External Crisis Prediction Using Machine Learning: An Early Warning System for Argentina

Augusto Alejandro Comisso

UNIVERSIDAD NACIONAL DE SAN MARTIN

INSTITUTO DE ALTOS ESTUDIOS SOCIALES

MAESTRIA EN DESARROLLO ECONOMICO

2023

Abstract

External Early Warning Systems (EWSs) have traditionally been designed to warn policymakers of potential balance of payments crises. This study compares the performance of three different EWSs Machine Learning models using Classification and Regression Trees (CART), Bagging or Bootstrap Aggregating (BA) and Random Forest (RF). The three models have been trained with Argentinian data between 2003 and 2023 and have shown high out-of-sample performance. In addition, using 103 macroeconomic variables, the models highlight the relevance of certain variables in the fiscal, external, real and monetary sectors as potential predictors of future currency crises in Argentina.

Keywords: Early Warning System; balance of payment crises; machine learning; Classification and Regression Trees; Bootstrap Aggregating; Random Forest

JEL Classification: C14, C45, E00, E47, F31

Contents

1 Introduction

1.1 Motivation

Currency crises have significant economic, social, and political consequences. As Keynes (1919) stated for Europe in the early 20th century, the effects of a currency crisis are severe. Argentinians have experienced these consequences firsthand across different generations.

Is it possible to predict an external crisis? There is often a lack of consensus among economists regarding the likelihood of a future balance of payments crisis. However, an Early Warning System (EWS) can contribute to the analysis of external risks in emerging markets (IMF, 2021). These models predict the likelihood that a country will face a currency or balance of payments crisis. EWSs have outperformed alternative warning systems such as credit ratings and bond spreads (Berg and Pattillo, 1999).

The primary role of the EWS is to identify underlying vulnerabilities that increase the likelihood of a crisis. The ultimate goal is to inform decisions on corrective policies. Economists combine the results of EWSs assessments with their specific knowledge and judgement (IMF, 2012). The EWS is not a substitute for judgement. However, unlike individual judgement, EWS provides a quantifiable assessment of economic vulnerability (Bussiere, 2013).

Machine learning techniques offer new possibilities for risk assessment in macroeconomics. External crises are infrequent and involve non-linear relationships not usually seen in no-crisis periods. Among other machine learning techniques, Classification and Regression Trees (CART), Bagging (BA) and Random Forest (RF) models can capture these non-linear relationships better than traditional methods (IMF, 2021).

In 2021, the IMF updated its methodology for constructing EWSs. While in 2017 they relied on traditional models (Ahuja et al., 2017), in 2021 published the adoption of machine learning (ML) models, including tree-based ML models. IMF (2021) constructed external EWSs for a panel of 192 countries and identifies variables that are relevant to track. The question remains whether these models are useful for Argentina. Additionally, it is important to determine which of the three-based models has the best out-of-sample performance for Argentina. Lastly, it is worth considering whether the variables proposed by the IMF to track in emerging countries are relevant for Argentina.

This thesis aims to create and compare different tree-based machine-learning EWSs for predicting external crises in Argentina. The focus is on empirical work rather than theoretical discussions, as machine learning techniques are primarily used for prediction.

This is a novel empirical work. No previous research has been conducted on developing EWS using machine learning (ML) models specifically for the case of Argentina between 2003 and 2023. Despite Argentina's history of periodic balance of payments crises, and the new methodology proposed by the IMF (2021), Argentine policymakers currently lack modern models for external EWS.

On the other hand, Argentina offers an opportunity to evaluate EWSs. Accurate crisis forecasting often requires a large number of observations and positive cases. Argentina has experienced frequent currency crises, which provides a source of positive cases.

1.2 Thesis objective

This thesis focuses on developing and comparing different EWSs for forecasting external sector crises in Argentina. Machine learning techniques, including CART, BA, and RF, have been utilised exclusively for the purpose of out-of-sample prediction.

The research is based entirely on empirical evidence and does not discuss theory. It is important to note that machine learning models do not assume the ability to resolve theoretical disputes. The study does not claim to establish causality.

I have employed these techniques in this thesis for three main reasons. Firstly, they have a proven track record of being a useful methodology with good out-of-sample performance. Secondly, they add originality to the research. Thirdly, they provide a valuable tool for policymakers in Argentina.

The period under consideration is from 2003 to 2023. Although analyzing previous periods, such as the convertibility period, would be interesting, it would result in the loss of relevant explainable variables for more recent periods and the near future. Therefore, the selected period of study is restricted by data availability and aimed at providing a valuable tool for policymakers in Argentina.

To construct EWS, this thesis only considers tree-based models, which could be a potential methodological criticism. Exploring other machine learning techniques, such as boosting and neural networks could be useful to address the research questions of this thesis. Furthermore, implementing unsupervised machine learning models could provide a deeper analysis of crisis definition. Therefore, this thesis's possible future developments include using these models (Section 6.7). Nevertheless, it is worth noting that the tree-based methods achieved high out-of-sample performance while saving on design and computational costs compared to alternative methods.

1.3 Research questions, hypotheses and methodology

The research aimed to answer three questions. First, to investigate whether machine learning models based on trees could be used to create a EWS that could accurately predict external crises in Argentina. The hypothesis was that these methods could achieve high out-of-sample performance in predicting balance of payments crises.

The second question was which tree-based method would best predict Argentina's external crises using data from 2003 to 2023. The hypothesis was that the RF technique would outperform the BA and CART techniques.

The third question focused on identifying the main instrumental variables that policymakers should track in order to be aware of potential external crises in Argentina. The hypothesis was that policymakers should focus on international and political variables.

1.4 Thesis structure

The thesis consists of six chapters and a conclusion. Chapter 2 explains the definition of an external crisis as used in this research. In short, defining a label means creating a dummy variable. However, this is not an easy task. Determining the meaning of a crisis may require judgement, as there is no consensus on this issue. For this research, we have chosen the definition commonly used by international organisations as the base case, taking a pragmatic approach. However, we also assess alternative crisis scenarios and signalling periods.

Chapter 3 discusses various theoretical frameworks for explaining external crises. It is important to note that economic theory can aid in selecting 'features' for machine learning models. In machine learning, the independent variables are referred to as 'features'. For example, in the case of a doctor, the features may include various indicators in a blood test. For a policymaker, the features could be macroeconomic variables that may explain a crisis. The features of our empirical exercise are outlined in section 6.1.

Chapter 4 argues that research on developing EWSs for Argentina using machine learning tools is lacking. This chapter examines existing literature on external EWSs and emphasises the role of machine learning techniques in the current EWSs of international organisations. The chapter states that, according to the IMF(2021), CART, BA, and RF models are more effective than traditional methods in capturing non-linear relationships and consequently have better out-of-sample performance.

Chapter 5 of the thesis presents the methodology used in this research. It explains the process of training, selecting, and evaluating machine learning models to achieve high out-of-sample performance. The section also describes and compares CART, BA and RF algorithms. The final section also presents the evaluation metrics commonly used in machine learning.

Chapter 6 presents the database, the training-validation-test sets, and the empirical results. The available empirical evidence is divided into three parts, one for each research question. Finally, a reflection on the robustness of the results and potential future extensions is presented.

Section 7 summarises the findings of the empirical study. The study shows that CART, BA, and RF have strong out-of-sample performance and can forecast a balance of payments crisis up to 24 months in advance. This conclusion holds for alternative crisis definitions and signalling periods. The BA and RF models outperform the CART model, while RF outperforms BA in narrow signalling periods. Contrary to the 'one-size-fits-all' view of the IMF (2021), the models highlighted different variables from the fiscal, monetary, real, and external sectors as potential predictors of currency crises for different signalling horizons and crisis definitions for the case of Argentina. This thesis presents innovative work by using tree-based models to develop an EWS for Argentina in the 21st century. It also provides a valuable tool for policymakers.

2 Crisis definition: the labels

Chapter 2 of the research defines an external crisis. To label it, a dummy variable is required. However, determining what constitutes a crisis requires judgment, and there is no consensus on this issue. For practical purposes, this research uses the definition employed by international organisations as the 'base case'. Alternative crisis scenarios are also taken into account.

2.1 The lack of consensus on the definition and measure of external crises

A currency crisis is not clearly defined. It occurs when currency speculators attack, causing a sudden depreciation in the exchange rate. However, several ways exist to prevent or mitigate the effects of such a depreciation. For instance, a central bank can defend the exchange rate by using international reserves or by raising interest rates. Alternatively, a government can stabilise the exchange rate by introducing exchange controls or borrowing money from international organisations, sovereigns or private banks. The definition of a crisis is unclear, and there is no consensus on this matter.

Nominal exchange rate depreciation According to Alves Jr et al. (1999), a currency crisis arises when the government is either unwilling or unable to defend the currency's valuation. Berg et al. (2005) explains that this is also the ideal definition of a crisis for private banks only interested in successful currency attacks.

Decline in international reserves Conversely, Kaminsky et al. (1998) has argued that a currency crisis should be characterised by a significant depreciation in the official exchange rate as well as a significant reduction in the central bank's international reserves. The author argues that a currency crisis can result from both successful and unsuccessful attacks on a currency, and in some cases can be resolved by a combination of both.

According to Kaminsky et al. (1998), a crisis can be identified by using an exchange market pressure index. This index is calculated as a weighted average of the monthly percentage change of the nominal exchange rate and the monthly change in international reserves measured in United States dollar (USD). The weights are chosen to ensure equal variance of the two components in the pooled sample. A crisis is said to have occurred when the index exceeds three standard deviations from the mean. However, it is important to note that the criteria for defining a crisis may vary depending on the specific characteristics of each country.

Rising interest rates According to Girton and Roper (1977) and Eichengreen et al. (1995), raising interest rates may help to mitigate the effects of a currency attack. To address this issue, they proposed an index that considers the percentage changes in the exchange rate, foreign exchange reserves, and interest rates. The three components are weighted so that their conditional volatilities are equal. Alonso-Alvarez and Molina (2023) suggests that the short-term interest rate is the key variable in predicting banking, sovereign and currency crises. This is because monetary authorities in emerging markets use it to defend exchange rate pegs and avoid capital outflows. However, the lack of data on interest rates is the reason why they are often left out of crisis definitions, as Berg et al. (2005) notes.

Exchange rate controls Kaminsky et al. (1998) also stresses that the imposition of exchange controls, or the closure of the official foreign exchange market, can define a currency crisis. Therefore, to define a crisis, it is necessary to assess developments in the foreign exchange market and add them to the currency pressure index.

IMF financial support The IMF (2021) has put forward two different definitions of external crises. The first one is known as the 'sudden stops currency crisis', which is typical for emerging markets with open capital accounts. On the other hand, 'exchange rate pressure events' include episodes of sharp exchange rate depreciation or reserve depletion, even if the realised capital outflows are not large.

For 'exchange rate pressure events', the IMF (2021) has proposed an index defined as a weighted average of the annual percentage change in the nominal exchange rate and the annual percentage change in reserves as a percentage of GDP. The weights are chosen so that the variance of the two components is equal in the pooled sample. Crisis events occur when the index is in the lower 15th percentile of the full panel of countries [1].

The IMF (2021) also defines a crisis event as one in which exchange controls are imposed or if the country is approved for substantial IMF financial support. This definition aims to include scenarios where exchange rate controls or significant IMF support prevent large exchange rate depreciations or significant reserve losses. IMF support is considered substantial when the amount agreed upon is at least five times the country's quota in the IMF and is provided in response to a currency crisis.

Other sources of discrepancies Discrepancies may also arise from different indices. International reserve changes and nominal depreciation can be measured monthly or annually. Discrepancies may also occur due to variations in data frequency and weighting.

Additionally, defining a crisis as a discrete event can be problematic. When defined through an index, a crisis is actually a continuum. Therefore, a threshold is necessary to define a crisis accurately. A low threshold may classify mild events as crises, while a high threshold may miss actual crisis events. As a result, the definition of a crisis may vary depending on the chosen threshold.

Figure 1 provides a clear example of the impact of threshold selection. The adopted criteria [2] indicate a crisis in January 2006, during which there was a significant decline

[1] The cross-currency rate is defined as units of USD per unit of the local currency. A lower rate indicates a local currency depreciation against the USD.

[2] Figure 1 displays Argentina's crisis periods during the 21st century following the criteria outlined by Kaminsky et al. (2009). The index is calculated by taking a weighted average of the monthly percentage depreciation of the nominal exchange rate (Oficial ARS-USD exchange rate (A3500)) and the monthly

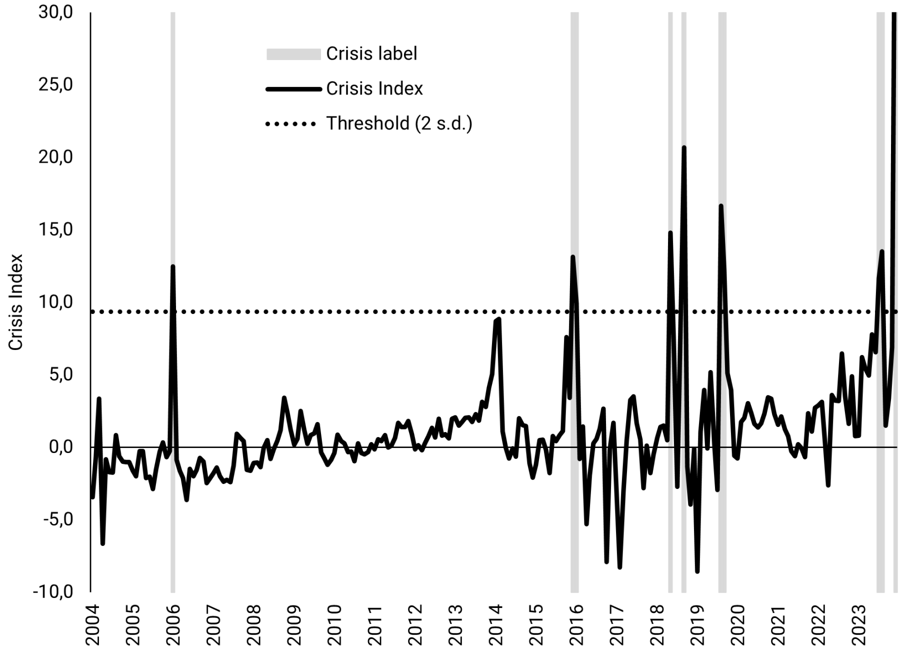


Figure 1: Crisis Index, threshold (2 s.d.) and crisis label for Argentina (2004-2023) based on Kaminsky et al. (2009). Source: own based on BCRA.

in international reserves due to Argentina's early debt payment to the IMF. A more restrictive threshold could be adopted to avoid including this payment in the calculations. However, this would come at the cost of ignoring the currency depreciation of 2014.

2.2 The definition of external crisis in this thesis

This section outlines the process of identifying a crisis episode for this thesis. It is crucial to acknowledge that no universal formula can accurately identify a crisis in all countries and situations. To identify a crisis, it is crucial to merge the outcomes of a quantitative definition with country-specific knowledge of exchange market developments, as outlined in Berg et al. (2005). Ultimately, the analyst's expertise and judgment will determine whether a currency crisis has occurred. For practical reasons, the definition of this thesis's 'base case' is similar to that of the IMF (2021).

Figure 2 illustrates the crisis periods that Argentina has experienced in the 21st

decline in international reserves measured in USD. The weights are based on the variance of the two components. A crisis is said to have occurred when the index exceeds two standard deviations (s.d.) above the mean. Exchange rate controls and IMF support are excluded from the calculations.

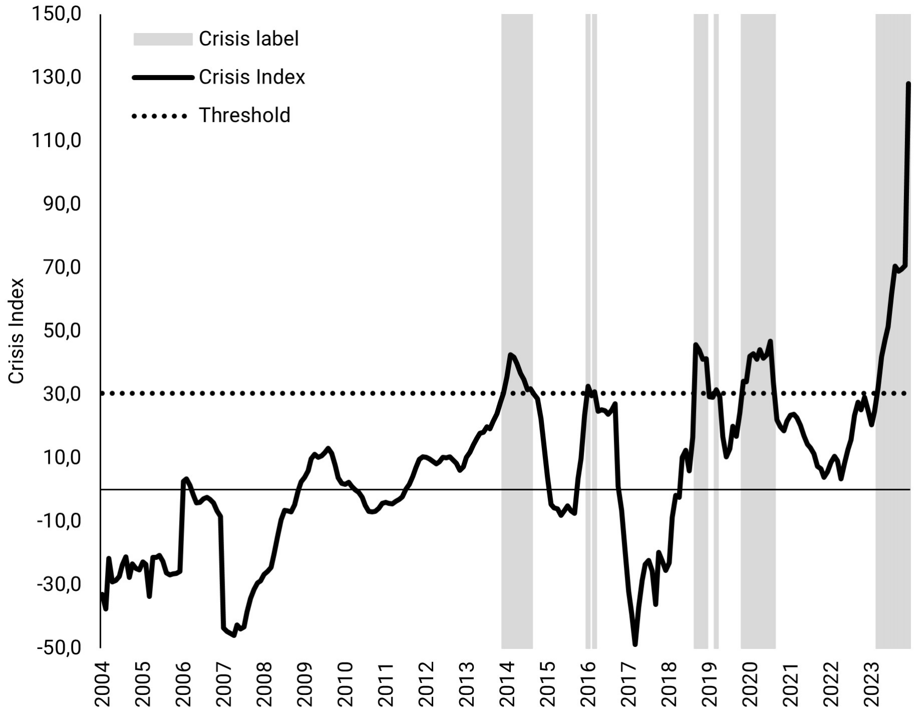


Figure 2: Crisis Index, threshold (upper 85th percentile) and crisis label for Argentina (2004-2023) based on IMF (2021). Source: own based on BCRA.

century based on the criteria provided by the IMF (2021). The index is a weighted average of the year-on-year percentage variation of the nominal exchange rate A3500 and the year-on-year variation in international reserves measured in USD. The weights are chosen based on the variance of the two components. If the index is in the upper 85th percentile, it is considered a crisis [3].

Figure 3 adds to Figure 2 the periods that include both exchange rate controls and IMF support with a 12-month signalling horizon. For this analysis, we define exchange rate controls as being in place when the implicit exchange rate in the American Depositary Receipt is 20 per cent or more above the official exchange rate. In addition, this figure includes a Stand-By Arrangement (SBA) of USD 50 billion that was signed between Argentina and the IMF in June 2018. It is worth noting that this amount represents 1'110 per cent of Argentina's quota in the IMF.

The thesis proposes a 'base case' definition of crisis similar to that proposed by IMF

[3]The cross-currency rate is defined as the number of Argentine pesos (ARS) per unit of USD. A higher rate indicates a depreciation of the Argentinian currency against the USD. Afterwards, the index subtracts the variation in international reserves.

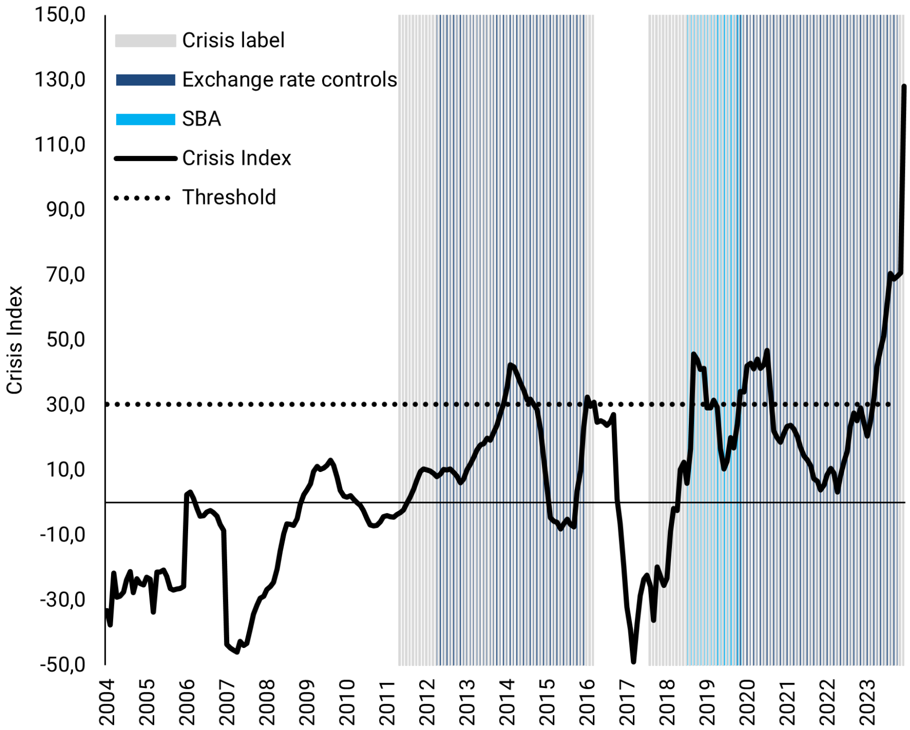


Figure 3: Crisis Index and crisis labels for Argentina (2004-2023). Based on IMF (2021) including SBA and exchange rate controls (12-month signalling period). Source: own based on BCRA.

(2021). According to this criteria, a crisis is identified when the following conditions are met: i) the weighted average of the year-on-year percentage variation of the nominal exchange rate A3500 and the year-on-year variation in international reserves measured in USD is in the upper 85th percentile; ii) exchange controls are in place, meaning that the implicit exchange rate in the American Depositary Receipt is 20 per cent or more above the official exchange rate; and iii) the country is receiving IMF support at an agreed level of at least five times its quota.

This thesis has also adopted as the 'base case' the 24-month signalling horizon proposed by Kaminsky et al. (2009) and IMF (2013). According to this approach, if a signal is followed by a crisis within 24 months, it is considered a reliable prediction. The main objective of this method is to give policymakers enough time to make informed economic policy decisions and to prevent future external crises.

However, the models will be evaluated using different labels. We assessed alternative crisis scenarios due to the controversial nature of any crisis definition. Additionally, we tested tree-based machine learning models in different crisis scenarios as a robustness exercise. We started with the IMF's 'base case' broad definition of a crisis and a lengthy signalling period. As we progressed, we gradually narrowed down the definition of a crisis and the signalling period. The following alternative scenarios were considered:

(A) The crisis definition included exchange rate controls, SBA, and a signalling period of 24 months. A total of 103 variables were considered;

(B) Same as (A), but only 88 variables were considered, excluding nominal variables;

(C) Same as (B), but without considering exchange rate controls and SBA as indicators of a crisis;

(D) Same as (C), but with a signalling period of 12 months;

(E) Same as (D), but with a signalling period of one month.

3 Theoretical frameworks: background of the features

The feature selection of the EWS machine learning models is based on theoretical frameworks from balance of payments crisis theories. In other words, the theory serves as a conceptual provider of features in this thesis. It is important to emphasise that machine learning models are unable to resolve theoretical disputes.

This section discusses various theoretical frameworks for explaining external crises. The literature has been classified according to the criteria presented by Alves Jr et al. (1999). 'First-generation models' focus on poor domestic fundamentals resulting from inconsistencies between fiscal and exchange rate policies. 'Second-generation models' emphasise policy makers decision and competitive devaluations. 'Third-generation models' focus on the impact of international financial markets on currency markets. As an antecedent, the structuralist tradition is resumed.

3.1 Structuralist models

In the traditional structuralist balance of payments crisis models, the main reason for a currency crisis is the lack of foreign exchange to finance industrial accumulation. This means that the traditional export sector, such as agriculture, cannot generate enough foreign exchange to meet the demand of the inward-looking industrial sector. This has been discussed in various studies, such as those conducted by Alejandro (1963); Braun and Joy (1968); Olivera (1977); Krugman and Taylor (1978). [4]

Classical Argentine works have analysed the country's external crises and emphasised the structural characteristics of its economy. Studies by Ferrer (1963); Villanueva (1964); Sidrauski (1968); Braun and Joy (1968); Diamand (1972); Canitrot (1983) have shown that the combination of inward-looking industry and insufficient agricultural exports leads to successive external crises and a stop-go dynamic in economic activity. While Braun and Joy (1968) stresses the insufficient agricultural exports as the ultimate cause of external crises, Diamand (1972) points out that the high productivity of pampas agriculture causes the exchange rate to appreciate, making it impossible for the industry to compete in exports. The result is an unbalanced production structure, which explains the external crises.

Other authors in this Marxian tradition agree with the concept of structural external restrictions, but emphasise the class struggle over the distribution of the surplus (Arceo, 2008; González, 2011; Basualdo, 2020). [5]According to a Marxian strand of literature, Argentina's external crises are caused by class conflict. Some works within this tradition emphasise the impact of class struggle on relative prices, as in the works of O'donnell

[4]During the mature phase of the Latin American import substitution model, there were several balance of payments crises. The dependency literature, rooted in Structuralism and Marxism, emphasised Latin America's cultural and technological dependence on developed countries. The imported cultural patterns of consumption and the dependence on external technology and machinery had a serious impact on the balance of payments (Furtado, 2020; Pinto Santa Cruz, 1953; Cardoso and Faletto, 2000; Tavares and Serra, 1972; Baran, 1957; Frank, 1967).

[5]According to Basualdo (2020), the ultimate reason for the external constraints on economic growth can be traced back to the behaviour of the capitalist class. This class captures the surplus and leads to successive processes of external public indebtedness. It then drains the surplus abroad instead of investing it in productive ventures to displace the external constraint, such as energy production. Furthermore, in analysing the impact of internal absorption on the external sector, researchers in this tradition have highlighted the different impacts of capitalist consumption and the consumption of higher income earners on imports of goods and services, e.g. in tourism. This has been discussed in studies by Cortés and Marshall (1986, 2003) and in a more recent study by González and Fernández (2021).

(1977); Lindenboim et al. (2007); Graña et al. (2009); Fiorito (2015) [6].

Recent works in the Structuralist-Keynesian tradition have added the role of the financial account of the balance of payments in exchange rate crises [7]. For example, according to Cimoli et al. (2016), the canonical sequence commences with high international liquidity and low interest rates in central countries. This, in turn, attracts a surge of foreign capital inflows, including both FDI and short-term inflows, into the Latin American economy. Consequently, the currency appreciates, leading to higher real wages and a short-lived consumption boom. Simultaneously, external debt increases. However, foreign capital becomes increasingly reluctant to lend to highly indebted economies due to concerns about the sustainability of the external sector. After, negative news triggers a significant depreciation of the currency.

3.2 First-generation models

In the first-generation of models (Krugman, 1979; Flood and Garber, 1984), the collapse of an exchange rate regime is attributed to unsustainable fiscal policies. In a typical balance of payments crisis, a fiscal deficit is financed by money creation, leading to a persistent loss of international reserves and eventually to currency depreciation [8]. The core message of the IMF's surveillance of Latin American emerging markets is still based on this view. [9]

[6]For example, in O'donnell (1977), the oligopolistic bourgeoisie, together with the landowning class, successively introduced a policy of devaluation, thus reducing the wages of the working class. Fiorito (2015) adopted the Kaleckian idea of the political cycle to explain the relative price fluctuations in Argentina. Currency devaluations are, therefore, a manifestation of ongoing class conflict.

[7]Several authors analyse the impact of international finance on the external sector of the periphery (Thirlwall and Hussain, 1982; da Conceição Tavares, 2000; Vernengo, 2006; Moreno-Brid, 1998; Caldentey and Vernengo, 2013; Cimoli et al., 2010, 2016; Bortz, 2018), even resorting to Minskyan categories (Médici, 2017; Dvoskin and Feldman, 2018; Zeolla and Médici, 2022) and the Kaleckian framework (Bortz, 2018; Bortz et al., 2020).

[8]As a precursor to Krugman (1979); Flood and Garber (1984), the 'Monetary Approach to the Balance of Payments' stressed that, in a small country with a fixed exchange rate regime, if the supply of money exceeds the demand for money, there will be a proportional loss of international reserves, which can eventually lead to a currency crisis. This has been discussed in studies by Johnson (1972, 1977); Olivera (1977).

[9]A branch of literature argues that external crises in Latin America are the result of macroeconomic populism. As immediate precursors, we identify the works of Alejandro (1963); Canitrot (1975); Diaz-Alejandro (1979); Llach and Sánchez (1984); Sachs (1989); Pereira (1991); Dornbusch and Edwards (1991), which mainly refer to the import substitution phase. For the period of economic and financial globalisation, we can cite the works of Dornbusch (1996); Edwards (2019); Acemoglu et al. (2013); Rodrik (2018); Gerchunoff and Rapetti (2016); Gerchunoff et al. (2020). According to this literature,

In this type of crisis, financial markets play a secondary role. When reserves fall to critical levels due to fiscal deficits, the market runs against the currency, forcing it to depreciate. However, the main cause of currency crises has been the lack of sound fundamentals (Krugman et al., 1999).

3.3 Second-generation models

In 1992, the United Kingdom, Italy and Spain devalued their currencies and left the exchange rate mechanism of the European Monetary System. In 1993, the French franc also devalued. However, the devaluations of the European countries in 1992 and 1993 did not fit the first-generation models. They had access to international markets and could avoid devaluation by selling reserves and raising interest rates. Nevertheless, their central banks chose not to do so because they wanted a competitive exchange rate to promote economic recovery and employment growth (Krugman et al., 1999).

In the first-generation of models, the role of the central bank is passive. For example, the central bank does not use monetary policy to prevent the currency from collapsing. Nor does the central bank decide to devalue the currency for competitive reasons. This is the main flaw in the first-generation of models (Alves Jr et al., 1999).

According to the model proposed by Obstfeld (1996), central bank behaviour is less mechanistic than in first-generation models. While the central bank may aim to prevent currency depreciation due to its impact on trade, investment, and inflation, policymakers may also aim to avoid the negative effects of higher interest rates and wages on growth and employment. A central bank may choose to devalue its currency when the real exchange rate has significantly appreciated, nominal wages are inflexible in nominal terms, and there are real costs to activity and employment. This devaluation can occur even without a speculative attack on the currency. On the other hand, the cost of defending the parity increases as agents anticipate future depreciation and run against the currency.

external crises are caused by excessive internal absorption, mainly due to excessive domestic private consumption and fiscal deficits, leading to a currency crisis. Ultimately, policies are determined by the cultural patterns of the Argentine working class and its unions (Gerchunoff and Rapetti, 2016).

3.4 Third-generation models

The message of the first-generation of models was also inconsistent with the Asian crises. Fiscal positions in all Asian countries were strong, exchange rates were not overvalued, and inflation had been moderated for several years. The Asian crises then taught that, in the absence of a change in fundamentals, a sudden deterioration in private agents' expectations can lead to a currency run and hence exchange rate collapse (Ghosh and Ghosh, 2003).

According to the argument presented in Krugman (1998), the Asian crisis cannot be explained by first or second-generation models. In other words, the Asian devaluations were not the result of fiscal deficits or competitive devaluations. The root cause of the crisis was a financial crisis caused by capital inflows and the practice of local banks making bad loans in the knowledge that they would be bailed out. A moral hazard problem led to a financial crisis, which led to a currency crisis. So the currency crisis was not the real problem, but a symptom of the underlying problem.

For Krugman et al. (1999) there is a range of fundamentals where a currency crisis cannot happen and a range of fundamentals where a currency crisis will happen. In between, there is a range of fundamentals where a crisis could happen but not necessarily will [10]. In this intermediate scenario, the author considers four main reasons. First, 'self-fulfilling' crises due to investor pessimism, where any event could trigger it. The second is 'herding', where investors sell the currency as its price falls (as in Calvo and Mendoza (1996)). Third is 'contagion', where a currency crisis in one country causes a crisis in another related country (e.g., the Asian and Tequila crises). Fourth is 'market manipulation', where a large investor can force a currency crisis (e.g. Soros and the pound).

Other models emphasise the impact of poorly regulated banking systems on balance of payments crises. Diaz-Alejandro (1985), Velasco (1987), and Calvo (1998b) argue that central bank bailouts of financial institutions through money printing could trigger

[10]Third-generation models explained the balance of payments crises by emphasising self-fulfilling events. For example, Eichengreen et al. (1995) argues that currency crises are self-fulfilling and that currency attacks can occur even when the fiscal stance is consistent with exchange rate policy. Third-generation models, however, require weak fundamentals for it to occur (Alves Jr et al., 1999).

currency crises.

Empirical studies have also highlighted the importance of the Federal Reserve's monetary policy. The authors were inspired by the successive Latin American crises in the era of globalisation. They argued that domestic macroeconomic imbalances were associated with the Federal Reserve's accommodative monetary policy and the large capital inflows of the previous years. The sudden stop of capital inflows and the resulting depreciation of domestic currencies were caused by the subsequent increase in US dollar interest rates by the FED (Calvo et al., 1993; Calvo, 1998a; Dornbusch et al., 1995; Calvo, 1998b; Ferretti and Razin, 2000; Mendoza, 2002; Calvo et al., 2004, 2006; Adalet and Eichengreen, 2007). In other words, the currency crisis was viable in the context of balanced budgets and real exchange rates and was due to external shocks and financial factors. The authors were inspired by the successive balance of payments crises in the era of financial globalisation [11]. Contagion effects are also part of the history. Private agents may not have distinguished between the fundamentals of different countries, leading them to run against all emerging currencies after a particular currency crisis (Calvo and Reinhart, 1996; Masson, 1999; Claessens and Forbes, 2004).

4 External EWSs: literature review

Section 4.3 highlights the lack of previous research on external EWSs for Argentina using machine learning tools. Sections 4.1 and 4.2 divide the existing literature on external EWS into pre-ML and post-ML literature. It also highlights the importance of machine learning techniques in the current EWSs of international organisations.

4.1 Pre machine learning EWSs literature

The literature on EWSs dates back to the 1970s, following a series of currency crises in the post-Bretton Woods era. Researchers sought to develop leading indicators that could predict these crises. In the early days of risk analysis, ratios were assessed manually

[11] While some argue that the latter is due to the lack of capital controls (Rey, 2015), others argue for dollarisation (Reinhart and Calvo, 2000), and others for floating exchange rate regimes to better absorb shocks (Edwards, 2004).

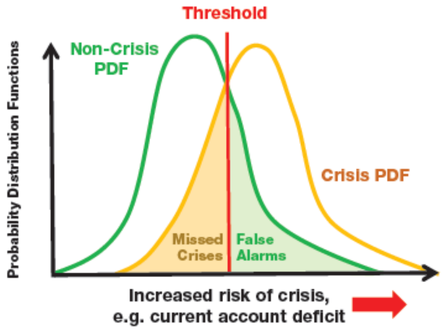


Figure 4: Example of signalling approach: crisis and non-crisis Probability Distribution Functions (PDF). Source: IMF (2013).

rather than using sophisticated statistical methods on computers (Holopainen and Sarlin, 2017).

Later, the risk analysis literature was influenced by the Latin American debt crisis of the 1980s. Using an empirical model-based approach, the early EWSs focused on the sudden stops in capital inflows to emerging markets. This narrow focus was a major drawback of the earlier risk analysis literature (Währungsfonds, 2010).

During the 1990s, the literature on EWSs gained momentum due to several emerging market balance of payments crises, such as the Mexican, Asian, Russian and Argentinean crises. Institutions such as the Federal Reserve System (FED) and the Bank for International Settlements (BIS) developed EWSs. The International Monetary Fund (IMF) also implemented a version of the Kaminsky et al. (1998) model. In addition, private sector firms such as Lehman Brothers, Citicorp, JP Morgan, Goldman Sachs and Credit Suisse First Boston developed their own EWS models to inform foreign exchange trading strategies (Berg et al., 2005). These private EWS models had excellent in-sample performance but poor out-of-sample performance, according to IMF (2013).

In the early 2000s, various models were developed to address the increasing volatility of emerging markets. Deutsche Bank and Morgan Stanley created their models for private use, the IMF implemented various versions of traditional models on the public side (Berg et al., 2005). These models varied depending on the risk assessment for an emerging, low-income, or advanced economy, as well as the sector being assessed, such as the external, fiscal, financial, or real sector (Ahuja et al., 2017).

Since then, two methodological approaches have become standard. The first is the 'signalling approach' (Figure 4), used in studies such as Kaminsky et al. (1998); Berg and Patillo (1998); Kaminsky (1999); Herrera and Garcia (1999). The second is the 'categorical dependent variable regression' as used in Eichengreen et al. (1995); Frankel and Rose (1996); Berg and Patillo (1998). Both approaches have been central to IMF surveillance for more than a decade.

However, these systems have been criticised for generating false alarms and missing real crises out-of-sample (IMF, 2013). For example, the IMF's probit model signalled a crisis in Argentina in March 2001, after the crisis had already begun. Moreover, the IMF's signal model almost failed to detect the 2001 crisis in Argentina (IMF, 2013). More recently, Boonman et al. (2017) presented a EWS multinomial ordered logit model to predict currency crises for Argentina and Brazil from 1990-2007. The authors also found that the model performed better for Brazil than for Argentina.

As a result, recent literature has criticised the traditional approaches for their limitations. According to Borio and Lowe (2002), the signalling approach may underestimate the probability of a crisis when the relevant factors are close but below their thresholds. The IMF (2021) also notes that discrete choice regression struggles with a large number of regressors and has a significant gap between in-sample and out-of-sample performance. Kaminsky (2006) argues that these methods take a one-size-fits-all approach that may not work effectively in all scenarios. In addition, Kaminsky (2006) and IMF (2021) found that neither the standard regression framework nor the signal approach can effectively capture non-linearities, which become more likely as the number of fragilities increases.

4.2 Machine learning EWS literature

Artificial intelligence has introduced new models for risk assessment. Machine learning techniques such as CART, BA, RF, boosting techniques and neural networks can provide valuable support in complex and dynamic non-linear environments. According to IMF (2021), these techniques are considered to be the next generation of risk assessment models. [12]

[12] The global financial crisis gave impetus to EWSs application to financial crises in developed countries. Manasse and Roubini (2009) pioneered the use of CART to predict financial crises. Fouliard et al. (2021)

There was an early use of Neural Networks in the field of external EWSs. Nag and Mitra (1999) studied currency crises in Indonesia, Malaysia, and Thailand from 1980 to 1998. They concluded that their neural network model outperformed the signal approach model out of the sample. [13]

Alternatively, Ghosh and Ghosh (2003) introduced CART to explain currency crises in 42 countries from 1987 to 1999. According to the authors, CART can capture the non-linear relationship between macroeconomic imbalances and structural vulnerabilities. They conclude that weak institutions and weak fundamentals can be combined to predict a currency crisis.

Contemporaneously, Kaminsky (2006) classified currency crises into six varieties using CART. The study looked at 96 crises in 20 countries from 1970 to 2001. The author found four types of currency crises associated with domestic economic fragility: current account deterioration, fiscal imbalances, financial excesses or unsustainable external debt. A fifth type of crisis, known as the sudden stop phenomenon, was also identified. In addition, the study found that currency crises can occur even in economies with sound fundamentals. The author concludes that the second generation of EWS should include methods such as regression tree analysis (Figure 5). [14]

The IMF has recently incorporated machine learning models to conduct vulnerability exercises (IMF, 2021). They monitored 79 variables for 192 countries from 1990 to 2017. After analysing emerging market currency crises, they conclude that a boosting model has better out-of-sample performance [15]. However, the (IMF, 2021) has used a RF model

developed a ML model to predict out-of-sample financial crises. Similarly, the IMF (2021) developed EWSs to forecast financial crises in advanced economies.

[13] In a more recent study, Lin et al. (2008) developed a fuzzy logic neural network model to predict currency crises in 20 countries from January 1970 to June 1998. The results showed that their model outperformed the traditional logit model. In another study, Sevim et al. (2014) developed models using artificial neural networks and decision trees that covered the period 1992-2011 of the Turkish economy. They used thirty-two macroeconomic indicators as independent variables. The models could predict the Turkish crises of 1994 and 2001 twelve months in advance. However, there is a common criticism of artificial neural network models: the risk of over-fitting, the computational and design costs, and the fact that they are black boxes.

[14] More recently, Joy et al. (2017) used CART and RF to predict banking and currency crises in 36 developed economies from 1970 to 2010. The study found that high domestic short-term interest rates and overvalued exchange rates are the most significant indicators of currency crises in the short run. The research also showed that domestic factors have a greater impact on currency crises than international ones.

[15] The signal extraction method showed better out-of-sample performance in the case of sudden stops, but no single model performed better in the case of exchange market pressure events. In the latter case,

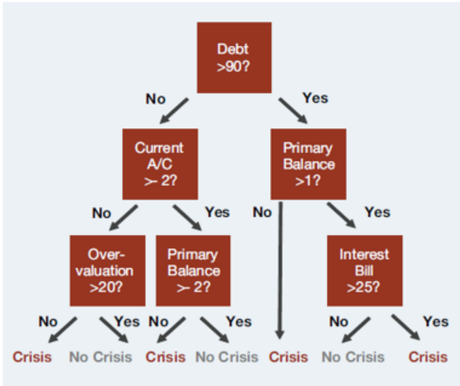


Figure 5: Example of CART. Source: IMF (2021).

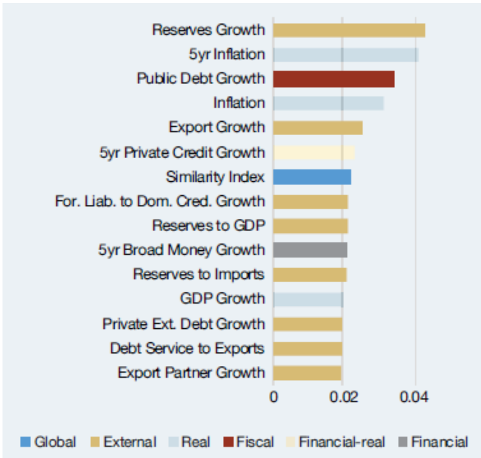


Figure 6: Variable importance of the IMF's RF model for emerging countries. Source: IMF (2021).

because it is more transparent than the boosting 'black box' model. They have also found that, in most cases, RF models are the most effective in assessing the near-term risk of a crisis in the fiscal, real and financial sectors.

The IMF (2021) has identified several variables that need monitoring based on the crisis type and country. The importance of international variables and domestic asset bubbles was highlighted for 'sudden-stop crises' in all countries between 1990 and 2017. In the context of 'exchange market pressure events' in emerging countries, the first five key variables were: reserve growth, five-year inflation, public debt growth, inflation, and export growth (Figure 6).

Nevertheless, the IMF (2021) has recognised that the causes of external crises differ

the signal extraction method performed better for advanced economies, while RF was more effective for low-income economies, and boosting techniques had the best performance for emerging markets (IMF, 2021).

from country to country. Abiad (2003) also stressed that different indicators are relevant for different countries. Therefore, models that attempt to generalise may not perform well when applied to a particular country. Moreover, not all crises are the same (Kaminsky, 2006; Joy et al., 2017; IMF, 2021). Therefore, the development of customised EWS tailored to a specific country may perform better than a multi-country model.

4.3 Machine learning EWS literature on Argentina

Pre-ML models have been applied to the Argentine case with low out-of-sample performance. Various international organisations, private banks and researchers have developed EWS models to predict external crises and applied them to Argentina. However, these models have shown poor out-of-sample performance. [16]

This thesis presents pioneering work in the field by developing an EWS for Argentina's external sector between 2003 and 2023 using CART, BA, and RF models. It is worth noting that while two previous studies have examined Argentina's case using ML techniques [17], no studies have been conducted to develop an external EWS for Argentina using these models.

[16] Alvarez-Plata and Schrooten (2004) and IMF (2013) have shown that the forecasting quality of the Kaminsky et al. (1998) approach could have been better in the case of Argentina. Marongiu et al. (2005) found that the leading indicator approach applied to the Argentine currency crisis failed due to the strong influence of the period considered on the threshold. IMF (2013) also admit that their probit model failed to predict the Argentine crisis of 2001 in a timely manner. Boonman et al. (2017) acknowledge that their logit model performed better for Brazil than for Argentina.

[17] Frankel and Wei (2004) analysed currency crises in 19 countries, including Argentina, using CART. According to the study, a currency crisis will likely occur when short-term debt-to-reserves exceeds 157 per cent and inflation exceeds 17.2 per cent. It is worth noting that the tree was evaluated in-sample, and the authors did not validate the hyper-parameters. Chamon et al. (2007) also used CART to predict capital account crises in 49 countries from 1995 to 2005. The model was able to predict the 2001 crisis in Argentina out-of-sample by identifying relevant thresholds of a current account deficit of -2.9 of GDP and a ratio of external debt to international reserves above 41 per cent. It is important to note that this is a multi-country study that focuses on sudden-stop currency crises.

5 Supervised Machine learning: tree-based methods, model selection and assessment

5.1 Model selection and assessment

The chosen model should have the minimum error when making predictions on unseen data. But how can we train and select this model to achieve this? If we select our model using the same data that was used to train it, there is no guarantee that the model will accurately predict future unseen data. This is because a model that predicts seen crises does not necessarily predict future unseen crises.

This section explains how a model is trained and selected to achieve high out-of-sample performance. First, it explains the problem of the trade-off between variance and bias and how this affects the error that the model wants to avoid. In other words, it discusses the impact of an over-fitted and under-fitted model on the out-of-sample error. Secondly, it emphasises the importance of splitting the data into three sets: training, validation and test. Finally, it discusses the potential problems with this strategy and the solution provided by techniques such as the validation set approach, k-fold cross-validation and bootstrapping.

5.1.1 Bias-variance trade-off

Prediction error can be divided into three parts: irreducible error, bias and variance (Tibshirani et al., 2009). The irreducible error is the noise that cannot be reduced by any model. Bias is the difference between the average prediction after the model has been trained over several independent datasets and the true value [18]. As with bias, the variance also supposes that the model has been trained over several independent datasets. Intuitively, the variance is the amount by which the model's predictions would change if

[18]
$$\text{Bias}^2(\hat{f}(x)) = \left(E[\hat{f}(x)] - f(x)\right)^2$$

it were trained on a different sample from the same population [19]. Fortmann-Roe (2012) provides intuition about the concepts using a bull's-eye diagram (Figure 7). The author also shows that the expected difference between the true population value and the value predicted by the model is equal to the expression:

$$\text{Expected Prediction Error} = \text{Irreducible Error} + \text{Bias}^2 + \text{Variance}$$

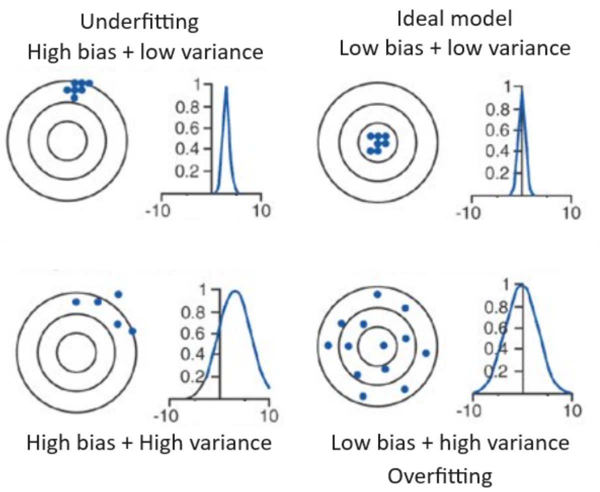


Figure 7: Bias and variance graphically explained. Source: own based on Fortmann-Roe (2012) and der Aalst (2017).

There is a trade-off between bias and variance [20]. The problem of dealing with the bias-variance trade-off is the problem of dealing with under-fitting and over-fitting (Fortmann-Roe, 2012). The more complex the model, the lower the bias, but the higher the variance. The less complex the model, the lower the variance, but the higher the bias. Figure 8 gives a graphical intuition of the over-fitting and under-fitting problems.

An over-fitted model remembers noise instead of learning the underlying trends in the data set, resulting in poor out-of-sample performance (Figure 9). In other words, the model may perform well on the data it was trained on, but not on new, unseen data. Therefore, while it may have successfully predicted past crises, it cannot be relied upon to

[19]
$$\text{Variance}(\hat{f}(x)) = E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right]$$

[20] In real practice, we have finite data and an imperfect model, which means that reducing both variance and bias to zero, leaving only the irreducible error, is improbable. Instead, there is a trade-off between bias and variance (Fortmann-Roe, 2012).

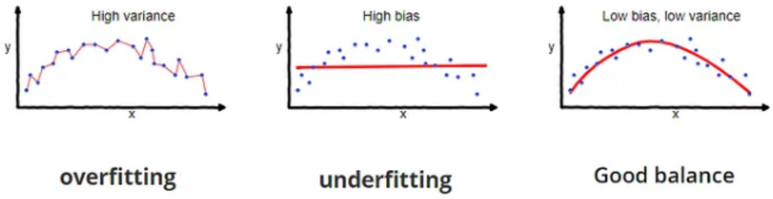


Figure 8: Over-fitting and under-fitting graphically explained. Source: ITBodhi (2020).

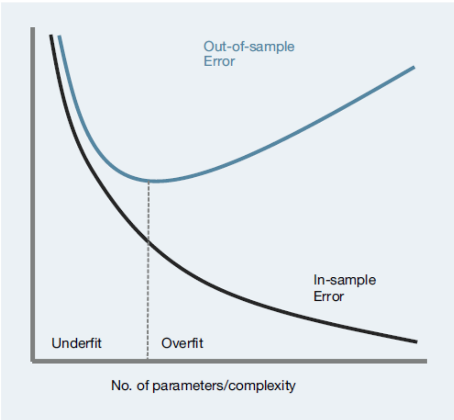


Figure 9: Example of under-fitting and over-fitting problem. Source: IMF (2021).

predict future crises. On the other hand, an under-fitted model has to be more complex to capture the relationships accurately (IMF, 2021).

5.1.2 The Validation Set Approach

Consequently, the goal of minimising forecasting error is to balance over-fitting and under-fitting (Fortmann-Roe, 2012). A model must be built neither too simple nor too complex to predict a future crisis. The ideal model should balance bias and variance to achieve minimum error [21]. Additionally, there could be an irreducible error that we have to accept. [22]

[21] Economists tend to prefer a model without bias at the expense of assuming a higher error due to a higher variance. Indeed, this kind of model performs well in the theoretical long term. However, statisticians have taught economists that they only have one or a few realisations of the dataset. Then, in practice, the long run does not exist. The error could be reduced by accepting bias in exchange for less variance.

[22] Academic articles stress that the best predictive model is the one that has asymptotic consistency and asymptotic efficiency. A model is asymptotically consistent if its bias tends to zero as the sample size tends to infinity. And a model is asymptotically efficient if it has the lowest variance of all possible models. In practice, samples do not tend to infinity. And in a small sample, a model that is asymptotically consistent and efficient could perform worse than a model that is not asymptotically consistent and efficient (Fortmann-Roe, 2012).

Figure 10: Example of train, validation and test sets. Source: Tibshirani et al. (2009).

In a data-rich environment, the best practice for obtaining an equilibrated model that achieves minimum out-of-sample error is randomly splitting the data into three independent sets (Tibshirani et al., 2009):

1. A 'training set' to train the model to learn potential underlying relationships;
2. A 'validation set' to evaluate performance and select between different model types and hyper-parameter choices;
3. A 'test set' to evaluate the model's performance on unseen data, also known as out-of-sample performance.

Each set has a different role. The 'training set' is the segment of the original set used to train the model to learn potential underlying relationships. The training set should be unbiased and representative of the population. The 'validation set' is used to evaluate performance and select between different types of models. It is an independent and unbiased set of data to predict which model will have the best out-of-sample performance. Finally, the 'test set' is used to evaluate the model's performance on unseen data, also known as out-of-sample performance. It's important to note that the 'test set' should not be used to select the model, as this can lead to over-fitting. Instead, the 'validation set' should be used to identify the best model, and the 'test set' should be used as a final evaluation to ensure that the model performs well on new, unseen data. Figure 10 illustrates this splitting strategy (Tibshirani et al., 2009).

Implementing the validation-set approach helps avoid over-fitting and under-fitting. This differs from evaluating models solely on their goodness of fit within the sample, which can lead to over-fitting and harm future predictions. In this way, models are selected and evaluated based on their predictive performance outside the sample (James et al., 2013; IMF, 2021).

However, using the splitting strategy to train a model has two problems. Firstly, the validation set is based on only about a quarter of the total observations, which reduces the number of observations available for training. This can negatively impact the accuracy of the model [23]. Second, it is important to ensure that the datasets used for training and validation are representative of the population. Would the chosen model change significantly if we had chosen a different split between training and validation? The error in the model can vary greatly depending on which observations are included in the validation test and which are included in the training test (Wilber and Croome, 2012).

5.1.3 k-Fold Cross-Validation and the Bootstrap

Alternative methods could be used to address the problems of the validation-set approach. On the one hand, there exist analytical methods, e.g. the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Alternatively, resampling methods like cross-validation and bootstrap are used to estimate extra sample prediction error and select models.

In practice, two of the most commonly used statistical resampling methods are cross-validation and the bootstrap (James et al., 2013; Kohavi et al., 1995). For nonlinear algorithms like trees, the estimation of the effective number of parameters is very difficult. Consequently, methods like AIC and BIC are impractical and leave us with cross-validation or bootstrap as the methods of choice (Tibshirani et al., 2009).

The first method, k-fold Cross-Validation (CV), is the simplest and most widely used (Tibshirani et al., 2009). The data not used for testing are randomly divided into k equal-sized folds. For each estimation, $k-1$ folds are used as the training set, and one fold is reserved as the validation set. This process is repeated k times, with each fold serving as a validation set. Finally, the average of the scores of the validation sets is used to select the best model. Thereafter, we can eliminate the need to decide which split to use by using all the data for training and validation. Figure 11 shows an example

[23]To prevent loss of observations, an alternative approach is to retrain the model using all the data from the original training and validation sets, except for the test set, once it has been selected. This ensures that the test set is the only data that is not used to train the model. However, this approach still results in the loss of approximately 25 per cent of the observations. This is the trade-off for not having to wait for new data for out-of-sample evaluation.

of a 5-fold cross-validation. In practice, it is common to use k=5 or k=10, as stated in Tibshirani et al. (2009).

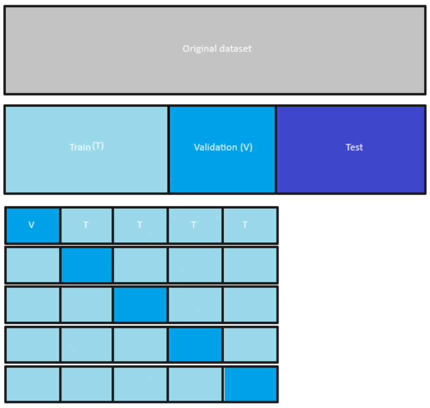


Figure 11: Example of 5-fold cross-validation. Source: own based on Tibshirani et al. (2009).

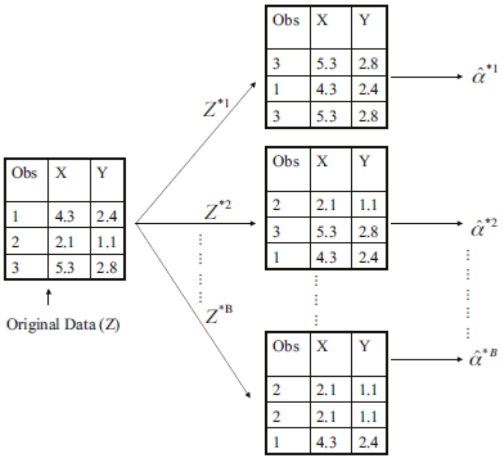


Figure 12: Example of the bootstrap approach on a toy sample containing three observations. Source: Tibshirani et al. (2009).

The bootstrap technique is a useful alternative to cross-validation for training and selecting the model (Tibshirani et al., 2009; Fortmann-Roe, 2012; IMF, 2021). The aim is to train the model by randomly sampling with replacement from the non-test set. Each random sample from the original set of n elements should also have n elements. It is important to note that this sampling is done with replacement. Validation is performed using the non-selected, i.e. Out-of-bag (OOB) observations for each sample. In practice, the model is trained on approximately two-thirds of the observations and tested on the remaining third. This process is repeated several times. The OOB error represents the model's error in the training validation session. A graphical example of the bootstrap approach on a toy sample is provided in Figure 12. [24]

5.2 Methods

This section describes traditional methods used in external EWS and compares three tree-based machine-learning methods: CART, BA and RF. CART is a supervised machine learning technique first introduced by Breiman (1984). While CART is a transparent model ('crystal box') and has a low bias, it suffers from high variance. BA is a supervised machine learning technique originally proposed by Efron (1979) and Tibshirani and Efron (1993). The BA algorithm's objective is to enhance CART's performance by reducing its variance while maintaining its bias. However, the primary drawback of BA is the low variance reduction due to the high similarity between trees. To address the issue of tree correlation, Breiman (2001) proposed the RF algorithm, which is a variant of the BA algorithm.

5.2.1 Traditional methods

EWS have long been used for country surveillance in international organisations. The traditional EWS literature has relied on two approaches: logit or probit regressions and the signal extraction approach (IMF, 2021).

Logit and probit models. Logistic and probit regressions (LR) are used to predict the class of the dependent variable (i.e. crisis or no-crisis) based on a linear predictor (z) transformed by a sigmoid or normal function, respectively. The parameters β are estimated using either maximum likelihood estimation or a gradient descent algorithm (James et al., 2013; Gelman et al., 2020). For instance, the equation below is utilised

[24] Resampling approaches can be computationally expensive. This is because they involve training the same algorithm multiple times on different subsets of the training data. Today, however, the computational requirements of resampling methods are generally affordable (James et al., 2013). Nevertheless, the validation-set approach is still widely used when resource constraints do not allow alternatives that require resampling, i.e. cross-validation and bootstrap (Wilber and Croome, 2012).

to estimate the likelihood of y being part of a specific category (1 or 0; i.e. crisis or no-crisis) based on inputs X. This approach is widely accepted and easy to understand (IMF, 2013).

$$P(y=1|X) = \frac{1}{1+e^{-z}}, \quad \text{where} \quad z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

However, logistic regression has three disadvantages. Firstly, logit and probit models may lose observations in the presence of missing data (Alessi et al., 2015) [25]. Secondly, this approach struggles to handle a large number of predictors [26]. Thirdly, the main issue with the LR approach is the significant difference in performance between in-sample and out-of-sample data (IMF, 2021). For instance, IMF (2013) found that its probit model failed to predict the 2001 Argentine crisis in a timely manner.

Signal extraction approach. Alternatively, the Signal Extraction method (SE) is a non-parametric method that compares the value of a variable in a given period to a threshold defined as a percentile of its distribution. This technique assigns thresholds on an indicator-by-indicator basis to maximise accuracy, i.e. to minimise the sum of false positives and false negatives. The country receives a score of one if the indicator exceeds the threshold, or zero otherwise. Figure 13 shows the aggregation of results from different indicators. The weights could be calculated based on formal criteria or determined through judgement. This method is widely accepted, including by international organisations. Also, it is a method that facilitates interpretation and communication to policymakers (IMF, 2013).

[25] Refer to Section 5.4 for imputation of missing values

[26] To address this issue, dimension reduction techniques such as principal component analysis and variable selection techniques such as ridge and lasso can be employed. However, using principal component analysis may complicate interpretation, and ridge and lasso could leave us with variables with no theoretical relevance.

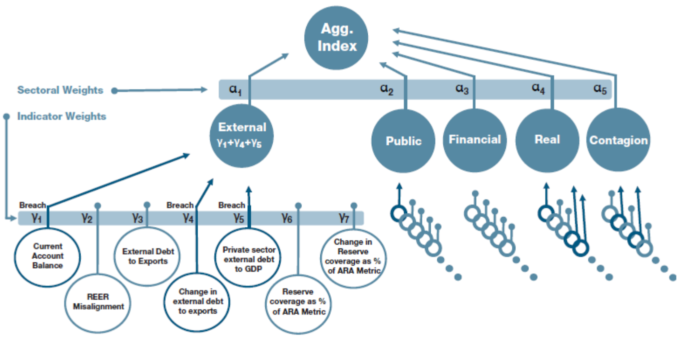


Figure 13: Example of aggregation in the SE approach. Source: IMF (2013).

However, the SE technique may not perform well when dealing with non-linear interactions between multiple predictors (IMF, 2021). For example, Figure 14 displays the non-linear relation between the output (EMAE) gap estimated as a 12-month mobile average (v78), the International Reserve year-over-year growth (v86), and the U.S. Dollar Index (v3). Signal extraction models cannot capture such non-linear interactions.

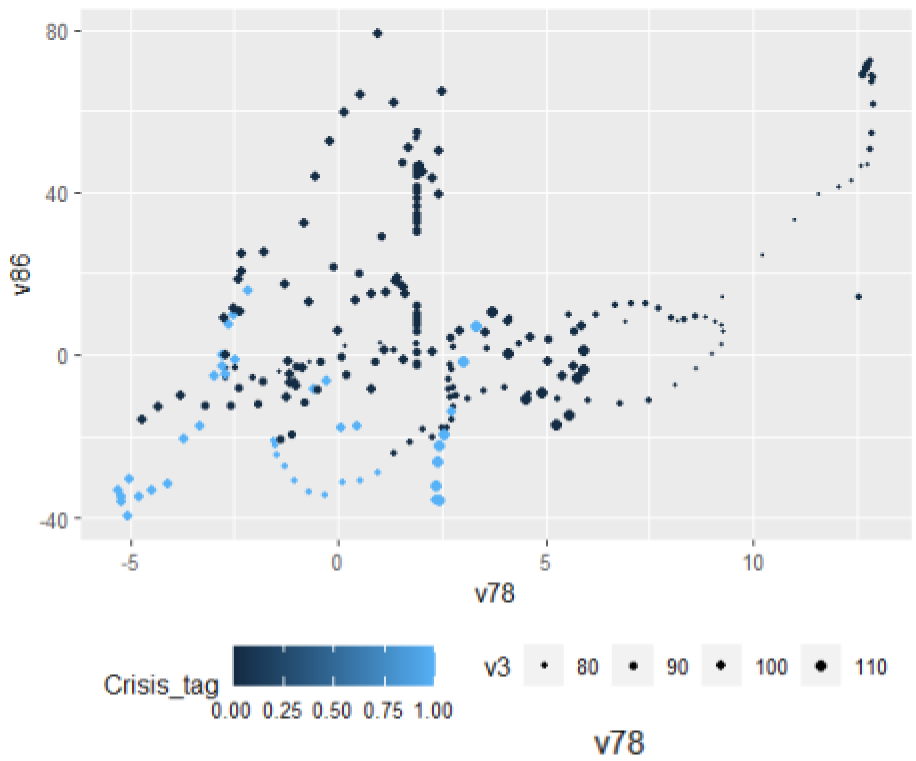


Figure 14: Example of non-linear relationships: Monthly output gap estimated as a 12-month mobile average (v78), International Reserve year-over-year growth (v86), DXY (v3), and crisis labels (1-month signal horizon), Argentina 2003-2023.

5.2.2 Classification and Regression Trees (CART)

CART is a supervised machine learning technique first introduced by Breiman (1984). CART produces a recursive, one-dimensional sample partition into disjoint subsets of crisis and no-crisis subsets, as shown in Figure 5.

The algorithm works by choosing, at each step, the variable that best discriminates between observations that have experienced a crisis and those that have not. The chosen variable minimises a loss function that decreases as the sub-nodes become more homogeneous or pure. This indicates that the observations assigned to each subnode have similar labels. The loss function reaches its minimum when all labels in each sub-node are the same (James et al., 2013).

Different criteria could be used to measure the purity of a node. Firstly, the classification error rate [27], i.e. the fraction of the training observations (p) that do not belong to the majority class. Secondly, the Gini Index [28] takes a small value if p is close to zero or is close to one. Thirdly, the Entropy [29] will also take a value near zero if p is either near zero or one (Tibshirani et al., 2009).

Consequently, the CART algorithm first calculates the Gini for each feature in each node. It then identifies the feature that gives the lowest Gini. A decision node is then created using this feature. When no more splits are possible, a final node ('leaf node') is created and labelled with the most frequent class (Tibshirani et al., 2009).

There are various reasons why the splitting process may stop when creating a decision tree. For example, no further splitting would be required if a node is pure. In addition, stopping conditions could be imposed to avoid over-fitting. This is when the tree becomes

---

27
$$\text{Error} = 1 - max(\hat{p})$$

28
$$\text{Gini} = \sum_{k=1}^{K} \hat{p}(1-\hat{p})$$

29
$$\text{Entropy} = -\sum_{k=1}^{K} \hat{p}log(\hat{p})$$

too deep and learns too much from the noise of the training examples, resulting in less generalisable rules (refer to Section 5.1.1).

To tackle this problem, one proposed solution is to prune the tree by imposing a penalty for an overly complex structure. The ideal level of complexity is chosen using cross-validation techniques. However, even after pruning, the problem of over-fitting may still persist (Tibshirani et al., 2009).

The objective is to select a sub-tree with the lowest test error rate by pruning the tree to the right extent. Estimating the cross-validation error for every possible sub-tree would be too costly. So, the cost complexity pruning formula is used [30]. T indicates the number of terminal nodes of the tree. When α is zero, there is no pruning, having the 'maximal tree'. However, as α increases, there is a penalty to pay for having a tree with many terminal nodes. The tuning parameter α controls a trade-off between the sub-tree's complexity and its fit to the training data. We can select a value of α using a validation set or using cross-validation (Tibshirani et al., 2009).

Thereafter, the CART algorithm has several steps. First, choose a criteria to measure the purity of the node and use recursive binary splitting to grow a large tree on the training data. An additional criterion of stopping only when each terminal node has fewer than some minimum number of observations could be used. The second step is to prune the tree. To do that, it is necessary to cross-validate the hyper-parameter α. The third step, once we have the pruned tree, is to assess the tree prediction out-of-sample (Tibshirani et al., 2009). [31]

CART has several advantages. First, it makes no assumptions about the distribution of the independent variables, allowing it to capture complex non-linear relationships between them. Furthermore, as a non-parametric method, it can deal with data with non-parametric distributions without suffering from specification problems. Finally, it

[30]

$$\mathrm{CC} = \sum_{i=1}^{|T|} Gini + \alpha|T|$$

[31] To prevent over-fitting, we use a cross-validation strategy as described in Section 5.1.3. The training sample creates the classification tree, while the validation sample selects a tree. Once a tree is trained and selected, we can predict out-of-sample whether an observation belongs to a 'crisis' or 'no-crisis' category by passing it through the tree until it reaches a leaf node (IMF, 2021; Tibshirani et al., 2009).

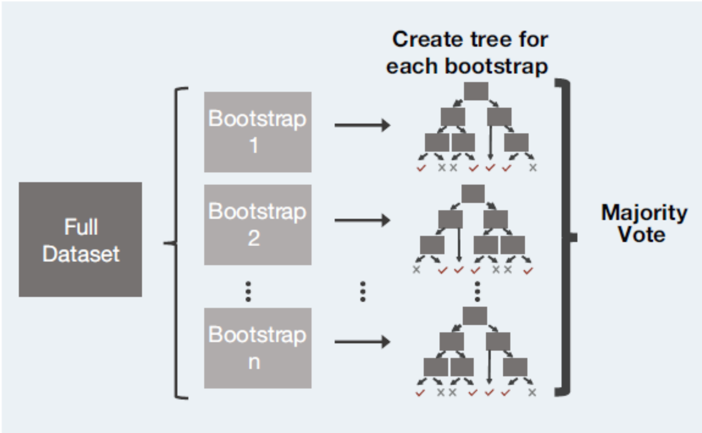


Figure 15: Bootstrap Aggregating (BA). Source: IMF (2021).

is an algorithm that can be quickly understood by a non-technical user (IMF, 2021; Tibshirani et al., 2009).

However, while CART has low bias, it suffers from high variance (Tibshirani et al., 2009). Different partitioning of the set into training and testing will result in varying trees and, consequently, different out-of-sample predictions. In other words, decision trees may have low out-of-sample predictive performance due to high variance (James et al., 2013).

5.2.3 Bootstrap Aggregating (BA)

BA is a supervised machine learning technique originally proposed by Efron (1979) and Tibshirani and Efron (1993). BA algorithm aims to improve the performance of CART by reducing its variance while preserving its bias (Tibshirani et al., 2009; Fortmann-Roe, 2012).

The main idea is to grow many unpruned trees and ensemble [32] their results. Each tree is constructed from a random sample with replacement ('bootstrap sample') from the original dataset. By the majority vote of many weak, unpruned trees, the algorithm makes an aggregate prediction. Figure 15 illustrates this procedure.

The algorithm uses the law of large numbers to average out the noise and preserve the signal. Outside of statistics, this idea is known as the 'wisdom of crowds' [33]. A diverse

[32] In machine learning, the aggregation of a prediction of multiple models to obtain an aggregate prediction is called 'ensemble learning'. BA, and RF are examples of ensemble learning.

[33] Francis Galton became aware that under the right circumstances -i.e. independence and diversity of

and independent group's collective knowledge usually exceeds any individual's knowledge. In statistics, this could be achieved by voting (Tibshirani et al., 2009). Consequently, bootstrap aggregation averages predictions trees constructed over a collection of bootstrap samples, thereby reducing its variance (Tibshirani et al., 2009; James et al., 2013).

The main disadvantage of BA is low variance reduction because the trees are very similar. Intuitively, if all trees were equal, then ensembling multiple trees is the same as having only one tree. The 'wisdom of crowds' assumes that the voters are independent and diverse, but bagged trees are not. In other words, the trees are highly correlated, and therefore, the variance is inflated by their covariance, which affects out-of-sample performance (Tibshirani et al., 2009).

5.2.4 Random Forest (RF)

To mitigate the tree correlation problem, Breiman (2001) proposed the RF algorithm. This algorithm retains the proposal of averaging the predictions of several trees generated by bootstrapping samples from the original sample but aims to reduce the correlation between them (Tibshirani et al., 2009).

To reduce tree correlation, it intervenes in their construction. Each partition does not consider all the variables but randomly selects a small subset to choose the optimal partition. In this way, RF guarantees that the grown trees have less correlation (Tibshirani et al., 2009). This hyper-parameter of the RF model should be selected by cross-validation. Alternatively, a heuristic rule is to use a number of variables equal to the square root of the total number of variables.

By combining many weak trees with high variance but low correlation, the algorithm reduces the overall variance and provides an aggregate prediction with better out-of-sample performance (Tibshirani et al., 2009; James et al., 2013). It is worth noting that the bias of the aggregated model is equivalent to the bias of a single tree, but the whole model has a much lower variance due to the aggregation of uncorrelated trees (Fortmann-Roe, 2012).

individuals- the median prediction of a crowd is usually better than the prediction of the smarter single person in it (Surowiecki, 2005).

RF is one of the most widely used ML algorithms (IMF, 2021). RF allows for non-linear interactions among the many explanatory variables, can handle large datasets, is computationally cheap, and does not require parametric assumptions. Finally, and most importantly, RF gives models with low bias and low variance, making good out-of-sample predictions.

5.3 Evaluation metrics

This section explains the metrics commonly used to measure model performance, including the Accuracy Rate (AR), Recall Rate (RR), Precision Rate (PR), and F1 score. Additionally, the concept of the receiver operating characteristic (ROC) curve and the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) will be discussed (Quinlan, 1986; Ward-Powers, 2011; IMF, 2021).

5.3.1 The confusion matrix

For each outcome, there are four possible responses from the model. First, the model could correctly predict a crisis, i.e. a True Positive (TP) case. Second, it could correctly predict a no-crisis, i.e. a True Negative (TN) case. Third, the model could incorrectly predict a crisis, i.e. a False Positive (FP) case. Finally, the model could fail to predict a crisis, i.e. False Negative (FN) case. A confusion matrix provides a visual representation of these four alternatives and the performance of the model (Table 1).

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	TN	FP
	Crisis	FN	TP

Table 1: Confusion matrix.

5.3.2 Accuracy, sensibility, and precision rates and F1 score

The most cited performance metric in the ML literature is the AR. The AR is the sum of forecasted TP plus TN cases divided by total cases. Hence, AR is a measure of overall performance. The higher the AR, the better. However, this performance measure is not useful in the case of unbalanced sets. For instance, if we have a crisis of 1 year in a set of 100 years, and the model fails to predict the crisis but correctly predicts the 99 non-crises, the AR is equal to 0,99.

$$\text{Accuracy (AR)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

An alternative measure is the RR, also known as the sensibility rate. The RR is the ratio of correctly predicted positive crises to all cases that are, in fact, positive. Hence, RR is important when FN cases are more critical than FP cases. For instance, we could prefer that our model does not miss to forecast the one crisis in 100 years. If the model fails to do it, then the RR will be zero; otherwise, it will be 1. However, this measure could also be misleading. Imagine that the model predicts 100 crises in 100 years. Then its RR will equal one but erroneously predict 99 out of 100 cases.

$$\text{Recall (RR)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Consequently, an additional performance rate is the PR. The PR is the ratio of correctly predicted positive classes to all cases predicted to be positive. Hence, PR is essential when it is crucial to avoid false diagnoses. Even so, this ratio could also be misleading. Imagine the country suffers ten crises in 100 years, and the model predicts only one crisis. Then, there will not be FP cases, and the PR will equal one.

$$\text{Precision (PR)} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

The F1 score summarises both precision and recall. It takes values between zero and one, with values closer to one indicating better performance. If either PR or RR are low, the F1 score will also be low. Only when both PR and RR tend to be one does the F1 score also tend to be one.

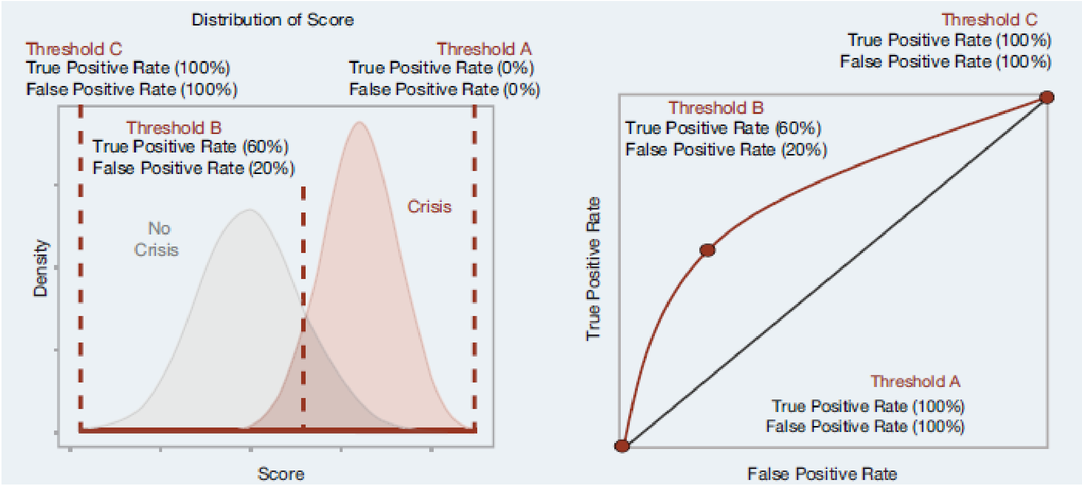


Figure 16: Example ROC curve and AUC. Source: IMF (2021).

$$F1 \text{ score} = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Ideally, the model would have precision and recall rates and accuracy equal to one. However, a trade-off between RR and PR often exists in practice. A greater recall usually means a lower precision. For instance, if a model predicts that every year will be a crisis, it will have a perfect recall but poor precision. Conversely, if a model never predicts a crisis, it will have great precision but poor recall. According to the problem, one performance measure or the other will be more relevant. For instance, we might prefer a recall over precision in the case of external crisis prediction. However, both measures help to understand the performance of the model.

5.3.3 ROC curve and AUC

When a threshold to classify between crisis and non-crises is selected, a point in this trade-off is also chosen. If the threshold is closer to zero, most cases will be classified as a crisis, and then we will gain in recall at the expense of precision. Conversely, if the threshold is near 1, then we rarely classify a case as a crisis, and hence, we will gain precision at the cost of recall. The preference for one or another type of error will guide the selection of the threshold.

The receiver operating characteristic (ROC) curve visually explains the RR and PR trade-offs (Figure 16). In essence, there exists a trade-off between true positive cases and false positive cases. The ROC curve shows how performance changes as a function of the model's classification threshold. The perfect model will always have a TP rate equal to one regardless of the value of the FP rate. So, to decrease the FP rate, it will not be necessary to sacrifice the TP rate. The closer the 'ROC curve of the real model' is to the 'ROC curve of the perfect classifier', the better (IMF, 2021).

The area under the ROC curve (AUC) also gives a sense of the model's proficiency. A perfect model has an AUC equal to 1. In this sense, the AUC provides a single measure of model performance. However, this metric treats all classification errors equally. For example, in crisis detection, a policymaker may want to maximise true positives even at the cost of increasing false positives (IMF, 2021).

5.4 Imputation of missing values

According to IMF (2021), the researcher has four options when there is missing data: i) drop all observations with missing variables [34]; ii) drop all variables with missing observations; iii) simple imputation, i.e. replaces a variable's missing observations with the variable's mean or median for numeric variables, or the mode for qualitative variables; iv) model-based imputation.

Model-based imputation has advantages and disadvantages. On the one hand, it is a sophisticated way of imputing missing values using the whole dataset. These models can be simple autoregressive models or more sophisticated models. For example, for imputing missing values, Stekhoven and Bühlmann (2012) proposed a nonparametric method that can handle both continuous and categorical variables simultaneously [35]. Machine learning models could also be used to impute values. The main disadvantage of model-based imputation is that it is more difficult to communicate the results of the model (IMF, 2021).

In practice, IMF models use a combination of the second and third approaches. Then, variables are dropped if more than a third of their observations are missing. The remain-

[34] This option usually reduces the data set dramatically.

[35] For implementation in R, see Wright and Ziegler (2015).

ing missing observations are imputed using the median of the variable, which is robust to outliers (IMF, 2021).

This work follows the approach of (IMF, 2021). Variables are dropped if more than a third of their observations are missing. The remaining missing observations are imputed using the median of the variable. In our database, there were only a few missing values for the remaining variables. Therefore, the imputation approach adopted did not impact the final results.

6 External Machine learning EWS for Argentina

6.1 Database

The theories of balance of payments crises (Section 3) provide the inspiration for the selected predictive variables in EWS. Given the objective of forecasting and providing policy messages, these variables should be theoretically relevant. We have also selected variables that are traditionally used in EWSs of international organisations [36]. However,

[36] According to IMF (2013), the probit model used in the IMF includes the variables: i) real exchange rate overvaluation, i.e. v80; ii) current account, i.e. proxy of v83; iii) foreign exchange rate loses, i.e. v86; iv) export growth, i.e. v87; v) ratio of short term debt to foreign exchange reserves, i.e. proxied by v96. The signal model of the IMF uses the first four variables of its probit model plus: vi) foreign exchange reserves to M2 and the growth of that ratio, i.e. proxy of v61; vii) domestic credit growth, i.e. v62 to v65; viii) change in the money multiplier, i.e. proxied by v55 and v68; ix) real interest rate, i.e. proxied by v48; and x) M1 excess over money demand, i.e. proxied by v93 and v95. For a panel of countries, as labels for ML models, IMF (2021) considered the variables (our variable proxies are between brackets): Fiscal balance as a percentage of GDP (v40), 5-year change in M2/GDP (v68), reserves/M2 (v61), reserves/GDP (v94), dummies for hard peg and float and for parallel market (v95), change in unemployment rate (v79), real GDP growth (v77), share of non-investment grade debt (v46 to v51), current account balance/GDP (v88), amortisation (v35), FX share of public debt (v35), debt service/exports (v96), FX share of household and non-financial corporate credit (v93), political violence, successful coup and political variables (v100 to v103), private credit/GDP (v64 and v65), EMBI spread level and growth (v46 to v51), corporate sector return (v72), probability of default (v46 to v51), interest coverage ratio, price-earnings ratio, bank returns (v72), share of non-performing loans (v73), banks' capital adequacy ratio (v74), loan-to-deposit ratio (v69), primary gap/GDP (v39), inflation (v75), REER acceleration (v97), real house price acceleration (v75), Real stock price acceleration (v41 and v42), FFR level and growth (v15), VIX (v11), US NEER change (v3), US yield spread (v12), TED spread (v14), 5-year cumulative inflation (v13), private sector credit growth (v71), house price growth (v75), stock price growth (v41 to v43), REER growth (v80), cross-border liabilities of banks to GDP growth (v92), financial contribution to GDP, contribution of construction to GDP, real growth of exports (v87), change in ToT (v6 and v7), Change in non-fuel TOT, Change in oil price, Absolute oil balance/GDP (v99), Change in growth of export partner relative to 5-year trend (v10), Bank-to-bank liabilities to AEs with financial crisis/GDP, Frequency of banking crises in AEs, Similarity to last year's crises, External debt/GDP and exports, Private external debt/GDP, Bank external debt/GDP, Private non-bank external debt/GDP,

the choice is constrained by data availability. The period covered is February 2003 to October 2023. The frequency of the variables is monthly. Where data were missing, they were imputed according to section 5.4. The 103 selected characteristics were classified into six groups:

1. Global variables: Implicit exchange rate in American Depositary Receipt (ADR)s – Bloomberg (v1); Coronavirus disease (COVID) dummy (v2); U.S. Dollar Index (DXY) Dollar Index – Bloomberg (v3); DXY gap over simple average long-run trend – Bloomberg (v4); Commodity Prices Index - BCRA (v5); Terms of trade Index - The National Institute of Statistics and Censuses of Argentina (INDEC) (v6); Terms of Trade year-over-year variation - INDEC (v7); Terms of Trade gap over simple average long-run trend - INDEC (v8); Economic Activity Index of Brazil seasonally adjusted – Central Bank of Brazil (v9); Economic Activity Index of Brazil output gap over long-run trend– Central Bank of Brazil (v10); VIX Index - Chicago Board Options Exchange (v11); 10-Year Treasury Constant Maturity Minus 2-Year Treasury Constant Maturity - Federal Reserve Bank of St. Louis (FRED) (v12); 5-Year Breakeven Inflation Rate Percent, Not Seasonally Adjusted - FRED (v13); TED Spread Percent, Not Seasonally Adjusted - FRED (v14); Federal Funds Effective Rate - Board of Governors of the Federal Reserve System (v15).

2. Fiscal sector: Revenue excluding BCRA (v16); Expense (v17); Interest (v18); Subsidies to the private sector (v19); Social benefits (v20); Compensation to employees (v21); Net Operating Balance (v22); Net Lending / Borrowing (v23); Revenue excluding BCRA in USD A3500 (v24); Expense in USD A3500 (v25); Interest in USD A3500 (v26); Subsidies to the private sector in USD A3500 (v27); Social benefits in USD A3500 (v28); Compensation to employees in USD A3500 (v29); Net Operating Balance in USD A3500 (v30); Net Lending / Borrowing in USD A3500 (v31); Debt in foreign and local currency year-over-year growth (v32); Revenue excluding

Total and external public debt/GDP, Cross-border liabilities of banks to GDP growth (v92), Contribution of financial intermediation to GDP, Contribution of construction to GDP, Real growth of exports (v87), Change in ToT (v6 and v7), Change in TOT of non-fuel commodities, Change in oil price, Absolute oil balance/GDP (v99), Change in growth of export partners relative to 5-year trend (v10), Bank-to-bank liabilities to AEs with financial crisis/GDP, Frequency of banking crises in AEs, similarity to previous year's crises, external debt/GDP and exports, private external debt/GDP, bank external debt/GDP, private non-bank external debt/GDP, total and external government debt/GDP, cross-border bank-to-bank liabilities/GDP, household liabilities/GDP, external liabilities/domestic credit, change in total debt/GDP in debt shocks, external debt/GDP growth.

BCRA in USD A3500, 12 month moving average (v33); Expense in USD A3500, 12 month moving average (v34); Interest in USD A3500, 12 month moving average (v35); Subsidies to the private sector in USD A3500, 12 month moving average (v36); Social benefits in USD A3500, 12 month moving average (v37); Compensation to employees in USD A3500, 12 month moving average (v38); Net Operating Balance in USD A3500, 12-month moving average (v39); Net Lending / Borrowing in USD A3500, 12 month moving average (v40). All the variables refer to the national nonfinancial public sector and have the Ministry of Economy of Argentina and BCRA as a source.

3. Financial and monetary sector: Argentine stock market (MERVAL) in ARS – Bloomberg (v41); MERVAL in real terms – Bloomberg (v42); MERVAL in implicit ARS-USD exchange rate in ADRs – Bloomberg (v43); Short run Bonds in ARS - Argentine Institute of Capital Markets (AICM) (v44); Long run Bonds nominated in ARS - AICM (v45); Short run Bonds nominated in USD - AICM (v46); Long run Bonds nominated in USD - AICM (v47); Short run Bonds in ARS expressed at official USD A3500 - AICM (v48); Long run Bonds nominated in ARS expressed at official USD A3500 - AICM (v49); Short run Bonds nominated in USD expressed at official USD A3500 - AICM (v50); Long run Bonds nominated in USD expressed at official USD A3500 - AICM (v51); Accumulated Monetary Base expansion in USD A3500 due to Net Foreign Exchange Purchases to the Private Sector – BCRA (v52); Accumulated Monetary Base expansion in USD A3500 due to overnight and short-term net credits to Banks – BCRA (v53); Accumulated Monetary Base expansion in USD A3500 due to Net Foreign Exchange Purchases to the National Treasury – BCRA (v54); Accumulated Monetary Base expansion in USD A3500 due to Other Operations with the National Treasury – BCRA (v55); Accumulated Monetary Base total expansion in USD A3500 – BCRA (v56); Accumulated Monetary Base expansion in USD A3500 due to Net Foreign Exchange Purchases to the Private Sector in terms of the stock of International Reserves – BCRA (v57); Accumulated Monetary Base expansion in USD A3500 due to overnight and short-term net credits to Banks in terms of the stock of International Reserves – BCRA (v58); Accumulated Monetary Base expansion in USD A3500 due to Net Foreign Exchange Purchases to the National Treasury in terms of the stock of International

Reserves – BCRA (v59); Accumulated Monetary Base expansion in USD A3500 due to Other Operations with the National Treasury in terms of the stock of International Reserves – BCRA (v60); Accumulated Monetary Base total expansion in USD A3500 in terms of the stock of International Reserves – BCRA (v61); Loans to the Private Sector in USD – BCRA (v62); Loans to the Private Sector in ARS expressed in USD A3500 – BCRA (v63); Loans to the Private Sector in USD to International Reserves – BCRA (v64); Loans to the Private Sector in ARS expressed in USD A3500 to International Reserves– BCRA (v65); Deposits in USD from the private sector– BCRA (v66); Deposits in ARS from the private sector in USD A3500 – BCRA (v67); M2 monetary aggregate expressed in USD A3500 – BCRA (v68); Loan deposit ratio – BCRA (v69); Financial system liquidity in percentage of total assets – BCRA (v70); Financial system credit to the public sector in percentage of total assets – BCRA (v71); Financial system Return Over Equity in percentage – BCRA (v72); Financial system Non Performing Loans to the Private Sector to total loans to the private sector in percentage – BCRA (v73); Capital Integration Level 1 Credit Risk – BCRA (v74).

4. Real sector: Consumer Price Index – INDEC and BCRA (v75); Consumer Price Index year-over-year variation – INDEC (v76); Monthly Economic Activity Estimator (EMAE) INDEC (v77); Monthly output gap estimated as a 12-month mobile average – EMAE INDEC (v78); Average Taxable Remuneration of Stable Workers– RIPTE Ministry of Labor and Social Security of Argentina (MTEySS)- divided by Consumer Price Index – INDEC (v79).

5. External sector: REER gap over long-run trend estimated as a simple average – BCRA (v80); Goods Imports – Argentine Commercial Exchange (ICA) INDEC (v81); Goods Exports – ICA INDEC (v82); Goods Trade Balance – ICA INDEC (v83); Export prices gap over trend estimated as a simple average – INDEC (v84); Import quantities gap over trend estimated as simple average – INDEC (v85), International Reserve year-over-year growth – BCRA (v86); Goods Exports year-over-year growth – ICA INDEC (v87), Goods Trade Balance at the exchange market in USD – BCRA (v88); Service Trade Balance at the exchange market in USD – BCRA (v89); Primary Income at the exchange market in USD – BCRA (v90);

Non-resident direct and portfolio investment in USD – BCRA (v91); Financial loans, debt securities and lines of credit at the exchange market in USD – BCRA (v92); Formation of external assets in USD – BCRA (v93); International Reserves to Imports ICA – BCRA, INDEC (v94); Official and alternative exchange rate gap in percentage – BCRA, INDEC (v95); Primary Income to Goods Trade Balance at the exchange market in USD – BCRA (v96), Multilateral Real Exchange Rate – BCRA (v97); Non Energy Goods Trade Balance – ICA INDEC (v98); Energy Goods Trade Balance – ICA INDEC (v99).

6. Political variables: a dummy variable for the ruling party of the government, assigning a value of 1 for Peronist governments, and 0 otherwise (v100); a dummy variable for electoral elections, assigning a value of 1 when presidential or congress elections are taking place within the next 12 months, and 0 otherwise (v101); a dummy variable for presidential elections, assigning a value of 1 when presidential elections are taking place within the next 12 months, and 0 otherwise (v102); a dummy variable for congress elections, assigning a value of 1 when congress elections are taking place within the next 12 months, and 0 otherwise (v103).

6.2 Training-validation-test sets

The balance of data in ML is a crucial aspect that can have a significant impact on model performance. In this context, we analyse a dataset consisting of 238 observations, with a relatively balanced number of crisis and non-crisis observations [37]. It is worth noting that the dataset was already balanced, and therefore, no downsampling was required.

The dataset was divided into two samples: a training sample of 178 observations and a test sample of 60 observations. The sample was divided using a stratified and random sampling method. First, we divided our sample into two sub-samples based on the crisis label. We then randomly selected 75 per cent of the crisis sub-sample and 75 per cent

[37] The data set is relatively balanced due to the high frequency of external crises in Argentina. The number of crises in Argentina between 2003 and 2023 depends on how a crisis is defined and the length of the signalling period. Adopting the IMF (2021) crisis definition, including SBA and exchange controls and 24 months of signalling, results in 158 monthly crisis episodes. Excluding SBAs, exchange rate controls, and 24 months of signalling resulted in 131 crises. Excluding SBA and exchange controls and 12 months of signalling gives 91 crisis episodes. Excluding SBA and exchange controls and 1-month signalling results in 41 crises.

of the non-crisis sub-sample. These selected sub-samples were combined to form the training validation set. The remaining observations formed the test set. This method ensured that both the training and test sets had a proportional representation of crisis and non-crisis observations.

6.3 Empirical findings: research question I

As mentioned in Section 1.2, the research aimed to investigate the feasibility of using machine learning models based on trees to develop an EWS that could accurately predict external crises in Argentina. The hypothesis was that these methods could achieve high out-of-sample performance in predicting balance of payments crises.

This section provides empirical evidence for each method. The evaluation considered different crisis definitions. The following alternative cases were evaluated:

(A) The crisis index of Section 2.2 including exchange rate controls, SBA and a 24-month signalling period. A total of 103 variables were included;

(B) Same as (A), but only 88 variables were considered, excluding nominal variables;

(C) Same as (B), but without considering exchange rate controls and SBA;

(D) Same as (C), but with a signalling period of 12 months;

(E) Same as (D), but with a signalling period of one month.

Do tree-based methods maintain good out-of-sample performance when the main selected variables are eliminated? After presenting evidence for each scenario, the methods are reassessed by removing the previously selected main features.

6.3.1 Classification and Regression Trees (CART)

This section presents empirical evidence demonstrating that CART models exhibit strong out-of-sample performance and can predict a balance of payments crisis up to 24 months in advance. This is also applicable to alternative crisis definitions and signalling periods. CART models showed excellent performance for broad crisis definitions and longer signalling periods. Despite the crisis definition becoming progressively narrower and the signalling horizon shorter, the CART performance remains high but lower than before. Furthermore, excluding alternative features did not significantly affect the out-of-sample performance of these models.

Different criteria were used to construct different trees. The maximal tree was constructed using information, entropy and Gini criteria. Then, 10-fold cross-validation strategy was used to determine the hyper-parameter α and to prune the maximal tree [38]. After selection, each tree was subjected to an out-of-sample evaluation. The resulting predictions were presented in a confusion matrix.

Case A: including exchange rate controls, SBA and a 24-month signalling period. Alternative trees were generated using different criteria and 103 variables [39]. The first variable selected for all cases was v34, i.e. the 12-month moving average of national non-financial public sector expenditure in USD at the official exchange rate A3500. These trees further divide this node using different variables under different criteria [40]. Each tree has four nodes and five leaves.

Then, the trees were pruned. The hyper-parameter α was cross-validated using both 5 and 10 folds. In each and every case, only one split was selected. Figure 17 shows the pruned tree. All trees select a threshold of USD 5564 million per month for v34. If v34 is above this threshold, there is a 97 per cent probability of a crisis. If v34 is below this threshold, there is a 100 per cent probability of no-crisis.

Table 2 shows the confidence matrix of the CART model in predicting events outside the training data. The tree correctly classified all 40 instances of months with a crisis label in the test set. In addition, the tree correctly predicted all 20 no-crisis months in

[38] Other hyper-parameters were determined in a way that did not affect the tree construction. The minimum number of observations required for a split was set to 2, the minimum observations required for a bucket was set to 1, and the maximum tree depth was set to 30 splits.

[39] Information criteria and 5-fold cross-validation, information criteria and 10-fold cross-validation, entropy criteria and 5-fold cross-validation, entropy criteria and 10-fold cross-validation, Gini criteria and 5-fold cross-validation, Gini criteria and 10-fold cross-validation.

[40] The information criteria selected the international variables v9, v1 and v5. The entropy and Gini criteria selected v92, v90 and v13 to subdivide successive nodes.

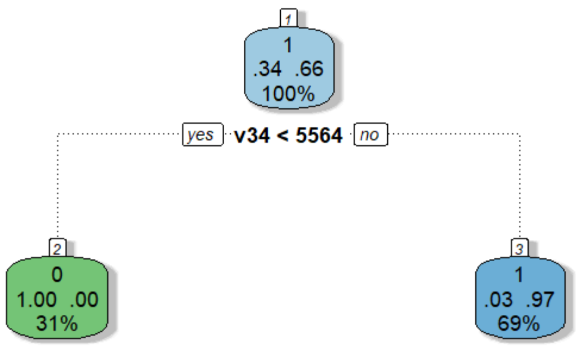


Figure 17: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 1. Variables: 103. Case: A.

the test set. The models achieved 100 per cent accuracy, recall, precision, and F1 score, as demonstrated in Figure 18.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	20	0
	Crisis	0	40

Table 2: Confusion matrix: CART model. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 1. Variables: 103. Case: A. Set: test.

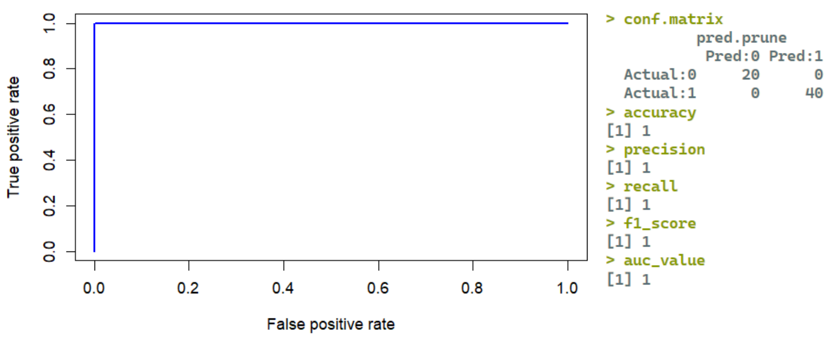


Figure 18: CART pruned tree out-of-sample performance: ROC curve and metrics. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 1. Variables: 103. Case: A.

Case B: same as (A), but only 88 variables were considered, excluding nominal variables. Different criteria and 88 variables were used to generate alternative trees. The first variable selected for all cases, as in case A, was v34. These trees further divided this node using different variables under different criteria. Each 'maximal tree' has four nodes and five leaves.

The trees were then pruned, and the hyper-parameter α was cross-validated using both 5 and 10 folds. In each case, only one split was selected. All trees use a threshold of USD 5564 million per month for v34. As in figure Figure 17, if v34 exceeds this threshold, there is a 97 per cent probability of a crisis. If v34 is below this threshold, there is a 100 per cent probability of no-crisis.

Table 3 displays the confidence matrix of the CART model in predicting events outside the training data. The tree accurately classified all 40 crisis labels and all 20 non-crisis months in the test set. The models achieved 100 per cent accuracy, recall, precision, and F1 score, as in Figure 18.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	20	0
	Crisis	0	40

Table 3: Confusion matrix: CART model. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 1. Variables: 88. Case: B. Set: test.

Case C: same as (B), but without considering exchange rate controls and SBA. The first variable selected for all alternative specifications was v37, which represents the 12-month moving average of social benefits in official USD. All trees were assigned a no-crisis label if social benefits were less than USD 2891 million per month (Figure 19 illustrates the case).

In contrast to cases A and B, trees in case C have more splits. Under the information criteria, v39 [41] and v78 [42] are part of the pruned tree (under both 5 and 10 cross-validation). Under the entropy and Gini criteria, v57 [43] is in the pruned tree.

[41] Net operating balance in USD A3500, 12-month moving average.

[42] Monthly output gap estimated as a 12-month moving average - EMAE INDEC.

[43] Accumulated monetary base expansion in USD A3500 due to net foreign exchange purchases by the

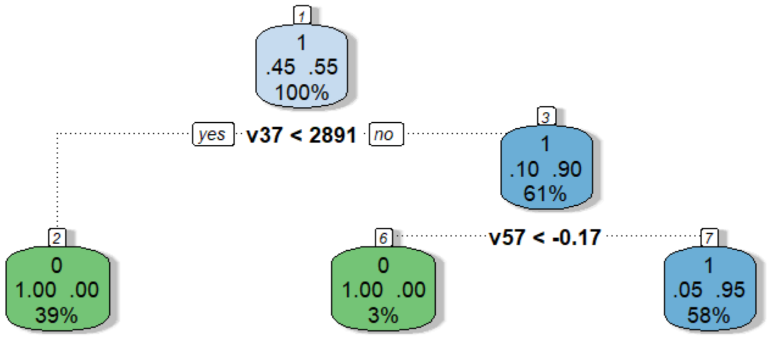


Figure 19: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.025. Selected number of splits: 2. Variables: 88. Case: C.

Regarding out-of-sample performance, the CART model performs less well under Scenario C than Scenarios A and B, as shown in Figures 20 and Table 4. In the case of Scenario C, the performance is lower but still high.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	26	1
	Crisis	1	32

Table 4: Confusion matrix: CART model. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.025. Selected number of splits: 2. Variables: 88. Case: C. Set: test.

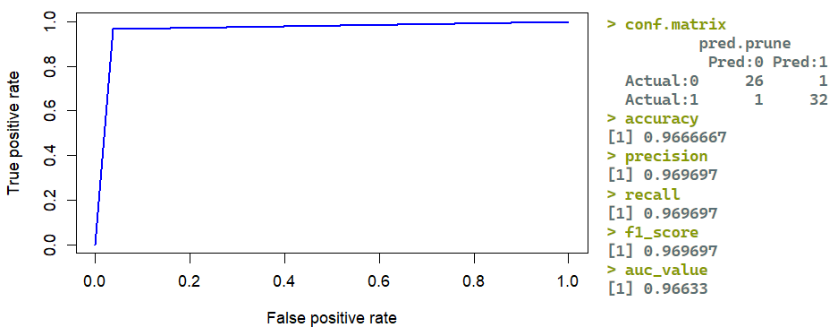


Figure 20: CART pruned tree out-of-sample performance: ROC curve and metrics. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.025. Selected number of splits: 2. Variables: 88. Case: C.

private sector in relation to the stock of international reserves - BCRA.

Case D: same as (C), but with a signalling period of 12 months. The first variable selected for all alternative specifications was v37. Compared to scenarios A and B, the trees in scenario D have more splits. The pruned tree in scenario D, under both 5 and 10 cross-validations, includes v12 [44] and v63 [45] based on the information criteria. Under the entropy and Gini criteria, the pruned tree (Figure 21) includes v32, v35, v48 [46] and v3 [47] (under both 5 and 10 cross-validation).

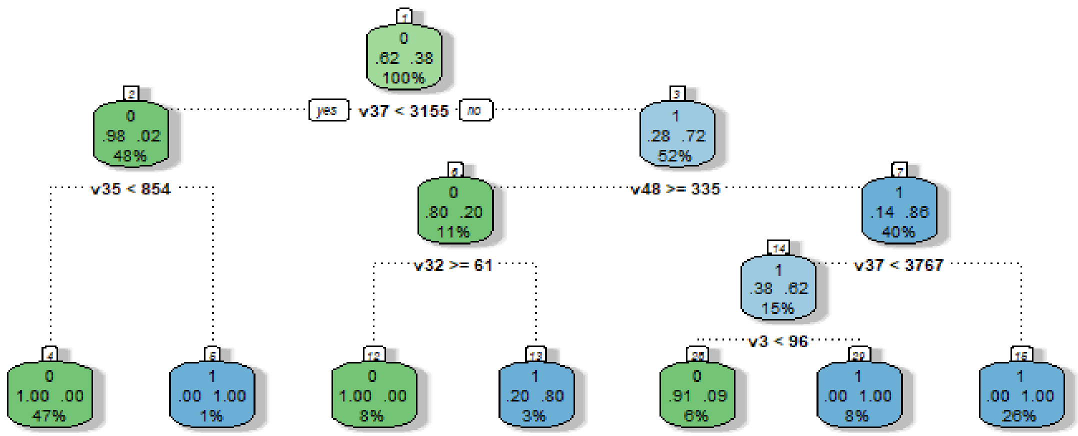


Figure 21: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.01470588. Selected number of splits: 6. Variables: 88. Case: D.

The CART model's out-of-sample performance is lower in scenario D compared to scenarios A and B, but still high, as shown in Figure 22 and Table 5.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	36	1
	Crisis	0	23

Table 5: Confusion matrix: CART model. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.01470588. Selected number of splits: 6. Variables: 88. Case: D. Set: test.

[44] 10-Year Treasury Constant Maturity Minus 2-Year Treasury Constant Maturity - Federal Reserve Bank of St. Louis (FRED).

[45] Loans to the private sector in ARS expressed in USD A3500 - BCRA.

[46] Short run bonds in ARS expressed in official USD A3500 - AICM.

[47] U.S. Dollar Index (DXY) Dollar Index - Bloomberg.

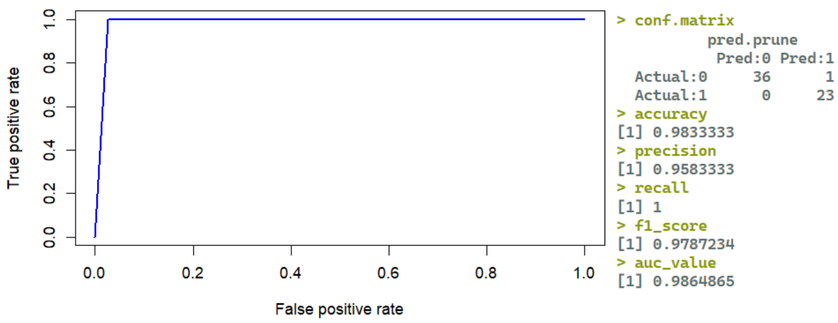


Figure 22: CART pruned tree out-of-sample performance: ROC curve and metrics. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.01470588. Selected number of splits: 6. Variables: 88. Case: D.

Case E: same as (D), but with a signalling period of one month. The first variable selected for all alternative specifications was v86, i.e. the year-on-year growth in the central bank's international reserves. Other relevant variables were v60 [48], v52 [49] and v88 [50].

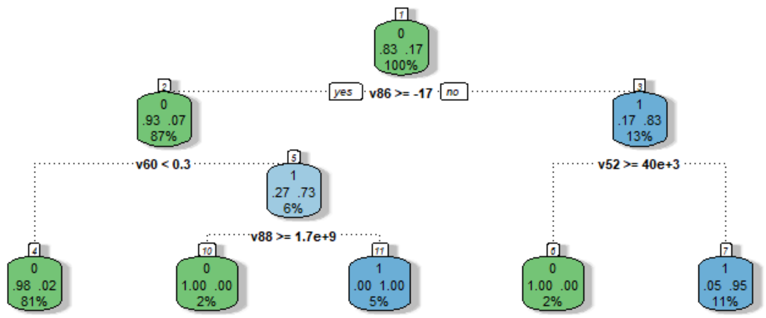


Figure 23: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 4. Variables: 88. Case: E.

The worst performance of the CART model is in scenario E as shown in Table 6. Nevertheless, the CART model still performed well in terms of out-of-sample performance,

[48]Accumulated monetary base expansion in USD A3500 due to other operations with the National Treasury in terms of the stock of international reserves - BCRA.

[49]Accumulated monetary base expansion in USD A3500 due to net foreign exchange purchases from the private sector - BCRA.

[50]Goods Trade Balance at the exchange market in USD - BCRA.

as shown in Figure 24.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	49	1
	Crisis	2	9

Table 6: Confusion matrix: CART model. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 4. Variables: 88. Case: E. Set: test.

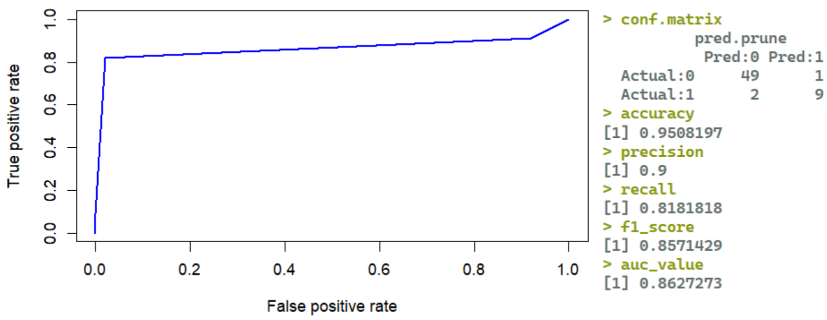


Figure 24: CART pruned tree out-of-sample performance: ROC curve and metrics. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 4. Variables: 88. Case: E.

Removal of key features Does the CART method still show good out-of-sample performance after removing the main selected variables? In this exercise, the moving average fiscal variables v34, v35, v36, v37, v38, v39, v40 and the accumulated monetary variables v54, v55, v57, v59, v60 are excluded. As a result, the number of features has been reduced from 88 to 76.

For scenario B with 76 variables, the pruned tree selected only one split and variable v77. Table 7 demonstrates that CART maintains its out-of-sample performance, as in case B, with 88 variables and v34 as the primary variable.

For Scenario C with 76 variables, the pruned tree (Figure 25) selected 8 splits, with the first split given by Social benefits in USD (v28). The model performs less but similar to Scenario C with 88 variables. Table 8 shows its out-of-sample performance, and Figure 26 shows the ROC curve.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	20	0
	Crisis	0	40

Table 7: Confusion matrix: CART model. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 1. Variables: 76. Case: B. Set: test.

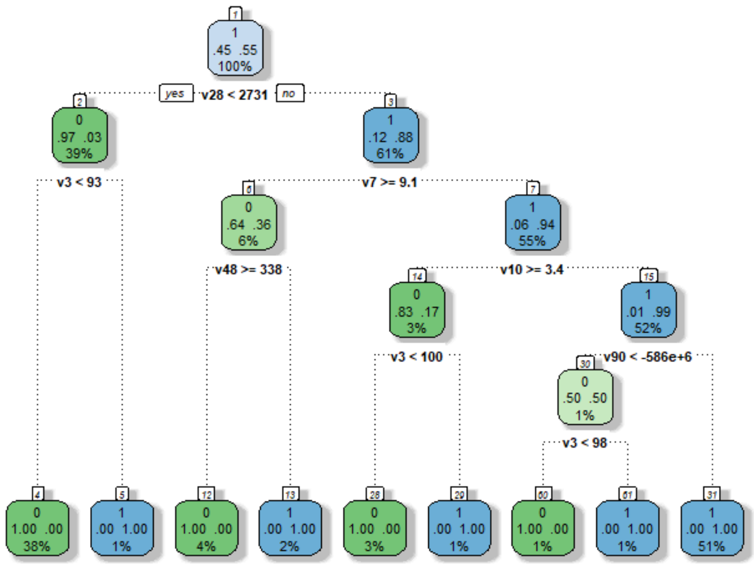


Figure 25: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.0001. Selected number of splits: 8. Variables: 76. Case: B.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	25	2
	Crisis	2	31

Table 8: Confusion matrix: CART model. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.0001. Selected number of splits: 8. Variables: 76 (over 76). Case: C. Set: test.

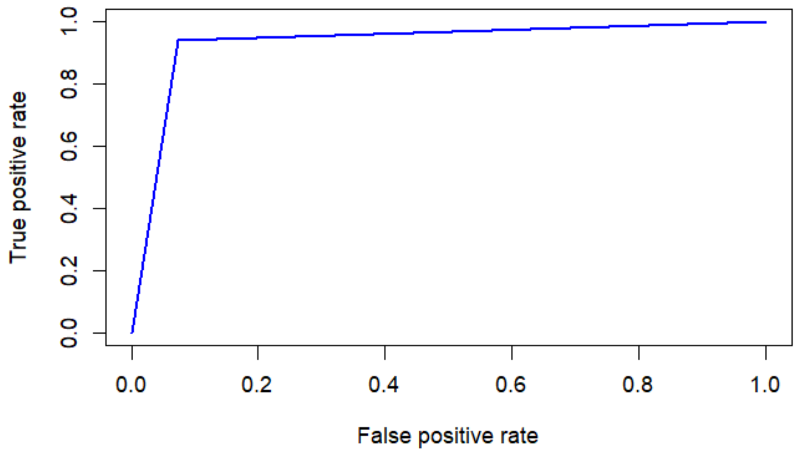


Figure 26: CART pruned tree out-of-sample performance: ROC curve. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.0001. Selected number of splits: 8. Variables: 76. Case: C.

For Scenario D with 76 variables, the pruned tree (Figure 27) shows a lower but still high performance compared to Scenario D with 88 variables. Table 9 shows its out-of-sample performance and Figure 28 shows the ROC curve.

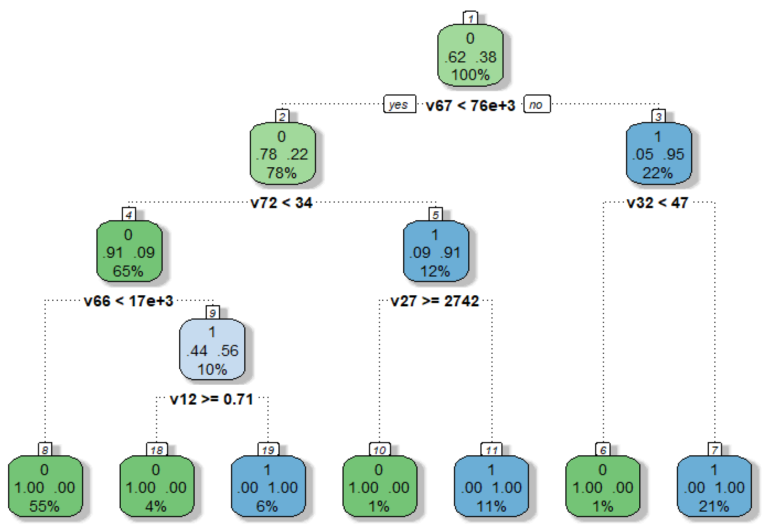


Figure 27: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.0001. Selected number of splits: 6. Variables: 76. Case: D.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	34	1
	Crisis	2	23

Table 9: Confusion matrix: CART model. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.0001. Selected number of splits: 6. Variables: 76. Case: D. Set: test.

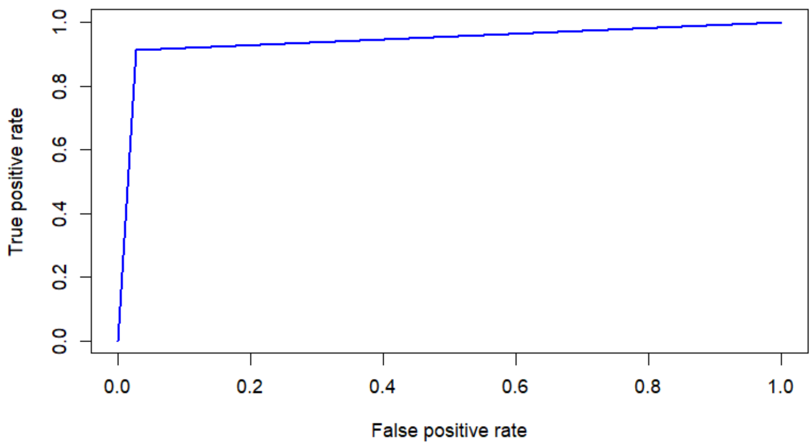


Figure 28: CART pruned tree out-of-sample performance: ROC curve. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.0001. Selected number of splits: 6. Variables: 76. Case: D.

Finally, for Scenario E with 76 variables, the pruned tree (Figure 29) performs identically to Scenario E with 88 variables. Table 10 shows its out-of-sample performance, and Figure 30 shows the ROC curve.

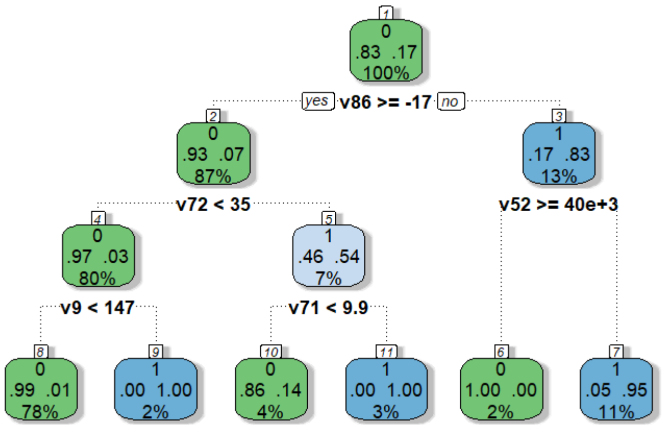


Figure 29: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 5. Variables: 76. Case: E.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	49	1
	Crisis	2	9

Table 10: Confusion matrix: CART model. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 5. Variables: 76. Case: E. Set: test.

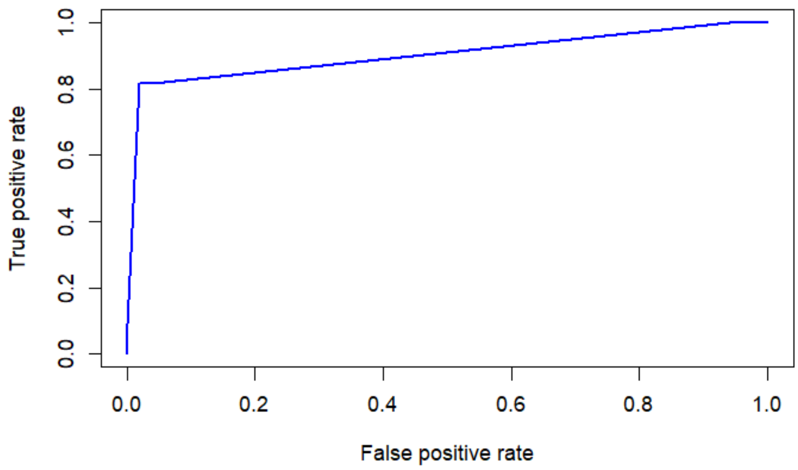


Figure 30: CART pruned tree out-of-sample performance: ROC curve. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 5. Variables: 76. Case: E.

6.3.2 Bootstrap Aggregating (BA)

This section presents empirical evidence to demonstrate the high effectiveness of BA models in predicting a balance of payments crisis up to 24 months in advance. This pertains to alternative definitions of crises and periods of signalling. Although the definition of the crisis has become increasingly narrow and the signalling horizon has become shorter, the BA models still perform well, albeit slightly lower than before. Additionally, excluding alternative features did not significantly affect the out-of-sample performance of these models.

To create a BA model, a set number of bootstrap samples are generated, and each sample is utilised to construct a decision tree. Additional bootstrap samples are taken if required to ensure that OOB errors stabilise at a minimum. The algorithm scores all 103 (case A) or 88 (cases B, C, D, and E) variables in each node to build each tree. The OOB error is presented as a measure of the model's in-sample performance.

After constructing models with the training validation sample, bagged trees are used to score the test observations. The number of bagged trees can be tens of thousands or even a million. The trees vote on whether the observations indicate a crisis or not. If over 50 per cent of the trees indicate a crisis, the observation is labelled as 'crisis'. Otherwise, it is labelled as 'no-crisis'. Upon completion of the voting process, the model's out-of-sample performance is evaluated by computing the confusion matrix and performance metrics.

Case A: including exchange rate controls, SBA and a 24-month signalling period. Bagged models have a perfect out-of-sample performance in case A. Figure 31 shows that 10 thousand trees had an OOB error of 2.81 per cent. Subsequently, models trained using the BA technique performed exceptionally well when tested on unseen data, as demonstrated in Figure 32 and Table 11.

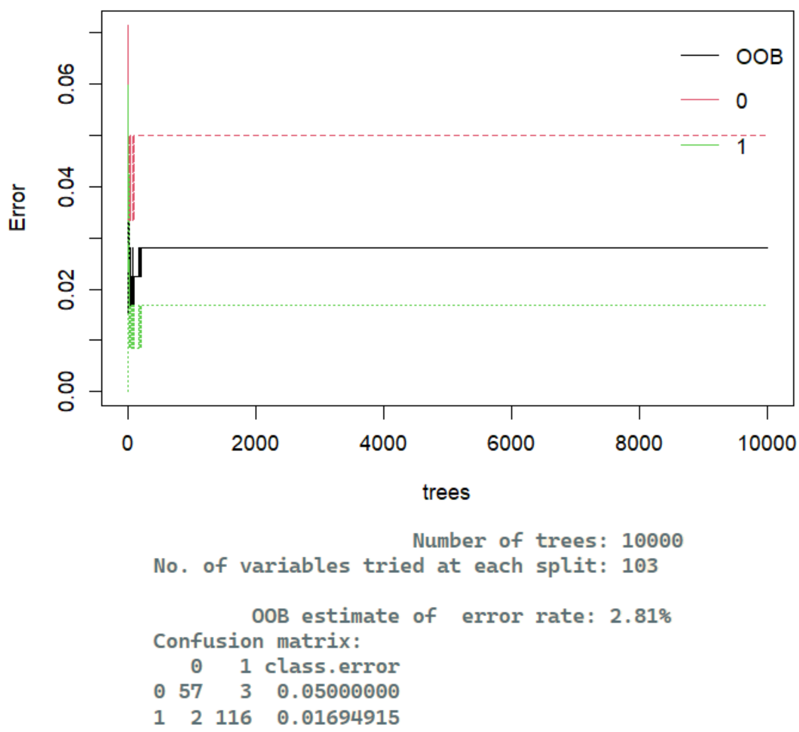


Figure 31: BA FP (1), FN (0) and total OOB errors. Variables: 103 (over 103). Trees: 10’000. OOB error: 2.81 per cent. Case A.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	20	0
	Crisis	0	40

Table 11: Confusion matrix: BA model. Variables: 103 (over 103). Case A. Set: test.

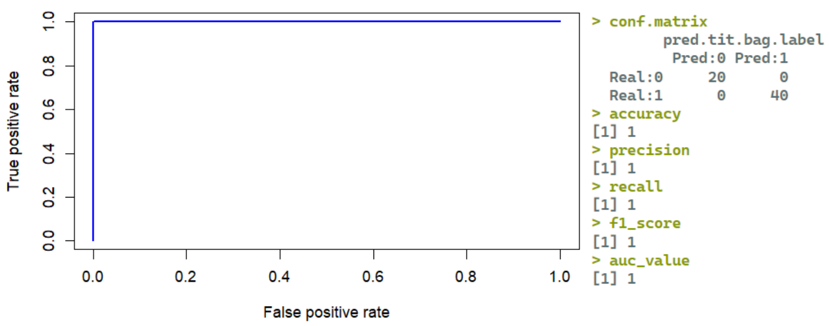
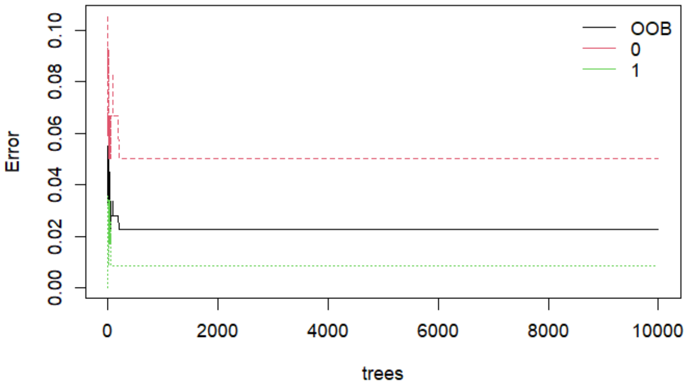


Figure 32: BA out-of-sample performance: ROC curve and metrics. Variables: 103 (over 103). Trees: 10’000. OOB error: 2.81 per cent. Case A.

Case B: same as (A), but only 88 variables were considered, excluding nominal variables. In case B, the bagged model also exhibits perfect out-of-sample performance. Figure 33 illustrates that 10'000 trees had an OOB error rate of 2.25 per cent. The models trained using BA demonstrate exceptional out-of-sample performance, as depicted in Figure 34 and Table 12.




```
                     Number of trees: 10000
No. of variables tried at each split: 88

        OOB estimate of  error rate: 2.25%
Confusion matrix:
   0   1 class.error
0 57   3 0.050000000
1  1 117 0.008474576
```

Figure 33: BA FP (1), FN (0) and total OOB errors. Variables: 88 (over 88). Trees: 10'000. OOB error: 2.25 per cent. Case B.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	20	0
	Crisis	0	40

Table 12: Confusion matrix: BA model. Variables: 88 (over 88). Case B. Set: test.

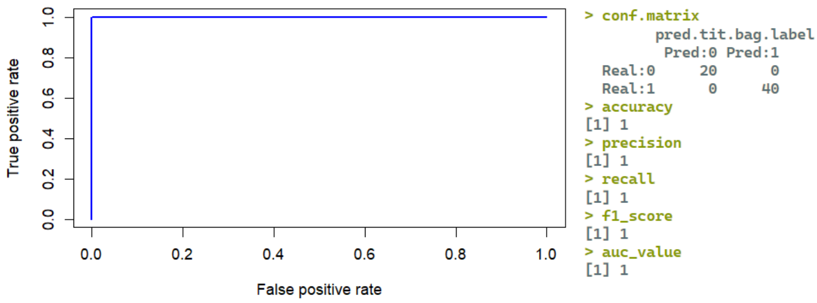


Figure 34: BA out-of-sample performance: ROC curve and metrics. Variables: 88 (over 88). Trees: 10'000. OOB error: 2.25 per cent. Case B.

Case C: same as (B), but without considering exchange rate controls and SBA. In case C, the bagged models have a lower but still high out-of-sample performance. Ten thousand trees had an OOB error of 2.25 per cent (Figure 35). The models trained using BA show high out-of-sample performance, as shown in Figure 36 and Table 13.

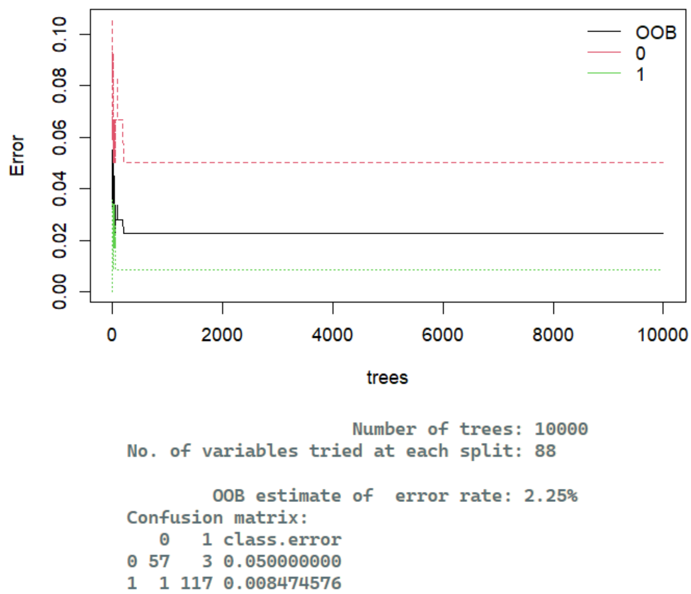


Figure 35: BA FP (1), FN (0) and total OOB errors. Variables: 88 (over 88). Trees: 10'000. OOB error: 2.25 per cent. Case C.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	26	1
	Crisis	0	33

Table 13: Confusion matrix: BA model. Variables: 88 (over 88). Case C. Set: test.

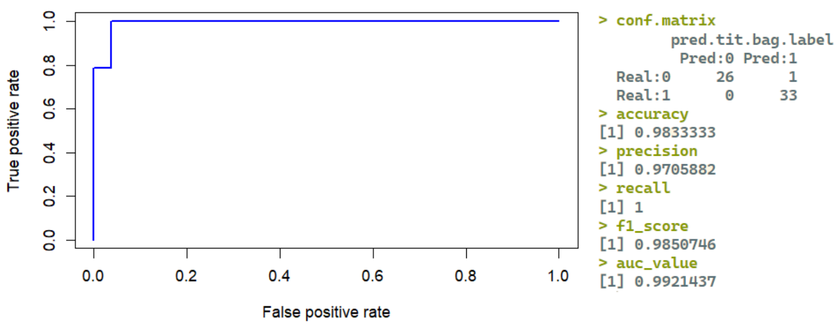


Figure 36: BA out-of-sample performance: ROC curve and metrics. Variables: 88 (over 88). Trees: 10'000. OOB error: 2.25 per cent. Case: C.

Case D: same as (C), but with a signalling period of 12 months. The BA model maintains a high out-of-sample performance for scenario D. Figure 37 illustrates that 10'000 trees had an OOB error rate of 2.25 per cent. The models trained using BA show exceptional out-of-sample performance, as shown in Figure 38 and Table 14.

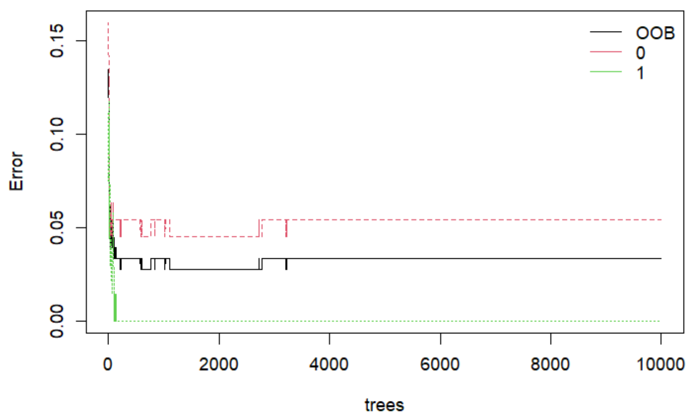


Figure 37: BA FP (1), FN (0) and total OOB errors. Variables: 88 (over 88). Trees: 10'000. OOB error: 2.25 per cent. Case: D

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	37	0
	Crisis	0	23

Table 14: Confusion matrix: BA model. Variables: 88 (over 88). Case D. Set: test.

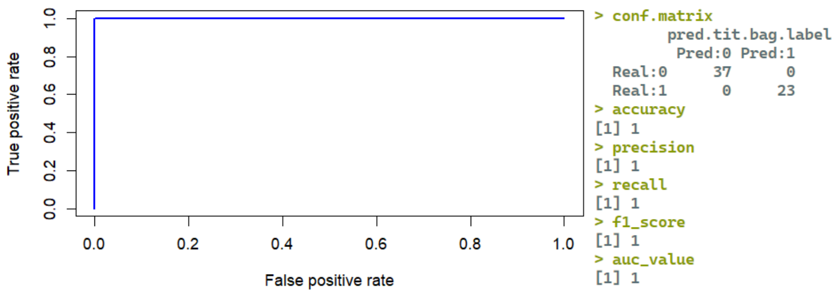


Figure 38: BA out-of-sample performance: ROC curve and metrics. Variables: 88 (over 88). Trees: 10'000. OOB error: 2.25 per cent. Case: D.

Case E: same as (D), but with a signalling period of one month. In the case of BA models, as in the case of CART models, the worst performance is observed in scenario E. Figure 39 illustrates that 100'000 trees had an OOB error rate of 6.78 per cent. Although the out-of-sample performance is lower in this scenario, it is still high, as shown in Figure 40 and Table 15.

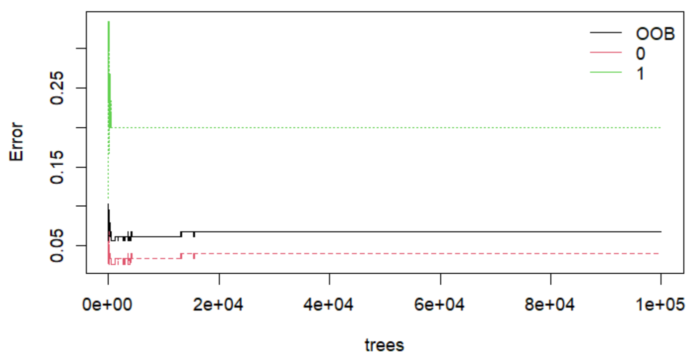


Figure 39: BA FP (1), FN (0) and total OOB errors. Variables: 88 (over 88). Trees: 100'000. OOB error: 6.78 per cent. Case: E.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	49	1
	Crisis	1	10

Table 15: Confusion matrix: BA model. Variables: 88 (over 88). Case E. Set: test.

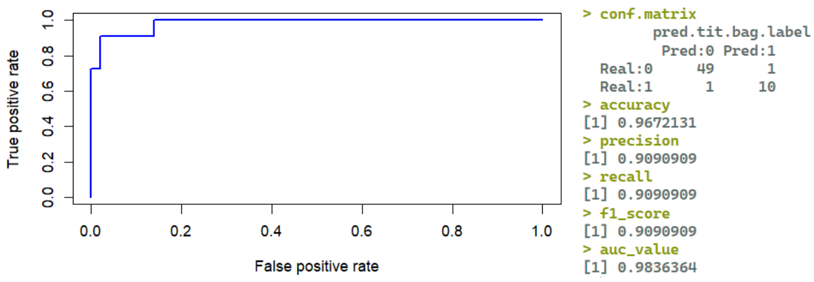


Figure 40: BA out-of-sample performance: ROC curve and metrics. Variables: 88 (over 88). Trees: 100'000. OOB error: 6.78 per cent. Case: E

Removal of key features Do BA methods still show good out-of-sample performance after removing the main independent variables? In this excersise, the moving average fiscal variables v34, v35, v36, v37, v38, v39, v40 and the accumulated monetary variables v54, v55, v57, v59, v60 are excluded. As a result, the number of features has been reduced from 88 to 76. For scenario B, the BA model had a similar confidence matrix (Table 16) as in case B with 88 variables.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	20	0
	Crisis	0	40

Table 16: Confusion matrix: BA model. Variables: 76 (over 76). Case: B. Set: test.

For Scenario C with 76 variables, the BA model performs worse than Scenario C with 88 variables but still with a high out-of-sample score. Table 17 shows its out-of-sample performance, and Figure 41 shows the ROC curve.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	25	2
	Crisis	1	32

Table 17: Confusion matrix: BA model. Variables: 76 (over 76). Case: C. Set: test.

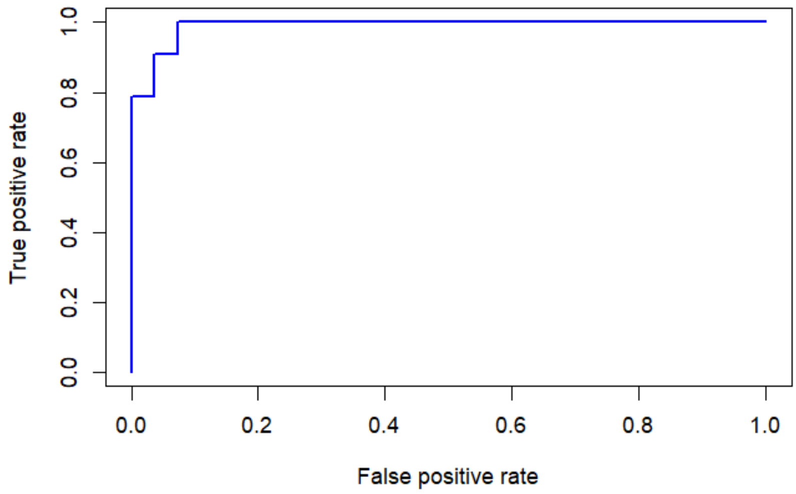


Figure 41: BA out-of-sample performance: ROC curve. Variables: 76 (over 76). Trees: 10'000. OOB error: 5.06 per cent. Case: C.

For Scenario D with 76 variables, the BA model performs similarly to Scenario C with 88 variables. Table 18 shows its out-of-sample performance, and Figure 42 shows the ROC curve.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	36	1
	Crisis	1	22

Table 18: Confusion matrix: BA model. Variables: 76 (over 76). Case: D. Set: test.

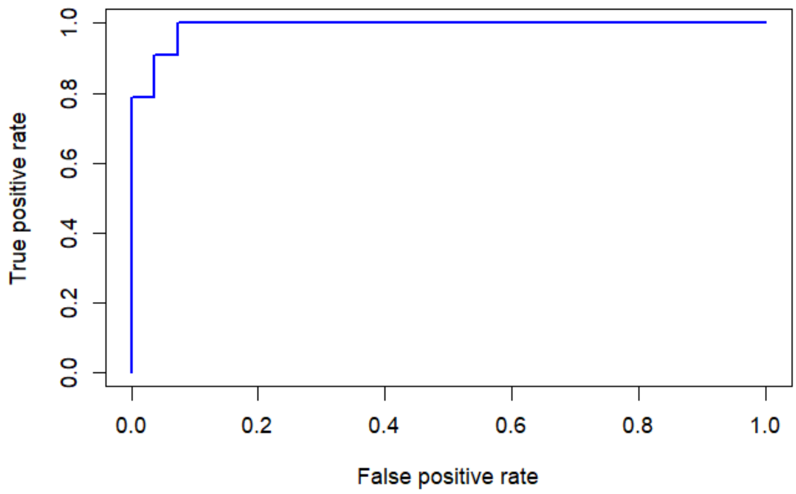


Figure 42: BA out-of-sample performance: ROC curve. Variables: 76 (over 76). Trees: 10'000. OOB error: 3.37 per cent. Case: D.

For Scenario E with 76 variables, the BA model performs worse than Scenario E with 88 variables but still well. Table 19 shows its out-of-sample performance, and Figure 43 shows the ROC curve.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	48	2
	Crisis	2	9

Table 19: Confusion matrix: BA model. Variables: 76 (over 76). Case: E. Set: test.

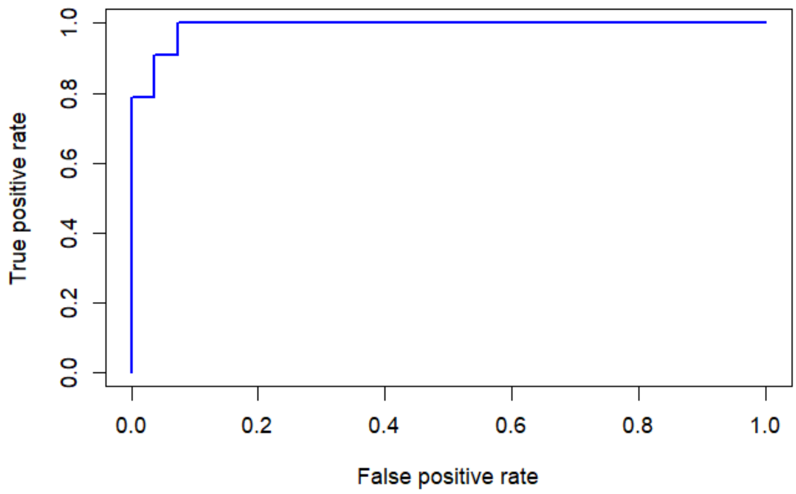


Figure 43: BA out-of-sample performance: ROC curve and metrics. Variables: 76 (over 76). Trees: 10'000. OOB error: 6.78 per cent. Case: E.

6.3.3 Random Forest (RF)

This section presents empirical evidence demonstrating the strong out-of-sample performance of RF models in predicting a balance of payments crisis up to 24 months in advance. The models exhibit excellent performance for broad crisis definitions and longer signalling periods, which holds true for alternative crisis definitions and signalling periods as well. Despite the progressively narrower crisis definition and shorter signalling horizon, the RF performance remains high, albeit slightly lower than before. Additionally, excluding alternative features has no significant impact on the out-of-sample performance of these models.

The process begins by generating a predetermined number of bootstrap samples. Each sample is then used to construct a decision tree. If required, additional bootstrap samples are taken to ensure that the OOB error stabilises at a minimum. The number of variables assigned to each node is cross-validated to ensure optimal results.

After completing the training, the RF models are evaluated out-of-sample. The selected model's trees vote on whether the observations indicate a crisis. If over 50 per cent of the trees indicate a crisis, the observation is labelled as 'crisis'. If less, the observation is labelled as 'no-crisis'.

Case A: including exchange rate controls, SBA and a 24-month signalling period. Using 10-fold cross-validation, we obtained 52 variables as the hyper-parameter. According to Figure 44, 100'000 trees had an in-sample OOB error rate OF 3.37 per cent. In case A, the RF model show high out-of-sample performance, as shown in Figures 45 and Table 20.

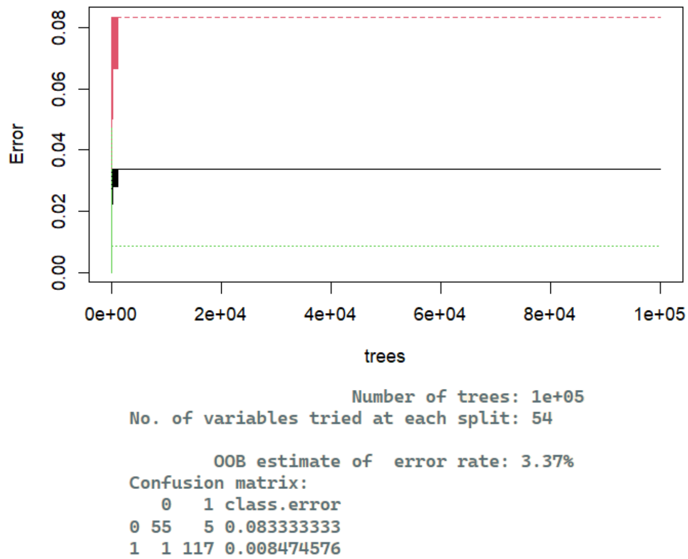


Figure 44: RF FP (1), FN (0) and total OOB errors. Cross-validation folds: 10. Variables selected by CV: 52 (over 103). Trees: 100'000. OOB error: 3.37 per cent. Case: A

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	20	0
	Crisis	0	40

Table 20: Confusion matrix: RF model. Cross-validation folds: 10. Variables selected by CV: 52 (over 103). Case: A. Set: test.

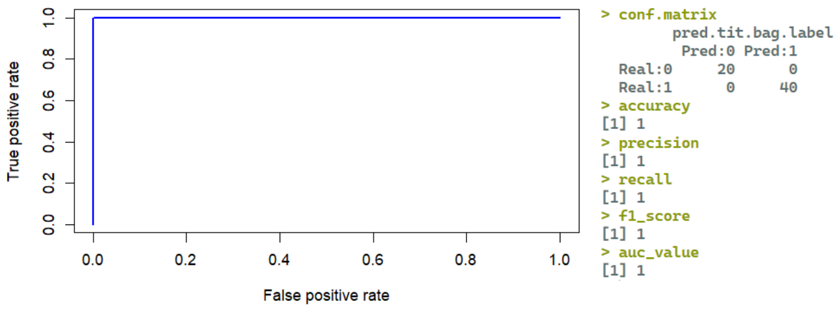


Figure 45: RF out-of-sample performance: ROC curve and metrics. Cross-validation folds: 10. CV variables: 52 (over 103). Trees: 100'000. OOB error: 3.37 per cent. Case: A

Case B: same as (A), but only 88 variables were considered, excluding nominal variables. Using 10-fold cross-validation, we obtained 44 variables as the hyperparameter. According to Figure 46, 10'000 trees had an in-sample OOB error rate of 2.81 per cent. As in case A, in case B, the RF model show high out-of-sample performance, as shown in Figures 47 and Table 21.

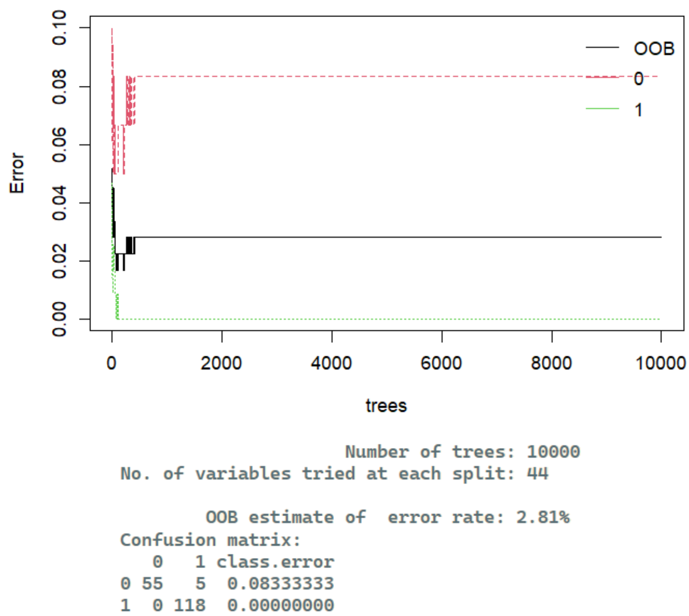


Figure 46: RF FP (1), FN (0) and total OOB errors. Cross-validation folds: 10. Variables selected by CV: 44 (over 88). Trees: 10'000. OOB error: 2.81 per cent. Case: B.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	20	0
	Crisis	0	40

Table 21: Confusion matrix: RF model Cross-validation folds: 10. Variables selected by CV: 44 (over 88). Case: B. Set: test.

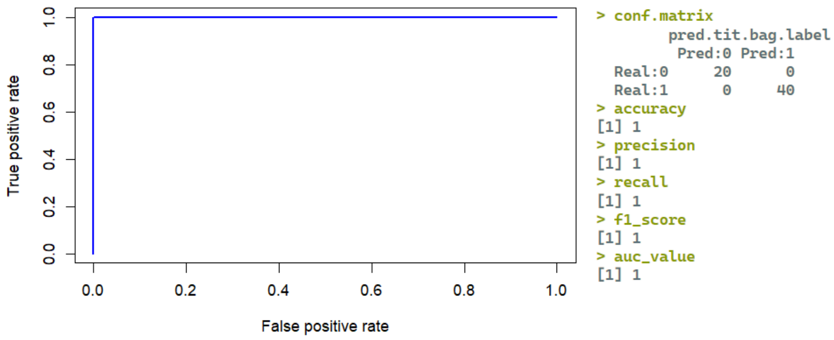


Figure 47: RF out-of-sample performance: ROC curve and metrics. Cross-validation folds: 10. CV variables: 44 (over 88). Trees: 10'000. OOB error: 2.81 per cent. Case: B.

Case C: same as (B), but without considering exchange rate controls and SBA. Using 10-fold cross-validation, we obtained 44 variables as the hyper-parameter. According to Figure 48, 10'000 trees had an in-sample OOB error rate of 2.81 per cent. The RF model shows high out-of-sample performance, as shown in Figures 49 and Table 22.

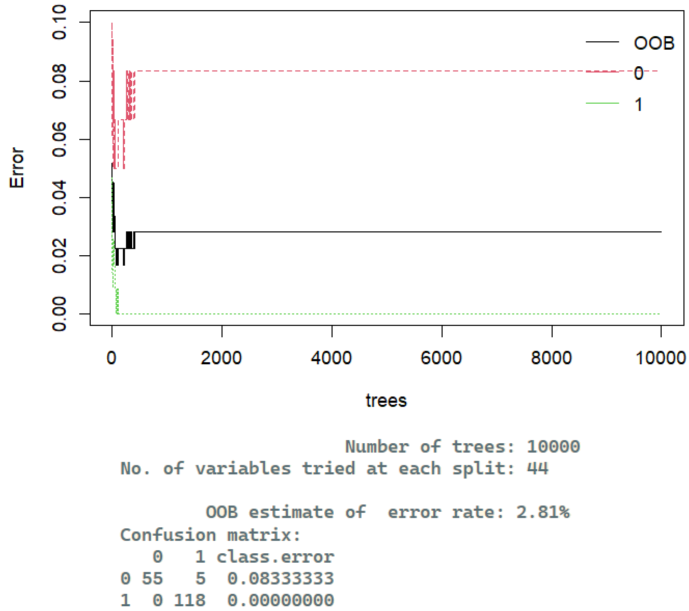


Figure 48: RF FP (1), FN (0) and total OOB errors. Cross-validation folds: 10. Variables selected by CV: 44 (over 88). Trees: 10'000. OOB error: 2.81 per cent. Case: C.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	27	0
	Crisis	1	32

Table 22: Confusion matrix: RF model. Cross-validation folds: 10. Variables selected by CV: 44 (over 88). Case: C. Set: test.

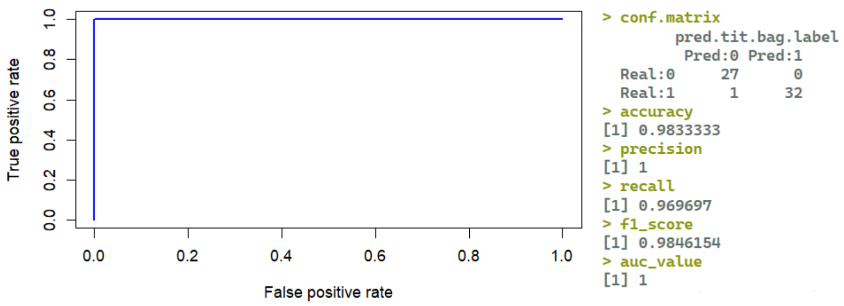


Figure 49: RF out-of-sample performance: ROC curve and metrics. Cross-validation folds: 10. CV variables: 44 (over 88). Trees: 10'000. OOB error: 2.81 per cent. Case: C.

Case D: same as (C), but with a signalling period of 12 months. Using 10-fold cross-validation, we obtained 59 variables as the hyper-parameter. According to Figure 50, 10'000 trees had an in-sample OOB error rate of 5.65 per cent. The RF model shows lower but still high out-of-sample performance, as shown in Figures 51 and Table 23.

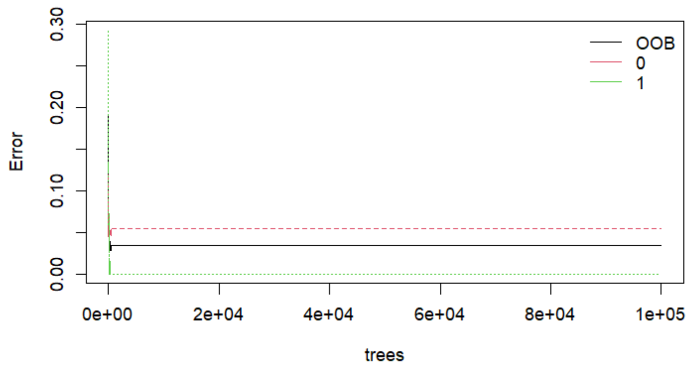


Figure 50: RF FP (1), FN (0) and total OOB errors. Cross-validation folds: 10. Variables selected by CV: 59 (over 88). Trees: 10'000. OOB error: 5.65 per cent. Case: D.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	35	2
	Crisis	1	22

Table 23: Confusion matrix: RF model. Cross-validation folds: 10. Variables selected by CV: 59 (over 88). Case D. Set: test.

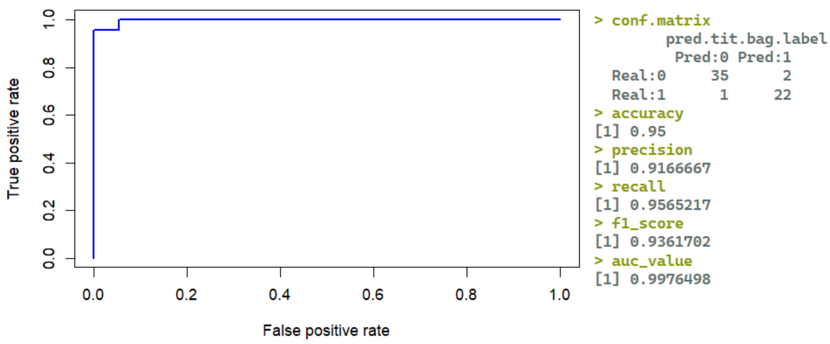


Figure 51: RF out-of-sample performance: ROC curve and metrics. Cross-validation folds: 10. CV variables: 59 (over 88). Trees: 10'000. OOB error: 5.65 per cent. Case: D.

Case E: same as (D), but with a signalling period of one month. Using 10-fold cross-validation, we obtained 6 variables as the hyper-parameter. According to Figure 52, 10'000 trees had an in-sample OOB error rate of 6.21 per cent. The RF model shows high out-of-sample performance, as shown in Figures 53 and Table 24.

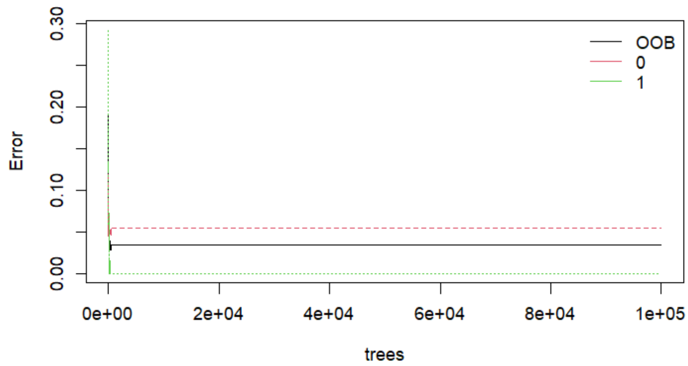


Figure 52: RF FP (1), FN (0) and total OOB errors. Cross-validation folds: 10. Variables selected by CV: 6 (over 88). Trees: 10'000. OOB error: 6.21 per cent. Case: E.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	49	1
	Crisis	1	10

Table 24: Confusion matrix: RF model. Cross-validation folds: 10. Variables selected by CV: 6 (over 88). Case E. Set: test.

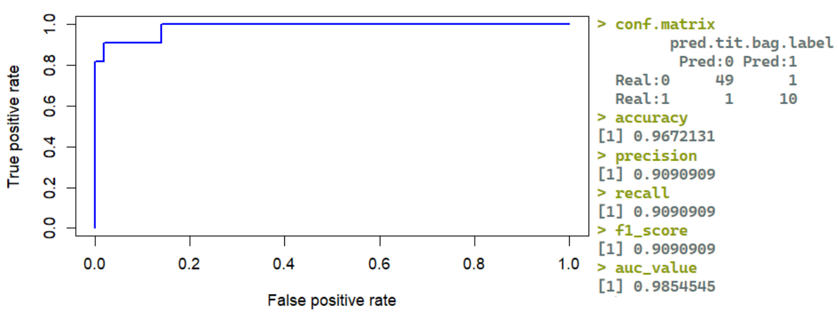


Figure 53: RF out-of-sample performance: ROC curve and metrics. Cross-validation folds: 10. CV variables: 6 (over 88). Trees: 10'000. OOB error: 6.21 per cent. Case: E.

Removal of key features Do RF methods still show good out-of-sample performance after removing the main independent variables? As in previous excerises, the moving average fiscal variables v34, v35, v36, v37, v38, v39, v40 and the accumulated monetary variables v54, v55, v57, v59, v60 are excluded. For scenario B, the cross-validated hyper-parameter for the quantity of variables is 70 out of 76. The RF model had a similar confidence matrix (Table 25) as in case B.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	20	0
	Crisis	0	40

Table 25: Confusion matrix: RF model. Variables: 70 (over 76). Case: B. Set: test.

For Scenario C with 76 variables, the RF model performs similarly to Scenario C with 88 variables. For scenario C, the cross-validated hyper-parameter for the quantity of variables is 6 out of 76. Table 26 shows its out-of-sample performance and Figure 54 shows the ROC curve.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	27	0
	Crisis	1	32

Table 26: Confusion matrix: RF model. Variables: 6 (over 76). Case: C. Set: test.

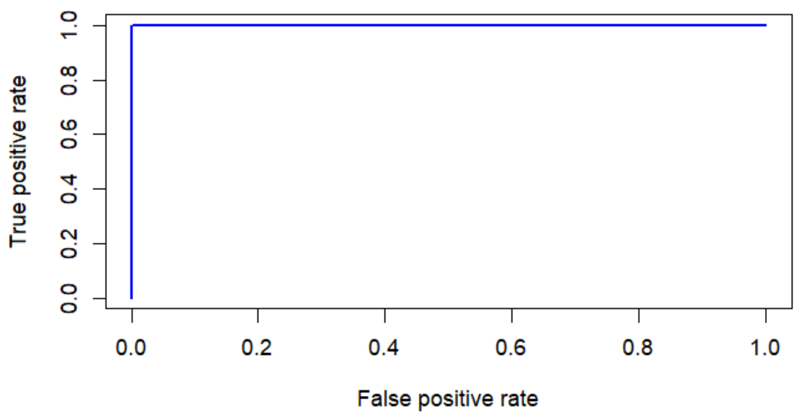


Figure 54: RF out-of-sample performance: ROC curve and metrics. Variables: 6 (over 76). Trees: 10’000. OOB error: 2.25 per cent. Case: C.

For Scenario D with 76 variables, the BA model performs similarly to Scenario C with 88 variables. For scenario D, the cross-validated hyper-parameter for the quantity of variables is 11 out of 76. Table 27 shows its out-of-sample performance, and Figure 55 shows the ROC curve.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	36	1
	Crisis	1	22

Table 27: Confusion matrix: RF model. Variables: 11 (over 76). Case: D. Set: test.

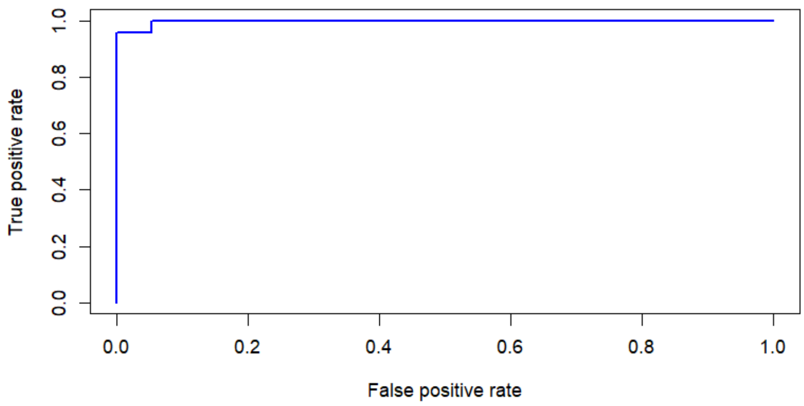


Figure 55: RF out-of-sample performance: ROC curve and metrics. Variables: 11 (over 76). Trees: 10'000. OOB error: 3.37 per cent. Case: D.

For Scenario E with 76 variables, the RF model performs similarly to Scenario E with 88 variables. For scenario E, the cross-validated hyper-parameter for the quantity of variables is 28 out of 76. Table 28 shows its out-of-sample performance, and Figure 56 shows the ROC curve.

		Predicted	
		No-crisis	Crisis
Actual	No-crisis	49	1
	Crisis	1	10

Table 28: Confusion matrix: RF model. Variables: 28 (over 76). Case: E. Set: test.

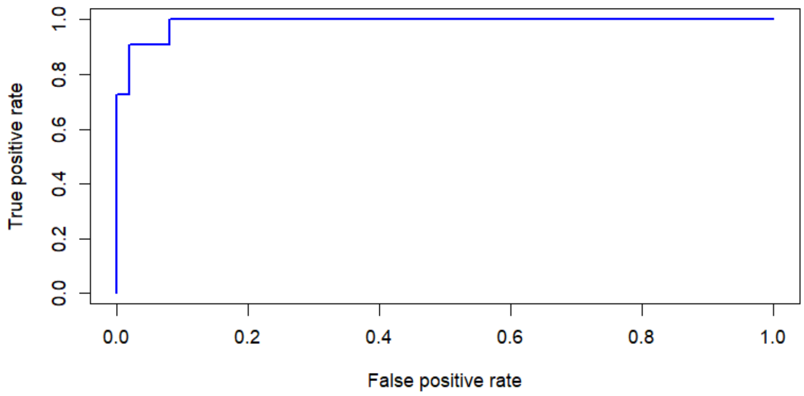


Figure 56: RF out-of-sample performance: ROC curve and metrics. Variables: 28 (over 76). Trees: 10'000. OOB error: 6.78 per cent. Case: D.

6.3.4 External Risk Index

The aggregate risk is tracked by the average external risk index constructed based on the winning models (IMF, 2021). This section briefly presents the external risk index for alternative definitions of crisis and for both BA and RF models. In all cases, 76 variables were considered.

For instance, Figure 57 shows a significant risk increase in May 2010. In April 2012 (24 months later), the parallel exchange rate was more than 20 per cent above the official exchange rate (i.e. the threshold adopted to indicate exchange controls). Following the unification of the exchange rate, the external risk index shows a period of relative calm (also in line with the crisis label). However, a new peak is observed in August 2016. In July 2018 (24 months later), Argentina requested USD 50 billion from the IMF. Figure 58 shows the index constructed with the RF model for the same scenario.

In summary, the external risk index closely tracks crisis labels. Despite the progressively narrower crisis definition and shorter signalling horizon, the external risk index constructed based on BA and RF models closely tracks labelled periods (as shown in Figures 59, 60, 61, 62, 63 and 64).

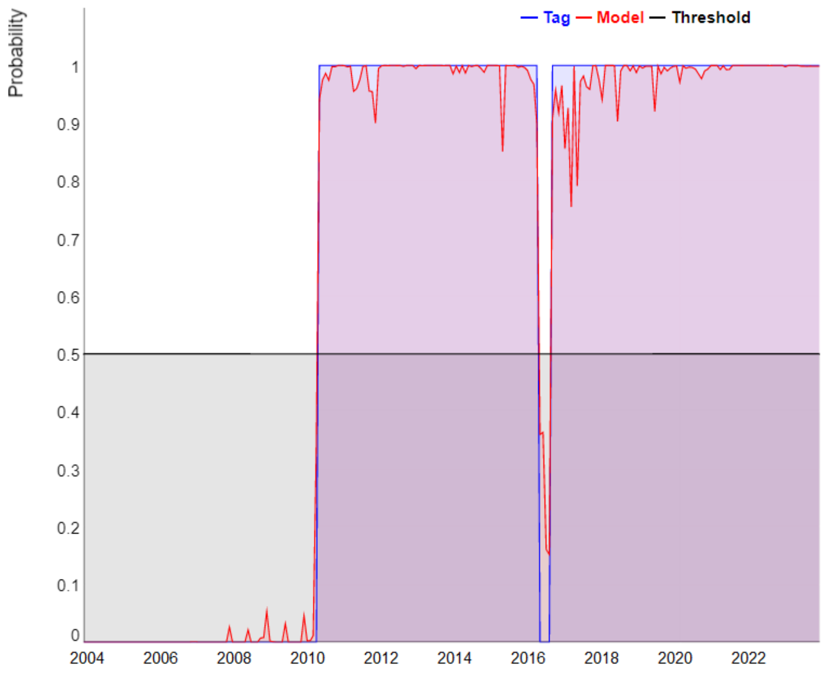


Figure 57: Labels, threshold and External Risk Index - BA model (trained with 76 variables). Scenario B.

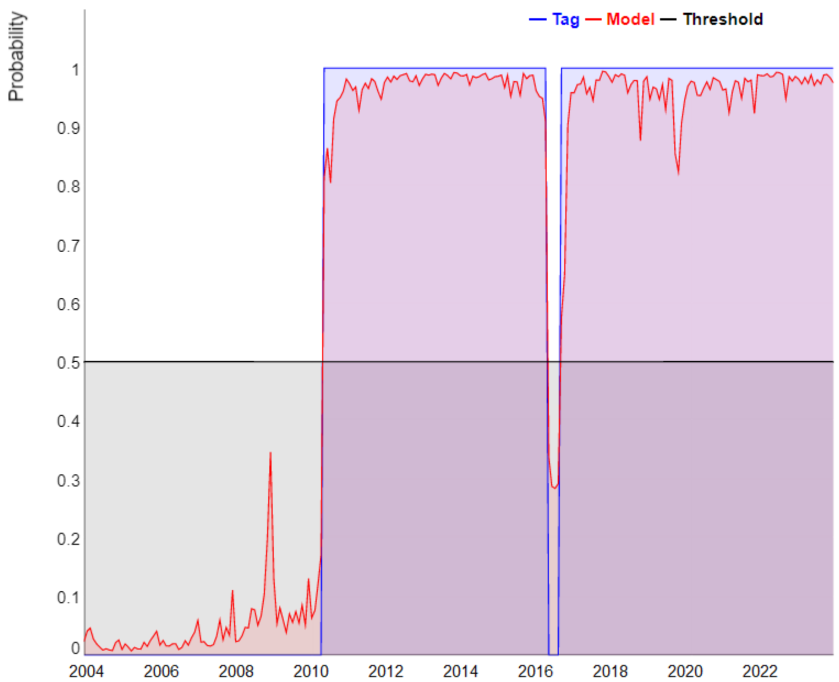


Figure 58: Labels, threshold and External Risk Index - RF model (trained with 76 variables). Scenario B.

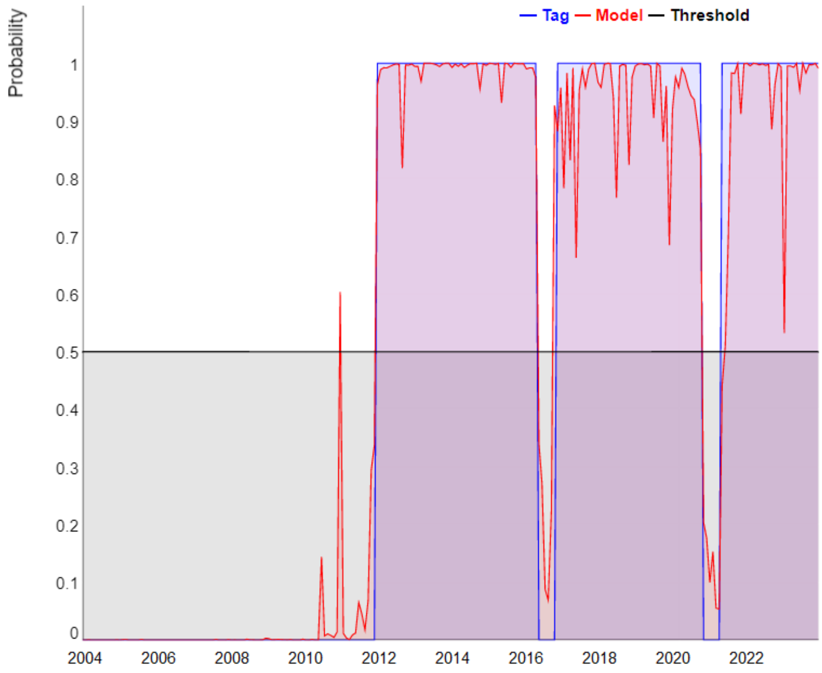


Figure 59: Labels, threshold and External Risk Index - BA model (trained with 76 variables). Scenario C.

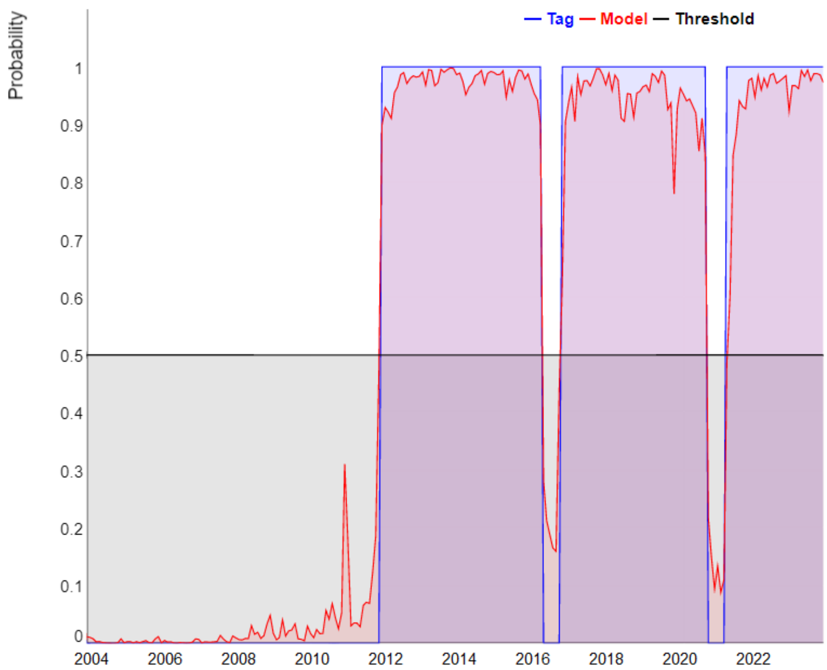


Figure 60: Labels, threshold and External Risk Index - RF model (trained with 76 variables). Scenario C.

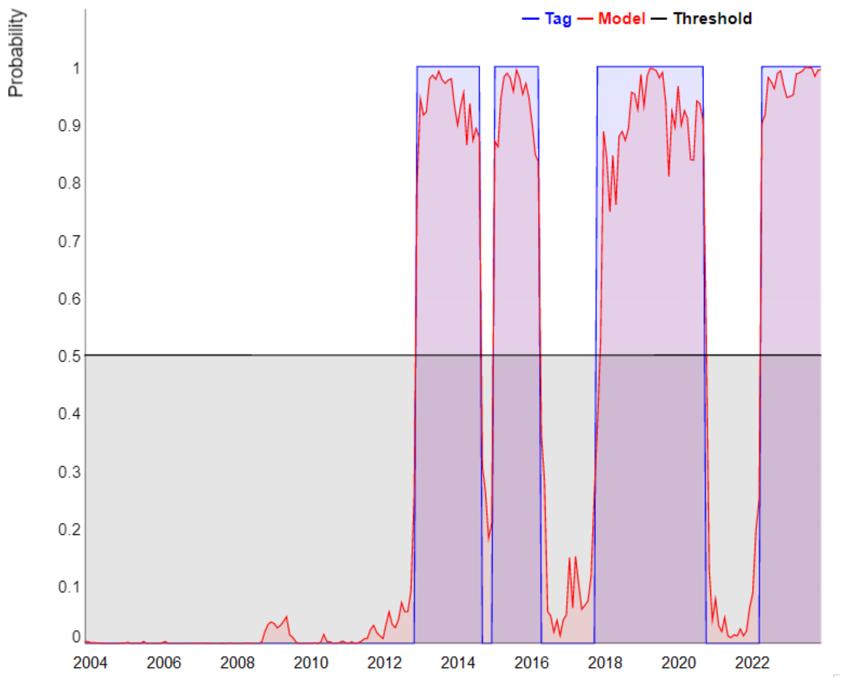


Figure 61: Labels, threshold and External Risk Index - BA model (trained with 76 variables). Scenario D.

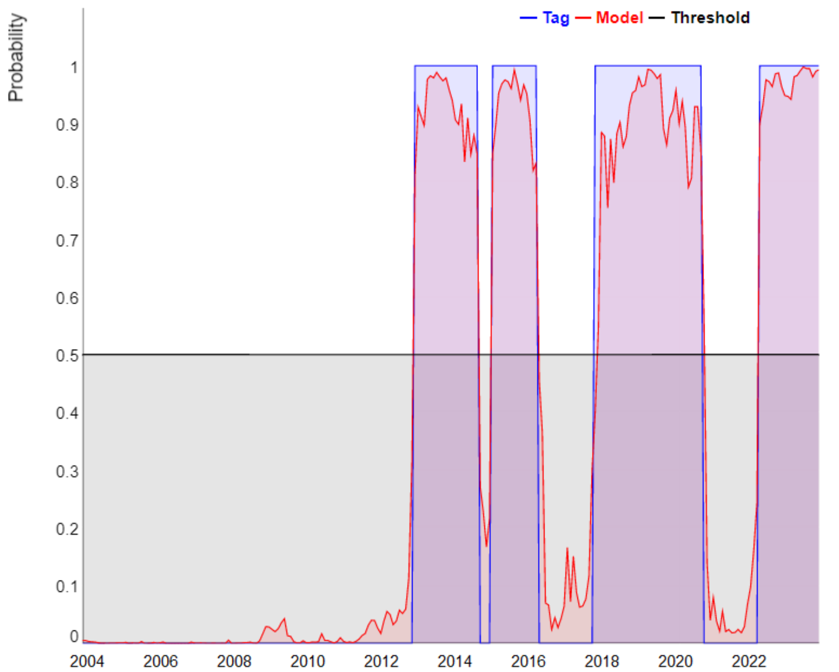


Figure 62: Labels, threshold and External Risk Index - RF model (trained with 76 variables). Scenario D.

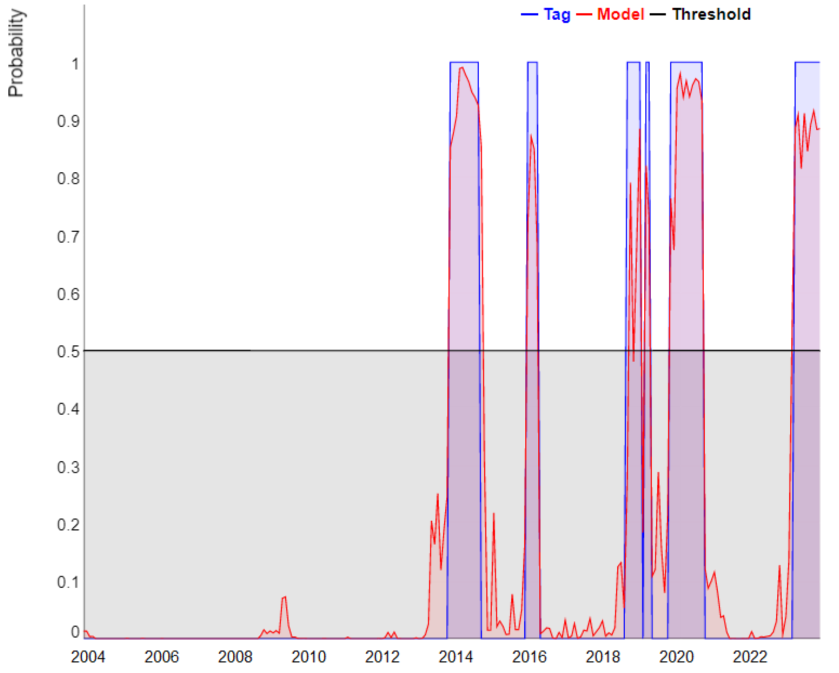


Figure 63: Labels, threshold and External Risk Index - BA model (trained with 76 variables). Scenario E.

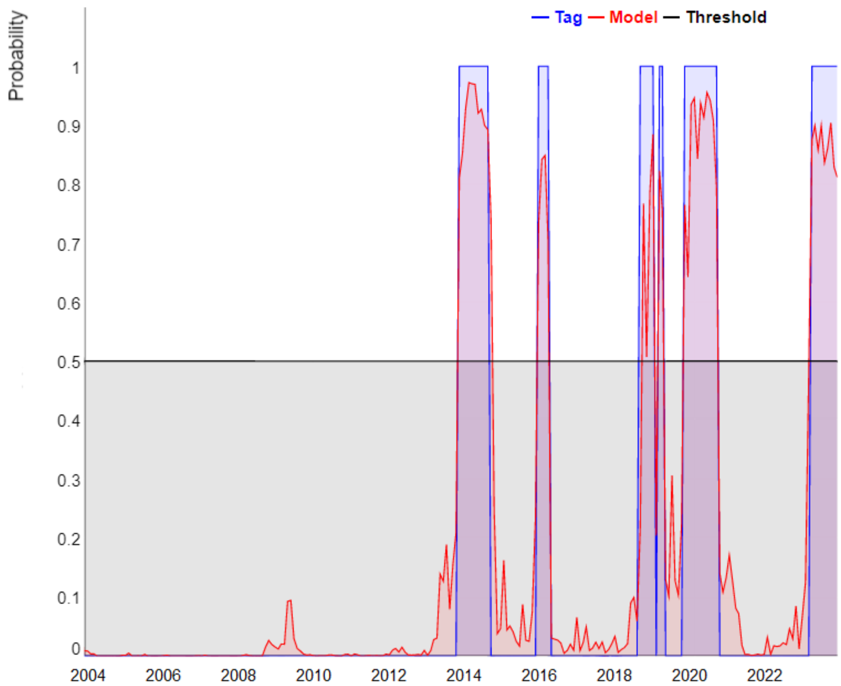


Figure 64: Labels, threshold and External Risk Index - RF model (trained with 76 variables). Scenario E.

6.4 Empirical findings: research question II

The second question was which tree-based method would best predict Argentina's external crises using data from 2003 to 2023. The CART, BA, and RF models performed similarly for long signalling periods. However, for short signalling periods, the 'wisdom of crowds' worked well, with the BA and RF models outperforming the CART model. Finally, the BA and RF models performed similarly. RF outperformed BA for short signalling periods.

Case A: including exchange rate controls, SBA and a 24-month signalling period. Figure 65 presents a comparison of the out-of-sample performance of CART, BA, and RF models. All three models achieved 100 per cent accuracy, recall, precision, and F1 score. The ROC curve was ideal, and the AUC was 1.

		TP	FN	TN	FP	AR	PR	RR	F1	AUC
CART Information (cv=5 folds; alfa=0.01666667, leafs=5)	CART1	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
CART Information (cv=10 folds; alfa=0.025; leafs=2)	CART2	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
CART Entropy (cv=5 folds; alfa=0.03333333; leafs=2)	CART3	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
CART Entropy (cv=10 folds, alfa=0.03333333; leafs=2)	CART4	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
CART Gini (cv=5 folds, alfa=0.03333333; leafs=2)	CART5	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
CART Gini (cv=10 folds, alfa=0.03333333; leafs=2)	CART6	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
BA Accuracy (#variables= 103, #trees=10'000, OOB error=2.81%)	BA1	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
BA Gini (#variables= 103, #trees=10'000, OOB error=2.81%)	BA2	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
RF Accuracy (cv 5 folds #variables=53, #trees=10'000, OOB error=3.37%)	RF1	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
RF Gini (cv 5 folds #variables=53, #trees=10'000, OOB error=3.37%)	RF2	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
RF Accuracy (cv 10 folds #variables=52, #trees=100'000, OOB error=3.37%)	RF3	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
RF Gini (cv 10 folds #variables=52, #trees=100'000, OOB error=3.37%)	RF4	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
RF Accuracy (heuristic #variables=10, #trees=10'000, OOB error=2.81%)	RF5	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
RF Gini (heuristic #variables=10, #trees=10'000, OOB error=2.81%)	RF6	40	0	20	0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000

Figure 65: Out-of-sample performance table. Features: 103 (including nominal variables). Case: A. Threshold: 0.5.

Case B: same as (A), but only 88 variables were considered, excluding nominal variables. Figure 66 compares the out-of-sample performance of CART, BA, and RF models in scenario B. All three models achieved 100 per cent accuracy, recall, precision, and F1 score. The ROC curve was ideal, and the AUC was 1, except for the no cross-validated (heuristic) RF model.

		TP	FN	TN	FP	AR	PR	RR	F1	AUC
CART Information (cv=5 folds; alfa=0.025, leafs=2)	CART1	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
CART Information (cv=10 folds; alfa=0.025; leafs=2)	CART2	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
CART Entropy (cv=5 folds; alfa=0.03333333; leafs=2)	CART3	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
CART Entropy (cv=10 folds, alfa=0.03333333; leafs=2)	CART4	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
CART Gini (cv=5 folds, alfa=0.03333333; leafs=2)	CART5	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
CART Gini (cv=10 folds, alfa=0.03333333; leafs=2)	CART6	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
BA Accuracy (#variables= 88, #trees=10'000, OOB error=2.25%)	BA1	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
BA Gini (#variables= 88, #trees=10'000, OOB error=2.25%)	BA2	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
RF Accuracy (cv 5 folds #variables=72, #trees=10'000, OOB error=2.25%)	RF1	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
RF Gini (cv 5 folds #variables=72, #trees=10'000, OOB error=2.25%)	RF2	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
RF Accuracy (cv 10 folds #variables=78, #trees=10'000, OOB error=1,69%)	RF3	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
RF Gini (cv 10 folds #variables=78, #trees=10'000, OOB error=1,69%)	RF4	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
RF Accuracy (heuristic #variables=9, #trees=10'000, OOB error=2.25%)	RF5	39	1	20	0	0,983333	1,000000	0,975000	0,987342	0,990000
RF Gini (heuristic #variables=9, #trees=10'000, OOB error=2.25%)	RF6	39	1	20	0	0,983333	1,000000	0,975000	0,987342	0,990000

Figure 66: Out-of-sample performance table. Features: 88 (excluding nominal variables). Case: B. Threshold: 0.5.

Case C: same as (B), but without considering exchange rate controls and SBA. Figure 67 shows that when exchange rate controls and the SBA are removed, performance remains high but is lower than in cases A and B for all models. In addition, the 'wisdom of crowds' indicates that the BA and RF models have higher out-of-sample performance than the CART models. A lower correlation between the trees of the RF models positively impacts the overall performance. However, this improvement is marginal compared to the BA models.

		TP	FN	TN	FP	AR	PR	RR	F1	AUC
CART Information (cv=5 folds; alfa=1e-04; leafs=4)	CART1	31	2	27	0	0,966667	1,000000	0,939394	0,968750	0,969697
CART Information (cv=10 folds; alfa=1e-04; leafs=4)	CART2	31	2	27	0	0,966667	1,000000	0,939394	0,968750	0,969697
CART Entropy (cv=5 folds; alfa=0.025; leafs=3)	CART3	32	1	26	1	0,966667	0,969697	0,969697	0,969697	0,966330
CART Entropy (cv=10 folds, alfa=0.025; leafs=3)	CART4	32	1	26	1	0,966667	0,969697	0,969697	0,969697	0,966330
CART Gini (cv=5 folds, alfa=0.025; leafs=3)	CART5	32	1	26	1	0,966667	0,969697	0,969697	0,969697	0,966330
CART Gini (cv=10 folds, alfa=0.025; leafs=3)	CART6	32	1	26	1	0,966667	0,969697	0,969697	0,969697	0,966330
BA Accuracy (#variables= 88, #trees=1'000'000, OOB error=3.37%)	BA1	33	0	26	1	0,983333	0,970588	1,000000	0,985075	0,992100
BA Gini (#variables= 88, #trees=1'000'000, OOB error=3.37%)	BA2	33	0	26	1	0,983333	0,970588	1,000000	0,985075	0,992100
RF Accuracy (cv 5 folds #variables=2, #trees=10'000, OOB error=1.69%)	RF1	32	1	27	0	0,983333	1,000000	0,969697	0,984615	1,000000
RF Gini (cv 5 folds #variables=2, #trees=10'000, OOB error=1.69%)	RF2	32	1	27	0	0,983333	1,000000	0,969697	0,984615	1,000000
RF Accuracy (cv 10 folds #variables=5, #trees=10'000, OOB error=1.69%)	RF3	32	1	27	0	0,983333	1,000000	0,969697	0,984615	1,000000
RF Gini (cv 10 folds #variables=5, #trees=10'000, OOB error=1.69%)	RF4	32	1	27	0	0,983333	1,000000	0,969697	0,984615	1,000000
RF Accuracy (heuristic #variables=9, #trees=10'000, OOB error=1.12%)	RF5	32	1	27	0	0,983333	1,000000	0,969697	0,984615	1,000000
RF Gini (heuristic #variables=9, #trees=10'000, OOB error=1.12%)	RF6	32	1	27	0	0,983333	1,000000	0,969697	0,984615	1,000000

Figure 67: Out-of-sample performance table. Features: 88 (excluding nominal features). Case: C. Threshold: 0.5.

Case D: same as (C), but with a signalling period of 12 months. Figure 68 illustrates that shortening the signalling period to 12 months results in high performance, albeit lower than in cases A and B, for all models. Furthermore, according to the 'wisdom of crowds', the BA and RF models exhibit higher out-of-sample performance than the CART models. The lower correlation between the trees in the RF models does not positively impact overall performance.

		TP	FN	TN	FP	AR	PR	RR	F1	AUC
CART Information (cv=5 folds; alfa=0.01470588; leafs=6)	CART1	23	0	35	2	0,966667	0,920000	1,000000	0,958333	0,994125
CART Information (cv=10 folds; alfa=1e-04; leafs=8)	CART2	23	0	36	1	0,983333	0,958333	1,000000	0,978723	0,986487
CART Entropy (cv=5 folds; alfa=1e-04; leafs=9)	CART3	23	0	35	2	0,966667	0,920000	1,000000	0,958333	0,972973
CART Entropy (cv=10 folds, alfa=0.01470588; leafs=7)	CART4	23	0	35	2	0,983333	0,958333	1,000000	0,978723	0,984136
CART Gini (cv=5 folds, alfa=1e-04; leafs=9)	CART5	23	0	35	2	0,966667	0,920000	1,000000	0,958333	0,972973
CART Gini (cv=10 folds, alfa=0.01470588; leafs=7)	CART6	23	0	36	1	0,983333	0,958333	1,000000	0,978723	0,984136
BA Accuracy (#variables= 88, #trees=1'000'000, OOB error=3.37%)	BA1	23	0	37	0	1,000000	1,000000	1,000000	1,000000	1,000000
BA Gini (#variables= 88, #trees=1'000'000, OOB error=3.37%)	BA2	23	0	37	0	1,000000	1,000000	1,000000	1,000000	1,000000
RF Accuracy (cv 5 folds #variables=55, #trees=100'000, OOB error=2.81%)	RF1	23	0	37	0	1,000000	1,000000	1,000000	1,000000	1,000000
RF Gini (cv 5 folds #variables=55, #trees=100'000, OOB error=2.25%)	RF2	23	0	37	0	1,000000	1,000000	1,000000	1,000000	1,000000
RF Accuracy (cv 10 folds #variables=55, #trees=10'000, OOB error=2.25%)	RF3	23	0	37	0	1,000000	1,000000	1,000000	1,000000	1,000000
RF Gini (cv 10 folds #variables=55, #trees=10'000, OOB error=2.25%)	RF4	23	0	37	0	1,000000	1,000000	1,000000	1,000000	1,000000
RF Accuracy (heuristic #variables=9, #trees=10'000, OOB error=2.25%)	RF5	22	1	35	2	0,950000	0,916667	0,956522	0,936170	0,997600
RF Gini (heuristic #variables=9, #trees=10'000, OOB error=2.25%)	RF6	22	1	35	2	0,950000	0,916667	0,956522	0,936170	0,997600

Figure 68: Out-of-sample performance table. Features: 88 (excluding nominal features). Case: D. Threshold: 0.5.

Case E: same as (D), but with a signalling period of one month. Figure 69 shows that when the signalling period is reduced to 1 month, performance remains high but lower than in all previous cases. The 'wisdom of crowds' still applies, as BA and RF models have higher overall performance than CART models. However, there is no evidence of greater performance of RF models compared to BA models due to lower tree correlation.

		TP	FN	TN	FP	AR	PR	RR	F1	AUC
CART Information (cv=5 folds; alfa=0.1666667; leafs=2)	CART1	9	2	47	3	0,918033	0,750000	0,818182	0,782609	0,865455
CART Information (cv=10 folds; alfa=1e-04; leafs=9)	CART2	9	2	49	1	0,950820	0,900000	0,818182	0,857143	0,862727
CART Entropy (cv=5 folds; alfa=0.03333333; leafs=5)	CART3	10	1	45	5	0,901639	0,666667	0,909091	0,769231	0,904546
CART Entropy (cv=10 folds, alfa=1e-04; leafs=9)	CART4	9	2	49	1	0,950820	0,900000	0,818182	0,857143	0,899091
CART Gini (cv=5 folds, alfa=0.03333333; leafs=2)	CART5	8	3	46	4	0,885246	0,666667	0,727273	0,695652	0,823636
CART Gini (cv=10 folds, alfa=1e-04; leafs=5)	CART6	9	2	49	1	0,950820	0,900000	0,818182	0,857143	0,899091
BA Accuracy (#variables= 88, #trees=10'000, OOB error=6.78%)	BA1	10	1	49	1	0,967213	0,909091	0,909091	0,909091	0,983600
BA Gini (#variables= 88, #trees=10'000, OOB error=6.78%)	BA2	10	1	49	1	0,967213	0,909091	0,909091	0,909091	0,983600
RF Accuracy (cv 5 folds #variables=18, #trees=10'000, OOB error=6.78%)	RF1	10	1	49	1	0,967213	0,909091	0,909091	0,909091	0,985455
RF Gini (cv 5 folds #variables=18, #trees=10'000, OOB error=6.78%)	RF2	10	1	49	1	0,967213	0,909091	0,909091	0,909091	0,985455
RF Accuracy (cv 10 folds #variables=6, #trees=10'000, OOB error=6.21%)	RF3	10	1	49	1	0,967213	0,909091	0,909091	0,909091	0,992727
RF Gini (cv 10 folds #variables=6, #trees=10'000, OOB error=6.21%)	RF4	10	1	49	1	0,967213	0,909091	0,909091	0,909091	0,992727
RF Accuracy (heuristic #variables=9, #trees=10'000, OOB error=5.65%)	RF5	10	1	49	1	0,967213	0,909091	0,909091	0,909091	0,992727
RF Gini (heuristic #variables=9, #trees=10'000, OOB error=5.65%)	RF6	10	1	49	1	0,967213	0,909091	0,909091	0,909091	0,992727

Figure 69: Out-of-sample performance table. Features: 88 (excluding nominal features). Case: E. Threshold: 0.5.

Removal of key features It is worth noting that the main conclusion remains when the moving average fiscal variables v34, v35, v36, v37, v38, v39, v40 and the accumulated monetary variables v54, v55, v57, v59, v60 are excluded (Figure 70). The CART, BA and RF models performed similarly for long reporting periods. However, for short signalling periods, the wisdom of crowds worked well, with the BA and RF models outperforming the CART model. Due to lower tree correlation, RF outperformed BA.

Scenario	Model	TP	FN	TN	FP	AR	PR	RR	F1	AUC
B	CART	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
B	BA	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
B	RF	40	0	20	0	1,000000	1,000000	1,000000	1,000000	1,000000
C	CART	31	2	25	2	0,933333	0,939394	0,939394	0,939394	0,932660
C	BA	32	1	25	2	0,950000	0,941177	0,969697	0,955224	0,988777
C	RF	32	1	27	0	0,983333	1,000000	0,969697	0,984615	1,000000
D	CART	21	2	36	1	0,950000	0,954546	0,913044	0,933333	0,943008
D	BA	22	1	36	1	0,966667	0,956522	0,956522	0,956522	0,997650
D	RF	22	1	36	1	0,966667	0,956522	0,956522	0,956522	0,997650
E	CART	9	2	48	2	0,934426	0,818182	0,818182	0,818182	0,987273
E	BA	9	2	48	2	0,934426	0,818182	0,818182	0,818182	0,987273
E	RF	10	1	49	1	0,967213	0,909091	0,909091	0,909091	0,989091

Figure 70: Out-of-sample performance table. Features: 76 (excluding nominal features, moving average fiscal variables and accumulated monetary variables). Crisis definition: cases B, C, D and E. Threshold: 0.5.

6.5 Empirical findings: research question III

The third question aimed to identify the main instrumental variables that policymakers should monitor in order to be aware of potential external crises in Argentina. One hundred and three macroeconomic variables were used, taking into account the balance of payments crisis theories. For different signalling horizons and different features included, the models highlighted the alternative importance of fiscal, monetary, real and external variables as potential predictors of currency crises. [51]

Section 1.2 specifies that this research is based on empirical evidence only and does not discuss theory. Machine learning models are designed to solve supervised classification problems and cannot resolve theoretical disputes. The identifying variables are only relevant as instrumental variables to signal currency crises.

Case A: including exchange rate controls, SBA and a 24-month signalling period. Research suggests that fiscal and external variables can be significant predictors of crises, especially in situations with longer-term signalling periods and more crisis observations. In scenario A, the CART model (Figure 71) have identified v34 (Figure 74) as a significant variable with a monthly threshold of USD 5564 million. Additionally,

[51] It is important to note that the IMF's RF model for emerging countries (IMF, 2021) has identified a set of key variables that are important in signalling a crisis (Figure 6). The first five variables are reserve growth, five-year inflation, public debt growth, inflation and export growth. For the case of Argentina, this research presents different variables depending on the crisis and signalling period scenarios and the features included.

the BA (Figure 72) and RF models (Figure 73) have identified v36 (Figure 75) and v37 (Figure 76) as significant variables according to the mean decrease accuracy and mean decrease Gini criteria [52]

It should be noted that while the variables mentioned have fiscal implications, they are not solely fiscal in nature. Variables such as v34 (total expenditures in USD), v36 (subsidies to the private sector in USD), and v37 (social benefits in USD) are measured using the official exchange rate A3500, meaning that exchange rate policies can significantly impact these variables. Higher levels of fiscal expenditure and exchange rate appreciation increase the probability of signalling a crisis. Furthermore, the external sector variable v92 (see Figure 78) is significant.

In scenario A, BA and RF models also prioritise specific real variables, such as the Consumer Price Index (v75) and Monthly Economic Activity Estimator (v77 - Figure 77), which may indicate an impending crisis. Considering all factors, if there are high levels of expenses in USD terms (e.g. due to expansive fiscal policy and exchange rate appreciation), coupled with high levels of economic activity and inflation, it may indicate an impending crisis within a 24-month time-frame.

[52] The importance of a variable in the model increases with the value of mean decrease accuracy or mean decrease Gini score. The Mean Decrease Accuracy expresses how much accuracy the model losses by excluding each variable. The more the accuracy decreases, the more important the variable is for successful classification. The mean decrease in Gini coefficient measures how each variable contributes to the homogeneity of the nodes and leaves in the resulting random forest. Whenever a node is split on variable X, the Gini impurity criterion for the two descendant nodes is lower than that of the parent node. Calculating the average Gini decreases across all trees in the forest provides a measure of variable importance based on the Gini criteria (Breiman and Cutler, 2001). Breiman proposed to evaluate the importance of a variable X_m by adding up the weighted impurity decreases $p(t)\Delta i(s_t, t)$ for all nodes t averaged over all N_T trees in the forest (Louppe et al., 2013).

$$\text{Importance}(X_m) = \frac{1}{N_T} \sum_{T} \sum_{t \in T: v(s_t) = X_m} p(t)\Delta i(s_t, t)$$

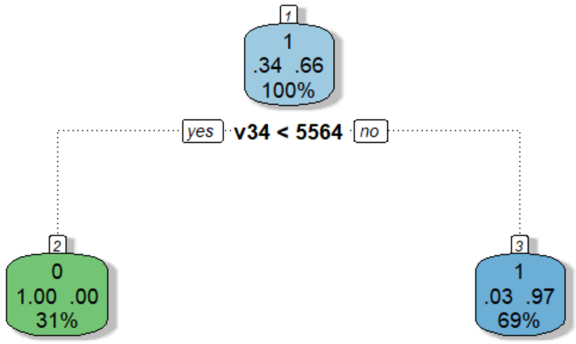


Figure 71: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 1. Variables: 103. Case: A.

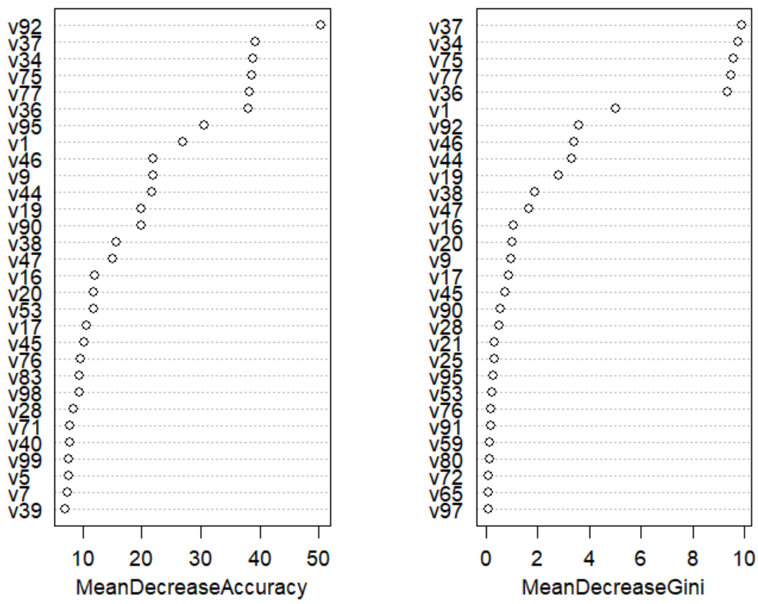


Figure 72: BA variable importance: Accuracy and Gini criteria. Variables: 103 (over 103). Trees: 10'000. OOB error: 2.81 per cent. Case: A.

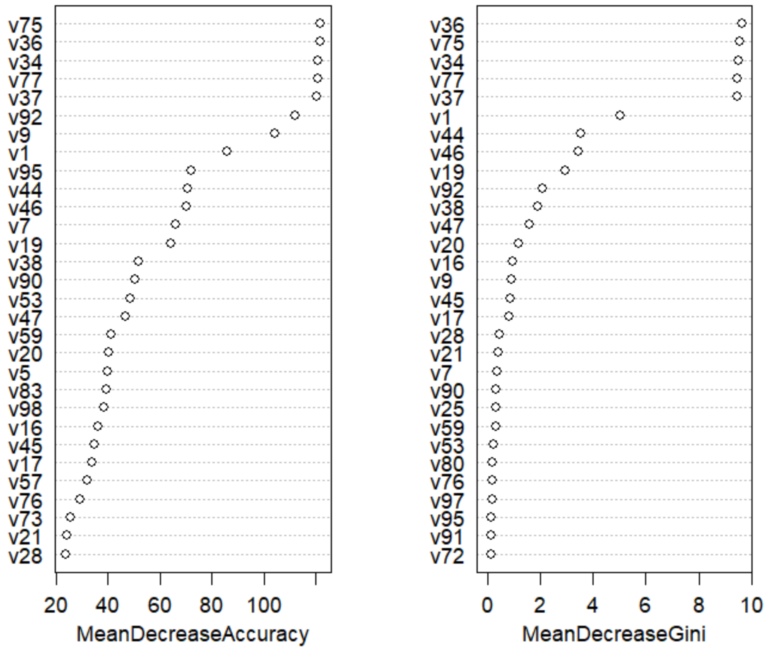


Figure 73: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 52 (over 103). Trees: 100'000. OOB error: 3.37 per cent. Case: A.

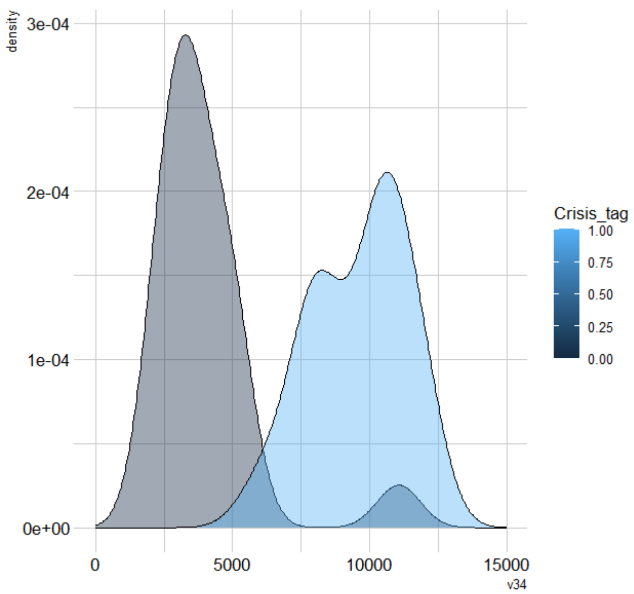


Figure 74: Non-parametric adjusted distributions of Expenditures (v34) in official USD millions for NFPS under crisis and no-crisis labels in scenario A. Source: own based on BCRA and MEcon.

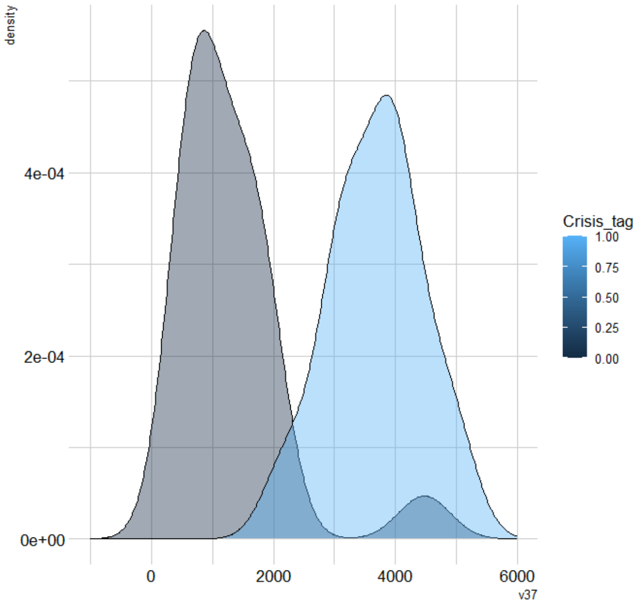


Figure 76: Non-parametric adjusted distributions of social benefits (v37) USD millions for NFPS under crisis and no-crisis labels in scenario A. Source: own based on BCRA and MEcon.

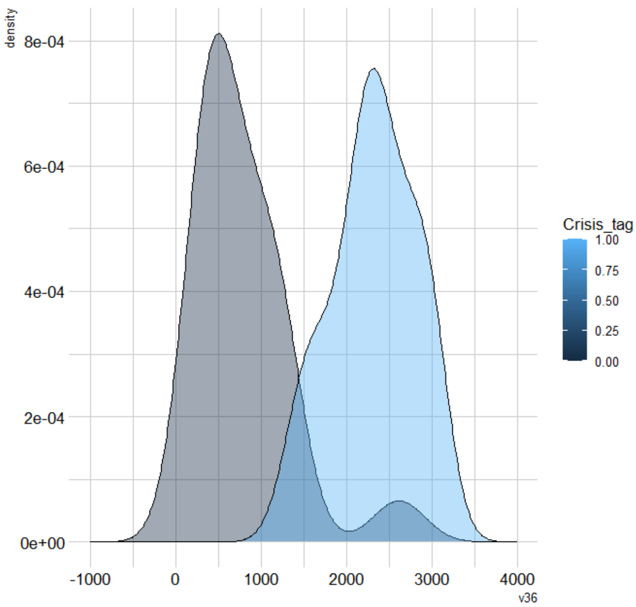


Figure 75: Non-parametric adjusted distributions of public subsidies to the private sector (v36) USD millions for NFPS under crisis and no-crisis labels in scenario A. Source: own based on BCRA and MEcon.

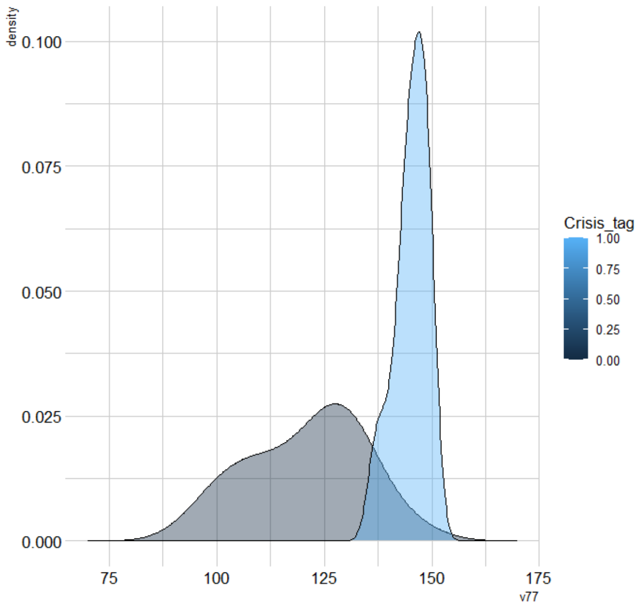


Figure 77: Non-parametric adjusted distributions of Monthly Economic Activity Estimator, EMAE index (v77) under crisis and no-crisis labels in scenario A. Source: own based on INDEC.

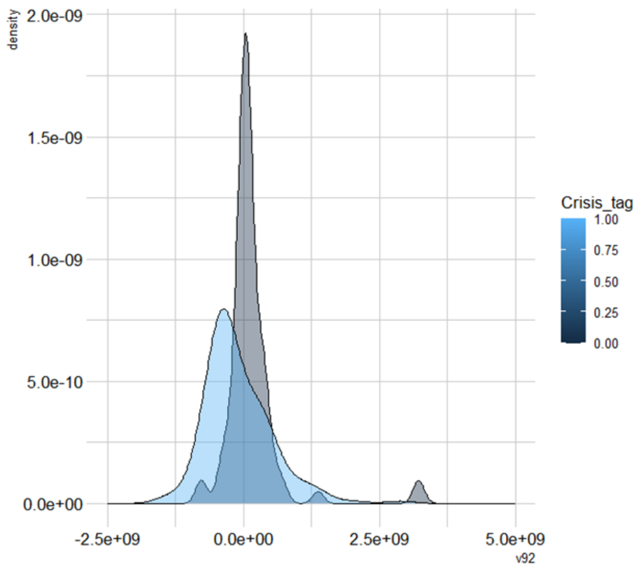


Figure 78: Non-parametric adjusted distributions of Financial loans, debt securities and lines of credit at the exchange market (v92) in USD under crisis and no-crisis labels in scenario A. Source: own based on BCRA.

Case B: same as (A), but only 88 variables were considered, excluding nominal variables. When excluding nominal variables in scenario B, the CART (Figure 71), BA (Figure 79), and RF (Figure 80) models still consider 'fiscal-external' variables as relevant. However, there are two main differences worth noting. Firstly, the real variable Monthly Economic Activity Estimator (v77) gains explanatory power. Secondly, v38 (Figure 81) also gains explanatory power. As with case A, high levels of expenses in USD terms due to expansive fiscal policy and exchange rate appreciation, coupled with high levels of economic activity, may indicate an impending crisis within a 24-month time-frame.

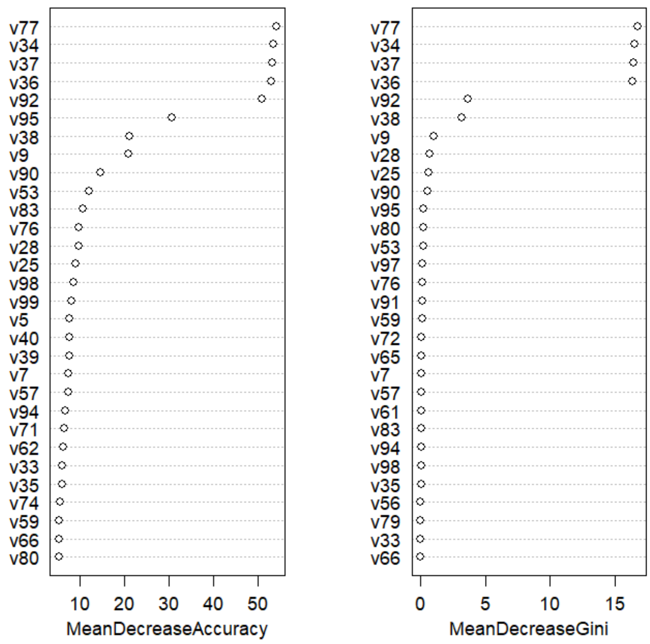


Figure 79: BA variable importance: Accuracy and Gini criteria. Variables: 88 (over 88). Trees: 10'000. OOB error: 2.25 per cent. Case: B.

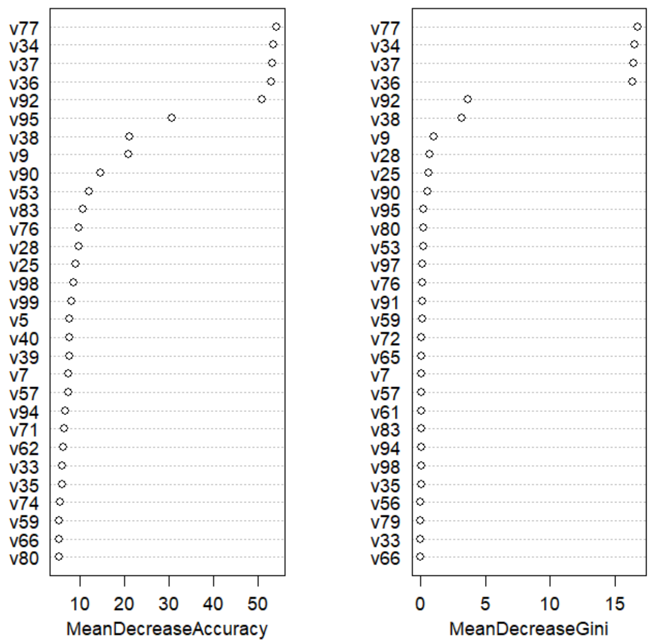


Figure 80: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 78 (over 88). Trees: 10'000. OOB error: 1.69 per cent. Case: B.

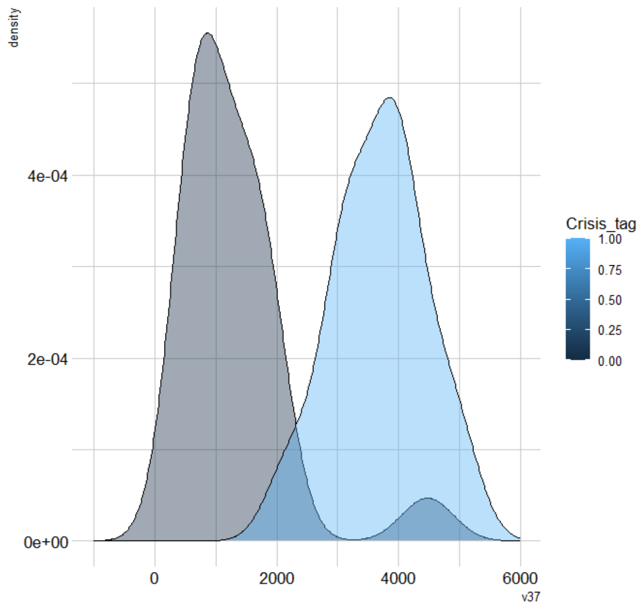


Figure 81: Non-parametric adjusted distributions of Compensation to Employees (v38) in official USD millions for NFPS under crisis and no-crisis labels in scenario B. Source: own based on BCRA and MEcon.

It is worth noting that variables v77 (Figure 77) and v92 (Figure 78) are the most important variables to follow when excluding the moving average fiscal variables v34, v35, v36, v37, v38, v39, v40 and the accumulated monetary variables v54, v55, v57, v59, v60. In particular, the 'fiscal' variables lose importance when only the remaining 76 characteristics are considered. Graph 82 shows the main variables for the BA model, and Graph 83 shows the main variables for the RF model when these variables are excluded.

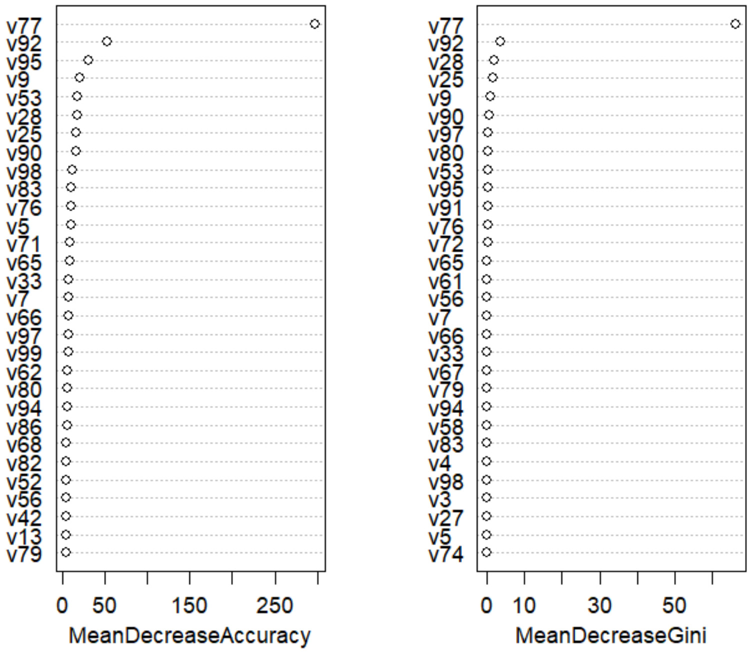


Figure 82: BA variable importance: Accuracy and Gini criterias. Variables: 76 (over 76). Trees: 10'000. OOB error: 2.25 per cent. Case: B.

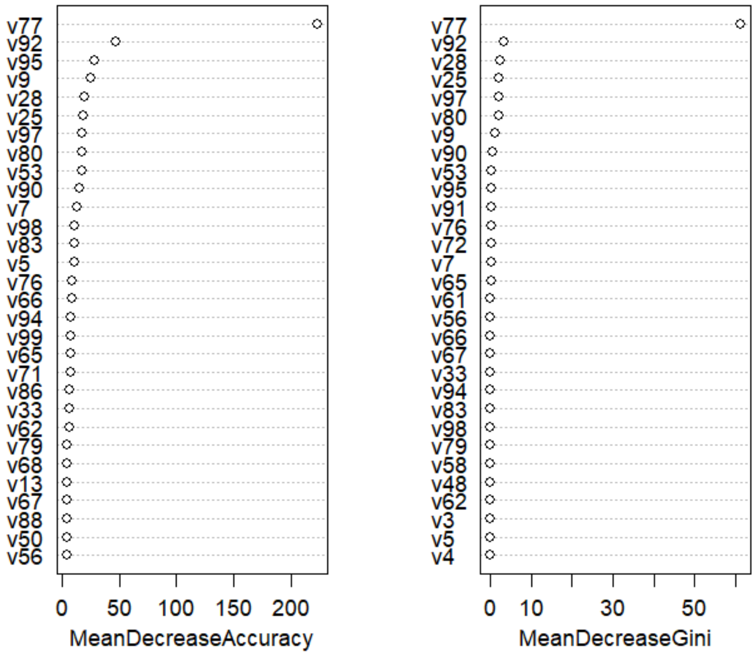


Figure 83: RF variable importance: Accuracy and Gini criterias. Cross-validation folds: 10. CV variables: 70 (over 76). Trees: 10'000. OOB error: 2.25 per cent. Case: B.

Case C: same as (B), but without considering exchange rate controls and SBA. Considering 88 features, 'fiscal-external' and real variables remain relevant if we exclude exchange rate controls and the SBA from the crisis definition and maintain the 24-month signalling (scenario C). This is true for CART (Figure 84), BA (Figure 85) and RF (Figure 86) models.

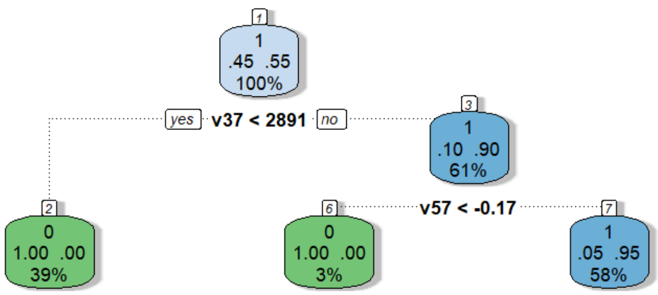


Figure 84: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.025. Selected number of splits: 2. Variables: 88. Case C.

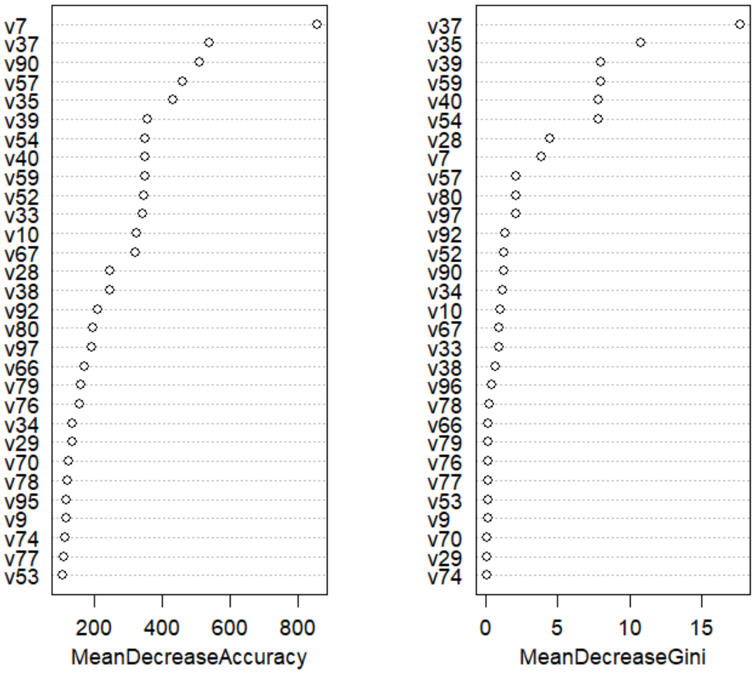


Figure 85: BA variable importance: Accuracy and Gini criteria. Variables: 88 (over 88). Trees: 1'000'000. OOB error: 3.37 per cent. Case: C.

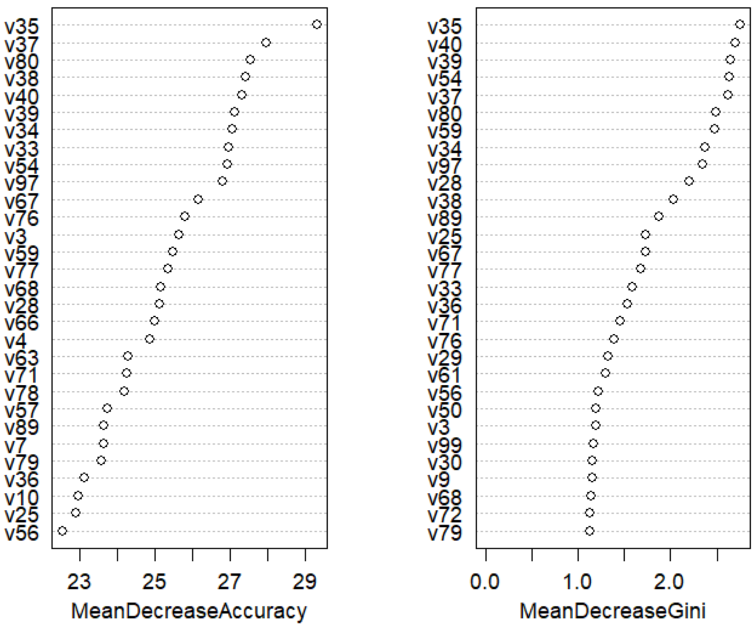


Figure 86: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 5 (over 88). Trees: 10'000. OOB error: 1.69 per cent. Case: C.

However, there have been changes in the predictive power of variables. Firstly, real variable v77 (Monthly Economic Activity Estimator EMAE) has lost its predictive power, while monetary-external variables v54 (Figure 87), v59 (Figure 88), and v57 (Figure 89) have gained predictive power. The Central Bank's creation of monetary base to finance the Treasury and destruction of monetary base by selling foreign exchange to private entities may be a clue of an incoming crisis.

In other words, when exchange rate controls are considered, the Monthly Economic Activity Estimator (v77) becomes a more useful variable in explaining crises. However, it loses its explanatory power when exchange rate controls are not considered. On the other hand, monetary-external and fiscal-external variables show the opposite pattern.

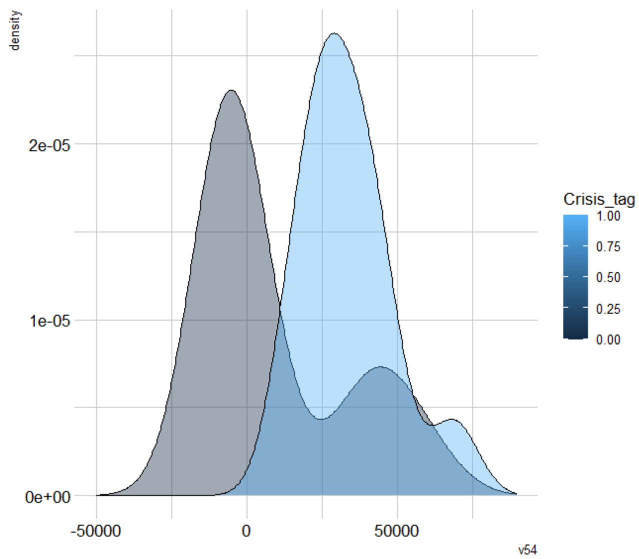


Figure 87: Non-parametric adjusted distributions of accumulated Monetary Base expansion due to net foreign exchange purchases to the National Treasury (v54) in official USD millions for NFPS under crisis and no-crisis labels in scenario C. Source: own based on BCRA.

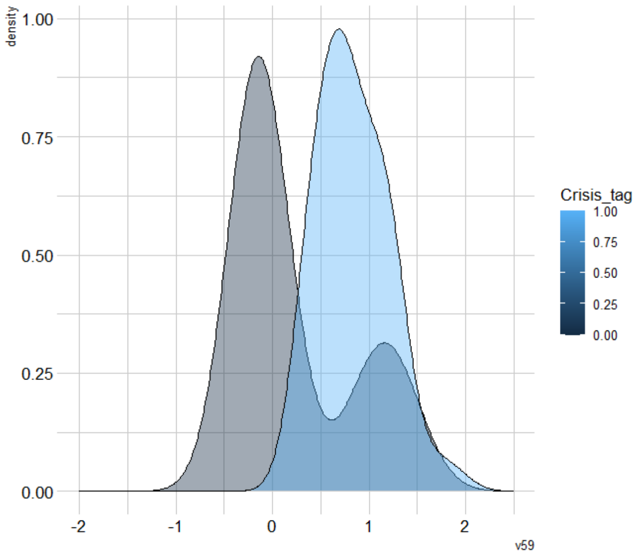


Figure 88: Non-parametric adjusted distributions of accumulated Monetary Base expansion due to net foreign exchange purchases to the National Treasury as measured by the stock of International Reserves (v59) under crisis and no-crisis labels in scenario C. Source: own based on BCRA.

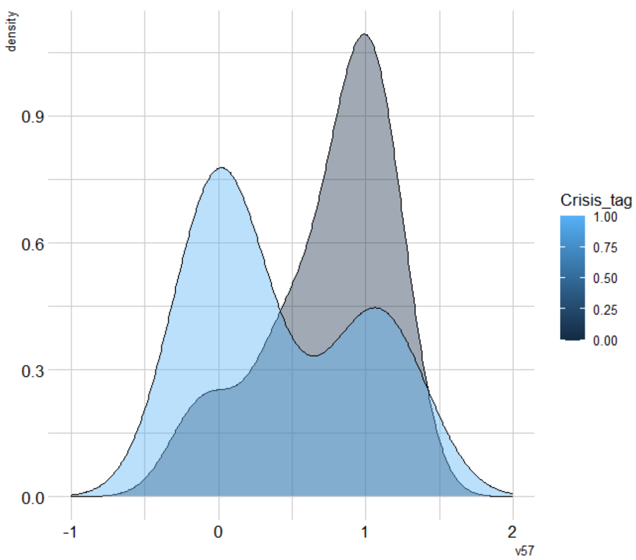


Figure 89: Non-parametric adjusted distributions of accumulated Monetary Base expansion due to Net Foreign Exchange Purchases to the Private Sector as measured by the stock of International Reserves (v57) under crisis and no-crisis labels in scenario C. Source: own based on BCRA.

On the external front, the terms of trade (v7 - Figure 90) and the Real Exchange Rate gap over the long-run trend estimated as a simple average (v80 - Figure 91) gain predictive power.

On the fiscal-external front, the fiscal deficit (v40 - Figure 92), interest payments (v35 - Figure 93) and the primary income payments through the official exchange market (v90) gain more predictive power.

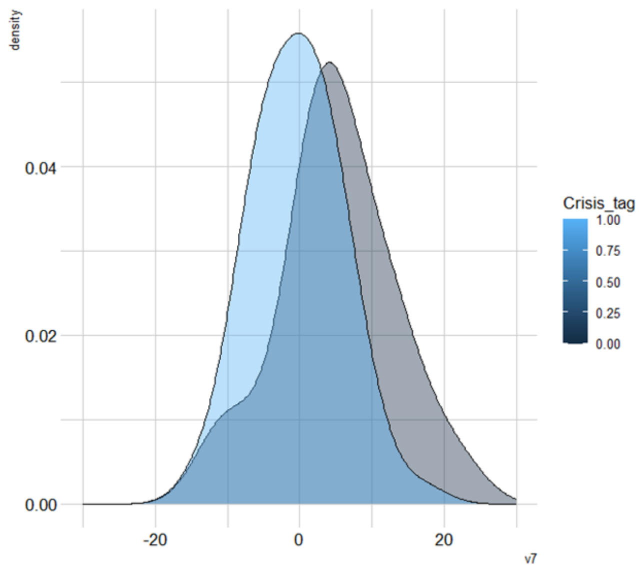


Figure 90: Non-parametric adjusted distributions of Terms of Trade year-over-year variation (v7) under crisis and no-crisis labels in scenario C. Source: own based on INDEC.

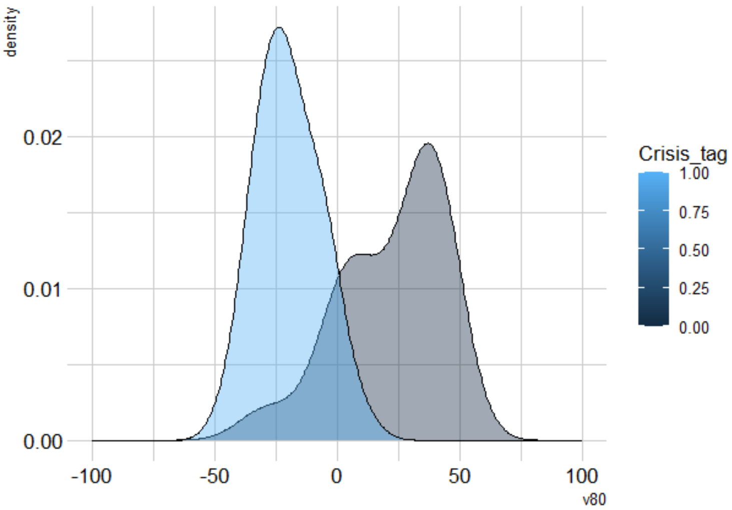


Figure 91: Non-parametric adjusted distributions of Real Exchange Rate gap over the long-run trend estimated as a simple average (v80) under crisis and no-crisis labels in scenario C. Source: own based on BCRA.

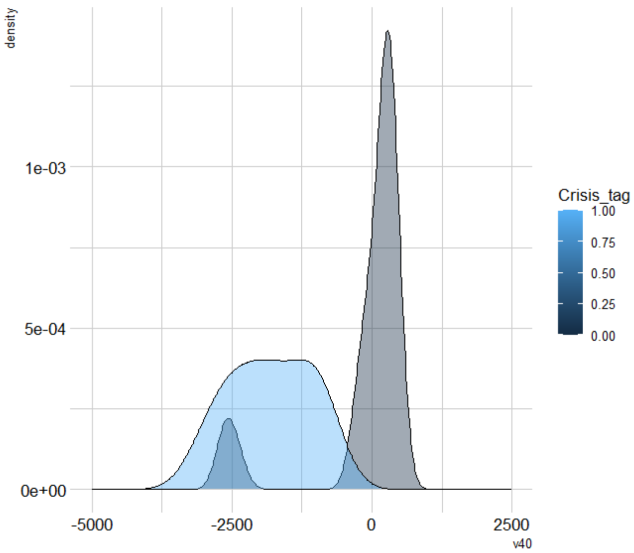


Figure 92: Non-parametric adjusted distributions of 12-month moving average Net Lending / Borrowing (v40) in official USD millions for NFPS under crisis and no-crisis labels in scenario C. Source: own based on BCRA.

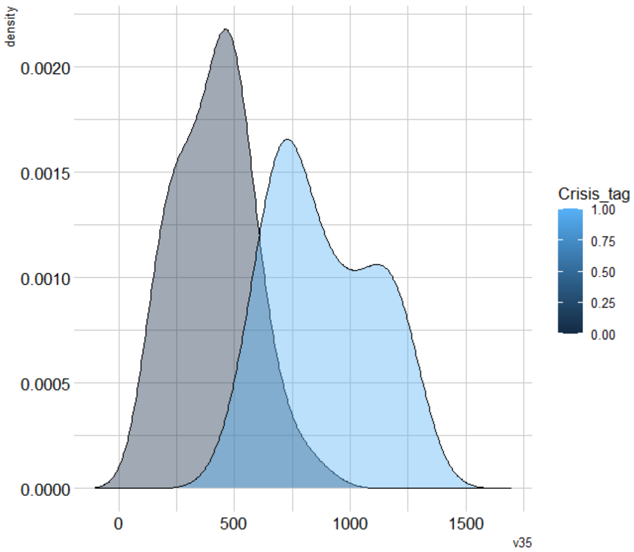


Figure 93: Non-parametric adjusted distributions of 12-month moving average Interest (v35) in official USD millions for NFPS under crisis and no-crisis labels in scenario C. Source: own based on BCRA.

It is worth noting that when considering 76 features, 'fiscal variables' remain important. In fact, v28 (Figure 96) is the most important variable to follow when excluding moving average fiscal variables v34, v35, v36, v37, v38, v39, v40 and accumulated monetary features v54, v55, v57, v59, v60.

When excluding these variables, the external variables Real Exchange Rate gap over the long-run trend estimated as a simple average (v80 - Figure 91) and the Multilateral Real Exchange Rate (v97 - Figure 97)) gain importance. Graph 94 shows the main variables for the BA model, and Graph 95 shows the main variables for the RF model when these variables are excluded.

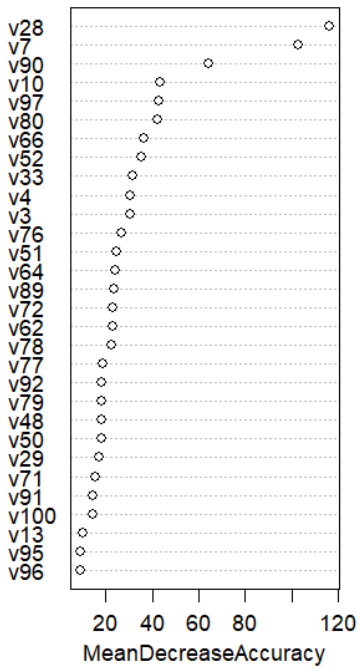


Figure 94: BA variable importance: Accuracy and Gini criteria. Variables: 76 (over 76). Trees: 10’000. OOB error: 5.06 per cent. Case: C.

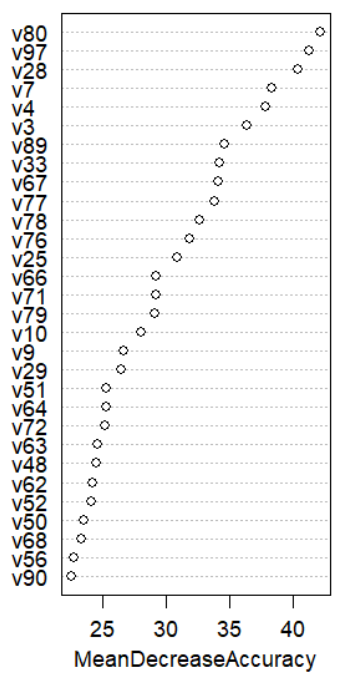


Figure 95: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 6 (over 76). Trees: 10’000. OOB error: 2.25 per cent. Case: C.

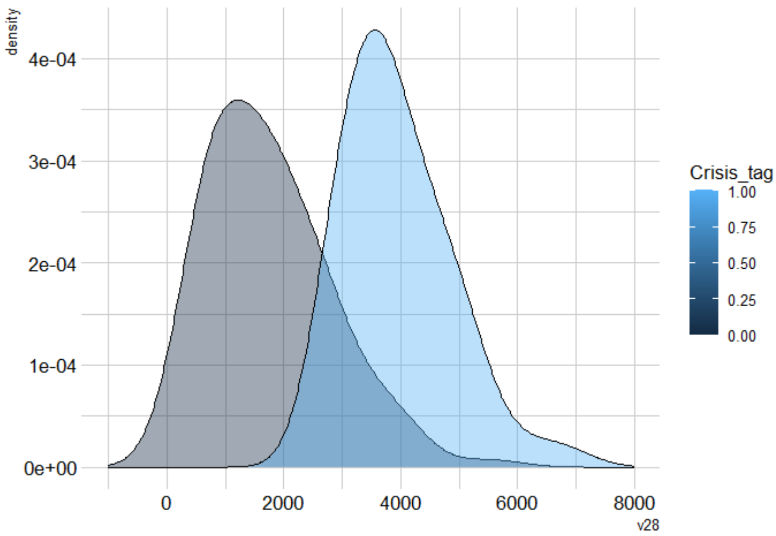


Figure 96: Non-parametric adjusted distributions of social benefits (v28) in official USD millions for NFPS under crisis and no-crisis labels in scenario C. Source: own based on INDEC.

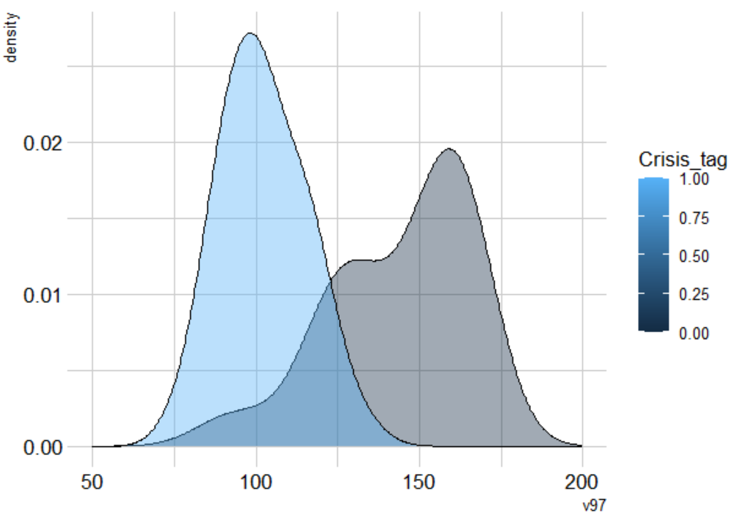


Figure 97: Non-parametric adjusted distributions of Multilateral Real Exchange Rate (v97) under crisis and no-crisis labels in scenario C. Source: BCRA.

Case D: same as (C), but with a signalling period of 12 months. Considering 88 features, 'Fiscal-external' and real variables remain relevant if we exclude exchange rate controls and the SBA from the crisis definition and reduce to 12-month signalling (scenario D). In particular, the 12-month moving average of social benefits in official USD (v37) is the most important variable. This is true for CART (Figure 98), BA (Figure 99) and RF (Figure 100) models. The main novelty is variable v48 (Short run Bonds in ARS expressed at the official exchange rate) in CART and BA models.

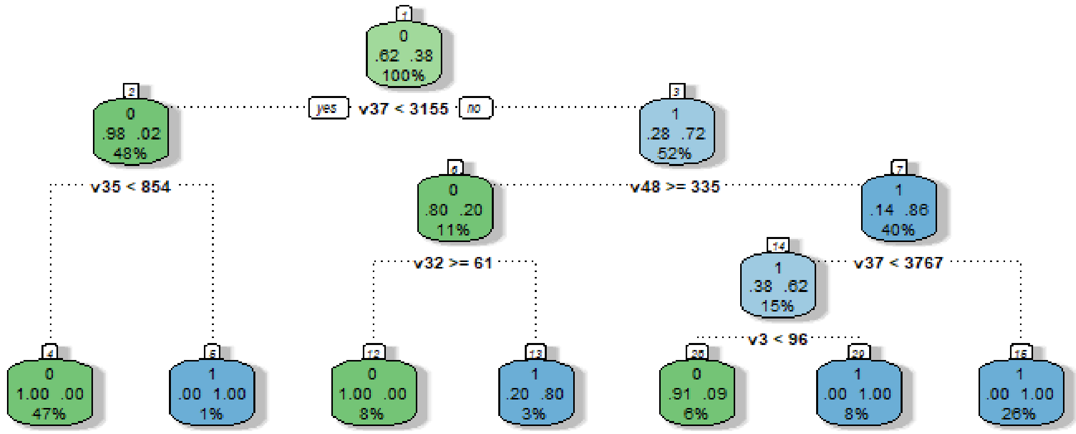


Figure 98: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.01470588. Selected number of splits: 6. Variables: 88. Case: D.

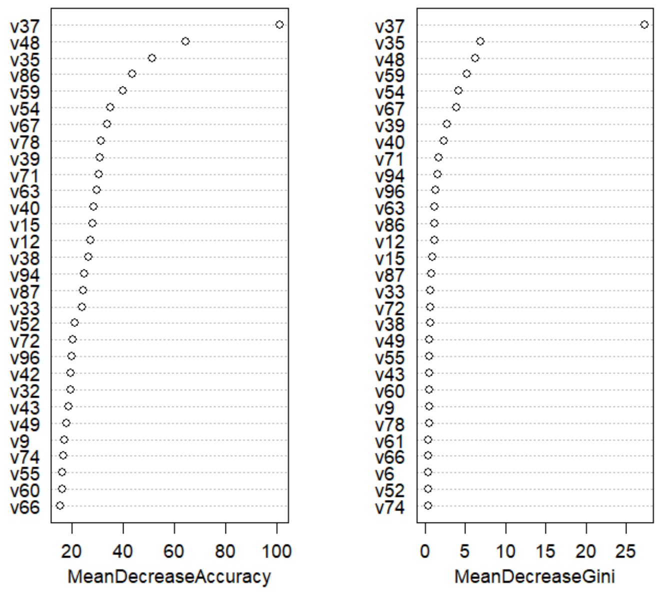


Figure 99: BA variable importance: Accuracy and Gini criteria. Variables: 88 (over 88). Trees: 1’000’000. OOB error: 3.37 per cent. Case: D.

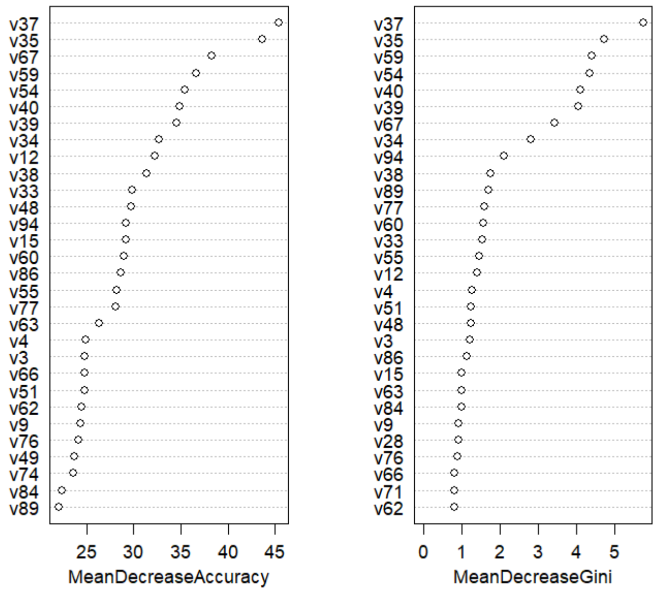


Figure 100: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 8 (over 88). Trees: 10'000. OOB error: 2.25 per cent. Case: D

However, it is important to note that social benefits (v37) and fiscal variables lose their significance in scenario D when excluding moving average fiscal variables v34 to v40 and accumulated monetary features. Graph 101 displays the main variables for the BA model, and Graph 102 displays the main variables for the RF model when these variables are excluded.

On the other hand, we can gain a better understanding considering external sector variables, year-over-year growth of International Reserves (v86 - Figure 107) and the Service Trade Balance at the official exchange market (v89 - Figure 113). Also, private sector deposits in ARS expressed in official USD (v67 - Figure 103) and the Financial System Return Over Equity (v72 - Figure 109) provide insight from the monetary sector. Finally, the Monthly output EMAE gap estimated as a 12-month mobile average (v78) is a useful indicator from the real sector.

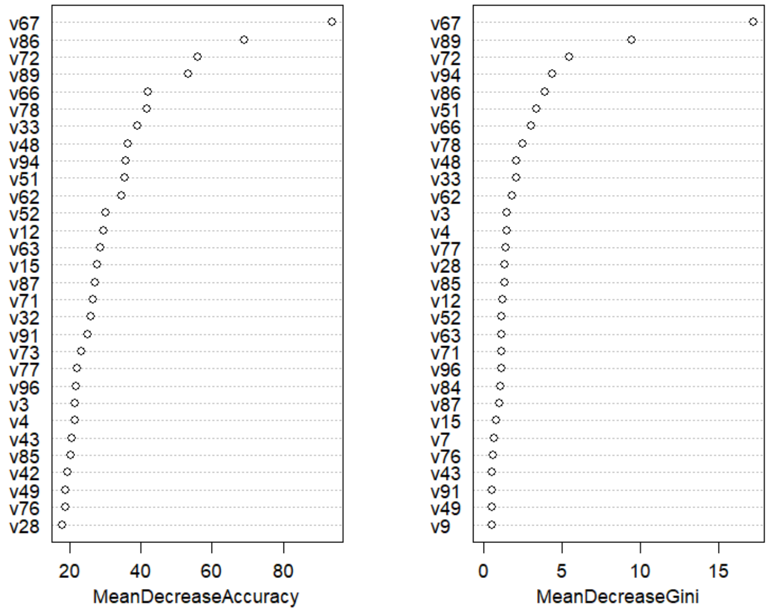


Figure 101: BA variable importance: Accuracy and Gini criteria. Variables: 76 (over 76). Trees: 10'000. OOB error: 3.37 per cent. Case: D.

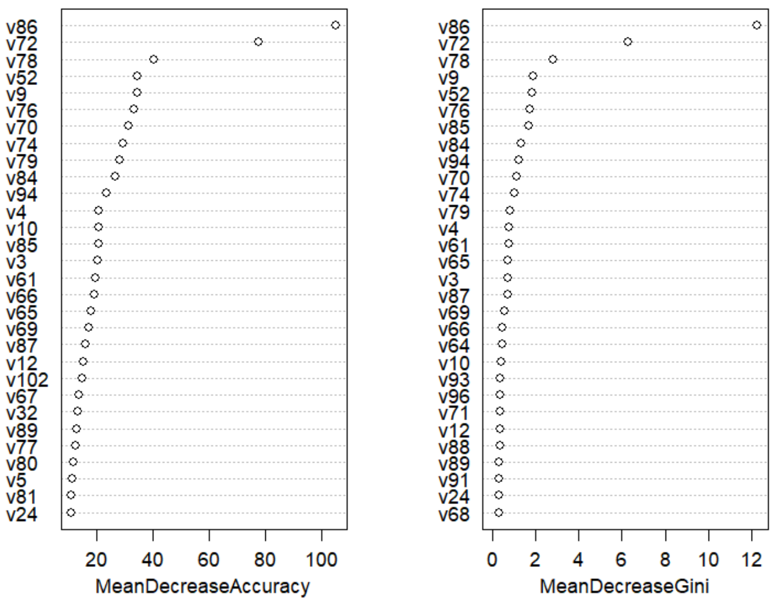


Figure 102: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 11 (over 76). Trees: 10'000. OOB error: 2.81 per cent. Case: D.

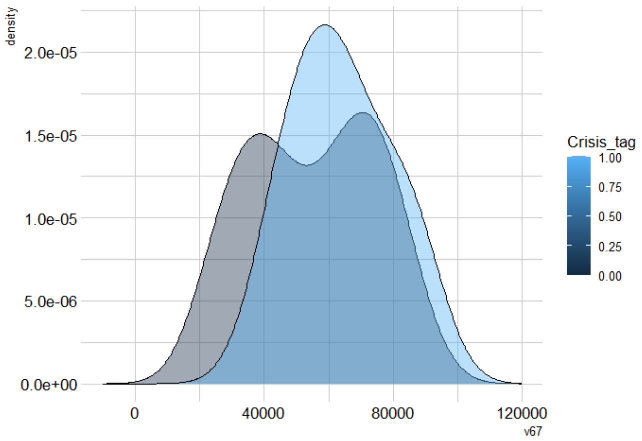


Figure 103: Non-parametric adjusted distributions of ARS Deposits from the private sector (v67) in official USD millions under crisis and no-crisis labels in scenario D. Source: own based on BCRA.

Case E: same as (D), but with a signalling period of one month. The significance of external variables increases for CART, BA, and RF models when the signalling period is shorter, as shown in Figure 104, Figure 105 and Figure 106. In scenario E, monitoring specific variables related to international reserves is important. The variables in question are the year-over-year growth of the International Reserve (v86 - Figure 107) and the accumulated expansion of the Monetary Base in USD A3500 due to the National Treasury's operations in relation to the stock of International Reserves (v60 - Figure 108). Additionally, the Financial System Return Over Equity (v72 - Figure 109) is a key variable in the monetary sector and is of particular importance.

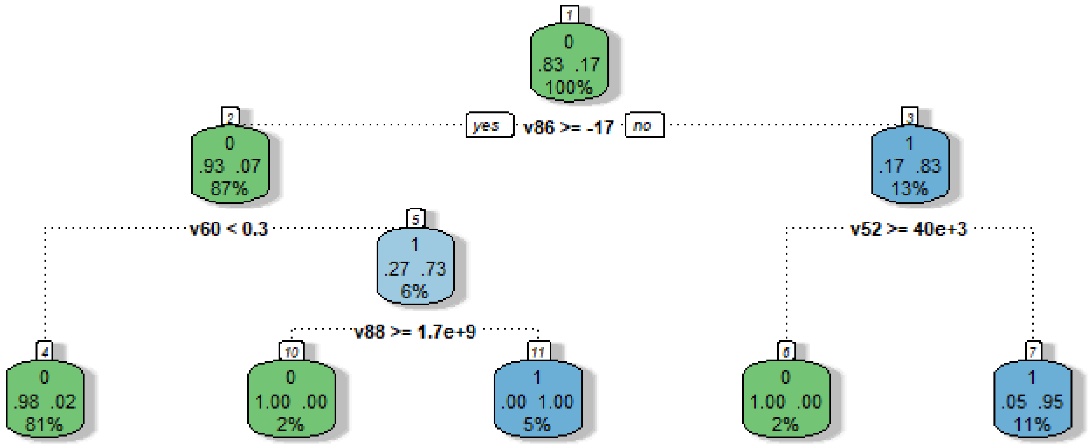


Figure 104: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 4. Variables: 88. Case: E.

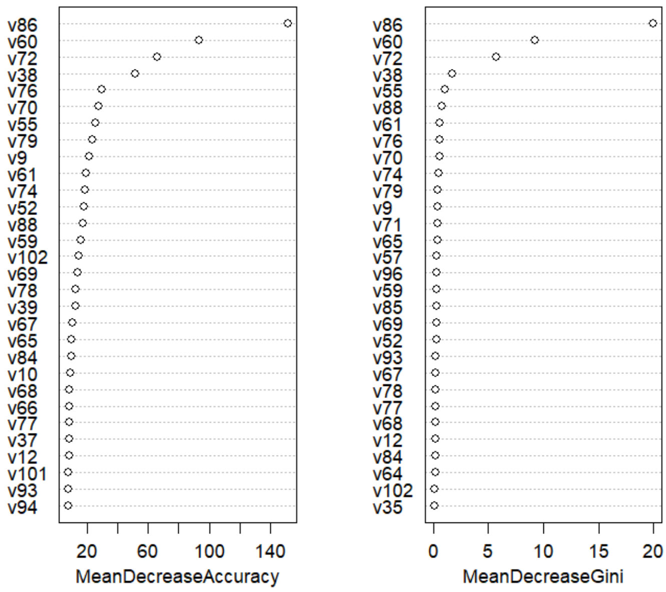


Figure 105: BA variable importance: Accuracy and Gini criterias. Variables: 88 (over 88). Trees: 10’000. OOB error: 6.78 per cent. Case: E.

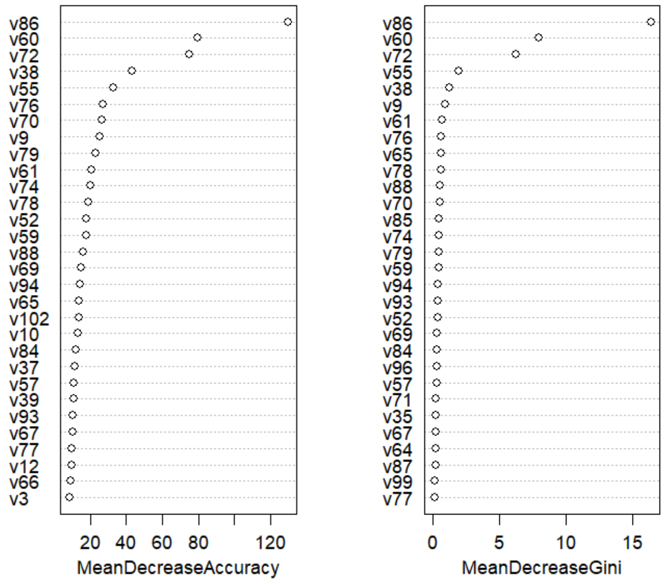


Figure 106: RF variable importance: Accuracy and Gini criterias. Cross-validation folds: 10. CV variables: 6 (over 88). Trees: 10'000. OOB error: 6.21 per cent. Case: E.

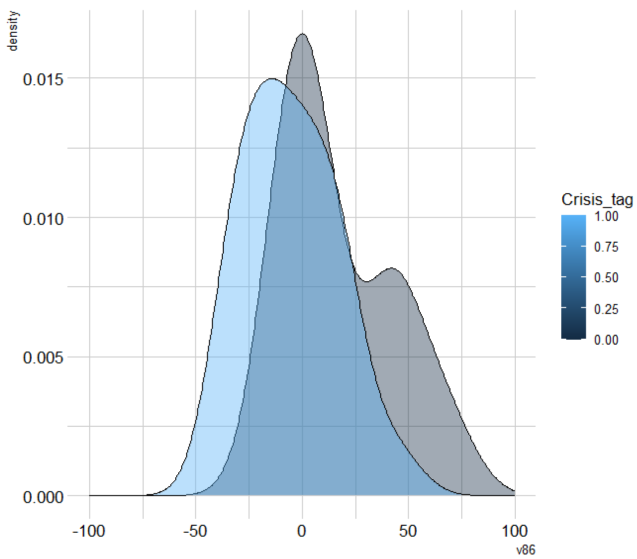


Figure 107: Non-parametric adjusted distributions of International Reserves year-over-year per cent growth (v86) under crisis and no-crisis labels in scenario E. Source: own based on BCRA.

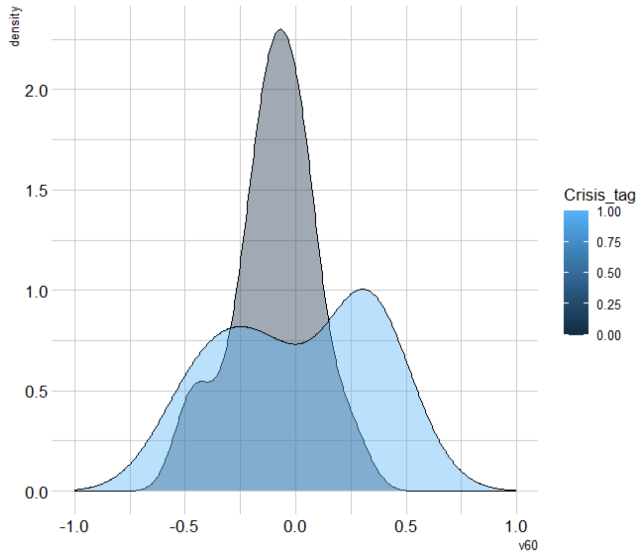


Figure 108: Non-parametric adjusted distributions of accumulated Monetary Base expansion in USD A3500 due to Other Operations with the National Treasury as measured by the stock of International Reserves (v60) under crisis and no-crisis labels in scenario E. Source: own based on BCRA.

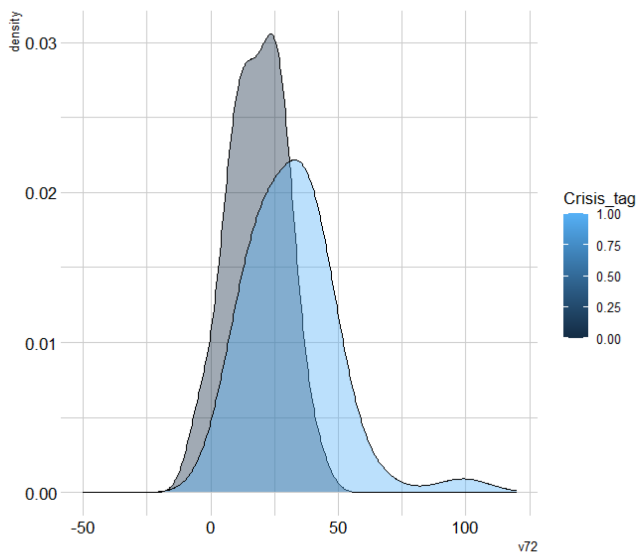


Figure 109: Non-parametric adjusted distributions of Financial system Return Over Equity in percentage (v72) under crisis and no-crisis labels in scenario E. Source: own based on BCRA.

It is worth noting that variables v86 and v72 are still the main variables to track when excluding moving average fiscal variables v34, v35, v36, v37, v38, v39, v40 and accumulated monetary features v54, v55, v57, v59, v60. Graph 110 shows the main variables for the BA model, and Graph 111 shows the main variables for the RF model when these variables are excluded.

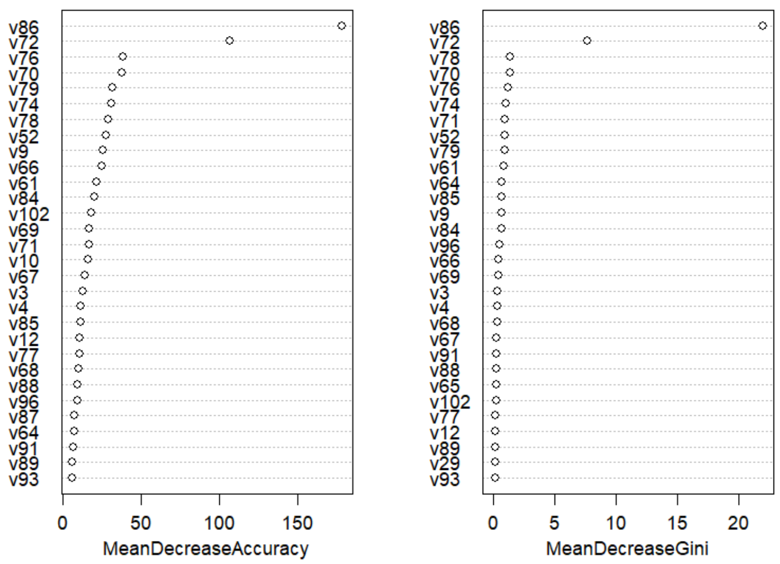


Figure 110: BA variable importance: Accuracy and Gini criteria. Variables: 76 (over 76). Trees: 10’000. OOB error: 6.78 per cent. Case: E.

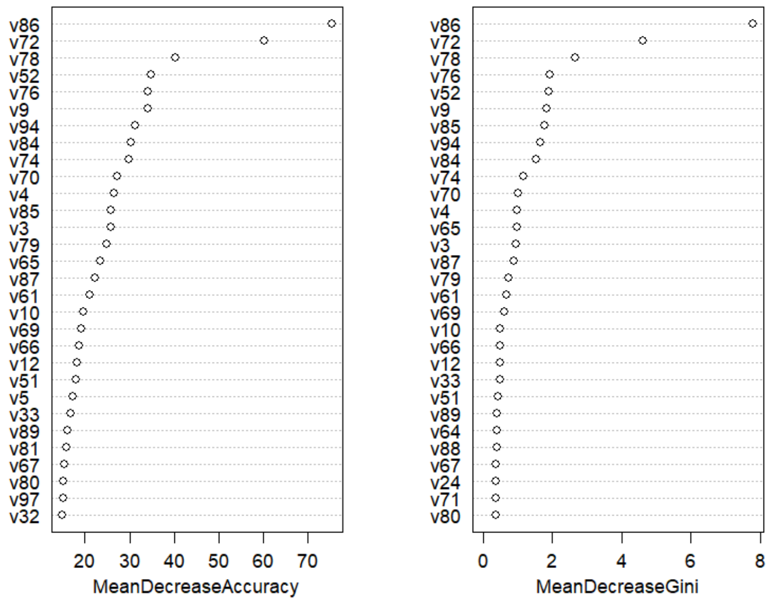


Figure 111: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 28 (over 76). Trees: 10'000. OOB error: 6.78 per cent. Case: E.

Scenario of interest: excluding exchange rate controls and signalling period of 12 months. A scenario of interest is the case of not considering exchange rate controls as a crisis. Some economists consider it a policy tool used to achieve industrial and redistributive objectives. Therefore, this is a special case that should be taken into account. On the other hand, the SBA period is still included as a crisis.

The assessment will evaluate the out-of-sample performance by removing nominal variables and moving average fiscal variables v33 to v40, as well as the accumulated monetary variables v54, v55, v57, and v59-60. This reduces the number of features from 103 to 75.

The variables to consider are mainly from the monetary and external sectors in CART (Figure 112), BA (Figure 114) and RF models (Figure 115). Important variables from the monetary sector include deposits in ARS from the private sector in USD A3500 (v67) and deposits in USD from the private sector (v66). From the external sector, it is important to consider the year-over-year growth of International Reserves (v86) and the Service Trade Balance at the official exchange market in USD (v89 - Figure 113). Meanwhile, the Monthly Economic Activity Estimator (EMAE) and social benefits in USD remain relevant variables to consider from the real and fiscal-external sectors, respectively.

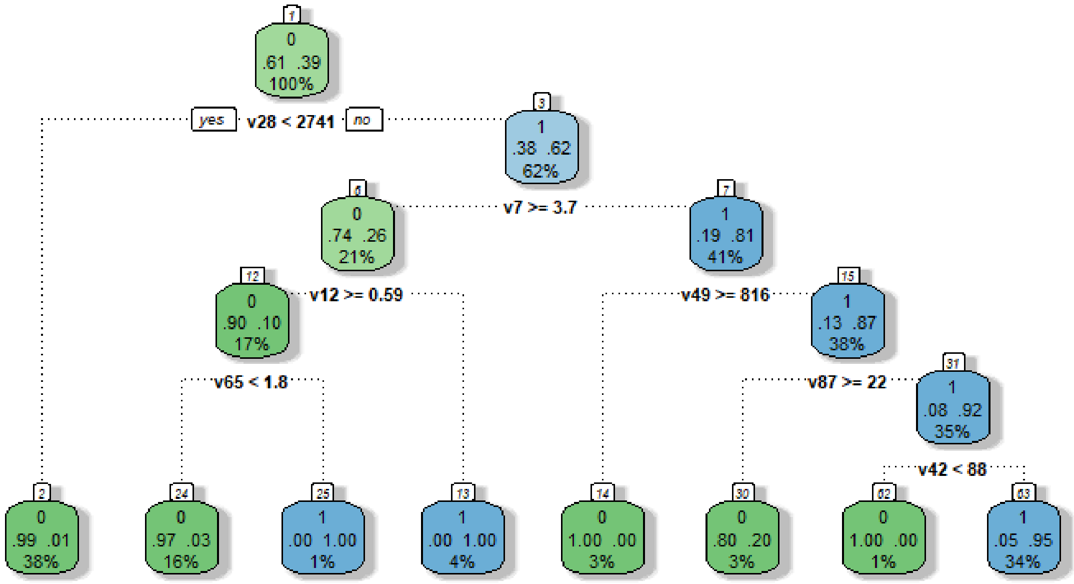


Figure 112: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.01428571. Selected number of splits: 7. Variables: 75. Scenario without exchange rate controls. Signalling period: 12 months.

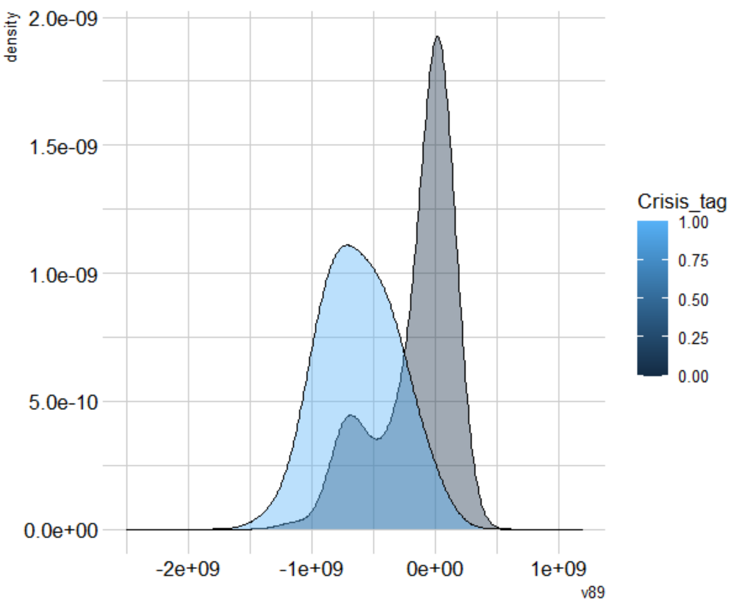


Figure 113: Non-parametric adjusted distributions of Service Trade Balance at the exchange market in USD (v89) under crisis and no-crisis labels. Scenario without exchange rate controls. Signalling period: 12 months. Source: own based on BCRA.

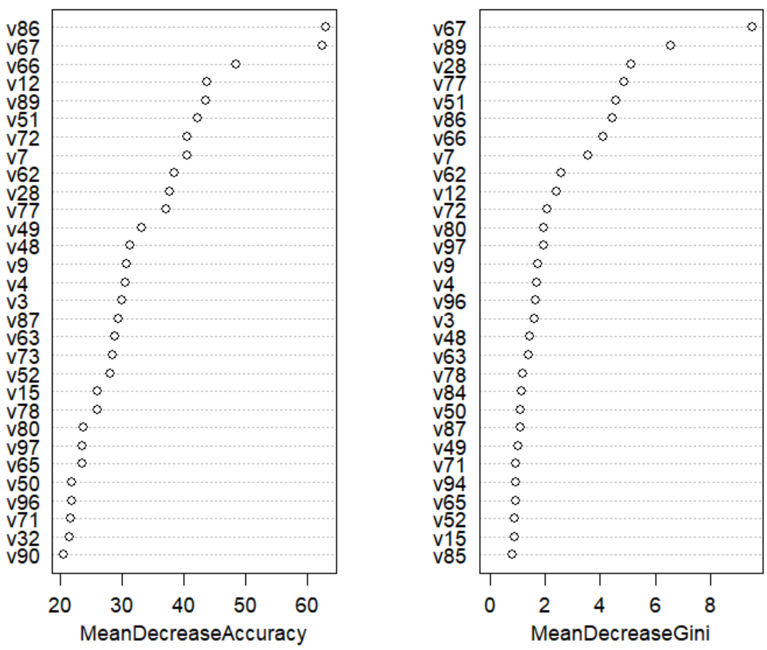


Figure 114: BA variable importance: Accuracy and Gini criteria. Variables: 75 (over 75). Trees: 10'000. OOB error: 4.49 per cent. Scenario without exchange rate controls. Signalling period: 12 months.

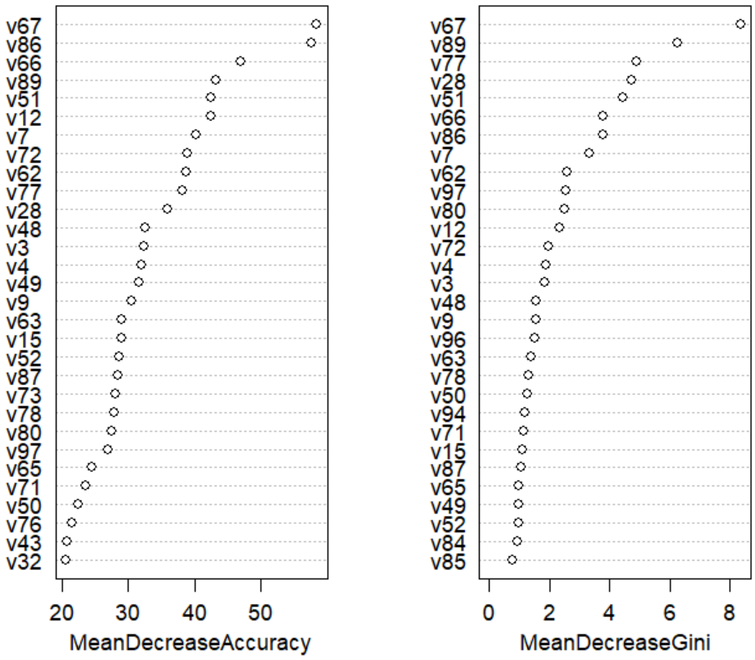


Figure 115: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 55 (over 75). Trees: 10'000. OOB error: 4.49 per cent. Scenario without exchange rate controls. Signalling period: 12 months.

6.6 Robustness checks

To ensure accurate and reliable results, three robustness checks were conducted. These included alternative model specifications using the same labels and features, alternative labels, and alternative features.

Alternative model specifications The CART, BA, and RF models produced similar performance results using the same labels and features. It is worth noting that the final results were not significantly affected by the choice of model specifications. Various partitioning criteria, such as information, entropy, and Gini, were used for CART models. Trees were pruned using both five and ten-fold cross-validation methods, and no significant differences were found between different specifications. Different numbers of trees and variable importance criteria, such as accuracy and Gini, were used for BA models. For RF models, variable selection was performed using either 5- or 10-fold cross-validation. The results remained consistent across different specifications when cross-validated.

Alternative labels The study assessed the out-of-sample performance of tree-based models across various crisis definitions and signalling periods to ensure the robustness of the findings. The models performed well for all metrics in all cases, but their performance decreased when there were more no-crisis observations than crisis observations. To address this issue in shorter periods, it may be worthwhile to explore the use of boosting and neural network models.

Alternative features The models were initially trained using a total of 103 features. These features were taken from different theories related to balance of payment crisis. The study also presented various versions of the same variable, such as year-on-year and monthly changes, as well as moving averages and gaps around a long-term trend. After this, nominal variables were excluded to test the results. In the second step, moving averages and accumulated variables that were commonly selected were also excluded. Finally, other variables that were usually selected were also excluded.

6.7 Extensions

Unsupervised machine learning: cluster analysis A potential extension of this thesis is to explore the application of unsupervised machine learning methods to aid in defining a crisis, specifically in the context of Argentina. Unsupervised learning involves working with unlabeled data, which in this case means that the crisis is not defined. Then, researchers can use clustering to label data and then feed a supervised classifier (Waggoner, 2020). Defining a crisis can be challenging due to differing opinions. To help researchers reach a consensus, it is important to implement objective methods. However, it is still essential to exercise sound judgement in cluster analysis.

Boosting and Neural Network models Future research on EWS for Argentina could benefit from exploring the integration of other machine learning models, such as boosting models (e.g. adaptive and gradient boosting) and neural networks. These models may perform better than tree-based methods for narrow crisis definitions and short signalling periods. Boosting models have been shown to better predict out-of-sample data than other algorithms. When non-informative features are present, boosting techniques may outperform the RF algorithm. However, it is important to note that boosting and neural networks require more computational and design effort and are often considered opaque compared to the RF algorithm.

Traditional models One possible extension is to evaluate the out-of-sample performance of linear machine learning models and traditional methods (i.e. logistic regressions and the signalling approach) for Argentina from 2003 to 2023. These classifiers may perform well when crises are defined in a broad sense and for long signalling periods. This clearly contrasts the main message of IMF (2021).

A previous assessment using a lasso penalised logit model resulted in high out-of-sample performance, although lower than BA and RF. The performance remains consistent when the crisis definition is 'broad', but widens as the definition 'narrows'. Figure 116 displays the External Risk Index constructed using a lasso-logit model for a 12-month signalling period without exchange rate controls.

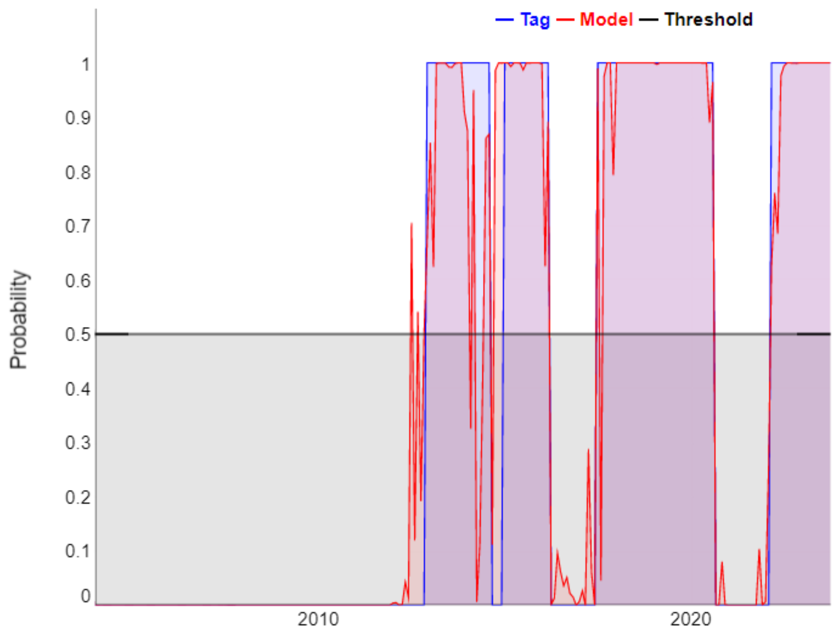


Figure 116: Labels, threshold and External Risk Index - lasso penalised logit model. Scenario 12-month signalling period without exchange rate controls.

7 Conclusions

This thesis uses machine learning tree-based models to present innovative work by developing an Early Warning System (EWS) for Argentina's external sector in the 21st century. The motivation behind this work was to provide a valuable tool for policymakers in Argentina by addressing a relevant subject in an innovative way.

Research question I. The study aimed to determine whether machine learning models based on trees could accurately predict external crises in Argentina. To achieve this, the study compared the performance of three ML models: CART, BA and RF. The study concludes that all three models have robust out-of-sample performance and can forecast a balance of payments crisis up to 24 months in advance.

This conclusion holds for alternative crisis definitions and signalling periods. The tree-based models showed excellent performance for broad crisis definitions and longer signalling periods. However, traditional methods and linear machine learning classification could also perform well when crises are defined as in the 'base case' given by IMF (2021). Thus, BA and RF models showed their true potential in the context of more specific crisis definitions and shorter signalling horizons. Even when the crisis definition was progressively narrower and the signalling horizon shorter, the BA and RF models still achieved high out-of-sample performance. Moreover, the exclusion of various characteristics did not affect the out-of-sample performance.

Research question II. The second research question addressed which tree-based method would best predict Argentina's external crises using data from 2003 to 2023. The CART, BA, and RF models performed similarly for broad crisis definitions and long signalling periods. However, for narrow crisis definitions and short signalling periods, the 'wisdom of crowds' worked well, with the BA and RF models outperforming the CART model. Additionally, RF has outperformed BA when considering short signalling periods.

Research question III. The third research question aimed at identifying the main instrumental variables that policymakers should monitor in order to be aware of potential external crises in Argentina. Considering the balance of payments crisis theories, one hundred and three macroeconomic variables were evaluated under different crisis definitions and signalling horizons. The identified variables are relevant only as instrumental variables to signal currency crises.

This research is empirical and does not discuss theory. An economist might be tempted to find the confirmation of his previous beliefs and theoretical framework in this empirical evidence. For example, some economists might find confirmation of a 'first-generation crisis' in the broad definitions of crisis and the signalled fiscal external variables (v34, v36 and v37). However, other economists could see a classic 'structural crisis' by emphasising the role of the EMAE (v77) and the IPC (v75). Therefore, it is important to stress that the machine learning models do not assume the ability to resolve theoretical disputes. The study does not claim to establish causality but only to identify 'signalling variables'.

Contrary to the 'one-size-fits-all' view of the IMF (2021), the models highlighted different variables from the fiscal, monetary, real, and external sectors as potential predictors of currency crises for different signalling horizons and crisis definitions in the case of Argentina. Depending on the crisis definition and signalling horizon adopted, different sets of variables have different degrees of importance. The set of characteristics included also changes the relative importance of key variables.

Research question III (case A). Based on broad crisis definitions and long signalling horizons, fiscal expenditure in official USD (v34, v36 and v37) and real variables such as EMAE (v77) and IPC (v75) were found to be the most important of the 103 variables. High levels of expenditure in USD, combined with high levels of economic activity and inflation, could signal an impending crisis within a 24-month time horizon. In other words, an expansionary fiscal policy combined with an exchange rate appreciation beyond the thresholds signals an impending crisis within a 24-month time horizon.

Research question III (case B). Excluding the nominal variables and the moving average fiscal-external variables, the key variables were the real variable v77 (EMAE) and the external variable v92 (financial loans, debt securities and lines of credit on the foreign exchange market in USD).

Research question III (case C). When exchange rate controls and the stand-by arrangement (SBA) are not considered as a crisis, the real variable v77 (EMAE) loses its predictive power, while the monetary-fiscal-external variables (v54, v59 and v57) gain predictive power. The creation of monetary base by the central bank to finance the treasury and the destruction of the monetary base by selling foreign exchange to private entities could be a sign of an impending crisis. External factors such as the real exchange rate (v97) and its appreciation relative to its long-term trend (v80), as well as the terms of trade (v7) and primary income payments through the official foreign exchange market (v90) have predictive power and should be taken into account. On the fiscal-external side, social benefits in USD (v28), interest payments (v35) and the budget deficit (v40) are also important predictors.

Research question III (case D and E). Alternative variables can provide predictive power for a shorter time horizon and a narrower crisis definition. For a 1-month signalling period, for tree-based models, it is worth considering the year-on-year growth in international reserves (v86), accumulated expansion of the Monetary Base in USD A3500 due to the National Treasury's operations in relation to the stock of International Reserves (v60), and financial system return on equity (v72) to gain insight. For 12-month signalling, it is also important to consider the services trade balance on the official exchange market (v89), private sector deposits in ARS expressed in official USD (v67) and the monthly output EMAE gap estimated as a 12-month mobile average (v78).

Research question III (exchange rate control not classified as a crisis). Finally, we considered a case where exchange rate controls are not classified as a currency crisis. To predict a crisis 12 months in advance, tree-based models selected relevant variables from the monetary sector: ARS deposits of the private sector in USD (v67) and USD deposits of the private sector (v66). From the external sector, the year-on-year growth of international reserves (v86) and the services trade balance on the official exchange market in USD (v89) were considered. Meanwhile, considering the Monthly Economic Activity Estimator EMAE (v77) and social benefits in USD (v28) as relevant variables from the real and fiscal-external sectors, respectively, is important.

Research question III (initial hypothesis). It is important to note that, contrary to our previous hypothesis, international and political variables did not play a central role as instrumental variables in identifying external crises. This research did not find political variables to be significant. Meanwhile, international variables played a secondary role. The only exception is v7, which refers to the year-over-year variation in terms of trade. It played an important role in identifying crises when the crisis definition did not consider exchange rate controls and the SBA.

Divergences with the IMF (2021). This thesis focused on the case of Argentina and found two broad differences with the IMF (2021), which is focused on a panel of emerging countries.

Coinciding with IMF (2021), our research finds that tree-based models had excellent out-of-sample performance in the IMF's 'base case' crisis definition. However, whether machine learning models significantly outperform traditional methods in this scenario is unclear. In contrast, in the case of Argentina, the IMF's perspective may be supported by 'narrow' crisis definitions. Further research is needed to provide a comprehensive answer to this question.

Secondly, the variable importance of the IMF's RF model for emerging countries is being questioned due to its 'one-size-fits-all' approach. This research demonstrates that different variables are significant for Argentina 2003-2023, depending on the crisis definition and signalled periods. For example, while the IMF (2021) identified reserve growth as the most important signalling variable, it is only significant for short signalling periods in the case of Argentina. Also, while GDP growth had a secondary importance in the IMF's emerging panel, it was a central signalling variable for the case of Argentina when the crisis is defined in a broad sense as proposed by the IMF (2021).

Extensions. This research raises new questions and suggests potential avenues for future research. One potential extension of this thesis is to investigate the use of unsupervised machine learning methods to assist in defining a crisis. Another possible extension is to assess the out-of-sample performance of linear machine learning models and traditional methods, such as logistic regressions and the signalling approach, for Argentina between 2003 and 2023. The third line of research on EWSs for Argentina could benefit from exploring the integration of other machine learning models, such as boosting models and neural networks.

References

Abiad, M. A. (2003). *Early warning systems: A survey and a regime-switching approach.* International Monetary Fund.

Acemoglu, D., Egorov, G., and Sonin, K. (2013). A political theory of populism. *The quarterly journal of economics*, 128(2):771–805.

Adalet, M. and Eichengreen, B. (2007). Current account reversals: always a problem? In *G7 Current Account Imbalances: Sustainability and Adjustment*, pages 205–246. University of Chicago Press.

Ahuja, M. A., Wiseman, K., and Syed, M. M. H. (2017). *Assessing country risk: Selected approaches.* International Monetary Fund.

Alejandro, C. F. D. (1963). A note on the impact of devaluation and the redistributive effect. *Journal of Political economy*, 71(6):577–580.

Alessi, L., Antunes, A., Babeckỳ, J., Baltussen, S., Behn, M., Bonfim, D., Bush, O., Detken, C., Frost, J., Guimaraes, R., et al. (2015). Comparing different early warning systems: Results from a horse race competition among members of the macro-prudential research network.

Alonso-Alvarez, I. and Molina, L. (2023). How to foresee crises? a new synthetic index of vulnerabilities for emerging economies. *Economic Modelling*, 125:106304.

Alvarez-Plata, P. and Schrooten, M. (2004). Misleading indicators? the argentinean currency crisis. *Journal of Policy Modeling*, 26(5):587–603.

Alves Jr, A. J., Ferrari Jr, F., and De Paula, L. F. (1999). The post keynesian critique of conventional currency crisis models and davidson's proposal to reform the international monetary system. *Journal of Post Keynesian Economics*, 22(2):207–225.

Arceo, N. (2008). *Empleo y salarios en la Argentina: una visión de largo plazo.* Capital intelectual.

Baran, P. A. (1957). The political economy of growth by.

Basualdo, E. (2020). *Endeudar y fugar: Un análisis de la historia económica argentina, desde Martínez de Hoz hasta Macri*. Siglo XXI Editores.

Berg, A., Borensztein, E., and Pattillo, C. (2005). Assessing early warning systems: how have they worked in practice? *IMF staff papers*, 52(3):462–502.

Berg, A. and Patillo, C. (1998). Are currency crises predicable. *A Test, "IMF Working Paper, WP/98/154*.

Berg, A. and Pattillo, C. (1999). Predicting currency crises:: The indicators approach and an alternative. *Journal of international Money and Finance*, 18(4):561–586.

Boonman, T. M., Jacobs, J. P., and Kuper, G. H. (2017). An early warning system for currency crises in argentina and brazil 1990-2009. *EconoQuantum*, 14(2):47–68.

Borio, C. E. and Lowe, P. W. (2002). Asset prices, financial and monetary stability: exploring the nexus.

Bortz, P. G. (2018). Flujos de capital y endeudamiento externo: Algunas reflexiones para américa latina. *Estudios sobre financierizacion en America Latina*, pages 295–321.

Bortz, P. G., Michelena, G., and Toledo, F. (2020). Exogenous shocks and external indebtedness. effects on growth and distribution in emerging and developing economies. *El trimestre económico*, 87(346):403–436.

Braun, O. and Joy, L. (1968). A model of economic stagnation—a case study of the argentine economy. *The Economic Journal*, 78(312):868–887.

Breiman, L. (1984). *Classification and regression trees*. Routledge.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

Breiman, L. and Cutler, A. (2001). Random forests. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_graphics.htm [Accessed: 2024/29/02].

Bussiere, M. (2013). In defense of early warning signals.

Caldentey, E. P. and Vernengo, M. (2013). Wage and profit-led growth: the limits to neo-kaleckian models and a kaldorian proposal. *Levy Economics Institute, Working Papers*, (775).

Calvo, G. A. (1998a). Capital flows and capital-market crises: the simple economics of sudden stops. *Journal of applied Economics*, 1(1):35–54.

Calvo, G. A. (1998b). Varieties of capital-market crises. In *The debt burden and its consequences for monetary policy: proceedings of a conference held by the International Economic Association at the Deutsche Bundesbank, Frankfurt, Germany*, pages 181–207. Springer.

Calvo, G. A., Izquierdo, A., and Mejia, L.-F. (2004). On the empirics of sudden stops: the relevance of balance-sheet effects.

Calvo, G. A., Izquierdo, A., and Talvi, E. (2006). Phoenix miracles in emerging markets: recovering without credit from systemic financial crises.

Calvo, G. A., Leiderman, L., and Reinhart, C. M. (1993). Capital inflows and real exchange rate appreciation in latin america: the role of external factors. *Staff Papers*, 40(1):108–151.

Calvo, G. A. and Mendoza, E. G. (1996). Petty crime and cruel punishment: lessons from the mexican debacle. *The American Economic Review*, 86(2):170–175.

Calvo, S. G. and Reinhart, C. (1996). Capital flows to latin america: is there evidence of contagion effects? *Available at SSRN 636120*.

Canitrot, A. (1975). La experiencia populista de redistribución de ingresos. *Desarrollo económico*, pages 331–351.

Canitrot, A. (1983). El salario real y la restricción externa de la economía. *Desarrollo económico*, pages 423–427.

Cardoso, F. H. and Faletto, E. (2000). Dependência e desenvolvimento na américa latina. *En: Cinqüenta anos de pensamento na CEPAL-Rio de Janeiro: Record/CEPAL, 2000-v. 2, p. 495-519*.

Chamon, M., Manasse, P., and Prati, A. (2007). Can we predict the next capital account crisis? *IMF Staff Papers*, 54(2):270–305.

Cimoli, M., Ocampo, J., and Porcile, G. (2016). Financial cycles and the macroeconomic dynamics of developing economies. Technical report, ISIGrowth Working Paper 29.

Cimoli, M., Porcile, G., and Rovira, S. (2010). Structural change and the bop-constraint: why did latin america fail to converge? *Cambridge Journal of Economics*, 34(2):389–411.

Claessens, S. and Forbes, K. (2004). International financial contagion: The theory, evidence and policy implications. In *Conference "The IMF's role in emerging market economies: Reassessing the adequacy of its resources", Amsterdam*.

Cortés, R. and Marshall, A. (1986). Salario real, composición del consumo y balanza comercial. *Desarrollo económico*, pages 71–88.

Cortés, R. and Marshall, A. (2003). Salarios, desigualdad y sector externo bajo distintos regímenes macroeconómicos. *Realidad Económica*, 195.

da Conceição Tavares, M. (2000). Subdesenvolvimento, dominação e luta de classes. *Celso Furtado e o Brasil*, page 129.

der Aalst, W. V. (2017). Hitting the bulls-eye with a low bias and low variance. https://twitter.com/wvdaalst/status/841353585229209600 [Accessed: 2024/29/02].

Diamand, M. (1972). La estructura productiva desequilibrada argentina y el tipo de cambio. *Desarrollo económico*, 12(45):25–47.

Diaz-Alejandro, C. (1985). Good-bye financial repression, hello financial crash. *Journal of development Economics*, 19(1-2):1–24.

Diaz-Alejandro, C. F. (1979). Southern cone stabilization plans.

Dornbusch, R. (1996). Latin triangle. Massachusetts Institute of Technology.

Dornbusch, R. and Edwards, S. (1991). The macroeconomics of populism. In *The macroeconomics of populism in Latin America*, pages 7–13. University of Chicago Press.

Dornbusch, R., Goldfajn, I., Valdés, R. O., Edwards, S., and Bruno, M. (1995). Currency crises and collapses. *Brookings papers on economic activity*, 1995(2):219–293.

Dvoskin, A. and Feldman, G. D. (2018). Income distribution and the balance of payments: a formal reconstruction of some argentinian structuralist contributions part ii: Financial dependency: Part ii: Financial dependency. *Review of Keynesian Economics*, 6(3):369–386.

Edwards, S. (2004). Financial openness, sudden stops, and current-account reversals. *American Economic Review*, 94(2):59–64.

Edwards, S. (2019). On latin american populism, and its echoes around the world. *Journal of Economic Perspectives*, 33(4):76–99.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, pages 569–593. Springer.

Eichengreen, B., Rose, A. K., and Wyplosz, C. (1995). Exchange market mayhem: the antecedents and aftermath of speculative attacks. *Economic policy*, 10(21):249–312.

Ferrer, A. (1963). Devaluación, redistribución de ingresos y el proceso de desarticulación industrial en la argentina. *Desarrollo económico*, pages 5–18.

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

Fiorito, A. (2015). Patrones de desarrollo y distribución del ingreso en la argentina. *Documento de Trabajo del Cefidar*, 70.

Flood, R. P. and Garber, P. M. (1984). Collapsing exchange-rate regimes: some linear examples. *Journal of international Economics*, 17(1-2):1–13.

Fortmann-Roe, S. (2012). Understanding the bias-variance tradeoff. *URL: http://scott. fortmann-roe. com/docs/BiasVariance. html (hämtad 2019-03-27)*.

Fouliard, J., Howell, M., and Rey, H. (2021). Answering the queen: Machine learning and financial crises. Technical report, National Bureau of Economic Research.

Frank, A. G. (1967). *Capitalism and underdevelopment in Latin America*, volume 93. NYU Press.

Frankel, J. A. and Rose, A. K. (1996). Currency crashes in emerging markets: An empirical treatment. *Journal of international Economics*, 41(3-4):351–366.

Frankel, J. A. and Wei, S.-J. (2004). Managing macroeconomic crises.

Furtado, C. (2020). *Formação econômica do Brasil*. Companhia das Letras.

Gelman, A., Hill, J., and Vehtari, A. (2020). *Regression and other stories.* Cambridge University Press.

Gerchunoff, P. and Rapetti, M. (2016). La economía argentina y su conflicto distributivo estructural (1930-2015). *El trimestre económico*, 83(330):225–272.

Gerchunoff, P., Rapetti, M., and De León, G. (2020). La paradoja populista. *Desarrollo económico*, 59(229):299–328.

Ghosh, S. R. and Ghosh, A. R. (2003). Structural vulnerabilities and currency crises. *IMF Staff Papers*, 50(3):481–506.

Girton, L. and Roper, D. (1977). A monetary model of exchange market pressure applied to the postwar canadian experience. *The American economic review*, 67(4):537–548.

González, M. L. (2011). La relación de largo plazo entre patrones de crecimiento y mercado de trabajo: un análisis sobre los salarios en argentina entre 1950 y 2006.

González, M. L. and Fernández, M. (2021). *¿ Por qué siempre faltan dólares?: Las causas estructurales de la restricción externa en la economía argentina del siglo XXI.* Siglo XXI Editores.

Graña, J. M., Kennedy, D., et al. (2009). " salarios eran los de antes...": salario, productividad y acumulación de capital en la argetnina en el último medio siglo.

Herrera, S. and Garcia, C. (1999). *User's guide to an early warning system for macroeconomic vulnerability in Latin American countries*, volume 2233. World Bank Publications.

Holopainen, M. and Sarlin, P. (2017). Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Quantitative Finance*, 17(12):1933–1963.

IMF (2012). *The IMF-FSB Early Warning Exercise: Design and Methodological Toolkit.* International Monetary Fund.

IMF (2013). *Early Warning System Models: The Next Steps Forward.* International Monetary Fund.

IMF (2021). *How to Assess Country Risk: The Vulnerability Exercise Approach Using Machine Learning.* International Monetary Fund.

ITBodhi (2020). Bias and variance trade off. https://medium.com/@itbodhi/bias-and-variance-trade-off-542b57ac7ff4 [Accessed: 2024/29/02].

James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning*, volume 112. Springer.

Johnson, H. G. (1972). The monetary approach to balance-of-payments theory. *Journal of financial and quantitative analysis*, 7(2):1555–1572.

Johnson, H. G. (1977). The monetary approach to balance of payments theory and policy: Explanation and policy implications. *Economica*, 44(175):217–229.

Joy, M., Rusnák, M., Šmídková, K., and Vašíček, B. (2017). Banking and currency crises: Differential diagnostics for developed countries. *International Journal of Finance & Economics*, 22(1):44–67.

Kaminsky, G., Lizondo, S., and Reinhart, C. M. (1998). Leading indicators of currency crises. *Staff Papers*, 45(1):1–48.

Kaminsky, G., Mati, A., and Choueiri, N. (2009). Thirty years of currency crises in argentina: External shocks or domestic fragility? Technical report, National Bureau of Economic Research.

Kaminsky, G. L. (1999). *Currency and banking crises: the early warnings of distress.* International Monetary Fund.

Kaminsky, G. L. (2006). Currency crises: Are they all the same? *Journal of International Money and Finance*, 25(3):503–527.

Keynes, J. M. (1919). *The economic consequences of the peace.* Routledge.

Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada.

Krugman, P. (1979). A model of balance-of-payments crises. *Journal of money, credit and banking*, 11(3):311–325.

Krugman, P. (1998). What happened to asia. In *Global competition and integration*, pages 315–327. Springer.

Krugman, P. and Taylor, L. (1978). Contractionary effects of devaluation. *Journal of international economics*, 8(3):445–456.

Krugman, P. R., Rogoff, K. S., Fischer, S., and McDonough, W. J. (1999). Currency crises. In *International capital flows*, pages 421–466. University of Chicago Press.

Lin, C.-S., Khan, H. A., Chang, R.-Y., and Wang, Y.-C. (2008). A new approach to modeling early warning systems for currency crises: Can a machine-learning fuzzy expert system predict the currency crises effectively? *Journal of International Money and Finance*, 27(7):1098–1121.

Lindenboim, J., Kennedy, D., and Graña, J. M. (2007). Salarios y productividad: contenido de la distribución funcional. diferenciales sectoriales y de tamaño en argentina en el período 1993-2006. In *publicado en el CD del 8 Congreso Nacional de Estudios del Trabajo*, volume 8.

Llach, J. J. and Sánchez, C. E. (1984). Los determinantes del salario en la argentina. un diagnóstico de largo plazo y propuestas de políticas. *Estudios*, 7(29):3–47.

Louppe, G., Wehenkel, L., Sutera, A., and Geurts, P. (2013). Understanding variable importances in forests of randomized trees. *Advances in neural information processing systems*, 26.

Manasse, P. and Roubini, N. (2009). "rules of thumb" for sovereign debt crises. *Journal of International Economics*, 78(2):192–205.

Marongiu, F. et al. (2005). Towards a new set of leading indicators of currency crisis for developing countries: an application to argentina. *Draft of*, page 3.

Masson, P. (1999). Contagion:: macroeconomic models with multiple equilibria. *Journal of international Money and Finance*, 18(4):587–602.

Médici, F. (2017). *Discusiones sobre el tipo de cambio. El eterno retorno de lo mismo.* UNM.

Mendoza, E. G. (2002). Credit, prices, and crashes: Business cycles with a sudden stop. In *Preventing currency crises in emerging markets*, pages 335–392. University of Chicago Press.

Moreno-Brid, J. C. (1998). On capital flows and the balance-of-payments-constrained growth model. *Journal of Post Keynesian Economics*, 21(2):283–298.

Nag, A. and Mitra, A. (1999). Neural networks and early warning indicators of currency crisis. *Reserve Bank of India Occasional Papers*, 20(2):183–222.

Obstfeld, M. (1996). Models of currency crises with self-fulfilling features. *European economic review*, 40(3-5):1037–1047.

O'donnell, G. (1977). Estado y alianzas en la argentina, 1956-1976. *Desarrollo ecońomico*, 16(64):523–554.

Olivera, J. H. (1977). Monetarismo vs. estructuralismo. In *conferencia reproducida en Olivera*.

Pereira, L. C. B. (1991). Populism and economic policy in brazil. *Journal of Interamerican Studies and World Affairs*, 33(2):1–22.

Pinto Santa Cruz, A. (1953). Chile, un caso de desarrollo frustrado. *(No Title)*.

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1:81–106.

Reinhart, C. and Calvo, G. (2000). When capital inflows come to a sudden stop: Consequences and policy options.

Rey, H. (2015). Dilemma not trilemma: the global financial cycle and monetary policy independence. Technical report, National Bureau of Economic Research.

Rodrik, D. (2018). Is populism necessarily bad economics? In *AEA Papers and proceedings*, volume 108, pages 196–199. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Sachs, J. D. (1989). Social conflict and populist policies in latin america.

Sevim, C., Oztekin, A., Bali, O., Gumus, S., and Guresen, E. (2014). Developing an early warning system to predict currency crises. *European Journal of Operational Research*, 237(3):1095–1104.

Sidrauski, M. (1968). Devaluación, inflación y desempleo. *Económica*, 14.

Stekhoven, D. J. and Bühlmann, P. (2012). Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118.

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Tavares, M. d. C. and Serra, J. (1972). Além da estagnação. *Da substituição de importações ao capitalismo financeiro*, 153:207.

Thirlwall, A. P. and Hussain, M. N. (1982). The balance of payments constraint, capital flows and growth rate differences between developing countries. *Oxford economic papers*, 34(3):498–510.

Tibshirani, R., Hastie, T., , Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Tibshirani, R. J. and Efron, B. (1993). An introduction to the bootstrap. *Monographs on statistics and applied probability*, 57(1).

Velasco, A. (1987). Financial crises and balance of payments crises: a simple model of the southern cone experience. *Journal of development Economics*, 27(1-2):263–283.

Vernengo, M. (2006). Technology, finance, and dependency: Latin american radical political economy in retrospect. *Review of Radical Political Economics*, 38(4):551–568.

Villanueva, J. (1964). Problemas de industrialización con restricciones en el sector externo. *Desarrollo Económico*, pages 171–182.

Waggoner, P. D. (2020). *Unsupervised machine learning for clustering in political and social research*. Cambridge University Press.

Währungsfonds, I. (2010). The imf-fsb early warning exercise. *Design and Methodological Toolkit, September*.

Ward-Powers, D. (2011). Evaluation: from precision, recall and f-factor to roc, informedness, markedness & correlation. *Journal of Machine Learning Research*, 2:37–63.

Wilber, J. and Croome, J. (2012). *Cross validation. Reduce, Reuse, Resample.* https://mlu-explain.github.io/cross-validation/.

Wright, M. N. and Ziegler, A. (2015). ranger: A fast implementation of random forests for high dimensional data in c++ and r. *arXiv preprint arXiv:1508.04409*.

Zeolla, N. H. and Médici, F. (2022). Desregulación, endeudamiento y fragilidad financiera externa: un enfoque minskiano para la crisis argentina 2018-2019. *Ensayos de Economía*, 32(60):66–90.

List of Tables

List of Figures

List of Acronyms

A3500 Oficial ARS-USD exchange rate

ADR American Depositary Receipt

AICM Argentine Institute of Capital Markets

AR Accuracy Rate

ARS Argentine pesos

AUC Area Under the ROC Curve

BA Bagging or Bootstrap Aggregating

BIS Bank for International Settlements

BCRA Central Bank of the Argentine Republic

CART Classification and Regression Trees

COVID Coronavirus disease

CV Cross-Validation

DXY U.S. Dollar Index

EMAE Monthly Economic Activity Estimator

EWS Early Warning System

FED Federal Reserve System

FRED Federal Reserve Bank of St. Louis

ICA Argentine Commercial Exchange

IMF International Monetary Fund

INDEC The National Institute of Statistics and Censuses of Argentina

LR Logistic and probit regressions

MEcon Treasury of Argentina

MERVAL Argentine stock market

ML machine learning

MTEySS Ministry of Labor and Social Security of Argentina

NFPS Non-Financial Public Sector

NOB Net Operating Balance

RF Random Forest

OOB Out-of-bag

PR Precision Rate

ROC Receiver Operating Characteristic

SBA Stand-By Arrangement between Argentina and the IMF

SE Signal Extraction method

FN False Negative

FP False Positive

RR Recall Rate

TN True Negative

TP True Positive

USD United States dollar