Universidad Nacional
de San Martín

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

A. A. Comisso

UNIVERSIDAD NACIONAL DE SAN MARTIN
MAESTRIA EN DESARROLLO ECONOMICO

March 2024

Index

Motivation and thesis objective

* This study compares the out-of-sample performance of three different EWSs Machine Learning models using CART, BA and RF. In addition, using 103 macroeconomic variables, the models highlight the relevance of certain variables as potential predictors of future currency crises in Argentina.
* **Motivation**: The topic strike a balance between being relevant, original, and practical.
* **Relevance (I)**: Keynes expressed it straightforwardly in 1919.
* **Relevance (II)**: Argentina experienced at least a currency crisis in 36 out of 240 months from 2003 to 2023.
* **Relevance (III)**: IMF(2021) has developed a machine learning (ML) model for country surveillance, but as of now, Argentina does not have one.
* **Originality**: tree-based machine learning models for the case of Argentina (no previous research).
* **Practical tool**: Art. IV Consultations.

Motivation

Figure: Changing methodologies for external EWSs: IMF(2017) vs. IMF(2021)

Research questions and hypotheses

1. First, to investigate whether machine learning models based on trees could be used to create 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.
2. 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.
3. 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.

Lack of consensus on crisis

- External crisis: a dummy variable is required (the 'label' in supervised Machine Learning).
- The lack of consensus on the definition and measure of external crises
  - Nominal exchange rate depreciation
  - Decline in international reserves
  - Rising interest rates
  - Exchange rate controls
  - IMF financial support
  - Other sources of discrepancies (e.g. the threshold). A currency crisis in 2006? Moving the threshold, we lose the 2014 crisis.

Lack of consensus on crisis

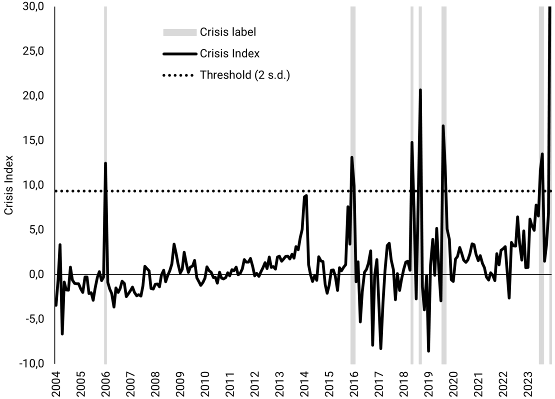

Figure: Crisis Index, threshold (2 s.d.) and crisis label for Argentina (2004-2023) based on Kaminsky et al. (2009). Source: own based on BCRA.

This thesis crisis definition

* **'Base case' broad** definition gradually **narrowed down**
  * **'Base case' broad** crisis definition similar to that proposed by IMF(2021) as pragmatic approach.
  * **Narrowed down** as a robustness exercise
  * Narrowed down due the lack of consensus
* **'Base case' broad** definition of crisis: i) the weighted average of the yoy percentage variation of A3500 and the yoy variation in IR is in the upper 85th percentile; ii) exchange controls, i.e. ADR is 20 per cent or more above the official exchange rate; and iii) IMF support at least five times the quota.
  (A) **'Base case'** and signalling 24 months (103 variables);
  (B) (A) but only 88 variables (excluding nominal variables);
  (C) (B) without exchange rate controls and SBA;
  (D) (C) but signalling period 12 months;
  (E) (D) but signalling period 1 month (**'index only'**).

Crisis labels for Argentina (2004-2023)

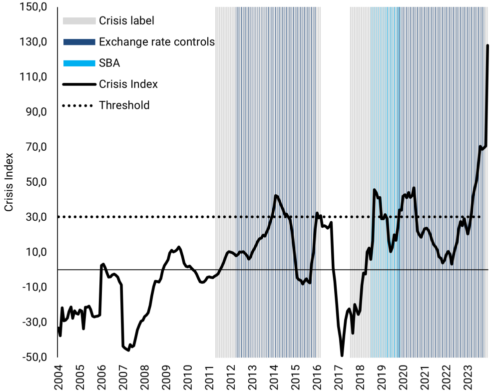

Figure: Crisis Index and crisis labels for Argentina (2004-2023). Based on IMF (2021) including SBA and exchange rate controls (12-month signalling period). Source: own based on BCRA.

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.
* Classified criteria by Alves Jr. et al. (1999)
* **Structuralist** models (inward-looking industry)
* **First**-generation models (fiscal policy)
* **Second**-generation models (real exchange misalignment)
* **Third**-generation models (financial crisis - **sudden stops**)

Database

* The **theories** of balance of payments crises provide the inspiration for the selected predictive variables.
* We have also selected **variables that are traditionally used** in EWSs of international organisations.
* However, the choice is **constrained by data availability**. The period covered is February 2003 to October 2023. The frequency of the variables is monthly.
* The 103 selected characteristics were classified into six groups: i) global variables, ii) fiscal sector variables, iii) financial and monetary sector variables, iv) real sector variables, v) external sector, and vi) political variables.

Global variables

Global variables: SPY – Bloomberg (v0); 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

Fiscal variables

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 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 as a source.

Financial and monetary sector variables

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)

Financial and monetary sector variables

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)

Financial and monetary sector variables

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)

Financial and monetary sector variables

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).

External sector variables

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).

Real sector and political variables

Real sector: Consumer Price Index – INDEC (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).

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).

External EWSs: literature review

* **Pre-machine learning external EWSs literature**
  * 70s, 80s, 90s, 2000s two methodological approaches have become standard in private banks and public sectors (IMF, 2017):
  1. ‘signalling approach’
  2. ‘categorical dependent variable regression’
* **Machine learning external EWS literature**
  * Neural Networks pioneers
  * CART pioneers
  * According to IMF(2021) CART, Random Forest and Boosting techniques are considered to be the next generation of risk assessment models. External EWS for 192 countries between 1990-2017 using 79 variables.
* **Machine learning external EWS literature on Argentina**: no studies have been conducted to develop an external EWS for Argentina ML tree-based models.

Bias-variance trade-off

* How can I effectively **train and choose** a machine learning model that provides the most accurate predictions when tested with new data? (i.e. low **out-of-sample** prediction error)
* **Prediction error** can be divided into three parts: irreducible error, bias and variance.

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

* **Trade-off** between bias and variance: 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.

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

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

Under and over-fitting trade-off

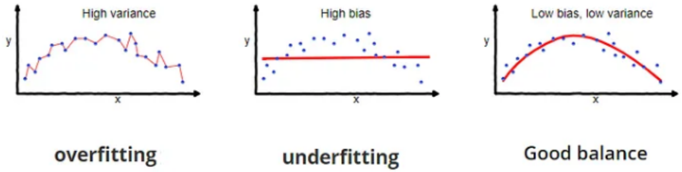

Figure: The problem of dealing with the bias-variance trade-off is the problem of dealing with over-fitting and under-fitting. An **over-fitted model** remembers noise instead of learning the underlying trends in the data set, resulting in poor out-of-sample performance. In other words, the model may perform well on the data it was trained on, but not on new, unseen data. On the other hand, an **under-fitted model** has to be more complex to capture the relationships accurately. Source: ITBodhi (2020)

Under and over-fitting trade-off

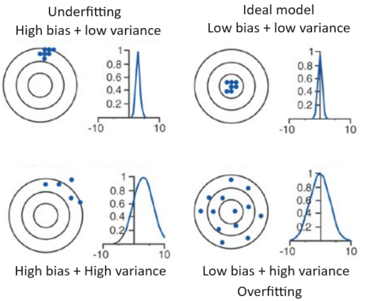

Figure: Bias and variance graphically explained. Source: own based on Fortmann-Roe (2012) and der Aalst (2017).

The Validation Set Approach

Figure: A **‘training set’** to train the model to learn potential underlying relationships; a **‘validation set’** to evaluate performance and select between different model types and hyper-parameter choices; and a **‘test set’** to evaluate the model’s performance on unseen data, also known as out-of-sample performance. **Advantage**: computationally cheap (vs. resampling approaches). **Disadvantage**: Would the chosen model change significantly if we had chosen a different split between training and validation?

k-Fold Cross-Validation

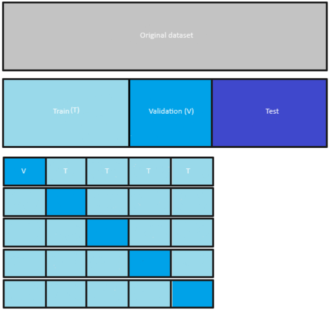

Figure: The data not used for testing are randomly divided into k equal-sized **folds**. For each estimation, $k-1$ folds are used as the **training set**, and one fold is reserved as the **validation set**. This process is repeated k times, with each fold serving as a validation set. Finally, the **average** 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.

Bootstrap

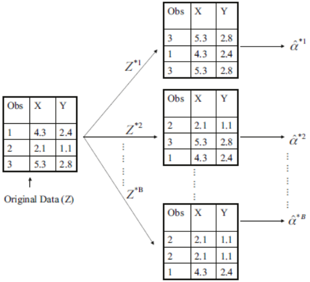

Figure: Consist in **training** the model by **randomly sampling with replacement** from the non-test set. Each random sample from the original set of *n* elements should also have *n* elements. It is important to note that this sampling is done with replacement. **Validation** is performed using the non-selected, i.e. OOB observations for each sample. Source: Hastie et al. (2009)

Methods: CART, BA and RF

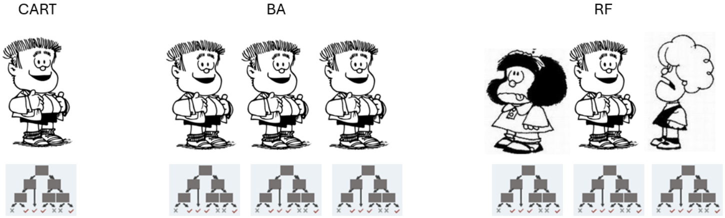

Figure: CART (low bias but high variance: impact on out-of-sample performance), BA (wisdom of crowds and lower variance but high tree correlation), RF (lower tree correlation). Source: www.quino.com.ar

Classification and Regression Trees (CART)

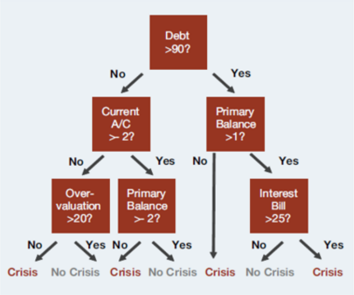

Figure: CART is a supervised machine learning technique first introduced by Breiman (1984). CART produces a **recursive**, one-dimensional sample **partition into disjoint subsets** of crisis and no-crisis. Source: IMF (2021)

BA and Random Forest (RF)

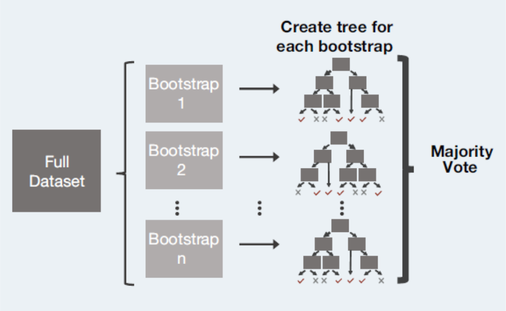

Figure: By the **majority vote** of many weak unpruned **trees**, the BA and RF algorithm makes an aggregate prediction ('ensemble learning'). Outside of statistics, this idea is known as the **'wisdom of crowds'**. Source: IMF (2021)

Evaluation metrics

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

Table: There are **four possible responses** from the model. First, the model could correctly predict a crisis, i.e. a **TP** case. Second, it could correctly predict a no-crisis, i.e. a **TN** case. Third, the model could incorrectly predict a crisis, i.e. a **FP** case. Finally, the model could fail to predict a crisis, i.e. **FN** case. A confusion matrix provides a visual representation of these four alternatives.

Evaluation metrics

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

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

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

Evaluation metrics

* If the country 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.
* Imagine that the model predicts 100 crises in 100 years. Then its **RR** will equal one (but erroneously predict 99 out of 100 cases).
* 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.
* A **trade-off** between **RR** and **PR** often exists in practice. 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.
* Only when both **PR** and **RR** tend to be one does the **F1** score also tend to be one.

Evaluation metrics: ROC and AUC

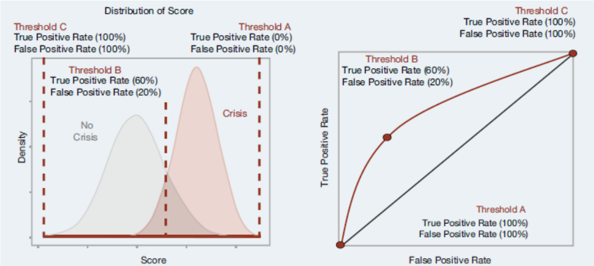

Figure: The receiver operating characteristic (**ROC**) curve visually explains the RR and PR trade-offs. In essence, there exists a trade-off between true positive cases and false positive cases. The area under the ROC curve (**AUC**) also gives a sense of the model's proficiency. Source: IMF (2021).

Empirical findings: research question I

* The study aimed to determine whether machine learning CART, BA and RF models could accurately predict external crises in Argentina. 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.
* **Robustness (labels)**: this conclusion holds for alternative crisis definitions and signalling periods. 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.
* **Robustness (features)**: Moreover, the exclusion of various features did not affect the out-of-sample performance.

Empirical findings: research question I

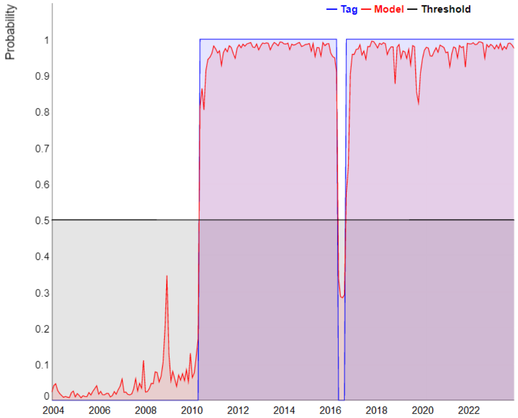

Figure: Labels, threshold and External Risk Index - RF model. Scenario A' and B' (broad crisis definition) but trained with 76 variables (24 months signalling with exchange rate controls and SBA).

Empirical findings: research question I

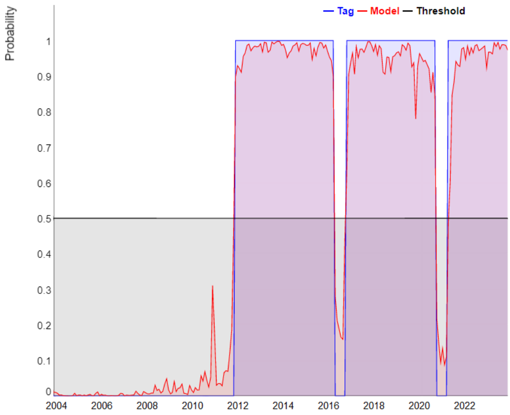

Figure: Labels, threshold and External Risk Index - RF model (trained with 76 variables). Scenario C' (24 months signalling without exchange rate controls and SBA).

Empirical findings: research question I

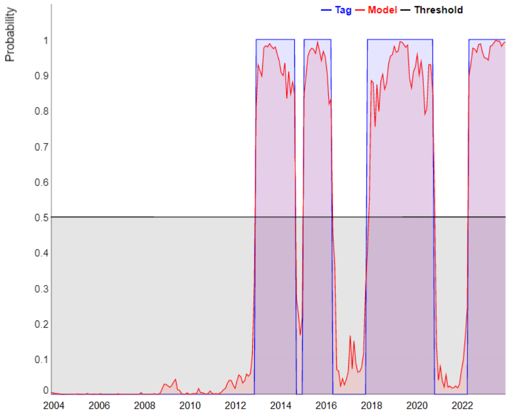

Figure: Labels, threshold and External Risk Index - RF model (trained with 76 variables). Scenario D' (12 months signalling without exchange rate controls and SBA.

Empirical findings: research question I

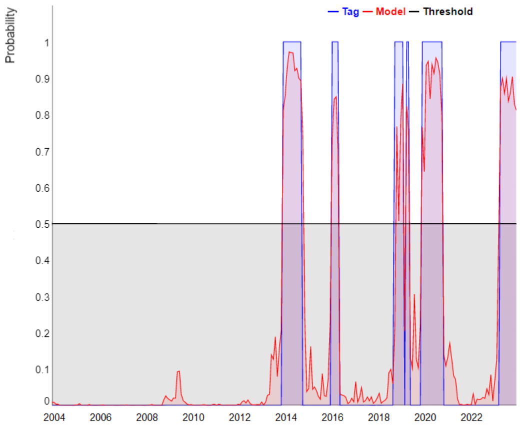

Figure: Labels, threshold and External Risk Index - RF model (trained with 76 variables). Scenario E' (1 month signalling without exchange rate controls and SBA).

Empirical findings: 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.
* 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, due to lower tree correlation, **RF has outperformed BA** when considering **short signalling periods**.

Empirical findings: research question II

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: 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.

Empirical findings: 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. 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'.

Empirical findings: Research question III (case A)

* **Case A**: including exchange rate controls, SBA and a 24-month signalling period.
* 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** and **nominality** beyond the thresholds signals an impending crisis within a 24-month time horizon.

Empirical findings: Research question III (case A)

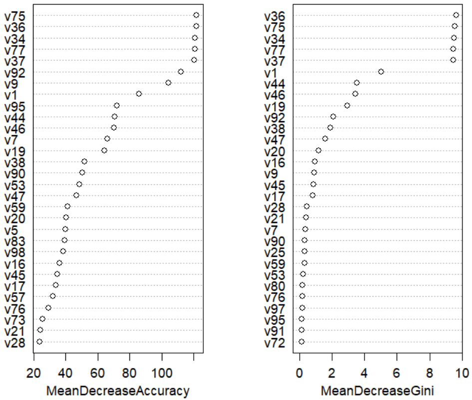

Figure: **RF variable importance**: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 52 (over 103). Trees: 100'000. OOB error: 3.37 per cent. Case: A.

Empirical findings: Research question III (case A)

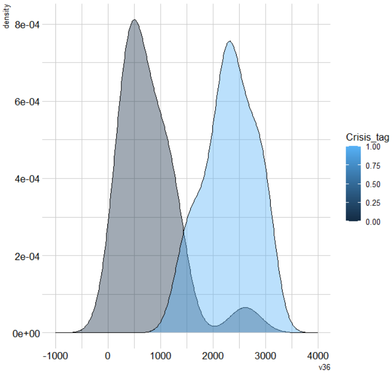

Figure: Non-parametric adjusted distributions of **public subsidies** to the private sector (v36) USD millions for NFPS under crisis and no-crisis labels in scenario A. Source: own based on BCRA and MEcon.

Empirical findings: Research question III (case A)

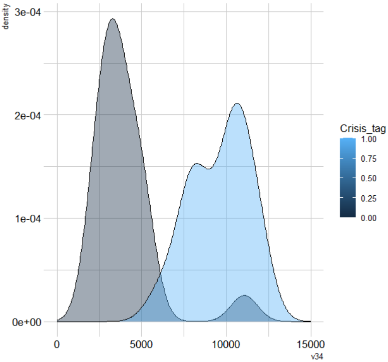

Figure: Non-parametric adjusted distributions of **total expenditure** (v34) in official USD millions for NFPS under crisis and no-crisis labels in scenario A. Source: own based on BCRA and MEcon.

Empirical findings: Research question III (case B)

* **Case B**: same as (A), but only 88 variables were considered, **excluding nominal variables**.
* Firstly, the real variable Monthly Economic Activity Estimator **EMAE** (v77) gains explanatory power.
* Secondly, **Compensation to Employees** (v38) 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.
* **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).

Empirical findings: Research question III (case B)

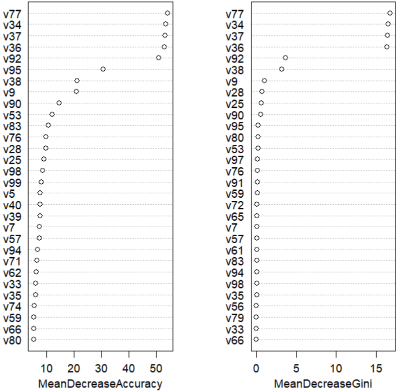

Figure: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 78 (over 88). Trees: 10'000. OOB error: 1.69 per cent. Case: B.

Empirical findings: Research question III (case B)

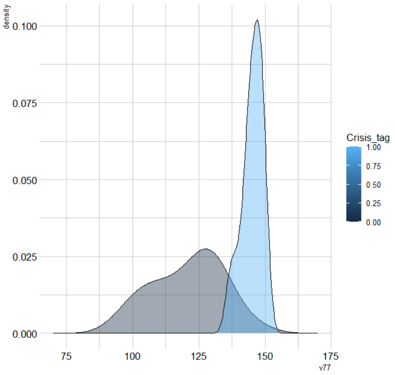

Figure: Non-parametric adjusted distributions of Monthly Economic Activity Estimator, **EMAE** index (v77) under crisis and no-crisis labels in scenario A. Source: own based on INDEC.

Empirical findings: Research question III (case B)

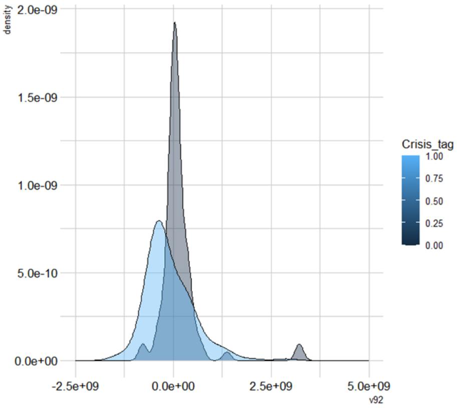

Figure: Non-parametric adjusted distributions of **Financial loans**, debt securities and lines of credit at the exchange market (v92) in USD under crisis and no-crisis labels in scenario A. Source: own based on BCRA.

Empirical findings: Research question III (case C)

* **Case C**: When **FX controls and 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 and v37), **interest** payments (v35), and the **budget deficit** (v40) are also important predictors.

Empirical findings: Research question III (case C)

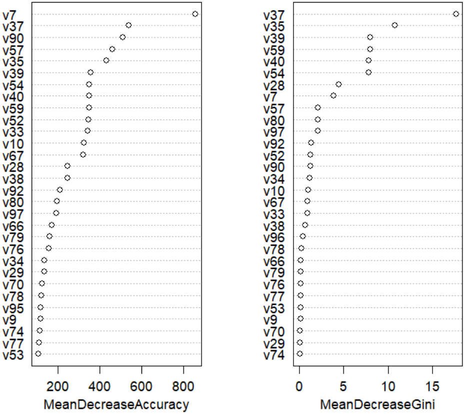

Figure: BA variable importance: Accuracy and Gini criteria. Variables: 88 (over 88). Trees: 1'000'000. OOB error: 3.37 per cent. Case: C.

Empirical findings: Research question III (case C)

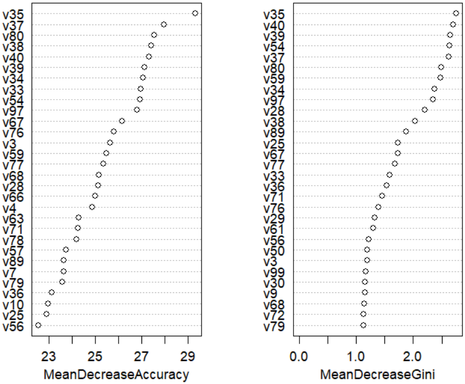

Figure: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 5 (over 88). Trees: 10'000. OOB error: 1.69 per cent. Case: C.

Empirical findings: Research question III (case C)

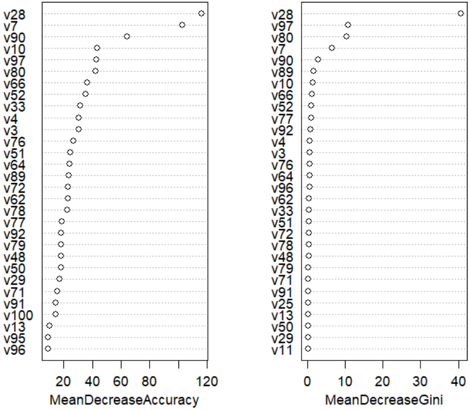

Figure: BA variable importance: Accuracy and Gini criteria. Variables: 76 (over 76). Trees: 10'000. OOB error: 5.06 per cent. Case: C.

Empirical findings: Research question III (case C)

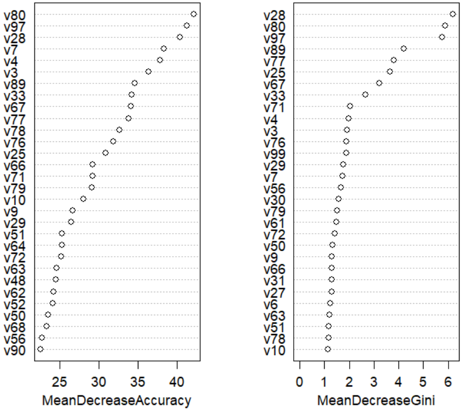

Figure: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 6 (over 76). Trees: 10'000. OOB error: 2.25 per cent. Case: C.

Empirical findings: Research question III (case C)

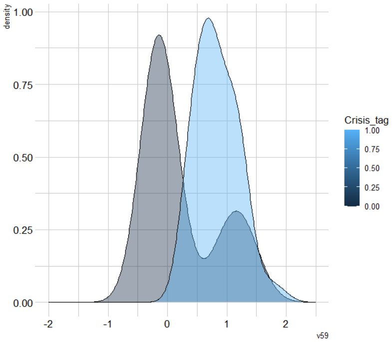

Figure: Non-parametric adjusted distributions of accumulated **Monetary Base expansion** due to net **foreign exchange purchases to the National Treasury** as measured by the stock of International Reserves (v59) under crisis and no-crisis labels in scenario C. Source: own based on BCRA.

Empirical findings: Research question III (case C)

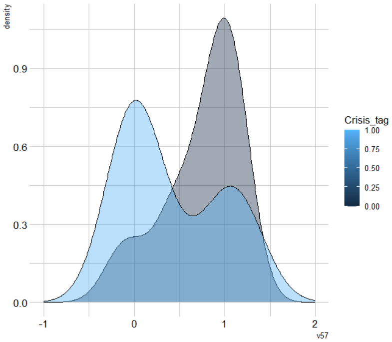

Figure: Non-parametric adjusted distributions of accumulated **Monetary Base expansion** due to Net **Foreign Exchange Purchases to the Private Sector** as measured by the stock of International Reserves (v57) under crisis and no-crisis labels in scenario C. Source: own based on BCRA.

Empirical findings: Research question III (case C)

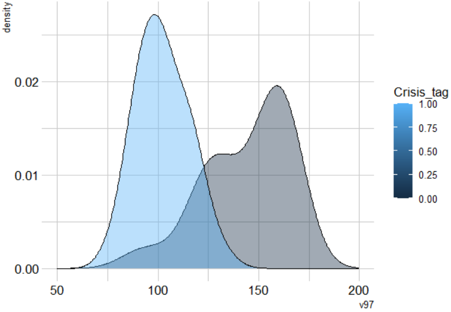

Figure: Non-parametric adjusted distributions of **Multilateral Real Exchange Rate** (v97) under crisis and no-crisis labels in scenario C. Source: BCRA.

Empirical findings: Research question III (case C)

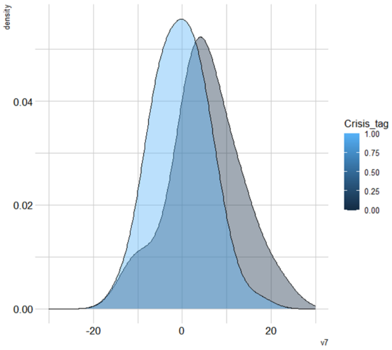

Figure: Non-parametric adjusted distributions of **Terms of Trade** year-over-year variation (v7) under crisis and no-crisis labels in scenario C. Source: own based on INDEC.

Empirical findings: Research question III (D-E)

* For **12-month signalling** period (under 76 variables, i.e. excluding 'fiscal-external' variables) consider the private sector **deposits in ARS** expressed in official USD (v67), year-on-year growth in **international reserves** (v86), **services trade balance** on the official exchange market (v89), **financial system return on equity** (v72), and the monthly output **EMAE gap** estimated as a 12-month mobile average (v78).
* For a **1-month signalling** period (under 88 variables) 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. It should be noted that **fiscal variables** (e.g. v38) play a secondary role in this scenario.

Empirical findings: Research question III (D)

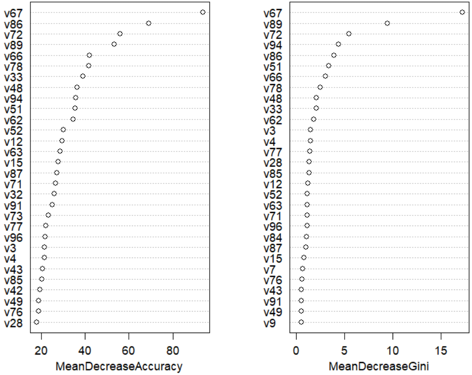

Figure: BA variable importance: Accuracy and Gini criteria. Variables: 76 (over 76). Trees: 10'000. OOB error: 3.37 per cent. Case: D.

Empirical findings: Research question III (D)

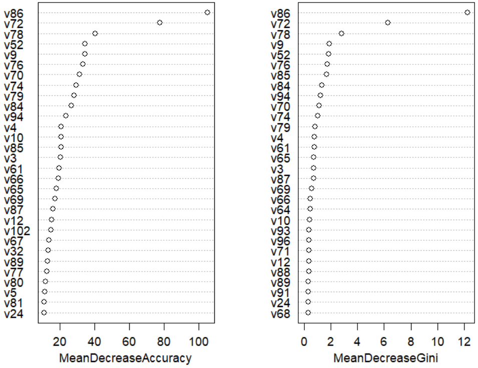

Figure: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 11 (over 76). Trees: 10'000. OOB error: 2.81 per cent. Case: D.

Empirical findings: Research question III (E)

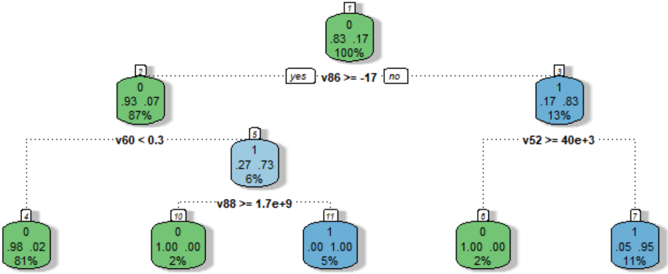

Figure: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.03333333. Selected number of splits: 4. Variables: 88. Case: E.

Empirical findings: Research question III (E)

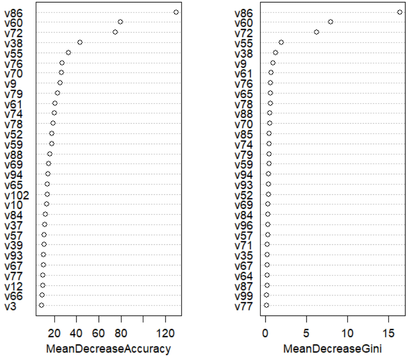

Figure: RF variable importance: Accuracy and Gini criterias. Cross-validation folds: 10. CV variables: 6 (over 88). Trees: 10'000. OOB error: 6.21 per cent. Case: E.

Empirical findings: Research question III (D)

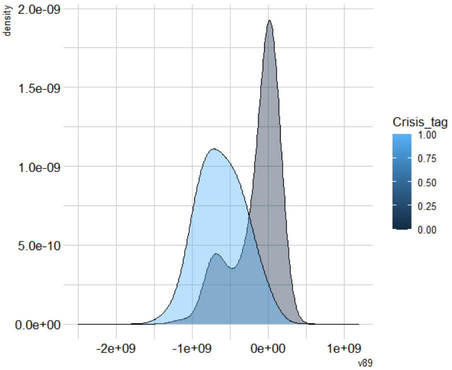

Figure: Non-parametric adjusted distributions of **Service Trade Balance** at the exchange market in USD (v89) under crisis and no-crisis labels. Scenario without exchange rate controls. Signalling period: 12 months. Source: own based on BCRA.

Empirical findings: Research question III (D)

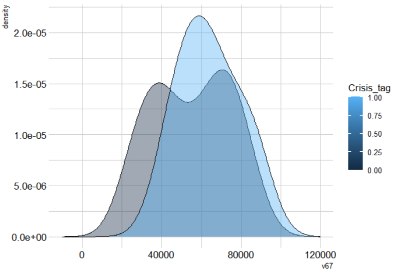

Figure: Non-parametric adjusted distributions of **ARS Deposits** from the private sector (v67) in official USD millions under crisis and no-crisis labels in scenario D. Source: own based on BCRA.

Empirical findings: Research question III (D-E)

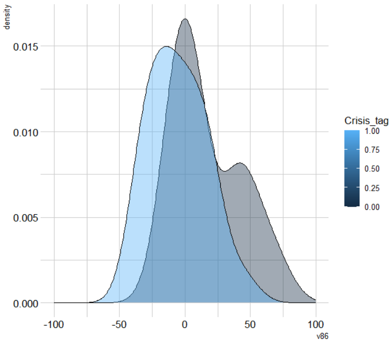

Figure: Non-parametric adjusted distributions of **International Reserve** year-over-year per cent growth (v86) under crisis and no-crisis labels in scenario E. Source: own based on BCRA.

Empirical findings: Research question III (D-E)

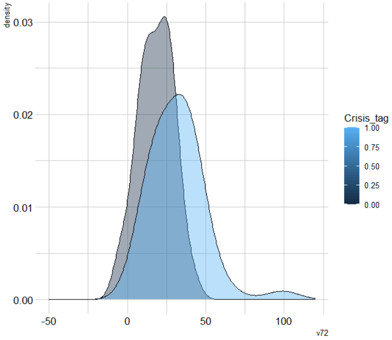

Figure: Non-parametric adjusted distributions of **Financial system Return** Over Equity in percentage (v72) under crisis and no-crisis labels in scenario E. Source: own based on BCRA.

Research question III (only 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 (only exchange rate control not classified as a crisis).

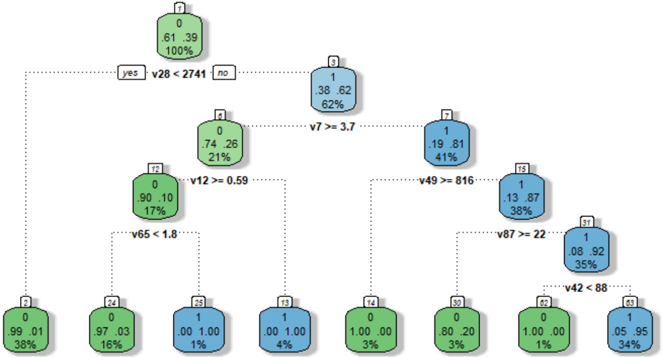

Figure: CART pruned tree. Criteria: Gini. Cross-validation folds: 10. Selected α: 0.01428571. Selected number of splits: 7. Variables: 75. Scenario without exchange rate controls. Signalling period: 12 months.

Research question III (only exchange rate control not classified as a crisis).

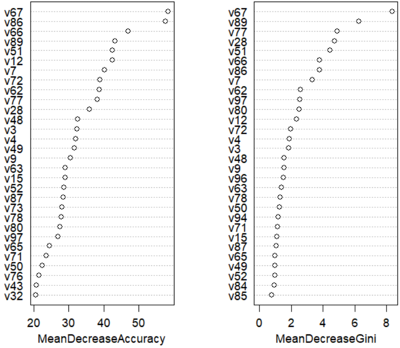

Figure: RF variable importance: Accuracy and Gini criteria. Cross-validation folds: 10. CV variables: 55 (over 75). Trees: 10'000. OOB error: 4.49 per cent. Scenario without exchange rate controls. Signalling period: 12 months.

Divergences with the IMF (2021)

* IMF(2021) is focused on a **panel of 192 countries**.
* 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.
* For example, while the IMF(2021) identified **international reserves** growth as the most important signalling variable, it is only significant for **short signalling periods** in the case of Argentina. Also, while **GDP** had a secondary importance in the IMF's emerging panel, it was a central signalling variable when the crisis is defined in a **broad sense**.
* Whether machine learning models significantly outperform traditional methods in a 'broad' scenario because of **non-linearities is unclear**. In contrast, in the case of Argentina, the IMF's perspective may be supported by 'narrow' crisis definitions. Further research is needed.

Variable importance IMF(2021)

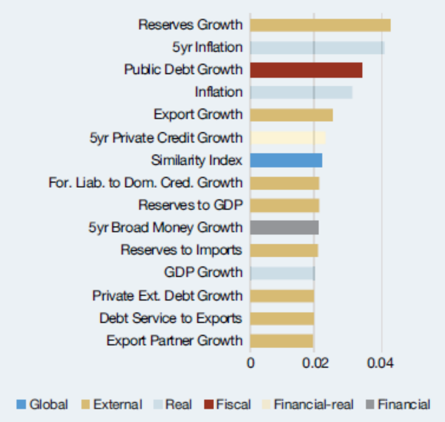

Figure: Variable importance of the IMF's RF model for emerging countries. Source: IMF(2021).

Variable importance IMF(2021)

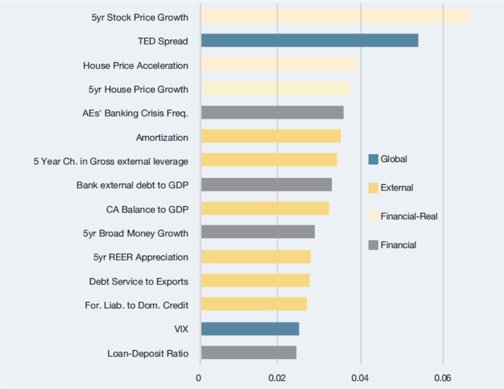

Figure: Variable importance of the IMF's model to detect episodes of 'sudden stops with growth impact' (SSGI). Source: IMF(2021).

Variable importance IMF(2021)

SSGIs are driven by variables from all sectors. Estimated on the full SSGI sample (1990-2017), the "winning" signal extraction model puts emphasis on asset bubble building and busting, reflected in stock price and housing price variables. In addition, global shocks and contagion effects (proxied by TED spread and the frequency of banking crises in AEs) play an important role, particularly in explaining GFC events. Country-specific external variables are also highly important: half of the top 15 are external variables, capturing external debt shocks and current account shocks.

Figure: Drivers of episodes of 'sudden stops with growth impact' (SSGI) according to IMF's signal extraction model. Source: IMF(2021), p. 22.

Robustness checks

* **Alternative labels**: Model performance was high for alternative labels. However, performance declined when there were more non-crisis observations than crisis observations. To address this issue in shorter time periods, it may be worth exploring the use of boosting and neural network models.
* **Alternative features**: i) the models were initially trained using a total of 103 features; ii) after this, nominal variables were excluded to test the results; iii) moving averages and accumulated variables that were commonly selected were also excluded; iv) finally, other variables that were usually selected were also excluded.
* **Alternative model specifications**: It is worth noting that the final results were not significantly affected by the choice of model specifications.

Extensions

* This research raises new questions and suggests potential avenues for future research.

1 **Unsupervised machine learning: cluster analysis**. However, it is still essential to exercise sound judgement in cluster analysis.

2 'Horse race' with **Boosting** and **Neural Network** models. However, it is important to note that boosting and neural networks require more computational and design effort and are often considered 'black boxes' compared to the RF algorithm.

3 'Horse race' with **traditional models**. 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'.

Extensions

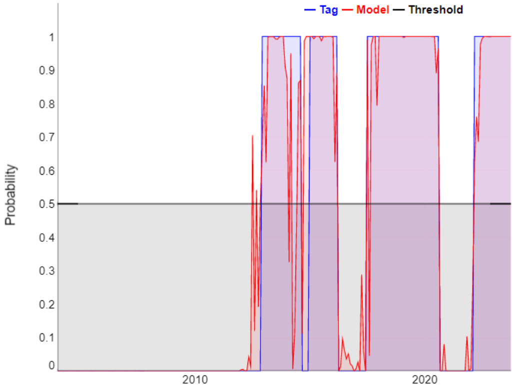

Figure: Labels, threshold and External Risk Index - **lasso penalised logit model**. Scenario 12-month signalling period without exchange rate controls.

Extensions

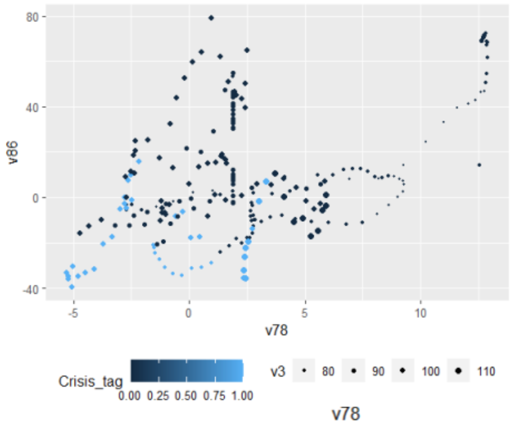

Figure: Example of non-linear relationships: Monthly output gap estimated as a 12-month mobile average (v78), International Reserve year-over-year growth (v86), DXY (v3), and crisis labels (1-month signal horizon), Argentina 2003-2023.

Conclusions

* The motivation behind this work was to provide a valuable **tool** for policymakers in Argentina by addressing a **relevant subject** in an **innovative way**.
* The study demonstrated 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 is **robust** to different labels and features. However, the **performance is lower** for narrower crisis definition and shorter signalling periods (neural network and boosting models a solution?).
* The CART, BA and RF models **performed similarly** for **broad** crisis definitions and **long** signalling periods. 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** (lower tree correlation) when considering **short signalling periods**.

Conclusions

* 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.
* **Contrary to the IMF (2021)**, for the case of Argentina, it is not clear that machine learning models significantly outperform traditional methods in 'broad' crisis scenarios - as defined by the IMF (2021) - due to **non-linearities** (given that in most cases one variable could explain most of the classification problem). Conversely, non-linearities and the higher performance of ML models can be seen in 'narrow' (non-IMF) crisis definitions. However, this question is beyond the scope of this thesis.

Conclusions

* **Contrary to my 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.
* **A lesson in humility for economists?** Economists frequently use their theoretical framework to predict the viability of a 'base case scenario'. In a discussion among economists with different theoretical backgrounds about the sustainability of the current scenario, which of them would consider checking the ROE of the financial system as a potential predictor? This task is performed by machines. As the joke goes, economists have predicted 10 of the last 5 crises. In contrast, **machines learn from empirical evidence and classify with flawless performance**.
* **Machine learning models do not assume the ability to resolve theoretical disputes**. ML models do not claim to establish causality but only to identify 'signalling variables'.

Gracias

References

* 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.
* Breiman, L. (1984). Classification and regression trees. Routledge.
* Breiman, L. (2001). Random forests. Machine learning, 45:5–32.
* der Aalst, W. V. (2017). Hitting the bulls-eye with a low bias and low variance. https://twitter.com/wvdaalst/status.
* Fortmann-Roe, S. (2012). Understanding the bias-variance tradeoff. URL: http://scott. fortmann-roe. com.
* IMF (2017). Assessing country risk: Selected approaches. Ahuja, M. A., Wiseman, K., and Syed, M. M. H. International Monetary Fund.

References

* 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/ biasand- variance- trade-off-542b57ac7ff4.
* James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). An introduction to statistical learning, volume 112. Springer.
* 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.
* Keynes, J. M. (1919). The economic consequences of the peace. Routledge.
* 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.