree-based machine learning models over time.

This study contributes to the emerging literature on the effectiveness of ML methods in macroeconomic forecasting through an interdisciplinary approach, examining whether these algorithms enhance out-of-sample predictions for Sudden Stops.

Keywords: Machine Learning, Sudden Stop Crises, Deep Learning, Out-of -Sample Prediction.

ÖZ

MAKİNE ÖĞRENİMİ İLE DIŞ MAKROEKONOMİK KRİZLERİN TAHMİNİ

CELIK, SONGUL SIVA
Doktora, İktisat Bölümü
Tez Yöneticisi: Dr. Öğr. Üyesi Kağan Parmaksız

Mart 2024, 221 sayfa

Bu çalışmanın temel amacı, Makine Öğrenmesi (MÖ) yöntemlerini kullanarak Ani Duruşları öngörmek ve bu yöntemlerin örnek dışı tahmin yeteneklerini değerlendirmektir. İki aşamada yürütülen çalışmanın ilk aşamasında, Ani Duruş tahmininde kullanılan geleneksel modellerinden birini kullanan ve kısıtlı dışsal değişken içeren Forbes ve Warnock 'un (2021) modelini temel aldık. Bu modelin örnek dışı tahmin gücünü ölçtük. Ardından aynı veri setini, Esnek Ağ (Elastic Net), Rassal Ormanlar (Random Forests), Destek Vektör Makineleri (Support Vector Machines), XGBoost, k-En Yakın Komşular (kNN), AdaBoost ve Çok Katmanlı Algılayıcı (MLP) gibi çeşitli MÖ yöntemlerini uygulayarak hem makina öğrenimi yöntemlerinin hem de temel modelin örnek dışı performans güçlerini modelle doğruluk, hassasiyet, hatırlama, F1 puanı ve ROC eğrisi gibi ölçütlerle kapsamlı bir şekilde karşılaştırdık.

İkinci aşamada, veri kümesini sadece verinin ulaşılabilirliği kriteriyle IMF'nin çeşitli veri setlerini kullanarak genişlettik. Makine öğrenimi yöntemlerini kullanarak, Ani Duruş Krizlerinin tahmininde daha etkili olan dışsal değişkenleri belirlemek için bu genişletilmiş veri setini daralttık. Seçilen dışsal değişkenlerle, Esnek Ağ, Rassal Ormanlar, Destek Vektör Makineleri, XGBoost, Lojistik Regresyon ve modern derin

öğrenme yöntemi olan Uzun Kısa Vadeli Hafıza (LSTM) MÖ yöntemleriyle tahminler gerçekleştirdik.

Sonuçlar, XGBoost ve Rassal Ormanların her iki veri setinde de yüksek tahmin performansına sahip olduğunu, ancak Uzun Kısa Vadeli Hafıza (LSTM) yönteminin zaman içindeki bağlantıları göz önüne alarak geleneksel ağaç tabanlı MÖ modellerinden daha iyi performans sergilediğini ortaya koymaktadır.

Bu şekilde, disiplinler arası bir yaklaşım ile, makine öğrenimi yöntemlerinin makroekonomik tahminlemedeki etkinliği üzerine ile ortaya çıkan literatüre katkıda bulunmayı amaçladık.

Anahtar Kelimeler: Makine Öğrenmesi, Ani Duruş, Derin Öğrenme, Örneklem Dışı Tahmin

To my husband, Serdar

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Sudden Stops (SS) are economic fluctuations defined by a set of empirical regularities associated with a large, sudden reversal of capital inflows (i.e., a sudden "loss of access" to international financial markets). The defining characteristic of a Sudden Stop is a sharp reversal in external capital inflows, which is often measured by a sudden jump in the current account. At about the same time as the access to foreign financing is lost, or shortly after, the economies affected by Sudden Stops experience deep recessions, sharp depreciations in the real exchange rate (RER), declines in asset prices. Moreover, they are often preceded by expansion periods; high credit growths, large current account deficits, appreciated RER and asset price booms.

Mendoza and Korinek (2013) states that the Sudden Stops observed in emerging markets during the 1990s were a precursor to the Global Financial Crisis 2008-2009. Sudden Stop (SS) crises were a primary concern for Emerging Markets (EM), as they had significant impacts during the 1990s, particularly affecting EMs. However, the scope of this crisis phenomenon expanded beyond EMs, influencing both advanced and developed countries. This perspective is reinforced by Bianchi and Mendoza (2020), which documented a total of 58 recorded SS events by the close of 2016. Remarkably, 35 of these events were witnessed in emerging markets, while 23 of them SS events unfolded in advanced economies. This underscores the widespread occurrence of SS events and underscores the need to address them effectively across diverse economic landscapes. Comparing the period before 2000 with the period between 2001 and 2014, in a sample of both advanced economies (AE) and Emerging Markets (EM), Eichengreen and Gupta (2016) shows that the frequency and duration of sudden stops have remained largely unchanged since 2002. They conclude that sudden stops remain as a significant economic problem.

Although studies on Sudden Stops began with Calvo (1998) and continued thereafter, they have regained attention and increased after the Global Financial Crisis (GFC)[1]. The GFC has shifted focus to credit boom-bust cycles, capital flow volatility, and Sudden Stops, prompting discussions on macroprudential policies. During the peak of the GFC, specifically spanning from the fourth quarter of 2008 (2008Q4) to the first quarter of 2009 (2009Q1), Forbes and Warnock (2012) discerned a significant occurrence of Sudden Stop (SS) episodes in capital flows across a spectrum of nations. Notably, this analysis highlighted a total of 22 countries that encountered such SS phenomena during this turbulent period. Among these countries were Argentina, Brazil, Estonia, Iceland, India, Latvia, Norway, Peru, Romania, Russia, Greece, and Turkey. This identification sheds light on the far-reaching impacts of the GFC, as economies across diverse geographical regions faced abrupt reversals in capital flows, underscoring the pervasive nature of the crisis and its implications for global financial stability.

In relation with credit booms and Sudden Stops, Mendoza and Terrones (2012) associate credit booms with economic expansion, real estate growth, and external deficits, often followed by downturns. These booms exhibit global synchrony, clustering around major events like the 2008 GFC. Mendoza and Terrones (2012) note that while not all booms lead to crises, when they do, banking, currency crises, and Sudden Stops may follow. Reinhart and Reinhart (2008) concur, stressing the negative consequences of capital inflow bonanzas, particularly for emerging markets. Jorda et al. (2012) further highlight that recessions followed by financial crisis are more severe, with credit-intensive expansions preceding deeper recessions. This emphasizes the amplified risks associated with credit booms and Sudden Stop Crisis, particularly in emerging markets.

After the Global Financial Crisis, unconventional monetary policies and near-zero interest rates in the United States facilitated the flow of international capital into emerging economies. However, episodes like the "taper tantrum" in 2013, sparked by concerns that the Federal Reserve might reduce its purchases of securities, and the

---

[1] The expression "Sudden Stop" was initially mentioned in this context by Dornbusch, Goldfajn, and Valdés in their 1995 paper, drawing inspiration from an old adage in banking.

"normalization" episode in 2015, fueled by expectations of rising U.S. interest rates, highlighted the possibility that Sudden Stops may be becoming more frequent or potentially more disruptive (Eichengreen and Gupta, 2016). Forbes and Warnock (2021) indicate that while the massive waves of capital flows observed during the 2008-2009 crisis have transformed into more controllable ripples, a notable surge occurred in 2015. This was driven by investor anticipation of the U.S. Federal Reserve's interest rate hike after a prolonged period. In this context, around 27% of countries in their sample encountered sudden stops in capital flows, while 22% confronted retrenchment. Furthermore, Forbes and Warnock (2021) demonstrate that while international capital flow volatility and extreme capital flow events have decreased globally since the Global Financial Crisis (GFC), they remain relatively high and comparable to levels before 2008. This underscores the continued significance of Sudden Stops, particularly for Emerging Markets (EMs).

Moreover, policymakers and academia have intensified their focus on capital regulations and macroprudential studies to enhance macro-financial stability.

In 2012, the IMF altered its perspective and highlighted the potential advantages of capital controls, calling for further research on a comprehensive, flexible, and balanced approach to managing capital flows. More recently, the IMF (2022) emphasized the importance of countries having the option to proactively restrict debt inflows to safeguard macroeconomic and financial stability. This recommendation is particularly relevant in cases where external liabilities pose risks, especially when they generate currency mismatches due to external debt denominated in foreign currency without appropriate foreign currency assets or hedges.

This renewed emphasis underscores the crucial need to predict and preempt the implications of Sudden Stops, especially in the context of Emerging Markets (EMs) and their vulnerability to severe capital flow disruptions. In empirical studies aiming to predict Sudden Stop Crises, traditional econometric methods such as logistic regression, probit, or complementary log-log methods have been commonly employed. However, these methods have limitations that can affect their accuracy and generalizability. A significant constraint is associated with the linearity

assumption in these predictive models, which may hinder the discovery of complex relationships between variables, including non-linearities and interactions inherent in a crisis event. ***As an alternative to these methods, and by embracing a data-intensive approach, our objective is to explore the potential of machine learning methods in predicting Sudden Stop events. We aim to investigate whether these methods can effectively address these limitations and provide higher out-of-sample performance, leading to more generalizable models***. While machine learning presents various valuable tools for different economic purposes, such as leveraging big data, constructing unconventional datasets, or utilizing clustering methods, our primary focus is on the superior performance of machine learning in out-of-sample prediction within the field of macroeconomic forecasting.

Machine learning offers several key advantages over traditional econometric models. First and foremost, it offers superior performance in out-of-sample prediction because of its ability to generalize, i.e., to learn robust models from data. Second, it can make use of big data with many variables abstracting out the variable selection process. Third, ML performance gets better with more samples of data, leading to the improved models with the addition of new data.

Machine learning methods are often recognized for their high accuracy and out-of-sample prediction power, which means more generalizable models. Most ML methods do not impose strong assumptions on the distribution of the data and the linearity of the relationship between variables and outcomes. ***This flexibility helps reveal the complexities inherent in macroeconomic crises***. Moreover, the methodological properties of ML methods can enhance out-of-sample prediction, as they can be trained to achieve higher prediction accuracy rather than focusing solely on parameter estimation and in-sample prediction. In-sample prediction performance, which measures the ability to fit the model to the training data, is often high for traditional econometric models due to their simplicity and structure. However, out-of-sample prediction performance, which measures the ability to generalize to new, unseen data, is typically higher for ML models due to their ability to capture more complex relationships in the data.

While the primary strength of ML lies in its flexibility and ability to learn from data, capturing complexities in the relationships between variables, it is essential to acknowledge limitations that apply to both methods—traditional econometrics and ML methods [2]. First, the property of a small sample size in macroeconomic datasets, typically a few thousand observations at most, can impact predictive ability. The inherent small sample sizes may lead to issues such as *overfitting* (i.e., discovering patterns that are specific to the sample used for estimation and do not extend to other samples) or, in traditional statistical terms, spurious regression. Additionally, the infrequency of crises contributes to imbalances and may result in biased estimation, collectively leading to a decrease in prediction accuracy. While these challenges are common to any predictive model, whether traditional econometric models or ML models used for predicting Sudden Stop (SS) events, they are not the sole focal points of concern.

Moreover, ML models, due to their complexity, generally require larger datasets to effectively learn intricate relationships within the data. In contrast, simpler models, particularly traditional linear models, may perform better when the dataset is small. It's important to note that ML methods are not entirely immune to these challenges; however, ML models are equipped with tools such as regularization, cross-validation, and hyperparameter tuning to mitigate these issues. By leveraging these tools, ML methods aim to minimize problems arising from small sample sizes and rarity of events. Therefore, our interest lies in investigating these ML methods, not only in the context of small sample sizes and imbalanced datasets but more prominently, for their specific capabilities in revealing complexities in Sudden Stop crises.

In this study, our primary objective is to predict Sudden Stops using ML methods and evaluate their out-of-sample prediction performances. To do so, we conduct our analysis in two parts. In the first phase, to establish a baseline

---

[2] Goulet Coulombe et al. (2020) demonstrate that the capacity to handle non-linearities constitutes the primary factor contributing to the improved performance of machine learning methods in macroeconomic forecasting applications. Hellwig (2021) provides a discussion on improved accuracy of ML methods and limitations.

representing traditional methods, we use Forbes and Warnock(2021) as our baseline model. We replicate and calculate out-of-sample prediction of this baseline model. Following this, we employ a set of ML methods, utilizing the same variables and dataset. The traditional ML methods in our analysis include Elastic Net, Random Forests, Support Vector Machines, kNN (k-Nearest Neighbors), AdaBoost (Adaptive Boosting), and XGBoost (Extreme Gradient Boosting), along with Multi-Layer Perceptron (MLP). We conduct a comprehensive comparison of the out-of-sample performance of these ML methods against the base model. This comparison utilizes various performance metrics such as accuracy, precision, recall, F1-score, and the receiver-operating characteristic (ROC) curve to assess their predictive power. Through this analysis, we aim to explore whether ML algorithms can contribute to enhancing out-of-sample performances in predicting Sudden Stops.

In the second phase, we extend the dataset sourced from the International Monetary Fund (IMF), with the sole criterion being data availability. After extension of the dataset, we employ ML methods for feature selection and subsequently use the selected features for ML estimation. The out-of-sample comparisons of the methods are then conducted. This part involves a range of traditional ML methods, including Elastic Net, Random Forests, Support Vector Machines, XGBoost, Logistic Regression, and a modern ML approach – the deep learning method known as Long Short-Term Memory (LSTM). This comprehensive analysis aims to explore the potential improvements in out-of-sample performances facilitated by ML algorithms and feature selection techniques.

In this way, we contribute to the emerging literature on the effectiveness of machine learning (ML) methods in macroeconomic forecasting through an interdisciplinary approach.

In the remainder of this section, we offer an overview of our approach and highlight key themes covered. The subsequent chapter delves into the literature review, encompassing empirical and theoretical studies on the Sudden Stop Crisis, providing a comprehensive understanding of the topic. Following this, in Chapter 3, we explore the fusion of machine learning (ML) with economics.

Within this chapter, we navigate through various facets of this integration, including the *subsection "Revolutionizing Statistical Paradigms: Navigating Data Complexity with Algorithmic Insights."* Here, we explore the factors that may have contributed to the gradual adoption of ML within economics. Drawing insights from influential figures such as statistician Leo Breiman, economists Hal R. Varian, we uncover the shifting cultural dynamics that facilitated ML's acceptance in economics. Moreover, we elucidate the potential benefits arising from this fusion.

Continuing in the same vein, our focus shifts to the role of ML in addressing challenges within macroeconomic forecasting. This exploration takes shape in the subsection "Machine Learning's Role in Overcoming Challenges in Macroeconomic Forecasting: Tackling Complexity, Small Sample Sizes, and Rare Events.

Further narrowing our focus, we examine the enhancement of macroeconomic forecasting in the context of predicting Sudden Stop Crises. The subsection "Enhancing Macroeconomic Forecasting Through Machine Learning: Addressing Challenges in Predicting Sudden Stop Crises" specifically scrutinizes the complexities surrounding forecasting these crises and how ML can adeptly navigate these obstacles, leading to improved out-of-sample predictions. In this context, we also refer to the related literature.

Transitioning to the subsection "***Predicting Sudden Stops Using Supervised Machine Learning,"*** we outline our strategy for forecasting Sudden Stops through supervised ML techniques. We explain how to set our prediction problem in ML setting. Starting from setting the prediction problem as binary classification problem, we briefly covered the main necessary steps such as splitting the dataset as test and training subsets, model selection. Emphasizing the significance of model generalization, we explore cross-validation and hyperparameter tuning. Moreover, we explain the error types—Type 1 (False Positive) and Type 2 (False Negative) and the interpretation in Sudden Stops—and how they impact predictive accuracy.

Understanding errors is crucial in predictive modeling, especially in binary classification. Two significant types of errors in this context are Type 1 (False

Positive) and Type 2 (False Negative). Type 1 Error occurs when the model predicts a positive outcome that doesn't happen. For Sudden Stops (SS), this could mean wrongly predicting an upcoming SS event. Conversely, Type 2 Error happens when the model fails to predict a positive outcome that does occur. In SS, this could involve overlooking signs of an impending crisis.

These errors in SS have substantial implications, affecting the economy and decision-makers. A Type 1 Error, or "False Positive," can disrupt financial markets, institutions, and public perception, leading to economic restrictions and panic. Ironically, measures meant to prevent issues might intensify volatility. On the other hand, a Type 2 Error, or "False Negative," affects the economy's stability by missing a real Sudden Stop. Lack of readiness and missed intervention chances can worsen a crisis, leading to recession, diminished investor confidence, and sovereign debt distress.

After explaining the types of errors, we outline the process of model selection and introduced performance metrics such as accuracy, precision, recall, specificity, F1-Score, AUC-ROC, and Precision-Recall Curve, emphasizing their roles in guiding our model selection. Performance metrics play a crucial role in predictive modeling, quantifying, and evaluating model effectiveness. They provide a structured approach to assess how well a model's outputs align with real-world outcomes, assisting decision-makers in refining approaches. In the context of binary classification, where the goal is to categorize data into two classes, these metrics are particularly crucial. As mentioned earlier, model selection involves comparing various models using performance metrics. The choice of metrics is influenced by policy goals and the specific types of errors one intends to minimize. In other words, when evaluating different models, the selection of performance metrics is tailored to the objectives and priorities set by the policies in place, considering the types of errors that need to be reduced or avoided. For instance, a policymaker prioritizing minimizing false negatives (Type II error) is better off checking recall, F1 score, or AUC-ROC scores. Suppose the other case, if a policymaker prioritizes minimizing false positives (Type I error), it is better to compare precision, F1 score, or AUC-ROC scores. This

subsection, in summary, offers a structured and concise strategy for Sudden Stop prediction, unifying key concepts and systematic steps.

In Chapter 4, our primary aim is to conduct a comparative analysis of the out-of-sample performance of Sudden Stop prediction among several Machine Learning methods and a conventional statistical approach, specifically the complementary log-log method (cloglog), within the Generalized Linear Model (GLM) framework. To commence, we utilize the capital flow dataset curated by Forbes and Warnock (2021) to identify Sudden Stop Crises in 59 countries, encompassing both advanced and emerging economies during the period from 1978 Q1 to 2020 Q3.

We then explain the Sudden Stop Identification methodology, which uses gross flows, as established by Forbes and Warnock (2012), and implement it on the dataset. The subsequent section is dedicated to presenting the results, where key findings are highlighted. Following this, we replicate the estimation process for the base case, as presented in their analysis, employing the complementary log-log model. The estimation period spans from 1986 Q1 to 2018 Q4, with exogenous variables including global liquidity, global risk (VIX), global growth, the average long-run interest rates of the UK, US, Euro Area, and Japan, along with contagion and local real GDP growth. Afterwards, we scrutinize its out-of-sample performance, establishing it as our baseline scenario. Subsequently, employing the same dataset, we implement a range of supervised Machine Learning methods and conduct a comparative analysis of their respective out-of-sample performances. Before delving into the estimation results, we also provide non-technical brief explanations for the ML methods used in this chapter, including Elastic Net, Random Forests, Support Vector Machines, kNN (k-Nearest Neighbors), AdaBoost (Adaptive Boosting), and XGBoost, along with Multi-Layer Perceptron (MLP).

In Chapter 5, we present machine learning-based solutions designed to tackle the Sudden Stop prediction problem. Our methodology involves an expansionary approach to the dataset employed in the preceding chapter, integrating comprehensive quarterly data sourced from the International Monetary Fund (IMF).

This enriched dataset is then harmonized with the model-selected variables utilized in our earlier estimations, along with additional variables identified from Forbes and Warnock (2021), thereby significantly augmenting the dataset.

While the initial dataset comprises model-selected variables widely used in empirical studies, the extension process is principally motivated by data availability. Variable selection is conducted independently of their alignment with theoretical foundations or common usage in predicting Sudden Stop events.

It is essential to underscore that the dataset extension does not involve prioritizing specific data-driven variables. Instead, it involves the inclusion of a broader set of variables without pre-selection. In this context, conventional statistical criteria, such as collinearity considerations, are set aside. Our methodology is firmly grounded in a robust reliance on machine learning (ML) *feature selection* methods. These methods are chosen for their demonstrated ability to objectively identify and prioritize relevant variables within the expanded dataset.[3]

Our objectives in this chapter encompass a dual focus. First, we strive to identify the pivotal variables influencing the occurrence of Sudden Stop events. Secondly, leveraging the selected variables, our aim is to construct highly discriminative ML models that exhibit superior performance on out-of-sample data. This two-fold approach is geared towards enhancing the precision and efficiency of our predictive models in forecasting Sudden Stops.[4]

In **Chapter 5**, the subsequent actions can be outlined as follow: First, we explain how to extend our dataset and incorporate other data implications such as 1 quarter lagged versions, year-over-year percentage changes, and rate of change versions. From the IMF's Balance of Payment, International Financial Statistics, and Direction

---

[3] This underscores the merit of employing ML methods in discerning crucial variables. Consider a scenario with a multitude of variables where the significance of each is uncertain. Absent ML-driven feature selection, discerning the importance of individual variables becomes a challenging task.

[4] It is worth noting here that by using selected variables in the ML model, we reduce the risk of overfitting.

of Trade datasets, we obtain 192 variables. With the inclusion of other versions, the final set of variables derived from this process totals 768. Having excluded the current levels of the variables, we supplement this set with the 1-period lagged versions of the 28 variables selected from Forbes and Warnock (2021), resulting in a total exogenous variable set of 604. The dataset spans from 1994q1 to 2018q4 .Further details on this process are explained in this subsection.

Next, we elaborate on how we employ feature selection methods for each model. It's important to note that different feature selection methods are used for different models: Random Forest, XGBoost, and SVM utilize the Recursive Feature Elimination (RFE) method for feature selection, while Elastic Net employs the Lasso method for variable selection, incorporating regularization to promote sparsity. Logistic Regression adopts a hybrid approach, initially using Random Forest for feature ranking and subsequently employing stepwise Bayesian Information Criterion (BIC) for further selection.

Subsequently, we explain the estimation results after performing feature selection and retrain the models with the selected features. We then proceed to make out-of-sample prediction performances. In this section, we also provide a brief non-technical explanation of the Long Short-Term Memory (LSTM) method.

In the culmination of our comprehensive exploration into predicting Sudden Stop events using a variety of analytical methodologies, ***Chapter 6*** serves as the final reflection on our findings and contributions to the field. We delve into the key takeaways, implications, and potential avenues for future research based on the outcomes of our investigation.

First, we provide a concise summary of the main results derived from our analysis of Machine Learning methods and traditional statistical approaches. The comparative performance across different models sheds light on the strengths and limitations of each methodology in forecasting Sudden Stop events. This comprehensive evaluation allows us to draw meaningful conclusions regarding the effectiveness and applicability of these techniques in the realm of macroeconomic forecasting.

Furthermore, we discuss the practical implications of our findings for policymakers, economists, and other stakeholders involved in decision-making processes. The insights gained from our study can inform strategies to mitigate the impact of Sudden Stop events, fostering more robust economic policies and risk management frameworks.

Additionally, we reflect on the challenges encountered during our research and the methodologies employed to address them. A critical examination of these challenges enhances the transparency of our study and provides valuable insights for researchers undertaking similar endeavors in the future.

In essence, Chapter 6 encapsulates the culmination of our research, offering a synthesis of our findings, their practical implications, and directions for future exploration. Through this comprehensive conclusion, we aim to contribute not only to the existing body of knowledge but also to the practical applications of predictive modeling in macroeconomics.

CHAPTER 2

LITERATURE REVIEW IN SUDDEN STOPS

Since the Global Financial Crisis of 2008-2009, there has been a growing concern regarding credit boom-bust cycles, international capital flow volatility, and reversals. Policymakers have increasingly turned to macroprudential policies to ensure financial stability and macroeconomic stability. One area of interest that has garnered significant attention is the phenomenon known as the Sudden Stop Crisis. Mendoza and Korinek (2013) state that the Sudden Stops observed in emerging markets during the 1990s were a precursor to the global financial crisis of 2008. Bianchi and Mendoza (2020) report that by the end of 2016 recorded a total of 58 Sudden Stop events recorded worldwide, with 35 taking place in emerging markets and 23 in advanced economies. During Sudden Stops, countries experienced a loss of access to credit, leading to abrupt reversals in current account balances and triggering severe economic downturns, often referred to as Great Recessions.

The capital flow pattern during the 2000s and early 2010s witnessed fluctuations, with a drying up of capital flows in late 2001, followed by a surge in the mid-2000s driven by low-interest rates and relaxed lending standards, and a sharp contraction during the Global Financial Crisis in 2008, as investors became risk-averse. However, there was a quick rebound in 2010 as the global economy began recovering.

After the Global Financial Crisis, unconventional monetary policies and near-zero interest rates in the United States facilitated the flow of international capital into emerging economies. However, episodes like the "taper tantrum" in 2013, sparked by concerns that the Federal Reserve might reduce its purchases of securities, and the "normalization" episode in 2015, fueled by expectations of rising U.S. interest rates, highlighted the possibility that Sudden Stops may be becoming more frequent or potentially more disruptive (Eichengreen and Gupta, 2016).

13

Forbes and Warnock (2021) asserts that the waves of global capital flows that resembled giant tidal waves during the 2008-2009 crisis have now transitioned into more manageable "ripples." Notably, the most significant ripples occurred in 2015 when investors anticipated the U.S. Federal Reserve's interest rate hike after nearly a decade. During this period, approximately 27% of countries experienced sudden stops in capital flows, and 22% faced retrenchment. These rates were considerably lower than the peaks observed before the crisis (34% and 32%, respectively) and during the 2008-2009 crisis (80% and 63%, respectively). However, when examining Emerging Market Economies (EMEs), extreme capital flow episodes were often higher and closer to pre-2008 levels, particularly in 2015. This suggests that EMEs remain more vulnerable to sharp capital flow movements, including Sudden Stops, compared to advanced economies.

Given the significance of these issues, policymakers and academia have intensified their focus on capital regulations and macroprudential studies to enhance macro-financial stability. In 2012, the IMF altered its perspective and highlighted the potential advantages of capital controls, calling for further research on a comprehensive, flexible, and balanced approach to managing capital flows. More recently, the IMF (2022) emphasized the importance of countries having the option to proactively restrict debt inflows to safeguard macroeconomic and financial stability. This recommendation is particularly relevant in cases where external liabilities pose risks, especially when they generate currency mismatches due to external debt denominated in foreign currency without appropriate foreign currency assets or hedges.

2.1. Background and Analytical Approach to Sudden Stops

Calvo (1998) introduced the concept of "Sudden Stops" in capital account to address the deficiencies of traditional economic models in explaining the distinctive characteristics of the crises witnessed in the 1990s. The Mexico crisis in 1994-1995 (known as the Tequila Crisis) and the Asian financial crises in 1997-1998 presented puzzling challenges that could not be fully explained by the existing theories. While the Mexico crisis was initially linked to fiscal and current account deficits, the Asian

economies, with their high saving rates and often-low current account deficits or even surpluses, experienced even deeper and more protracted crises. The crises in Mexico and Asia showed that even countries with seemingly strong fundamentals could be vulnerable to sudden and severe disruptions in capital flows, leading to a reassessment of the factors contributing to financial fragility. These differing characteristics raised doubts about the effectiveness of previous models in capturing the complexities of these events. Relatedly, Krugman (1999)'s critique of existing models of "third-generation currency crisis" highlights the inadequacies of these models in explaining the perplexing observations during the crises. Krugman (1999) identifies two main views in the post-crisis theoretical literature: one suggesting hidden subsidies leading to reckless spending and the other arguing that the countries' investments were fundamentally sound but vulnerable to "financial fragility" and self-fulfilling pessimism from international lenders. However, he points out that neither of these views fully accounts for the severity of the crisis, failing to explain factors like contagion, the transfer problem, and the role of balance sheet problems in constraining firms. Krugman (1999) emphasizes the need for a comprehensive "third-generation" model that considers factors beyond the banking system, such as the role of companies' balance sheets in influencing investment capacity and the impact of capital flows on the real exchange rate. This approach, known as the balance-sheet effect, has become essential in explaining the mechanism behind Sudden Stops, as highlighted in studies by Calvo and Reinhart (2000), Calvo et al (2008), Mendoza (2002), and Bianchi (2011) and others.

Calvo (1998) provides an analytical approach to Sudden Stops in Net Capital Flows. It uses basic accounting identities and economic principles to discuss the mechanics of a sudden stop in capital inflows (it need not result in capital outflow) and show that this could have large deleterious effects on the economy, validating the pessimistic conjectures that likely led to the initial stop.

2.1.1. Effects of a Capital Inflows Slowdown: The Non-Monetary Economy

In a non-monetary economy, by abstracting from errors and omissions, capital inflows (KI) and the current account deficit (CAD) are related through the accounting identity:

$$KI = CAD \qquad\qquad (2.1)$$

Additionally, in both monetary and non-monetary economies, considering tradable and nontradables goods, the following identity holds:

$$CAD = Z - GNP = Z^* - GDP - NFTA \qquad\qquad (2.2)$$

,where Z, Z^*, GNP, GDP^*, and NFTA represent aggregate demand, demand for tradables, gross national product, gross domestic product of tradables, and net factor transfers abroad.

During a capital-inflows episode, KI experiences a sharp and sustained increase, leading to high CADs. A sudden stop in KI results in a sudden contraction in CAD, which could, in theory, be accommodated by reducing the demand for tradable goods without affecting output. However, this is unlikely to be the case. A lower demand for tradable goods (Z^*) is likely to be accompanied by a lower demand for nontradables goods ($Z - Z^*$). In a flexible-prices world, this implies a higher real exchange rate, which could lead to unexpected problems, such as nonperforming loans in the nontradables sector (e.g., real estate), causing widespread bankruptcies.

(I)The severity of the damage caused by a sudden stop in capital inflows depends on how easily the associated fall in CAD can be accommodated. The conjecture (I) put forth is that the larger the share of consumption in total expenditure (Z), particularly on tradable goods (Z^*), the more pronounced the damage to the real economy from a fall in CAD. This conjecture is based on the assumption that consumption of tradable goods is more labor-intensive than investment in tradable goods. As labor is predominantly a nontradable good, a reduction in aggregate demand for tradable goods would result in a larger cut in the demand for nontradables, leading to a deeper real devaluation and more significant financial turmoil. This differs from the traditional observation about a country's solvency based on investment's impact on debt repayment.

The analysis of sudden stops discussed above did not consider the maturity structure of a country's debt, a factor that has gained significant attention in the aftermath of

recent crises. However, in principle, the theory of sudden stops remains independent of the debt maturity structure. For instance, if a country's current account deficit (CAD) is entirely financed by foreign direct investment (FDI), sudden stops might not be a major concern if all FDI is directed towards new investments. On the other hand, if FDI involves the purchase of existing firms, it may not necessarily lead to a higher CAD. The transaction could result in equal-value asset accumulation in the opposite direction, leaving the CAD unchanged. The key factor determining the impact of a cut in FDI lies in how the proceeds are spent (as indicated by conjecture 1).

While the impact of a cut in capital inflows (KI) is theoretically independent of the debt maturity structure, the actual size of the cut may not be. Debt maturity structure, particularly the residual debt maturity (i.e., the time profile of maturing debt), becomes relevant in assessing the potential reversal of capital flows, particularly the largest possible short-run fall in KI. Additionally, the probability of debt refinancing plays a crucial role, which is influenced by the country's standing with institutions like the IMF and key G7 countries, and its ability to respond if official refinancing is not available. It is reasonable to conjecture that countries with shorter residual debt maturity structures may be more vulnerable to sudden stop crises, which leads to the Conjecture 2 claiming that *(II) The shorter is the residual maturity structure of a country's debt, the more fertile will be the ground for a sudden stop crisis.*

While Calvo (1998) acknowledges that the initial trigger for sudden stops in capital inflows is often external, he emphasizes the significant role played by *self-fulfilling mechanisms* in exacerbating the crisis. He explains how the conjectures (I) &(II) that initially lead to a sudden stop can become self-fulfilling prophecies. Two lines of reasoning are presented for this perspective:

I. The first line of reasoning suggests that a capital inflows slowdown could potentially push the economy into insolvency. However, this point appears less relevant, particularly in the case of Asian countries. Nevertheless, there might be situations where the shock does not lead to bankruptcy explicitly, but it affects investors' perception of the country's willingness to pay, impacting investor confidence and leading to similar outcomes.

II. The second line of reasoning, which will be the main focus, revolves around the drastic lowering of the "average and marginal productivity of physical capital" due to socially costly bankruptcy battles following sharp and unexpected changes in relative prices. Sudden stops are likely to trigger widespread bankruptcies across various sectors of the economy.

The process of bankruptcy brings forth several challenges to the economy. One significant aspect is the destruction of specific human capital within firms. Bankruptcy interferes with the fulfillment of implicit contracts, such as internal promotion schemes tied to track records. New owners, taking over distressed firms, may repudiate these contracts and debts, eroding incentives within the firm and reducing the effectiveness of the labor force. In extreme cases, as seen after the breakdown of the former Soviet Union, firms can be cannibalized by employees and managers, leading to severe losses in human capital.

Moreover, bankruptcies have negative externalities beyond individual firms. Many firms depend on credit, and a surge in bankruptcies raises concerns about the solvency of not only the directly affected firms but also those connected through the credit channel. This heightened uncertainty prompts the need for additional information to assess firms' creditworthiness, diverting human capital from productive activities to financial matters, thus depressing the productivity of physical capital.

This destruction of human capital and the credit channels has profound consequences. It impedes consumption smoothing, and the new temporary equilibrium experiences a sharp decrease in the relative price of nontradable goods compared to tradable goods, leading to real depreciation.

Adding to the challenges, pro-cyclical policies can exacerbate the self-fulfilling output collapse prophecy. Many affected countries, influenced by IMF-sponsored programs, adopt tight fiscal and monetary policies. Tight fiscal policy further depresses the relative price of nontradables, potentially contributing to deeper and more widespread bankruptcies. Meanwhile, tight monetary policy intensifies the problem of credit destruction.

In conclusion, the sudden stop crisis, triggered initially by exogenous factors like capital inflows slowdown, can set in motion self-fulfilling mechanisms that further deteriorate the economy. The destruction of human capital, negative externalities of bankruptcies, and pro-cyclical policies all combine to deepen the economic turmoil, making the road to recovery more challenging.

2.1.2. The Monetary Economy

A monetary economy differs from the "real" one in that, instead of identity (2.1), we have:

$$KI = CAD + RA \qquad (2.3)$$

,where KI stands for capital inflows, CAD for the current account deficit, and RA for the accumulation of international reserves per unit of time.

The previous discussion on the self-fulfilling prophecy mechanism still applies to the monetary case. However, a key difference is that a slowdown in capital inflows (i.e., a reduction in KI) could now be met by a loss of international reserves (i.e., a fall in RA). This implies that the output and credit collapse associated with a contraction in the CAD could potentially be cushioned by a loss of international reserves. However, this apparent solution is largely illusory in practice.

Consider the central bank balance-sheet identity (in terms of tradable goods):

$$RA + NDA = H \qquad (2.4)$$

,where RA denotes international reserves, H represents high-powered money, and NDA stands for net domestic assets, including the central bank's certificates of deposit, net worth, and government deposits at the central bank, among other things

In the scenario of an exogenous fall in capital inflows (KI), if the central bank holds onto its reserves, the economy would undergo a similar adjustment as in the non-

monetary case. However, to achieve a better outcome, the central bank needs to implement measures that release international reserves (RA) and allow the current account deficit (CAD) to decrease by a smaller amount than the decline in KI. One way to achieve this is by extending loans to firms and individuals facing reduced access to international credit. However, in practice, this approach is challenging due to credit rationing. When the country faces credit constraints, individuals and firms would find it advantageous to claim that they have lost their international credit lines, making it difficult for the central bank to allocate loans effectively.

During a slowdown in capital inflows, domestic interest rates typically increase. To counteract this, the central bank often boosts NDA (via methods like a discount window) to mitigate the interest rate hike. A higher *NDA* (keeping international reserves constant), in turn, results in an increase in the stock of high-powered money, *H*, and devaluation, i.e., a rise in the nominal exchange rate (i.e., the price of foreign in term of domestic currency). However, this approach alone doesn't fully resolve the adjustment problem since international reserves remain unchanged. To make this strategy effective, the central bank must intervene in the foreign exchange market and release international reserves.

To properly release reserves and address the KI contraction, the central bank might need to intervene in the foreign exchange market. Yet, such a move could expose the central bank to a speculative attack, especially if the country is committed to a fixed (or semi-fixed) exchange rate.

Even if the country is not committed to defending the currency, policies that involve reserves losses to mitigate the slowdown in capital inflows might require a departure from pure floating. Releasing reserves might offer temporary relief, but it could trigger further KI contractions and exert additional downward pressure on the real economy.

In summary, while releasing international reserves may appear to ease the impact of a sudden cut in capital inflows, implementing this strategy can be challenging and risky in practice, as evidenced by experiences and basic economic reasoning.

- *Impact of Sticky Prices and Wages in a Monetary Economy: Keynesian Channels and Limitations of Currency Devaluation in Addressing Current Account Deficits*: In a monetary economy, the possibility of sticky prices and wages arises, introducing a Keynesian channel through which a reduction in the current account deficit can have a depressive effect on output, independent of the channels discussed earlier. Keynesian considerations often suggest the option of countercyclical monetary policy or currency devaluation to mitigate major output losses. While devaluation can help adjust relative prices associated with lower capital inflows (KI) more quickly, it may not necessarily prevent financial crises, especially if debt is denominated in foreign exchange. Currency devaluation can raise the relative price of tradable goods compared to nontradables, but if the nontradables sector holds dollar-denominated debt, the problem remains similar to that discussed in the previous section. Devaluation alone may not be sufficient to bypass financial crises when debt is denominated in *foreign currency.* Additional measures may be necessary to address the impact on the financial sector and potential debt-related vulnerabilities.

- **Dollar-Denominated Debt vs. Domestic-Currency Denominated Debt:**
  If debt in the nontradables sector was denominated in domestic currency, it would not necessarily lead to a significantly better situation during a crisis. Devaluations triggered by crises are often accompanied by higher nominal and real interest rates, as seen in past instances like Mexico in 1995 and Korea in 1998. These devaluations are typically involuntary and raise doubts about the government's ability to control key macroeconomic variables. Consequently, both nontradables and tradables experience higher real interest rates, which implies a mounting debt burden. The main difference between dollar-denominated and domestic-currency denominated debt lies in the timing of financial distress. With dollar debts, the negative impact caused by devaluation is immediate, whereas with domestic-currency denominated debt, it may take several months or even years to surface.

In conclusion, Sudden Stops characterized by abrupt and significant swings in capital inflows pose severe risks to economies, leading to bankruptcies, human capital losses, and disruptions in local credit channels. Large current account deficits,

regardless of financing methods, are risky as they necessitate continuous inflows of capital, making them vulnerable during sudden stops. The negative impact of reduced capital inflows is likely to be greater when the propensity to spend on nontradables is higher. Short-term financing arrangements can exacerbate risks, contributing to larger capital outflows and destabilizing the economy further.

In the policy discussion, several key points are emphasized. The financial sector's role is crucial in addressing capital inflow fluctuations and related challenges. Controls on international capital flows should be accompanied by domestic capital market regulations. Comprehensive financial sector policies should cover all aspects, including domestic transactions overseen by the central bank, to manage potential fiscal burdens during financial crises. For financially closed and underdeveloped systems, gradual financial reform is recommended, focusing on firms' leverage ratios. Financially open systems should be maintained with strong rules to resemble stable countries. Financial liberalization should be mindful of potential risks, with liberalized systems better prepared to manage capital flows. Efficient bankruptcy regulations are essential to prevent significant destruction of human capital during liquidity crises. After a sudden stop in capital inflows, traditional monetary and fiscal policies may have limited impact, advocating a prudent approach focused on price stability and fiscal responsibility for a sustainable recovery.

2.2. Intertemporal Approach to Sudden Stops: Fisherian Sudden Stop Models

Highlighting the unique characteristics of Sudden Stops - deep recessions, sharp price corrections, and short duration, Mendoza (2002) suggests viewing them as a part of 'excess volatility,' making them short-lived phenomena in the cyclical dynamics of small open economies, distinct from regular business cycle patterns. Mendoza (2002)'s approach deviates from the conventional paradigm of perfect market assumptions in his analysis of Sudden Stops. He introduces financial frictions within an otherwise frictionless and flexible-price environment, giving rise to endogenous credit constraints. Crucially, the forward-looking behavior of economic agents serves as the catalyst for the distortions engendered by these constraints,

resulting in their binding or nonbinding nature contingent upon prevailing economic conditions.

Sudden Stops occur when credit constraints become binding, resulting in a sudden loss of access to international capital markets. While short-term in impact and not significantly affecting the long-run business cycle, they can have negative effects on social welfare. Policymakers must understand the dynamics and drivers of Sudden Stops to design effective preventive measures and mitigate their impact on economic stability and welfare. Various triggers, such as policy shocks, policy-credibility shocks, shocks to domestic productivity, or changes in international liquidity (e.g., world's real interest rate), can lead to the transition to a Sudden-Stop state. Understanding these factors is crucial for managing and responding to Sudden Stops efficiently.

Mendoza (2002) introduces a modification to the conventional flexible-price intertemporal approach used in studying current-account determination and business cycles in small open economies. He incorporates a credit friction that connects agents' borrowing ability to the endogenous dynamics of prices and income. This adjustment addresses empirical limitations in conventional models, which fail to account for the abrupt reversals in capital inflows and collapses of private consumption observed during Sudden Stops. The key reason for these counterfactual outcomes lies in the assumption of perfect credit markets, where agents can borrow or lend without constraints at the world-determined real interest rate based solely on their wealth (No-Ponzi-Game condition). The introduction of the credit friction relaxes this assumption and aims to capture important elements of credit frictions observed in the literature on emerging-markets crises, as studied by Calvo and Mendoza (2000). The proposed credit constraint, following the Fisherian approach, emphasizes the credit-market effects of price shocks within a neoclassical flexible-price environment. Notably, Mendoza demonstrates that Sudden Stops can be consistent with the optimal adjustment of a flexible-price economy, contingent upon the sudden binding of the credit constraint, which takes the form of a liquidity constraint that requires borrowers to finance a portion of their current obligations

from their current income—a common criterion used to screen borrowers in credit markets.

Mendoza's (2002) framework departs from conventional models in macroeconomics with financial frictions by introducing an occasionally binding endogenous collateral constraint. In contrast to the prevailing literature that typically adopts either the Keynesian setup featuring price or wage stickiness with an external financing premium (Bernanke, Gertler, and Gilchrist, 1998) or the Fisherian analysis of debt-deflations driven by collateral constraints (Kiyotaki and Moore, 1997) to model episodes akin to 'great depression,' Mendoza's approach considers Sudden Stops (SS) as an excess volatility phenomenon, offering a distinct perspective. Notably, the majority of existing studies in this domain tend to incorporate credit constraints that remain constantly binding along the equilibrium path. Consequently, such models face challenges in effectively accounting for the sudden and severe economic contractions observed during Sudden Stops, which emerge as atypical events within the smoother co-movements of regular business cycles. Mendoza's framework also stands apart from previous literature by emphasizing the intricate interplay among uncertainty, risk aversion, and incomplete contingent-claims markets, providing insight into the transmission mechanism that connects financial frictions to real economic outcomes. In alignment with models investigated by Aiyagari (1993), Aiyagari and Gertler (1999), and Eaton and Gersovitz (1981), the importance of precautionary saving and state-contingent risk premia is accentuated, underscoring their pivotal role in driving business cycle dynamics. Conversely, extant models of Sudden Stops based on the Kiyotaki-Moore or Bernanke-Gertler-Gilchrist frameworks often assume risk-neutral agents and perfect foresight, enabling analytical tractability through closed-form solutions and linear-approximation techniques. Nevertheless, these assumptions neglect critical elements, such as choice under uncertainty, risk aversion, and precautionary saving, which are deemed indispensable for a comprehensive analysis of economies characterized by imperfect credit markets.

"Liability dollarization," which refers to the prevalent practice of emerging economies having a significant portion of their debt denominated in U.S. dollars and

other strong currencies, constitutes a pivotal element in the transmission mechanism through which liquidity constraints exert influence on the real economy. Foreign debt, being denominated in the international unit of account (tradable goods), while being leveraged on income valued at a different relative price, engenders pronounced fluctuations in the production and relative price of nontradable goods. Consequently, such sharp fluctuations in output and nontradable prices arise as endogenous outcomes within the model, representing the economy's equilibrium adjustments in response to actual foreign or domestic shocks and policy uncertainties. Remarkably, this framework accommodates the occurrence of Sudden Stops, even in the absence of the traditional debt-deflation intertemporal channel and without resorting to the Keynesian hypothesis of rigid prices or wages or the postulation of multiple equilibria.

Mendoza's (2002) framework, 'Fisherian Sudden Stop Models with contractionary depreciations under liability dollarization,' was initially designed to model endogenous Sudden Stops in intertemporal settings within emerging markets. Subsequently, this framework has evolved to encompass Fisherian models with asset price deflation, proving applicable to analyze events like the Global Financial Crisis of 2008 and study Sudden Stops in advanced economies as well. These Fisherian Sudden Stop Models now serve as important tools in quantitative and normative investigations of Sudden Stops, providing valuable insights and advancing our comprehension of this economic phenomenon in both advanced and emerging economies.

2.3. Stylized Facts of Sudden Stops

Most empirical studies apply event analysis tools to cross-country panel datasets, using one or more filters to identify Sudden Stop events. Traditionally, many papers adopt the definition of Calvo (2004; 2008) when defining the sharp falls in net capital flows. A sufficiently large increase in the current account GDP ratio (ca/y) is widely used as the main identification filter, because the current account is the broadest measure of the flow of credit of an economy vis-a-vis the rest of the world, and hence a large increase in ca/y indicates a sharp contraction in credit (both private

and public) from abroad. This filter is often used together with a second filter that detects if the Sudden Stop is systemic across countries (e.g., using the EMBI+ index for emerging markets), and in some instances, other filters, such as large output drops to capture Sudden Stops with deep recessions (e.g., Calvo et al., 2013) are added.

Sudden Stops (SS) are economic fluctuations defined by a set of empirical regularities associated with a large, sudden reversal of capital inflows (i.e., a sudden "loss of access" to international financial markets). The defining characteristic of a Sudden Stop is a sharp reversal in external capital inflows, which is often measured by a sudden jump in the current account. At about the same time as the access to foreign financing is lost, or shortly after, the economies affected by Sudden Stops experience deep recessions, in many countries the largest since the Great Depression, sharp real depreciations and collapses in asset prices. This phenomenon occurs in both advanced and developed countries.

One of the recent event studies by Bianchi and Mendoza (2020) demonstrates the stylized facts, which is consistent with existing literature: i) across all countries, a typical SS event features a current-account reversal of 3.7 percentage points of GDP. The reversals are larger in EMs (4.4 percentage points) than in AEs (2.7 percentage points), ii) SS events are infrequent, but they are twice as likely to occur in emerging than in advanced economies. They found 51 Sudden Stops in total (2.4 percent frequency), of which 36 occurred in EMs (2.9 percent frequency) v. 15 in AEs (1.7 percent frequency). Hence, Sudden Stops are rare events that co-exist with typical business cycles. iii) SS events are clustered around "big events". They are not randomly distributed over time. For example, there are several years in which no Sudden Stops occur, while we observe 14 Sudden Stops in 1982 and 1983, when the Sovereign Debt Crisis of the early 1980s exploded, 13 in 1998 and 1999, when the Asian crisis occurred, and seven in 2009, with the Global Financial Crisis. iv) SS events are associated with sharp economic downturns, preceded by expansions, and followed by protracted recessions. For all countries combined, GDP and consumption are 2.5 and 1.6 percent below trend respectively. In EMs (AEs), GDP and consumption are 3.6 (1.1) and 1.5 (1.6) percent below trend, respectively.

Moreover, compared with the expansions that precede Sudden Stops, these downturns represent sharp reversals. Relative to the year before a Sudden Stop hits, the deviations from trend in GDP and consumption for all countries fall by 4.4 and 4 percentage points, respectively, and again the reversals are larger for EMs than AEs. Business cycles are also larger in EMs, so relative to the standard deviations of cyclical components, they are comparable. Investment shows a similar pattern, but with larger changes, since investment is also more volatile over the business cycle. For all countries, investment is nearly 11 percentage points below trend when a Sudden Stop hits, and this represents a reversal of nearly 19 percentage points relative to the year before. Two years after SS events, all three macro-aggregates remain significantly below trend. Across EMs and AEs, GDP and consumption are 1.5 to 2 percent below trend and investment 3.3 to 5.5 percent below trend.

The stylized fact demonstrates that credit boom-bust cycles are crucial in understanding sudden stop (SS) episodes in the economy. Credit booms are periods when credit to the private sector rises significantly above its long-run trend and often linked to economic turbulence. As highlighted by Mendoza and Terrones (2012), these credit booms are often associated with periods of economic expansion, rising equity and housing prices, real appreciation, and widening external deficits during the upswing phase, followed by opposite dynamics in the downswing. Credit booms demonstrate a global dimension as they tend to be synchronized internationally and are centered around major events, including the 1980s debt crisis, the 1992 ERM crisis, the 1990s Sudden Stops, and the 2008 Global Financial Crisis. However, not all credit booms end in crises, but when they do, they are frequently followed by banking crises, currency crises, or Sudden Stops with a similar frequency in both Emerging Markets (EMs) and Industrial Countries (ICs), approximately 20 to 25 percent for banking and currency crises, and 14 percent for Sudden Stops. Reinhart and Reinhart (2008) concur with Mendoza (2012)'s findings, adding that capital inflow bonanzas are no blessing for advanced or emerging market economies. In emerging markets, these bonanzas are associated with a higher likelihood of economic crises, while in developing countries; they are linked to procyclical fiscal policies and attempts to avoid exchange rate appreciation. On the other hand, the results are not as stark for advanced economies, but bonanzas are still associated with

more volatile macroeconomic outcomes for GDP growth, inflation, and external accounts. In line with the findings from Reinhart and Rogoff (2009) indicating that recoveries from recessions triggered by financial crises are slow, Jorda et al. (2012) contribute to the understanding of credit boom-bust cycles by showing that financial crisis recessions are more painful than normal recessions, and credit-intensive expansions tend to be followed by deeper recessions. They find that a large build-up of credit during an expansion is closely associated with the severity of subsequent recessions, both financial and normal. During credit booms, real credit per capita reaches approximately 30 percent above trend in the median of all emerging market credit booms, whereas in industrialized country credit booms, it peaks at 12 percent, highlighting the larger scale of credit expansion in emerging markets. Moreover, the aftermath of financial-crisis recessions, observed over a 5-year period, reveals a significant decline of about 5 percent in real GDP per capita compared to normal-recession paths, underscoring the higher costs associated with financial crises.

2.4. Determinants of Sudden Stops

A 'Sudden Stop,' as initially defined by Calvo (1998), refers to substantial negative swings in capital inflows, characterized by large and unexpected falls in net capital flows into a country. During Sudden Stops, countries lose its access to the international credit market, triggered by a loss of confidence from international lenders. Although the primary reason for a Sudden Stop is external and financial in nature, domestic vulnerabilities also play a crucial role in determining whether a country experiences such an abrupt capital flow reversal. These vulnerabilities encompass factors such as large current account deficits, liability dollarization, debt structures, and exchange rate regimes, among others.

To basically understand how this large and unexpected net capital flow falls affect real exchange rate, aggregate consumption and output, we can refer to Calvo and Reinhart (2000):

In national income accounting, capital inflows are equivalent to the current account deficit plus the accumulation of international reserves. During sudden stops,

countries experience either reserve losses or lower current account deficits, both of which have significant consequences. While a loss of international reserves increases a country's financial vulnerability, contractions in the current account deficit have serious effects on production and employment.

A sudden contraction in the current account deficit leads to a sharp decline in aggregate demand, resulting in reduced demand for both tradable and nontradable goods. While excess supply of tradables can be shipped abroad, nontradables remain domestically constrained, leading to a fall in their relative price and a real depreciation of the currency. This decline in output and employment occurs through two channels: the familiar Keynesian channel, which relies on inflexible downward prices and wages, and the less familiar but potentially more damaging Fisherian channel.

The Fisherian channel is based on financial contracts contingent on a few "states of nature," such as terms of trade and demand. The sudden stop calls for a lower relative price of nontradables, but since interest rates remain invariant, the ex post real interest rate faced by nontradables producers surges, resulting in an increased share of nonperforming loans. This can have severe implications for the financial sector, leading banks to become more cautious in lending, particularly to small and medium-sized firms. Consequently, enterprise and trade credits may dry up, contributing to a major and long-lasting recession.

Moreover, the impact of RER (Real Effective Exchange Rate) depreciation can be amplified in countries with a high degree of dollarization. Even in countries without significant dollarization, shorter maturity bank loans can exacerbate nonperforming loans due to revised upward real interest rates following the sudden stop. Calvo et al. (2008) also shows that the size of the increase in the Real Exchange Rate (RER) depends on the percentage fall in the absorption of tradables required to close the current account gap (CAD/Z). Therefore, *the main country-specific vulnerabilities to "Sudden Stops" are a large current account deficit, financial frictions like short-term debt maturity and non-state contingent debt, and liability dollarization, which amplifies the effects of such sudden capital flow reversals.*

29

Frankel (2005) also emphasizes the significance of *liability dollarization and debt structure* in rendering countries vulnerable to currency crises and sudden stops. Amid various mechanisms attempting to elucidate the contractionary effects of currency devaluations, especially in the context of the 1990s emerging market crisis, he highlights the paramount importance of the balance sheet mechanism. The presence of substantial foreign currency-denominated debts, particularly in U.S dollars, held by domestic banks and firms becomes a key factor in the vulnerability to sudden stops. These debts, manageable at previous exchange rates, become problematic to service after sharp increases in foreign exchange prices, leading to layoffs and bankruptcies. The susceptibility of countries to sudden stops is closely tied to the short-term debt structure and the extent of liability dollarization among firms and banks. The issue of "mismatch" arises between the currency denomination of a country's debts and the currency that its firms earn, adding to the precariousness of the situation.

2.5. Empirical Studies

2.5.1. A Review of Global and Country-Specific Variables in Sudden Stop Studies

The significance of the Sudden Stop crisis has drawn widespread attention, leading to increased efforts in understanding its determinants and vulnerabilities. Empirical studies in this field commonly employ traditional econometric methods like probit, logit, and complementary log-log model to examine the impact of various variables, encompassing both global factors (external variables) and country-specific indicators. Despite variations in time periods and country samples across studies, certain variables consistently emerge as key factors influencing the occurrence of Sudden Stops. In this context, this review aims to highlight the most frequently used variables and present some notable findings from empirical research in the field of Sudden Stops.

Current Account Deficits and Domestic Debt Structure: Most empirical studies consistently find that large current account deficits (CAD\GDP) are important in

understanding vulnerabilities associated with Sudden Stops. They are seen as a crucial factor indicating reliance on capital inflows and heightened risk of abrupt capital flow reversals. Similarly, most empirical studies identify domestic credit to GDP, domestic credit growth, and external debt to GDP as main predictors of Sudden Stops. These variables play a significant role in determining a country's vulnerability to abrupt reversals in capital flows. However, the effect of liability dollarization on the probability of Sudden Stops varies among studies. Calvo et al. (2008) emphasizes the importance of liability dollarization, particularly when combined with large current account deficits, as a crucial component of the balance sheet effect that can lead to a dangerous cocktail for Sudden Stops. On the other hand, studies by Cavallo and Frankel (2008) and Eichengreen et al. (2008) do not find liability dollarization to be statistically significant in predicting Sudden Stops.

Trade Openness: Trade openness is a widely used variable in studies examining the vulnerability of countries to Sudden Stops. Researchers often measure trade openness using the trade-to-GDP ratio, which reflects a country's level of integration into international trade. Cavallo and Frankel (2008) utilize the gravity method as an instrumental variable approach to address possible endogeneity issues in their analysis of trade openness. Their findings reveal a negative correlation between trade openness and the probability of sudden stops. Less open economies are more prone to experiencing sudden stops and currency crashes, consistent with the results of previous research like Edwards (2004). Additionally, they estimate that a 10-percentage-point increase in the trade to GDP ratio is associated with a 1-percentage-point reduction in the probability of a sudden stop. Similarly, Edwards (2004) also finds that the degree of trade openness influences the negative effects of sudden stops on economic growth. More open economies tend to be relatively less affected compared to less open ones, suggesting that trade integration can provide a certain level of resilience during sudden stops. Frankel (2005) supports the notion that increasing trade openness reduces the probability of sudden stops, indicating that greater trade integration may enhance a country's ability to navigate external shocks and maintain financial stability. However, contrasting findings from Milesi-Ferretti and Razin (1998, 2000) reveal conflicting evidence regarding the relationship between trade openness and current account reversals and currency crises.

31

Financial Openness and Financial Integration: Financial openness, or the degree of integration with global financial markets, has been a topic of interest in understanding vulnerability to Sudden Stops and external crises. Eichengreen et al. (2008) finds that financial integration with global markets reduces vulnerability to Sudden Stops. This suggests that countries with greater financial openness may be better equipped to manage external shocks and capital flow reversals. Contrary to critics of globalization, Edwards (2004) provides weak evidence suggesting that countries with a higher degree of financial openness have a lower probability of facing current account reversals. This finding implies that the presence of capital account restrictions may not effectively reduce the likelihood of external crises. Forbes and Warnock (2012) report that there is no evidence to support the notion that capital controls reduce the likelihood of having a surge or stop episode. In fact, the negative coefficient on financial integration suggests that countries more integrated with global financial markets are less likely to experience Sudden Stops. Calvo et al. (2008) emphasizes the significant role of financial integration in shaping the probability of Sudden Stops and highlights the importance of institutional development during the integration process: Using data from Lane and Milessi-Ferretti (2006), two measures of financial integration are considered: one for Foreign Direct Investment (FDI) and another for portfolio stocks. The results indicate that higher financial integration reduces the probability of a Sudden Stop. Interestingly, the relationship between portfolio integration and Sudden Stop probability differs between developing and developed countries. For developing countries, increased portfolio integration is associated with a higher likelihood of a Sudden Stop, while the opposite is observed in developed countries, where financial integration reduces the risk of such episodes.

Calvo et al. (2008) references Bordo (2007)'s idea that during financial integration, countries may be prone to crises, but these experiences can lead to the development of robust institutions, enhancing financial stability and reducing vulnerability to Sudden Stops. Furthermore, the study explores non-linearities in financial integration and finds that the relationship between portfolio integration and Sudden Stop probability is non-linear. This analysis aids in classifying emerging markets based on their integration levels and the likelihood of experiencing a Sudden Stop.

In summary, as countries move from low levels of financial integration, the probability of experiencing a Sudden Stop rises. However, with the progression of financial integration, this probability gradually diminishes and becomes negligible at high levels of integration. Notably, emerging markets find themselves in an intermediate position, situated between developed and other developing countries. Within this gray area, the probability of a Sudden Stop is observed to be the highest, indicating that financial integration can pose risks in the absence of supporting institutions for more sophisticated and credible financial instruments. This underscores the importance of balancing financial integration with the development of robust institutions, particularly for emerging markets, to ensure stability and minimize the potential risks associated with Sudden Stops.

International Reserves: The role of international reserves in mitigating vulnerability to Sudden Stops has been extensively studied. Edwards (2004) and Cavallo and Frankel (2008) find that countries with higher levels of net international reserves have a lower probability of experiencing a reversal. This implies that maintaining an adequate level of reserves can act as a buffer, helping countries navigate through external shocks and reducing the likelihood of sudden capital flow reversals. Their findings highlight the importance of prudent reserve management as a key protective measure for countries to enhance their financial stability. On the other hand, Calvo et al. (2008) does not find a significant relationship between the M2/reserve ratio and Sudden Stops, indicating that the specific measure of international reserves used in the study may not be a robust predictor of the likelihood of reversals. Overall, the majority of studies point to the crucial role of international reserves in bolstering a country's resilience to external crises, reinforcing the significance of prudent reserve policies as a crucial tool for countries to withstand the challenges posed by Sudden Stops and other external shocks.

Exchange Rate Regime: The literature provides mixed evidence regarding the impact of exchange rate regimes on the probability of Sudden Stops. Calvo et al. (2008) finds no significant role for exchange rate regimes in affecting the probability of Sudden Stops, suggesting that the specific type of exchange rate regime a country adopts may not have a significant impact on its vulnerability to abrupt capital flow

reversals. However, Edwards (2004), Cavallo and Frankel (2008), and Eichengreen et al. (2008) present contrasting results, indicating that countries with rigid exchange rate regimes are more vulnerable to experiencing Sudden Stops. This suggests that fixed or pegged exchange rate regimes may limit a country's ability to respond to external shocks and adjust to changing market conditions, potentially increasing the risk of abrupt capital flow reversals.

Contagion: Contagion, the phenomenon where Sudden Stops in one country can influence the likelihood of Sudden Stops in another country, has been a subject of investigation in the literature, with varying findings. Edwards (2004) and Eichengreen et al. (2008) report positive evidence of contagion, suggesting that the occurrence of Sudden Stops in one country positively affects the probability of Sudden Stops in other countries, consistent with the bunching feature of Sudden Stops. However, Cavallo and Frankel (2008) find the regional dummy, which captures the impact of Sudden Stops in neighboring countries, to be statistically insignificant, indicating that the proximity of Sudden Stops in neighboring countries may not significantly influence the probability of Sudden Stops in a particular country, according to their analysis. On the other hand, Forbes and Warnock (2012) emphasize the importance of contagion, especially for stops and retrenchment. Their findings indicate that countries are more likely to experience a Sudden Stop or retrenchment episode if their major trading or financial partners have recently experienced the same type of episode. Additionally, countries are also more likely to experience Sudden Stops and flight if their neighbors have faced similar episodes.

Fiscal Variables: Fiscal variables, including the budget deficit/GDP ratio and government external debt ratios, are another set of widely used country-specific variables in the literature to understand vulnerability to Sudden Stops. Eichengreen and Gupta(2016) and Calvo et al. (2008) suggest that while lower government deficits and a stronger fiscal stance can create some policy space for countries to respond to external shocks, these fiscal variables are not sufficient to prevent a country from experiencing a Sudden Stop. In other words, while sound fiscal policies may provide some degree of protection, they may not shield an economy from abrupt capital flow reversals or fully mitigate the risks associated with Sudden Stops.

Global variables: Global variables play a crucial role in understanding the incidence of Sudden Stops and other extreme capital flow events in emerging market economies. Eichengreen and Gupta (2016) compares two subperiods, 1990-2000 and 2001-2014, and finds that global risk aversion, as captured by the VIX index (which measures market volatility and investor sentiment), has become more important for the occurrence of Sudden Stops in the second subperiod.

This suggests that changes in market sentiment and increased global risk aversion have played a significant role in shaping capital flows and the likelihood of Sudden Stops in emerging market economies during the later period. In the first period, other global variables, such as G4 money supplies, world GDP growth, and the Federal Reserve's policy interest rate, were more significant. This highlights the dynamic nature of global factors in influencing the incidence of Sudden Stops over time. Similarly, Forbes and Warnock (2012) emphasize the significance of global risk measures, particularly the VIX, in predicting the probability of Sudden Stops and other extreme capital flow events, such as surges, retrenchment, and flight.

Additionally, they find that strong global economic growth is associated with a higher likelihood of surges and a lower likelihood of stops, indicating the influence of overall global economic conditions on capital flows in emerging markets. However, Forbes and Warnock (2021) reveal a shift in determinants after the Global Financial Crisis (GFC). They find that oil prices have become a more significant determinant of Sudden Stops and other extreme events, surpassing the importance of previously identified global risk measures. This suggests that changing global economic dynamics and external shocks can alter the relative importance of various global variables in shaping capital flows to emerging markets. Eichengreen et al. (2008) also identify oil prices as an important factor in capital flows.

They point out that many oil-exporting countries have increased domestic liquidity, leading to increased capital flows from oil-exporter countries to other emerging markets. This indicates the potential spillover effects of commodity prices and commodity-producing countries on capital flows in the global financial system.

2.5.2. 'Systemic' Sudden Stops and Examining the Significance of Balance Sheet Effects: Insights from Calvo et al. (2008)

Calvo et al. (2008) place significant emphasis on the balance sheet effect (BSE) and its relevance in understanding the likelihood of systemic sudden crises. To comprehend such episodes, it becomes crucial to consider market incompleteness with non-contingent assets and an external credit constraint, akin to Mendoza's approach (2002). The presence of these assets creates a "mismatch" between fixed liabilities and variable assets, depending on the state of the economy. The analysis of the BSE's relevance to the probability of a systemic Sudden Stop involves two main components: the change in relative prices, specifically the real exchange rate (RER) response to capital flow declines, and liability dollarization. Even without liability dollarization, the existence of non-state contingent assets attributes importance to the BSE in explaining how the initial external trigger affecting key relative prices eventually leads to a full-blown Sudden Stop Crisis.

The Balance Sheet Effect (BSE) refers to the phenomenon where a decline in key relative prices, such as the Real Exchange Rate (RER), negatively affects the net worth of firms and the entire private sector. This, in turn, leads to a reduction in borrowing capacity, resulting in a decrease in aggregate consumption of tradable goods and, consequently, non-tradable goods.

However, while Calvo et al. (2008) recognize the BSE as a crucial transmission mechanism, the study does not delve into the feedback loop associated with it. This feedback loop occurs when the initial trigger causing changes in relative prices leads to the deterioration of net worth, resulting in reduced borrowing and consumption. This decline in aggregate consumption further worsens relative prices, leading to additional deterioration in net worth. The interactions between declining relative prices, aggregate demand, and borrowing create a self-reinforcing cycle. This process of financial amplification through the feedback loop plays a central role in the Fisherian Models of Sudden Stops, initially proposed by Mendoza (2002; 2005).

'Systemic 'Sudden Stops (3S) are defined as sudden stops that take place in conjunction with a sharp rise in aggregate interest-rate spreads. An initial trigger that

is of a financial and external nature characterizes these episodes. Although 3S are initially triggered by factors that are exogenous to individual economies, whether this initial shock develops into a full-fledged Sudden Stop depends also on country-specific variables. These country-specific factors play a crucial role in determining the severity and impact of the initial shock on the domestic economy.

First, following Calvo (1998), a Sudden Stop is defined as a phase that meets the following conditions: i) It includes at least one observation where the year-on-year decline in capital flows (net flows) is at least two standard deviations below the mean value observed in the sample. This criterion addresses the requirement for the decline in capital flows to be considered "unexpected." ii) The Sudden Stop phase concludes when the annual change in capital flows surpasses one standard deviation below the mean value. This condition introduces the notion of persistence, a common characteristic of Sudden Stops, which often involve prolonged periods of declining capital flows. iii) To maintain symmetry in the definition, the initiation of a Sudden Stop phase is identified as the first instance when the annual change in capital flows falls one standard deviation below the mean value.

By applying these criteria, we can effectively identify and analyze phases of Sudden Stops based on their deviation from the typical patterns of capital inflows. Secondly, the interest here lies in the identification of "systemic'' Sudden Stops (3S), i.e., Sudden Stops with exogenous trigger. For this reason, it is required additionally that the detected SS windows coincide with a period of skyrocketing aggregate spreads. The same methodology is used to detect large changes in capital flows is used for aggregate spreads to detect periods of market turmoil.

To detect the empirical relevance of balance sheet effect, Calvo (1998) highlights two components: i) the sensitivity of RER to capital inflow falls and ii) liability dollarization (both in the private and public sectors). The sensitivity of RER to capital inflow falls is related to the size of the supply of tradable goods (Y) relative to demand for tradable good (Z). In other words, the size of the decrease in the RER depends on the percentage fall in the absorption of tradables needed to close the current account gap (CAD/Z): As a matter of fact, the less leveraged the absorption

of tradable goods is, the smaller will be the effect on the RER. To see this, rewrite CAD/Z as:

$$\frac{CAD}{Z} = \frac{Z-Y+S}{Z} = 1 - \frac{Y-S}{Z} = 1 - \omega \qquad (2.5)$$

where ω, defined as $\omega = \frac{Y-S}{Z}$, can be thought of as the un-leveraged absorption of tradables and S are international factor payments, remittances abroad, etc.

It is evident that the higher the supply of tradables (Y), the smaller will be financing from abroad (or leverage) of the absorption of tradables. Thus, high values of 1-ω mean that a country relies less on its own financing of the absorption of tradables, and is therefore more vulnerable to RER depreciation stemming from closure of the current account gap. Notice that the denominator in (2.5) is the absorption of tradables, and not GDP. This points to the fact that normalization of the current account deficit by the absorption of tradables may be more suitable than normalization by GDP when analyzing vulnerability to Sudden Stops.

ii) The second component of the balance sheet effect is 'liability dollarization': they focus on an even narrower concept of foreign-exchange denominated debt, namely, Domestic Liability Dollarization, (DLD), i.e., foreign-exchange denominated domestic debts towards the domestic banking system, as a share of GDP. The rationale behind this choice is that typically banks are at the heart of the economy's payment system and, thus, their bankruptcy or even temporary suspension of activities could trigger a serious supply shock.

With a sample of 110 countries, including 21 developed economies, and 89 developing countries for the period 1990- 2004, Calvo et al . (2008) adopts a panel Probit model that approximates the probability of falling into a full-fledged 3S episode as a function of lagged values of 1-ω and DLD, controlling for a set of macroeconomic variables typically used in the literature on determinants of crises.

To assess BSE, Calvo et al. (2008) examines the interaction between two key variables: ω, representing the un-leveraged absorption of tradable goods, and DLD

(Domestic Liability Dollarization), which measures the degree of foreign currency denominated domestic debts towards the domestic banking system as a share of GDP. ***The findings suggest that the effects of ω on the probability of a Sudden Stop depend significantly on the level of DLD.*** Countries with low ω values (indicating higher leverage of current account deficits) are more vulnerable to Sudden Stops, especially if they have high DLD levels. This interaction between ω and DLD has both statistical and economic significance. For example, when comparing two countries with ω values of 0.6 (the lowest measure in the sample), the one with high DLD (dollarized economy) experiences a substantially higher probability of a Sudden Stop compared to the country with low DLD (non-dollarized economy). The difference in the probability of a Sudden Stop between these two scenarios can be as much as 17 percentage points. Moreover, the non-linear nature of the relationship between ω and the probability of a Sudden Stop is evident. As ω approaches 1 (when the current account deficit is zero), the difference in the probability of a Sudden Stop between high and low DLD countries decreases to around 5 percentage points, representing about 30 percent of the difference observed at lower ω levels. ***The high non-linearity described by the data implies that low ω and high dollarization can be a very dangerous cocktail, as potential balance sheet effects become highly relevant in determining the probability of a Sudden Stop.*** Furthermore, the impact of DLD on the probability of a Sudden Stop is particularly pronounced for emerging markets. Before the Russian crisis in 1997, approximately 61 percent of EMBI+ countries (emerging markets) in the sample had DLD values above the dollarization median, while 80 percent of developed countries had DLD values below the median. This suggests that dollarization of liabilities is more prevalent and influential in determining the probability of a Sudden Stop for emerging markets compared to developed countries.

In summary, the probability of a Sudden Stop is significantly influenced by two main factors: a limited supply of tradable goods relative to their absorption and the extent of Domestic Liability Dollarization. These factors are closely tied to domestic policies, such as tariff and competitiveness policies affecting tradable goods supply, as well as fiscal and monetary mismanagement leading to Domestic Liability

Dollarization. While foreign creditors may test a country's resilience, vulnerability to Sudden Stops is primarily amplified by these domestic determinants. Furthermore, the impact of balance-sheet factors on the probability of a Sudden Stop exhibits non-linear characteristics. Specifically, when there is high leverage in tradable goods' absorption and substantial Domestic Liability Dollarization, the risk of a Sudden Stop becomes particularly pronounced, creating a potentially dangerous combination.

2.6. Fisherian Sudden Stop Models

The Fisherian Sudden Stop Models have emerged as a dominant approach in the literature, capable of providing both qualitative and quantitative predictions in line with the stylized facts of financial crises. This modeling approach is built on **occasionally binding collateral constraints**, leading to a financial amplification mechanism akin to Irving Fisher's debt deflation theory (1933).

Since the pioneering work of Mendoza (2002), the literature on Fisherian Sudden Stop Models has seen significant growth. This two sectors (tradable goods and non-tradable goods) small open economy-DSGE framework with occasionally binding collateral constraint explores Sudden Stop Crises as endogenous outcomes resulting from responses to standard shocks after periods of leverage build-up. Over time, researchers have extended and explored numerous variants of this framework.

An important characteristic of this framework is its incorporation of **pecuniary externalities**, providing a natural foundation for understanding and designing macroprudential policies. These policies are aimed at promoting financial stability and resilience against boom-bust cycles in international capital flows, a subject of great interest since the 2008 Global Financial Crisis (GFC). In the aftermath of the GFC, policymakers have shown increased focus on macroprudential policies, capital controls, and the formulation of optimal strategies to maintain macro-financial stability.

As a consequence, several influential papers have employed this framework to analyze *normative aspects of Sudden Stops* and optimal macroprudential policies.

For instance, Bianchi (2011), building upon Mendoza (2002) framework, investigates the concept of 'overborrowing' in emerging economies and its correlation with Sudden Stop Crises as endogenous outcomes. Overborrowing occurs due to agents' failure to fully consider the price effect of their current consumption-borrowing choices on their future borrowing capacity. In simpler terms, individuals cannot internalize the impact of today's borrowing decisions on the relative prices, i.e. the real exchange rate, which in turn affects the borrowing ability via occasionally binding-endogenous collateral constraint. Consequently, they are unable to internalize the price effect, leading to an undervaluation of the social marginal costs of borrowing and ultimately giving rise to overborrowing. By solving the problem for a social planner, who can internalize this pecuniary externality, Bianchi (2011) achieves a constraint-efficient allocation and proposes an optimal borrowing tax, serving as a form of capital inflow controls to decrease the frequency and severity of Sudden Stop Crises in emerging economies. Similarly, Benigno et al. (2016) contribute to the literature by discussing post-crisis policy alternatives, known as fiscal policy interventions. Benigno et al. (2016) explores the role of subsidies on non-tradable goods or taxes on tradable goods, which results in a change in the relative prices of non-tradable goods to tradable goods, i.e. the real exchange rate. The aim of these policy interventions is to alleviate the decrease in the real exchange rate, thereby mitigating the financial amplification effect.

Fisherian Sudden Stop Models are characterized by *the presence of an occasionally binding collateral constraint* for borrowers. Typically depicted as small open economies with a representative agent, these models allow agents to borrow from international credit markets up to a certain fraction of their collateral's market value, which is influenced by the endogenous aggregate states of the economy. This endogeneity of collateral value and borrowing capacity leads to the emergence of asymmetry and amplification of negative shocks, particularly evident during sudden stops when debt levels in the economy are high.

In the literature on financial frictions, collateral constraints are often directly imposed on the optimization problems of agents, rather than arising as an endogenous outcome of explicitly modeled contracts. This approach aligns with

common practice in macroeconomic studies on financial frictions, following influential works by Aiyagari and Gertler (1999) and Kiyotaki and Moore (1997). However, some studies take a different approach, incorporating collateral constraints derived from contractual setups, typically resulting from limited enforcement or costly state verification (e.g., Mendoza and Quadrini, 2010; Bianchi and Mendoza, 2018).

The endogenous nature of borrowing capacity and the presence of borrowing constraints are features shared with a broader class of financial frictions models, including the classical financial accelerator model by Bernanke and Gertler (1989). In this model, an external financing premium, determined by net worth, emerges endogenously as an optimal contract outcome. Nonetheless, the Fisherian models distinguish themselves due to the occasional binding of the borrowing constraint, setting them apart from other studies in financial frictions and macroeconomics, such as Kiyotaki and Moore (1997) and Bernanke and Gertler (1989).

While the collateral constraint typically remains loose during most periods, it becomes binding when triggered by a standard shock after an expansionary phase, initiating the Fisherian debt deflation mechanism. In these episodes, declines in relative prices (or asset/house prices, depending on the collateral type) and the market value of collateral interact and reinforce each other, along with aggregate demand and borrowing. This feedback loop gives rise to a vicious circle of declines, intensifying the severity of the sudden stop crisis.

Figure 2.1 illustrates the fundamental mechanics of financial amplification schematically: Consider an emerging economy that borrows from foreign sources, subject to a collateral constraint. As the current account tends to be countercyclical, periods of economic expansion also witness a buildup of leverage. Consequently, when leverage ratios reach sufficiently high levels, the collateral constraint becomes binding, compelling agents to reduce their expenditures. This, in turn, leads to a decrease in aggregate demand, causing declines in real exchange rates, relative prices, and asset prices. Since the value of collateral is closely linked to these relative prices, these declines further tighten the collateral constraint, prompting agents to curtail spending even more. Thus, a detrimental feedback loop is triggered.
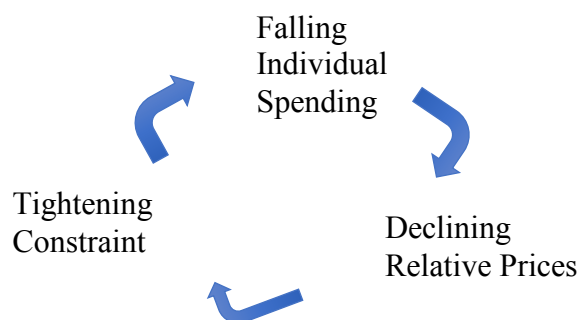
Figure 2. 1. Financial Amplification Effects

In a broader context, Mendoza and Korinek (2013) present a comprehensive framework of Fisherian Sudden Stop Models and discuss their applications in three categories:(i) Sudden Stop Crisis in Emerging Markets: Focuses on contractionary real exchange rate devaluations under 'liability dollarization' (i.e., debts denominated in different units than incomes and collateral)(ii) Sudden Stop Crisis in Advanced Economies: Examines scenarios with asset price deflations, particularly relevant for cases like the Global Financial Crisis of 2007-2008.(iii)Equilibrium Business Cycle Model: Explores a full-blown equilibrium model encompassing various aspects of the business cycle. Through these applications, the Fisherian Sudden Stop Models offer valuable insights into the dynamics of financial crises and their implications for both emerging markets and advanced economies.

2.6.1. General Structure of Fisherian Sudden Stop Models

Consider a small open economy in infinite time $t = 1,2,3.....$ The economy is inhabited by a representative agent who receives a stochastic endowment every period and values consumption according to a standard time-separable expected utility function:

$$U = \sum \beta^t E[u(c_t)] \tag{2.6}$$

,where $\beta < 1$ subjective discount factor and u(c_t) is a standard twice-continuously differentiable, strictly concave period utility that satisfies the Inada conditions. Foreign creditors are large compared to the small open economy and trade one-

period non-state contingent discount bonds with the domestic agent. International bonds carry an exogenous, time- and state-invariant price of $\frac{1}{R}$, where R is the gross world real interest rate. Since the simple setup is the only internationally traded asset, the also defines the country's net foreign asset (NFA) position. The period budget constraint is

$$c_t + \frac{b_{t+1}}{R} = y_t + b_t \tag{2.7}$$

where $\frac{b_{t+1}}{R}$ is the value of bond purchases carried as savings into the ensuing period and b_t is the repayment on the bond holdings of the home agent at the beginning of period t.

The assumption that bonds are not state-contingent implies that risk markets are **incomplete,** hence the small open economy has an incentive to self-insure. Moral hazard problem limits domestic agents' borrowing ability. After borrowing in period t, borrowers have the option to abscond. Lenders can identify this behavior, and if they act promptly, they can recover up to $\overline{b}$ units of the lent amount; otherwise, the entire loan is lost, and lenders have no further recourse or methods of punishment.

To prevent borrowers from absconding, lenders set a limit on lending, denoted by $\overline{b}$. The borrowing limit $\overline{b}$ generally depends on the aggregate state of the economy: In a booming economy with an appreciated exchange rate and elevated asset prices, lenders will have a higher likelihood of recovering funds than in a depressed economy characterized by low exchange rates and asset prices. This leads to another assumption that the financial constraint depends on the aggregate consumption. Therefore, the dependence of borrowing consumption on the aggregate conditions is expressed by

$$\frac{b_{t+1}}{R} \geq -\overline{b}\,(C_t) \tag{2.8}$$

,where $\overline{b}\,'(C_t) > 0,$ i.e., higher consumption increases borrowing capacity.

This reduced form constraint (2.8), at the most general level, is at the heart of Fisherian effects. The variants of this setting based on relative price changes that are associated with the declines in aggregate consumption are examined: i) Fisherian models in which falling consumption causes real exchange rate, i.e. Fisherian Sudden Stop Models with contractionary devaluations under 'liability dollarization'' ii) Fisherian models in which falling consumption causes the asset price declines.

The combination of **non-state-contingent debt** and **the collateral constraint** plays a crucial role in generating Sudden Stops as an equilibrium outcome in this framework. These financial imperfections create a *mismatch* between the denomination of the agent's financial liabilities and their borrowing capacity. This asymmetry leads to financial amplification effects, wherein the liabilities of the agent are non-state-contingent, while the borrowing limit fluctuates in parallel with aggregate states over the business cycle. In the event of adverse shocks, the borrowing limit tightens, but the level of debt remains unchanged, preventing the agents from smoothing the impact of these shocks over time. Consequently, the representative agents experience a Sudden Stop. In essence, a Fisherian model of financial amplification necessitates a relative price that links the value of collateral with borrowing ability.

2.6.2. Fisherian Models with Contractionary Devaluations under 'Liability Dollarization'

First proposed by Mendoza (2002), the representative agent in small open economy faces an occasionally binding debt-to income constraint. It assumes that financial liabilities in emerging economies are often denominated in hard currencies (or tradable goods) but backed up by income or assets from the non-traded sector of the economy. Hence, the relative price between liabilities and the value of collateral is the relative price of non-traded to traded goods, i.e., the real exchange rate.

To *introduce liability dollarization*, the general model is extended to include traded and a non-traded good. Agents are assumed to receive both tradable and non-tradable endowments. Then, the agents ` borrowing constraint becomes consisting of tradable and non-tradable income, which is a function relative prices. In other words, agents

can borrow against their income, which is itself a function of the relative prices. Since the price of the tradable goods is assumed to be stable, the changes in the price of non-tradables change the relative prices, i.e., the real exchange rate. Therefore, the key ingredient for financial amplification is the real exchange rate that connects borrowing ability to the market value of collateral.

The debt- to income (DTI) constraint of the representative agent is given by

$$\frac{b_{t+1}}{R} \geq -\kappa_t(y_t^T + y_t^N \cdot p_t^N) \qquad (2.9)$$

,where y_t^T, y_t^N are tradables and non-tradables income respectively, and p_t^N is the relative price of non-tradables to non-tradables, which is considered as the real exchange rate as the price of tradables is assumed to be 1.

When the constraint becomes binding, the representative agent experiences a shock to net worth or endowment income of sufficient magnitude similar amplification dynamics is set in motion. However, the dynamics now occur through movements in the country`s real exchange rate. A negative shock forces the agent to contract consumption of traded goods because he is unable to borrow the amount needed to support the unconstrained allocation. For the economy to absorb the available supply of non-traded goods, the real exchange rate must depreciate. But this reduces the value of the agents' income and collateral, and tightens the financial constraint, which forces further cutbacks in consumption, and leads to a feedback loop. The balance sheet effect, which connects constrained borrowing to tradables demand and real depreciation, is extensively employed in the Sudden Stops literature, beginning with Calvo (1998). However, the financial amplification of this effect through the Fisherian debt deflation mechanism is only present in models of the class. This feedback mechanism amplifies the impact of shocks when the credit constraint is always binding compared to situations where the constraint is not binding.

Fisherian models present a distinct approach from other credit-constraint models with always binding constraints. While the latter models consistently impose a

negative effect on aggregate demand via the balance sheet effect, Fisherian models introduce a dynamic feedback loop that triggers the Fisherian financial amplification mechanism when the credit constraint becomes binding. In Fisherian Models of Sudden Stops, the binding credit constraint is directly tied to market prices, particularly the value of collateral. This connection sets in motion a series of interactions that intensify the impact of shocks on the economy. When the constraint binds, agents resort to fire sales, leading to declining prices.

As prices fall, the value of collateral also diminishes, further tightening the credit constraint and amplifying the downturn in economic activity. This nonlinear effect causes a more severe decline in aggregate demand during sudden stops. In contrast, other credit-constraint models exhibit financial amplification but lack the intricate feedback loops and interplay between declining prices, borrowing capacity, and collateral values characteristic of Fisherian models.

Mendoza and Rojas (2018) discuss that `liability dollarization`, when introduced into the traditional model Fisherian Sudden Stop Model, brings about three significant effects that play a crucial role in shaping the dynamics of the economy:

i.   *Impact on Debt Burden*: Fluctuations in real exchange rates have a direct impact on the burden of repaying existing debt. As the value of the domestic currency fluctuates relative to foreign currencies, the cost of servicing debt denominated in foreign currency can vary. If the domestic currency depreciates, the burden of repaying foreign-denominated debt increases, making it more challenging for borrowers to meet their debt obligations. On the other hand, if the domestic currency appreciates, the burden of repayment may ease. This effect adds an additional layer of complexity to the debt dynamics in the model.

ii. *Influence on Domestic Bond Prices and Real Interest Rates*: Expected changes in real exchange rates also have an impact on domestic bond prices and real interest rates. As investors anticipate future fluctuations in the exchange rate, they adjust their expectations of the risk associated with

holding domestic bonds. This, in turn, affects the prices of domestic bonds in the market. Furthermore, expected changes in real exchange rates also influence the real interest rates in the economy, which can affect investment and borrowing decisions. These changes in bond prices and real interest rates add another dimension to the financial market dynamics.

iii. *Incentive for Risk-Taking*: Liability dollarization creates a risk-taking incentive represented by a negative premium on the ex-ante domestic real interest rate. This negative premium is a consequence of the positive relationship between the real exchange rate and aggregate consumption. As a result, there is a negative conditional co-variance between future marginal utility and future real exchange rates, leading to a reduction in the marginal cost of borrowing for domestic agents. This, in turn, strengthens borrowing incentives, encouraging domestic agents to take on more debt.

The positive findings demonstrate that the debt-repayment-burden effect plays a crucial role, leading to two important implications: first, it mitigates the severity of Sudden Stops, and second, it makes multiple equilibria harder to achieve. However, achieving multiplicity requires much higher limits on debt-to-income ratios and income realizations, which must fall within a narrower range of relatively high values.

The Debt-to-Income (DTI) constraint (2.9), initially proposed by Mendoza (2002) to examine the relationship between real-exchange-rate movements and borrowing capacity, has become a fundamental component in various economic models. Notably, the literature extensively employs this formulation of the credit constraint in a wide range of contexts. For instance, Durdu et al. (2009) and Arce et al. (2019) utilize the DTI constraint to study reserve accumulation models, while Bianchi (2011) explores macroprudential policy models. Additionally, Benigno et al. (2013) investigate real-exchange-rate stabilization policies, and Hernandez and Mendoza (2017) analyze ex-post intervention with industrial policy, all employing the DTI constraint. Moreover, Schmitt-Grohé and Uribe (2021) delve into self-fulfilling crises, Bianchi et al. (2016) examine noisy news and regime-switching shocks, and

Flemming et al. (2019) as well as Seoane and Yurdagul (2019) study trend shocks, all using the DTI constraint.

Furthermore, Bengui and Bianchi (2018) consider imperfect enforcement in capital-flow management policies, while Mendoza and Rojas (2019) investigate models with banks intermediating capital inflows in tradable units for domestic loans in units of the domestic Consumer Price Index (CPI), both incorporating the DTI constraint. Additionally, Ottonello (2015) and Farhi and Werning (2016) explore exchange-rate policy models with nominal rigidities and credit frictions, all utilizing the DTI constraint as a crucial element in their analyses.

Moreover, Fisherian models explore various adverse supply-side effects that can manifest during periods of financial instability. These effects encompass declining values of marginal products of inputs due to price deflation (Durdu et al., 2009), binding credit limits affecting working capital (Bianchi and Mendoza, 2018), and declining investment triggered by collapsing equity prices (Mendoza, 2010). By incorporating these supply-side dynamics, Fisherian models offer a more comprehensive understanding of how financial crises can influence not only aggregate demand but also the supply of goods and services in the economy.

Additionally, some Fisherian models extend their analysis to explore international spillover effects, accounting for factors such as international asset trading, short-selling constraints, and mark-to-market capital requirements (Mendoza and Smith, 2006; Mendoza and Quadrini, 2010). These international dimensions are crucial in understanding how financial disturbances in one country can propagate globally, contributing to the systemic nature of financial instability. Bianchi and Mendoza (2020) further enhance the Fisherian model by incorporating tradable and nontradables goods and introducing investment goods production using both types of inputs. The model captures households' consumption choices, investment decisions, and their constraints on borrowing, which is limited to a fraction of the market value of the capital stock. The model's focus on the real exchange rate and market price of capital, determined by the relative price of nontradables to tradables, plays a central

role in the financial amplification mechanism, magnifying the effects of shocks on the economy through changes in borrowing capacity.

2.6.3. Fisherian Models with Asset Price Deflation

In this type models, a representative agent borrowing is limited by the market value of asset prices. When the collateral constraint is binding, i.e., sudden stop crisis, similar Fisherian financial amplification is set in motion: Agents fire-sale assets to meet the constraint , which causes asset prices to decline. The decline in the asset prices reduces the market value of collateral, which in turn deteriorates the borrowing ability, aggregate demand. The declines in the asset prices, borrowing, and consumption mutually reinforce each other. Mendoza and Smith (2006), Bianchi and Mendoza (2010) and Jeanne and Korinek (2010) adopt this type of constraints in their Sudden Stop analyses.

2.6.4. Mean-Converting (transitory), Conventional Shocks vs. Unconventional Shocks, 'News Shocks' in Fisherian Models

Bianchi et al. (2018) investigate the crucial role of unconventional shocks in generating Sudden Stops and driving financial crises by using Fisherian Sudden Stop model with liability dollarization (Mendoza 2002; Bianchi 2011). These unconventional shocks come in the form of news about future economic fundamentals and regime changes in world interest rates, and they interact with collateral constraints to influence the dynamics of financial crises. Notably, during favorable economic conditions, when "good news" aligns with a low-world-interest-rate regime, borrowing incentives are strengthened. However, this heightened borrowing activity during prosperous times increases vulnerability to financial crises and magnifies the impact of collateral constraints.

While Fisherian Sudden Stop models have demonstrated substantial amplification and asymmetry in response to standard-size shocks, such as TFP and terms-of-trade shocks, they have predominantly focused on conventional shocks—usually TFP or interest-rate shocks—that follow symmetric probabilistic processes known to agents.

As a result, two critical sources of financial volatility, namely noisy news about future economic fundamentals and regime shifts in global liquidity, have been overlooked in the analysis of macroprudential policy. This stands in contrast to empirical studies on credit cycles and financial crises, which highlight the importance of considering such factors as essential determinants of credit dynamics and their interaction with the real economy (Calvo et al., 1996; Shin, 2013; Bruno and Shin, 2015; Mendoza and Terrones, 2012; Borio, 2014; Reinhart and Rogoff, 2014; Schularick and Taylor, 2012).

To bridge this gap, Bianchi et al. (2018) aim to fill these research limitations by introducing both news shocks and regime switches in global liquidity into a Fisherian model of macroprudential policy. They incorporate noisy yet informative news about future income shocks, following recent advances in the macroeconomic literature on news and economic fluctuations (e.g., Beaudry and Portier, 2006; Schmitt-Grohé and Uribe, 2012; Jaimovich and Rebelo, 2009; Christiano et al., 2010; Blanchard et al., 2013). Additionally, they consider shifts in global liquidity as a regime-switching process in the evolution of world interest rates or leverage limits, recognizing their significance in driving capital inflows and domestic credit in emerging economies, as documented in prior studies (Calvo et al., 1996; Shin, 2013; Eichengreen and Gupta, 2016).

Through quantitative experiments calibrated using data for Argentina, the authors reveal the significant effects of news shocks and global liquidity regimes on the Fisherian financial amplification mechanism. Specifically, they find that good news and low interest rates fuel credit booms, which can lead to severe financial crises if positive shocks fail to materialize or if sudden shifts in financial regimes occur. Moreover, the precision of news plays a crucial role, as higher information accuracy results in agents accumulating fewer precautionary savings, leading to less frequent but more severe financial crises.

In a related paper, Akinci and Chahrour (2018) conducted a study investigating the role of "good news" about future productivity in the dynamics of Sudden Stops and financial amplification. They employed the small open economy RBC model of

García-Cicco et al. (2010), with the addition of an occasionally binding collateral constraint as in Mendoza (2010). In their estimated model, positive news about productivity leads to an increase in leverage, thereby raising the probability of a Sudden Stop occurring in the future. Prior to the Sudden Stop, the economy experiences a boom period characterized by consumption and investment levels above trend, which aligns well with empirical data. During the Sudden Stop, the nonlinear effects of the constraint cause consumption and investment to significantly fall below trend, while the trade balance undergoes a sharp reversal, mirroring real-world observations. Remarkably, the study highlights the substantial risk posed by good news, with almost 90% of Sudden Stops occurring after positive news shocks. This suggests that financial crises can be triggered by positive news followed by adverse outcomes, even in the absence of any actual changes in the underlying fundamentals. Furthermore, the incorporation of news shocks about future productivity enables their model to better-fit patterns of negative trade balances and rising debt, which have proven to be strong predictors of financial crises in previous research (Gourinchas and Obstfeld, 2012; Schularick and Taylor, 2012).

2.6.5. Mean-Converting (transitory), Conventional Shocks vs. Trend Shocks in Fisherian Models

Seona and Yurdagul (2019) investigates sudden-stop-like crises in small open economies subject to collateral constraints. They find that Fisherian Sudden Stop Models with liability dollarization fail to produce the observed persistence and sluggishness in the recovery after the crisis. To address this, they extend Bianchi (2011), following Mendoza (2002), by introducing "trend shocks". Their study highlights the importance of permanent income shocks in generating plausible Sudden Stop dynamics. By including both transitory and trend shocks, based on Aguiar and Gopinath (2006), they estimate these shocks using Argentinean data from 1876 to 2004.The inclusion of trend shocks in the model successfully captures the dynamics of Sudden Stops, supporting the conclusions of previous studies by Aguiar and Gopinath (2006), Aguiar and Gopinath (2007), and García-Cicco et al. (2010). These studies have highlighted the crucial role of trend shocks and financial frictions in generating the observed business cycles dynamics in emerging economies.

52

The positive result of including trend shocks has significant economic implications, as it leads to different debt dynamics in small open economy models compared to transitory shocks alone. The Permanent Income Hypothesis underlies these differences, where a negative transitory shock increases debt to smooth consumption, while a negative trend shock results in a permanent decline in consumption. Consequently, the trend shock model exhibits a deleveraging effect after Sudden Stops, reflecting households' recognition of a permanent decrease in wealth. In contrast, the model with only transitory shocks experiences deleveraging due to the tightened borrowing constraint resulting from low output levels.

Moreover, the model with trend shocks demonstrates overborrowing compared to the constrained efficient economy, aligning with Bianchi (2011) `s findings. This means that the competitive equilibrium exhibits higher debt levels than the constrained planner's solution. However, including trend shocks introduces a new aspect of overborrowing. In contrast to a model with only mean-reverting shocks, the economy experiences more overborrowing during bad times when debt is issued to smooth consumption. Conversely, with trend shocks, households tend to overborrow during good times to increase present consumption, expecting higher future income. This behavior leads to increased borrowing during prosperous periods. Overall, Seona and Yurdagul (2019) shed light on the importance of permanent income shocks in generating Sudden Stops and highlights the role of trend shocks in explaining debt dynamics and overborrowing in open economies with collateral constraints.

CHAPTER 3

MACHINE LEARNING AND ECONOMICS

Machine Learning (ML) methods are computational techniques that enable computers to learn from data and improve their performance on a specific task over time. These methods are designed to recognize patterns, relationships, and trends within data, allowing them to make predictions, classifications, and decisions without being explicitly programmed for each specific scenario. ML algorithms adapt and refine themselves based on the data they are exposed to, enabling them to handle complex and dynamic situations.

ML methods can be categorized into two main types: supervised learning and unsupervised learning. In supervised learning, algorithms are trained on labeled data, where the correct outcomes are provided for input examples. The algorithm learns to map inputs to outputs and can make predictions on new, unseen data. Common supervised learning techniques include regression for predicting numerical values and classification for categorizing data into classes.

Unsupervised learning, on the other hand, deals with unlabeled data. The goal here is to find hidden patterns and structures within the data. Clustering is a common unsupervised learning technique that groups similar data points together, while dimensionality reduction methods help in reducing the complexity of data by retaining essential features.

Additionally, there are more advanced techniques like reinforcement learning, which focuses on training agents to make sequential decisions in an environment, and deep learning, a subset of ML that utilizes neural networks with multiple layers to learn complex patterns.

Overall, ML methods have a wide range of applications, from image and speech recognition to recommendation systems, medical diagnosis, financial forecasting,

and more. They enable computers to handle large and complex datasets, discover insights, and make accurate predictions, greatly enhancing our ability to solve complex problems in various fields.

3.1. The Fusion of Machine Learning and Economics: Navigating Complexity and Shaping Decisions

The convergence of machine learning (ML) and economics has ushered in a transformative era, redefining the landscape of economic analysis, policy formulation, and decision-making. This dynamic fusion has given rise to a myriad of innovative applications across diverse economic domains, empowering economists, policymakers, and businesses to navigate intricate challenges with unprecedented insight and foresight. As the digital era evolves, the integration of ML into economics offers promising avenues for augmented forecasting, policy articulation, risk management, resource allocation, crisis prediction, and strategic decision-making.

At its core, ML's predictive power takes center stage. Its proficiency in capturing non-linearities and interaction terms within complex economic systems is a game-changer. In the realm of macroeconomic forecasting, this prowess becomes particularly vital, enabling the prediction and management of major economic crises, including banking and financial crises. Armed with historical economic data, ML algorithms demonstrate a unique capacity to forecast future trends with unparalleled precision. These forecasts transcend conventional parameters, encompassing a broad spectrum of variables such as GDP growth, stock market dynamics, inflation rates, and more. By extracting complex patterns from extensive datasets, ML empowers economists to sharpen their predictions and make decisions anchored in a profound comprehension of economic intricacies. This predictive capability becomes especially relevant in anticipating potential crises, like those within the banking sector, where early detection of anomalies and patterns could provide crucial time for policymakers and institutions to strategize and enact preventive measures, thereby potentially mitigating the impact of such crises. The flexibility and adaptability of

ML allow economists to navigate the intricate web of economic variables and dynamics, enhancing the resilience of economies in an ever-changing global landscape.

The realm of financial markets, known for their volatility and intricacies, reaps substantial benefits from ML's capabilities. Algorithms adept at deciphering nuanced market data patterns have revolutionized financial analysis, enabling investors to identify trends and anomalies that eluded conventional methodologies. The optimization of portfolios and risk management reaches new heights as ML-driven insights guide investment strategies, infusing them with an enhanced awareness of potential market fluctuations. Hoang and Wietgratz (2022) offer an overview of machine learning's role in finance along with recent applications.

However, the synergistic alliance between ML and economics extends well beyond the confines of financial sectors. One of ML's pivotal strengths lies in policy analysis, where it becomes an invaluable instrument for simulating and evaluating the consequences of economic policies. Policymakers can model diverse scenarios and preempt potential outcomes, fostering well-informed decisions and preempting potential risks. This proactive methodology mitigates uncertainties associated with policy changes and establishes a foundation for an adaptive, responsive economic landscape. Kleinberg et al. (2015) delve into the significance of machine learning methods due to their capacity for enhanced predictions and their role in influencing policy decisions within the field of economics.

An often underestimated yet potent facet of ML lies in its proficiency to analyze textual data through Natural Language Processing (NLP). This capability proves particularly significant in gauging public sentiment, a pivotal determinant of economic conduct. By sifting through voluminous textual content from news articles, social media, and myriad sources, ML algorithms unearth sentiment trends and predict their potential influence on economic activities. This equips decision-makers with real-time insights into public sentiments, further enriching their strategic responses. For example, Chen et al. (2023) use a variety of machine learning techniques on multiple sources of textual data to identify and predict financial crises.

56

Furthermore, the impact of machine learning also reaches resource allocation, ensuring resources are used well across sectors, reducing waste, and boosting efficiency. This technology also plays a role in analyzing monetary policies by uncovering hidden patterns in complex monetary data, aiding in creating effective policies.

The enormity of modern economic data, often referred to as "big data," finds a natural ally in ML. With its capacity to process vast datasets, ML illuminates hidden patterns and relationships, enabling economists to discern insights that traditional methods might overlook. Techniques like clustering unravel complex structures within data, offering novel perspectives and enhancing the depth of economic analysis.

In summary, the blending of machine learning and economics has brought about a significant change. It's changing how we look at economic data, make policies, make choices, distribute resources, and predict crises. This teamwork allows researchers in various fields to use big sets of data effectively, get practical insights, and handle economic complexities with great accuracy. This connection has far-reaching effects and is set to create a future marked by well-informed, adaptable, and strong economies.

3.2. Revolutionizing Statistical Paradigms: Navigating Data Complexity with Algorithmic Insights

In this section, we explore the factors that have caused a slower adoption of machine learning (ML) within economics. By drawing on insights from notable figures such as statistician Leo Breiman, Hal R. Varian, and economists Athey, Imbens, and Kleinberg, we aim to understand why this integration took time to unfold. Through their perspectives, we will uncover the changing cultural dynamics that eventually paved the way for ML's acceptance in economics. Furthermore, we'll explore the potential benefits that this fusion brings to the field.

In 2001, the eminent statistician Leo Breiman offered thought-provoking insights into the necessity of reshaping the utilization of statistical modeling to extract

meaningful conclusions from data. He categorized two distinct cultures within the statistical realm: one rooted in the assumption that data follows a predefined stochastic model, and the other characterized by the utilization of algorithmic approaches that remain agnostic about the underlying data mechanism. Breiman candidly critiqued the prevalent tendency within the statistical community to predominantly rely on data models, noting that such a commitment led *to "irrelevant theory, questionable conclusions, and kept statisticians from working on a larger range of interesting current problems."*

In his assertion, Breiman (2001) urged for a more flexible incorporation of algorithm-based tools into the traditional model-driven approach. He articulated that the model-driven perspective should adapt more readily to the rapidly developing landscape of algorithmic modeling, which demonstrates rapid progress and applicability in various fields outside of statistics. Breiman recognized that the strength of algorithmic modeling lies in its adeptness at accommodating the complexities of both large and intricate datasets and serving as a compelling alternative to data models for certain scenarios, particularly on "smaller data sets." And he adds *'If our goal as a field is to use data to solve problems , hence we need to move away from exclusive dependence on data models and adopt a more diverse set of tools. '*

Provocatively critiquing the data-driven approach, Breiman (2001) notes that the insistence on data models has led to the stagnation of multivariate analysis tools, relegating them to discriminant analysis and logistic regression in classification and multiple linear regression in regression. He highlighted the inherent mismatch between the multivariate normal assumption and the reality of multivariate data. Breiman's skepticism centered on the imposed simplicity of parametric models for complex systems such as medical or financial data, which he argued *"result[s] in a loss of accuracy and information compared to algorithmic models."*

Breiman (2001) eloquently expressed how this adherence to data models could inadvertently limit statisticians' problem-solving capabilities by binding them to a

particular perspective. He invoked the well-known saying, *"If all a man has is a hammer, then every problem looks like a nail,"* suggesting that such narrowness of approach hindered statisticians' ability to tackle complex and varied problems that were emerging as data and computational capabilities advanced.

In terms of predictive power, he claims that higher predictive accuracy is associated with more reliable information about the underlying data mechanism. Weak predictive accuracy can lead to questionable conclusions. Algorithmic models, on the other hand, can give better predictive accuracy than data models and provide better information about the underlying mechanism.

Consequently, Breiman(2001) emphasized that the selection of a suitable approach, whether a data model or an algorithmic model, should be guided by the nature of the problem and the characteristics of the data. He cautioned against the assumption that a data model is always the best fit. Instead, he championed a holistic focus on the problem and the data's characteristics to determine the most fitting approach.

> *'The goals in statistics are to use data to predict and to get information about the underlying data mechanism. Nowhere is it written on a stone tablet what kind of model should be used to solve problems involving data. To make my position clear, I am not against data models per se. In some situations, they are the most appropriate way to solve the problem. But the emphasis needs to be on the problem and on the data.'*

In his final remarks, Breiman(2001) underscored that the field of statistics should be geared toward using data to predict outcomes and uncover insights about the underlying data mechanism. He envisioned a return to statistics' foundational roots, wherein working with real-world data and collaborating across disciplines would be pivotal for the field's vitality and growth. Breiman's vision embraced the challenges posed by complex data and the potential of algorithmic tools to provide deeper insights, envisioning a more dynamic and collaborative future for the discipline.

> *' The roots of statistics, as in science, lie in working with data and checking theory against data. I hope in this century our field will return to its roots. There are signs that this hope is not illusory. Over the last ten years, there*

has been a noticeable move toward statistical work on real world problems and reaching out by statisticians toward collaborative work with other disciplines. I believe this trend will continue and, in fact, has to continue if we are to survive as an energetic and creative field.'

Aligning with Breiman's perspective (2001), Athey and Imbens (2019) highlight that the key factor impeding the adoption of potent Machine Learning (ML) techniques in economics is the cultural reliance on model-driven statistical methodologies. Athey (2019) affirms that "A significant aspect of this phenomenon may indeed stem from the culture, as Breiman alludes. Economics literature emphasizes methods endowed with formal traits akin to those absent in many ML methods. These encompass attributes related to estimators and tests in large samples, encompassing attributes like consistency, normality, and efficiency. "

Within traditional econometrics, the primary focus is on parameter estimation, aiming for unbiased parameter estimations. Additionally, relatedly, the field addresses hypothesis testing and the creation of confidence intervals. Conversely, ML algorithms prioritize accurate prediction, shedding the stringent assumptions and restrictions of traditional methods, striving for more broadly applicable models with reduced variance. The contrast in underlying assumptions between traditional statistical methodologies and ML has hindered the integration of ML into economics. Furthermore, while traditional econometrics centers on parameter interpretations, this feature is frequently absent in ML methods.

In addition to cultural reliance, another factor contributing to the delayed integration of ML methods in economics pertains to the inherent nature of these methods. ML methods possess the ability to automatically capture non-linearities and interactions, enabling a transition from rigid assumptions to more adaptable models, or from basic models to complex ones. This significantly empowers ML in prediction. Nevertheless, these advancements often lack explicit causal explanations, leading to the emergence of the concept of a "black box" characteristic. This interpretational deficiency has historically hindered the incorporation of ML in economics. However, this perspective has recently shifted, with numerous studies integrating ML into economics and finance. Kleinberg et al. (2015) argues that in certain cases,

prediction supersedes parameter estimation. He asserts that "Empirical policy research frequently revolves around causal inference. Given that policy decisions often depend on understanding counterfactual scenarios—what occurs with or without a policy—this close connection between causality and policy becomes apparent. While this link holds in many instances, we contend that there are policy applications where causal inference is not central or even necessary ". Kleinberg et al. (2015) advocates for the adoption of ML due to its potential for superior outcomes and underscores its importance in optimal resource allocation and policy impact assessment. Furthermore, recent studies have incorporated Shapley Values to assess the individual impact of variables, as demonstrated by Bluwstein et al. (2023). The Shapley Value, originating from game theory, has found application in machine learning to discern the specific contributions of features within predictions. By deconstructing complex models, it assigns importance to each feature through the consideration of various combinations. This approach aids in the interpretation of predictions, particularly in intricate models, thus enhancing transparency and uncovering feature interactions. Additionally, other techniques have emerged and gained widespread usage, such as 'permutation feature importance' and 'partial dependency plots,' serving the same purpose. Moreover, 'surrogate models' can handle the inherent black box nature of ML methods. These models simplify complex machine learning predictions, offering understandable insights. They act as bridges between complex algorithms and the need for clear understanding. Economists can also benefit from surrogate models to comprehend relationships between variables and predictions. These simpler models assist economists in making informed decisions by revealing the impact of different variables. By extracting insights from complex models, surrogate models empower economists to combine accuracy with interpretability for effective decision-making.

Similarly, Athey and Imbens (2019) acknowledge that obtaining causal inference through ML methods can often be challenging, particularly in the context of microeconometric analysis. They emphasize, however, that ML's primary strength lies in its capacity to enhance policy analysis by providing improved predictions for instrumental variables. Despite the inherent difficulty in achieving causal inference

with ML, its ability to yield enhanced predictive results for instrumental variables holds paramount significance in the realm of microeconometric analysis. In another study ,Athey et al. (2019), they thoroughly explore this matter and offer guidance to economists on leveraging ML methods to achieve enhanced results in the realm of causal inference. They assert that "*ML tools are progressively becoming standard across various disciplines, necessitating economists to adapt their toolkit while preserving the enduring strengths of applied econometrics.*" The authors present a selection of tools and methods within ML that they propose should be integral to the toolkit of empirical economists and should be integrated into core econometrics graduate courses.

Moreover, Varian (2014) also strongly advocates for a broader utilization of ML methods in econometrics. He advises graduate students to enroll in machine learning courses within computer science departments, emphasizing the productive collaborations between computer scientists and statisticians, and anticipating similar productivity between computer scientists and econometricians. From the viewpoint of a statistician/econometrician , Varian (2014) provides a comprehensive explanation of the implementation of ML methods and underscores the methodological disparities between traditional econometrics and algorithmic approaches. He introduces the strategies he labels as "tricks for econometrics." In terms of addressing prediction challenges, he recommends the application of ML methods, elucidating specific techniques that are analogous to those familiar to econometricians. These techniques include Classification Trees, Random Forests, and Regularized Logistic Regression.

Furthermore, Athey and Imbens (2019) and Mullainathan and Spiess (2017) offer explanations of ML methods tailored for economists, approaching machine learning as an applied econometric approach. These authors collectively reference the influential book "Elements of Statistical Learning" authored by Hastie, Tibshirani, and Friedman (2009), which holds a pivotal place in their work.

In conclusion, the evolution of statistical methodologies and the integration of machine learning (ML) into economics reflect a paradigm shift towards more

adaptable and holistic approaches. Leo Breiman's critique of the dominance of data models challenged the statistical community to embrace algorithmic tools and consider problem-specific characteristics. This resonates in the efforts to incorporate ML techniques into economics, overcoming cultural reliance on data models and navigating the complexities of real-world scenarios. As economists increasingly harness the power of ML to enhance policy analysis, predict outcomes, and unravel intricate relationships, the discipline is moving closer to Breiman's vision of statistics returning to its foundational roots—working with diverse data sources, fostering interdisciplinary collaboration, and addressing complex problems with versatile tools. This transition towards a more dynamic and integrative framework underscores the potential for meaningful progress and innovation in both the realms of statistics and economics.

3.3. Machine Learning's Role in Overcoming Challenges in Macroeconomic Forecasting: Tackling Complexity, Small Sample Sizes, and Rare Events

Although machine learning can offer numerous valuable tools for various purposes in economics, such as benefiting from big data, constructing unconventional data, or utilizing clustering methods, our primary focus lies in the superiority of machine learning's performance in out-of-sample prediction within the field of macroeconomic forecasting.

Machine learning methods hold a significant allure due to their ability to address prominent challenges in the realm of macroeconomic forecasting. To begin with, the dynamics preceding economic crises are inherently intricate. Basic linear or threshold models, although offering intuitive narratives, often encounter difficulties in accurately capturing the depth of these complexities.

Secondly, due to the limited sample sizes in macroeconomic panels, which often contain only a few thousand correlated observations at best, it becomes easy to spot patterns or explanations for past crises that are actually random and won't help predict future crises (Hellwig, 2021). This risk of finding patterns that are specific to the sample being analyzed but don't apply to other samples is called 'overfitting'.

This concern is particularly relevant for widely-used methods among applied econometric methods, such as maximum-likelihood and least-squares. These methods aim to fit the data as closely as possible to generate unbiased estimates of coefficients rather than minimizing prediction errors (Kleinberg et al., 2015). The dilemma between underfitting and overfitting is well-known in predictive modeling: while predicting macroeconomic crises might require complex models, increasing complexity also heightens the risk of overfitting. Machine learning algorithms are adept at striking a balance between these two concerns, making them the preferred choice for various prediction tasks.

Considering the issues mentioned earlier, the accuracy of predictions is mainly associated with the following factors:

i) Methodological Difference: Machine learning (ML) pursues the optimization of the bias-variance trade-off. In ML, the dataset is partitioned into a training sample and a test sample. The model estimates the training sample but uses the test sample for predictions. The optimization of methods focuses on minimizing the test error. A low-test error implies greater generalizability of the model. ML methods prioritize low variance, while simultaneously avoiding high bias, as high bias would result in poor performance on both samples. Therefore, ML is designed to strike a balance between in-sample bias and out-of-sample variance.

ii) ML`s power in dealing with non-linearities and interactions of variables: ML demonstrates an ability to capture non-linearities, interactions, and hidden intricate relationships among variables. This capability allows ML to navigate the complexities of the data more effectively and often derive accurate predictions.

iii) ML`s ability to minimize the problems arose by small sample via various methods such as regularization, hyperparameter tuning and imbalanced data sets due to nature of the rarity of macroeconomic crises.

iv) Ensemble Methods for Enhanced Predictions: Machine learning introduces ensemble methods such as bagging, boosting, bootstrap techniques, and the

aggregation of predictions from various methods. These ensemble approaches harness the collective predictive power of multiple models, leading to improved accuracy and robustness.

3.4. Enhancing Macroeconomic Forecasting Through Machine Learning: Addressing Challenges in Predicting Sudden Stop Crises

In this section, our focus is directed towards Sudden Stop Crises, the potential challenges associated with their forecasting, and how machine learning (ML) can navigate these difficulties to improve out-of-sample predictions.

i. The first benefit of applying ML methods relates to methodological differences mentioned earlier: in-sample error minimization (reducing bias) versus out-of-sample error minimization (achieving low variance) through a balanced approach. This leads to the search for optimal model complexity, including interactions and nonlinear terms. Machine learning (ML) aims to optimize the bias-variance trade-off by partitioning data into training and test samples. The model estimates using training data and predicts with the test data, minimizing test error for greater generalizability. In summary, ML prioritizes low variance and avoids high bias to strike a balance between in-sample bias and out-of-sample variance.

ii. Complexities of Non-linearities and Interactions in Sudden Stop Crises:

In the context of Sudden Stop Crises, the literature reveals the presence of numerous interactions and non-linearities, akin to other types of crises. For instance, Calvo et al. (2008) propose a non-linear impact of financial integration on the likelihood of Sudden Stop Crises. Additionally, Calvo et al. (2008) underscores the significant role of the balance sheet effect in influencing Sudden Stop probabilities.

The balance sheet effect operates as a transmission mechanism, conveying significant and unforeseen reductions in net capital flows from international credit market to the real economy. When combined with the presence of financial frictions,

notably imperfect financial markets and incomplete asset markets, sudden fluctuations in relative prices have disparate impacts on the net worth of assets and liabilities of firms and individuals. This phenomenon becomes more pronounced as the values of individual income, goods prices, and collateral assets decrease, while the burden of debt increases. As a result, the financial amplification mechanism comes into play, setting off a vicious circle of contractions in aggregate demand, occurrences of business insolvency, and an increased stringency in the lending standards enforced by financial institutions. The initial disruption in the relative prices, originating from the external financial market, propagates and affects the real economy.

Calvo et al. (2008) assess the impact of balance sheet effect by using the interaction of two variables. This effect consists of two main components: firstly, large and unexpected net capital flow falls causing the depreciation of the real exchange rate; and secondly, the subsequent contraction of the economy due to financial market imperfections, including credit constraints, non-contingent debt, and liability dollarization in emerging market countries. The second variable involved is liability dollarization. In the probit regression analysis, the balance sheet effect is evaluated through the interaction of liability dollarization and the limited supply of tradable goods relative to domestic absorption of tradables (Current Account Deficit/Domestic Absorption of Tradable goods). The findings highlight that the impact of leveraging the current account deficit is contingent upon the extent of liability dollarization within the economy. The combination of high leverage of the current account deficit with liability dollarization risk poses a considerable probability of experiencing a Sudden Stop Crisis.

Furthermore, Fisherian Sudden Stop Models shed light on the crisis through the lens of the Fisherian amplification mechanism (Mendoza 2002; Mendoza and Korinek 2013; Bianchi 2011; Bianchi and Mendoza 2018, among others). According to these models, when external debt levels are high, a shock to key economic prices, such as the real exchange rate, triggers a chain reaction of declining collateral market values, deteriorating borrowing capability, reduced aggregate consumption, and a

subsequent further decline in the key prices. In essence, the initial decline in the real exchange rate erodes borrowing capacity due to the diminished value of pledged collaterals for obtaining credit, consequently leading to reduced borrowing and aggregate consumption. This decreased aggregate consumption, in turn, deepens the decline in the real exchange rate, exacerbating borrowing constraints and diminishing aggregate consumption even more. To put it simply, at elevated debt levels, a shock to real exchange rates instigates a financial amplification mechanism, setting off a cyclic sequence of falling prices, consumption contraction, and borrowing constraints. This mechanism underscores that Sudden Stops are a product of interactions between numerous financial imperfections and macroeconomic variables, including asymmetric information, moral hazard, endogenous borrowing constraints based on economic conditions, and incomplete asset markets with non-contingent debt. This mismatch, where income or goods and collateral value depend on the state, while debt remains non-state contingent, plays a crucial role in how external shocks from global financial markets affect the local real economy. This phenomenon represents the financial amplification of the balance sheet effect highlighted by Calvo et al. (2008), Calvo and Reinhart (2000), Krugman (1999) and others.

Consequently, it can be inferred from the literature that the complexities of non-linearities and interactions are evident, with the potential for hidden patterns and intricate relationships to be further explored using machine learning techniques.

iii.   Small data problem: Datasets are relatively small in Sudden Stop Crises, just as in macroeconomic forecasting of crises where panel datasets are also limited in size. Furthermore, there could be an additional constraint given that Sudden Stop crises are defined in the aftermath of the Mexican Crisis in 1994.These challenges include the risk of overfitting and the potential for spurious regression, which can ultimately lead to low model generalizability and poor out-of-sample prediction. This is especially relevant since traditional econometrics primarily emphasizes the estimation of unbiased parameters achieved through minimizing loss functions or maximizing likelihoods across the entire sample.

Looking at empirical studies in the context of Sudden Stops using traditional econometrics, we may observe diverse results concerning the influence of variables on the probability of Sudden Stops. For instance, some studies indicate that trade openness has a negative effect, while others show different outcomes. The same variability exists for the impact of exchange rate regimes on the probability. This could suggest that traditional econometric methods have limited generalizability, resulting in models with poor out-of-sample prediction performance.

While this challenge is common to any predictive model, ML models can also be negatively affected by this challenge, similar to traditional econometric models used for predicting Sudden Stop (SS) events. Moreover, due to their complexity, ML models generally require larger datasets to effectively learn intricate relationships within the data. In contrast, simpler models, particularly traditional linear models, may perform better when the dataset is small. It's important to note that ML methods are not entirely immune to these challenges; however, ML models are equipped with tools such as regularization, cross-validation, and hyperparameter tuning to mitigate these issues.

Dealing with small sample sizes is achievable through techniques such as regularization, hyperparameter cross-validation, ensemble methods like bagging and boosting algorithms, as well as stacking various ML techniques and averaging their results to minimize out-of-sample errors. Additionally, generating synthetic data resembling the existing dataset is feasible using methods like the GAN method. A detailed discussion of ML's strategies for addressing small data will be covered in a subsequent section. Furthermore, imputing missing values can also expand the available dataset.

iv. Rare Event: Sudden Stops are infrequent occurrences. Consequently, datasets designed to predict Sudden Stop crises often suffer from imbalanced distribution. An imbalanced dataset emerges when the distribution of classes or outcomes being predicted is skewed, with one class significantly outnumbering the others. Imbalanced datasets can compromise prediction accuracy. In the context of traditional econometrics, an imbalanced dataset

can lead to bias in favor of predicting the majority class. This stems from the fact that many statistical methods, including those used in traditional econometrics, seek to minimize error and naturally prioritize predicting the dominant class due to its greater representation in the data. Consequently, the minority class might be overlooked or misclassified, leading to diminished accuracy and potentially misleading outcomes.

Machine Learning (ML) addresses this issue through various approaches. One method involves optimizing hyperparameters within a specific algorithm through cross-validation. Another approach is modifying the ML model's loss functions by weighting each class contribution based on its occurrence rate. Alternatively, there are resampling techniques available, such as SMOTE (Synthetic Minority Over-sampling Technique). SMOTE is a valuable strategy for handling imbalanced datasets. By generating synthetic instances for the minority class, SMOTE tackles this challenge. It identifies neighboring instances, interpolates their features, and generates new samples. This results in a more balanced dataset, enhancing the model's capacity to capture patterns from the minority class

v.   Variable selection: The ability of machine learning (ML) methods to handle variable selection is particularly useful for predicting Sudden Stop (SS) crises. In the context of macroeconomic forecasting for SS events, there could be numerous variables that might have an impact on the occurrence of these rare events. However, identifying the most relevant variables while avoiding noise and unnecessary complexity is crucial for accurate predictions.

ML methods excel in this aspect by employing techniques like regularization (Lasso penalization, Ridge regression. or elastic net) and principal component analysis (PCA) to effectively select the most important variables. These methods help in reducing the risk of overfitting and enhancing the model's generalizability. By focusing on the most influential variables and their potential forms (lagged, growth, current values), ML methods can create predictive models that capture the essential dynamics leading up to a Sudden Stop crisis.

Traditional econometric methods, on the other hand, might struggle with the curse of dimensionality when dealing with a large number of variables. This could lead to difficulties in identifying the most relevant factors and result in suboptimal predictions. ML's ability to navigate variable selection challenges provides a distinct advantage in constructing robust and accurate predictive models for SS crises, contributing to improved macroeconomic forecasting capabilities.

vi. ML enables the utilization of unconventional data, including textual data. For instance, Sudden Stops (SS) are often triggered by shocks in the international credit market, potentially stemming from the loss of investor confidence in that market. Such events can be influenced by changes in beliefs related to global market conditions, domestic economic fundamentals, political stability, or a country's credibility in monetary and fiscal policies. This loss of confidence can lead to sudden and substantial falls in capital inflows, i.e., Sudden Stops.

Incorporating textual data from sources like Twitter, central bank speeches, or reports from global risk assessment firms can introduce a new variable that influences economic beliefs and, consequently, the probability of SS occurrences. Through Natural Language Processing (NLP), an unsupervised ML technique, unconventional data can be harnessed. This extracted information can serve as an additional variable for predicting Sudden Stop crises. Although not the primary focus of our study, it's worth noting that NLP can be applied in innovative ways.

Furthermore, the monetary policies of major economies, such as decisions made by entities like the U.S. Federal Reserve (FED), can exert considerable influence on smaller economies. For instance, the 'taper tantrum' of 2013 raised concerns about the potential for Sudden Stops. In such cases, sentiment analysis and derived unconventional variables could aid in predicting the ripple effects of major policy decisions on smaller economies. These applications highlight the versatility of ML methods in incorporating unconventional data sources to enhance macroeconomic forecasting accuracy.

70

To recap, Machine Learning methods have emerged as a promising solution to tackle the complex difficulties that come with predicting macroeconomic crisis. These challenges encompass the intricate patterns that come before economic crises, which are often too complex for simple linear models to grasp. Additionally, the limitation of having a small amount of data for analysis in macroeconomic panels adds to the problem. Furthermore, the risks of overfitting, where models become too tailored to specific data, and underfitting, where models fail to capture genuine relationships, add another layer of complexity to the task of prediction.

Machine learning stands out as an alternative to navigate through these challenges, making it a viable approach for various prediction tasks, especially those that involve rare events like financial crises, fiscal challenges, or Sudden Stop crises. Aimed at striking the right balance between bias and variance, machine learning focuses on achieving a broader understanding rather than narrow precision, leading to better predictions when tested on new data. Machine learning is adept at capturing complex relationships, interactions, and hidden trends that traditional methods may fail to identify. This added complexity may come at the cost of overfitting, similar to traditional methods that may suffer from overfitting or spurious regressions, resulting in low model generality and suboptimal out-of-sample predictions, where models become too specific to the training data and perform poorly on new data. However, some machine learning methods, such as cross-validation and hyperparameter tuning, help mitigate the overfitting problem. Furthermore, machine learning tackles the issue of limited data in cases like Sudden Stop Crises, where macroeconomic panel datasets in are notably small. Machine learning methodologies offer viable strategies to address these challenges, involving techniques like regularization, hyperparameter cross-validation, ensemble methods, and the generation of synthetic data. Synthetic data generation techniques, such as Generative Adversarial Networks (GANs), expand the available dataset, improving model robustness.

Additionally, the rarity of events, such as Sudden Stops, further exacerbates the challenge of small datasets. Traditional econometric methods commonly used for Sudden Stop prediction may exhibit bias toward predicting the majority class due to their emphasis on error minimization. Machine Learning (ML) addresses this issue

through various approaches like Synthetic Minority Over-sampling Technique (SMOTE), which create synthetic instances of the minority class, thus enhancing the model's ability to capture patterns from underrepresented classes.

Furthermore, machine learning offers effective variable selection methods, a crucial aspect of managing data dimensionality, a significant concern in empirical studies. Beyond conventional data sources, machine learning embraces unconventional data like textual information, leveraging techniques from natural language processing (NLP) to extract insights from sources such as social media, central bank speeches, and risk assessment reports. While this approach may extend beyond this study here, it's noteworthy that machine learning can repurpose NLP techniques for different applications.

In conclusion, the role of machine learning in macroeconomic forecasting is innovative, providing solutions to the intricate challenges posed by economic complexities, limited data availability, and rare event occurrences.

3.5. Related Literature

Hellwig (2021) notes that despite macroeconomic crises prediction having a longstanding history, as evidenced by works such as Frank and Cline (1971), it wasn't until the aftermath of the late 1990s Asian crises that the field of early warning systems (EWS) underwent a significant resurgence. Early studies often centered on currency and financial crises.

Empirical investigations in forecasting macroeconomic crises can be broadly categorized into two main streams:

i) Limited Dependent Variable Regression Models/Parametric Approach:

This category employs techniques like logit or probit regression, which fall under the umbrella of Generalized Linear Models (GLM). Focusing on Sudden Stop Crises, the aim is not necessarily to create an EWS, but rather to comprehend the determinants

of such crises. Numerous papers have explored this avenue, including works by Edwards (2004), Calvo et al. (2008), Eichengreen and Gupta (2006, 2016), Cavallo and Frankel (2008), and Forbes and Warnock (2012, 2021).

ii) Non-parametric Methods:

a) Signaling Approach: Popularized by Kaminsky et al. (1998) and Kaminsky and Reinhart (1999), the signaling approach involves issuing a warning signal when an indicator surpasses a threshold based on its own distribution percentile. This method aims to optimize variable thresholds for signaling crises in the subsequent 24 months, ranking variables based on their noise-to-signal ratios. This approach is akin to a concise classification tree, as noted by Hellwig (2021). Noteworthy examples of this approach include Kaminsky et al. (1998) for currency crises, Borio and Lowe (2002, 2004), and Borio and Drehmann (2009a, 2009b) for banking crises, and Alessi and Detken (2011) for asset price boom-bust cycles. The choice between signaling and discrete choice models, according to Alessi and Detken (2011), depends on the expected non-linearity between indicators and event variables.

b) Tree-based Methods: Employing techniques like Decision Trees or the CART method developed by Breiman (1984), Binary Recursive Trees are utilized for macroeconomic crisis prediction. Examples include Duttagupta and Cashin (2008) for banking crises, Manessa et al. (2003), Manessa and Rubini (2009) for sovereign crises. Random Forest developed by Breiman(2001), is also employed in studies such as Jarmulska (2020) for fiscal crises, Joy et al. (2015) for banking and currency crises, Alessi and Detken (2018) for banking crises, and Savona et al. (2015), Savona and Vezoli (2013) for sovereign debt crises.

Although Machine Learning techniques have been increasingly recognized for their enhanced predictive accuracy and robustness in the fields of econometrics and economics and finance, their application to predicting rare economic events, such as fiscal crises, banking crises, and financial crises, remains limited. Only a handful of studies have ventured into the incorporation of ML algorithms for such predictive purposes.

In the realm of Sudden Stop Crises, the available empirical research primarily hinges on two main methodologies: the utilization of limited dependent variable techniques (Generalized Linear Models such as probit or logit) as demonstrated in the aforementioned studies; or the implementation of event studies, exemplified by the works of Mendoza and Korinek (2013), Bianchi (2011), Bianchi and Mendoza (2018), among others. As of our current understanding, there is a notable absence of studies employing Machine Learning (ML) approaches to investigate Sudden Stop Crises.

Diverging from prior studies that solely employed Random Forests to address different crises, our focus is on employing and comparing a diverse range of powerful Machine Learning (ML) techniques, including XG Boost. In this regard, our research shares similarities with the works of Hellwig (2021), Bluwstein et al. (2023), and Coffinet and Kien (2019):

Hellwig (2021) applies Elastic Net, Random Forests, Support Vector Machines (SVM), and Gradient Boosted Trees to predict fiscal crises out of sample data. Similarly, Bluwstein et al. (2023) harness a variety of ML algorithms, including SVM, Random Forests, and Gradient Boosting, to forecast financial crises. Their approach involves the creation of early warning models through the application of machine learning techniques to macrofinancial data spanning 17 countries from 1870 to 2016. Notably, nonlinear ML models demonstrate superior out-of-sample predictive capabilities compared to logistic regression. The authors extensively compare various machine learning models, such as decision trees, random forests, extremely randomized trees, support vector machines (SVM), and artificial neural networks, against logistic regression. Their findings highlight the robust predictive power of most machine learning models, outperforming logistic regression except for individual decision trees.

Coffinet and Kien (2019) propose a machine learning toolkit designed to detect rare events, particularly banking crises. They incorporate multiple algorithms, including Random Forest, Gradient Boosting, and Deep Neural Networks, to construct the toolkit, which demonstrate enhanced performance in detecting rare events compared

to traditional econometric models. By applying the toolkit to predict banking crises, the authors achieve a high level of accuracy. This study provides valuable insights into the application of machine learning for predicting infrequent events and underscores the advantage of combining multiple algorithms over conventional econometric models.

Ultimately, the research highlights the potential of machine learning to enhance the ability to predict significant rare events, such as financial crises. Consequently, our study aligns with these works in terms of adopting diverse ML methods and evaluating their out-of-sample prediction power.

3.6. Predicting Sudden Stops Using Supervised Machine Learning

Supervised learning stands as a cornerstone of modern predictive analytics, empowering us to make informed decisions by uncovering patterns within data. At its core, supervised learning is a machine learning paradigm in which we train a model on a labeled dataset, enabling it to learn the relationships between input features and corresponding target labels. These relationships are then used to make predictions on new, unseen data.

3.6.1. Applying Supervised Learning to Predicting Sudden Stops

The application of supervised learning to predicting Sudden Stops involves transforming the problem into a binary classification task. Sudden Stops are categorized as either "Sudden Stop" or "No Sudden Stop" events based on historical data and relevant features. These features could include country-specific variables and global variables. Each instance is labeled according to whether a Sudden Stop occurred during the specified period.

By casting the Sudden Stop prediction problem into a binary classification framework, we enable supervised learning algorithms to identify patterns within the data that distinguish between instances of Sudden Stops and instances of non-Sudden Stops. The goal is to build a model that can generalize from the labeled training data and accurately predict whether an event will lead to a Sudden Stop or not.

3.6.1.1. Labeling Variables and Train-Test Split

To enable the supervised learning process, we require a dataset with labeled instances. In the context of predicting Sudden Stops, this involves historical data where each data point includes a set of features and a corresponding label indicating whether a Sudden Stop occurred during that period. These labels serve as ground truth, allowing the model to learn the relationships between features and outcomes.

To evaluate the model's effectiveness, we partition the dataset into two subsets: the training set and the testing set. The training set is used to teach the model patterns and relationships present in the data, while the testing set is kept separate and unseen during training. This enables us to assess the model's ability to generalize its learned patterns to new, unseen data, thereby gauging its predictive power for future Sudden Stop events.

3.6.1.2. Enhancing Model Performance Beyond Training Data

Upon splitting the dataset into training and testing subsets, the pursuit of accurate predictions evolves into a quest for broader generalization. The emphasis shifts from minimizing errors within the training sample to achieving low errors on unseen data. This transition underscores the model's adaptability to diverse contexts, crucial for predicting Sudden Stops (SS).

Central to this phase is the concept of optimizing the model's predictive power. By leveraging the training set, the model immerses itself in the intricacies of economic indicators, variables tied to SS. With each iteration, the model refines its internal parameters, learning to decipher patterns that correlate with Sudden Stop occurrences. The process mirrors a musician refining their technique, enhancing predictive nuances over time.

However, the model's true capability shines when it encounters unfamiliar, unseen data. This two-fold process, involving both cross-validation and hyperparameter tuning, serves to strengthen and enhance the model's ability to adapt.

◈ **Amplifying Generalization through Cross-Validation and Hyperparameter Tuning:**

Cross-Validation and K-Fold Technique:

Much like stress-testing a hypothesis from multiple angles, cross-validation validates the model's robustness. In the K-fold technique, the training data is divided into K subsets, or 'folds.' The model is then trained K times, using each fold as a validation set once. This process rigorously examines the model's adaptability to diverse data distributions, effectively simulating the uncertainties of real-world scenarios. The K-fold approach significantly impacts the model by preventing it from becoming overly tailored to the characteristics of a single training subset. This ensures the model's generalization to various situations, enhancing its reliability and performance.

Hyperparameter Tuning and Bias-Variance Tradeoff:

Hyperparameters orchestrate the learning process. Think of them as conductor's batons that shape the symphony of model training. Hyperparameter tuning involves adjusting these settings to fine-tune the model's performance. For instance, in the case of decision trees, one crucial hyperparameter is the depth of the tree. A shallow tree might oversimplify the model (high bias), while an overly deep tree might lead to overfitting (high variance). Hyperparameter tuning strikes a balance—a depth that captures nuanced patterns without succumbing to noise.

Another prime example of hyperparameter tuning's importance lies in the realm of regularization. Regularization is a technique crucial for balancing model complexity and curbing overfitting tendencies. Within this context, one pivotal hyperparameter takes center stage—'alpha'. This hyperparameter controls the degree of regularization applied to the model.

To illustrate, consider the scenario of predicting Sudden Stops using a logistic regression model. In this case, hyperparameter tuning involves adjusting the 'alpha' value to find the optimal point of regularization. A higher 'alpha' value tightens the

reins on the model's coefficients, allowing it to generalize its insights beyond the training dataset.

However, navigating the world of 'alpha' isn't without its challenges. Setting 'alpha' too high might lead to underfitting—a situation where the model fails to capture critical patterns associated with Sudden Stops. Conversely, setting 'alpha' too low could result in overfitting, where the model becomes overly sensitive to the noise in the training data and struggles to generalize to new data.

Striking the right equilibrium is the key. Hyperparameter tuning for 'alpha' aims to harness the benefits of regularization while avoiding the pitfalls of underfitting and overfitting. By attaining this delicate balance, the model becomes an effective tool for predicting Sudden Stops, capable of extrapolating insights from the training data to real-world scenarios. This meticulous tuning process underscores the intersection of model complexity, overfitting, and predictive precision in the pursuit of enhanced accuracy.

In our study, each machine learning method we employ comes with its own set of hyperparameters that significantly influence the model's performance. To ensure optimal results, we utilize cross-validation techniques to fine-tune these hyperparameters for each method. This iterative process enables us to strike the right balance between model complexity and predictive accuracy, enhancing the models' ability to anticipate Sudden Stops effectively.

3.6.1.3. Types of Errors, Relevance in Sudden Stops, and Model Comparison

In the realm of predictive modeling, comprehending the types of errors is a fundamental endeavor. In the context of binary classification, two prominent types of errors emerge: Type 1 Error and Type 2 Error.

Type 1 Error (False Positive) occurs when the model predicts a positive outcome that does not materialize in reality. When dealing with Sudden Stops (SS), a Type 1 Error could involve predicting a forthcoming SS that doesn't actually occur.

Type 2 Error (False Negative), conversely, a Type 2 Error takes place when the model fails to predict a positive outcome that does, in fact, occur. In the realm of SS, a Type 2 Error could involve the model overlooking the signs of a forthcoming crisis.

The effects of these errors during Sudden Stops have significant implications for the economy and hold importance for decision-makers and stakeholders alike. A Type 1 Error, often termed a "False Positive," has the potential to create waves across financial markets, economic institutions, and public perception. Beyond affecting statistics, its consequences can translate into disruptive shifts in market sentiment, investor confidence, and consumer behavior.

When the model commits a Type 1 Error by wrongly predicting a Sudden Stop that doesn't occur, a chain of consequences may be set in motion. The unnecessary alarms could lead to abrupt economic restrictions and investor panic. Hastily made policy choices may trigger a cascading effect that destabilizes the situation. This might lead to reduced economic activity, capital flight, and increased overall uncertainty. Ironically, the measures intended to prevent issues might inadvertently intensify volatility.

On the other hand, a Type 2 Error, called a "False Negative," has significant effects on the economy's stability. If the model misses a real Sudden Stop because it didn't predict it, the results are noticeable. This error, marked by lack of readiness and missed chances for intervention, can significantly worsen the effects of a developing crisis. The absence of timely action and decisive policy measures could escalate an already critical situation into a full-blown economic turmoil, leading to extended periods of recession, diminished investor confidence, and even sovereign debt distress.

Both types of errors have an impact on how the economy progresses, stressing the value of predictive models and their application. These errors show the need to find a balance between false alarms and missed opportunities. The details of Type 1 and Type 2 Errors emphasize the importance of accurate predictions for addressing

Sudden Stops. This urges caution in model creation, comparison, and policy implementation.

3.6.1.4. Performance Metrics and Model Selection

Model selection involves comparing different models using relevant performance metrics. The choice of metrics depends on policy goals and the error types one aims to minimize.

Performance metrics play a pivotal role in predictive modeling and data analysis, serving as essential tools to quantify and evaluate the effectiveness of models in making accurate predictions. These metrics provide a structured and quantifiable way to assess how well a model's outputs align with real-world outcomes, enabling data scientists, researchers, and policymakers to make informed decisions and refine their approaches.

Performance metrics are particularly crucial in the context of binary classification problems, where the goal is to categorize data instances into two distinct classes. Whether it's predicting medical diagnoses, financial market trends, or, as in our case, anticipating economic Sudden Stops, the ability to measure the accuracy of predictions is vital for effective decision-making.

By employing a range of performance metrics, we can systematically evaluate a model's strengths and weaknesses, identify the types of errors it might make, and gauge its ability to generalize to new, unseen data. These metrics provide a common language for quantifying the model's performance, enabling comparisons between different algorithms, hyperparameters, and methodologies.

- **Confusion Matrix:**

In classification modeling, the confusion matrix becomes a strong tool that goes beyond just measuring accuracy. It offers a comprehensive and structured way to understand the performance of a model in the context of binary classification tasks. The confusion matrix provides a visual breakdown of predicted outcomes versus actual outcomes, shedding light on both correct and erroneous predictions.

At its core, the confusion matrix serves as a guidepost for evaluating the efficacy of predictive models. By organizing the outcomes into categories such as True Positives, True Negatives, False Positives, and False Negatives, it enables us to quantify how well a model distinguishes between different classes. This breakdown is pivotal in understanding the strengths and weaknesses of the model, empowering data analysts, researchers, and policymakers to make more informed decisions.

The importance of the confusion matrix lies in its ability to highlight not just overall predictive accuracy, but the specific types of errors a model might commit. This insight is particularly crucial when different types of errors hold varying real-world consequences. Whether it's avoiding unwarranted alarms or ensuring timely interventions, the confusion matrix equips us with a sharper understanding of a model's behavior.

Confusion matrix for a binary classification prediction:

Table 3. 1. Confusion Matrix

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

The components of the confusion matrix are as the following:

True Positives (TP): True Positives are instances where the model correctly predicts the positive class, and the actual outcome is also positive. In the context of Sudden Stops, a true positive signifies that the model accurately identifies an impending economic crisis.

True Negatives (TN): True Negatives are cases where the model correctly predicts the negative class, and the actual outcome is also negative. For Sudden Stops, a true negative indicates that the model accurately predicts the absence of an economic crisis.

False Positives (FP): False Positives occur when the model incorrectly predicts the positive class, but the actual outcome is negative. This corresponds to predicting a Sudden Stop that doesn't occur (Type 1 Error).

False Negatives (FN): False Negatives involve incorrect negative predictions, where the model wrongly predicts the negative class, but the actual outcome is positive. This aligns with failing to predict an actual Sudden Stop (Type 2 Error).

These components form the cornerstone of the confusion matrix, providing a structured framework to assess a model's predictive accuracy and error tendencies. Understanding these terms is crucial to interpreting the outcomes of the matrix and gaining insights into the model's performance, particularly in contexts such as predicting economic crises like Sudden Stops.

3.6.1.5. Common Performance Metrics

- *Accuracy:*

Accuracy measures the proportion of correctly predicted instances to the total number of instances. (TP+TN/TP+TN+FP+FN)

- *Precision (Positive Predictive Value):*

Precision quantifies the ratio of correctly predicted positive instances to the total instances predicted as positive. (TP/TP+FP)

- *Recall (Sensitivity, True Positive Rate):*

Recall calculates the proportion of correctly predicted positive instances to the total actual positive instances. (TP/TP+FN)

- *Specificity (True Negative Rate):*

Specificity computes the ratio of correctly predicted negative instances to the total actual negative instances. (TN/TN+FP)

- *F1-Score:*

The F1-Score is the harmonic mean of precision and recall, offering a balanced measure of accuracy.

- *Area Under the Receiver Operating Characteristic Curve (AUC-ROC):*
AUC-ROC quantifies the model's ability to distinguish between classes across various probability thresholds. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity).

- *Precision-Recall Curve:*
The Precision-Recall Curve illustrates the trade-off between precision and recall as the classification threshold changes.

3.6.2. Summary and Navigating Model Performance Metrics in the Context of Sudden Stops

In the complex realm of predictive modeling, grasping performance metrics extends beyond numbers. This holds particular significance for rare occurrences like Sudden Stops, where relying solely on accuracy isn't sufficient. These metrics act as threads weaving a detailed image of a model's performance, guiding decisions and enhancing its efficiency.

At the core of these metrics lies deeper significance. Precision entails more than just accurate positives; it's about avoiding wrong positive predictions. Similarly, recall isn't solely about capturing positives; it also minimizes missed positive cases. These metrics unveil how well the model balances caution and optimism, showcasing its true strengths.

These metrics are closely tied to the types of errors – Type 1 and Type 2. Optimizing isn't a uniform approach; it's a delicate equilibrium. Prioritizing precision can prevent unnecessary alerts, while focusing on recall better prepare us for crises. The choice hinges on the situation and the balance between error types.

For rare events like Sudden Stops, traditional accuracy can mislead. The infrequency of crises affects accuracy calculations. The Area Under the ROC Curve , recall, and f1 metrics are more appropriate in our problem of Sudden Stop prediction.

Ultimately, these metrics serve as compasses, guiding model optimization. Beyond evaluation tools, they steer strategic choices. The chosen path mirrors the interplay of error types and outcomes. In the domain of Sudden Stops, these metrics illuminate the way forward, aiding models in navigating the uncertainties of economic crises.

CHAPTER 4

IMPLEMENTATION OF ML METHODS FOR SUDDEN STOP CRISIS PREDICTION WITH A LIMITED SET OF PRE-SELECTED VARIABLES

In this section, our primary aim is to conduct a comparative analysis of the out-of-sample performance of Sudden Stop prediction among several Machine Learning methods and a conventional statistical approach, specifically the complementary log-log method (cloglog), within the Generalized Linear Model (GLM) framework. The GLM, encompassing techniques such as logistic regression, probit, and cloglog, is widely utilized for predicting macroeconomic crises. In this study, we designate it as our representative traditional statistical approach. The foundational model proposed by Forbes and Warnock (2021) for predicting Sudden Stop crises serves as our benchmark. While their analysis comprehensively addresses four distinct extreme capital events—surges, stops, retrenchment, and flight—our specific focus centers on the segments associated with stop events.

To commence, we utilize the capital flow dataset curated by Forbes and Warnock (2021) to identify Sudden Stop Crises in 59 countries, encompassing both advanced and emerging nations, spanning the period from 1978 Q1 to 2020 Q3. Following this, we replicate the estimation process for the base case as presented in their analysis, employing the complementary log-log model, and proceed to scrutinize its out-of-sample performance, establishing it as our baseline scenario. Subsequently, employing the same dataset, we implement a range of supervised Machine Learning methods and conduct a comparative analysis of their respective out-of-sample performances.

In the subsections of this section, we first elucidate the definition of a Sudden Stop and provide an overview of the identification of Sudden Stops following the methodology outlined by Forbes and Warnock (2012; 2021). Next, we present the

replication of the base scenario from their paper. Following that, we detail the Machine Learning estimation strategy. Finally, we conduct a comparative analysis between these models and the traditional GLM method, specifically the complementary log-log model.

4.1. Sudden Stop Definition and Identification

Following Forbes and Warnock (2021), we employ their Sudden Stop identification methodology, initially developed in Forbes and Warnock (2012). This methodology diverges from the conventional approach, as established *by Calvo, Izquierdo, and Mejía (2004)*, by utilizing quarterly gross capital flows rather than relying on net-flows. Unlike the latter, which depends on current-account-based proxies for flows, Forbes and Warnock(2012; 2021)'s method utilizes actual flow data. In their definition, a Sudden Stop is characterized by a sharp decrease in gross capital inflows from foreign sources. To identify Sudden Stop episodes, we compute year-over-year changes in four-quarter gross capital inflows. Three criteria are applied to define these episodes:

   i.   The current year-over-year change in four-quarter gross capital inflows is more than two standard deviations below the historic (5-year moving) average during at least one quarter of the episode.

   ii.   The episode persists for all consecutive quarters in which the year-over-year change in annual gross capital flows is more than one standard deviation below the historical average.

  iii.   The episode's duration exceeds one quarter.

Let C_t be the 4-quarter moving sum of gross capital inflows (GINFLOW) and compute annual year-over-year changes in C_t:

$$C_t = \sum_{i=1}^{3} \text{GINFLOW}_{t-1} \text{ with t=1,2,3,......,N and} \tag{4.1}$$

$$\Delta C_t = C_t - C_{t-4} \text{ with t=5,6,.,N.} \tag{4.2}$$

Next, we compute rolling means and standard deviations of C_t over the last 5 years. A stop episode is a period when gross inflows fall one standard deviation below its mean, provided it reaches two standard deviations below at some point. The episode ends when gross inflows are no longer at least one standard deviation below its 5-year rolling mean.

In **Figure 4.1,** we present an illustrative example of the identification of Sudden Stop crises by using Turkey's Gross Capital Inflow spanning from the second quarter of 1978 (Q2 1978) to the third quarter of 2020 (Q3 2020). As depicted in **Figure 4.1**, Turkey experienced several episodes of sudden stops in its gross capital inflows, marked by disruptions occurring across quarters. The first sudden stop occurred from the third quarter of 1991 (Q3 1991) to the fourth quarter of the same year (Q4 1991). Subsequently, another episode extended from the second quarter of 1994 (Q2 1994) to the first quarter of 1995 (Q1 1995). In the early 2000s, Turkey faced a sudden stop from the first quarter of 2001 (Q1 2001) to the fourth quarter of the same year (Q4 2001). Notably, there was a sudden stop from the fourth quarter of 2007 (Q4 2007) to the first quarter of 2008 (Q1 2008), and another significant episode from the fourth quarter of 2008 (Q4 2008) to the fourth quarter of 2009 (Q4 2009).
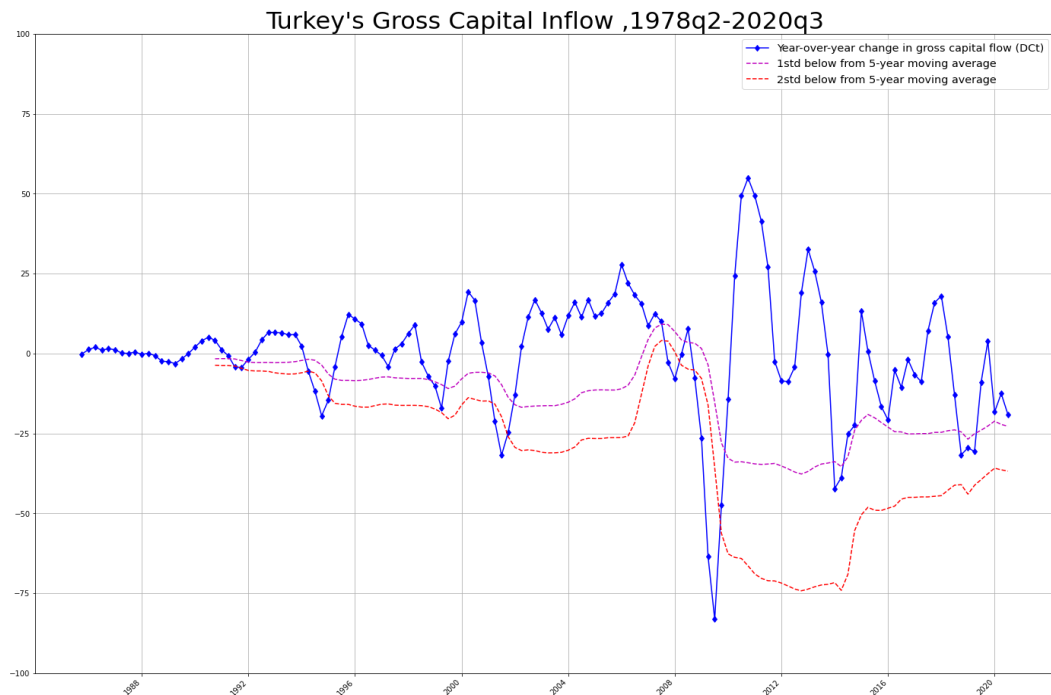
Figure 4. 1. Turkey`s Gross Capital Inflow, 1978q2-2020q3

The comprehensive analysis of Sudden Stops (SS) across 59 countries, based on the dataset constructed by Forbes and Warnock(2021), yields significant insights into the global economic landscape. **Figure 4.2** highlights the widespread impact of SS events, with Argentina standing out as the most affected. Intriguingly, the presence of advanced economies such as Norway, Finland, and Denmark in the top 10 underscores the varied susceptibility of nations to economic disruptions.
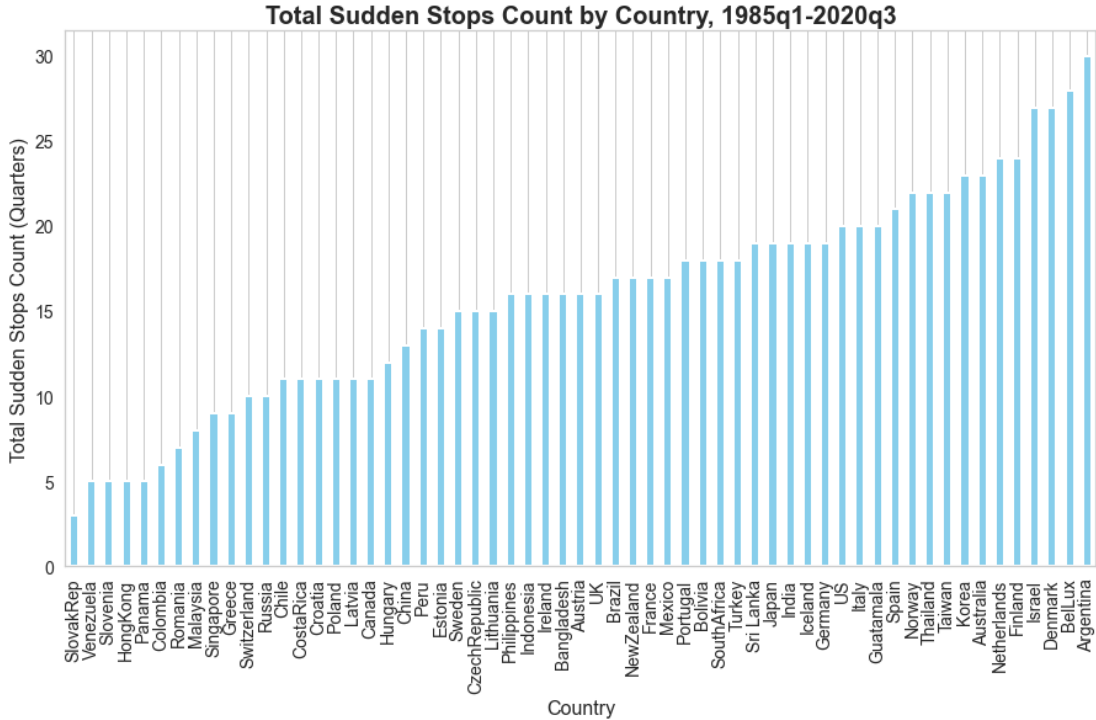
Figure 4. 2. Total Sudden Stop Quarters of Countries in the Sample, 1985q1-2020q3

The dataset reveals a substantial 930 Sudden Stop Crises events during this extensive period. When examining the temporal distribution, we observe a consistent frequency of these events. Before 2008, there were 458 instances of Sudden Stops, and after 2008 until the end of 2020 Q3, the frequency stands at 472. Despite the temporal transition, this consistency suggests a continuous susceptibility to economic disruptions, warranting further investigation into the underlying factors contributing to this persistent trend. As shown in the Figure 4.3, in the aftermath of the Global Financial Crisis, between 2013 and 2016, we observe an increase in Sudden Stops, which may be attributed to expectations of 'tapering' and 'normalization' policies by the Federal Reserve in the United States.
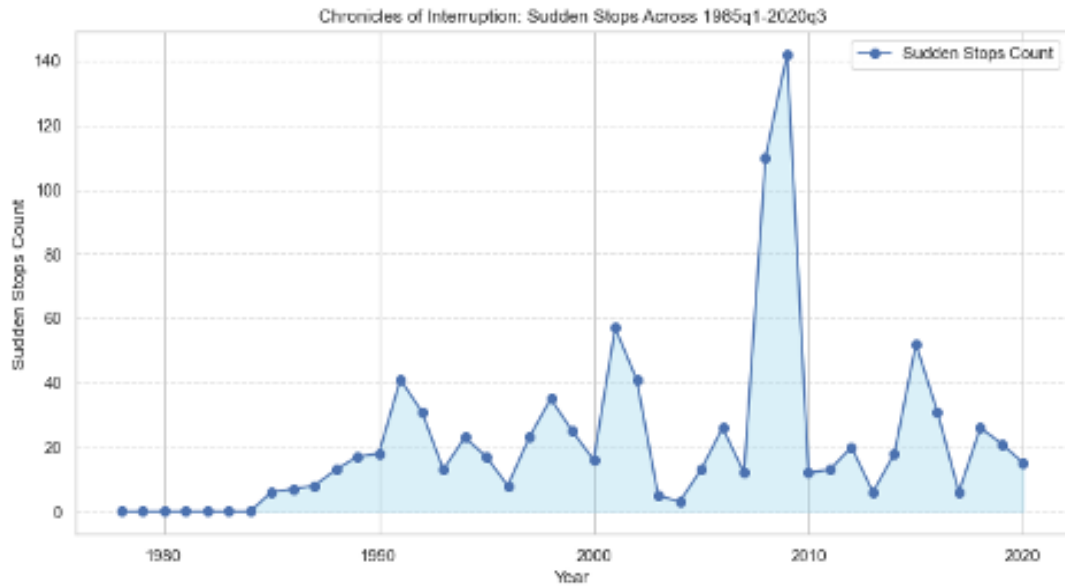
Figure 4. 3. The Occurrence of Sudden Stops (SS) Spanning From 1985q1 to 2020q3

Note: Each data point on the graph represents the total number of quarters with Sudden Stop events for each year, reflecting the total count of SS events for all countries.

The dataset discloses a significant number of Sudden Stop events during the Global Financial Crisis (GFC) period from 2008 to 2010, totaling 264 instances across various countries. **Figure 4.4** provides a visual representation of the Sudden Stop crisis during the GFC, showcasing the impact on different nations.

Notably, Norway experienced the highest number of Sudden Stops during this challenging period, enduring for 8 quarters. Following closely are Romania, France, Argentina, Turkey, and Spain, each facing 7 quarters of Sudden Stops. This distribution underscores the widespread consequences of the GFC across diverse regions and economies.

Moreover, the analysis extends to the post-2010 period, revealing a continued occurrence of Sudden Stops. The dataset indicates 208 instances after 2010, highlighting economic vulnerabilities persisted even beyond the immediate aftermath of the GFC. The examination of these post-2010 events contributes to a comprehensive understanding of the enduring challenges faced by various countries in the aftermath of the global economic downturn.
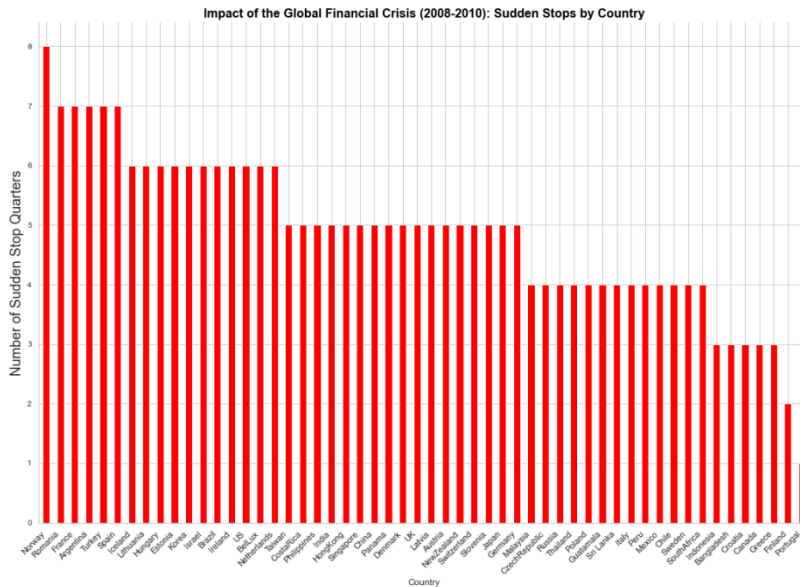
Figure 4. 4. Impact of the Global Financial Crisis (2008-2010)

In **Figure 4.5**, we analyze the distribution of Sudden Stop (SS) quarters and associated countries between 2011 and 2020. Following 2010, there were a total of 208 Sudden Stops, indicating a sustained vulnerability to economic disruptions during this period. Notably, within the top 10 affected countries, we observe a mix of both advanced and emerging countries, including the Netherlands and China.
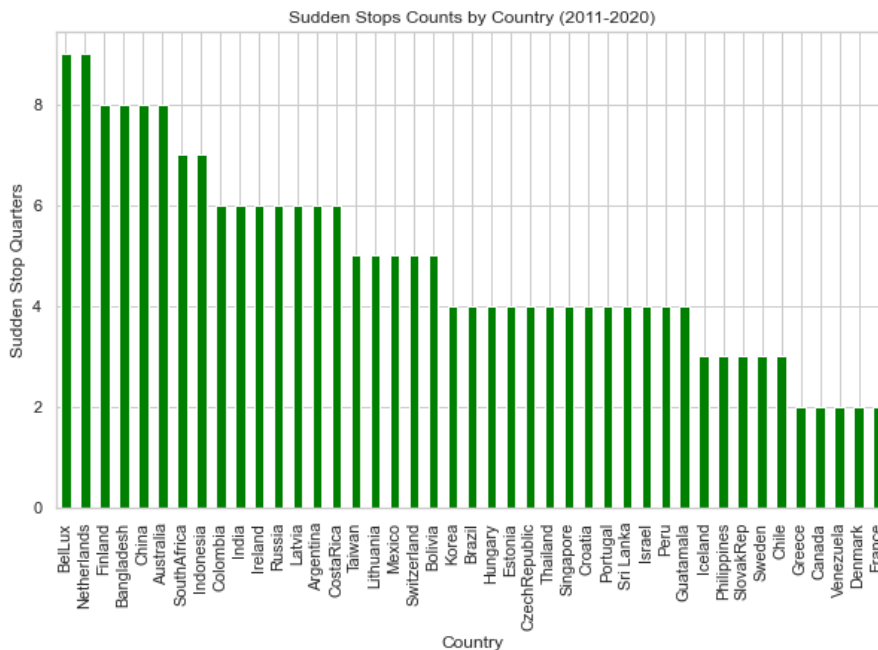
Figure 4. 5. Sudden Stops After the GFC, 2011q1-2020q3

After identifying instances of Sudden Stops and conducting a comprehensive examination of these events, the ensuing section focuses on replicating the base model. In-sample estimation is a common practice; however, we assess the out-of-sample prediction performance of the model, in order to gauge its generalizability.

4.2. The Base Model: Forbes& Warnock(2021)

Forbes&Warnock (2021) assess the role of global, contagion, and domestic variables on the conditional probability of having a surge, stop, flight, retrenchment episode each quarter and estimate the model:

$$Prob \; (e_{it} = 1) = F \left(\phi_{t-1}^{Global} \, B_G + \phi_{i, \; t-1}^{Contagion} \, B_C + \phi_{i, \; t-1,}^{Domestic} \, B_D \right) \qquad (4.3)$$

,where e_{it} is an episode dummy variable that takes the value of 1 if country i experiencing an episode in quarter t ; Φ^{Global}_{t-1} is a vector of global factors lagged by one quarter; $\Phi^{Domestic}_{i,t-1}$ is a vector of domestic variables lagged by one quarter; $\Phi^{Contagion}_{I,t-1}$ is a vector of contagion variables lagged by one quarter. They estimate the equation (3) using the complementary logarithmic (or cloglog) framework, which assumes that F(.) is the cumulative distribution function (cdf) of the extreme value distribution. The estimation strategy assumes that :

$$F(z) = 1 - \exp[- exp(z)] \qquad\qquad (4.4)$$

They initially estimate the model using only six variables: global liquidity, long-term interest rates, VXO (global risk), year-over-year global GDP growth from the IMF's World Economic Outlook dataset (global growth), a dummy variable equal to one if a country in the same region has the same type of episode (to capture contagion through geographical proximity), and local real GDP growth. Global risk represents the year-over-year change in the VXO. Global liquidity is the average of year-over-year percentage change in the broad money supply of the US, UK, Euro Area, and Japan. Long-run interest rates are the average long run interest rates for the US, UK, Euro Area, and Japan.

Their focus is on understanding how the impact of these variables has changed with respect to all types of events, namely surge, stop, flight, and retrenchment, since the Global Financial Crisis (GFC). They also make a comparison of the results of their own paper previously published by Forbes & Warnock (2012), extending the time horizon. However, our focus is solely on stop episodes, encompassing the entire sample, as our objective is to compare the out-of-sample prediction performance of different models. Initially, we replicated their model for the complete sample related to stop episodes. The outcomes of this replication, representing in-sample predictions, are presented below:

Table 4. 1. Complementary Log -Log (Cloglog) Estimation For 1986-2018

	COEFFICIENT	STANDARD DEVIATION	Z	P>|Z|	CONFIDENCE INTERVALS [0.025,0.975]
CONTAGION	0.7386	0.164	4.509	0.000	[0.418,1.060]
LONG-RUN INTEREST RATE	0.1003	0.026	3.913	0.000	[0.050,0.151]
GLOBAL GROWTH	-0.1271	0.045	-2.832	0.005	[-0.215,-0.039]
RISK	0.0457	0.006	7.609	0.000	[0.034,0.058]
LIQUIDITY	-0.0049	0.010	-0.482	0.630	[-0.025, 0.015]
REAL GDP GROWTH	-0.087	0.014	-6.286	0.000	[-0.115,-0.060]
OBS.	**4644**				

Notes: The dependent variable is a 0-1 variable indicating if there is a Sudden Stop episode. Estimates are obtained by clog log model, and we used robust standard errors clustered by country. The countries in the sample are added to the Appendix A.

The estimation results highlight the contagion variable as the most pivotal factor influencing the likelihood of a Sudden Stop episode. A one-unit increase in the contagion variable is associated with a substantial 2.09-fold rise in the odds of experiencing a Sudden Stop episode, underscoring its significant impact on the event's probability. Similarly, a one-unit increase in the long-run interest rate is linked to a notable 1.11 times increase in the odds of encountering a Sudden Stop episode. Conversely, an increase in global growth by one unit is associated with 0.88

times decrease in the odds of a Sudden Stop episode, reflecting its mitigating effect. The risk variable, capturing the year-over-year change in VXO, demonstrates 1.05 times increase in the odds of a Sudden Stop episode with a one-unit increment. Furthermore, a one-unit increase in real GDP growth is correlated with 0.92 times decrease in the odds of facing a Sudden Stop episode. All these variables exhibit statistical significance at the 95% confidence level. However, the liquidity variable, with a coefficient of -0.0049, does not attain statistical significance, as indicated by a relatively high p-value of 0.630.

Out- of -Sample Prediction Performance of the Clog log Model:

We assess the out-of-sample prediction performance of the base model using the AUC-ROC curve and confusion matrix. The dataset is partitioned into training and test sets, with 80% of the data allocated for training (in-sample prediction). The AUC-ROC curve, depicted in Figure 4.6, attains a value of 0.73, indicating the model's ability in distinguishing between classes. ROC curve serves as a benchmark for evaluating the model's discriminatory power. It represents the trade-off between the true positive rate and the false positive rate. We can pick an operating point on this curve by specifying the maximum allowable FPR (e.g., 0.2) and read the TPR (e.g., 0.52) as the detection performance of the model. Furthermore, the AUC of ROC is a commonly used measure of accuracy to evaluate and compare the performance of various classification models. A higher AUC-ROC value, closer to 1, signifies superior model performance.
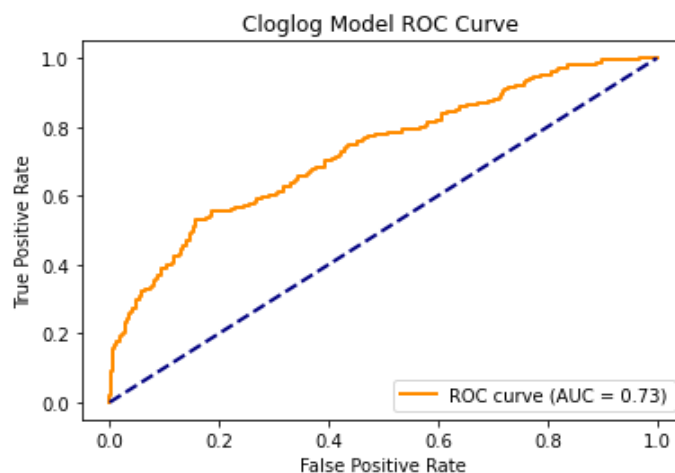
Figure 4. 6. AUC_ROC Curve of The Base Model

Transitioning to Figure 4.7, we delve into the confusion matrix, a tool employed to assess the model's classification performance. Simply put, the confusion matrix dissects the model's predictions into four categories: true positives, true negatives, false positives, and false negatives. In this specific context, the model fails to identify 141 true Sudden Stop crises out of the 168 Sudden Stop events that occurred (Type II error). Conversely, it exhibits a low number of false positive cases, accurately identifying only 9 cases out of 993 true non-crisis instances (Type I error).
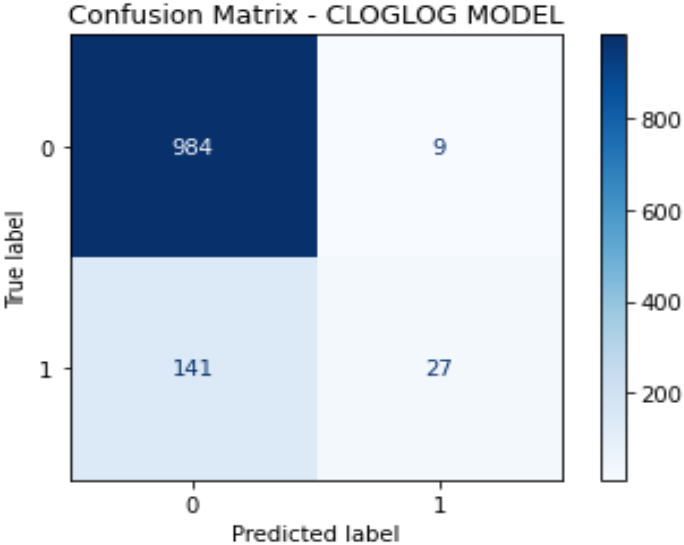
Figure 4. 7. Confusion Matrix of the Base Model

Hereafter we will use the ROC curve, AUC score and CM to evaluate and compare the out-of-sample performance of various ML algorithms.

With these evaluations in mind, we now shift our focus to the implementation of a set of Machine Learning Algorithms. Upon concluding this chapter, we will undertake a comparative analysis of the out-of-sample performance exhibited by these ML algorithms in comparison to the baseline model.

4.3. Machine Learning Methods and The Estimation Strategy

We employed the identical dataset as that of the base model and implemented various Machine Learning (ML) algorithms. The task was formulated as a binary

classification, with the target variable identified as Sudden Stops. Initially, we considered a traditional approach by partitioning the data into training and test sets, preserving a ratio of 0.8 for training and 0.2 for the test set. However, to obtain an objective out of sample performance from our relatively small size data (~4000 data points) we adopted a cross-validation strategy, where the data is divided into independent K-folds (e.g., 10), and each time the model is trained on K-1 folds and tested on the remaining one-fold .[5] The model performance is then evaluated by averaging the performance across all the folds. This procedure, commonly used strategy to deal with small size datasets, ensures that all the available data is used for training and testing while preserving the independence of the training and test datasets. Furthermore, because Sudden Stop crises are rare events, our dataset is heavily imbalanced. We then extended our cross-validation to "stratified k-fold cross -validation" where each fold maintains the class balance (the ratios of SS class to negative class).

In a standard k-fold cross-validation, data is randomly partitioned into folds. However, in the case of imbalanced datasets where one class significantly outnumbers the other (e.g., a rare event like a Sudden Stop Crisis), standard k-fold cross validation may lead to some folds having an insufficient representation of the minority class.[6]

Stratified k-fold addresses this issue by ensuring that each fold maintains the same class distribution as the entire dataset. In other words, it preserves the proportion of different classes within each fold, making it particularly beneficial when dealing with an imbalanced dataset.

---

[5] k-fold cross-validation is a common method used for splitting data into training and test sets to compare methods in terms of out-of-sample prediction, i.e., generalizability. For instance, in a macroeconomic context, Alessi and Detken (2018) employ k-fold cross-validation to compare the out-of-sample prediction power of traditional logit method and Random Forest in predicting banking crises.

[6] Within each observation, temporal order is maintained by lagging the exogenous variables by one quarter. This technique guarantees that the model is trained on historical data preceding the target period and tested on future data, effectively simulating a time-split validation approach, despite not explicitly utilizing one. While time-based split methods offer an alternative, we opted for stratified k-fold cross-validation due to the dataset's imbalance, which could cause potentially bias results. For a related discussion, see Bluwstein et al. (2023).

In addition, we applied feature scaling (normalization) for SVM, KNN, Multi-Layer Perceptron (MLP), and Elastic Net methods, as these models are sensitive to scales of features, favoring features with large values over features of small values.[7]

Furthermore, a hyperparameter tuning process was adopted for each model, utilizing 'stratified 5-fold cross-validation' during the grid search process. In the context of machine learning model optimization, hyperparameter tuning refers to the systematic selection enhancement of configuration settings, known as hyperparameters, which are predefined before the model training process. The objective is to improve the model's performance by strategically selecting the most effective combination of hyperparameter values.

A common method employed for hyperparameter tuning is grid search, where an exhaustive search is conducted over a predetermined set, or "grid," of hyperparameter combinations. For each combination, the model is trained and evaluated, and the optimal set of hyperparameters is determined based on a specified performance metric. Grid search provides an efficient and thorough approach to navigate the hyperparameter space and identify the configuration that maximizes the model's predictive capabilities. Hyperparameter tuning procedure is applied to each model under consideration to obtain best performance out of that model.

After identifying the best hyper-parameters, we trained each model using the 'stratified k-fold cross-validation' procedure. Subsequently, we evaluated the performance metrics (mean ROC curves and AUC scores) for each model. The comparison of models was conducted by initially comparing the mean ROC curve and mean AUC score out of k ROC curves and AUC scores obtained from the k-folds. Based on the mean ROC curve, we also recorded the TPR (recall) for a tolerable FPR (e.g., 0.2) as model's sensitivity.

After identifying the best parameters, we trained the model using 'stratified k-fold' again. Subsequently, we evaluated the performance metrics and AUC-ROC curves

---

[7] StandardScaler is implemented with careful consideration of data leakage prevention. Feature scaling is performed after the train-test split, ensuring that the StandardScaler is fit and transformed on the training set and only fitted on the test set to maintain the integrity of the model evaluation process.

for each model. The comparison of models was conducted by initially comparing the mean AUC-ROC and then assessing the recall ratios for each model.

We implemented Elastic Net, Random Forest, XGBoost, AdaBoost, SVM, kNN, and Multi-Layer Perceptron(MLP). In the next subsection, we provide a short non-technical summary on the models.[8]

It is important to note that In the field of machine learning, model selection often involves a process of trial and experimentation with various algorithms, rather than a pre-selection of a single algorithm. Researchers explore the strengths and weaknesses of different models, considering factors such as performance, interpretability, and scalability, before ultimately choosing a model based on empirical evidence gathered from experimentation. This approach promotes transparency and ensures that the selected model is well-suited to the data and problem domain. Therefore, we utilize a diverse set of machine learning algorithms to thoroughly explore our dataset and identify the most suitable model for our task.

4.3.1. Brief Summaries on the Selected Machine Learning Methods

In the ever-evolving landscape of technology, the realm of machine learning stands as a beacon of innovation and problem-solving. As we delve into the intricacies of various machine learning methods, it becomes imperative to understand their collective significance in shaping the way computers learn from data and make intelligent decisions.

Machine learning, a subset of artificial intelligence, empowers systems to learn patterns, adapt to changing scenarios, and improve their performance over time. This transformative field has garnered immense importance across industries, offering

---

[8] Without a predetermined rationale for selecting specific algorithms for each dataset, there would be no prior knowledge of which method is best suited for a particular dataset. This uncertainty is common in machine learning experimentation, as the performance of algorithms can vary depending on the characteristics of the dataset and the complexity of the problem being addressed. Therefore, researchers often employ a trial-and-error approach, experimenting with various algorithms to determine which ones perform best for their specific task.

solutions to complex problems in diverse domains. From predicting financial trends and analyzing medical data to recognizing patterns in vast datasets, machine learning has become an indispensable tool.

The exploration will focus on several prominent machine learning methods, each with its unique strengths and applications. These methods include Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Multi-Layer Perceptron (MLP), and ensemble methods like Random Forest, XGBoost (Extreme Gradient Boosting), and AdaBoost (Adaptive Boosting). Each method brings a distinctive approach to the table, addressing various challenges and catering to specific types of data.

Before delving into the technical intricacies, a brief comparison of these methods will highlight their significance. Support Vector Machines (SVM) excel in finding optimal hyperplanes for classification tasks, while k-Nearest Neighbors relies on proximity for classification. Ensemble methods like Random Forest and XGBoost leverage the power of multiple decision trees, offering enhanced accuracy and robustness. AdaBoost, with its adaptive learning, focuses on refining the accuracy of weak learners. Multi-Layer Perceptron, a cornerstone of neural networks, excels in capturing complex relationships in data.

This enlightening journey into the world of machine learning unveils a landscape where algorithms learn, adapt, and contribute to the ever-expanding realm of technological marvels.

4.3.1.1. Elastic Net

Elastic Net is a regularization and variable selection method used in machine learning for regression and classification, initially proposed by Zou, H., & Hastie, T. (2005). It combines L1 and L2 penalties to prevent overfitting and address multicollinearity. In classification tasks, Elastic Net is often applied to logistic regression, introducing sparsity (like Lasso, proposed first by Tibshirani, R. (1996)) and controlling coefficient magnitudes (like Ridge, first introduced by Hoerl, A. E.,

& Kennard, R. W. (1970)). This regularization technique is valuable in high-dimensional datasets, automatically selecting relevant features and improving predictive performance by striking a balance between simplicity and accuracy. Elastic Net regularization operates by adding penalty terms to the standard objective function in machine learning models. These penalty terms influence the optimization process during model training, aiming at preventing overfitting and enhancing model generalization. In the context of regression, such as linear regression or logistic regression, Elastic Net introduces two specific regularization terms: L1 and L2 penalties.

The L1 penalty (Lasso) is incorporated by adding the sum of the absolute values of the coefficients to the objective function. This inclusion promotes sparsity within the model, compelling certain coefficients to be precisely zero. In simpler terms, it facilitates automatic feature selection by effectively disregarding less impactful features. Conversely, the L2 penalty (Ridge) is introduced through the addition of the sum of the squared values of the coefficients to the objective function. This imposition serves to control the overall magnitude of the coefficients, averting them from reaching excessive values. Such control proves beneficial in addressing multicollinearity, particularly in situations where predictor variables exhibit correlation.

$$Elastic\ Net\ Penalty = \ \alpha(\lambda_1\|\beta\|) + (1-\alpha)\left(\frac{\lambda_2}{2}\|\beta\|_2^2\right) \qquad (4.5)$$

,where α controls the mix between L1 and L2 penalties, ranging from 0 to 1. λ_1, λ_2 are regularization parameters for L1 and L2 penalties, respectively. β represents the vector of model coefficients. Regularization parameters (λ_1, λ_2)and mixing parameter(α) are hyperparameters of the model, they need to be tuned for higher accuracy.

We incorporated the Elastic Net Penalty into logistic regression by augmenting the regular maximization of the likelihood problem, thereby transforming it into a Penalized/Regularized Logistic Regression. While the maximum likelihood estimator

seeks to fit the data as closely as possible, the penalty term constrains the model's capacity to fit the data by discouraging larger slope coefficients, introducing a trade-off in the likelihood maximization problem.

From the standpoint of economists and econometricians, Varian (2014), Athey & Imbens (2019), and Mullainathan & Spiess (2017) offer valuable insights. For a more in-depth understanding, consult the work of Zou, H., & Hastie, T. (2005), as well as the widely referenced textbook 'The Elements of Statistical Learning' by Hastie, Tibshirani, and Friedman (2009).

4.3.1.2. Tree-Based Methods

While one of our primary focus is on employing tree-based models as a class of nonlinear ML models (Random Forest, XGBoost, and AdaBoost), it is essential to provide concise explanations of the tree-based methods in general. Tree-based methods, forming a family within supervised machine learning, execute classification and regression tasks by constructing tree-like structures to predict the target class or value based on input features.

Tree-based machine learning methods, from most simple to complex, can be categorized as follows:

- ***Single Tree Model***:

A single Decision Tree undergoes division into two branches at each depth level, starting from the top node. The end branches, or leaves, where no further splitting occurs, represent the final decisions. Conditions based on feature values guide the binary choices determining the next branch. This process continues until one of the leaves is reached.

A single decision tree holds notable advantages, particularly in its high interpretability and transparency. The straightforward, graphical representation allows for an intuitive understanding of the decision-making process, making it an

effective tool for communicating insights to various stakeholders. Additionally, decision trees are non-parametric and make no assumptions about the data distribution, demonstrating flexibility in handling both numerical and categorical features without extensive pre-processing. The provision of feature importance information is valuable for identifying key variables influencing model decisions. However, these advantages come with inherent drawbacks. Single decision trees are susceptible to overfitting, especially in cases of model complexity, leading to potential challenges in generalizing to new data. They exhibit sensitivity to small variations in the training data, resulting in different tree structures for similar datasets. Additionally, their lack of robustness in the face of noisy data or outliers and limited expressiveness in capturing complex patterns and nonlinear relationships may limit their overall predictive capabilities. Therefore, the decision to employ a single decision tree should be made judiciously, considering the specific characteristics of the data and the balance between interpretability and model performance.

- **Ensemble of Decision Trees (Random Forest):**

Random Forest utilizes an ensemble of decision trees created independently based on subsets of training data. Positioned at a higher level compared to individual Decision Trees, the Random Forest ensemble method typically exhibits higher accuracy.

- **Gradient Boosted Trees:**

The boosting method progressively creates models, starting with the first model that learns from the training data. Subsequent models then incorporate knowledge from both the training data and the errors of preceding models. This sequential process enhances model performance compared to ensemble trees. Gradient Boosting builds trees sequentially, with each new tree aiming to correct the errors of the combined ensemble. This iterative approach aids in reducing overfitting, as each subsequent tree focuses on areas where the ensemble has performed poorly. Gradient Boosted Trees often incorporate a shrinkage parameter, also known as the learning rate. A

smaller learning rate penalizes the contribution of each individual tree, providing a mechanism to control overfitting. Additionally, parameters governing the depth of individual trees and regularization terms are frequently fine-tuned to further prevent overfitting. In essence, Gradient Boosted Trees present a powerful approach for improving model accuracy while effectively addressing concerns related to overfitting.

- ***Extreme Gradient Boosting (XGBoost):***

XGBoost, a variant of Gradient Boosting, is designed to address overfitting issues.

Further explanation on Random Forest, XGBoost, AdaBoost:

- ***Random Forest Method***

A Random Forest is an ensemble machine learning model that operates by constructing a multitude of decision trees during training and outputting the mode of the classes (classification) or the mean prediction (regression) of the individual trees. It belongs to the class of ensemble learning methods, which combine the predictions of multiple models to improve overall performance and robustness.

Following is a breakdown of key components:

Decision Trees: The basic building blocks of a Random Forest are decision trees. Each tree is constructed by recursively partitioning the data into subsets based on features, making binary decisions at each node. The decisions lead to the assignment of a target value or class at the tree's terminal nodes (leaves).

Ensemble Learning: A Random Forest builds multiple decision trees independently during training. Each tree is trained on a random subset of the data, and the randomness is introduced by selecting a random subset of features at each split. This diversity helps the model generalize well to unseen data.

Voting or Averaging: For classification tasks, the final prediction of the Random Forest is determined by a majority vote from the individual trees. In regression tasks, it's the average of the predictions.

Bootstrapping: Random Forest employs bootstrapped sampling (sampling with replacement) to create different subsets of the training data for training each tree. This introduces variability in the training process.

Feature Randomness: At each split in a tree, only a random subset of features is considered for making the decision. This further enhances the diversity among the trees.

The concept of Random Forests was introduced by Leo Breiman, a statistician and professor at the University of California, Berkeley. In this seminal work, Breiman(2001) outlined the principles behind Random Forests and highlighted their effectiveness in improving the accuracy and robustness of machine learning models. Since then, Random Forests have become a widely used and influential approach in the field of machine learning and data science. Renowned for their robustness, versatility, and capacity to handle high-dimensional datasets, Random Forest models are extensively employed in classification and regression tasks, consistently delivering high accuracy and generalization performance.

Optimizing the performance of a Random Forest model necessitates meticulous tuning of key hyperparameters. Among these, the number of trees in the forest (n_estimators), the maximum depth of each tree (max_depth), and the minimum number of samples required to split an internal node (min_samples_split) play pivotal roles in balancing model complexity. Additionally, parameters like min_samples_leaf, controlling the minimum number of samples in a leaf node, and max_features, determining the maximum features considered for node splitting, are vital for mitigating overfitting. The criterion for assessing split quality (criterion), and the assignment of weights to different classes for imbalanced datasets (class_weight) are other considerations. Hyperparameter tuning involves

systematically adjusting these factors, often employing techniques like grid search or random search, combined with cross-validation, to identify the optimal configuration tailored to the specific dataset and problem at hand.

From the standpoint of economists and econometricians, Varian (2014), Athey & Imbens (2019), and Mullainathan & Spiess (2017) offer valuable insights. For a more in-depth understanding, consult the work Breiman(2001) as well as the widely referenced textbook 'The Elements of Statistical Learning' by Hastie, Tibshirani, and Friedman (2009) and the second edition of the textbook Hastie, Tibshirani, and Friedman (2017) .

- **XGBOOST (Extreme Gradient Boosting):**

Developed by Chen & Guestrin (2016), XGBoost has emerged as a standout methodology within ensemble learning, particularly in the gradient boosting framework. Employing a boosting framework, XGBoost builds an ensemble of weak learners sequentially, with each aimed at correcting errors made by the existing ensemble. In XGBoost, weak learners are typically shallow decision trees or stumps, known for their limited predictive power individually.

The process of error correction in XGBoost involves several key mechanisms. The algorithm minimizes an objective function comprising a loss term and a regularization term. The loss term quantifies prediction errors, while the regularization term controls model complexity. Iterative optimization of this objective function occurs during training, facilitated by gradient descent. XGBoost introduces L1 and L2 regularization terms into the objective function, penalizing complexity and contributing to tree pruning, preventing overfitting, and promoting generalized model development. Weighted updates from weak learners further refine the model, with more accurate learners receiving higher weights, emphasizing their influence on the ensemble. The inclusion of a shrinkage parameter, or learning rate, scales the contribution of each weak learner, offering more conservative updates and often enhancing generalization.

In summary, XGBoost error correction strategy integrates objective function optimization, gradient descent, regularization, and weighted updates from weak learners. This ensemble learning approach, coupled with meticulous hyperparameter tuning, positions XGBoost as a favored choice for creating highly accurate and robust models across diverse machine learning applications.

Tuning XGBoost hyperparameters is critical for model optimization and preventing overfitting. The learning rate dictates the optimization step size, with lower values enhancing robustness. The number of boosting rounds influences accuracy, necessitating a balance for efficient computation. Maximum tree depth (max_depth) controls complexity, while minimum child weight (min_child_weight) influences algorithm conservatism. Subsample introduces randomness, and colsample_bytree varies feature selection. Gamma determines minimum loss reduction for node partitioning, and regularization terms (alpha and lambda) mitigate overfitting. Scale_pos_weight addresses class imbalance, and specifying the objective function is crucial. Systematic tuning, often via techniques like grid or randomized search with cross-validation, is essential for finding optimal hyperparameter combinations tailored to specific datasets and tasks.

- *AdaBoost (Adaptive Boosting):*

AdaBoost, an abbreviation for Adaptive Boosting, introduced by Yoav Freund and Robert Schapire in 1996, distinguishes itself as a potent ensemble method with a unique approach compared to Gradient Boosted Trees. Freund and Schapire(1996) developed AdaBoost as a method to boost the performance of weak learners, particularly for binary classification problems. Their work on AdaBoost has had a significant impact on the field of machine learning and ensemble methods.

At its core, AdaBoost focuses on the iterative adjustment of sample weights, giving prominence to misclassified instances to progressively enhance model accuracy. This algorithm assigns equal weights to all data points initially, but as each iteration unfolds, it elevates the weights of misclassified samples, emphasizing their

importance in subsequent rounds. The sequential construction of weak learners characterizes AdaBoost, with each learner concentrating on rectifying the misclassifications made in previous rounds. The final model emerges as a weighted sum of these learners, their contributions based on accuracy. AdaBoost's adaptability, ability to achieve high accuracy with simple weak learners, and a minimal set of hyperparameters contribute to its appeal. However, its sensitivity to noisy data and the potential for overfitting, especially with complex base models, pose challenges. In summary, AdaBoost stands as a versatile ensemble method, adept at iteratively refining model accuracy by strategically adjusting its focus, making it a valuable tool in classification tasks, particularly when simplicity and adaptability are crucial.

In the process of tuning AdaBoost for optimal performance, several key hyperparameters come into consideration. The number of estimators (n_estimators) stands out as a crucial parameter, representing the count of sequentially trained weak learners to be combined in the ensemble. While increasing the number of estimators can enhance performance, it's essential to balance this against potential longer training times. The learning rate (learning_rate) determines the contribution of each weak learner to the final combination, with a lower learning rate requiring more estimators but potentially improving generalization. The choice of the base estimator (base_estimator) is another pivotal decision, often involving decision trees, and its parameters, such as maximum depth (max_depth) or minimum samples required for splitting (min_samples_split), should be fine-tuned.

For detailed mathematical and technical information, the textbook "Elements of Statistical Learning (2017)" (ESL) offers comprehensive explanations on Support Vector Machines (SVM) in Chapter 10 as well as the article by Freund and Schapire(1996).

4.3.1.3. SVM (Support Vector Machines)

Support Vector Machine (SVM) is a supervised machine learning algorithm designed for classification and regression tasks. Developed by Vladimir Vapnik and his colleague Alexey Chervonenkis in the 1960s and 1970s, SVM's primary objective is

to find a hyperplane in a high-dimensional space that maximally separates data points into different classes. Initially intended for binary classification, SVM has been extended for multi-class scenarios and regression tasks.

The core aim of SVM is to identify an optimal hyperplane with the largest margin, where the margin represents the distance between the hyperplane and the nearest data point from any class. This emphasis on a large-margin hyperplane not only facilitates effective data separation but also enhances generalization to unseen data.

The optimization process in SVM involves determining hyperplane parameters that maximize the margin while adhering to specific constraints. This task centers around minimizing a cost function, which incorporates a term penalizing misclassification error and a regularization term for model complexity control. Quadratic Programming or Sequential Minimal Optimization are common techniques employed to solve the optimization problem.

An intrinsic feature of SVM is its ability to handle non-linear relationships through the kernel trick. The kernel function allows SVM to implicitly map input data into a higher-dimensional space, enabling the discovery of a hyperplane for effective separation between classes in non-linear scenarios.

While SVM offers advantages such as effectiveness in high-dimensional spaces and applicability when dimensions exceed samples, challenges include sensitivity to kernel and parameter choices, as well as potential computational expenses when training on large datasets. In summary, SVM stands out as a versatile algorithm with a well-defined objective, making it widely utilized in various machine learning applications.

Optimizing a Support Vector Machine (SVM) involves tuning several key parameters to ensure the model's performance aligns with the characteristics of the dataset and the nature of the problem. Among these parameters, the regularization parameter (C) stands out, determining the balance between achieving a smooth

decision boundary and accurate classification. The choice of kernel type, including linear, polynomial, radial basis function (RBF), and sigmoid, plays a pivotal role in mapping input data into higher-dimensional spaces. Additionally, kernel coefficients (gamma for RBF, coef0 for polynomial/sigmoid) influence the complexity of decision boundaries, with higher values introducing more intricate patterns.

Degree, specific to polynomial kernels, determines the degree of the polynomial function, affecting the complexity of the decision boundary. The shrinking heuristic parameter enables a faster training process by skipping certain support vectors, particularly beneficial for large datasets. Class weights offer a means of addressing imbalances in datasets by assigning different weights to different classes. The decision function shape parameter handles multi-class problems, specifying whether the decision functions are structured in a one-vs-one (ovo) or one-vs-rest (ovr) format.

For detailed mathematical and technical information, the textbook "Elements of Statistical Learning (2017)" (ESL) offers comprehensive explanations on Support Vector Machines (SVM) in both Chapter 4 and Chapter 12 as well as the article by Cortes & Vapnik(1995).

4.3.1.4. KNN (k-Nearest Neighbors)

The k-Nearest Neighbors (KNN) algorithm is a classic and versatile supervised machine learning approach widely used for classification and regression tasks. While the concept of KNN dates back to the early works in pattern recognition, there is no specific attribution to its introduction. At its core, KNN relies on the concept of proximity, where a data point is classified based on the majority class of its k nearest neighbors in the feature space. The main components influencing the algorithm's behavior include the choice of the number of neighbors (k), the distance metric used to measure similarity between data points, and the weighting scheme for neighbors during predictions. The optimization process for KNN involves tuning these parameters through techniques such as grid search or randomized search combined with cross-validation to enhance predictive performance on specific datasets.

In a classification scenario, when a new data point needs classification, KNN calculates distances between the point and all others in the training set, identifies the k nearest neighbors, and assigns the majority class of these neighbors to the new data point. KNN excels in scenarios with irregular decision boundaries and clustered data points of the same class. Its non-parametric nature and adaptability to various data types make KNN a valuable tool in machine learning, data mining, and pattern recognition. Overall, KNN's simplicity and effectiveness contribute to its popularity in scenarios where the underlying data structure is complex and not explicitly defined.

The most crucial parameter in optimizing the k-Nearest Neighbors (KNN) algorithm is the number of neighbors (k). The choice of k has a profound impact on the model's tendency toward overfitting or underfitting. A smaller value of k, such as 1 or 3, tends to create a more complex decision boundary, making the model sensitive to noise and potentially leading to overfitting, especially in the presence of outliers or localized irregularities. On the other hand, a larger k, say 10 or 20, results in a smoother decision boundary, which might lead to underfitting by oversimplifying the underlying relationships in the data. Balancing the choice of k is essential to strike a suitable trade-off between capturing local variations and maintaining generalization to ensure optimal predictive performance. While other parameters, such as the distance metric and weighting scheme, are important for fine-tuning, selecting an appropriate value for k remains a primary consideration for mitigating overfitting and underfitting in KNN models.

The k-Nearest Neighbors (KNN) algorithm, a potent tool in certain contexts, is not without its limitations. First and foremost, KNN can be computationally demanding, particularly as datasets grow in size, as it requires calculating the distance between the query instance and all training instances. Additionally, KNN is highly sensitive to noisy data and outliers, which can significantly impact the accuracy of predictions. The curse of dimensionality poses another challenge, as the algorithm's effectiveness diminishes with an increasing number of features. Selecting the optimal value for 'k,' the number of neighbors, is a critical decision, influencing the algorithm's sensitivity

to noise and patterns in the data. Moreover, imbalanced datasets can bias KNN towards the majority class, leading to less accurate predictions for minority classes. Despite these limitations, KNN remains a valuable tool, especially in scenarios with well-behaved, noise-free data and when computational resources permit. Acknowledging these drawbacks is essential for informed decision-making when considering the application of the KNN algorithm.

For detailed mathematical and technical information, the textbook "Elements of Statistical Learning (2017)" (ESL) offers comprehensive explanation in Chapter13.

4.3.1.5. Multi-Layer Perceptron (MLP)

A Multi-Layer Perceptron (MLP) is a class of artificial neural network designed for supervised learning tasks, such as classification and regression. The concept of the perceptron, which forms the foundational unit of the Multi-Layer Perceptron (MLP), was introduced by Frank Rosenblatt in 1958. However, the development of the broader MLP architecture and the backpropagation algorithm for training multi-layer networks occurred over subsequent years with contributions from various researchers. The advancement and popularization of MLPs in the field of neural networks have involved the work of multiple scientists and researchers, making it a collaborative effort over time rather than being attributed to a single individual. Comprising multiple layers, an MLP consists of an input layer, one or more hidden layers, and an output layer. Each layer contains nodes, or neurons, connected to nodes in adjacent layers through weighted connections. The information flows forward from the input layer, where input features are fed, through the hidden layers, where complex relationships are learned, to the output layer, which produces the final predictions. The nodes in each layer apply an activation function to the weighted sum of their inputs, introducing non-linearity and enabling the network to model intricate patterns in data. Training an MLP involves adjusting the weights during a process called backpropagation, where the difference between predicted and actual outputs is minimized using optimization techniques. MLPs are known for their capability to capture complex relationships in data and are foundational to deep

learning networks, playing a key role in various applications within machine learning and artificial intelligence.

When fine-tuning a Multi-Layer Perceptron (MLP) neural network, critical hyperparameters must be carefully adjusted to optimize its performance. Among these, the number of hidden layers and neurons shapes the network's architecture, influencing its capacity to model complex relationships while balancing the risk of overfitting. The selection of activation functions, such as ReLU or Sigmoid, introduces non-linearity and impacts the network's ability to capture intricate patterns. The learning rate governs the size of optimization steps, requiring a delicate balance to avoid overshooting or slow convergence. Batch size determines the number of training samples processed in each iteration, affecting the noise level in optimization updates. Epochs represent the number of passes through the entire training dataset during training, with too few leading to underfitting and too many causing overfitting.

For detailed mathematical and technical information, the textbook "Elements of Statistical Learning (2017) (ESL) " offers comprehensive explanations in Chapter11.

In conclusion, this discourse has provided an overview of selected machine learning methodologies, emphasizing their distinctive characteristics and applications. The discussion began with Elastic Net, variable selection and a regularization method combining L1 and L2 penalties, highlighting its utility in regression and classification tasks. Subsequently, the focus shifted to tree-based methods, specifically Random Forest, XGBoost, and AdaBoost, elucidating their foundational principles and contributions to ensemble learning.

The discourse then delved into Support Vector Machines (SVM), emphasizing its role as a supervised learning algorithm for classification and regression problems. SVM's core objective of identifying an optimal hyperplane in high-dimensional space and its ability to handle non-linear relationships through the kernel trick were underscored. Additionally, key parameters required for tuning SVMs to obtain optimal SVM performance were highlighted.

The discussion extended to the k-Nearest Neighbors (kNN) algorithm, exploring its reliance on proximity for classification and regression tasks. Considerations for optimizing kNN parameters, such as the number of neighbors (k), distance metric, and weighting scheme, were presented. Despite its computational demands and sensitivity to noisy data, kNN was acknowledged as a valuable tool in machine learning.

Finally, the discourse touched upon Multi-Layer Perceptron (MLP), an artificial neural network designed for supervised learning tasks. MLP's architecture, rooted in the work of Frank Rosenblatt(1957), was outlined, along with critical hyperparameters governing its optimization.

For a more comprehensive exploration of the topics covered, further references and in-depth readings from seminal works and pioneering researchers in the field are recommended. This scholarly endeavor aspires to make a meaningful contribution to the ongoing discourse surrounding the practical applications of machine learning methodologies.

The forthcoming stages of this research will involve the practical implementation of these methodologies, accompanied by a comparative analysis of their out-of-sample performances. Optimization efforts will include meticulous tuning of hyperparameters to tailor each methodology to the unique characteristics of the datasets. Through this research, we aim to contribute to a deeper understanding of the practical implications and efficiency of these machine learning methodologies in macroeconomic forecasting.

4.3.2. Model Selection: A Comparative Analysis of Machine Learning Methods and the Traditional Model

In the process of model selection, we conduct a comparative analysis based on the mean Area Under the Receiver Operating Characteristic Curve (AUC-ROC) values. While AUC-ROC serves as the initial metric for comparison, it is crucial to assess

the models' performance using additional metrics to ensure a comprehensive evaluation. Specifically, we examine recall (the detection performance of SS) as a selected additional performance metric in the second phase of investigation. Subsequently, we delve into a third phase where we investigate recall at a fixed mean false positive rate.

In other words, the strategy adopted in this study comprises a sequential comparison of mean AUC, recall -obtained by default class probability threshold of 0.5-, and a targeted assessment of recall at a fixed false positive rate.

This approach represents a holistic and methodical means of model selection. It is designed to reveal both global and nuanced insights, providing a comprehensive understanding of each algorithm's strengths and weaknesses in the context of predicting rare and impactful events.

While we have previously defined AUC-ROC, it is worth emphasizing its pivotal role in the selection process. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) serves as a crucial performance metric in binary classification tasks, including the modeling of Sudden Stop Crises. The ROC curve graphically represents the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various classification thresholds.

AUC-ROC quantifies the discriminatory power of a classification model by measuring the area under this curve. Higher AUC-ROC values, ranging from 0 to 1, indicate superior model performance.

The ROC curve aids in model selection by illustrating the trade-offs between sensitivity and specificity at different threshold settings. A model with better discriminatory ability exhibits an ROC curve that approaches the upper-left corner of the plot, leading to a higher AUC-ROC value.

Figure 4.8 below shows the performance of various ML models under study in terms of their mean AUC scores after 10-Fold cross validation procedure.
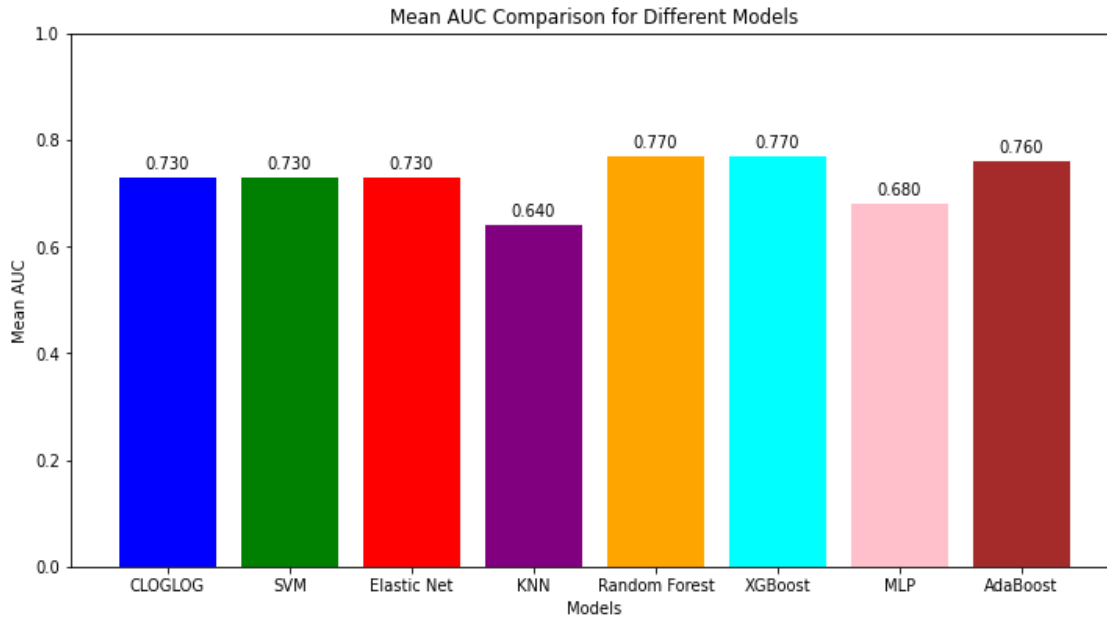
Figure 4. 8. Mean AUC Comparison for Different Models.

*Note*: we provide the AUC_ROC curves for each model in the Appendix B.

As depicted in Figure 4.8, it is evident that kNN and MLP are inferior to the base model (cloglog model), while SVM and Elastic Net demonstrate similar mean AUC scores. In contrast, tree-based models, -Random Forest, XGBoost, and AdaBoost-, outperform the base model[9]. This suggests that, while the base model shows relative proficiency in distinguishing between positive and negative classes, its overall performance is moderately good. Notably, it exhibits poor performance in recall, registering at 0.16, indicating a high likelihood of misidentifying crisis times as non-crisis. This emphasizes a notable Type II error, where the model fails to identify instances of actual crises, impacting its recall performance negatively. In terms of recall, the base model performs the worst, as shown in **Figure 4.9** and **Figure 4.10**, which display the models along with their associated accuracy, precision, recall, and F1 scores. The best-performing model in terms of recall emerges as Elastic Net, with a recall score of 0.7012, which is 337.5% higher than that of the Cloglog method.[10]

---

[9] In terms of AUC scores, the performances of Random Forest and XGBoost are almost 5.5% higher than that of Cloglog .

[10] Recall scores are calculated at the default threshold, 0.5. In this context, 'threshold' refers to the point at which class determination is made, with values above the threshold classified as positive and those below classified as negative.
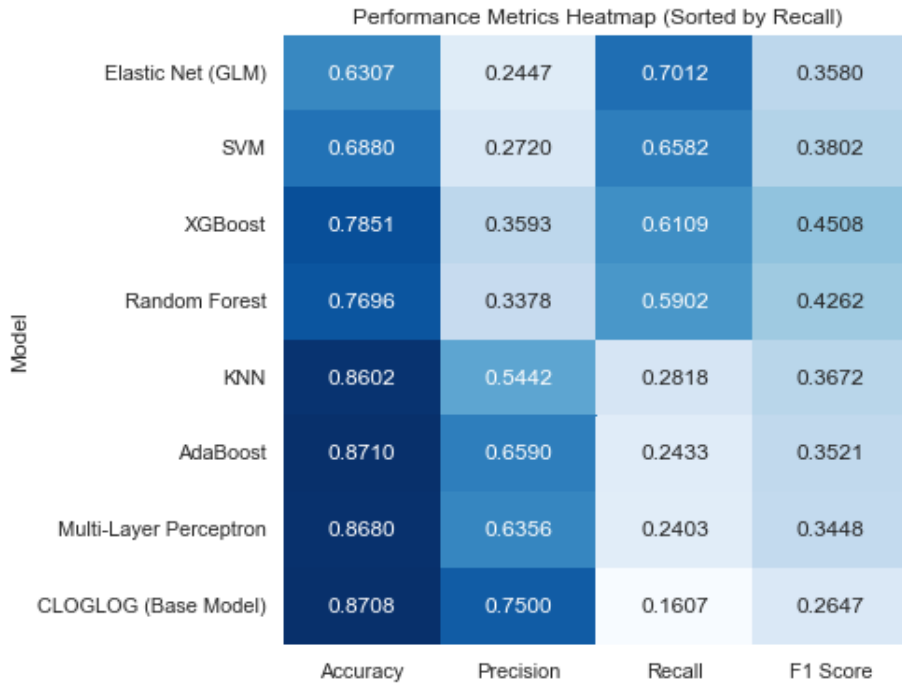
Figure 4. 9. Performance Metrics Heatmap (Sorted by Recall)
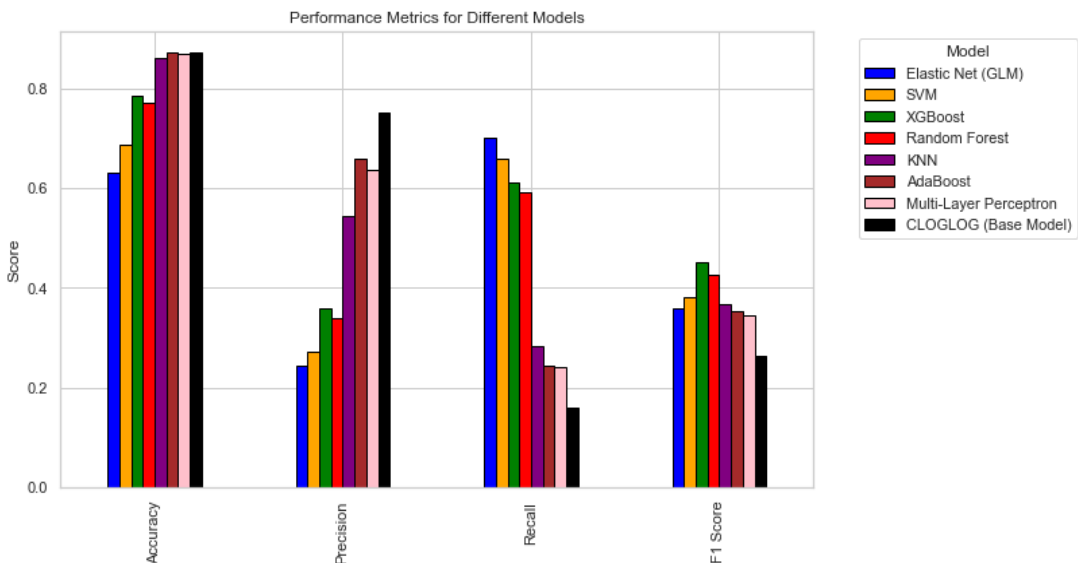
Figure 4. 10. Performance Metrics for Different Models

In the context of predicting rare events, illustrated by the occurrence of Sudden Stop Crises, conventional metrics such as accuracy and precision may be unsuitable for model selection due to inherent imbalances within the dataset. When the event of interest is significantly outnumbered by non-events, accuracy becomes a potentially misleading metric, as a model achieving high accuracy may primarily predict the

115

majority class. Precision, accentuating the accuracy of positive predictions, may similarly offer a distorted evaluation when the positive class is scarce. The inherent trade-off between precision and recall is exacerbated in such scenarios, where optimizing for one metric may lead to a compromise in the other. Given the potential consequences associated with false negatives, particularly in critical scenarios like identifying Sudden Stop Crises, prioritizing recall over precision becomes imperative. Metrics that comprehensively consider both false positives and false negatives, such as recall, f1 or the area under the ROC curve (AUC-ROC), are recommended for a more nuanced and accurate model evaluation in these circumstances.

Furthermore, since the ROC curve illustrates the trade-off between the true positive rate (sensitivity) and false positive rate (1 - specificity), it is essential to understand the inherent balance between these two parameters. This trade-off arises from the fact that adjusting the classification threshold influences the model's ability to correctly identify positive instances while simultaneously misclassifying negative instances. A typical threshold setting is 0.5, meaning that instances with predicted probabilities greater than or equal to 0.5 are classified as positive. However, this threshold is not universally optimal for all scenarios. Consequently, the choice of threshold has a direct impact on the model's operating point in the trade-off space. For instance, setting a specific false positive rate, such as 0.2, allows us to explore how different models perform in terms of recall, thereby providing insights into their ability to identify true positive instances while controlling the false positive rate according to a predefined criterion.

As the final step in model selection, we fixed the mean false positive rate at 0.2, setting a tolerance level of only 20 percent misclassification of non-crisis (normal) events as crises. This will allow the model on average to misidentify non-crisis times as Sudden Stop crises 20% of the time. Subsequently, we examined the corresponding mean true positive rates (recall) on their individual ROC curves.[11]

---

[11] Similar to Bluwstein et al. (2023). In their model comparison, they adjust the models to accurately identify 80% of financial crises. Subsequently, they assess the false alarm rates across models,

The choice to set the mean false positive rate at 0.2 during model evaluation is driven by several considerations specific to the practical requirements of the application. In the context of imbalanced data, particularly when dealing with rare events like sudden stop crises, the default threshold of 0.5 may inadequately address the need for nuanced performance evaluation. By fixing the false positive rate at a certain value (e.g., 0.2), the assessment becomes tailored to scenarios where controlling false positives is paramount. This approach not only aligns with practical considerations of limiting misclassifications but also allows for a more nuanced examination of a model's performance under conditions reflective of the real-world application.

Figure 4.11 illustrates the SS detection performance of the studied ML methods in terms of the recall scores, offering insight into each model's ability to capture true positive instances at a fixed False Alarm Rate. The results align with the earlier comparison of mean AUC scores, confirming the superior performance of tree-based methods over other models and the base model.

Model	Fixed False Positive Rate	Corresponding Mean True Positive Rate
XGBOOST	0.2	0.627446
Adaboost	0.2	0.604197
Random Forest	0.2	0.599864
CLOGLOG(BASE MODEL)	0.2	0.553571
SVM	0.2	0.517208
Elastic Net	0.2	0.514794
KNN	0.2	0.411598
MLP	0.2	0.359350

Figure 4. 11. Detection Performance of Studied Models: Recall Scores at FPR 0.2

In conclusion, the comprehensive assessment, considering mean AUC, recall, and the fixed mean false positive rate, positions tree-based models as more favorable choices for predicting sudden stop crises. However, the ultimate model choice should align

representing the proportion of times a crisis is signaled but doesn't occur. Conversely, our methodology fixes the false alarm rate and focuses on comparing the recall, specifically the true positive rate. By ensuring a predetermined level of false alarms, we can effectively control for the risk of unnecessary panic or disruption caused by false crisis signals.

with specific application priorities, accounting for trade-offs and considering potential consequences, such as false negatives, in achieving the overall goals of the predictive modeling task.
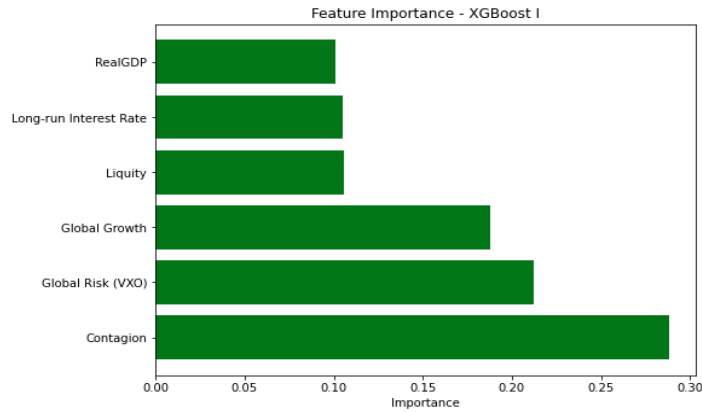
4.3.3. Feature importance and Shapley Values

After the comparison, it is evident that XGBoost surpasses all the selected methods, including the base model. Although we refrain from conducting a causality analysis, delving into the feature importance of XGBoost provides valuable insights. Feature importance elucidates the contribution of each variable to the predictive power of the model, leveraging the Gini impurity metric. It is crucial to note that this metric doesn't infer the causes of Sudden Stop Crises but rather identifies predictors.

In the realm of model interpretability, both feature importance and Shapley values serve as crucial tools, each offering distinct insights into the predictive dynamics of a model. Feature importance, often computed using metrics like Gini impurity in the case of XGBoost, highlights the relative significance of different variables in contributing to the model's overall predictive power. The resulting scores range between 0 and 1, with a higher value indicating a more substantial impact on predictions. On the other hand, Shapley Values, rooted in cooperative game theory, take a collaborative approach to understanding variable contributions. They meticulously consider the interactions and non-linearities among features, providing a more nuanced perspective compared to traditional feature importance metrics. Shapley values distribute the predictive contribution of each variable across all possible combinations, allowing for a comprehensive evaluation of individual and joint impacts.

While feature importance gives a quick overview of influential predictors, Shapley values excel in capturing intricate relationships and dependencies within the model. Moreover, Shapley values inherently ensure that the sum of contributions across all variables aligns with the model's overall prediction, providing a more internally consistent measure.

Figure 4. 12. Feature Importance- XGBoost I

In practice, the choice between feature importance and Shapley values depends on the specific goals of the analysis. If a quick understanding of influential predictors suffices, feature importance may be the go-to metric. However, for a more thorough exploration of variable interdependencies and nuanced contributions, Shapley values offer a deeper and more context-aware examination. By leveraging both these tools, analysts can gain a comprehensive understanding of the factors shaping their predictive models.As illustrated in **Figure 4. 12**, the feature importance of the trained XGBoost model highlights Contagion, Global Growth, and Global Risk (VXO) as the most influential predictors. These variables play a significant role in shaping the model's predictive capabilities, offering valuable information for understanding and interpreting the factors contributing to Sudden Stop Crises.
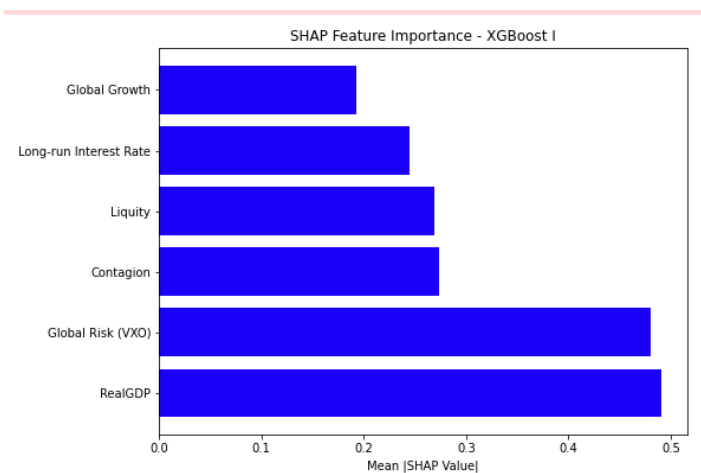
Figure 4. 13. Shapley Feature Importance-XGBoost I[12]

---

[12] Similar to Bluwstein et al. (2023). In their problem to predict financial crises ,they implement Shapley Values to the best performing model in their case, which is Extremely Randomized Trees.

In Figure 4.13, Shapley values reveal that Real GDP, Global Risk (VXO), and Contagion are the best predictors. This gives a higher value for Real GDP in contrast to the feature importance, but the other two most important variables remain the same: Contagion and Global Risk (VXO).

4.3.4. Summary on Machine Learning Methods and The Estimation Strategy

In this chapter, we undertook several key steps in the prediction of Sudden Stop events. Firstly, we leveraged the capital flow dataset curated by Forbes and Warnock (2021) to identify Sudden Stop Crises in 59 countries. This comprehensive dataset spans both advanced and emerging nations, covering the period from 1978 Q1 to 2020 Q3. Following this, we replicated the estimation process for the base case presented in their analysis. This involved employing the complementary log-log model and scrutinizing its out-of-sample performance, establishing it as our baseline scenario.

Secondly, we explicitly framed our prediction problem as a comparison of out-of-sample performances, distinct from a parameter estimation or causal inference problem. We aimed to highlight the implementation and estimation strategies for several key classes of ML methods, briefly touching upon sample division, cross-validation techniques, hyperparameter tuning, and feature scaling.

Thirdly, we provided concise, non-technical summaries of selected methods, including Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Multi-Layer Perceptron (MLP), and ensemble methods such as Random Forest, XGBoost, and AdaBoost. Each method brings a unique approach to addressing various challenges and caters to specific data types. Technical references for these methods were also provided.

The implementation of these selected methods was carried out using the Python programming language. Subsequently, we explained our model selection strategy, referring to relevant metrics. Sequentially, we assessed the mean AUC scores, examined recall as an additional performance metric, and compared recall (true

positive rate) while fixing the false alarm rate. It's worth noting that while we conducted evaluations in three sequential steps, the ultimate and definitive step lies in rectifying false alarms and ensuring recall. The results unveiled the superior performance of tree-based methods, notably with XGBoost emerging as the top-performing model. It attains a recall score of 0.627, marking a remarkable improvement of nearly 14% compared to the recall score of 0.55 achieved by the complementary log-log model, when allowing for a 20% false alarm rate.

Furthermore, after identifying XGBoost as the top-performing model, we delved into understanding which variables were most influential in predicting Sudden Stops. We utilized both the feature importance method and Shapley values[13], acknowledging differences in methodology and results. Commonly identified as the best predictors were Contagion, indicating Sudden Stops occurring in the same region as the specific country but with a one-quarter lag, and Global Risk, represented by the change in VXO with a one-quarter lag.

---

[13] To compute Shapley importances, we utilize the SHAP library in Python.

CHAPTER 5

MACHINE LEARNING SOLUTIONS FOR SUDDEN STOP PREDICTION: DATASET EXPANSION, VARIABLE SELECTION, AND MODEL DEVELOPMENT

In this chapter, we introduce machine learning-based solutions to address the Sudden Stop prediction problem. Our approach involves expanding the dataset utilized in the previous chapter by incorporating extensive quarterly data sourced from the International Monetary Fund (IMF). This dataset is then merged with the model-selected variables used in our earlier estimations, and additional model selected variables from Forbes & Warnock (2021), effectively extending the dataset.

While the initial dataset includes model-selected and widely used variables, the extension process is primarily driven by data availability. Notably, we do not take collinearity into account, and the selection of variables is independent of their alignment with theoretical foundations or common usage in empirical studies predicting Sudden Stop events.

It is crucial to emphasize that the extension of the dataset does not involve prioritizing specific data-driven variables. Instead, it encompasses the inclusion of a broader set of variables without pre-selection. In this context, traditional statistical criteria such as collinearity considerations are set aside. Our methodology is anchored in a robust reliance on machine learning (ML) feature selection methods. These methods are chosen for their proven ability to objectively identify and prioritize relevant variables within the expanded dataset.

Our objective is two-fold. First, we aim to identify the important variables that play a role in the identification of SS events. Second, using the selected variables we aim to build highly discriminative ML models that perform well on the out-of-sample data.

The data selection criterion is driven by data availability, prioritizing variables with fewer than 10% missing values. We meticulously examined all available quarterly datasets from the International Monetary Fund (IMF) website for the same 51 countries as in our previous estimation within the consistent time span from 1986 to 2018. The Balance of Payments (BOP), International Financial Statistics (IFS), and Direction of Trade (DOT) datasets proved to be the most suitable datasets for our analysis. For data augmentation, we integrated features from BOP, IFS, and DOT. To avoid redundancy, we excluded variables present in both BOP and IFS, resulting in a refined set of 192 variables. Through country-specific mean imputation for missing values, we ensured the creation of a comprehensive dataset with minimal gaps. Additional enhancements included deriving 1-quarter lag versions of the primary variables, integrating year-over-year percentage changes, and capturing the rate of change from the previous quarter.

This comprehensive feature enhancement approach utilized one-period lagged versions, rate of change versions, and year-over-year versions to enrich the dataset, resulting in a total of 768 exogenous variables.

Within the BOP and IFS datasets, we obtained a variety of variables, encompassing components of the Balance of Payments such as capital account details, current account details, as well as variables like total international reserves and international liquidity. Unfortunately, exchange rate-related variables and variables related to unemployment and GDP components in IFS had more than 30% missing values for our period of interest and country sample. From the DOT dataset, we extracted trade-related variables for each country, including the value of imports, exports, and trade balance, all measured in US dollars. These variables were aggregated across different markets, including trade with Advanced Economies, Emerging and Developing Economies, Emerging and Developing Europe, Emerging and Developing Asia, as well as the total global trade.

In line with the previous analysis, our objective was to maintain consistency by retaining data from as many countries as possible. However, we had to exclude Hong

Kong and Norway from analysis due to their sizeable missing data, leaving us with 49 countries for continued analysis.

The augmentation phase also entailed merging variables from previous estimations, including global liquidity, global risk, long-run interest rates, global growth, local GDP growth, and contagion. Additional model-based variables were also sourced from Forbes & Warnock (2021),which included shadow short-run rates, oil prices, commodity prices, global inflation rate, and dummy variables associated with region, income group, and EM or Advanced Economy classification. The entire variable list is given in the Appendix A.

Our ML approach involved the development diverse machine learning models, including Random Forest, XGBoost, Support Vector Machines (SVM), Elastic Net, and Logistic Regression.[14] To perform variable selection, we applied model-specific methods. For these models, we excluded the current values of variables and utilized the 1-period lagged, rate of change in the previous period, and year-over-year percentage change of the variables from the initial 192 obtained from the IMF, as mentioned earlier. Consequently, the final set of variables derived from this process totaled at 576. We supplemented this set with the 1-period lagged versions of the 28 variables selected from Forbes & Warnock (2021), resulting in a total exogenous variable set of 604.

For Random Forest, XGBoost, and SVM, we employed the Recursive Feature Elimination (RFE) method for feature selection. Specifically, we opted for the RFE with Cross-Validation (RFECV) method, integrating cross-validation into model-based feature selection for a more robust performance assessment, avoiding

---

[14] The decision not to employ the k-Nearest Neighbors (k-NN) algorithm for the dataset containing 604 variables stemmed from multiple factors. High dimensionality, computational complexity, and the curse of dimensionality pose significant challenges to the effectiveness of k-NN in such scenarios. Furthermore, concerns regarding model interpretability and the availability of more suitable algorithms for high-dimensional data played a role in guiding the decision-making process. k-NN typically struggles in high-dimensional spaces due to the curse of dimensionality. As the number of dimensions increases, the meaningfulness of distances between data points diminishes, thereby hindering k-NN's ability to accurately identify nearest neighbors. Despite its previous use in analyses involving only 6 variables, the limitations of k-NN become more pronounced in datasets with significantly higher dimensionality.

overfitting, enhancing generalization capability, and reducing bias in feature importance estimation.[15]

On the other hand, Elastic Net utilized the Lasso method for variable selection, employing regularization to promote sparsity. Logistic Regression adopted a hybrid approach, initially employing Random Forest for feature ranking and subsequently utilizing stepwise Bayesian Information Criterion (BIC) for further selection.[16]

After variable selection, our estimation methods as before, included Stratified K-fold Cross Validation and grid search for hyperparameter tuning. Details of the tuned hyperparameters for each model can be found in the Appendix B.

In addition to the above traditional methods, we introduced the Long Short-Term Memory (LSTM) model, a powerful deep learning technique. Emphasizing its ability to learn temporal dependencies, LSTM's forecasting strategy sets it apart from traditional models. The model excels in capturing temporal long and short-term dependencies in time-ordered data, providing a nuanced understanding of the evolution of dynamic processes like the Sudden Stop events.

---

[15] In many real-world applications, the costs associated with prediction errors can vary significantly depending on the specific context. For example, in medical diagnosis, the cost of failing to detect a disease (false negative) may be much higher than incorrectly diagnosing a healthy individual (false positive). By incorporating cost-sensitive considerations into the feature selection process, Recursive Feature Elimination (RFE) can prioritize features that are most relevant for minimizing the overall cost of errors. RFE, along with its cross-validated counterpart RFECV (Recursive Feature Elimination with Cross-Validation), is a popular feature selection technique known for its effectiveness in optimizing model performance while considering asymmetric error costs.

[16] For datasets with a high dimensionality, such as those with 604 variables, direct application of the Bayesian Information Criterion (BIC) can be computationally demanding. In response, various strategies have been proposed to address this challenge. For example, Hellwig (2021) proposes a method that begins with a minimal model (e.g., containing only an intercept) and iteratively expands it by adding variables that maximize the model fit, as assessed by BIC, until no further improvement is observed. In our study, we adopt a different yet complementary approach. Initially, we leverage the Random Forest algorithm to train a model and rank the variables based on their feature importance. This allows us to identify the variables that contribute most significantly to the predictive power of the model. Subsequently, we implement a forward selection procedure, where variables are added one at a time based on their importance ranking. At each step, we calculate the BIC value for the augmented model and compare it to the previous iteration. The process continues until there is no further improvement in BIC, indicating that the addition of additional variables does not enhance model fit. By prioritizing variables according to their importance ranking from Random Forest and evaluating their incremental contribution to model fit using BIC, our approach effectively balances computational efficiency with model selection accuracy. This hybrid methodology enables us to identify a parsimonious set of variables that collectively capture the underlying structure of the data, facilitating robust and interpretable model development.

The following subsections offer a concise summary of the Recursive Feature Elimination with Cross-Validation (RFECV) process, providing a general overview. Following this summary, we delve into detailed insights for each model, culminating in a comprehensive model evaluation and comparison.

5.1. Recursive Feature Elimination with Cross-Validation (RFECV) Process

For the variable selection process, we chose the Recursive Feature Elimination with Cross-Validation (RFECV) method for Random Forest, XGBoost, and SVM. The incorporation of cross-validation within the RFECV framework is crucial for several reasons. Cross-validation involves dividing the dataset into multiple folds, training the model on distinct subsets, and evaluating its performance on separate test sets. The significance of cross-validation in feature selection lies in

• *Robust Performance Assessment:*

Cross-validation ensures a more robust evaluation of model performance by assessing it across different data subsets. This guards against overfitting to a specific training set and provides a reliable estimate of how well the model generalizes to unseen data.

• *Avoiding Overfitting in Feature Selection:*

The iterative nature of RFECV, coupled with cross-validation, helps prevent overfitting to idiosyncrasies in any single partition of the data. This safeguards against the selection of features that may be influential only in certain subsets.

• *Enhancing Generalization Capability:*

RFECV with cross-validation contributes to the selection of features that consistently exhibit importance across various data subsets. This improves the model's ability to generalize well to new and unseen data, promoting its overall reliability.

• Reducing Bias in Feature Importance Estimation:

Cross-validation aids in reducing bias in the estimation of feature importance. It ensures that the ranking of features is based on their impact on model performance across different training and validation sets, providing an unbiased selection.

In summary, the integration of cross-validation within RFECV is pivotal for ensuring a robust and unbiased variable selection process, ultimately leading to the identification of a subset of features that enhances model interpretability and generalization.

5.2. Estimation Results

5.2.1. Random Forest

In the variable selection phase, we employed the Recursive Feature Elimination with Cross-Validation (RFECV) method. During this step, RFECV for Random Forest was configured to choose a minimum of 100 features, optimizing for detection sensitivity (i.e., recall score). This resulted in the selection of 131 variables from the initial pool of 604. Subsequently, we retrained the model using a stratified 10-fold cross-validation, incorporating the best parameters obtained through the grid search.

Post-retraining, we delved into the identification of the most influential features using the Random Forest`s *'feature importance'* method which serves as a valuable technique to evaluate the contribution of each variable (feature) in predicting the target outcome. In Random Forest:

The feature importance is determined by assessing how much each feature decreases the model's impurity, measured by Gini impurity or information gain metrics, during the construction of decision trees within the forest.

During the training of a Random Forest model, decision trees are built by considering a random subset of features at each split. The feature importance is then calculated based on the average decrease in impurity across all decision trees.

Features that lead to a more significant decrease in impurity are assigned higher importance scores. These scores are normalized to establish a relative ranking of feature importance.

The interpretation of feature importance scores lies in understanding that a higher score indicates greater influence in making accurate predictions. This method helps to identify which features contribute the most to the overall predictive power of the model. In essence, the 'feature importance' method offers insights into the relative significance of each feature, aiding in the understanding and interpretation of the Random Forest model's predictive capabilities.

As depicted in **Figure 5.2,** the algorithm pinpointed the year-over-year percentage change in the country's trade with advanced economies ('Advanced Economies - Goods, Value of Imports, Cost, Insurance, Freight (CIF), US Dollars') as the foremost predictor. Close behind were variables associated with the year-over-year percentage change in the country's 'Current Account, Total, Debit, US Dollars,' along with another trade variable indicating the country's total value of imports worldwide ('World - Goods, Value of Imports, Cost, Insurance, Freight (CIF)').

Among the top 30 features, the pre-selected model-based variables (the variables from Forbes&Warnock,2021) encompassed the average of short-run interest rates for Japan, EU, and the UK, global growth, VXO change, global inflation, contagion variable (cont_stop), global liquidity (money_global), and long-run interest rates for the EU and US.

Using the selected features, the mean ROC curve for the designed Random Forest model is depicted in the **Figure 5.1**. The model achieves a mean AUC score of 0.79 after 10-fold stratified cross-validation.

In selecting an operating point at the ROC curve, we fixed the false positive rate at 0.2 and achieved the true positive rate, i.e., recall, of 0.65. This operation point is found to be a good compromise between Sudden Stop (SS) detection sensitivity and

false alarms. The 10-fold cross-validation ROC graph with mean and standard deviation of ROC has been included in the Appendix C for reference.
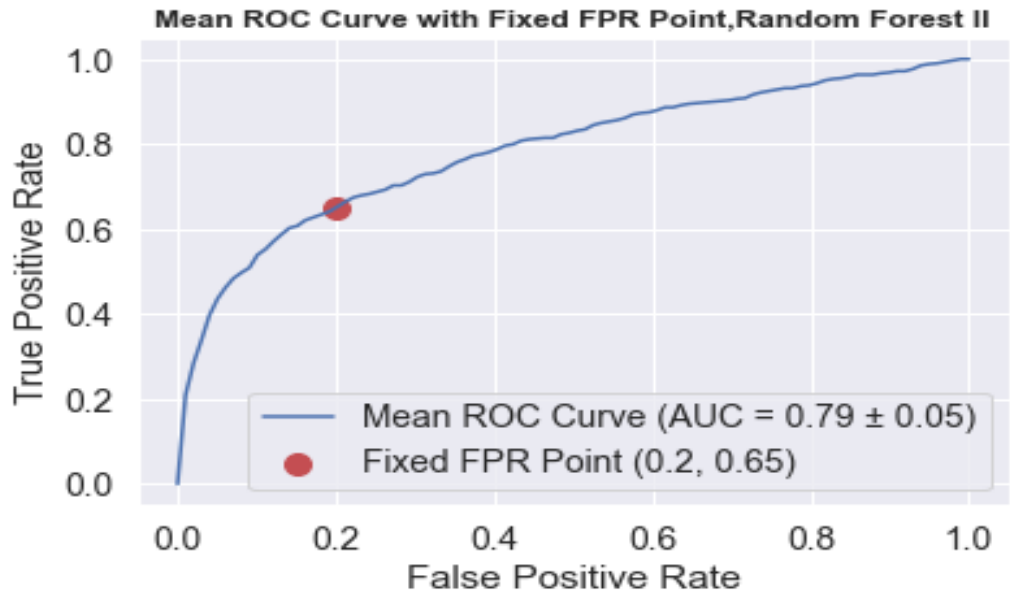
Figure 5. 1. Mean ROC Curve with Fixed FPR Point, Random Forest II
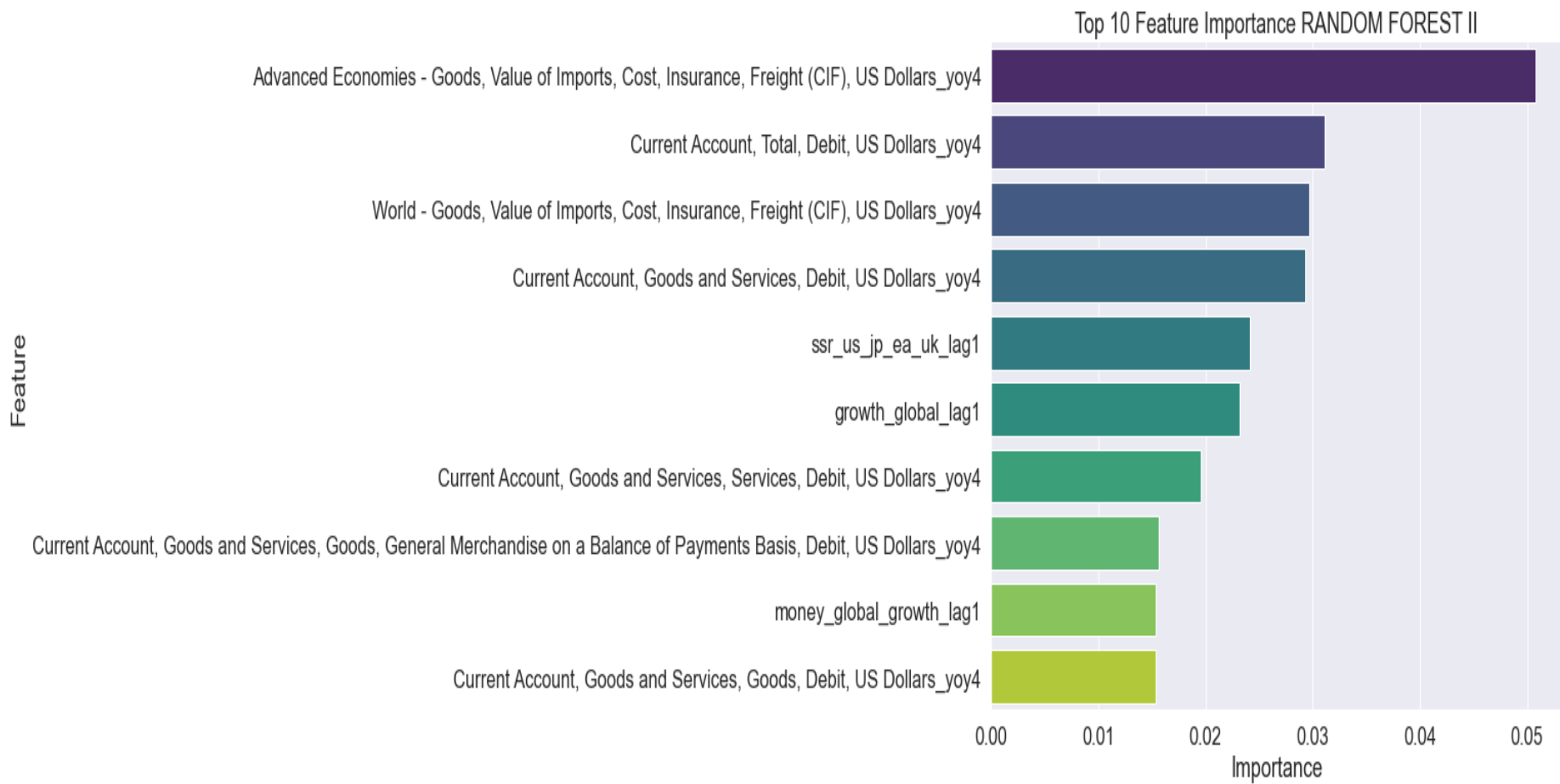
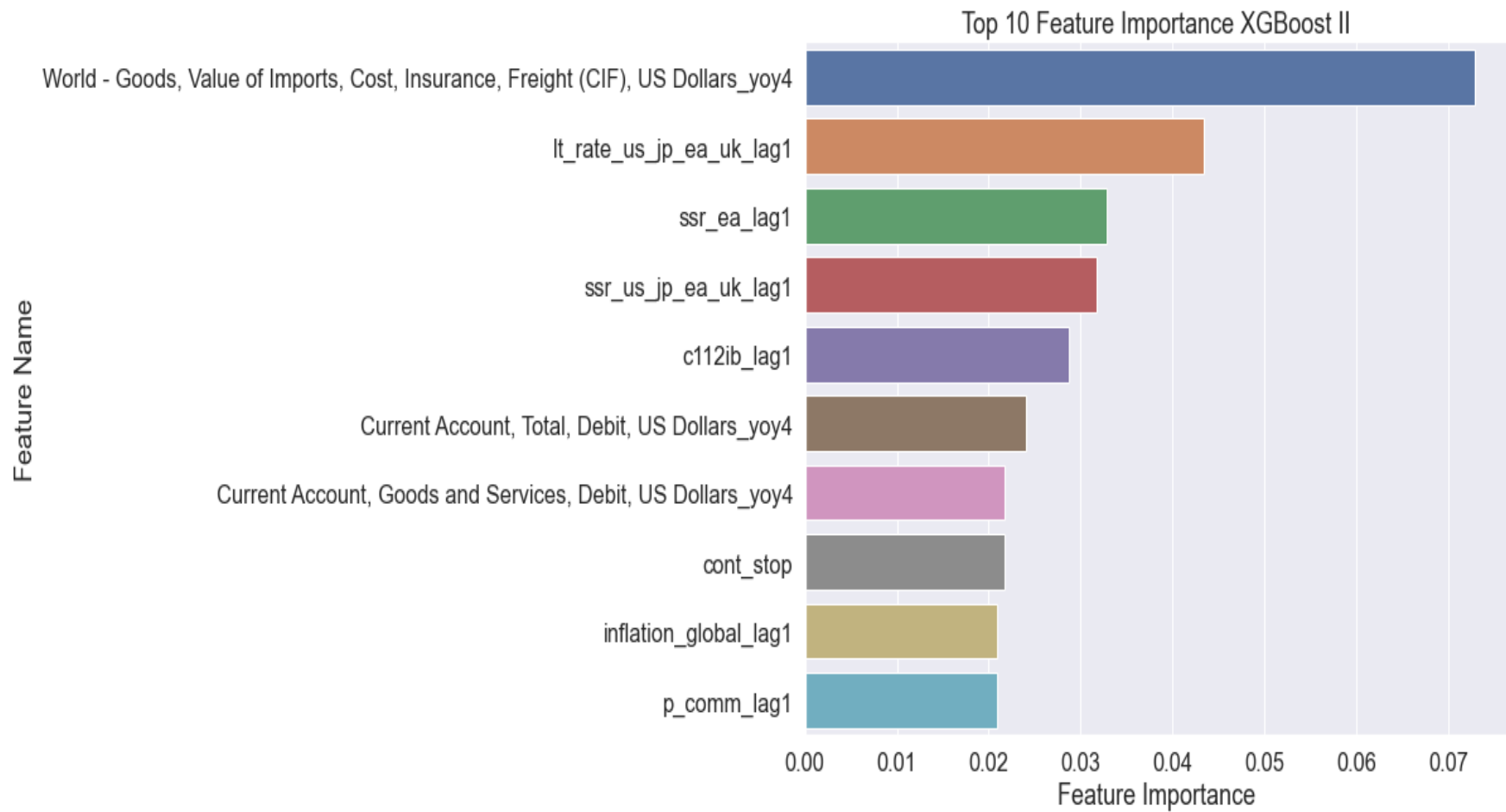**Figure 5. 2.** Top 10 Features by Feature Importance, Random Forest II

5.2.2. XGBoost (Extreme Gradient Boost)

Similarly, in the case of XGBoost, we employed Recursive Feature Elimination with Cross-Validation (RFECV) and determined the optimal number of features to be 100. After retraining the model with the best parameters obtained through hyperparameter tuning, we applied the feature importance method specific to XGBoost.

As illustrated in **Figure 5.4,** the most influential predictor identified is the year-over-year percentage change in the country's goods trade worldwide, specifically the variable 'World - Goods, Value of Imports, Cost, Insurance, Freight (CIF), US Dollars.' This is closely followed by the average of long-run interest rates for the EU, UK, US, and Japan, along with the shadow short-run interest rates of the EU and the average of shadow short-run interest rates for the EU, UK, US, and Japan.

The mean ROC curve for the designed XGBoost model is depicted in the **Figure 5.3**. The model achieves a mean AUC score of 0.83 after 10-fold stratified cross-validation. In selecting an

Operating point at the ROC curve, we fixed the false positive rate at 0.2 and achieved the true positive rate, i.e., recall, of 0.69. This operation point is found to be a good compromise between Sudden Stop (SS) detection sensitivity and false alarms. The 10-fold cross-validation ROC curves with mean ROC and its standard deviation has been included in the Appendix C for reference.

Figure 5. 3. Top 10 Features by Feature Importance, XGBoost II
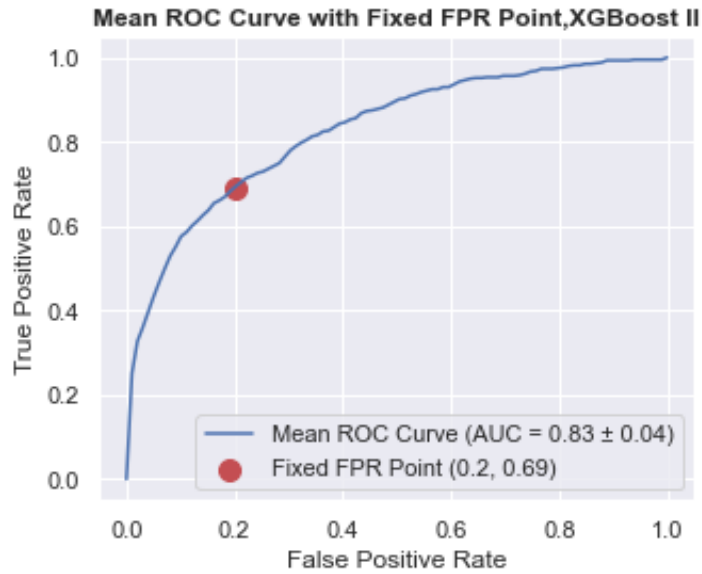
Figure 5. 4. Mean ROC Curve with Fixed FPR Point, XGBoost II

5.2.3. Support Vector Machines (SVM)

The RFECV for SVM optimizes with 50 features; however, SVM exhibits notably poor performance, with a mean AUC score of 0.51, as illustrated in **Figure 5.5**. It's crucial to note that SVM doesn't inherently provide a feature importance method. Nevertheless, we utilized a linear kernel through grid search. In an effort to discern variable importance, we compared the features based on the absolute values of coefficients. The detailed feature importance can be found in the Appendix C.
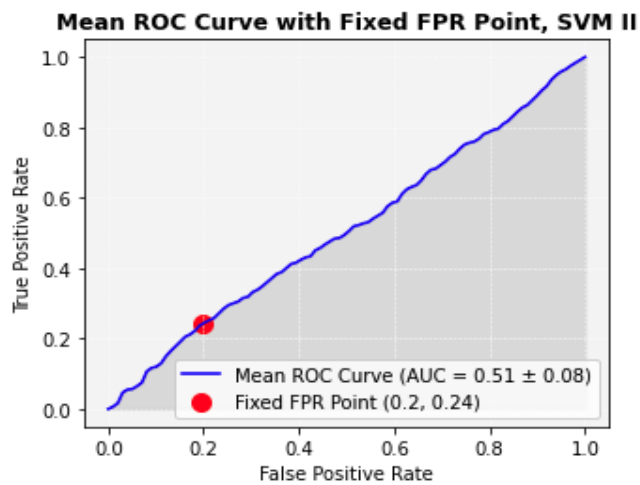
Figure 5. 5. Mean ROC Curve with Fixed FPR Point, SVM II

133

5.2.4. Elastic Net

For the Elastic Net method, a regularized logistic regression, we initially employed Lasso for variable selection, resulting in the choice of 56 variables. Subsequently, we retrained the Elastic Net. The feature importance analysis was conducted by comparing the absolute values of coefficients, with the top 10 variables presented in the Appendix C. Among the preselected model-based variables, the top 30 include shadow short-run interest rates of the UK, oil prices, global risk (VXO), contagion, real GDP, and the EM dummy variable. The mean AUC score is 0.60, as depicted in **Figure 5.6.**
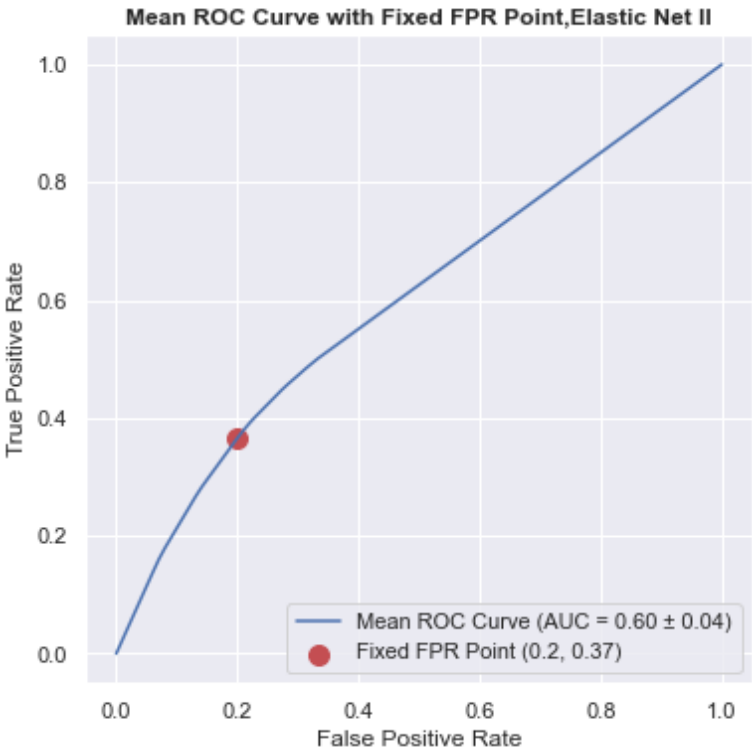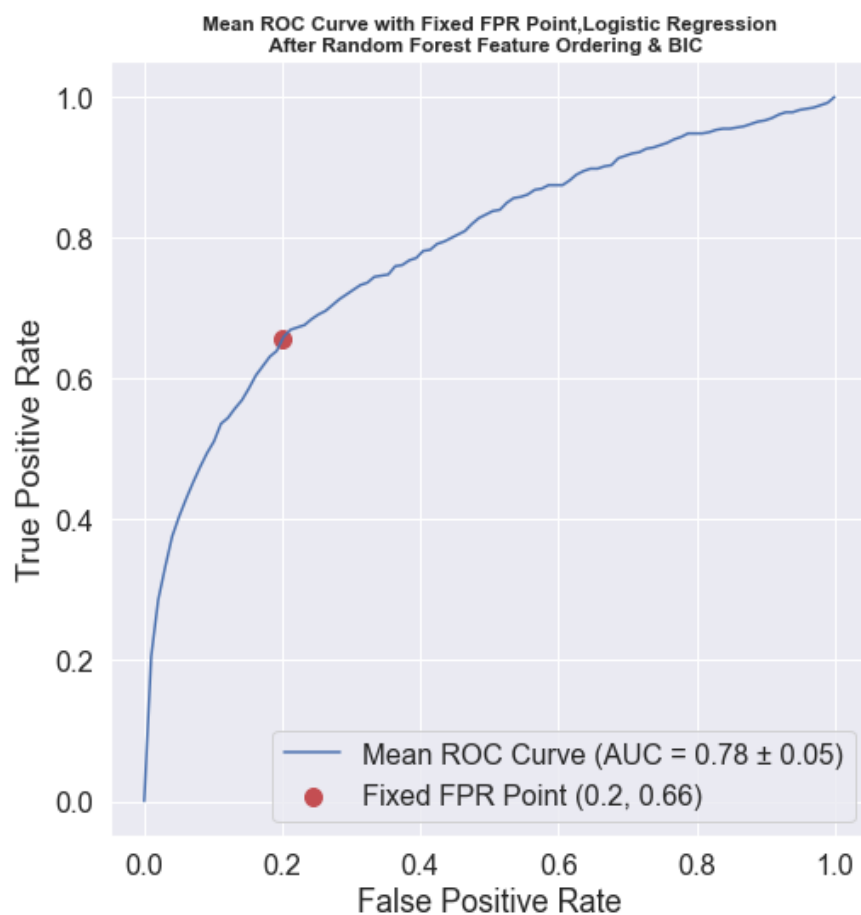

Figure 5. 6. Mean ROC Curve with Fixed FPR Point, Elastic Net II

5.2.5. Logistic Regression

We initially ranked the variables using Random Forest estimates, followed by a feature selection process to obtain the ranking. Subsequently, we applied the Bayesian Information Criteria (BIC), resulting in the selection of the first 16 features.

After performing logistic regression, we determined feature importance by considering the absolute values of coefficients. *'World - Goods, Value of Imports, Cost, Insurance, Freight (CIF), US Dollars'* emerged as the most crucial predictor. Notably, 9 out of the 16 selected variables were drawn from the predetermined set, including real GDP, global growth, the contagion variable (*cont_stop*), global risk (*VIX*), shadow short-run interest rates of Japan, the average of shadow short-run interest rates of Japan, EU, UK, US, and the long-run interest rate of the US, as well as global liquidity.

Figure 5. 7. Mean ROC Curve with Fixed FPR Point Logistic Regression

The mean ROC curve for the designed Logistic Regression model is depicted in the **Figure 5.7**. The model achieves a mean AUC score of 0.78 after 10-fold stratified cross-validation. In selecting an operating point at the ROC curve, we fixed the false positive rate at 0.2 and achieved the true positive rate, i.e., recall, of 0.66. This

operation point is found to be a good compromise between Sudden Stop (SS) detection sensitivity and false alarms.

5.2.6. Long Short Run Memory (LSTM)

Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber (1997), address the vanishing gradient problem encountered in training the traditional recurrent neural networks (RNNs) on long sequences of data. This architectural innovation has since become a cornerstone in deep learning for sequential data processing. Unlike conventional neural networks, LSTMs excel in decoding patterns within extended sequences.

At its core, an LSTM comprises a Cell State, functioning as persistent memory, and a Hidden State, serving as short-term memory. Working in tandem with Forget, Input, and Output Gates, these elements selectively retain or discard information, enhancing the model's understanding of temporal dependencies.

LSTM iteratively performs the following key operations:

Forget Gate Operation: Determines the relevance of information, selectively retaining or discarding details.

Input Gate Operation: Integrates new information to enhance the model's comprehension.

Updating the Cell State: Merges insights from the prior state with new information.

Output Gate Operation: Transfers pertinent information to the hidden state, shaping the model's immediate understanding.

Practically, LSTMs excel in unraveling complex patterns and dependencies within sequential data. They overcome challenges like the 'vanishing gradient,' making them crucial for tasks such as natural language processing, speech recognition, and time-based trend prediction.

136

In summary, LSTMs function as intelligent memory systems in neural networks, adeptly navigating sequential data intricacies and selectively retaining valuable insights over extended periods. This unique capability positions LSTMs as powerful tools in unraveling the nuances of sequential data, significantly contributing to the field of Machine Learning.

The estimation strategy employed by LSTM sets it apart from traditional models in terms of forecasting. Unlike traditional ML methods, which often utilize non-sequential features for training and validation, LSTM focuses on time-dependent sequences for prediction.

Traditional ML models typically employ cross-validation by splitting the dataset into subsets for training and validation randomly. This approach allows the model to learn from various parts of the data, providing a comprehensive understanding of its performance across different scenarios. In contrast, LSTM models, designed for sequential learning over time, may require a time series-aware cross-validation approach to preserve the temporal order of the data during validation.

Specifically, our LSTM model is designed for forecasting Sudden Stop events. The training strategy involves using data from four consecutive quarters leading up to a specific time 't'. This sequential training approach captures temporal dependencies and patterns within the data that may be crucial for accurate predictions.

In contrast to traditional models that may not explicitly consider time dependencies, LSTM's ability to analyze sequences makes it well-suited for tasks where historical context plays a significant role. The model learns patterns across multiple time steps, enabling it to make predictions based on the evolving nature of the data.

For LSTM training and testing, we constructed a time-series dataset comprising a total of 225 variables. These variables exclusively represent the current values of our extended dataset, spanning five consecutive quarters from the historical IMF dataset, covering the period from 1994 to 2018.

The whole training dataset is then separated into training and testing time-series datasets as part of our regular 10-fold cross-validation and testing procedure. We adopted a sequence-to sequence LSTM network architecture with 50 hidden units, followed by a fully-connected (dense) layer, and a softmax layer that makes a binary classification. For each fold, the LSTM model has been trained on the 90% of the whole data and tested on the remaining 10% of the independent data. We calculated the accuracy metrics and ROC Curves per fold, and averaged over all the folds as we did before.Therefore, the key difference lies in LSTM's focus on time-ordered data and its capability to capture temporal dependencies, providing a more nuanced understanding of dynamic processes such as Sudden Stop events. The LSTM model demonstrates a strong performance with an AUC score of 0.91, indicating its effectiveness in distinguishing between positive and negative instances. Additionally, at our reference false positive rate of 0.2, the model achieves an impressive true positive rate of 0.85. This high true positive rate suggests that the LSTM model excels at correctly identifying instances of interest while controlling the false positive rate at the specified threshold.

In summary, as depicted in **Figure 5.8**, the AUC score of 0.91 reflects the overall discriminatory power of the LSTM model, and the specific true positive rate of 0.85 at a fixed false positive rate of 0.2 highlights its capability to accurately predict Sudden Stop events, showcasing its robust performance in the task at hand.
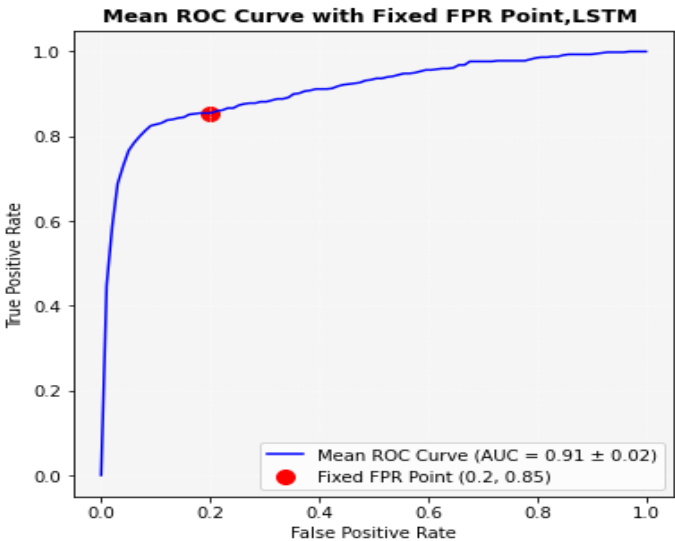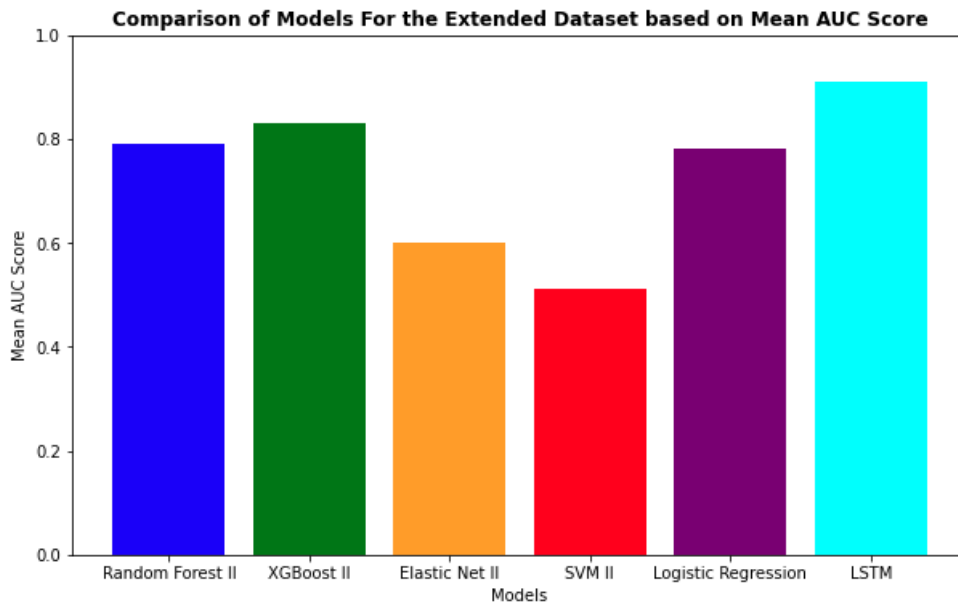

Figure 5. 8. Mean ROC Curve with Fixed FPR Point, LSTM

5.3. Model Selection

In the analysis of the expanded dataset, XGBoost stands out as a superior performer among traditional ML methods, demonstrating exceptional out-of-sample performance. It attains a noteworthy mean AUC score of 0.83, and a recall of 0.69 for a specified 0.2 false positive rate, as illustrated in **Figure 5.9** and **Figure 5.10**, respectively. These results align with our earlier findings when working with a limited number of features, where both XGBoost and RF were identified as top performers.

In contrast, SVM exhibits the poorest performance, yielding a mean AUC of 0.51, while Elastic Net falls short with a mean AUC of 0.60. Notably, a hybrid approach featuring Logistic Regression proves to be effective, securing the third-best performance with a mean AUC of 0.78. The second-best model is RF, boasting a mean AUC of 0.79.

These results suggest that in cases with a high number of features, traditional logistic regression, being a linear model, can be enhanced by incorporating powerful tree-based methods like Random Forest and XGBoost. These non-linear, non-parametric methods not only improve performance but also provide effective feature selection capabilities.

Figure 5. 9. Mean AUC Scores Across Different Models

Upon introducing temporal dependencies, the LSTM model takes the lead with a mean AUC of 0.91, surpassing all traditional models. This emphasizes its proficiency in capturing temporal patterns. Additionally, the LSTM model achieves a recall of 0.85 for a specified 0.2 false positive rate, highlighting its effectiveness. This underscores the LSTM's efficacy in providing a nuanced understanding of the probability of Sudden Stop events, emphasizing the need for advanced deep learning techniques in time-series forecasting. The findings contribute valuable insights to the financial domain, suggesting avenues for more sophisticated and accurate predictive models.

Model	True Positive Rate at FPR=0.2
LSTM	0.8548
XGBoost II	0.6922
Logistic Regression	0.6555
Random Forest II	0.6503
Elastic Net II	0.3657
SVM II	0.2437

Figure 5. 10. Recall Score Across Different Models for False Positive Rate of 0.2

5.4. Summary

In this chapter, we focus on leveraging Machine Learning Methods for variable selection using quarterly data from the International Monetary Fund (IMF). Variables are chosen based on data availability, preferring those with fewer than 10% missing values among selected countries from prior analysis.

The data augmentation process involves combining information from different datasets, resulting in 192 refined variables. Enhancements include deriving lag versions and capturing changes, thereby enriching the dataset to a total of 768 exogenous variables. Additionally, augmentation introduces variables from prior

estimations and other model-based variables from Forbes & Warnock (2021). The dataset spans from 1994q1 to 2018q4.

Except with the Norway and Hong Kong, the countries remain same with the previous analysis. The dataset comprises a total of 4066 instances, with 599 of them identified as Sudden Stop Events. Diverse machine learning models, including Random Forest, XGBoost, SVM, Elastic Net, and Logistic Regression, are deployed with model-specific variable selection methods. Post-selection, estimation methods remain consistent, employing Stratified K-fold Cross Validation and grid search for hyperparameter tuning. For these models, we excluded the current values of variables and utilized the 1-period lagged, rate of change in the previous period, and year-over-year percentage change of the variables from the initial 192 obtained from the IMF, as mentioned earlier. Consequently, the final set of variables derived from this process totaled 576. We supplemented this set with the 1-period lagged versions of the 28 variables selected from Forbes & Warnock (2021), resulting in a total exogenous variable set of 604.

The addition of the Long Short-Term Memory (LSTM) model, a deep learning technique, sets the chapter apart. LSTM's emphasis on temporal dependencies for forecasting distinguishes it from traditional models. Subsequent sections provide a concise RFECV summary and detailed insights into each model, concluding with a comprehensive comparison. For LSTM, we utilized a total of 225 variables with a 5-quarters span, which includes the current and past 4 quarters data. The LSTM model, at inference time, has used the past 4 quarters data along with the current quarters data to predict the SS event for the next quarter.

We employed various variable selection techniques for our models. In terms of feature selection, Random Forest identified 131 features, XGBoost selected 100 features, and SVM selected 50, all using the RFECV. Elastic Net, through using Lasso selection, identified 60 features, while Logistic Regression, guided by Bayesian Information Criteria (BIC), narrowed down to 16 features.

In the analysis of the expanded dataset, XGBoost emerges as a standout performer among traditional ML methods, showcasing superior out-of-sample performance. It

achieves a notable mean AUC score of 0.83 and a recall of 0.69 for a specified 0.2 false positive rate. This aligns with our earlier findings when working with a limited number of features, where both XGBoost and RF were identified as top performers.

In contrast, SVM exhibits the poorest performance, yielding a mean AUC of 0.51, while Elastic Net falls short with a mean AUC of 0.60. Notably, a hybrid approach featuring Logistic Regression proves to be effective, securing the third-best performance with a mean AUC of 0.78. The second-best model is RF, boasting a mean AUC of 0.79.

These results suggest that in cases with a high number of features, traditional logistic regression, being a linear model, can be enhanced by incorporating powerful tree-based methods like Random Forest and XGBoost. These non-linear, non-parametric methods not only improve performance but also provide effective feature selection capabilities.

Upon introducing temporal dependencies, the LSTM model takes the lead with a mean AUC of 0.91, surpassing all traditional models. This emphasizes its proficiency in capturing temporal patterns. Additionally, the LSTM model achieves a recall of 0.85 for a specified 0.2 false positive rate, highlighting its effectiveness. This underscores the LSTM's efficacy in providing a nuanced understanding of the probability of Sudden Stop events, emphasizing the need for advanced deep learning techniques in time-series forecasting.

CHAPTER 6

CONCLUSION

In this study, our objective was to leverage Machine Learning for predicting Sudden Stop Crises. Empirical studies in literature highlights that Sudden Stop remains a persistent issue for both advanced and emerging economies. Understanding the dynamics and predictive power behind it continues to be of paramount importance. Therefore, we aimed to explore machine learning methods, known for their superior out-of-sample prediction performance, as an alternative approach to addressing this problem.

To achieve this, we employed a diverse set of supervised machine learning techniques, comparing their out-of-sample prediction power and generalizability. The goal was to identify machine learning tools tailored to our specific problems and assess their ability to provide a highly predictive model. Our estimation is structured into two chapters. In the first part, in Chapter 4, we utilized a set of traditional machine learning methods with a limited number of pre-selected variables using Forbes & Warnock (2021) dataset, making comparisons using appropriate prediction performance metrics. Moving on to the second part, Chapter 5, we expanded the dataset sourced from the IMF. The extended dataset, driven by data availability, prioritized variables with fewer than 10% missing values from quarterly datasets of 51 countries (1986 to 2018) obtained from the IMF. The refined set of 192 variables was curated by integrating features from BOP, IFS, and DOT, excluding redundancies. Feature enhancements included 1-quarter lag versions, year-over-year percentage changes, and rate of change, resulting in 768 exogenous variables.

Variables from BOP and IFS covered components like the capital and current accounts, total international reserves, and international liquidity. Trade-related variables from the DOT dataset included imports, exports, and trade balance.

143

Consistency was maintained for 49 countries from 1994q1 to 2018q4. The dataset comprises a total of 4066 instances, with 599 of them identified as Sudden Stop Events. The augmentation phase incorporated variables from previous estimations, global liquidity, global risk, long-run interest rates, global growth, local GDP growth, contagion, and model-based variables from Forbes & Warnock (2021). The complete variable list is available in the Appendix A.

Here, we also capitalized on machine learning-based feature selection methods to identify important variables, conducting feature selections as well as building models with the selected features, using a set of machine learning techniques and making comparisons as previously outlined. The following paragraphs provide a summary, estimation results, and the model selection for each part subsequently.

In Chapter 4, we undertook several key steps in the prediction of Sudden Stop events. Firstly, we leverage the capital flow dataset curated by Forbes and Warnock (2021) to identify Sudden Stop Crises in 59 countries. This comprehensive dataset spans both advanced and emerging nations, covering the period from 1978 Q1 to 2020 Q3. Following this, we replicated the estimation process for the base case presented in their analysis. This involved employing the complementary log-log model and scrutinizing its out-of-sample performance, establishing it as our baseline scenario.

Secondly, we explicitly framed our prediction problem as a comparison of out-of-sample performances, distinct from a parameter estimation or causal inference problem. We aimed to highlight the implementation and estimation strategies for several key classes of ML methods, briefly touching upon sample division, cross-validation techniques, hyperparameter tuning, and feature scaling.

Thirdly, we provided concise, non-technical summaries of selected methods, including Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Multi-Layer Perceptron (MLP), and ensemble methods such as Random Forest, XGBoost, and AdaBoost. Each method brings a unique approach to addressing various challenges and caters to specific data types. Technical references for these methods were also provided.

After elucidating our model selection strategy and referencing performance metrics, we proceeded to evaluate mean AUC scores and scrutinize recall (sensitivity or true positive rate) as an additional performance metric. Additionally, we compare recall (true positive rate) while fixing the false alarm rate.

The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) serves as a crucial performance metric in binary classification tasks, including modeling sudden stop crises. The ROC curve graphically represents the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various classification thresholds. AUC-ROC quantifies the discriminatory power of a classification model by measuring the area under this curve. Higher AUC-ROC values, ranging from 0 to 1, indicate superior model performance. The ROC curve aids in model selection by illustrating the trade-offs between sensitivity and specificity at different threshold settings. A model with better discriminatory ability exhibits an ROC curve that approaches the upper-left corner of the plot, leading to a higher AUC-ROC value.

When comparing AUC-ROC scores, both k Neighbors (kNN) and Multi-Layer Perceptron (MLP) exhibit inferior performance compared to the base model (cloglog model), while SVM and Elastic Net demonstrate similar mean AUC scores. In contrast, tree-based models—Random Forest, XGBoost, and AdaBoost—outperform both the base model (traditional complementary log-log) and the other ML methods. This suggests that the base model shows relative proficiency in distinguishing between positive and negative classes, resulting in moderately good overall performance. Notably, it exhibits poor recall performance, registering at 0.16, indicating a high likelihood of misidentifying crisis times as non-crisis. This underscores a notable Type II error, where the model fails to identify instances of actual crises, negatively impacting its recall performance. In terms of recall, the base model performs the worst.

The rationale behind selecting recall as an additional metric for comparison, among others such as accuracy, precision, and F1, is as follows. First, Accuracy may not be a suitable metric for assessing the performance of models in predicting rare events, such as Sudden Stop Crises, due to the inherent imbalance in the dataset. In

situations where one class significantly outnumbers the other, as is common with crisis prediction, accuracy can be misleading. Models might achieve high accuracy by simply predicting the majority class, failing to capture the rare events of interest. This metric doesn't distinguish between correctly predicting non-events and correctly predicting events, providing a false impression of a well-performing model.

Similarly, Precision may not be a good metric to rely on in the case of rare events like predicting Sudden Stop Crises due to its sensitivity to imbalanced datasets. In situations where rare events are infrequent compared to non-events (as is often the case with crisis prediction), precision can be misleading.

Precision is calculated as the ratio of true positive predictions to the sum of true positives and false positives. In the context of rare events, where the majority of cases are non-events, a high precision score may still be achieved by correctly predicting a few rare events but misclassifying a significant number of non-events as positive.

This issue arises because precision doesn't account for the true negatives (correctly predicted non-events) and can give a falsely optimistic view of the model's performance. In the case of rare events, it's crucial to consider the overall performance, including the ability to correctly identify both positive and negative instances.

It is crucial to recognize the inherent tradeoff between recall and precision in classification tasks. This balance requires careful consideration of minimizing false positives and false negatives. Recall, measuring the model's ability to capture all actual positive instances, prioritizes avoiding false negatives. Precision, on the other hand, focuses on minimizing false positives, assessing the accuracy of positive predictions. The tradeoff emerges because enhancing one metric often comes at the cost of the other. A higher classification threshold improves precision but may reduce recall, making the model more selective. Conversely, a lower threshold enhances recall but might lower precision, leading to a more inclusive model. The decision between prioritizing high precision or high recall hinges on the specific

goals and constraints of the task. In the context of predicting Sudden Stop events, where missing such events is potentially more detrimental than false positive alarms, we prioritize recall as a key metric for comparison.

Furthermore, since the ROC curve illustrates the trade-off between the true positive rate (sensitivity) and false positive rate (1 - specificity), it is essential to understand the inherent balance between these two parameters. This trade-off arises from the fact that adjusting the classification threshold influences the model's ability to correctly identify positive instances while simultaneously misclassifying negative instances. A typical threshold setting is 0.5, meaning that instances with predicted probabilities greater than or equal to 0.5 are classified as positive. However, this threshold is not universally optimal for all scenarios. Consequently, the choice of threshold has a direct impact on the model's operating point in the trade-off space. For instance, setting a specific false positive rate, such as 0.2, allows us to explore how different models perform in terms of recall, thereby providing insights into their ability to identify true positive instances while controlling the false positive rate according to a predefined criterion.

As the final step in model selection, we fix the mean false positive rate at 0.2, setting a tolerance level of only 20 percent misclassification of non-crisis (normal) events as crises. This will allow the model, on average, to misidentify non-crisis times as Sudden Stop crises 20% of the time. Subsequently, we examined the corresponding mean true positive rates (recall) on their individual ROC curves.

The choice to set the mean false positive rate at 0.2 during model evaluation is driven by several considerations specific to the practical requirements of the application. In the context of imbalanced data, particularly when dealing with rare events like sudden stop crises, the default threshold of 0.5 may inadequately address the need for nuanced performance evaluation. By fixing the false positive rate at a certain value (e.g., 0.2), the assessment becomes tailored to scenarios where controlling false positives is paramount. This approach not only aligns with practical considerations of limiting misclassifications but also allows for a more nuanced examination of a model's performance under conditions reflective of the real-world application.

The results align with the earlier comparison of mean AUC scores, confirming the superior performance of tree-based methods over other models and the base model.

After identifying XGBoost as the top-performing model, we delve into understanding which variables are most influential in predicting Sudden Stops. Utilizing both the feature importance method and Shapley values, we acknowledge differences in methodology and results. Feature importance elucidates the contribution of each variable to the predictive power of the model, leveraging the Gini impurity metric. It is crucial to note that this metric doesn't infer the causes of Sudden Stop Crises but rather identifies predictors. The resulting scores range between 0 and 1, with a higher value indicating a more substantial impact on predictions. The feature importance of the trained XGBoost model highlights Contagion, Global Growth, and Global Risk (VXO) as the most influential predictors. These variables play a significant role in shaping the model's predictive capabilities, offering valuable information for understanding and interpreting the factors contributing to Sudden Stop Crises.

On the other hand, Shapley Values, rooted in cooperative game theory, take a collaborative approach to understanding variable contributions. They meticulously consider the interactions and non-linearities among features, providing a more nuanced perspective compared to traditional feature importance metrics. Shapley values distribute the predictive contribution of each variable across all possible combinations, allowing for a comprehensive evaluation of individual and joint impacts. While feature importance gives a quick overview of influential predictors, Shapley values excel in capturing intricate relationships and dependencies within the model. Moreover, Shapley values inherently ensure that the sum of contributions across all variables aligns with the model's overall prediction, providing a more internally consistent measure. Shapley values reveal that Real GDP, Global Risk (VXO), and Contagion are the best predictors, assigning a higher value to Real GDP in contrast to the feature importance.

In Chapter 5, we focus on leveraging Machine Learning Methods for variable selection using quarterly data from the International Monetary Fund (IMF). Variables are chosen based on data availability, preferring those with fewer than 10% missing values among selected countries from prior analysis.

The data augmentation process involves combining information from different datasets, resulting in 192 refined variables. Enhancements include deriving lag versions and capturing changes, thereby enriching the dataset to a total of 768 exogenous variables. Additionally, augmentation introduces variables from prior estimations and other model-based variables from Forbes & Warnock (2021). The dataset spans from 1994q1 to 2018q4.

Except with the Norway and Hong Kong, the countries remain same with the previous analysis. The dataset comprises a total of 4066 instances, with 599 of them identified as Sudden Stop Events. Diverse machine learning models, including Random Forest, XGBoost, SVM, Elastic Net, and Logistic Regression, are deployed with model-specific variable selection methods. Post-selection, estimation methods remain consistent, employing Stratified K-fold Cross Validation and grid search for hyperparameter tuning. For these models, we exclude the current values of variables and utilize the 1-period lagged, rate of change in the previous period, and year-over-year percentage change of the variables from the initial 192 obtained from the IMF, as mentioned earlier. Consequently, the final set of variables are derived from this process totaled 576. We supplement this set with the 1-period lagged versions of the 28 variables selected from Forbes & Warnock (2021), resulting in a total exogenous variable set of 604.

The addition of the Long Short-Term Memory (LSTM) model, a deep learning technique, sets the chapter apart. LSTM's emphasis on temporal dependencies for forecasting distinguishes it from traditional models. Subsequent sections provide a concise RFECV summary and detailed insights into each model, concluding with a comprehensive comparison. For LSTM, we utilize a total of 225 variables with a 5-quarters span, which includes the current and past 4 quarters data. The LSTM model, at inference time, has used the past 4 quarters data along with the current quarters data to predict the SS event for the next quarter.

We employ various variable selection techniques for our models. In terms of feature selection, Random Forest identified 131 features, XGBoost selected 100 features, and SVM selected 50, all using the RFECV. Elastic Net, through using Lasso

selection, identified 60 features, while Logistic Regression, guided by Bayesian Information Criteria (BIC), narrowed down to 16 features.

In the analysis of the expanded dataset, XGBoost emerges as a standout performer among traditional ML methods, showcasing superior out-of-sample performance. It achieves a notable mean AUC score of 0.83 and a recall of 0.69 for a specified 0.2 false positive rate. This aligns with our earlier findings when working with a limited number of features, where both XGBoost and RF were identified as top performers.

In contrast, SVM exhibits the poorest performance, yielding a mean AUC of 0.51, while Elastic Net falls short with a mean AUC of 0.60. Notably, a hybrid approach featuring Logistic Regression proves to be effective, securing the third-best performance with a mean AUC of 0.78. The second-best model is Random Forest, boasting a mean AUC of 0.79.

These results suggest that in cases with a high number of features, traditional logistic regression, being a linear model, can be enhanced by incorporating powerful tree-based methods like Random Forest and XGBoost. These non-linear, non-parametric methods not only improve performance but also provide effective feature selection capabilities.

Upon introducing temporal dependencies, the LSTM model takes the lead with a mean AUC of 0.91, surpassing all traditional models. This emphasizes its proficiency in capturing temporal patterns. Additionally, the LSTM model achieves a recall of 0.85 for a specified 0.2 false positive rate, highlighting its effectiveness. This underscores the LSTM's efficiency in providing a nuanced understanding of the probability of Sudden Stop events, emphasizing the need for advanced deep learning techniques in time-series forecasting.

In comparing common methods across two chapters of the thesis, a detailed analysis unveils intriguing insights into the impact of feature augmentation on model performance. In the initial chapter, where six model-driven exogenous variables were employed, four key methods—Random Forest, XGBoost, Elastic Net, and Support

Vector Machine (SVM)—underwent scrutiny for their predictive capabilities. Notably, Random Forest and XGBoost emerged as strong contenders, showcasing comparable performance with AUC scores of 0.77 and promising recall rates at a 0.2 false positive rate. However, Elastic Net and SVM exhibited slightly lower performance metrics, hinting at potential limitations in handling the complexity of the dataset. In the subsequent chapter, with the expansion of exogenous variables, the performance landscape shifted. While Random Forest and XGBoost continued to demonstrate enhanced predictive power, with notable increases in AUC scores to 0.79 and 0.83, respectively, Elastic Net experienced a decline in performance, evidenced by a reduced AUC score of 0.60. Similarly, SVM's performance diminished significantly, highlighting challenges in adapting to the expanded feature space. Notably, XGBoost emerged as the top-performing model, boasting both high AUC scores and superior recall rates. This comprehensive comparison underscores the nuanced interplay between feature selection, model complexity, and predictive performance, emphasizing the importance of iterative refinement and adaptation in predictive modeling endeavors.

The consistency between our findings and empirical studies in the literature regarding the performance of XGBoost and Random Forest is noteworthy. Across both chapters of our thesis, these algorithms have demonstrated robust predictive capabilities, as evidenced by their high AUC scores and recall rates. These results align with the prevailing understanding in the literature, which often highlights the effectiveness of XGBoost and Random Forest in various predictive modeling tasks. However, it is essential to note that in the second part, with an extended dataset, we explored the use of Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), for pure forecasting purposes. Remarkably, LSTM emerged as the top-performing model in this context, achieving an impressive AUC of 0.91 and a recall of 0.82 for a given 0.2 false positive rate (FPR). This highlights the potential of deep learning approaches, such as LSTM, in capturing temporal dependencies and achieving superior predictive performance, particularly in scenarios with extended datasets and time-series data. Therefore, while XGBoost and Random Forest remain established choices for predictive modeling, our findings

underscore the importance of considering alternative approaches, such as LSTM, for specific forecasting tasks, where temporal dynamics play a crucial role.[17]

In traditional binary classification tasks, it is common to evaluate models' sensitivity (i.e, True Postive Rate) using a fixed False Positive Rate (FPR) as a standard benchmark since there is always a trade-off between the models' sensitivity and false alarms. We initially assessed the performance of the trained models at the FPR of 0.2, and later extended the analysis to include additional FPR values of 0.05 and 0.1, for a more comprehensive assessment. By examining the performance of the models at three different FPRs—0.05, 0.1, and 0.2—we gained a deeper understanding of their behavior across varying decision boundaries.

When we fix the FPR at 0.05, for instance, we emphasize a stringent control over false positives, forcing the model to be more conservative in its predictions. Conversely, fixing the FPR at 0.2 allows for a higher tolerance of false positives, resulting in a higher True Positive Rate (TPR) but at the expense of more false alarms.

This broader analysis offers practical insights into the models' suitability for real-world applications where the costs or impacts of false positives and false negatives differ significantly. For instance, in medical diagnostics, a lower FPR is often crucial to avoid misdiagnosing a healthy patient as having a disease. On the other hand, in an email spam detection system, a higher FPR might be more tolerable to ensure that important emails are not incorrectly marked as spam.

Evaluating models at these various fixed FPRs enables us to provide nuanced recommendations based on the specific requirements of the task. In our analysis, we find that XGBoost and LSTM consistently exhibit high TPRs across different fixed FPRs, suggesting their effectiveness in capturing true positives while maintaining

---

[17] While methods like XGBoost provide powerful predictive capabilities and offer insights into feature importance through techniques like SHAP values, their interpretability allows for a deeper understanding of the underlying relationships in the data. On the other hand, deep learning methods like LSTM, while capable of achieving superior predictive performance, often operate as black boxes, making it challenging to interpret their inner workings. Thus, while the adoption of deep learning methods may lead to improved prediction accuracy, it comes at the cost of reduced interpretability.

acceptable false positive rates. This approach allows us to suggest the most suitable model for our classification task based on the application's needs and priorities. In conclusion, this comparison of models at different fixed FPRs provides actionable insights, enabling practitioners to make informed decisions about model selection for their specific applications.

In Chapter 4, the results revealed that XGBoost, ADABoost, and Random Forest consistently demonstrated higher recall rates compared to the complementary log-log model. These models also performed similarly to SVM and Elastic NET, showcasing their effectiveness in capturing true positives. On the other hand, MLP and KNN exhibited lower recall rates in comparison. Upon further analysis at False Positive Rates (FPRs) of 0.05 and 0.1, the ranking of these models remained unchanged, as shown in Table D.1, Table D.2 and Figure D.1 in the Appendix D. Specifically, XGBoost, ADABoost, and Random Forest maintained their top positions. To provide additional insight, mean Receiver Operating Characteristic (ROC) curves of the models at FPR 0.05 were shared, illustrating their comparative performance in distinguishing between true and false positives. This reaffirms the effectiveness of XGBoost, AdaBoost, and Random Forest in achieving higher recall rates, making them suitable candidates for the classification task. Further exploration into these models could lead to enhanced performance and valuable insights for practical applications.

In Chapter 5, the analysis revealed that LSTM outperformed all other models, with XGBoost and Random Forest closely following at the fixed alarm rate of 0.2. To further investigate model performance, we conducted additional evaluations at different false positive rates (FPRs) of 0.05 and 0.1. Remarkably, the model rankings remained consistent across these varying FPRs, as shown in Table D.3, Table D.4, and Figure D.2 in the Appendix D. This consistency in rankings provides robust evidence supporting the superior performance of LSTM, followed by XGBoost and Random Forest, for the given classification task.

In summary, our research endeavors to address the persistent challenge of predicting Sudden Stop Crises in both advanced and emerging economies through the

innovative application of Machine Learning (ML) techniques. As part of an ongoing discourse on the effectiveness of ML in macroeconomic forecasting, our study stands out by introducing a diverse range of ML methods and conducting a rigorous evaluation of their out-of-sample prediction power. Notably, we contribute to the existing body of knowledge by pioneering the use of advanced deep learning, specifically Long Short-Term Memory (LSTM), in the analysis of Sudden Stop events.

A distinctive feature of our research is the claim that, to the best of our knowledge, we are the first to employ such ML methods in the study of Sudden Stop phenomena. This assertion underscores the originality and novelty of our work, positioning it at the forefront of research in this domain.

Beyond the results of our study, it is pertinent to report the challenges encountered, elucidate the prospective avenues for future research within the context of this study, and discuss the implications derived from the findings.

First challenge was the class imbalance in the Sudden Stop dataset since the SS events occur rarely. To address this issue, we modified the ML model's loss functions by weighting each class contribution by its occurrence rate. This mitigated the class imbalance problem for the most part.

While our current approach proves to be a sound practice for handling imbalanced datasets, alternative methods can also be considered, such as oversampling techniques like SMOTE. However, caution must be exercised when implementing oversampling, as it has the potential to introduce noise into the dataset. This noise, if not controlled, can lead to a decrease in out-of-sample performance.

Secondly, we faced a challenge related to causality. Despite employing traditional machine learning methods that incorporate feature importance and Shapley Values, it's important to acknowledge that these metrics do not establish causation; rather, they quantify the importance of features in predicting outcomes. Even with these tools, we are limited in our ability to uncover the true causal relationships within the

154

data. Furthermore, as we delved into advanced techniques like Long Short-Term Memory (LSTM) networks, the complexity of the model increased to the point where we perceive them as black boxes, providing predictions without clear interpretability.

The challenge remains in reconciling the predictive power of these advanced models with the interpretability and understanding of causal mechanisms. Striking a balance between model complexity and interpretability is crucial for drawing meaningful insights from our study. As we explore advanced methods, we must remain mindful of the trade-offs and seek ways to enhance interpretability, potentially through the integration of causal inference methods or other techniques that shed light on the causal aspects of our findings.

Thirdly, the performance of ML models is inherently influenced by the characteristics of the dataset they are trained on. When confronted with a new dataset, it becomes imperative to carefully consider each model of interest individually. Also, it is important to make the datasets as big and as diverse as possible to mitigate the model dependency on the dataset.

The effectiveness of a model is contingent on how well its underlying patterns align with the patterns present in the new dataset. Models that may have excelled on one dataset might not necessarily perform optimally on another if the data distribution, characteristics, or underlying relationships differ.

Hence, it is crucial to conduct a thorough evaluation of each model when applying it to a new dataset. This involves assessing objective performance metrics such as accuracy, precision, recall, and F1 score, among others, to gauge how well the model generalizes to the specific characteristics of the new data.

Additionally, model recalibration or fine-tuning may be necessary to adapt the model to the nuances of the new dataset. This iterative process of evaluation and adjustment ensures that the chosen models are robust and reliable across diverse datasets, enhancing their applicability and generalization capability.

The implications of our study are far-reaching and hold significance for the field of economic analysis and crisis prediction. Firstly, our research provides a foundation for the establishment of an advanced Early Warning System (EWS) for Sudden Stop Crises, leveraging the capabilities of deep learning methods like LSTM. This suggests the potential for constructing a powerful system capable of issuing timely alerts and minimizing the impact of Sudden Stop Crises.

Secondly, the study advocates for a synergistic approach by combining machine learning methods with traditional economic techniques. The integration of ML-based feature selection methods points towards a hybrid model that capitalizes on the strengths of both approaches. In dealing with our extended dataset, characterized by its complexity and a multitude of potential variables, employing Logistic regression posed challenges. Instead, we opted for a more effective approach, utilizing Random Forest. We trained the model and employed feature importance to rank variables. Subsequently, we employed the Bayesian Information Criterion (BIC) to select the most important variables for further analysis and model development. This method proved to be more robust and suitable for our dataset's intricacies.

Moreover, our findings underscore the need for an interdisciplinary approach in economic analysis, encouraging collaboration between machine learning experts and economists to extract deeper insights into the factors influencing economic crises. Policymakers can benefit from the study's insights to formulate proactive policies, while risk management professionals can enhance their strategies for identifying and managing economic risks. The study's implications extend to advancing financial forecasting practices, guiding education and skill development initiatives, and fostering a holistic understanding of economic dynamics in an ever-changing landscape.

In our ongoing effort to enhance the study, we are planning to explore the state-of-the-art Large Language Models (e.g., LLM type models such as Chat-GPT), and extend our research by creating an unconventional dataset through sentiment analysis of central bank policies and global investors' confidence. These factors are crucial, given their potential significant impact on Sudden Stop (SS) crises. The primary

objective is to seamlessly integrate this sentiment-driven data into our predictive model, thereby contributing to the development of an advanced Early Warning System.

Furthermore, our goal is to form an even bigger and more diverse financial dataset for SS prediction and make this available to public so that other researchers use this data to further advance the field. To this aim, we will also explore more advanced imputation methods to make use of existing data sources with missing variables.

Furthermore, our future plans involve a deeper investigation into oversampling methods to refine our approach for handling imbalanced datasets. Thoughtful consideration and application of oversampling techniques can play a crucial role in contributing to more accurate predictions, particularly in mitigating the potential risks associated with both dataset imbalance and small sample sizes.

Looking ahead, we anticipate the integration of Explainable AI methods into our Long Short-Term Memory (LSTM) model as technological advancements continue. This integration is poised to offer a more comprehensive understanding of the relationships between variables in the model, shedding light on the intricacies of the predictive factors contributing to Sudden Stop events. Our commitment to continuous exploration and integration of emerging AI methodologies ensures that our predictive models remain at the forefront of innovation, providing valuable insights for economic forecasting and crisis prediction.

REFERENCES

Arce, F., Bengui, J., & Bianchi, J. (2019). A Macroprudential Theory of Foreign Reserve Accumulation (NBER Working Paper No. 26236).

Adrian, T., Gopinath, G., Gourinchas, P.-O., Pazarbasioglu, C., & Weeks-Brown, R. (2022, March 3). Why the IMF is Updating its View on Capital Flows. Retrieved from IMF Blog: https://www.imf.org/en/Blogs/Articles/2022/03/30/blog033122-why-the-imf-is-updating-its-view-on-capital-flows

Aguiar, M., & Gopinath, G. (2006). Defaultable debt, interest rates and the current account. Journal of International Economics, 69, 64-83.

Aguiar, M., & Gopinath, G. (2007). Emerging market business cycles: the cycle is the trend. Journal of Political Economy, 115, 69-102.

Aiyagari, S. R. (1993). Explaining Financial Market Facts: The Importance of Incomplete Markets and Transaction Costs. Federal Reserve Bank of Minneapolis Quarterly Review, 17, 17-31.

Aiyagari, S. R., & Gertler, M. (1999). 'Overreaction' of Asset Prices in General Equilibrium. Review of Economic Dynamics.

Akinci, O., & Chahrour, R. (2018). Good News, Leverage, and Sudden Stops. Journal of International Economics, 114, 362–375

Alessi, A., & Detken, C. (2018). Identifying excessive credit growth and leverage. Journal of Financial Stability, 35, 215-225.

Athey, S., Bayati, M., Imbens, G., & Qu, Z. (2019). Ensemble Methods for Causal Effects in Panel Data Settings. AEA Papers and Proceedings, 109: 65-70.

Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know About. Annual Review of Economics, 11(1), 685-725.

Bernanke, B., Gertler, M., & Gilchrist, S. (1998). The Financial Accelerator in a Quantitative Business Cycle Framework (NBER Working Paper No. 6455).

Benigno, G., Chen, H., Otrok, C., Rebucci, A., & Young, E. R. (2016). Optimal capital controls and real exchange rate policies: a pecuniary externality perspective. Journal of Monetary Economics, 84, 147–165.

Bianchi, J. (2011). Overborrowing and Systemic Externalities in the Business Cycle. American Economic Review, 101(7), 3400-3426.

Bianchi, J., Boz, E., & Mendoza, E. G. (2012). Macro-Prudential Policy in a Fisherian model of Financial Innovation. NBER Working Papers 18036, National Bureau of Economic Research, Inc.

Bianchi, J., & Mendoza, E. G. (2010). Overborrowing, Financial Crises and 'Macro-prudential' Taxes. NBER Working Papers 16091, National Bureau of Economic Research, Inc.

Bianchi, J., Liu, C., & Mendoza, E. G. (2016). Fundamentals news, global liquidity and macroprudential policy. Journal of International Economics, 99(S1), 2-15.

Bianchi, J., & Mendoza, E. (2018). "Optimal Time-Consistent Macroprudential Policy." Journal of Political Economy, 126(2), 588-634.

Bianchi, J., Boz, E., & Mendoza, E. G. (2020). A Fisherian Approach to Financial Crises: Lessons from the Sudden Stops Literature. Review of Economic Dynamics, 37, 254-283, August.

Bluwstein, K., Buckmann, M., Joseph A., Kapadia, S., & Şimşek, Ö. (2023). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. Journal of International Economics, 145, 103773.

Bernanke, B.,& Gertler, M. (1989). Agency costs, net worth, and business fluctuations. The American Economic Review, 79, 1–15.

Borio, C., & Lowe, P. (2002). Assessing the risk of banking crises. BIS Quart. Rev. 2002, 43–54.

Borio, C. (2003). Towards a macroprudential framework for financial supervision and regulation? CESifo Economic Studies, 49(2), 181–215.

Bengui, J., & Bianchi, J. (2014). Capital Flow Management When Capital Controls Leak. University of Montreal, Mimeo.

Benigno, G., Chen, H., Otrok, C., Rebucci, A., &Young, E. R. (2013). Financial crises and macroprudential policies. J. Int. Econ., 89(2), 453–470.

Bruno, V., & Shin, H.S. (2014). Cross-border banking and global liquidity*. Rev. Econ. Stud. (p. rdu042).

Blanchard, O. J., L'Huillier, J.-P., & Lorenzoni, G. (2013). News, noise, and fluctuations: an empirical exploration. American Economic Review, 103(7), 3045–3070.

Jaimovich, N., & Rebelo, S. (2009). Can news about the future drive the business cycle? American Economic Review, 99(4), 1097–1118.

Bordo, M. (2007). Growing up to Financial Stability (NBER Working Paper No. 12993).

Breiman, L. (2001). Statistical Modeling: The Two Cultures. Statistical Science, 16(3), 199–215.

Beaudry, P., & Portier, F. (2006). Stock prices, news, and economic fluctuations. American Economic Review, 96(4), 1293–1307.

Bruno, V., & Shin, H. S. (2015). Cross-Border Banking and Global Liquidity. Review of Economic Studies, 82(2), 535-564.

Calvo, G. A. (1998). Capital Flows and Capital-Market Crises: The Simple Economics of Sudden Stops. Journal of Applied Economics, 1(1), 35-54.

Calvo, G. A., Izquierdo, A., & Mejia, L. F. (2008). Systemic Sudden Stops: The relevance of Balance Sheet Effects and Financial Integration (NBER Working Paper 14026).

Chen, M., DeHaven, M., Kitschelt, I., Lee, S. J., & Sicilian, M. J. (2023). Identifying Financial Crises Using Machine Learning on Textual Data. International Finance Discussion Papers 1374. Washington: Board of Governors of the Federal Reserve System.

Coffinet, J., & Kien, J. N. (2019). Detection of rare events: A machine learning toolkit with an application to banking crises. The Journal of Finance and Data Science, 5(4), 183–207.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16) (pp. 785–794). Association for Computing Machinery.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273–297. https://doi.org/10.1007/BF00994018

Durdu, C. B., Mendoza, E. G., & Terrones, M. E. (2009). Precautionary demand for foreign assets in Sudden Stop economies: an assessment of the New Mercantilism. Journal of Development Economics, 89(2), 194–209.

Duttagupta, R., & Cashin, P. (2008). The Anatomy of Banking Crises (IMF Working Paper WP/08/93).

Edwards, S. (2005). Capital Controls, Sudden Stops and Current Account Reversals. NBER Working Papers 11170, National Bureau of Economic Research, Inc.

Eichengreen, B., Gupta, P., & Mody, A. (2008). Sudden Stops and IMF-Supported Programs. NBER Working Papers 12235, National Bureau of Economic Research, Inc.

Eichengreen, B., & Gupta, P. (2016). Managing Sudden Stops. In E. G. Mendoza, E. Pastén, & D. Saravia (Eds.), Monetary Policy and Global Spillovers: Mechanisms, Effects and Policy Measures, 009-047. Central Bank of Chile. Central Banking, Analysis, and Economic Policies Book Series.

Eaton, J., & Gersovitz, M. (1981). Debt with Potential Repudiation: Theoretical and Empirical Analysis. Review of Economic Studies, 48, 289-309.

Fisher, I. (1933). The debt-deflation theory of great depressions. Econometrica: Journal of the Econometric Society, 337–357.

Flemming, J., L'Huillier, J.-P., & Piguillem, F. (2019). Macro-prudential taxation in good times. Journal of International Economics, 121, 103251.

Forbes, K. J., & Warnock, F. E. (2012). Capital flow waves: Surges, stops, flight, and retrenchment. Journal of International Economics, 88(2), 235-251.

Forbes, K. J., & Warnock, F. E. (2021). Capital flow waves—or ripples? Extreme capital flow movements since the crisis. Journal of International Money and Finance, 116(C).

Frankel, J., & Cavallo, E. (2004). Does Openness to Trade Make Countries More Vulnerable to Sudden Stops, or Less? Using Gravity to Establish Causality. Working Paper Series rwp04-038, Harvard University, John F. Kennedy School of Government.

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, 55, 119–139. AT&T Labs, 180 Park Avenue, Florham Park, New Jersey 07932.

Garcia-Cicco, J., Pancrazi, R., & Uribe, M. (2010). Real business cycles in emerging countries? American Economic Review, 100(5), 2510–2531.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2020). How is Machine Learning Useful for Macroeconomic Forecasting?

Hellwig, K. P. (2021). Predicting Fiscal Crises: A Machine Learning Approach. IMF Working Paper No. 2021/150.

Hoang, D., & Wiegratz, K. (2023). Machine Learning Methods in Finance: Recent Applications and Prospects. European Financial Management. https://doi.org/10.1111/eufm.12408

Hernandez, J., & Mendoza, E. (2017). Optimal v. simple financial policy rules in a production economy with "liability dollarization". Ensayos sobre Política Económica, 35(82), 25–39.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12(1), 55–67. doi:10.2307/1267351

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. Neural Computation, 9, 1735-80. doi:10.1162/neco.19

IMF (November 14, 2012). The liberalization and management of capital flows: An institutional view.

Jarmulska, B. (2020). Random forest versus logit models: which offers better early warning of fiscal stress? ECB Working Paper Series No 2408.

Jeanne, O., & Korinek, A. (2010). Managing credit booms and busts: A Pigouvian taxation approach. NBER Working Paper No 16377.

Jordà, O., Schularick, M. H., & Taylor, A. M. (2012). When Credit Bites Back: Leverage, Business Cycles and Crises. Working Papers Series 20, Institute for New Economic Thinking.

Joy M., Rusnák M., Šmídková K., and Vašíček B. (2015). Banking and currency crises: differential diagnostics for developed countries, ECB Working Paper 1810.

Kaminsky, G., Lizondo, S., & Reinhart, C. M. (1998). Leading indicators of currency crises. IMF Staff Papers, 45(1).

Kiyotaki, N., & Moore, J. (1997). Credit cycles. Journal of Political Economy, 105(2), 211–248.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction Policy Problems. American Economic Review, 105(5), 491-495.

Krugman, P. (1999). Balance Sheets, the Transfer Problem, and Financial Crises. International Tax and Public Finance, 6(4), 459-472.

Korinek, A., & Mendoza, E. G. (2013). From Sudden Stops to Fisherian Deflation: Quantitative Theory and Policy Implications. NBER Working Papers 19362, National Bureau of Economic Research, Inc.

Manasse P., Roubini N., and Schimmelpfennig, A. (2003). Predicting Sovereign Debt Crises, IMF Working Paper WP/03/221.

Manasse, P., & Roubini, N. (2009). "Rules of thumb" for sovereign debt crises. Journal of International Economics, 78(2), 192-205.

Mendoza, E. G. (2002). Credit, Prices, and Crashes: Business Cycles with a Sudden Stop. In: Preventing Currency Crises in Emerging Markets, 335-392. NBER Chapters, National Bureau of Economic Research, Inc.

Mendoza, E. G. (2005). Real Exchange Rate Volatility and the Price of Nontradables in Sudden-Stop-Prone Economies. NBER Working Papers 11691, National Bureau of Economic Research, Inc.

Mendoza, E. G. (2010). Sudden Stops, Financial Crises, and Leverage. American Economic Review, 100(5), 1941-1966.

Mendoza, E. G., & Quadrini, V. (2010). Financial globalization, financial crises and contagion. Journal of Monetary Economics, 57(1), 24–39.

Mendoza, E. G., & Smith, K. A. (2006). Quantitative implications of a debt-deflation theory of Sudden Stops and asset prices. Journal of International Economics, 70(1), 82-114.

Mendoza, E. G., & Terrones, M. E. (2012). An Anatomy Of Credit Booms: Evidence From Macro Aggregates And Micro Data. NBER Working Papers 18379, National Bureau of Economic Research, Inc.

Mendoza, E. G., & Rojas, E. I. (2018). Positive and Normative Implications of Liability Dollarization for Sudden Stops Models of Macroprudential Policy. NBER Working Papers 24336, National Bureau of Economic Research, Inc.

Milesi Ferretti, G. M., & Razin, A. (2000). Current account reversals and currency crises: empirical regularities. In: Currency Crises. University of Chicago Press, pp. 285–323.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives, 31(2), 87-106.

Ottonello, P. (2015). Optimal exchange rate policy under collateral constraints and wage rigidity. Columbia University. Manuscript.

Reinhart, C., & Calvo, G. (2000). When Capital Inflows Come to a Sudden Stop: Consequences and Policy Options. MPRA Paper 6982, University Library of Munich, Germany.

Reinhart, C. M., & Rogoff, K. S. (2009). This Time Is Different: Eight Centuries of Financial Folly. Princeton University Press.

Reinhart, C. M., & Reinhart, V. R. (2009). Capital Flow Bonanzas: An Encompassing View of the Past and Present. In: NBER International Seminar on Macroeconomics 2008, 9-62, National Bureau of Economic Research, Inc.

Reinhart, C. M., & Rogoff, K. S. (2014). Recovery from financial crises: evidence from 100 episodes. American Economic Review, 104(5), 50–55.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386–408.

Savona, R., & Vezzoli, M. (2015). Fitting and Forecasting Sovereign Defaults using Multiple Risk Signals. Oxford Bulletin of Economics and Statistics, 77(1), 66-92.

Schmitt-Grohé, S., & Uribe, M. (2012). What's news in business cycles. Econometrica, 80(6), 2733–2764.

Schmitt-Grohé, S., & Uribe, M. (2021). Multiple Equilibria in Open Economies with Collateral Constraints. Review of Economic Studies, 88(4), 969–1001.

Schularick, M., & Taylor, A. M. (2012). Credit booms gone bust monetary policy, leverage cycles, and financial crises, 1870–2008. American Economic Review, 102(2), 1029.

Seoane, H. D., & Yurdagul, E. (2019). Trend shocks and sudden stops. Journal of International Economics, 121, 103252. ISSN 0022-1996.

Shin, H. S. (2013). The Second Phase of Global Liquidity and its Impact on Emerging Economies. Princeton University, Mimeo.

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. Journal of Economic Perspectives, 28(2), 3-28.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–88.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B, 67(2), 301–320.

APPENDICES

A. DATA SOURCES AND COUNTRY LISTS

I. BALANCE of PAYMENT and INTERNATIONAL FINANCIAL STATISTICS ,IMF

- Capital Account Total Credit US Dollars
- Capital Account Total Debit US Dollars
- Capital Account Total Net US Dollars
- Current Account Goods and Services Credit US Dollars
- Current Account Goods and Services Debit US Dollars
- Current Account Goods and Services Goods Credit US Dollars
- Current Account Goods and Services Goods Debit US Dollars
- Current Account Goods and Services Goods General Merchandise on a Balance of Payments Basis Credit US Dollars
- Current Account Goods and Services Goods General Merchandise on a Balance of Payments Basis Debit US Dollars
- Current Account Goods and Services Goods General Merchandise on a Balance of Payments Basis Net US Dollars
- Current Account Goods and Services Goods Net US Dollars
- Current Account Goods and Services Net US Dollars
- Current Account Goods and Services, Services Charges for the Use of Intellectual Property nie Credit US Dollars
- Current Account Goods and Services ,Services Charges for the Use of Intellectual Property nie Debit US Dollars
- Current Account Goods and Services ,Services Charges for the Use of Intellectual Property nie Net US Dollars
- Current Account Goods and Services ,Services Credit US Dollars
- Current Account Goods and Services ,Services Debit US Dollars
- Current Account Goods and Services ,Services Government Goods and Services nie Credit US Dollars
- Current Account Goods and Services ,Services Government Goods and Services nie Debit US Dollars
- Current Account Goods and Services ,Services Government Goods and Services nie Net US Dollars
- Current Account Goods and Services, Services Insurance and Pension Services Credit US Dollars
- Current Account Goods and Services ,Services Insurance and Pension Services Debit US Dollars

- Current Account Goods and Services ,Services Insurance and Pension Services Net US Dollars
- Current Account Goods and Services ,Services Net US Dollars
- Current Account Goods and Services ,Services Other Business Services Credit US Dollars
- Current Account Goods and Services ,Services Other Business Services Debit US Dollars
- Current Account Goods and Services, Services Other Business Services Net US Dollars
- Current Account Goods and Services ,Services Other Services Credit US Dollars
- Current Account Goods and Services ,Services Other Services Debit US Dollars
- Current Account Goods and Services ,Services Transport Credit US Dollars
- Current Account Goods and Services ,Services Transport Debit US Dollars
- Current Account Goods and Services ,Services Transport Net US Dollars
- Current Account Goods and Services ,Services Travel Credit US Dollars
- Current Account Goods and Services ,Services Travel Debit US Dollars
- Current Account Goods and Services ,Services Travel Net US Dollars
- Current Account Primary Income Compensation of Employees Debit US Dollars
- Current Account Primary Income Compensation of Employees Net US Dollars
- Current Account Primary Income Credit US Dollars
- Current Account Primary Income Debit US Dollars
- Current Account Primary Income Investment Income Credit US Dollars
- Current Account Primary Income Investment Income Debit US Dollars
- Current Account Primary Income Investment Income Direct Investment Credit US Dollars
- Current Account Primary Income Investment Income Direct Investment Debit US Dollars
- Current Account Primary Income Investment Income Direct Investment Income on Equity and Investment Fund Shares Credit US Dollars
- Current Account Primary Income Investment Income Direct Investment Income on Equity and Investment Fund Shares Debit US Dollars
- Current Account Primary Income Investment Income Direct Investment Income on Equity and Investment Fund Shares Dividends and Withdrawals from Income of Quasi-Corporations Credit US Dollars
- Current Account Primary Income Investment Income Direct Investment Income on Equity and Investment Fund Shares Dividends and Withdrawals from Income of Quasi-Corporations Debit US Dollars
- Current Account Primary Income Investment Income Direct Investment Income on Equity and Investment Fund Shares Dividends and Withdrawals from Income of Quasi-Corporations Net US Dollars
- Current Account Primary Income Investment Income Direct Investment Income on Equity and Investment Fund Shares Net US Dollars

- Current Account Primary Income Investment Income Direct Investment Net US Dollars
- Current Account Primary Income Investment Income Net US Dollars
- Current Account Primary Income Net US Dollars
- Current Account Secondary Income Credit US Dollars
- Current Account Secondary Income Debit US Dollars
- Current Account Secondary Income Financial Corporations Nonfinancial Corporations Households and NPISHs Credit US Dollars
- Current Account Secondary Income Financial Corporations Nonfinancial Corporations Households and NPISHs Debit US Dollars
- Current Account Secondary Income Financial Corporations Nonfinancial Corporations Households and NPISHs Net US Dollars
- Current Account Secondary Income General Government Credit US Dollars
- Current Account Secondary Income General Government Debit US Dollars
- Current Account Secondary Income General Government Net US Dollars
- Current Account Secondary Income Net US Dollars
- Current Account Total Credit US Dollars
- Current Account Total Debit US Dollars
- Current Account Total Net US Dollars
- Financial Account Net excluding exceptional financing other investment Net incurrence of liabilities excluding exceptional financing US Dollars
- Financial Account Net with Fund Record Other Investment Net with Fund Record Net incurrence of liabilities with Fund Record US Dollars
- Financial Account Net Lending Net Borrowing Balance from Financial Account Direct Investment Net Acquisition of Financial Assets Debt Instruments US Dollars
- Financial Account Net Lending Net Borrowing Balance from Financial Account Direct Investment Net Acquisition of Financial Assets Equity and Investment Fund Shares Equity Other Than Reinvestment of Earnings US Dollars
- Financial Account Net Lending Net Borrowing Balance from Financial Account Direct Investment Net Acquisition of Financial Assets Equity and Investment Fund Shares US Dollars
- Financial Account Net Lending Net Borrowing Balance from Financial Account Direct Investment Net Acquisition of Financial Assets US Dollars
- Financial Account Net Lending Net Borrowing Balance from Financial Account Direct Investment Net Incurrence of Liabilities Debt Instruments US Dollars
- Financial Account Net Lending Net Borrowing Balance from Financial Account Direct Investment Net Incurrence of Liabilities Equity and Investment Fund Shares Equity Other Than Reinvestment of Earnings US Dollars
- Financial Account Net Lending Net Borrowing Balance from Financial Account Direct Investment Net Incurrence of Liabilities Equity and Investment Fund Shares US Dollars

- Financial Account Net Lending Net Borrowing Balance from Financial Account Direct Investment Net Incurrence of Liabilities US Dollars
- Financial Account Net Lending Net Borrowing Balance from Financial Account Direct Investment US Dollars
- Financial Account Net Lending Net Borrowing Balance from Financial Account US Dollars
- Financial Account Other Investment Currency and Deposits Net Acquisition of Financial Assets Deposit-taking Corporations Except Central Bank US Dollars
- Financial Account Other Investment Currency and Deposits Net Acquisition of Financial Assets US Dollars
- Financial Account Other Investment Currency and Deposits Net Incurrence of Liabilities Deposit-taking Corporations Except Central Bank US Dollars
- Financial Account Other Investment Currency and Deposits Net Incurrence of Liabilities US Dollars
- Financial Account Other Investment Currency and Deposits US Dollars
- Financial Account Other Investment Loans Net Acquisition of Financial Assets US Dollars
- Financial Account Other Investment Loans Net Incurrence of Liabilities General Government US Dollars
- Financial Account Other Investment Loans Net Incurrence of Liabilities Other Sectors US Dollars
- Financial Account Other Investment Loans Net Incurrence of Liabilities US Dollars
- Financial Account Other Investment Loans US Dollars
- Financial Account Other Investment Net Acquisition of Financial Assets Debt Instruments Deposit-taking Corporations Except the Central Bank US Dollars
- Financial Account Other Investment Net Acquisition of Financial Assets Debt Instruments Other Sectors US Dollars
- Financial Account Other Investment Net Acquisition of Financial Assets Debt Instruments US Dollars
- Financial Account Other Investment Net Acquisition of Financial Assets US Dollars
- Financial Account Other Investment Net Incurrence of Liabilities Debt Instruments Deposit-taking Corporations Except the Central Bank US Dollars
- Financial Account Other Investment Net Incurrence of Liabilities Debt Instruments Other Sectors US Dollars
- Financial Account Other Investment Net Incurrence of Liabilities US Dollars
- Financial Account Other Investment Other Accounts Receivable/payable Net Acquisition of Financial Assets US Dollars
- Financial Account Other Investment Other Accounts Receivable/payable Net Incurrence of Liabilities US Dollars
- Financial Account Other Investment Other Accounts Receivable/payable US Dollars

- Financial Account Other Investment Other Equity Net Acquisition of Financial Assets Debt Instruments US Dollars
- Financial Account Other Investment US Dollars
- Financial Account Portfolio Investment Net Acquisition of Financial Assets Debt Securities US Dollars
- Financial Account Portfolio Investment Net Acquisition of Financial Assets Equity and Investment Fund Shares US Dollars
- Financial Account Portfolio Investment Net Acquisition of Financial Assets US Dollars
- Financial Account Portfolio Investment Net Incurrence of Liabilities Debt Securities US Dollars
- Financial Account Portfolio Investment Net Incurrence of Liabilities Equity and Investment Fund Shares US Dollars
- Financial Account Portfolio Investment Net Incurrence of Liabilities US Dollars
- Financial Account Portfolio Investment US Dollars
- Financial Account Reserve Assets Other Reserve Assets US Dollars
- Financial Account Reserve Assets Special Drawing Rights US Dollars
- Financial Account Reserve Assets US Dollars
- Fund Accounts Currency Holdings as of Quota Percent per annum
- Fund Accounts Currency Holdings SDRs
- Fund Accounts Currency Holdings US Dollars
- Fund Accounts Lending to the Fund SDRs
- Fund Accounts Lending to the Fund US Dollars
- Fund Accounts Net Cumulative Allocation Escrow account SDRs
- Fund Accounts Net Cumulative Allocation Escrow account US Dollars
- Fund Accounts Outstanding GRA SDRs
- Fund Accounts Outstanding GRA US Dollars
- Fund Accounts Outstanding Loans SDRs
- Fund Accounts Outstanding Loans US Dollars
- Fund Accounts Overdue Obligations SDRs
- Fund Accounts Overdue Obligations US Dollars
- Fund Accounts Quota SDRs
- Fund Accounts Quota US Dollars
- Fund Accounts SDR Holdings as of Allocation Percent per annum
- Fund Accounts SDR Holdings Allocations SDRs
- Fund Accounts SDR Holdings Allocations US Dollars
- Fund Accounts UFC Loans as of Quota Percent per annum
- Fund Accounts UFC Loans SDRs
- Fund Accounts UFC Loans US Dollars
- Fund Accounts Use of Fund Credit and Loans Reserve Tranche Position SDRs
- Fund Accounts Use of Fund Credit and Loans Reserve Tranche Position US Dollars
- International Liquidity Gold Holdings National Valuation SDRs
- International Liquidity Gold Holdings National Valuation US Dollars

170

- International Liquidity Total Reserves excluding Gold Foreign Exchange SDRs
- International Liquidity Total Reserves excluding Gold Foreign Exchange US Dollars
- International Liquidity Total Reserves excluding Gold SDRs
- International Liquidity Total Reserves excluding Gold US Dollars
- International Reserves Official Reserve Assets IMF Reserve Position SDRs
- International Reserves Official Reserve Assets IMF Reserve Position US Dollars
- International Reserves Official Reserve Assets Market Value SDRs
- International Reserves Official Reserve Assets SDRs
- International Reserves Official Reserve Assets SDRs SDRs
- International Reserves Official Reserve Assets SDRs US Dollars
- International Reserves Official Reserve Assets US Dollars
- Net Errors and Omissions US Dollars
- Net Lending Net Borrowing Balance from Current and Capital Account US Dollars
- Supplementary Items Balance on Goods Services and Income US Dollars
- Supplementary Items Capital Account Excludes Reserves and Related Items US Dollars
- Supplementary Items Capital Account Credit Excludes Reserves and Related Items US Dollars
- Supplementary Items Current Account Net Excluding Exceptional Financing US Dollars
- Supplementary Items Current Acct Capital Acct Financial Acct US Dollars
- Supplementary Items Direct Investment Net Incurrence of Liabilities Excluding Exceptional Financing US Dollars
- Supplementary Items Direct Investment Net Incurrence of Liabilities Equity and Investment Fund Shares Excluding Exceptional Financing US Dollars
- Supplementary Items Direct Investment Net Incurrence of Liabilities Debt Instruments Excluding Exceptional Financing US Dollars
- Supplementary Items Errors and Omissions with Fund Record US Dollars
- Supplementary Items Financial Account Net Excluding Exceptional Financing US Dollars
- Supplementary Items Financial Account Net with Fund Record US Dollars
- Supplementary Items Net Credit and Loans from the IMF Excluding Reserve Position US Dollars
- Supplementary Items Other Investment Net with Fund Record US Dollars
- Supplementary Items Other Investment Net Incurrence of Liabilities Debt Instruments Excluding Exceptional Financing US Dollars
- Supplementary Items Other Investment Net Incurrence of Liabilities Debt Instruments Of which Other Financial Corporations Of which Other Financial Corporations Excluding Exceptional Financing US Dollars
- Supplementary Items Other Investment Other Debt Instruments Net Incurrence of Liabilities with Fund Record US Dollars

- Supplementary Items Other Investment Other Debt Instruments Net Incurrence of Liabilities Central Bank with Fund Record US Dollars
- Supplementary Items Other Investment Other Debt Instruments Net Incurrence of Liabilities General Government with Fund Record US Dollars
- Supplementary Items Portfolio Investment Net Incurrence of Liabilities Excluding Exceptional Financing US Dollars
- Supplementary Items Portfolio Investment Net Incurrence of Liabilities Debt Securities Excluding Exceptional Financing US Dollars
- Supplementary Items Portfolio Investment Net Incurrence of Liabilities Equity Securities Excluding Exceptional Financing US Dollars
- Supplementary Items Reserve Assets with Fund Record US Dollars
- Supplementary Items Reserve Position in the Fund with Fund Record US Dollars
- Supplementary Items Reserves and Related items US Dollars
- Supplementary Items SDR Holdings with Fund Record US Dollars
- Supplementary Items Secondary Income Credit Excluding Exceptional Financing US Dollars
- Supplementary Items Special Drawing Rights Allocations with Fund Record US Dollars
- Supplementary Items Total Current Capital Account US Dollars
- Total International Reserves SDRs gold at 35 SDRs per ounce
- Total International Reserves US Dollars gold at 35 SDRs per ounce
- Total Reserves US Dollars Gold at Market Price

II. TRADE VARIABLES FROM DIRECTION OF TRADE STATISTICS (DOT), IMF

- Advanced Economies Goods Value of Exports Free on-board FOB US Dollars
- Advanced Economies Goods Value of Imports Cost Insurance Freight CIF US Dollars
- Advanced Economies Goods Value of Trade Balance US Dollars
- Emerging and Developing Asia Goods Value of Exports Free on-board FOB US Dollars
- Emerging and Developing Asia Goods Value of Imports Cost Insurance Freight CIF US Dollars
- Emerging and Developing Asia Goods Value of Trade Balance US Dollars
- Emerging and Developing Economies Goods Value of Exports Free on-board FOB US Dollars
- Emerging and Developing Economies Goods Value of Imports Cost Insurance Freight CIF US Dollars
- Emerging and Developing Economies Goods Value of Trade Balance US Dollars
- Emerging and Developing Europe Goods Value of Exports Free on-board FOB US Dollars
- Emerging and Developing Europe Goods Value of Imports Cost Insurance Freight CIF US Dollars

- Emerging and Developing Europe Goods Value of Trade Balance US Dollars
- World Goods Value of Exports Free on-board FOB US Dollars
- World Goods Value of Imports Cost Insurance Freight CIF US Dollars
- World Goods Value of Trade Balance US Dollars

III.   VARIABLES FROM FORBES AND WARNOCK (2021):

Dummy Variables:

- Emerging Market (EM)
- Income Group (high, middle, low)
- Region (Asia, Eastern Europe, Latin America, North America, Western Europe, Other)

Long Run Interest Rates:

- UK's long run interest rates   (c112ib)
- Euro Area (EA)'s long run interest rates   (lt_rate_ea)
- Japan's long run interest rates   (lt_rate_jp)
- US's long run interest rates   (lt_rate_us)
- US's long run interest rates, change in 1 period   (lt_rate_us_ch)
- Average of Japan's and EA's long run interest rates   (lt_rate_us_jp_ea)
- Average of US, UK, Japan, EA's long run interest rates (lt_rate_us_jp_ea_uk)
- Average of US, UK, Japan, EA's long run interest rates, change in 1 period (lt_rate_all_ch)

Shadow Short Run Rates:

- UK's short run interest rates (ssr_uk)
- Euro Area (EA)'s short run interest rates (ssr_ea)
- Japan's short run interest rates (ssr_jp)
- US's short run interest rates (ssr_us)
- US's short run interest rates, change in 1 period (ssr_us_ch)
- Average of US, UK, Japan, EA's short run interest rates (ssr_us_jp_ea_uk)
- Average of US, UK, Japan, EA's short run interest rates, change in 1 period (ssr_all_ch)

Global Prices:

- Commodity Prices Change, 1 period (comm_ch)
- Oil price (p_oil)
- Oil price Change, 1 period (p_oil_ch)

Global Risk Measures:

- Variance Risk Premium Average (vrp_avg)
- Change in Variance Risk Premium Average, 1 period (vrp_ch)
- CBOE Volatility Index (VXO) (vxo)
- Change in Variance Risk Premium Average, 1 period (vxo_ch)

Other Global Variables:

- Global growth (growth_global)
- Global Inflation (inflation_global)
- Global Liquidity (money_global)
- Growth in Global Liquidity (money_global_growth)

Regional Variables:
- Contagion Variable (cont_stop)

Local Variables:

- Real GDP Growth (realgdpyoy)

List of Countries in Sudden Stop Analysis:

Argentina, Australia, Austria, Bangladesh, BelLux, Bolivia, Brazil, Canada, Chile, China, Colombia, Costa Rica, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Guatemala, Hong Kong, Hungary, Iceland, India, Indonesia, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Malaysia, Mexico, Netherlands, New Zealand, Norway, Panama, Peru, Philippines, Poland, Portugal, Romania, Russia, Singapore, Slovak Republic, Slovenia, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Thailand, Turkey, UK, US, Venezuela.

List of Countries in Estimations:

Argentina, Australia, Austria, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, Indonesia, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Malaysia, Mexico, Netherlands, New Zealand, Norway, Panama, Peru, Philippines, Poland, Portugal, Romania, Russia, Slovak Republic, Slovenia, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Thailand, Turkey, UK, US.

BelLux, Guatemala, Hong Kong, India, China, Singapore, Taiwan, and Venezuela were excluded from the estimation process due to unavailability of some of the exogenous variables, specifically real GDP data as in Forbes and Warnock(2021).
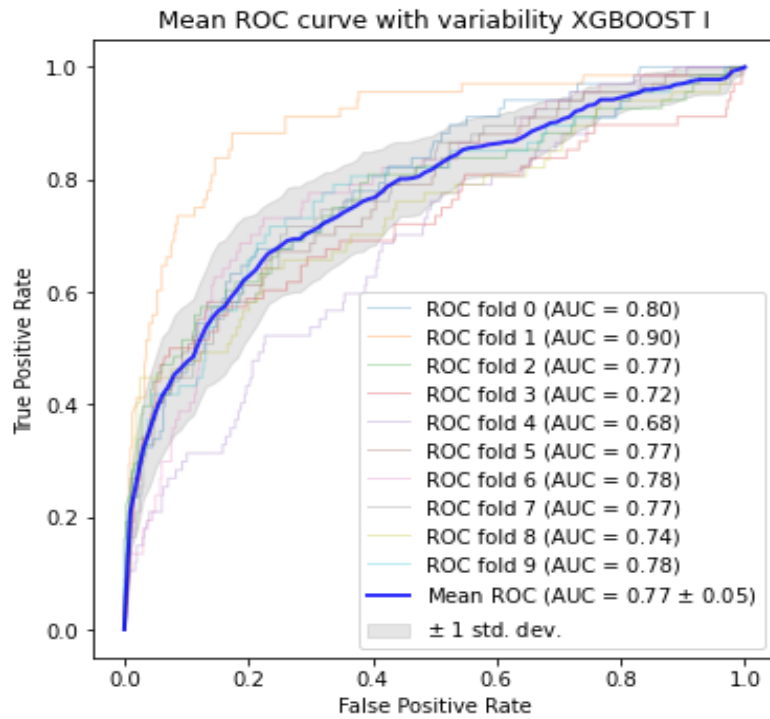
B. ADDITIONAL FIGURES FOR CHAPTER 4
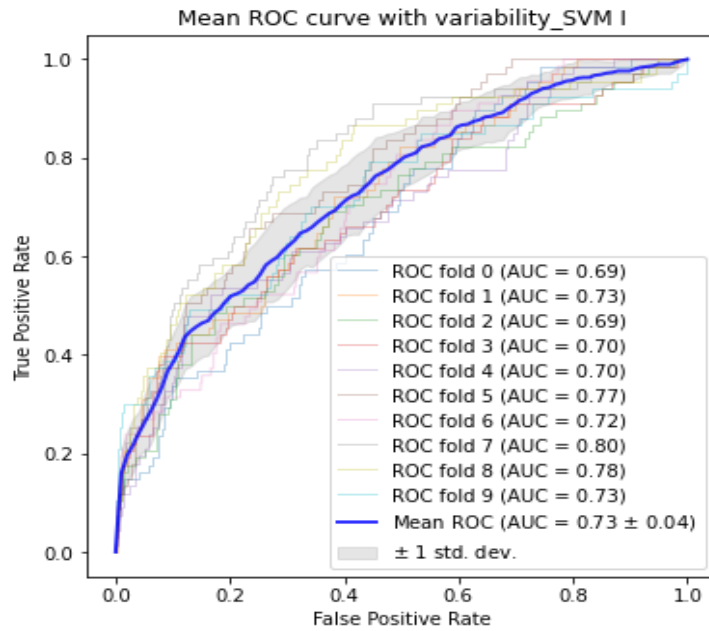
Figure B. 1. Mean Roc Curve with Variability SVM I
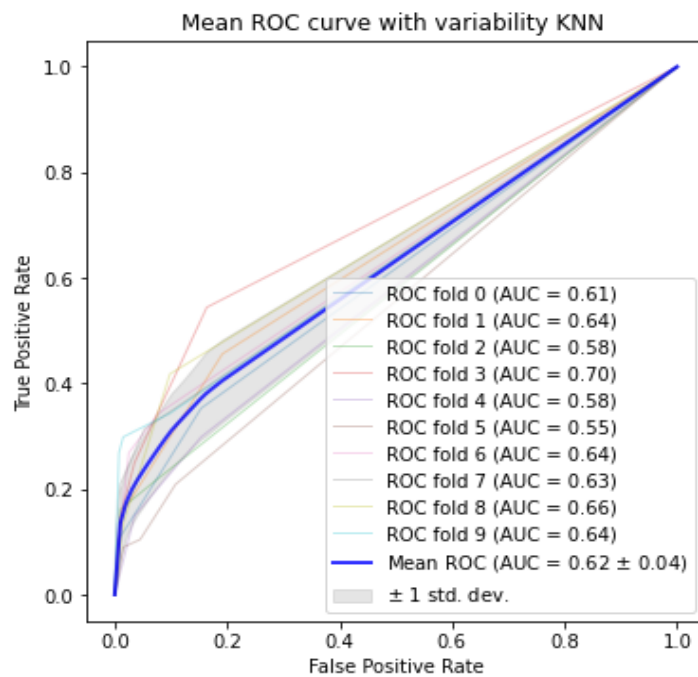
Figure B. 2. Mean Roc Curve with Variability KNN
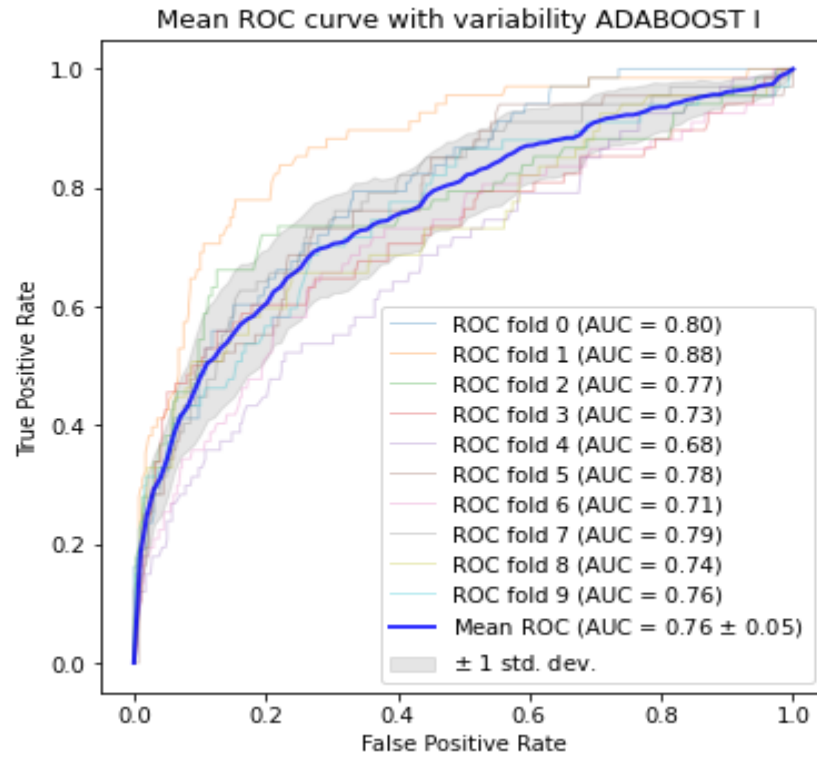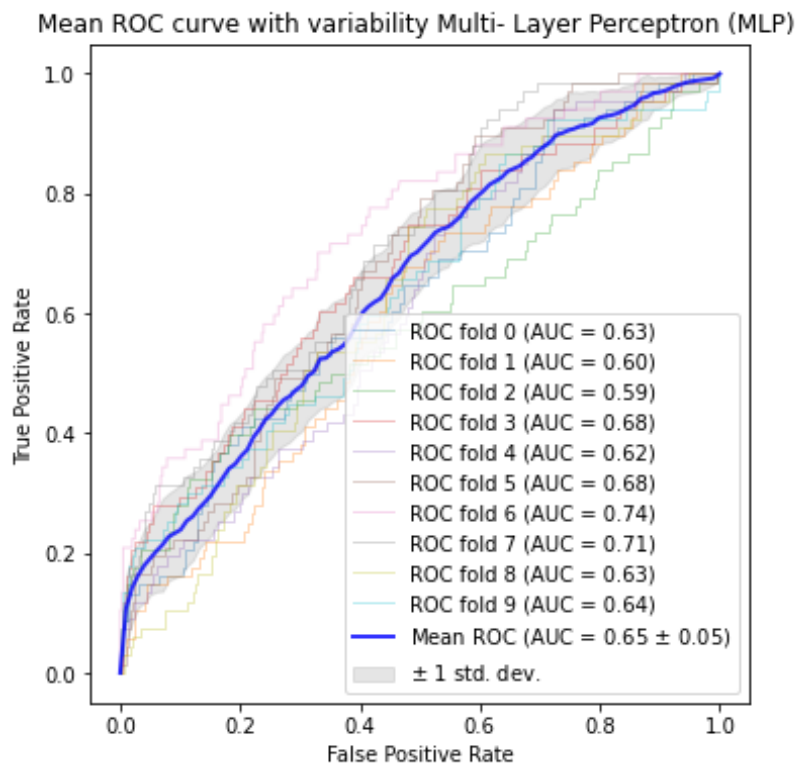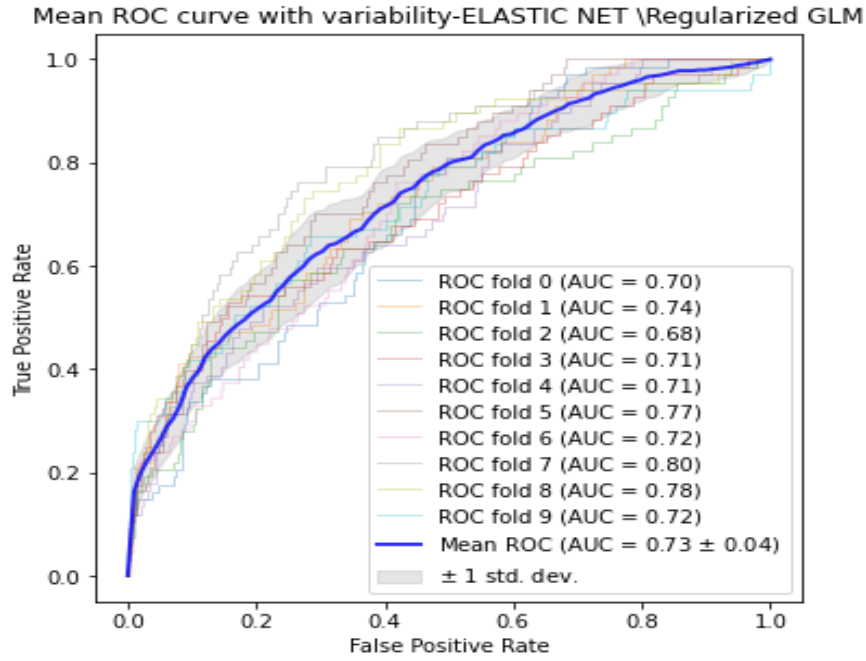
Figure B. 3. Mean Roc Curve with Variability AdaBoost I
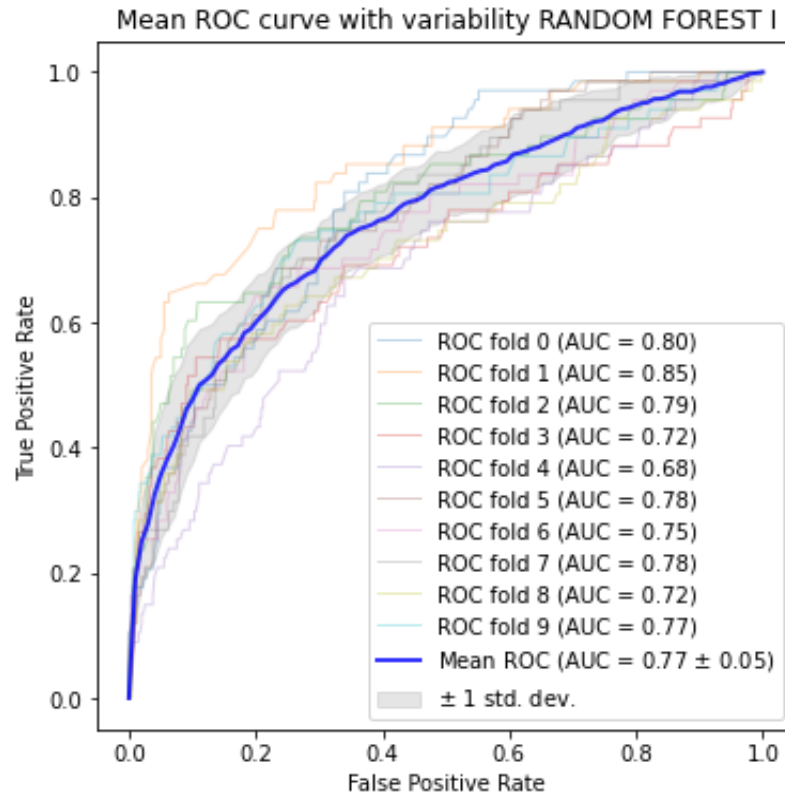
Figure B. 4. Mean Roc Curve with Variability Random Forest I

Figure B. 5. Mean Roc Curve with Variability XGBoost I

Figure B. 6. Mean Roc Curve with Variability MLP

178

Figure B. 7. Mean Roc Curve with Variability Elastic Net

Table B. 1. Grid Search Parameters for Chapter 4

Method	Grid Search Parameters
SVM	{'C': [15, 20, 25, 30, 40, 50], 'gamma': [1e-05, 5e-05, 0.0001, 0.0005]}
ELASTIC NET	{'penalty': ['l1', 'l2', 'elasticnet'], 'alpha': array([1.00000000e-04, 2.63665090e-04, 6.95192796e-04, 1.83298071e-03, 4.83293024e-03, 1.27427499e-02, 3.35981829e-02, 8.85866790e-02, 2.33572147e-01, 6.15848211e-01, 1.62377674e+00, 4.28133240e+00, 1.12883789e+01, 2.97635144e+01, 7.84759970e+01, 2.06913808e+02, 5.45559478e+02, 1.43844989e+03, 3.79269019e+03, 1.00000000e+04]), 'l1_ratio': [0.0, 0.1, 0.15, 0.2, 0.25, 0.4, 0.5, 0.75]}

KNN	{'n_neighbors': [3, 5, 7, 9, 11, 13]}
XGBOOST	{'learning_rate': [0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6], 'max_depth': [3, 4, 5], 'n_estimators': [50, 100, 150, 200, 250, 300, 400, 500, 600], 'reg_lambda': [0, 0.001, 0.01, 0.1, 1, 10, 100], 'max_features': [None], 'scale_pos_weight': [1, 2, 3, 4,5]}
RANDOM FOREST	{'n_estimators': [5, 10, 15, 20, 25, 30, 40, 45, 50], 'max_depth': [None, 2, 3, 4, 5, 6, 10], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4, 6, 8, 10], 'max_features': ['auto', 'sqrt']}
ADABOOST	{'n_estimators': [50, 100, 150, 200, 250, 300], 'learning_rate': [0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0, 3, 0.4, 0.5, 0, 8, 1.0]}
MLP	{'hidden_layer_sizes': [(256,), (128, 128), (128, 64, 64)], 'alpha': [0.0005, 0.001, 0.01], 'learning_rate_init': [0.001, 0.01, 0.1]}

Table B. 2. Grid Search Results for Chapter 4

Models	**Best Parameters by Grid Search CV**
SVM	{'C': 25, 'gamma': 0.0001}
Elastic Net	{'alpha': 0.0001, 'l1_ratio': 0.2, 'penalty': 'elasticnet'}
KNN	{'n_neighbors': 3}
XG Boost	{'learning_rate': 0.05, 'max_depth': 3, 'max_features': None, 'min_child_weight': 5, 'n_estimators': 400, 'reg_lambda': 10, 'scale_pos_weight': 5}

Table B.2. (cont'd)

Random Forest	{'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 6, 'min_samples_split': 2, 'n_estimators': 15}
AdaBoost	{'learning_rate': 1.0, 'n_estimators': 100}
MLP	'activation': 'relu', 'alpha': 0.001, 'hidden_layer_sizes': (256,), 'learning_rate_init': 0.01, 'max_iter': 1000

C. ADDITIONAL FIGURES FOR CHAPTER 5
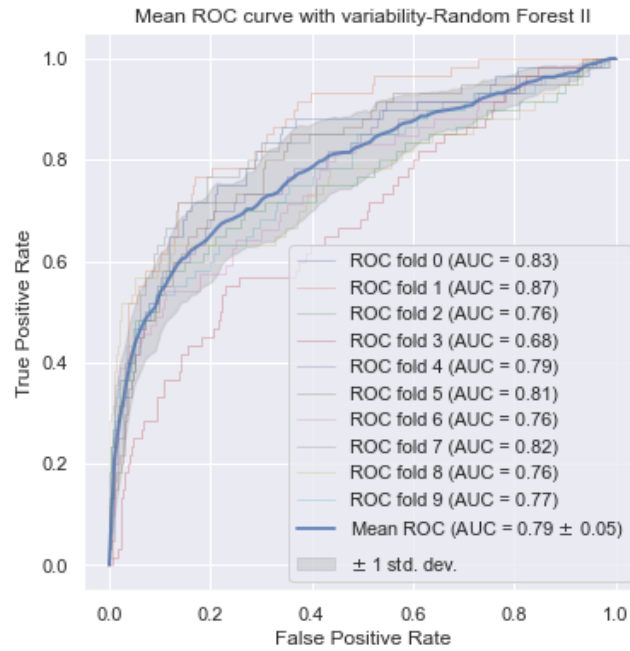
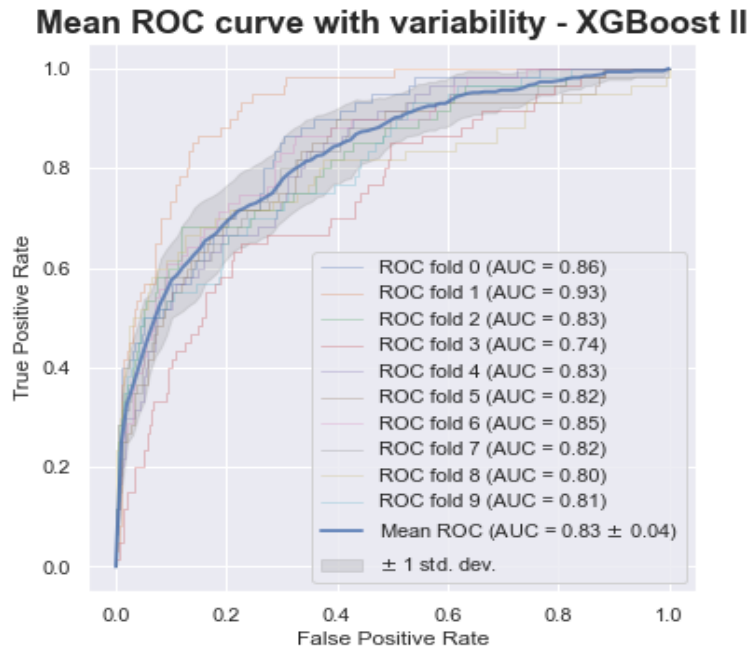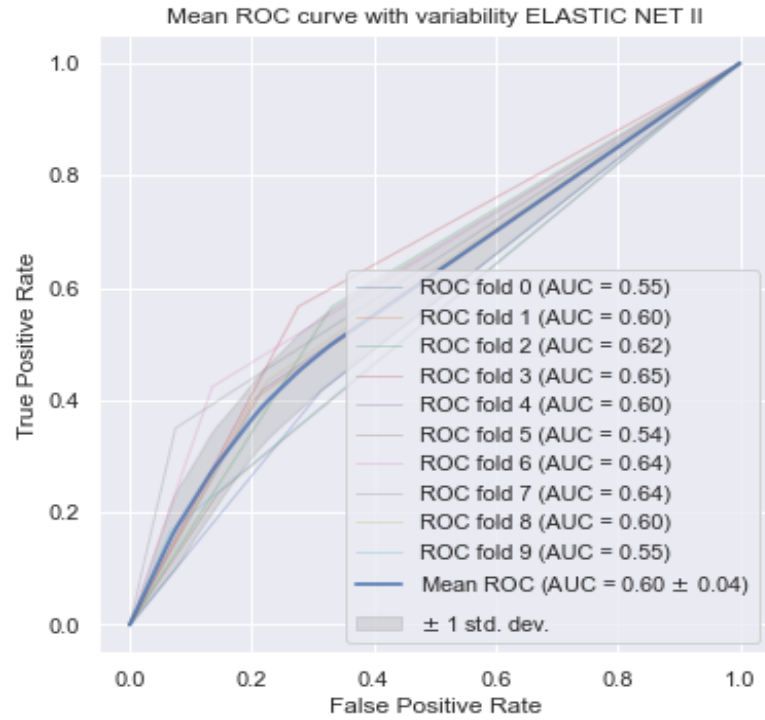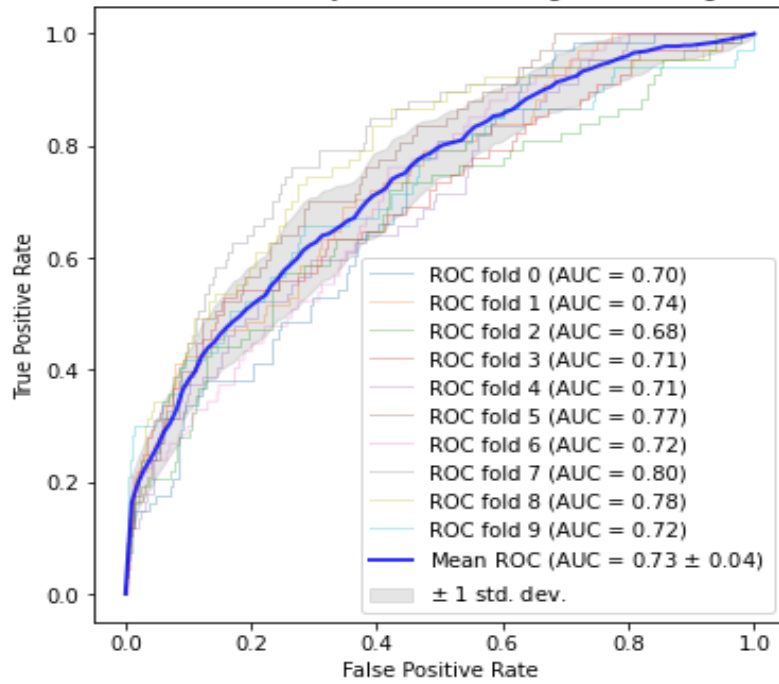**Figure C. 1.** Mean ROC Curves for Stratified 10-Fold Cross Validation Random Forest II

Figure C. 2. Mean ROC Curves for Stratified 10-Fold Cross Validation XGBoost II

Figure C. 3. Mean ROC Curves for Stratified 10-Fold Cross Validation Elastic Net II
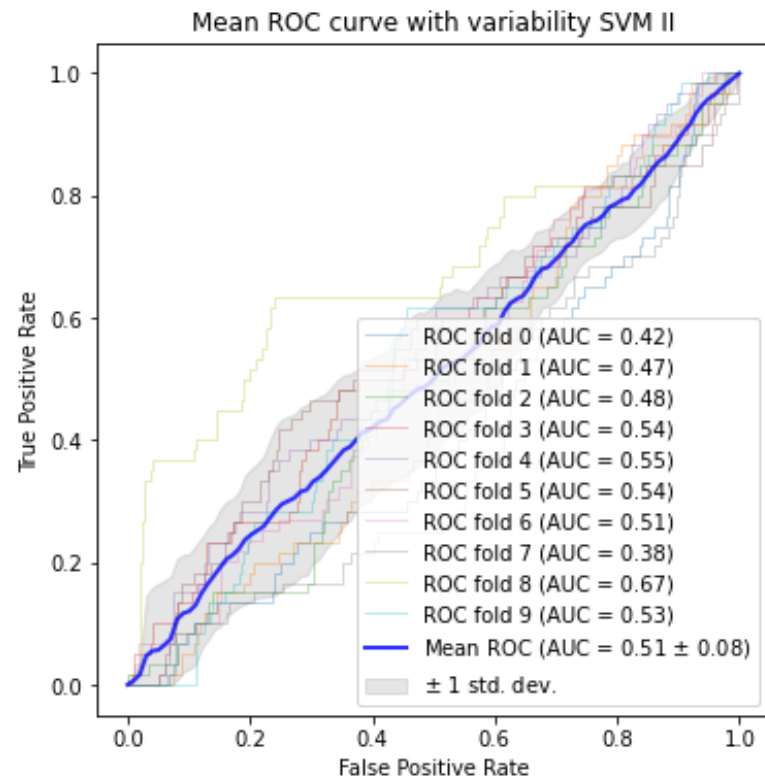
Figure C. 4. Mean ROC Curves for Stratified 10-Fold Cross Validation SVM II

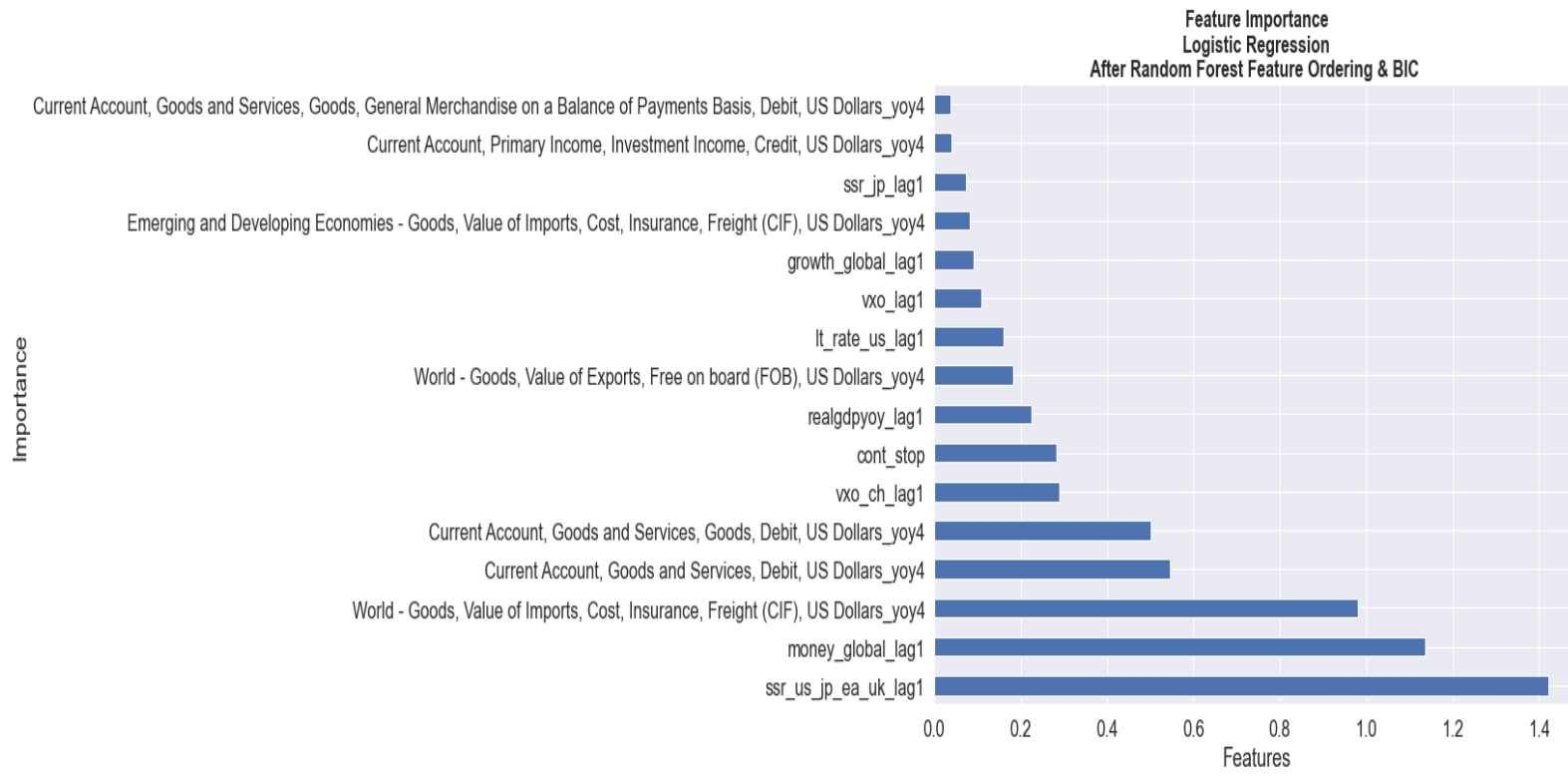**Figure C. 5.** Mean ROC Curves for Stratified 10-Fold Cross Validation Logistic Regression I
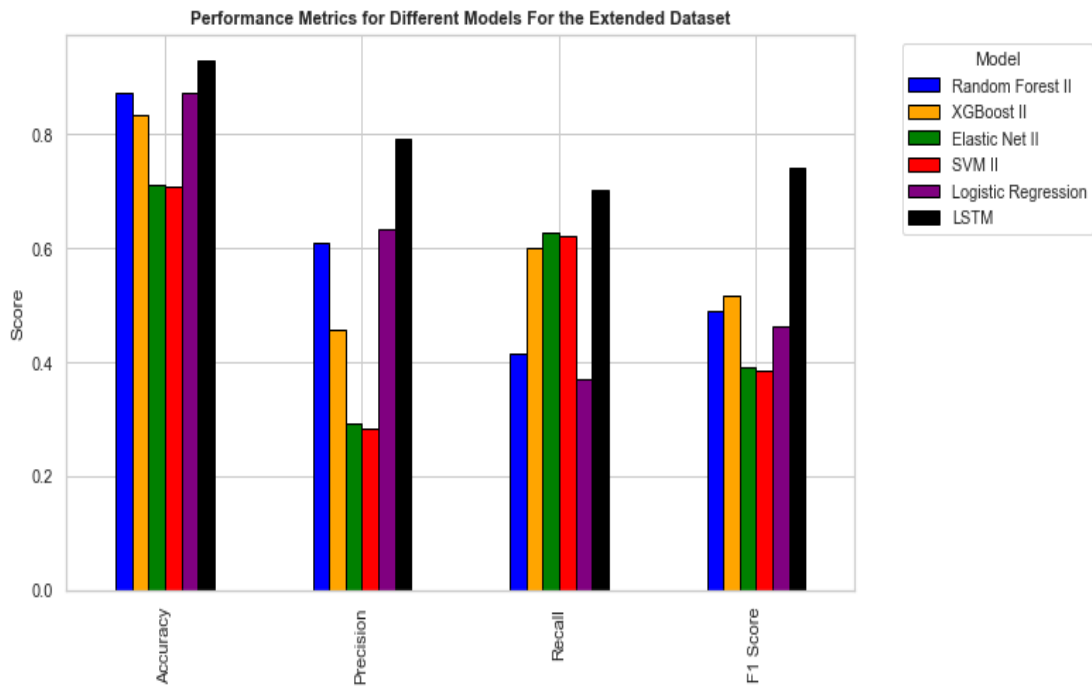
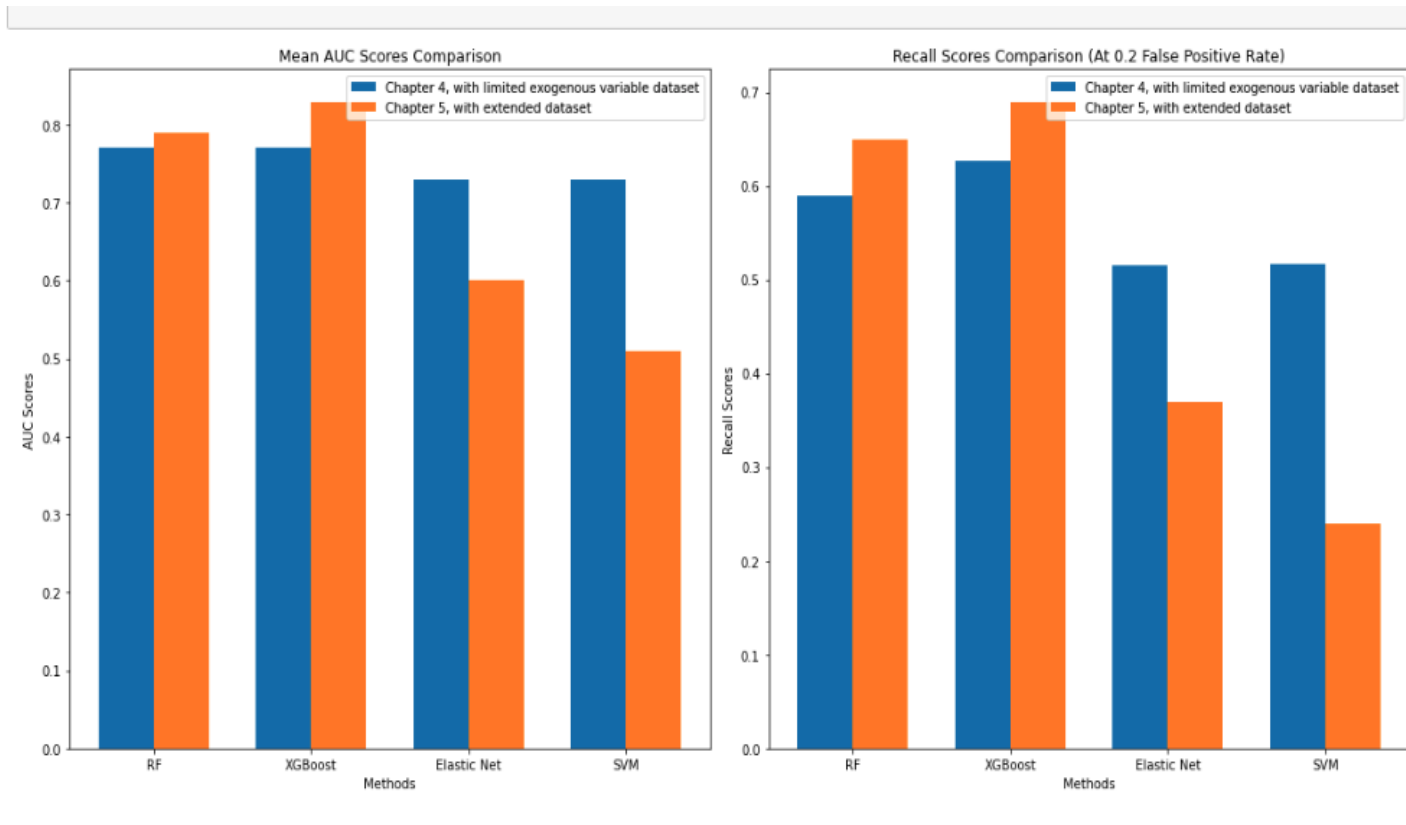**Figure C. 6.** Performance Metrics for the Extended Dataset

184
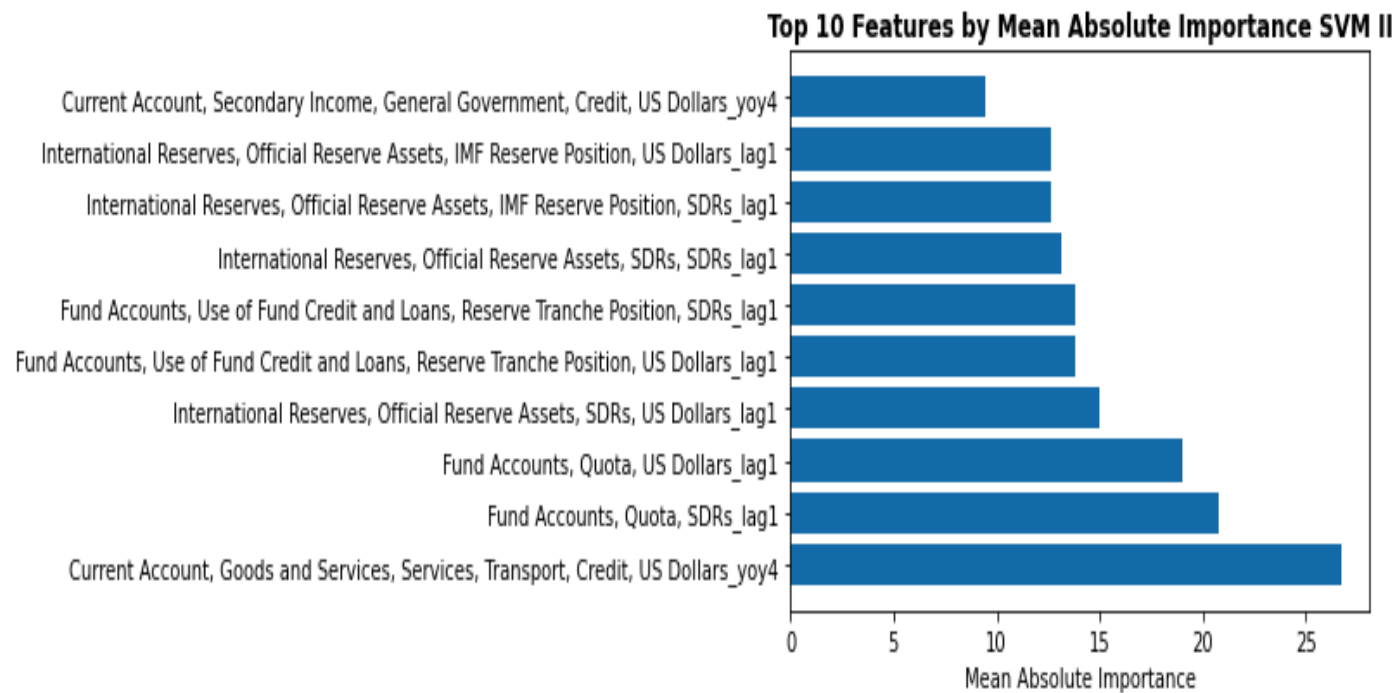
Figure C. 7. Feature Importance SVM II
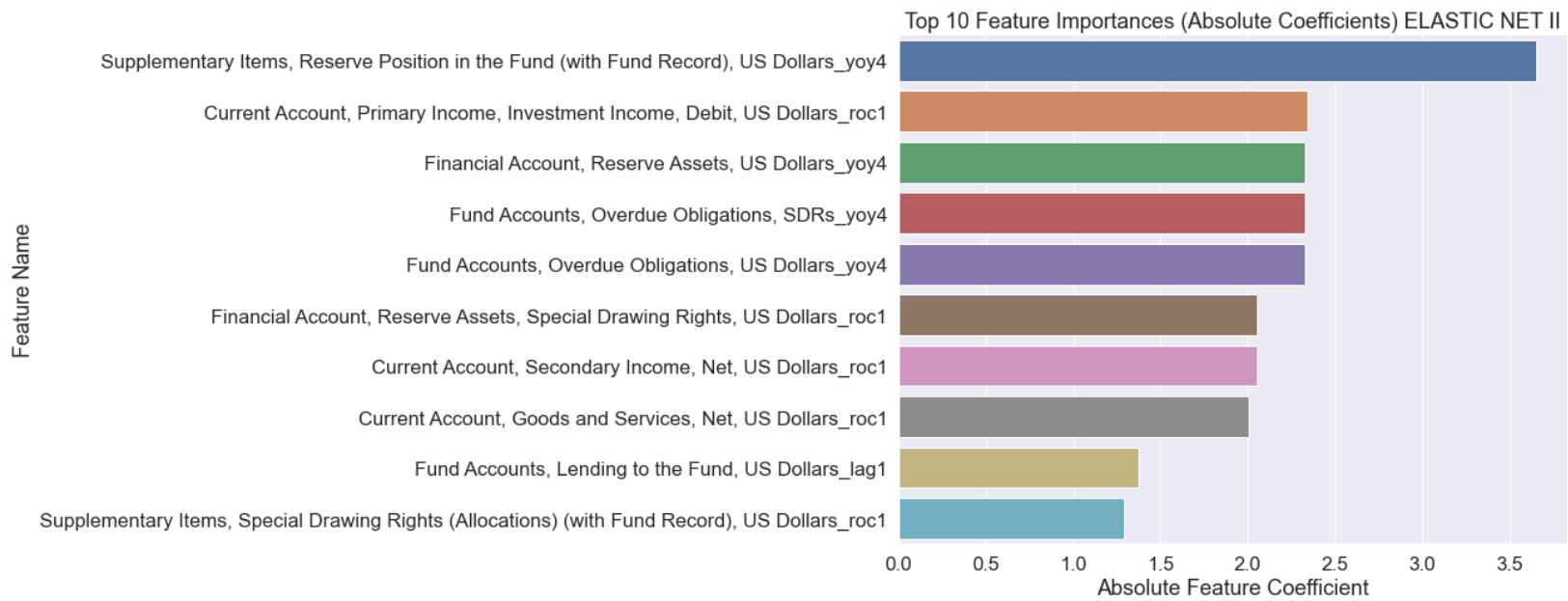
Figure C. 8. Feature Importance Elastic Net II

Figure C. 9. Feature Importance Logistic Regression

Note : Common methods used in both chapters with both datasets are RF, XGBoost, Elastic Net, SVM.

Figure C. 10. Comparison of Out of Sample Performances of Common ML Methods in Chapter 4 and Chapter 5

D. MODEL EVALUATION AT VARIOUS FIXED POSITIVE RATES (FPRs)

Table D. 1. Mean Recall Scores At Fixed False Positive Rate (FPRs) of 0.05
(Chapter 4)

Model	Fixed False Positive Rate	Mean True Positive Rate
XGBOOST I	0.05	0.3852
Random Forest I	0.05	0.3557
ADABoost	0.05	0.3424
Cloglog Model(Base Model)	0.05	0.2976
SVM I	0.05	0.2617
Elastic NET I	0.05	0.2617
KNN	0.05	0.2334
MLP	0.05	0.2187

Table D. 2. Mean Recall Scores At Fixed False Positive Rate (FPRs) of 0.1
(Chapter 4)

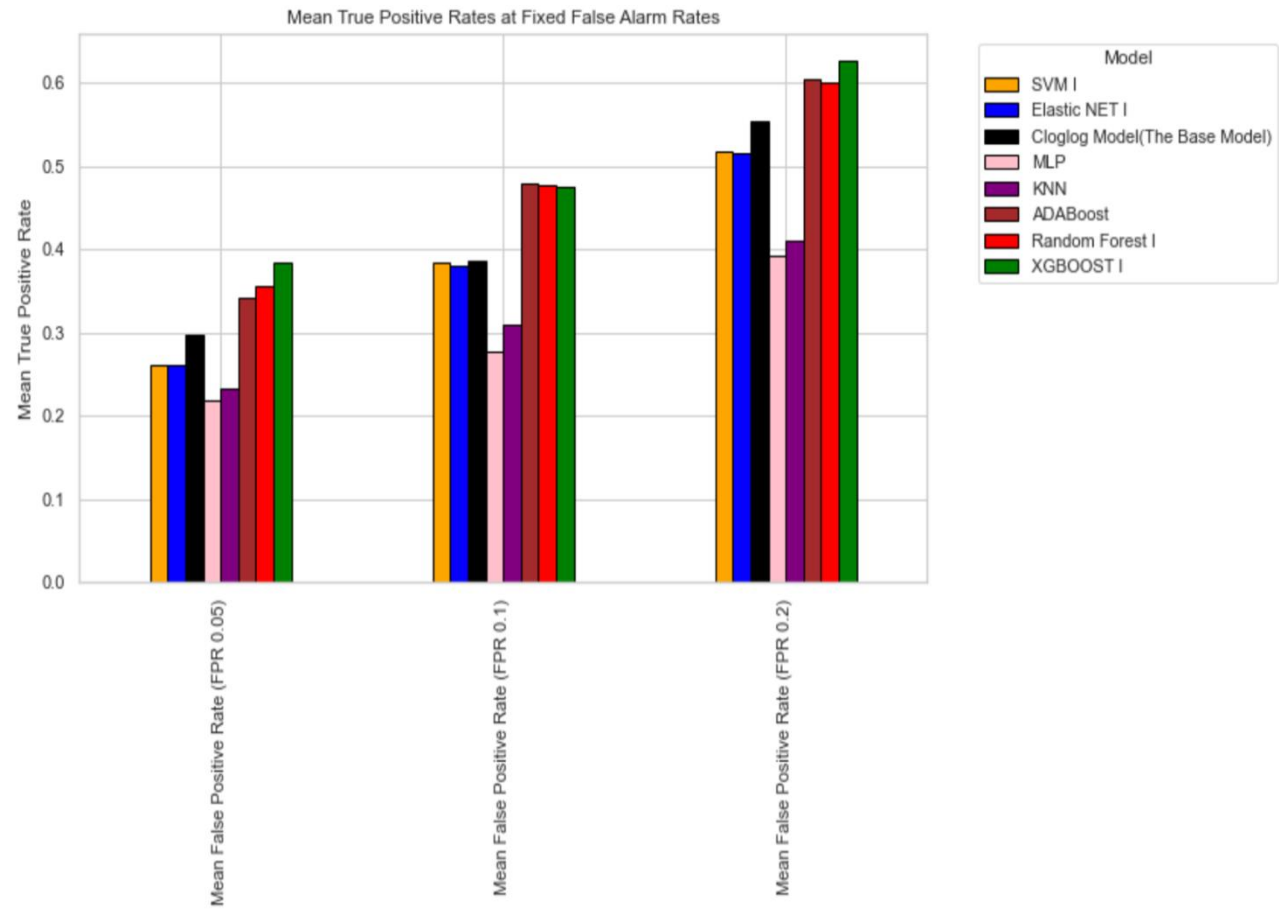| Model	Fixed False Positive Rate	Mean True Positive Rate
XGBOOST I	0.1	0.4791
Random Forest I	0.1	0.4764
ADABoost	0.1	0.4746
Cloglog Model(Base Model)	0.1	0.3870
SVM I	0.1	0.3840
Elastic NET I	0.1	0.3810
KNN	0.1	0.3095
MLP	0.1	0.2781

Table D. 3. Mean Recall Scores At Fixed False Positive Rate (FPRs) of 0.05
(Chapter 5)

| Model	Fixed Mean FPR	Mean TPR
LSTM	0.05	0.753606
Random Forest II	0.05	0.433816
XGBoost II	0.05	0.432177
Logistic Regression	0.05	0.402733
Elastic net II	0.05	0.114150
SVM II	0.05	0.056696

Table D. 4. Mean Recall Scores At Fixed False Positive Rate (FPRs) of 0.1
(Chapter 5)

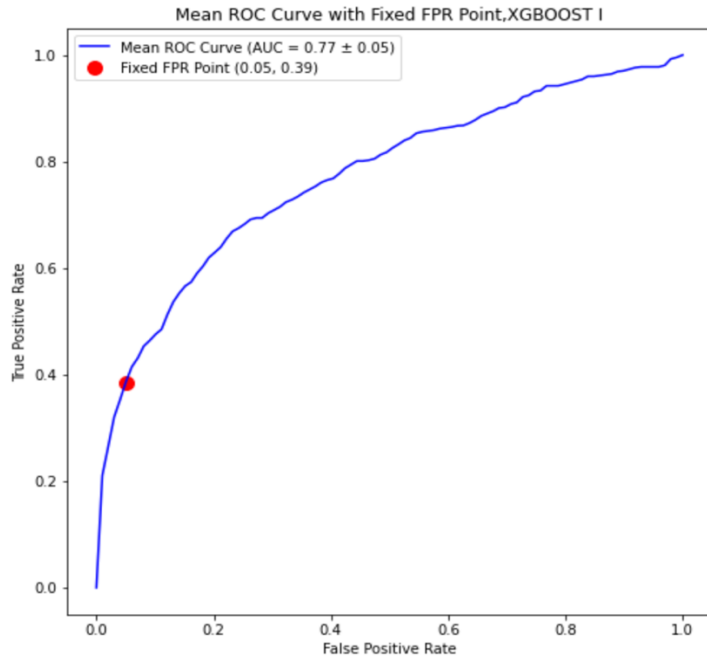| Model	Fixed Mean FPR	Mean TPR
LSTM	0.1	0.820884
XGBoost II	0.1	0.573514
Random Forest II	0.1	0.536150
Logistic Regression	0.1	0.509209
Elastic net II	0.1	0.213058
SVM II	0.1	0.119808

Figure D. 1. Mean Recall Scores at Various Fixed False Positive Rates (FPRs) of 0.05 ,0.1, 0.2 (Chapter 4)
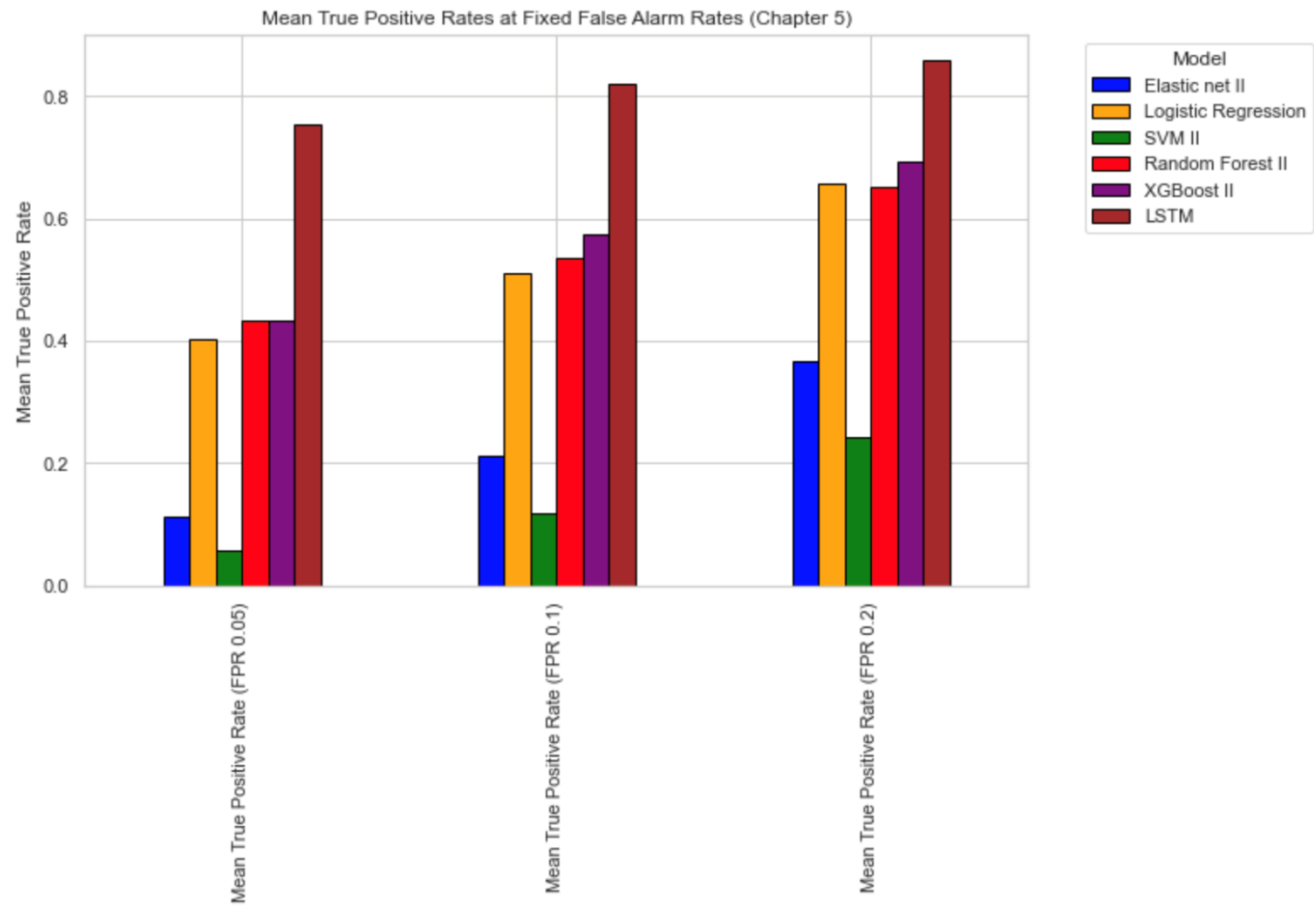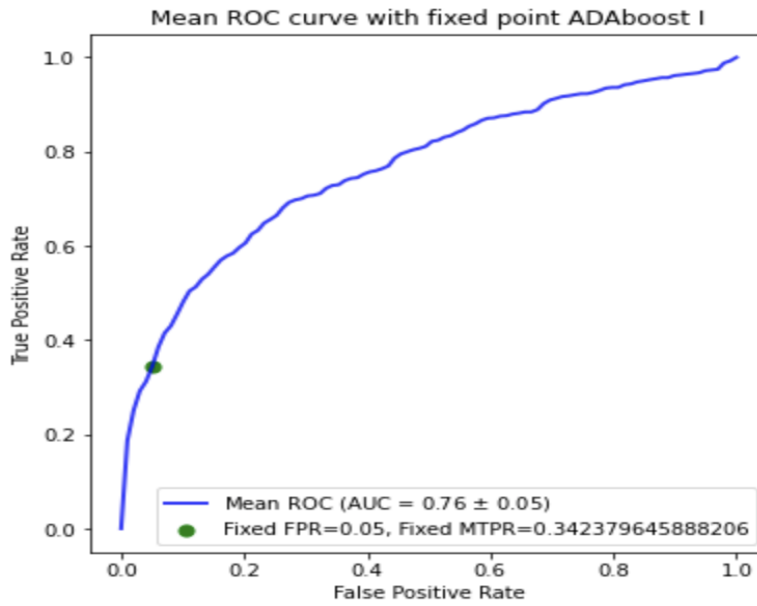
Figure D. 2. Mean Recall Scores at Various Fixed False Positive Rates (FPRs) of 0.05, 0.1, 0.2 (Chapter 5)

Figure D. 3. Mean ROC with the Operating Point at the Fixed FPR of 0.05, XGBoost (Chapter 4)
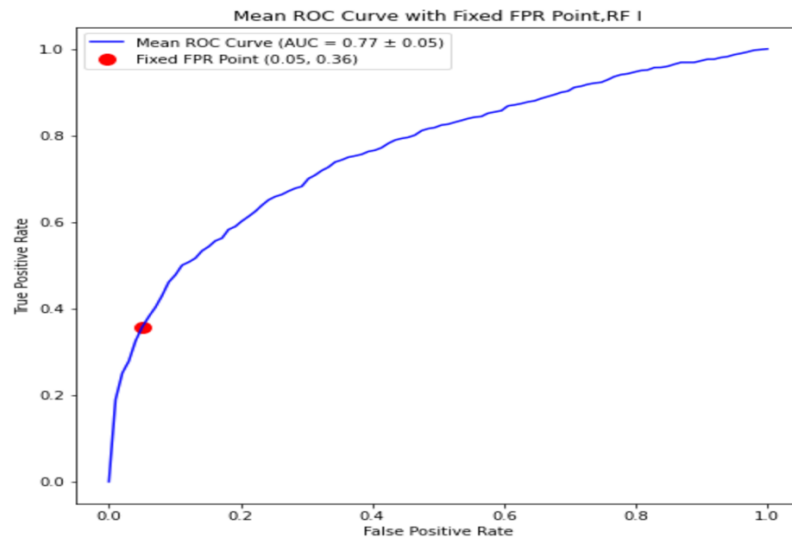
Figure D. 4. Mean ROC with the Operating Point at the Fixed FPR of 0.05, Random Forest (Chapter 4)

Figure D. 5. Mean ROC with the Operating Point at the Fixed FPR of 0.05, AdaBoost (Chapter 4)

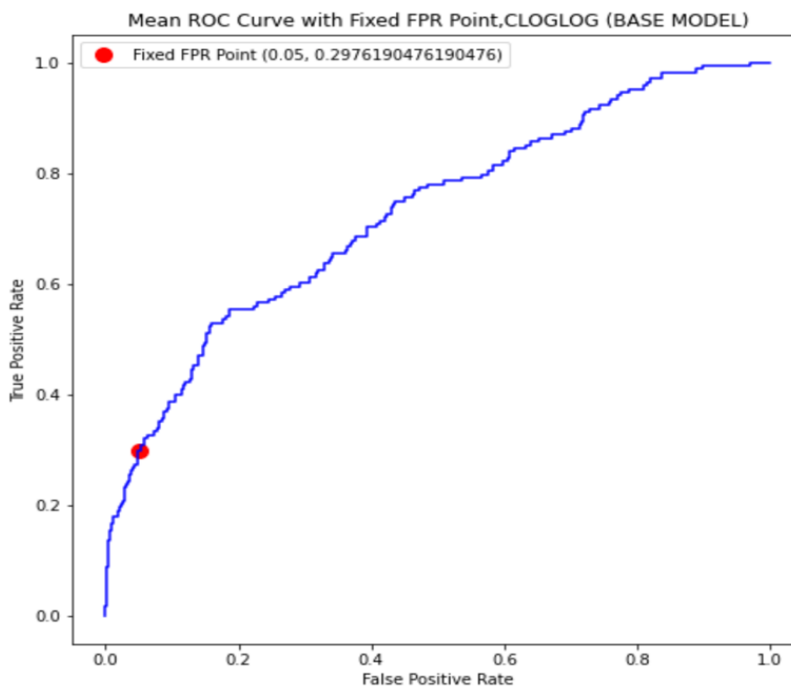**Figure D. 6.** Mean ROC with the Operating Point at the Fixed FPR of 0.05, Cloglog Model (Chapter 4)

Figure D. 7. Mean ROC with the Operating Point at the Fixed FPR of 0.05, SVM (Chapter 4)

Figure D. 8. Mean ROC with the Operating Point at the Fixed FPR of 0.05, MLP (Chapter 4)

Figure D. 9. Mean ROC with the Operating Point at the Fixed FPR of 0.05, Elastic Net (Chapter 4)

Figure D. 10. Mean ROC with the Operating Point at the Fixed FPR of 0.05, KNN (Chapter 4)

Figure D. 11. Mean ROC with the Operating Point at the Fixed FPR of 0.05, Elastic Net (Chapter 5)

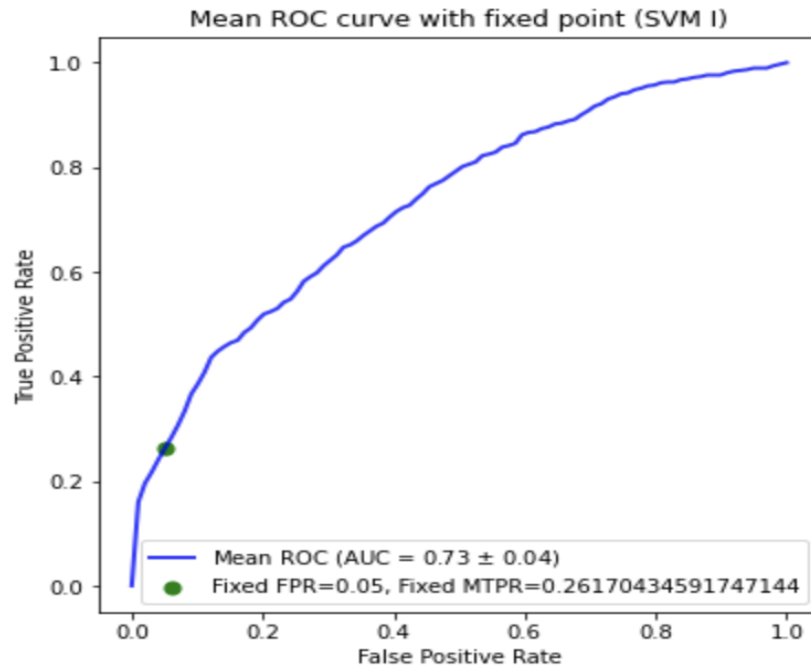**Figure D. 12.** Mean ROC with the Operating Point at the Fixed FPR of 0.05, SVM (Chapter 5)

Figure D. 13. Mean ROC with the Operating Point at the Fixed FPR of 0.05, Random Forest (Chapter 5)
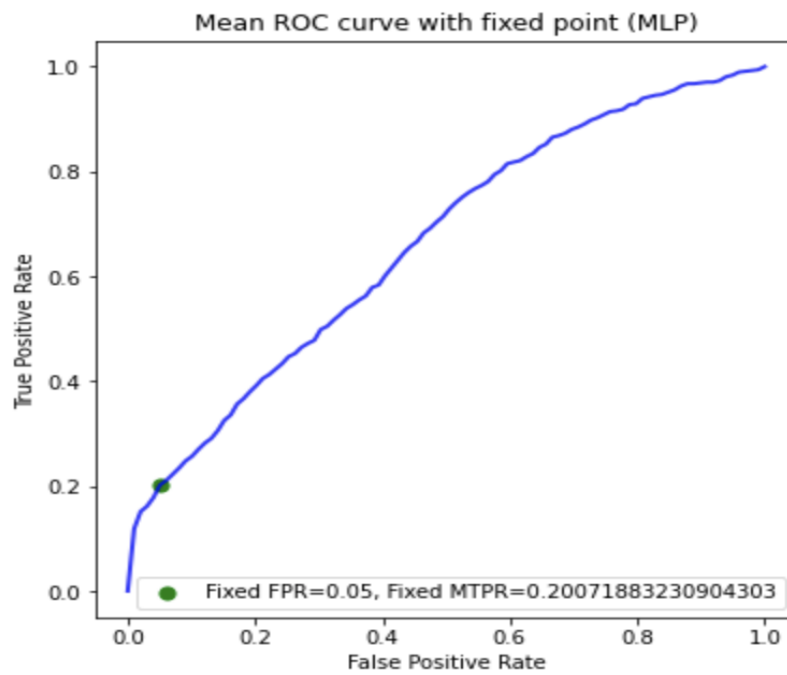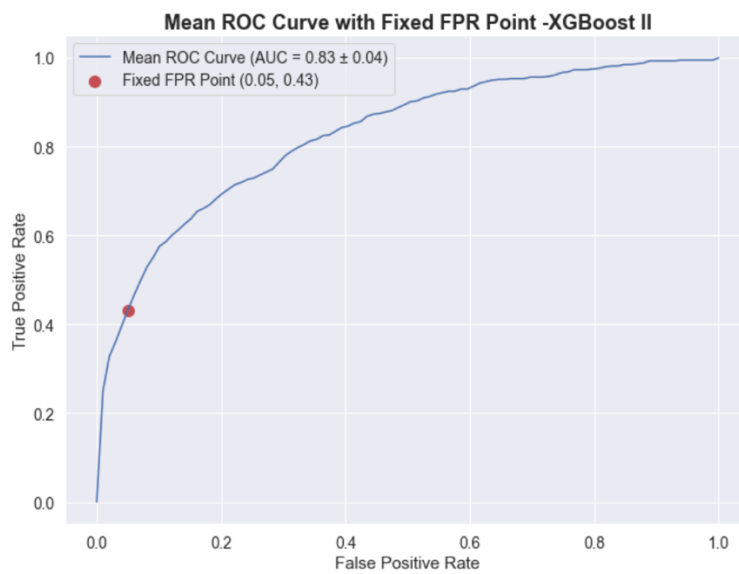

Figure D. 14. Mean ROC with the Operating Point at the Fixed FPR of 0.05, XGBoost (Chapter 5)
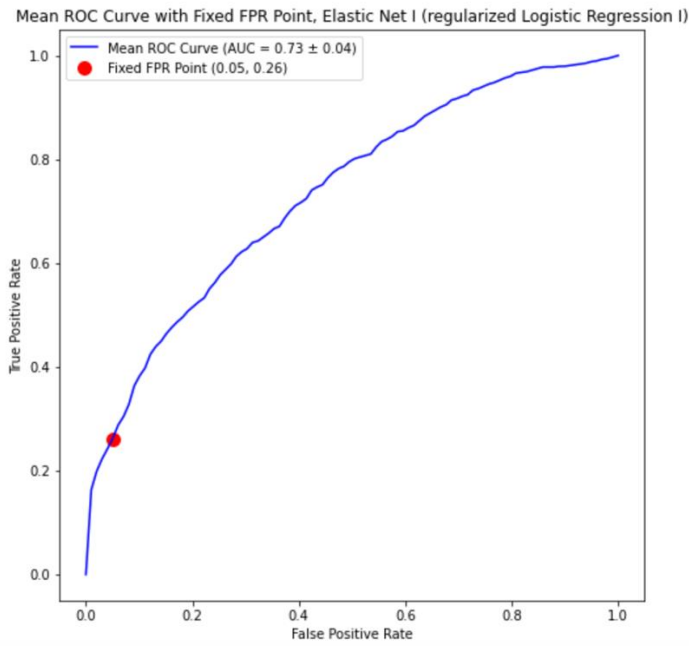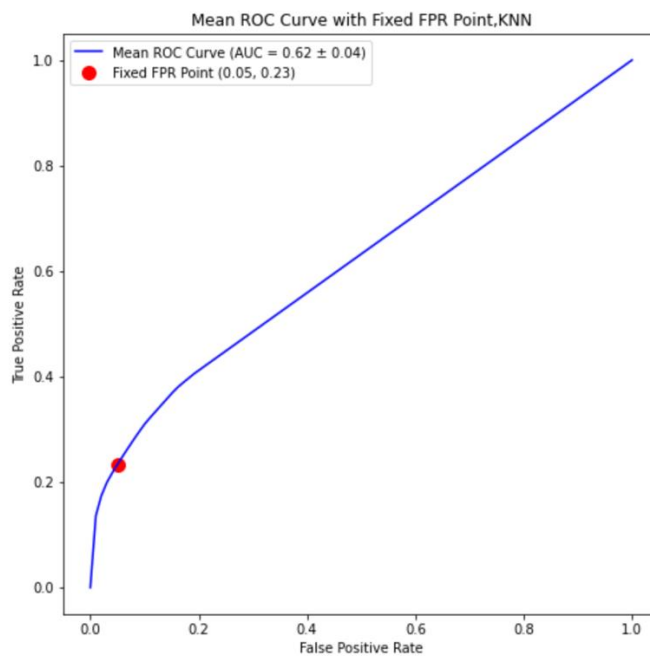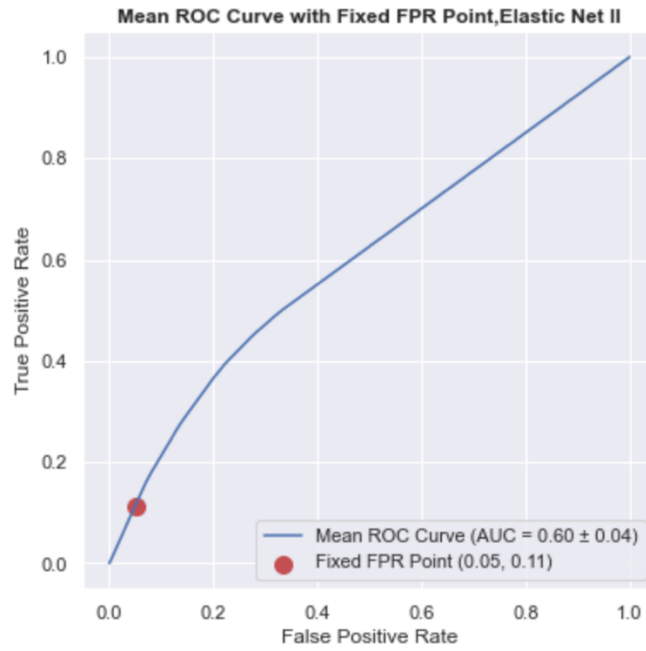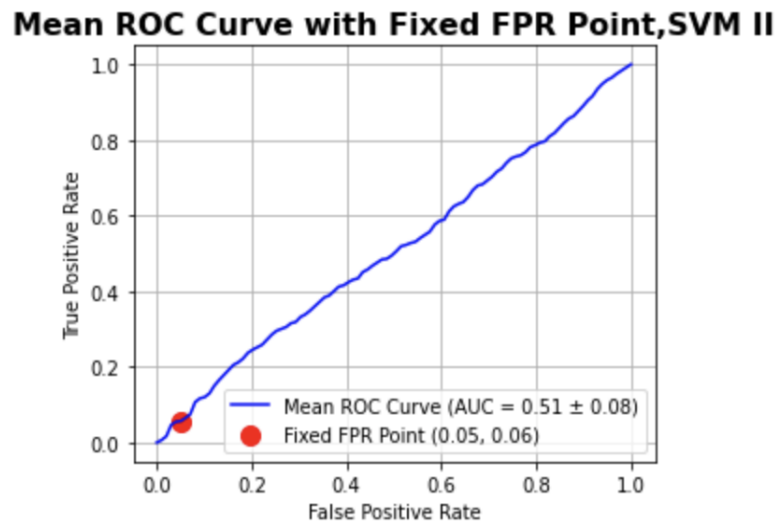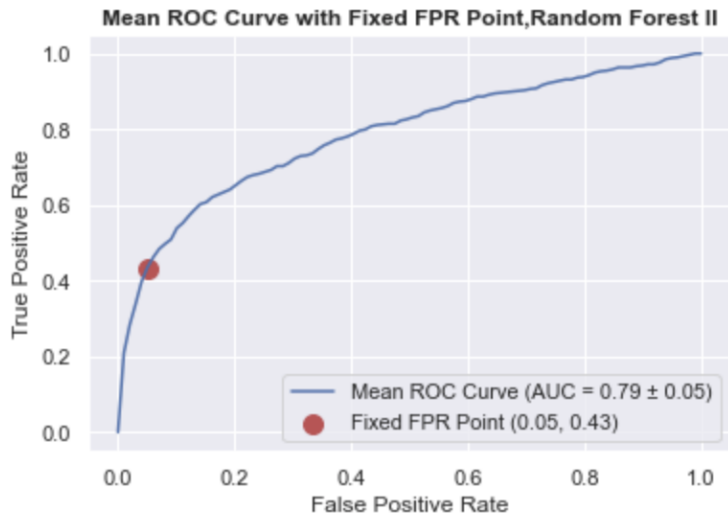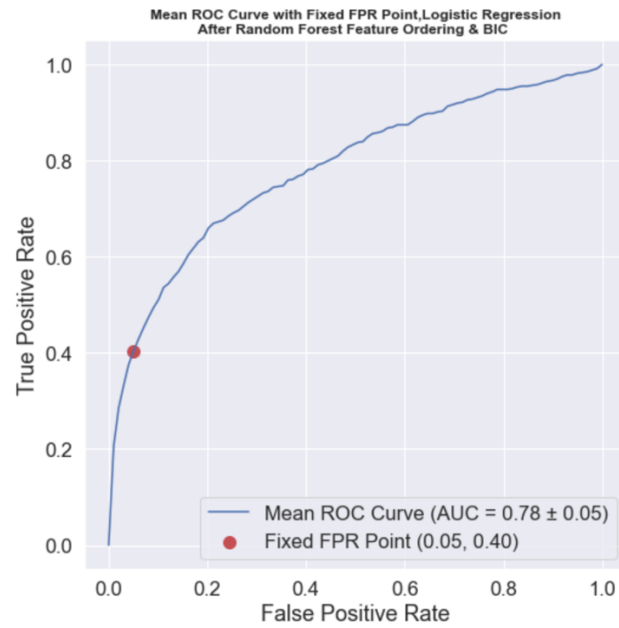
Figure D. 15. Mean ROC with the Operating Point at the Fixed FPR of 0.05, Logistic Regression (Chapter 5)
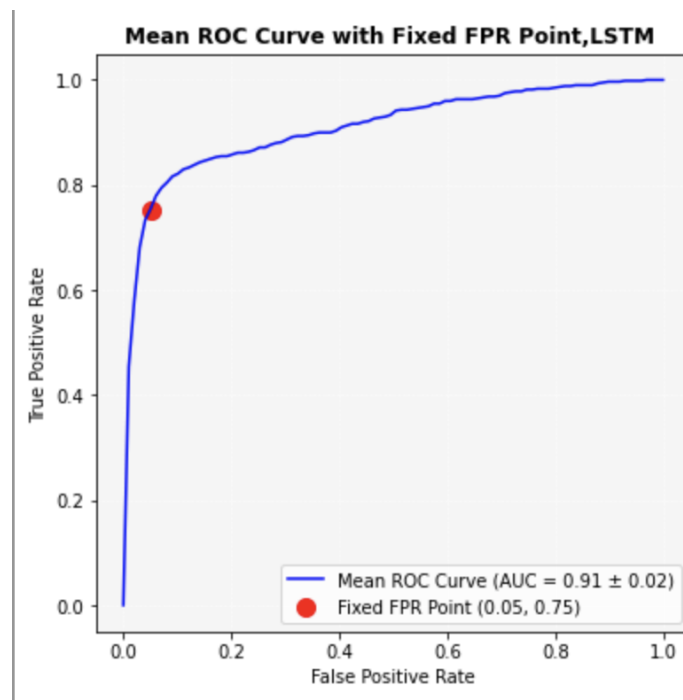
Figure D. 16. Mean ROC with the Operating Point at the Fixed FPR of 0.05, Logistic Regression (Chapter 5)

E. CURRICULUM VITAE

Personal Information

Surname, Name: CELIK, Songul Siva

Education

Ph.D. (2013-2024) Economics, Middle East Technical University, Ankara, Turkey.

M.A. (2011-2013) Economics, Bilkent University, Ankara, Turkey.

B.A. (2005-2010) Economics, Gazi University, Ankara, Turkey.

Exchange Student (2008-2009), FH Kufstein University of Applied Sciences (2008-2009), Kufstein, Austria.

Work Experience

Baskent University, Department of Economics, 2016-2020, Research Assistant.

Bilkent University , Department of Economics, 2011-2013, Teaching Assistant :

Macroeconomics I, Macroeconomics II, Econometrics I, Econometrics II.

Other

Master's Thesis: 'Is A Loan Tax Optimal In the Presence of Leaning Against Loan Winds?' (2013), supervised by Assoc. Prof. Dr. Refet S. Gürkaynak.

Research Interest

Macroeconomics, Monetary Economics, International Finance, Financial Economics.

Technical Skills

Python, Stata, Dynare, Matlab.

Languages

Turkish (Native), English (Fluent).

F. TURKISH SUMMARY / TÜRKÇE ÖZET

Ani Duruşlar (Sudden Stops), sermaye girişlerinin büyük miktarda ve aniden tersine dönmesiyle (yani uluslararası finans piyasalarına ani bir "erişim kaybı") ilişkili bir dizi ampirik düzenlilikler tarafından tanımlanan ekonomik dalgalanmalardır. Ani Duruşun tanımlayıcı özelliği, çoğunlukla cari işlemler hesabındaki ani bir sıçramayla ölçülen dış sermaye girişlerindeki keskin bir tersine dönüştür. Dış finansmana erişimin kaybolmasıyla hemen hemen aynı zamanda veya kısa bir süre sonra, Ani Duruşlardan etkilenen ekonomiler derin durgunluklar, reel döviz kurunda (RER) keskin değer kayıpları ve varlık fiyatlarında düşüşler yaşamaktadır. Üstelik bunların öncesinde genellikle genişleme dönemleri gelir; yüksek kredi büyümeleri, büyük cari işlemler açıkları, değer kazanan RER ve varlık fiyatlarındaki artışları gözlemlenir.

Mendoza ve Korinek (2013), 1990'lı yıllarda gelişmekte olan piyasalarda gözlemlenen Ani Duruşların, 2008-2009 Küresel Mali Krizinin habercisi olduğunu belirtmektedir. Ani Duruş (SS) krizleri, 1990'larda özellikle gelişen piyasaları etkileyen önemli etkileri olduğundan, Yükselen Piyasalar (EM) için birincil endişe kaynağıydı. Ancak bu kriz olgusunun kapsamı gelişmekte olan ülkelerin ötesine geçerek hem gelişmekte olan hem de gelişmiş ülkeleri etkilediği gözlemlendi. Bu durum, 2016 sonu itibarıyla toplam 58 kayıtlı Ani Duruş (SS) olayını belgeleyen Bianchi ve Mendoza (2020) tarafından da desteklenmektedir. Dikkat çekici bir şekilde, bu olaylardan 35'i gelişmekte olan pazarlarda görülürken, 23'ü gelişmiş ekonomilerde ortaya çıkmıştır. Bu durum, Ani Duruşların yaygın bir şekilde meydana geldiğinin ve bunların çeşitli ekonomik ortamlarda etkili bir şekilde ele alınması gerektiğinin altını çizmektedir. Eichengreen ve Gupta (2016), hem gelişmiş ekonomiler (AE) hem de Yükselen Piyasalar (EM) örnekleminde 2000 öncesi dönemi, 2001 ve 2014 arasındaki dönemle karşılaştırdığında, Ani Duruşların sıklığının ve süresinin o zamandan beri büyük ölçüde değişmediğini göstermektedir. Analizlerinin sonucunda, Eichengreen ve Gupta (2016), Ani Duruşların önemli bir ekonomik sorun olmaya devam ettiği sonucuna varmaktadırlar.

Ani Duruşlar ile ilgili çalışmalar Calvo (1998) ile başlayıp sonrasında devam etse de Küresel Mali Kriz (KMK) sonrasında yeniden ilgi görmeye başlamış ve artış göstermiştir. KMK odağını kredi patlaması-çöküş döngülerine, sermaye akışı oynaklığına ve Ani Duruşlara kaydırarak makro ihtiyati politikalar üzerine tartışmalara yol açtı. Forbes ve Warnock (2021), özellikle 2008'in dördüncü çeyreğinden (2008 Ç4) 2009'un ilk çeyreğine (2009 Ç1) kadar uzanan KMK`nın zirve yaptığı dönemde, çeşitli ülkelerde sermaye akışlarında Ani Duruş (SS) dönemlerinin önemli bir oluşumunu fark etti. Dikkat çekici bir şekilde, bu analiz, bu çalkantılı dönemde bu tür Ani Duruş (SS) olgusuyla karşı karşıya kalan toplam 22 ülkenin altını çizdi. Bu ülkeler arasında Arjantin, Brezilya, Estonya, İzlanda, Hindistan, Letonya, Norveç, Peru, Romanya, Rusya, Yunanistan ve Türkiye vardı. Bu çalışmalarında, Forbes ve Warnock (2021), farklı coğrafi bölgelerdeki ekonomilerin sermaye akışlarında ani tersine dönüşlerle karşı karşıya kalması nedeniyle KMK`nın geniş kapsamlı etkilerine ışık tutmakta ve krizin yaygın doğasının ve bunun küresel finansal istikrar üzerindeki etkilerinin altını çizmektedir.

Mendoza ve Terrones (2012), kredi patlamaları ve Ani Duruşlarla ilgili olarak, kredi patlamalarını ekonomik genişleme, gayrimenkul büyümesi ve dış açıklarla ve ardından genellikle gerilemelerle ilişkilendirmektedir. Bu patlamalar, 2008 Küresel Mali Kriz (KMK) gibi büyük olayların etrafında kümelenerek küresel bir eşzamanlılık sergilemektedir. Mendoza ve Terrones (2012), tüm yükselişlerin krize yol açmamasına rağmen, krize yol açtığında bunu bankacılık, döviz krizleri ve Ani Duruşların takip edebileceğini belirtmektedir. Reinhart ve Reinhart (2008) sermaye girişi bolluğunun özellikle gelişmekte olan piyasalar açısından olumsuz sonuçlarını vurgulayarak aynı fikirde olduklarını işaret etmektedirler. Jorda ve ark. (2012) ayrıca, finansal krizin ardından gelen durgunlukların daha şiddetli olduğunu, yoğun kredi genişlemelerin daha derin durgunlukları tetiklediğini vurgulamaktadır. Bu, özellikle gelişmekte olan piyasalarda kredi patlamaları ve Ani Duruş Kriziyle ilişkili artan riskleri vurgulamaktadır.

Küresel Mali Kriz 'in ardından Amerika Birleşik Devletleri'nde (ABD) uygulanan alışılmadık para politikaları ve sıfıra yakın faiz oranları, uluslararası sermayenin gelişmekte olan ekonomilere akışını kolaylaştırdı. Bununla birlikte, 2013'te ABD

Merkez Bankası'nın (FED) menkul kıymet alımlarını azaltabileceğine dair endişelerin yol açtığı " panik reaksiyonu (taper tantrum)" gibi olaylar ve 2015'te ABD faiz oranlarının artacağına dair beklentilerin körüklediği "normalleşme" dönemi, Ani Duruşların gerçekleşme ihtimalini vurguladı. Eichengreen ve Gupta (2016) bu duruma değinerek, Ani Duruşların daha sık hale gelebileceğini veya potansiyel olarak daha yıkıcı hale gelebileceğini belirtti. Forbes ve Warnock (2021), 2008-2009 krizinde gözlemlenen devasa sermaye akışı dalgalarının daha kontrol edilebilir dalgalanmalara dönüştüğünü ancak 2015 yılında kayda değer bir yükselişin meydana geldiğini belirtmektedir. Bu yükseliş, yatırımcıların ABD Merkez Bankası'nın faiz artırımına ilişkin beklentisinden kaynaklanmıştır. Bu bağlamda, örneklemdeki ülkelerin yaklaşık %27'si sermaye akışlarında Ani Duruşlarla karşılaşırken, %22'si ise daralmayla karşı karşıya kalmıştır. Ayrıca, Forbes ve Warnock (2021), Küresel Mali Krizden (KMK) bu yana uluslararası sermaye akışı oynaklığının ve aşırı sermaye akışı olaylarının küresel olarak azalmasına rağmen, bunların nispeten yüksek kaldığını ve 2008 öncesindeki seviyelerle karşılaştırılabilir düzeyde kaldığını göstermektedir. Bu, özellikle Gelişmekte Olan Piyasalar (EM'ler) için Ani Duruşların devam eden öneminin altını çizmektedir.

Ayrıca, politika yapıcılar ve akademisyenler, makro-finansal istikrarı artırmak için sermaye düzenlemeleri ve makro ihtiyati çalışmalara odaklanmalarını yoğunlaştırdılar. 2012 yılında IMF bakış açısını değiştirerek sermaye kontrollerinin potansiyel avantajlarını vurguladı ve sermaye akışlarının yönetilmesine yönelik kapsamlı, esnek ve dengeli bir yaklaşım konusunda daha fazla araştırma yapılması çağrısında bulundu. Daha yakın zamanda IMF (2022), makroekonomik ve finansal istikrarı korumak için ülkelerin borç girişlerini proaktif olarak kısıtlama seçeneğine sahip olmasının önemini vurguladı. Bu öneri özellikle dış yükümlülüklerin risk teşkil ettiği durumlarda, özellikle de uygun yabancı para varlıkları veya riskten korunma önlemleri olmadan yabancı para cinsinden dış borç nedeniyle para birimi uyumsuzluklarına neden oldukları durumlarda geçerlidir.

Bu yenilenen vurgu, özellikle Gelişmekte Olan Piyasalar (EM) ve bu piyasaların ciddi sermaye akışı kesintilerine karşı kırılganlıkları bağlamında Ani Duruşları tahmin etme ve önleme konusundaki hayati ihtiyacın altını çizmektedir. Ani Duruş

Krizlerini öngörmeyi amaçlayan ampirik çalışmalarda genellikle lojistik regresyon, probit veya tamamlayıcı logaritmik (cloglog) yöntemler gibi geleneksel ekonometrik yöntemler yaygın olarak kullanılmaktadır. Ancak bu yöntemlerin, sonuçların tahmin etme gücü ve genellenebilirliğini etkileyebilecek sınırlamaları vardır. Bu yöntemlere alternatif olarak ve veri yoğun bir yaklaşımı benimseyerek amacımız, Ani Duruş olaylarını tahmin etmede makine öğrenimi yöntemlerinin potansiyelini araştırmaktır. Bu yöntemlerin bu sınırlamaları etkili bir şekilde ele alıp alamayacağını ve daha yüksek örnek dışı performans sağlayıp sağlayamayacağını, böylece daha genelleştirilebilir modellere yol açıp açamayacağını araştırmayı amaçlıyoruz. Makine öğrenimi, büyük veriden yararlanmak, alışılmadık veri kümeleri oluşturmak veya kümeleme yöntemlerini kullanmak gibi farklı amaçlar için çeşitli değerli araçlar sunarken, temel odak noktamız, makroekonomik alanda örnek dışı tahmininde makine öğreniminin yüksek tahmin performansıdır.

Makine öğrenimi yöntemleri genellikle yüksek doğrulukları ve örneklem dışı tahmin güçleri ile tanınır; bu da daha genelleştirilebilir modeller anlamına gelir. Çoğu makine öğrenimi yöntemi, verilerin dağılımı ve değişkenler ile sonuçlar arasındaki ilişkinin doğrusallığı konusunda güçlü varsayımlar uygulamaz. Bu esneklik, makroekonomik krizlerin doğasında var olan karmaşıklıkların ortaya çıkarılmasına yardımcı olur. Üstelik, makine öğrenimi yöntemlerinin metodolojik özellikleri, yalnızca parametre tahminine ve örneklem içi tahmine odaklanmak yerine daha yüksek tahmin doğruluğu elde etmek üzere eğitilebildikleri için örneklem dışı tahmini geliştirebilirler. Geleneksel ekonometrik modeller, genellikle modeli eğitim verilerine (training set) tam olarak uydurma yetenekleri ile bilinirler, yani bu modeller örneklem içi tahmin performansı açısından genellikle yüksek doğruluk sağlarlar. Bunula birlikte, yeni, görünmeyen verilere genelleme yeteneğini ölçen örnek dışı tahmin performansı, verilerdeki daha karmaşık ilişkileri yakalama yeteneklerinden dolayı makine öğrenimi modelleri için genellikle daha yüksektir.

Makine Öğrenmesi (MÖ) temelde, değişkenler arasındaki karmaşıklıkları verilerden öğrenme esnekliği ile güçlü bir tahmin yeteneğine sahiptir. Ancak, her iki yöntemde (geleneksel ekonometri ve MÖ yöntemleri) de kabul edilmesi gereken sınırlamalar bulunmaktadır. Birincisi, makroekonomik veri setlerinde genellikle sınırlı sayıda

gözlem içeren küçük örneklem boyutları, tahmin yeteneğini olumsuz etkileyebilmektedir. Küçük örneklem boyutları, aşırı uydurma/uyum (yani, tahmin için kullanılan örneğe özgü olan ve diğer örnekleri kapsamayan kalıpların keşfedilmesi) veya geleneksel istatistik terimleriyle sahte regresyon (spurious regression) gibi sorunlara yol açabilmektedir. Ayrıca, krizlerin seyrek olması örneklemde dengesizliklere katkıda bulunabilir ve sapmalı tahminlere neden olabilir, bu da toplu tahmin doğruluğunu azaltabilir. Bu zorluklar, geleneksel ekonometrik modellerin veya MÖ modellerinin Ani Duruş (SS) olaylarını tahmin etmek için kullanılması durumunda her iki tahmin modelinde de ortak olabilir. Ancak, endişenin odak noktası bu sınırlamalar değildir.

Üstelik MÖ modelleri, karmaşıklıkları nedeniyle, veriler içindeki karmaşık ilişkileri etkili bir şekilde öğrenmek için genellikle daha büyük veri kümeleri gerektirmektedir. Bunun tersine, daha basit modeller, özellikle de geleneksel doğrusal modeller, veri kümesi küçük olduğunda daha iyi performans gösterebilmektedir. MÖ yöntemlerinin bu zorluklara karşı tamamen bağışık olmadığını belirtmek önemlidir; ancak MÖ modelleri, bu sorunları azaltmak için düzenleme, çapraz doğrulama ve hiper parametre ayarlama gibi araçlarla donatılmıştır. MÖ yöntemleri, bu araçlardan yararlanarak küçük örneklem boyutlarından ve nadir olaylardan kaynaklanan sorunları en aza indirmeyi amaçlamaktadır. Bu nedenle ilgimiz, bu MÖ yöntemlerini yalnızca küçük örneklem boyutları ve dengesiz veri kümeleri bağlamında değil, daha belirgin bir şekilde Ani Duruş krizlerindeki karmaşıklıkları ortaya çıkarmadaki spesifik yetenekleri açısından araştırılması üzerinedir.

Bu çalışmanın öncelikli hedefi, Ani Duruşları Makine Öğrenimi (MÖ) yöntemleri kullanarak tahmin etmek ve bu tahminlerin örnek dışı performansını değerlendirmektir. Analiz, iki ana bölümden oluşmaktadır. İlk aşamada, geleneksel yöntemleri temsil etmek amacıyla Forbes ve Warnock (2021)'un temel modelini kullanarak bir temel oluşturulmaktadır. Bu temel modelin örnek dışı tahmini hesaplanarak başlanmaktadır. Ardından, aynı değişkenleri ve veri kümesini kullanarak bir dizi Makine Öğrenimi yöntemini (Elastik Ağ, Rastgele Ormanlar,

Destek Vektör Makineleri, k-En Yakın Komşular, AdaBoost ve XGBoost ile Çok Katmanlı Algılayıcı) değerlendirmekteyiz.

Analizimizde, bu MÖ yöntemlerinin örnek dışı performansını temel modele göre kapsamlı bir şekilde karşılaştırmaktayız. Bu karşılaştırma, doğruluk, hassasiyet, kesinlik, geri çağırma, F1 puanı ve İşlem Karakteristik Eğrisi (ROC) eğrisi ile AUC_ROC (ROC Eğrisi Altındaki Alan) skoru gibi çeşitli performans ölçümlerini içermektedir. Bu analiz, MÖ algoritmalarının Ani Duruşları tahmin etme konusundaki örnek dışı performanslarına nasıl katkıda bulunabileceğini anlamak amacıyla gerçekleştirilmektedir.

İkinci aşamada, ilk aşamada kullandığımız veri setini, Uluslararası Para Fonu'nun (IMF) çeşitli çeyreklik veri setlerini tek bir kriter olan veri bulunabilirliği ilkesini dikkate alarak genişletiyoruz. Veri setini genişlettikten sonra, aşırı uyumu (overfitting) engellemek için en önemli bağımsız değişkenleri seçmek için Makine Öğrenimi (MÖ) yöntemlerini kullanarak ,tahmin için en az önemli değişkenleri eliyor ve daha sonra seçilen bağımsız değişkenleri ML tahmini için kullanıyoruz. Daha sonra, yöntemler arasında örnek dışı tahmin performans karşılaştırmaları yapmaktayız. Bu bölüm, Elastik Ağ, Rastgele Ormanlar, Destek Vektör Makineleri, XGBoost, Lojistik Regresyon ve derin öğrenme yöntemlerinden biri olan Uzun Kısa Süreli Bellek (LSTM) de dahil olmak üzere bir dizi MÖ yöntemini içermektedir. Bu kapsamlı analiz, makine öğrenimi algoritmaları ve bağımsız değişken seçme teknikleri tarafından desteklenen örnek dışı performanslardaki potansiyel iyileştirmeleri keşfetmeyi amaçlamaktadır.

2. Bölüm `de , Ani Duruş Krizi üzerine ampirik ve teorik çalışmaları kapsayan ve konunun kapsamlı bir şekilde anlaşılmasını sağlayan literatür taramasını derinlemesine incelemekteyiz. Bunu takiben Bölüm 3'te makine öğreniminin (MÖ) ekonomiyle birleşimini araştırıyoruz.

Bu bölümde, "İstatistiksel Paradigmalarda Devrim Yaratmak: Algoritmik İçgörülerle Veri Karmaşıklığında Gezinmek " ("Revolutionizing Statistical Paradigms: Navigating Data Complexity with Algorithmic Insights") alt bölümü de dahil olmak

üzere, bu entegrasyonun çeşitli yönlerini incelemekteyiz. Burada, makine öğreniminin ekonomide kademeli olarak benimsenmesine katkıda bulunmuş olabilecek faktörleri araştırmaktayız. İstatistikçi Leo Breiman ve iktisatçı Hal R. Varian gibi etkili şahsiyetlerden bilgiler alarak, makine öğreniminin ekonomide kabul edilmesini kolaylaştıran değişen kültürel dinamikleri ortaya çıkarıyoruz. Ayrıca bu birleşmeden doğan potansiyel faydaları da açıklıyoruz.

''Denetimli Makine Öğrenimini Kullanarak Ani Duruşları Tahmin Etme (Predicting Sudden Stops Using Supervised Machine Learning")'' alt bölümüne geçerek, denetimli MÖ teknikleri aracılığıyla Ani Duruşları tahmin etmeye yönelik stratejimizi özetliyoruz. MÖ ortamında tahmin sorunumuzu nasıl ayarlayacağımızı açıklıyoruz. Tahmin problemini ikili sınıflandırma problemi olarak ayarlamaktan başlayarak, veri setini test ve eğitim alt kümelerine bölme, model seçimi gibi gerekli temel adımları kısaca ele almaktayız. Model genellemesinin önemini vurgulayarak çapraz doğrulamayı ve hiperparametre ayarlamasını araştırmaktayız. Ayrıca, Tip 1 (Yanlış Pozitif) ve Tip 2 (Yanlış Negatif) hata türlerini ve Ani Duruşlardaki yorumlamayı ve bunların tahmin doğruluğunu nasıl etkilediğini açıklıyoruz.

Hata tiplerini anlamak, tahmine dayalı modellemede, özellikle ikili sınıflandırmada çok önemlidir. Bu bağlamda iki önemli hata türü Tip 1 (Yanlış Pozitif) ve Tip 2'dir (Yanlış Negatif). Tip 1 Hata, modelin gerçekleşmeyen olumlu bir sonuç öngörmesi durumunda ortaya çıkar. Ani Duruşlar (SS) için bu, yaklaşan bir Ani Duruşlar (SS) etkinliğini yanlış tahmin etmek anlamına gelebilir. Tersine, Tip 2 Hata, modelin ortaya çıkan olumlu bir sonucu tahmin etmede başarısız olması durumunda ortaya çıkar. Ani Duruşlar (SS) da bu, yaklaşmakta olan bir krizin işaretlerinin gözden kaçırılmasını içerebilir.

Ani Duruşlar (SS)`da bu hataların ekonomiyi ve karar vericileri etkileyen önemli sonuçları vardır. 1. Tip Hata veya "Yanlış Pozitif" finansal piyasaları, kurumları ve kamuoyunun algısını bozarak ekonomik kısıtlamalara ve paniğe yol açabilir. İronik bir şekilde, sorunları önlemeye yönelik tedbirler volatiliteyi yoğunlaştırabilir. Öte yandan, 2. Tip Hata veya "Yanlış Negatif", gerçek bir Ani Duruşu kaçırarak

ekonominin istikrarını etkiler. Hazırlık eksikliği ve kaçırılan müdahale fırsatları krizi daha da kötüleştirerek durgunluğa, yatırımcı güveninin azalmasına ve devlet borcu ödemelerinde sıkıntıya yol açabilir.

Hata türlerinin tanımlanması sonrasında, model seçimi sürecinin temel adımlarını belirliyor ve bu süreçte doğruluk (accuracy), hassasiyet(precision), geri çağırma (recall),F1-Score, AUC-ROC ve Precision-Recall Eğrisi gibi performans ölçütlerini açıklıyoruz. Bu performans ölçütlerinin, model seçimine yönelik rehberlik rollerini vurguluyoruz. Performans ölçütleri, tahminlere yönelik modellemede modelin etkinliğini nicelendirmek ve değerlendirmek açısından kritik bir rol oynamaktadırlar. Bu ölçütler, modellerin çıktılarının gerçek dünya sonuçlarıyla uyumunu değerlendirmek amacıyla yapılandırılmış bir yaklaşım sunarak, karar vericilere stratejik yaklaşımlar geliştirmelerinde yardımcı olur. İkili sınıflandırma bağlamında, verilerin iki sınıfa ayrılma hedefine odaklandığımızda, bu ölçütler özellikle önemlidir. Model seçimi süreci, performans ölçütlerini kullanarak çeşitli modelleri karşılaştırmayı içermekte ve bu karşılaştırmalar, en uygun modelin belirlenmesine katkı sağlamaktadır.

Ölçütlerin seçimi, politika hedeflerinden ve en aza indirilmeyi amaçlanan belirli hata türlerinden etkilenir. Başka bir deyişle, farklı modelleri değerlendirirken performans metriklerinin seçimi, azaltılması veya kaçınılması gereken hata türleri dikkate alınarak yürürlükteki politikaların belirlediği hedeflere ve önceliklere göre uyarlanır. Örneğin, yanlış negatifleri (Tip II hata) en aza indirmeye öncelik veren bir politika yapıcının geri çağırmayı (recall), F1 puanını veya AUC-ROC puanlarını kontrol etmesi daha doğrudur. Diğer durumu varsayalım, eğer bir politika yapıcı yanlış pozitifleri (Tip I hata) en aza indirmeye öncelik veriyorsa, hassasiyet (precision), F1 puanını veya AUC-ROC puanlarını karşılaştırması daha doğrudur. Bu alt bölüm, özet olarak, Ani Duruş tahmini için temel kavramları ve sistematik adımları birleştiren yapılandırılmış ve kısa bir strateji sunar.

Bölüm 4'te ana hedefimiz, çeşitli Makine Öğrenimi yöntemlerini ve geleneksel istatistiksel yaklaşım olan Genelleştirilmiş Doğrusal Model (GLM) çerçevesinde bulunan tamamlayıcı log-log yöntemini (cloglog) kullanarak Ani Duruş tahmininin

örnek dışı performansını karşılaştırmalı bir analiz yapmaktır. Başlangıç olarak, 1978'in 1. çeyreğinden 2020'nin 3. çeyreğine kadar olan dönemde hem gelişmiş hem de gelişmekte olan ekonomileri kapsayan 59 ülkedeki Ani Duruş Krizlerini belirlemek için Forbes ve Warnock (2021) tarafından oluşturulan sermaye akışı veri setini kullanıyoruz. Daha sonra Forbes ve Warnock (2012) tarafından oluşturulan brüt akışları kullanan Ani Duruş Tanımlama metodolojisini açıklayıp veri seti üzerinde uyguluyoruz. Sonraki bölüm, önemli bulguların vurgulandığı sonuçların sunulmasına ayrılmıştır. Bunu takiben, tamamlayıcı log-log modelini kullanarak, analizlerinde sunulduğu gibi temel durum için tahmin sürecini yineliyoruz. Tahmin dönemi, 1986'nın 1. çeyreğinden 2018'in 4. çeyreğine kadar uzanmakta olup dışsal değişkenler olarak küresel likidite, küresel risk (VIX), küresel büyüme, Birleşik Krallık, ABD, Euro Bölgesi ve Japonya'nın ortalama uzun vadeli faiz oranlarının yanı sıra yayılma ve yerel reel GSYİH büyümesi kullanılmaktadır. Daha sonra, örnek dışı performansını inceleyerek bunu temel senaryomuz olarak oluşturuyoruz.

Ardından, aynı veri kümesini kullanarak çeşitli denetimli Makine Öğrenimi yöntemlerini uyguluyoruz ve bu yöntemlerin ilgili örnek dışı performanslarına ilişkin karşılaştırmalı bir analiz sunuyoruz. Tahmin sonuçlarına geçmeden önce, bu bölümde kullanılan Elastik Ağ, Rastgele Ormanlar (Random Forest), Destek Vektör Makineleri (SVM), kNN (k-En Yakın Komşular), AdaBoost (Uyarlanabilir Arttırma), XGBoost ve Çok Katmanlı Algılayıcı (MLP) yöntemleri hakkında teknik olmayan kısa açıklamalar yapılmıştır.

Bölüm 5'te, Ani Duruş tahmin problemine çözüm getirmek amacıyla tasarlanmış makine öğrenimine dayalı çözümleri sunmaktayız. Metodolojimiz, Uluslararası Para Fonu'ndan (IMF) elde edilen kapsamlı üç aylık verileri entegre ederek, önceki bölümde kullanılan veri setine yönelik genişletici bir yaklaşım içermektedir. Bu zenginleştirilmiş veri seti, önceki tahminlerimizde kullanılan modelde seçilmiş değişkenlere ek olarak, Forbes ve Warnock (2021) tarafından tanımlanan ek değişkenlerle genişletilerek önemli ölçüde artırılmıştır.

Başlangıçtaki veri seti, ampirik çalışmalarda yaygın olarak kullanılan modele özgü değişkenleri içermektedir. Ancak genişletme süreci, temel motivasyonunu esas

210

olarak veri varlığından alır. Değişken seçimi, teorik temellere uygunluğundan veya Ani Duruş olaylarını tahmin etmede yaygın kullanımından bağımsız olarak gerçekleştirilmiştir.

Veri kümesi uzantısının belirli veri odaklı değişkenlerin önceliklendirilmesini içermediğinin altını çizmek önemlidir. Bunun yerine, ön seçim olmaksızın daha geniş bir değişken kümesinin dahil edilmesini içerir. Bu bağlamda, eşdoğrusallık hususları gibi geleneksel istatistiksel kriterler bir kenara bırakılmıştır. Metodolojimiz, makine öğrenimi (MÖ) değişken seçimi yöntemlerine sağlam bir şekilde güvenmeye dayanmaktadır. Bu yöntemler, genişletilmiş veri kümesindeki ilgili değişkenleri nesnel olarak tanımlama ve önceliklendirme konusundaki kanıtlanmış yetenekleri nedeniyle seçilmiştir.

Bu bölümdeki hedeflerimiz ikili bir odağı kapsamaktadır. İlk olarak Ani Duruş olaylarının oluşumunu etkileyen temel değişkenleri belirlemeye çalışıyoruz. İkinci olarak, seçilen değişkenlerden yararlanarak amacımız, örnek dışı veriler üzerinde üstün performans sergileyen, yüksek düzeyde ayırt edici makine öğrenimi modelleri oluşturmaktır. Bu iki yönlü yaklaşım, Ani Duruşları tahmin etmede tahmine dayalı modellerimizin hassasiyetini ve etkinliğini artırmaya yöneliktir.

Bölüm 5'te izlenecek adımları şu şekilde özetleyebiliriz: İlk olarak, veri setimizi genişletme sürecini detaylandırıyoruz. Bu genişletme kapsamında, eklediğimiz değişkenlere 1 çeyrek gecikmeli versiyonları, yıldan yıla yüzde değişimleri ve değişim oranı versiyonları gibi farklı veri versiyonlarını nasıl dahil edeceğimizi açıklıyoruz. IMF'nin Ödemeler Dengesi, Uluslararası Mali İstatistikler ve Ticaretin Yönü veri setlerinden toplamda 192 değişken elde ediyoruz. Diğer versiyonların da eklenmesiyle birlikte, bu süreçten elde edilen son değişken seti toplamda 768'e ulaşmaktadır. Değişkenlerin mevcut seviyelerini hariç tutarak, bu seti Forbes ve Warnock (2021) tarafından seçilen 28 değişkenin 1 dönem gecikmeli versiyonlarıyla tamamlıyoruz. Sonuç olarak, toplamda 604 dışsal değişken içeren bir set elde ediyoruz. Veri seti, 1994q1'den 2018q4'e kadar uzanmaktadır. Bu süreçle ilgili daha fazla ayrıntı, bu alt bölümde detaylı bir şekilde açıklanmaktadır.

Daha sonra, her model için dışsal değişken seçim yöntemlerini detaylı bir şekilde ele alıyoruz. Farklı modeller için farklı yöntemlerin kullanıldığına dikkat etmek önemlidir. Rastgele Orman, XGBoost ve SVM modelleri için dışsal değişken seçimi için Özyinelemeli Özellik Eliminasyonu (Recursive Feature Elimination) yöntemini tercih ederken, Elastik Ağ modelinde değişken seçimi için Lasso yöntemini kullanıyoruz. Lojistik Regresyon modelinde ise başlangıçta dışsal değişken sıralaması için Rastgele Orman kullanmakta ve daha sonra daha fazla seçim için adım adım Bayesian Bilgi Kriteri (BIC) kullanan hibrit bir yaklaşımı benimsemekteyiz.

Daha sonra, dışsal değişken seçimini tamamladıktan sonra tahmin sonuçlarını açıklıyor ve modelleri seçilen dışsal değişkenlere yeniden uyguluyoruz. Ardından, örnek dışı tahmin performanslarına odaklanarak analizimizi devam ettiriyoruz. Bu bölümde ayrıca Uzun Kısa Süreli Bellek (LSTM) yönteminin teknik olmayan kısa bir açıklamasını da sunuyoruz.

Ani Duruş olaylarını tahmin etmeye yönelik kapsamlı araştırmamızın son bölümünü oluşturan Bölüm 6, çeşitli analitik metodolojiler kullanarak elde ettiğimiz bulguları ve alana katkılarımız üzerine düşüncelerimizi sunmaktadır. Araştırmamızın sonuçlarına dayanarak, gelecekteki araştırmalar için temel çıkarımları, bulguları ve potansiyel araştırma yöntemlerini inceleyerek bu konudaki düşüncelerimizi paylaşıyoruz.

Bölüm 4'te Ani Duruş olaylarının tahmin edilmesinde birkaç önemli adım attık. İlk olarak, 59 ülkedeki Ani Duruş Krizlerini belirlemek için Forbes ve Warnock (2021) tarafından hazırlanan sermaye akışı veri setinden yararlanıyoruz. Bu kapsamlı veri seti, 1978'in 1. çeyreğinden 2020'nin 3. çeyreğine kadar olan dönemi kapsayan hem gelişmiş hem de gelişmekte olan ülkeleri kapsamaktadır. Bunu takiben, analizlerinde sunulan temel durum için tahmin sürecini tekrarladık. Bu, tamamlayıcı log-log modelinin kullanılmasını ve örnek dışı performansının incelenmesini ve bunu temel senaryomuz olarak oluşturmayı içeriyordu.

İkinci olarak, tahmin sorunumuzu, parametre tahmini veya nedensel çıkarım probleminden farklı olarak, örneklem dışı performansların karşılaştırılması olarak

açıkça çerçeveledik. Örnek bölme, çapraz doğrulama teknikleri, hiperparametre ayarlama ve özellik ölçeklendirme konularına kısaca değinerek, MÖ yöntemlerinin çeşitli temel sınıfları için uygulama ve tahmin stratejilerini vurgulamayı amaçladık.

Üçüncü olarak, Destek Vektör Makineleri (SVM), k-En Yakın Komşular (kNN), Çok Katmanlı Algılayıcı (MLP) ve Rastgele Orman gibi topluluk yöntemleri ile seçilen yöntemlerin kısa ve teknik olmayan özetlerini sunmuş bulunmaktayız. Ayrıca, XGBoost ve AdaBoost gibi topluluk yöntemlerine de değindik. Her bir yöntem, farklı zorluklara karşı benzersiz bir yaklaşım sunar ve belirli veri türlerine özel avantajlar sağlar. Bu yöntemlere dair teknik referanslar da sunulmuştur.

Model seçim stratejimizi açıkladıktan ve performans ölçümlerine referans verdikten sonra, ortalama AUC puanlarını değerlendirmeye ve ek bir performans ölçümü olarak geri çağırmayı (recall) incelemeye başladık. Ayrıca, yanlış alarm oranını sabitlerken geri çağırmayı (recall) karşılaştırdık.

Alıcı Çalışma Karakteristiği Eğrisi Altındaki Alan (AUC-ROC), özellikle Ani Duruş krizlerinin modellenmesi gibi ikili sınıflandırma görevlerinde önemli bir performans ölçüsü olarak hizmet eder. ROC eğrisi, çeşitli sınıflandırma eşiklerinde yanlış pozitif orana karşı gerçek pozitif oranı grafiksel olarak temsil eder. AUC-ROC, bu eğrinin altındaki alanı ölçerek bir sınıflandırma modelinin ayırt edici gücünü değerlendirir. 0 ila 1 arasında değişen daha yüksek AUC-ROC değerleri, üstün model performansını gösterir. ROC eğrisi, farklı eşik ayarlarında duyarlılık ve özgüllük arasındaki dengeyi göstererek model seçimine yardımcı olur. Daha iyi ayırt etme yeteneğine sahip bir model, grafiğin sol üst köşesine yaklaşan ve daha yüksek bir AUC-ROC değerine yol açan bir ROC eğrisi sergiler.

AUC-ROC puanları karşılaştırıldığında, k-En Yakın Komşular (kNN) ve Çok Katmanlı Algılayıcı (MLP) modelleri, temel modelle (cloglog modeli) kıyaslandığında daha düşük performans sergilemektedir. Diğer yandan, SVM ve Elastik Ağ modelleri benzer ortalama AUC puanlarına sahiptir. Ağaç tabanlı modeller (Random Forest, XGBoost ve AdaBoost), hem temel modelden (geleneksel

213

tamamlayıcı log-log) hem de diğer makine öğrenimi yöntemlerinden daha iyi performans göstermektedir. Bu durum, temel modelin pozitif ve negatif sınıflar arasında göreceli bir ayrım yeteneği sergilediğini ve bunun orta derecede iyi bir genel performansa neden olduğunu göstermektedir. Fakat, 0.16'da kaydedilen zayıf geri çağırma (recall) performansı, kriz zamanlarını kriz dışı olarak yanlış tanımlama olasılığının yüksek olduğunu göstermektedir. Bu durum, modelin gerçek kriz örneklerini tanımlamada başarısız olduğunu ve geri çağırma performansını olumsuz yönde etkilediğini göstermektedir. Geri çağırma açısından temel model en kötü performansı sergilemektedir.

Doğruluk, hassasiyet ve F1 gibi diğerlerinin yanı sıra karşılaştırma için ek bir ölçüm olarak geri çağırmanın seçilmesinin ardındaki mantık şöyledir: Birincisi, veri kümesindeki doğal dengesizlik nedeniyle, doğruluk (accuracy), Ani Duruş Krizleri gibi nadir olayları tahmin etmede modellerin performansını değerlendirmek için uygun bir ölçüm olmayabilir. Kriz tahminlerinde yaygın olduğu gibi, bir sınıfın diğerinden önemli ölçüde sayı olarak üstün olduğu durumlarda doğruluk yanıltıcı olabilir. Modeller, yalnızca çoğunluk sınıfını tahmin ederek, nadir görülen olayları yakalayamayarak yüksek doğruluk elde edebilir. Bu ölçüm, olay olmayanları doğru şekilde tahmin etmek ile olayları doğru şekilde tahmin etmek arasında ayrım yapmaz ve iyi performans gösteren bir modele ilişkin yanlış bir izlenim sağlar.

Benzer şekilde, Hassasiyet (Precision), nadir olaylar gibi dengesiz veri kümeleriyle uğraştığımız durumlarda güvenilir bir ölçüm olmayabilir, özellikle de Ani Duruş Krizlerini tahmin etme gibi nadir olaylar söz konusu olduğunda. Nadir olayların, olay olmayanlara göre daha az sıklıkta gerçekleştiği durumlarda (ki bu, kriz tahminleri için sıkça geçerlidir), Hassasiyet (Precision) ölçümü yanıltıcı olabilir.

Hassasiyet (Precision), gerçek pozitif tahminlerin, gerçek pozitiflerin ve yanlış pozitiflerin toplamına oranını gösterir. Nadir olayların çoğunlukta olduğu bağlamda, birkaç nadir olayın doğru bir şekilde tahmin edilmesi, ancak önemli sayıda olay dışı durumun yanlış olarak pozitif olarak sınıflandırılması yoluyla yüksek bir hassasiyet puanı elde edilebilir.

Bu sorun, Hassasiyet (Precision)` in gerçek negatifleri (doğru tahmin edilen olay olmayan durumları, diğer bir deyişle, kriz olmayan durumları) hesaba katmaması ve modelin performansına ilişkin yanlış iyimser bir görünüm sunabilmesi nedeniyle ortaya çıkar. Nadir olaylar söz konusu olduğunda hem negatif hem de pozitif örnekleri doğru şekilde tanımlama yeteneği de dahil olmak üzere genel performansı dikkate almak çok önemlidir.

Sınıflandırma görevlerinde geri çağırma (recall) ve hassasiyet (precision) arasındaki doğal dengeyi anlamak çok önemlidir. Bu denge, yanlış pozitifleri ve yanlış negatifleri en aza indirmenin dikkatlice değerlendirilmesini gerektirir. Geri çağırma (recall) ,modelin tüm gerçek pozitif örnekleri yakalama yeteneğini ölçer, bu nedenle yanlış negatiflerden kaçınmaya öncelik verir.

Öte yandan, hassasiyet (precision), yanlış pozitifleri en aza indirmeye ve pozitif tahminlerin doğruluğunu değerlendirmeye odaklanır. Bir ölçütün geliştirilmesi genellikle diğerinin maliyetine mal olduğundan, bu ödünleşme durumu ortaya çıkar. Daha yüksek bir sınıflandırma eşiği kesinliği artırabilir, ancak geri çağırma (recall) oranını azaltarak modeli daha seçici hale getirebilir. Tam tersi, daha düşük bir eşik, geri çağırma (recall) oranını artırabilir, ancak kesinliği düşürerek daha kapsayıcı bir modele yol açabilir.

Yüksek hassasiyet veya yüksek geri çağırma arasındaki seçimi yaparken, görevin belirli hedeflerine ve kısıtlamalarına odaklanmak önemlidir. Ani Duruş olaylarını tahmin etme bağlamında, bu tür olayları kaçırmanın yanlış pozitif alarmlardan potansiyel olarak daha zararlı olduğunu göz önünde bulundurarak hatırlama/geri çağırma (recall) ölçümüne öncelik veriyoruz.

Ayrıca, ROC eğrisi gerçek pozitif oranı (duyarlılık) ile yanlış pozitif oranı (1 - özgüllük) arasındaki dengeyi gösterdiğinden, bu iki parametre arasındaki doğal dengeyi anlamak önemlidir. Bu denge, sınıflandırma eşiğini ayarlamanın, modelin pozitif örnekleri doğru bir şekilde tanıma yeteneğini etkilerken aynı anda negatif örnekleri yanlış sınıflandırma durumundan kaynaklanır. Tipik bir eşik ayarı 0.5'tir,

bu da 0.5'ten büyük veya eşit olan tahmin olasılıklarına sahip örneklerin pozitif olarak sınıflandırılacağı anlamına gelir. Ancak, bu eşik tüm senaryolar için evrensel olarak optimal değildir. Sonuç olarak, eşik seçimi, modelin denge uzayındaki işletme noktasını doğrudan etkiler. Örneğin, belirli bir yanlış pozitif oranı, örneğin 0.2, belirlenmiş bir kriter doğrultusunda yanlış pozitif oranını kontrol ederken farklı modellerin geri çağırma performansını inceleme olanağı tanır.

Model seçimi sürecinin final aşamasında, ortalama yanlış pozitif oranını (false positive rate) 0.2 olarak belirliyoruz. Bu, normal olayları kriz olarak yanlış sınıflandırma oranını sadece %20'lik bir hoşgörü seviyesinde tutma anlamına gelir. Bu sayede model, ortalama olarak kriz dönemleri olmayan zamanları %20 oranında Ani Duruş krizi olarak yanlış tanımlayabilir. Daha sonra, ilgili ROC eğrilerindeki ortalama gerçek pozitif oranlarını (recall) detaylı bir şekilde inceledik.

Ortalama yanlış pozitif oranını (false positive rate) model değerlendirmesi sırasında %20 olarak belirleme kararı, uygulamanın pratik gereksinimlerine özgü bir dizi düşünceden kaynaklanmaktadır. Dengesiz veri bağlamında, özellikle Ani Duruş krizleri gibi nadir olaylarla uğraşırken, 0.5 varsayılan eşiği, nüanslı performans değerlendirmesi ihtiyacına yetersiz gelebilir. Yanlış pozitif oranını belirli bir değere (örneğin, %20) sabitleyerek, değerlendirme, yanlış pozitiflerin kontrolü önemli olan senaryolara özgü hale gelir. Bu yaklaşım, sınıflandırmalardaki sınırlı yanlış sınıflandırmaların pratik gereksinimlere uyum sağlamasının yanı sıra, bir modelin performansının gerçek dünya uygulamasını yansıtan koşullarda daha nüanslı bir şekilde incelenmesine olanak tanır.

Sonuçlar, önceki ortalama AUC puanları karşılaştırmasıyla uyumlu olup, ağaç tabanlı yöntemlerin diğer modellere ve temel modele göre üstün performans sergilediğini doğrulamaktadır.

XGBoost`u en üstün performans gösteren model olarak belirledikten sonra, Ani Duruş olaylarını tahmin etmede en etkili değişkenlerin neler olduğunu anlamaya yöneliyoruz. Metodoloji ve sonuçlardaki farklılıkları belirterek, hem Özellik Önem

metodu (feature importance method) hem de Shapley değerleri kullanıyoruz. Özellik Önemi (Feature Importance), her değişkenin modelin tahmin gücüne katkısını açıklar ve Gini belirsizlik metriğini kullanır. Bu metrik, Ani Duruş Krizlerinin nedenlerini çıkarmaz, ancak tahmin edicileri tanımlar. Elde edilen puanlar 0 ile 1 arasında değişir ve daha yüksek bir değer, tahminlere daha büyük bir etkinin olduğunu gösterir. Eğitilmiş XGBoost modelinin özellik önemi, Bulaşma, Küresel Büyüme ve Küresel Risk (VXO) değişkenlerini en etkili tahminciler olarak vurgular. Bu değişkenler, modelin tahmin yeteneklerini şekillendirmede önemli bir rol oynar ve Ani Duruş Krizlerine katkıda bulunan faktörleri anlama ve yorumlama konusunda değerli bilgiler sunar.

Diğer taraftan, işbirlikçi oyun teorisine dayanan Shapley Değerleri, değişken katkılarını anlamak için işbirlikçi bir yaklaşımı benimsemektedir. Bu değerler, açıklayıcı değişkenler arasındaki etkileşimleri ve doğrusal olmayan durumları titizlikle göz önünde bulundurarak, geleneksel özellik önemi (Feature Importance) ölçümlerine kıyasla daha ince bir bakış açısı sunar. Shapley değerleri, her bir değişkenin öngörücü katkısını tüm olası kombinasyonlara dağıtarak bireysel ve ortak etkileri kapsamlı bir şekilde değerlendirmemize olanak tanır. Özellik Önemi, etkili tahmin edicilere hızlı bir genel bakış sağlarken, Shapley değerleri model içindeki karmaşık ilişkileri ve bağımlılıkları yakalamada üstündür. Dahası, Shapley değerleri, doğası gereği, tüm değişkenler arasındaki katkıların toplamının modelin genel tahminiyle uyumlu olmasını sağlayarak içsel olarak daha tutarlı bir ölçüm sunar. Shapley değerleri, Reel GSYİH, Küresel Risk (VXO) ve Bulaşmanın en etkili tahminciler olduğunu göstermekte ve özellik öneminin aksine Reel GSYİH'ye daha yüksek bir ağırlık vermektedir.

Bölüm 5'te, Uluslararası Para Fonu'ndan (IMF) alınan üç aylık verileri kullanarak Makine Öğrenimi yöntemlerini değişken seçimi için kullanmaya odaklanıyoruz. Değişkenler, önceki analizden seçilen ülkeler arasında %10'dan az eksik veri içerenler tercih edilerek veri ulaşılabilirliğine dayanarak seçilmektedir.

Veri seti artırma süreci, IMF'nin çeşitli veri kümelerinden elde edilen bilgilerin birleştirilmesini içermektedir, bu süreç sonucunda 192 dışsal değişken elde

edilmektedir. Bu değişkenlerin 1 dönem gecikmeli halleri, yıllık değişimleri ve bir önceki çeyreğe göre değişim oranlarını da hesaplayarak, bu bilgileri dışsal değişken olarak kullanmaktayız. Bu, toplamda 768 dışsal değişkenle zenginleştirilmiş bir veri kümesine yol açmaktadır. Ayrıca, artırma süreci, önceki tahminlerde kullandığımız dışsal değişkenleri ve Forbes & Warnock (2021) tarafından belirlenen diğer model tabanlı değişkenleri içermektedir. Veri seti, 1994'ün 1. çeyreğinden 2018'in 4. çeyreğine kadar uzanmaktadır.

Norveç ve Hong Kong veri setindeki eksiklikler nedeniyle bu ülkeler örneklem dışı bırakılmıştır. Toplamda 4066 örnekten oluşan veri kümesinin 599'u Ani Duruş olaylarını içermektedir. Rastgele Orman, XGBoost, SVM, Elastik Ağ ve Lojistik Regresyon gibi çeşitli makine öğrenimi modelleri, modellere özgü değişken seçim yöntemleri kullanılarak devreye alınmıştır. Seçim sonrasında tahmin yöntemleri, daha önceki bölümde uygulanan metodolojiyle tutarlı bir şekilde sürdürülmüştür; Katmanlı K-katlı Çapraz Doğrulama ve hiperparametre ayarlaması için 'grid' taraması (grid search) uygulanmıştır. Bu modeller için, değişkenlerin mevcut değerleri hariç tutulmuş ve daha önce belirtildiği gibi, IMF'den elde edilen ilk 192 değişkenden 1 dönem gecikmeli, önceki dönemdeki değişim oranı ve değişkenlerin yıldan yıla yüzde değişimini içeren özel bir seçim yapılmıştır. Sonuç olarak, bu süreçten türetilen nihai değişken seti toplamda 576 değişkenden oluşmaktadır. Bu set, Forbes & Warnock (2021)'den seçilen 28 değişkenin 1 dönem gecikmeli versiyonları eklenerek tamamlanmış ve toplam 604 dışsal değişken setine ulaşılmıştır.

Modellerimiz için çeşitli değişken seçim teknikleri kullanmaktayız. Dışsal değişken seçimi açısından, bu 604 değişken içinden Rastgele Orman 131 özellik tanımlarken, XGBoost 100 özellik ve SVM 50 özellik seçmekte; bunların tümü RFECV (Recursive Feature Elimination with Cross-Validation) yöntemini kullanmıştır. Elastik Ağ, Lasso yöntemini kullanarak 60 özellik belirlerken, Bayes Bilgi Kriterleri (BIC) tarafından yönlendirilen Lojistik Regresyon, 16 özelliğe kadar daraltmıştır.

Uzun Kısa Süreli Bellek (LSTM) modelinin eklenmesi, bu bölümü diğerlerinden ayıran önemli bir özelliktir. Bu derin öğrenme tekniği, zamansal bağımlılıklara

odaklanarak geleneksel makine öğrenimi modellerden ayrılır. Sonraki bölümlerde, RFE-CV yöntemine kısa bir özet sunmakta ve her modelle ilgili detaylar vermekteyiz. LSTM modeli için, geçmiş 4 çeyrek ve mevcut çeyreğe ait verileri içeren toplam 225 değişken kullanılmıştır. LSTM modeli, çıkarım zamanında bir sonraki çeyreğin Ani Duruş olayını tahmin etmek için mevcut çeyrek verileriyle birlikte geçmiş 4 çeyreğin verilerini kullanmaktadır.

Genişletilmiş veri kümesinin analizinde, XGBoost, geleneksel makine öğrenimi yöntemleri arasında öne çıkan bir performans sergileyerek üstün örnek dışı performans elde etmiştir. Belirgin bir 0,2 yanlış pozitif oranı için 0,83'lük kayda değer bir ortalama AUC puanı ve 0,69'luk bir geri çağırma (recall) elde etmiştir. Bu sonuçlar, sınırlı sayıda açıklayıcı değişkenle çalışırken hem XGBoost'un hem de Rassal Orman'ın (Random Forest) en iyi performans gösterenler olarak belirlendiğini 4. bölümde belirttiğimiz önceki bulgularımızla uyumludur.

Buna karşılık, Destek Vektör Makineleri (SVM) 0,51 ortalama AUC değeri ile en düşük performansı sergilerken, Rassal Orman 0,79 ortalama AUC değeri ile ikinci en iyi performansı göstermektedir. Özellikle, hibrit bir yaklaşım içeren Lojistik Regresyon`un etkili olduğu kanıtlanmış ve ortalama 0,78 AUC ile üçüncü en iyi performansı sağlamıştır. Elastik Ağ ise 0,60 ortalama AUC değeri ile yetersiz kalmaktadır.

Bu sonuçlar, çok sayıda dışsal değişken olduğu durumlarda, doğrusal bir model olan geleneksel lojistik regresyonun, Random Forest ve XGBoost gibi güçlü ağaç tabanlı yöntemlerin dahil edilmesiyle geliştirilebileceğini göstermektedir. Bu doğrusal ve parametrik olmayan yöntemler yalnızca performansı artırmakla kalmamakta, aynı zamanda etkili özellik seçme yetenekleri de sağlamaktadırlar.

Zamana bağlı bağımlılıkların eklenmesiyle LSTM modeli, 0,91'lik ortalama AUC ile tüm geleneksel modelleri geride bırakarak liderliği ele geçirmektedir. Bu, zamansal kalıpları yakalama konusundaki yeterliliğini vurgulamaktadır. Ek olarak, LSTM modeli, belirli bir 0,2 yanlış pozitif oranı için 0,85'lik bir geri çağırma(recall) elde ederek etkinliğini vurgulamaktadır. Bu, LSTM'nin Ani Duruş olaylarının olasılığına

ilişkin incelikli bir anlayış sağlamadaki etkinliğini vurgulayarak, zaman serisi tahmininde ileri derin öğrenme tekniklerine olan ihtiyacı vurgulamaktadır.

Özetle, araştırmamız, Makine Öğrenimi (MÖ) tekniklerinin yenilikçi uygulaması yoluyla hem gelişmiş hem de gelişmekte olan ekonomilerdeki Ani Duruş Krizlerini tahmin etmeye çalışmaktadır. Makroekonomik tahminde makine öğreniminin etkinliği üzerine devam eden söylemin bir parçası olarak çalışmamız, çeşitli makine öğrenimi yöntemleri sunması ve bunların örneklem dışı tahmin güçlerinin titiz bir değerlendirmesini yapmasıyla öne çıkmaktadır. Özellikle Ani Duruş olaylarının analizinde gelişmiş derin öğrenmenin, özellikle Uzun Kısa Süreli Belleğin (LSTM) kullanılmasına öncülük ederek mevcut bilgi birikimine katkıda bulunuyoruz.

G. THESIS PERMISSION FORM / TEZ İZİN FORMU

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences ☐

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences ☒

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics ☐

Enformatik Enstitüsü / Graduate School of Informatics ☐

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences ☐

YAZARIN / AUTHOR

Soyadı / Surname   : CELIK
Adı / Name   : SONGUL SIVA
Bölümü / Department   : İktisat / Economics

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English): PREDICTING EXTERNAL MACROECONOMIC CRISES: MACHINE LEARNING PERSPECTIVE

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans** / Master ☐   **Doktora** / PhD ☒

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ☒

2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two years**. * ☐

3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. * ☐

* Enstitü Yönetim Kurulu kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir. / A copy of the decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.

Yazarın imzası / Signature ....   **Tarih** / Date ....