TO SUPERVISE OR TO SELF-SUPERVISE: ARE WE READY FOR THIS CHALLENGE?

A machine learning based comparison on credit supervision

ABSTRACT

This study investigates the need for credit supervision as conducted by the Central Bank of Brazil (CBB). It builds on a real bank on-site credit examination to compare the performance of a hypothetical self-supervision approach, in which banks themselves assess their loan portfolios without external intervention, with CBB's on-site banking supervision approach. The experiment trains three different machine learning algorithms to develop two different samples: the first one based on good and bad ratings informed by banks, and the second one based on past on-site credit portfolio examinations conducted by CBB's banking supervision. The findings show that overall performance of on-site supervision approach is consistently higher than the self-supervision approach, which justifies the need for on-site credit portfolio examination, as conducted by CBB. The study also argues that the poor performance of self-supervision approach derives from the use of loan loss provision to overcome an accounting-economic mismatch.

Keywords: Bank supervision; Machine learning; Loan loss provisions; On-site credit supervision.

JEL Classification: C45, G21, G28, M48

Tema 1: Inovação nos mercados financeiros e de capitais

1. Introduction

The need for banking supervision is the result of a chain of events triggered by the market failure that results in the financial system itself. Informational asymmetry between economic units makes allocation of resources inefficient without a hub to connect them. Hence, the financial intermediary is the prerequisite for financial intermediation. Having overcome the first market failure gives rise to the second one: principal-agent problem. Intermediating means capturing someone else's deposits and directing it to a third party at the intermediary's will. From the depositor's point of view, the sounder the bank, the safer the deposits. However, that may not be the case from the management's point of view, who can decide for a riskier, thus profitable, path. This environment can impede the alignment of interests between depositors and management, making financial intermediation inefficient without the presence of an independent external agent, namely banking supervision, which asserts the solvency of intermediaries.

The purpose of banking supervision is to keep the financial system sound and safe, ensuring that financial regulation, the set of rules that govern the financial system, is followed (Masciandaro and Quintyn, 2015)[1]. The flagship of financial regulation is the Basel Accords, policy directives prepared by the Basel Committee on Banking Supervision (BCBS), the high-level committee of the Bank of International Settlements (BIS), and adopted worldwide. The third Basel Accord, which emerged in the aftermath of the Great Financial Crisis (2008/2009), broadened the scope of prudential regulation and embraced liquidity and leverage as relevant

[1] Though complementary, banking regulation and banking supervision are separate activities, usually performed by different actors. The former concerns the rules governing the financial system, whereas the latter regards the enforcement of such rules (Masciandaro and Quintyn, 2015). In the Brazilian financial system, the National Monetary Council is responsible for banking regulation and the Central Bank of Brazil (CBB) is responsible for banking supervision.

microprudential issues. However, the solvency-based perspective remains as the focus of prudential regulation, highlighting the capital adequacy ratio as its leading indicator.

From the solvency perspective, keeping the financial system sound and safe means asserting what the bank's assets worth (Hellwig, 2014). In particular, the credit portfolio assessment, due to its relevance among assets, is an important task assigned to banking supervision. From the accounting standpoint, the credit portfolio is often measured by amortized cost deduced by loan loss provision. Loan loss provision is a combination of incurred and expected losses and is an accounting device designed to adjust credit portfolio value to its fair value. The role of banking supervision is to assess loan portfolios and check whether banks comply with rules and regulatory requirements, especially the adequacy of loan loss provision to the loan portfolio risk profile.

Although credit portfolio assessment is a classic banking supervision predicate, it is also true that the Great Financial Crisis interrupted a self-regulation process that gradually increased the reach of internal based models, allowing banks to replace regulatory standard models with proprietary versions developed internally. Continuous innovation in the financial system, brought about by technology revolution, may suggest this process can be reignited in the spirit of Stefanadis (2003). De Chiara, Livio and Ponce (2018) analyze the effect of tighter regulation and powerful supervision in the financial sector and the consequent social costs. The authors argue that the optimal supervisory architecture combines a supervisory regime where direct assessment by a supervisor is always required (Mandatory Supervision) with a Flexible Supervision regime where banks self-select the regulatory contract designed for their level of risk.

In this sense, this study investigates the need for credit supervision as conducted by Central Bank of Brazil (CBB). It builds on a case study to compare the performance of a hypothetical self-supervision approach, where banks themselves assess their loan portfolios without external intervention, with CBB's on-site banking supervision. To conduct this experiment, we used the proceeds of real case on-site credit portfolio examination to compare the performance of two different machine learning sampling approaches: the first one based on good and bad ratings informed by banks, and the second one based on past on-site loan portfolio examinations conducted by CBB's banking supervision.

The findings show that CBB's on-site supervision consistently outperforms the self-supervision approach, which justifies the necessity of on-site credit portfolio examination, as conducted by CBB. The study also argues that the poor performance of the self-supervision approach derives from the use of loan loss provision to overcome an accounting-economic mismatch.

The next section discusses the related literature on financial supervision and loan loss provisioning. Section 3 presents the empirical analysis comprising: (i) the machine learning algorithms used to develop sampling models based on on-site supervision and banks' experience; (ii) on-site examination procedure that produced the ground truth against which both supervisory approaches are compared; (iii) the analysis of the results. Section 4 concludes.

2. Banking supervision and loan loss provisioning regulation

The financial crisis casted doubts over policy certainties ranging from monetary policy to financial regulation and supervision. Barth *et al.* (2013) and Blanchard (2008) argue that the crisis was the result not only of incomplete regulation but also of ineffective supervision. Bernanke (2010) *apud* Tressel and Verdier (2014) ascertained that, based on evidence of declining lending standards during the boom, stronger regulation and supervision aimed at problems with underwriting practices and lender's risk management would have been a more effective and surgical approach to constraining the housing bubble than a general increase in interest rates.

The influence financial supervision exerts on bank's risk-taking is considered relevant by an extensive amount of studies. However, results are mixed when it comes to the effects of supervision on financial stability. Bhattacharya *et al.* (2002) conclude that intense supervision can improve the timeliness of supervisory intervention, while Delis and Staikouras (2011) show that intense supervision can limit banks' risk-taking. White (2006) defends that the best instruments to achieve financial stability are supervision and regulation, while Barth *et al.* (2004, 2008 and 2013) argue that the efficiency of financial intermediation, hence financial performance, is reduced by financial supervision.

In the Brazilian financial system, henceforth financial system, different types of financial institutions coexist, ranging from niche institutions, which explore specific types of activities, to universal banks, which gather many different activities in the same entity. The financial system is complex and well developed. In June/2019, it comprises 178 banks, mounting to 126% of GDP in assets, and 47% of GDP in credit[2], which makes Brazil an interesting case study. National Monetary Council (NMC) is the financial regulator and, differently from other jurisdictions, CBB is responsible for all aspects regarding financial institutions oversight, from entry to resolution, concordantly with Barth et al (2004) public interest view.

In Brazil, the supervisory process follows partially the Twin Peaks model (Group of Thirty, 2008; FSI, 2018), which recommends supervisory specialization by objectives: prudential monitoring of regulated institutions and oversight of business conduct. Though Twin Peaks model envisages two separate financial supervision authorities to tackle banking supervision,

[2] Data collected from the CBB website financial series repository: https://www3.bcb.gov.br/ifdata/

the Brazilian solution is a hybrid model, where an integrated supervisor, namely CBB, holds both objectives inside the same authority.

Prudential regulation (henceforth banking supervision) is the focus of our analysis. The objective of banking supervision is to assess the soundness of financial institutions, mainly commercial banks, and to assert that regulation is complied. It consists of two cornerstones: examination, or on-site supervision, and monitoring, or off-site supervision. On-site supervision follows a supervision cycle and involves sending supervisory staff to banks to conduct specific examinations. Off-site supervision is a permanent process that analyzes bank's performance and compliance to regulation based on multiple sources of data, as well as the outcomes of on-site supervision.

Brazilian financial regulator, the National Monetary Council (NMC), still has not adopted IFRS 9 as loan loss provisioning regulation for the financial system. To date, NMC resolution 2682/99 (NMC, 1999) defines loan loss provisioning regulation. It combines expected loss and incurred loss approaches in the same framework. Accordingly, financial intermediaries are bound to assign an individual rating to each credit operation booked in the loan portfolio. As presented in Table 1, there are nine different ratings, which mirror minimum and maximum provisions as percentage points of loans amount due. Whether the credit is due or past due defines the way ratings are assigned. For due credits, banks apply the expected loss approach, in which they assign ratings as they find best, as long as based on consistent credit risk assessment. The expected loss approach assigns ratings compatible to the loss banks expect to face in each credit operation along its lifetime. However, when a credit is past due, the incurred loss approach steps in and banks are deemed to assign ratings compatible to the extension of the delinquency, as determined by the regulation (see Table 1 for more detail).

Table 1. Ratings, provision and delinquency in Brazilian financial regulation

Ratings	Provison (% of the amount due)	Delinquency (days)
AA	< 0.5%	-
A	>=0.5%; <1%	-
B	>=1%; <3%	>= 15; <30
C	>=3%; <10%	>=30; <60
D	>=10%; <30%	>=60; <90
E	>=30%; <50%	>=90; <120
F	>=50%; <70%	>=120; <150
G	>=70%; <100%	>=150; <180
H	100%	>=180

Besides combining expected loss and incurred loss approaches in the same framework, Brazilian loan loss provisioning (LLP) regulation also differs from IFRS 9 for it does not follow the ED*PD*LGD rationale (LLP as the product of exposure at default, probability of default and loss-given-default) when computing incurred loss. According to NMC resolution 2682 (NMC, 1999), the amount of LLP determined for credits past due is just the product of the percentages presented in Table 1 and the amount due. Consequently, in case the credit is past due, the LLP regulation does not take into consideration the collateral that underlies the credit and hedges it. It assumes that the loss-given-default is 100%, which turns the percentages in Table 1 into probabilities of default. To make the point clearer, the probability of default of a credit 180-day past due is 100%, so provision equals the amount due.

In June 2019, the amount of loan loss provisions (LLP) in the Brazilian financial system equaled to 6.22% of credit portfolios and 18.2% of equity, reflecting the relevance of credit activity, hence loan portfolios, to the financial system[3]. The larger the loan portfolio, the more vulnerable banks are to an increase in loan default arising from deteriorating economic conditions (Laeven and Majnoni, 2003). Therefore, monitoring and supervising LLP is a crucial microprudential surveillance tool that bank supervisors use to assess banks' loan portfolio quality (Ozili and Outa, 2017).

[3] Data collected from the CBB website financial series repository: https://www3.bcb.gov.br/ifdata/

On-site prudential credit supervision works out under two different perspectives, namely credit management and credit risk. Credit management inspections focus on credit processes and compliance of internal credit policies to credit regulation and good practices. As for credit risk inspections, the objective is to assess the quality of credit portfolios and sufficiency of LLP. Banks that present underprovisioned loan portfolios are demanded to increase provisions in order to match their loan portfolios' risk.

In Brazil, on-site credit risk supervision as so conducted in banks is centered at the borrower's financial performance. Therefore, sampling procedures, as well as assessment of sampled borrowers' risk quality, involve intense cash flow analysis, in the spirit of Antunes *et al.* (2017) and Antunes *et al.* (2018).

From loan portfolio information banks file monthly at CBB's credit bureau repository, it is possible to derive elementary cash flow variables, such as expected cash flows, received cash flows and disbursed cash flows. Those variables are calculated at loan-level and on monthly basis. Since analysis is focused on borrowers, loan-level cash flow variables are aggregated and turned into borrower-level cash flow variables. Then, the following step is to calculate borrowers' financial performance indices, such as borrower's cash performance (BCP) and borrower's liquidity performance (BLP). These indices are calculated considering a six-month period before the starting date of analysis. Equations (1) and (2) below present indices' formulae.

$$BCP = \frac{\sum_{t_{-5}}^{t_0} Net\ Received\ Cash\ Flows}{\sum_{t_{-5}}^{t_0} Expected\ Cash\ Flows} \tag{1}$$

$$BLP = \frac{\sum_{t_{-5}}^{t_0} Net\ Received\ Cash\ Flows}{Maximum\ Borrower's\ Loan\ Portfolio} \tag{2}$$

Cash flow based sampling procedure selects borrowers in accordance with cash flow indices (1) and (2). Once sampled, on-site credit risk examination assesses borrowers in order to confirm the bad credit risk suggested and a possible LLP insufficiency. Borrowers presenting credits over 90-day past due are considered to be bad and the provision assigned by the bank is compared to regulation disposals.

This study investigates the need for credit supervision as conducted by CBB. It turns a natural experiment into a case study to compare the performance of CBB on-site supervision with a hypothetical self-supervision, where banks themselves assess their loan portfolios without external intervention. To conduct this experiment, we used results of a real case on-site credit portfolio examination to compare the performance of two different machine learning sampling approaches: the first one based on good and bad ratings informed by banks, and the second one based on past on-site loan portfolio examinations conducted by CBB's banking supervision.

3. Empirical analysis

Technology revolution reached the financial system. Although the extension and depth of its effects in the way business is done is yet to be fully realized, the only certainty is that business will not be as usual anymore. Alongside with the emergence of new entrants in the financial system, supervisory policymakers and standard setters around the world draw attention to risks and opportunities to financial stability. Technology-enabled innovation in financial services (FinTech) develops rapidly and demands a continuous assessment of the adequacy of regulatory frameworks (FSB, 2017).

Financial supervision gradually absorbs innovative technology approaches and the terms regtech and suptech were coined to capture a series of initiatives that use innovative technologies in financial supervision domain (FSI, 2018). While regtech accounts for innovative technologies used in support of compliance with financial regulation, suptech refers to the conduct of financial supervision underpinned by innovative technologies.

Concerning financial supervision, artificial intelligence techniques, mainly those involving machine learning algorithms, are the most used (FSI, 2018). Samuel (1959) defines machine learning as the *field of study that gives computers the ability to learn without being explicitly programmed*. In general, machine learning deals with (automated) optimization, prediction, and categorization, not with causal inference. In other words, classifying whether the borrower is a good or bad credit risk is a machine learning task. However, determining what factors drive the credit quality is not likely to be a machine learning challenge (FSB, 2017_A).

The different categories of machine learning algorithms relate to the extent of the human intervention required. In supervised learning, the algorithm receives a set of training data that contains labels that classify the observations. Contrarily, unsupervised learning detects patterns in the data through similar underlying characteristics, making labels needless. Two additional categories of machine learning algorithms fall in between supervised and unsupervised learning: reinforcement learning and deep learning. The former resorts to feedbacks that help the algorithm to learn. The latter works in layers inspired by the human brain, the reason why it is also known as artificial neural networks[4].

3.1 Methodology

[4] Comments on machine learning categories are limited to the purpose of the study. For additional information see FSB (2017_A).

We draw attention to the supervised learning algorithms used in this study, among the many different algorithms available in literature. The first statistic method used is random forest (RF), introduced by Breiman (2001) as an extension of the decision trees method (Breiman et al., 1984). Random forest consists of a large number of decision trees that operate as an ensemble. Each individual tree in the random forest is made of successive splits of the sample into two leaves, according to a single exogenous variable exceeding or not a threshold. The quality of each split is measured at the node by an impurity function, such as entropy or information gain. Each tree defines a class prediction, which equals to one vote. The most voted class is the model prediction.

The second statistic method used is extreme gradient-boosted-trees method (XG) introduced by Schapire (1990), who describes it as a method for *converting a weak learning algorithm into one that achieves arbitrarily high accuracy*. As Hastie et al. (2009) point out, the method works by sequentially applying weak learners to repeatedly re-weighted versions of the training data. Krauss, Do and Huck (2016) so explain the method:

> *After each boosting iteration, misclassified examples have their weights increased, and correctly classified examples their weights decreased. Hence, each successive classifier focuses on examples that have been hard to classify in the previous steps. After a number of iterations, the predictions of the series of weak classifiers are combined by a weighted majority vote into a final prediction.*

The third statistical method used is artificial neural network (ANN), which is the result of many ideas combined (Rosenblatt, 1958; Kelley, 1960; Bryson, 1961; Werbos, 1975; Schmidhuber, 1992; Hinton, 2006). ANN consists of a structure of neurons (also known as nodes or units) displayed in layers, namely input layer, hidden layers and output layer. The first layer, the input layer, receives the matrix of features. So, the number of neurons is equal to the number of features in the dataset. The last layer is the output layer and it holds the outcome of the model (good or bad borrower, "0" or "1"), the reason why a binary problem demands only one neuron in the output layer. In between the input and the output layers, there are hidden layers (whether there is only one hidden layer, the ANN is called a shallow learning; in case there are many hidden layers, the ANN is called deep learning).

Information fed in the input layer passes along to the hidden layer. All of the 26 features of each observation in the dataset is inputted in each neuron of the hidden layer, along with a random weight. Inside the neuron, all of the weighted features are summed up and then the neuron decides, based on a previously chosen activation function, if the signal is passed over to the next layer, where the same procedure is repeated, until the output layer. Reaching the output layer, the output value is compared to the actual value provided by the training set. This comparison triggers a loss-function, which minimization is the learning process. In order to minimize the cost-function, the result of the comparison between the output and actual values back propagates in the network and adjusts the weights assigned to input values, restarting the process. The new value obtained for the cost-function triggers another reassignment of weights. An epoch comprises a complete cycle and the number of epochs used in an ANN procedure is previously determined.

The study applies these three methods as classification devices to compare the performance of the self-supervision approach, built upon bank's experience, and CBB's on-site credit risk supervision approach. However, before applying a machine learning algorithm, one must train it on a dataset with known outcomes, namely a labeled training set. Therefore, to turn a machine learning algorithm into a classification device, the steps presented in Table 2 shall be followed:

Table 2 – Building a machine learning based classification device

Steps	Description
1	Define the exogenous variables that compound the datasets, known as the matrix of features.
2	Define the borrowers from which the matrices of features will be built and whose labels (good or bad borrowers, the endogenous variable) are known. In this study, we use two different sets of borrowers, belonging to the two supervisory approaches investigated.
3	Build the two datasets that will be used to train the algorithms, according to the two supervisory approaches analyzed.
4	Run (train) the algorithms on the datasets and evaluate their performance.
5	Build the validation set from the real case credit portfolio to be classified.
6	Apply the trained algorithms to the validation set and compare the outcomes.

Following the sequence presented in Table 2, the first step is to choose the exogenous variables to make up the datasets. Instead of adding every information about borrowers available in CBB's databases, which would lead us to a matrix of features with hundreds of variables and a computational consuming process, we opted by a parsimonious approach. We applied the experience of years of on-site supervision to choose which variables better inform about the quality of a borrower. In other words, we developed 26 proxies that reflect on-site experience in classifying good and bad borrowers (Table A.1 in Appendix A describes the matrix of features employed in the study).

The next stage is to choose the labeled borrowers whose data will form the datasets. The labeled borrowers are the endogenous variable of the datasets. In particular, "1" is assigned to bad borrowers and "0" is assigned to good borrowers. The first set of borrowers comprises 6581 samples of good and bad borrowers (5483 good and 1098 bad) derived from 12 previous on-site credit portfolio examinations conducted by CBB's banking supervision from 2015 to 2018. The second set of borrowers gathers 1.012.234 samples of good and bad borrowers (961544 good and 50690 bad) obtained from banks' experience and extracted from credit risk information banks file monthly in CBB's repositories.

As for the dataset built from banks' experience, Table 3 presents the parameters used to select borrowers, as well as the criteria applied to label them as good or bad.

Table 3 – Parameters and criteria adopted to label borrowers according to banks' experience

Parameters	Description
Banks	20 biggest loan portfolios in financial system at the starting date (june/2019)
Dates for extraction	December/ 2016; December, 2017; December 2018
Good borrowers	Rated as AA, A or B (LLP inferior to 3%) in each one of the last seven months previously to the dates chosen for extractions
Bad borrowers	Rated as F, G or H (LLP equal or higher than 50%) in each one of the last three months previously to the dates chosen for extractions. A bad borrower was excluded from the dataset if, in the six-month period after extraction dates: (i) the debt was paid; (ii) the debt was reduced; (iii) the rating assigned by the bank improved.
Materiality	Loans in excess of R$ 200.000[5]

After these preliminary steps, datasets are gathered through the selection of the variables that constitute the matrix of features for each of the labeled borrowers. In other words, the datasets are the merging of the endogenous variable and the exogenous variables. These datasets are used to train the algorithms.

The training procedure is to apply machine learning algorithms to the datasets. That allows algorithms to combine the matrix of features (26 fields of information about each borrower) and the labels to learn the general rule of classification to predict labels in any other out-of-sample dataset. When running the training, the dataset is split into two subsets, the training set and the test set. Following a usual rule of thumb, we assigned 70% of the dataset to form the training set and the remaining 30% to the test set.

We applied three different algorithms to the datasets: random forest (RF), extreme-gradient-boosted trees (XG) and artificial neural network (ANN) (Table A.2 in Appendix A details the settings used to tune the algorithms). The algorithms are coded in Python and run on the following packages: scikit-learn (RF), xgboost (XG) and Keras (ANN). All of the algorithms were subject to regularization procedures and K-fold cross validation where appropriate. Since

[5] Approximately US$ 40000.

the datasets are quite homogeneous, there is not much difference between the training sets and the test sets. Therefore, independently of the results obtained in the training phase, there is no guarantee that trained models will perform properly in an out-of-sample dataset.

Having trained three different models to classify good and bad borrowers, according to the supervisory approaches under comparison, the next step is to build a validation set, an out-of-sample dataset, and apply the trained models to it. Differently from the previous datasets, the validation set is not labeled, i.e., the endogenous variable is unknown. The role of the trained models is to apply the general rule of classification learned from on-site supervision previous experience, as well as banks' experience, and classify the validation set in good and bad borrowers.

The validation set is a real bank credit portfolio comprising 1338 borrowers, with a minimum amount due of R$ 10.000[6]. In order to establish a common ground truth against which the performance of both supervisory approaches can be assessed, the other front of analysis involves the mapping of good and bad borrowers in the validation set through an on-site credit examination. After excluding all borrowers already rated as "H" by the bank, i.e., 100% provisioned, on-site examination concluded that 1279 borrowers were considered to be good ("0") and 59, bad ("1"). Assuming that the results of on-site examination are the correct classification, i.e., the ground truth, is central to the analysis. Thus, to evaluate the supervisory approaches is just a matter of matching results.

As for the criteria used by on-site credit supervision to identify bad loans, it is rather straightforward. Whenever a credit is over 90-day past due (rating "E", or worse), it is considered to be a bad credit, hence it is labeled as "1", otherwise, "0". However, it is common

[6] Approximately U$ 2000

to find evergreened credits, i.e., credits artificially kept under the 90-day past due threshold through successive rollovers. Another practice used to evergreen credits is to distribute the expected cash flow asymmetrically. In other words, small installments, smaller than the interest accrued, are concentrated at the beginning of the credit cash flow, while principal and the remaining interest are placed long in the future. That makes the credit easy to be paid, though artificially. In both cases, the effect of these practices is disregarded and the borrowers are considered to be bad, thus labeled as “1”.

3.2 Results Analysis

If the comparison proves the self-supervision approach built upon bank’s experience outperforms CBB’s on-site credit risk supervision, there would be a strong argument in favor of revising the scope assigned to on-site credit risk supervision. Therefore, the last step is to compare the performance of on-site banking supervision and the hypothetical self-supervision approaches against the ground truth provided by on-site examination results.

As discussed before, the role assigned to on-site credit supervision is to detect bad borrowers classified as good ones and to quantify the consequent amount of insufficient loan loss provisions. Therefore, picking up bad borrowers is central to the analysis, which makes type-2 errors, i.e., classifying bad borrowers as good ones, much worse than type 1 errors, for good borrowers, even those mistakenly classified as such, are not revised during an on-site credit examination. Thus, from the supervisory standpoint, minimizing type-2 errors is crucial, even if the cost is maximizing type 1 errors, since these cases are revised and dumped during examination.

Another aspect to highlight is that the distribution of good and bad borrowers in credit portfolios is heavily unbalanced, for there are usually many more good borrowers than bad ones.

Accordingly, some measures used to assess machine learning algorithms performance may present the false sense of efficiency. The real bank case study under analysis proves the point. From the 1338 different borrowers portfolio, only 59, or 4.6% of the total, are classified as bad borrowers by on-site supervision staff, while 1279, 95.4% of the total, are considered to be good borrowers. An algorithm that classifies the whole portfolio as good borrowers is 95.4% accurate, even failing to catch a single bad borrower.

Tables 4 and 5 present the confusion matrix for both supervisory approaches, while Table 6 presents the efficiency measures. The results of the three algorithms used (RF, XG and ANN) are consolidated as one single result, in which every borrower labeled as “1” by any of the three algorithms is considered to be a bad borrower. Therefore, the consolidated result used to compare the two supervisory approaches is the aggregation of the three algorithms used.

The confusion matrix is a performance measurement device for machine learning classification algorithms. It combines actual and predicted values to produce the elementary outcomes, which allows one to compute the efficiency metrics used to assess performance. For this study consists of a binary classification problem, four outcomes are derived from the confusion matrix, namely True Positive (TP), True Negative (TN), False Positive (FP, also type 1 error) and False Negative (FN, also type-2 error). Tables A.3 and A.4 in Appendix A provides more detail on the basics of the confusion matrix and efficiency metrics.

From the confusion matrices, one can notice that the self-supervised approach sampled less borrowers (35) than on-site supervision approach (77), which is positive from the efficiency standpoint, for it demands less work hours to examine the loan portfolio. However, efficiency comes at a cost, for the narrower the sample, the harder it is to minimize type-2 error. On this matter, from the 59 bad borrowers in the loan portfolio, on-site supervision correctly classified 40, which results in a true positive rate of 0.68. As for the type-2 error, the approach failed to identify 19 out of 59 bad borrowers, leading to a false negative rate of 0.32. As for the self-supervised approach, only 22 bad borrowers are correctly classified, a true positive rate of 0.37 and a type-2 error of 0.63, which evidences the 37 bad borrowers it failed to identify.

For the sake of completeness, Table 6 also displays other performance measures. However, due to specificities of the borrowers' classification problem addressed in the study, the information they convey is minor. For the number of bad borrowers sampled is small, comments on the false positive rate are not relevant. Similarly, the heavily unbalanced distribution of good and bad borrowers in loan portfolios makes accuracy a fragile indicator. Regarding the precision measure, it focuses solely on the correctly classified bad borrowers and does not take into consideration false negatives, which, as commented before, is crucial for credit supervision. Therefore, though self-supervision approach presents higher precision, that does not mean much.

F1 score combines precision and recall (true positive rate) in the same measure. Hence, it informs how precise the classifier is, as well as how robust it is. The greater the F1 score, the better the performance. Though less precise, for it produces more false positives, on-site supervision approach presents a much better recall than the self-supervision approach, as the number of false negatives is smaller. Consequently, on-site supervision approach shows a better F1 score, i.e., a better performance, than self-supervision approach.

In brief, overall efficiency of CBB's supervisory approach is higher and the number of bad borrowers not identified by the self-supervised approach, the type-2 error, is nearly twice as big as CBB's approach. Apart from moral hazard issues, which do not belong to the scope of this analysis, the results of the self-supervision approach could be worse in the absence of CBB's on-site supervision, since ratings "F", "G" and "H" banks' assign to their credit portfolios are sometimes imposed by CBB's supervision.

Table 4: Confusion Matrix - on-site supervision approach

		Actual Values	
		Bad Borrower Positive (1)	Good Borrower Negative (0)
Predicted Values	Bad Borrower Positive (1)	40	37
	Good Borrower Negative (0)	19	1242

Table 5: Confusion Matrix – self-supervision approach

		Actual Values	
		Bad Borrower Positive (1)	Good Borrower Negative (0)
Predicted Values	Bad Borrower Positive (1)	22	13
	Good Borrower Negative (0)	37	1266

Table 6: Performance comparison – On-site supervision and self-supervision approaches

Efficiency Metrics	**On-site supervision approach**	**Self-supervision approach**
True Positive rate (Recall)	0.678	0.373
False Negative Rate	0.322	0.627
False Positive rate	0.029	0.010
Accuracy	0.958	0.963
Precision	0.519	0.629
F1 Score	0.588	0.468

3.3 Further analysis

Financial intermediation is the reason why a financial system is established and credit is the classic financial intermediary activity. Therefore, one cannot deny banks' expertise to grant credit and classify borrowers. However, in this study, expertise in doing credit does not translate into adequate borrowers' risk classification, for the performance of self-supervised approach is below expectations. A possible explanation is the presence of incentives that distort credit risk classification, in order to use loan loss provisions to meet other objectives.

Regarding the incentives that affect credit risk classification, hence, the amount of loan loss provisions constituted, literature is extensive on the subject. Ozili and Outa (2017) present a broad overview on bank loan loss provisions (LLP). Among the issues in which LLP is involved, we highlight the following: (i) the expectation component in provisioning behavior during business cycles and crisis periods (Laeven & Majnoni, 2003; El Sood, 2012; Agenor and Zilberman, 2015); (ii) the procyclicality of LLPs and the contribution to systemic risk and financial system instability (Borio, Furfine, & Lowe, 2001; Wong, Fong, & Choi, 2011); (iii) the role of LLP in bank earnings management and regulatory capital management (Lobo & Yang, 2001; Anandarajan, Hasan, & McCarthy, 2007; Perez, Salas-Fumas, & Saurina, 2008; Ozili, 2015); bank manager's provisioning discretion under different accounting and regulatory regimes (Alali and Jaggi, 2011; Kilic, Lobo, Ranasinghe, & Sivaramakrishnan, 2012; Leventis, Dimitropoulos, & Anandarajan, 2011; Marton & Runesson, 2017).

The analysis of the incentives that motivate the use of LLP to meet other objectives, rather than solely for credit risk purposes, is out of the scope of this study. However, the puzzle presented by the bad performance of the self-supervisory approach stimulates the search for a feasible explanation. Considering that the study addresses the Brazilian financial system, it is arguable that an idiosyncratic aspect of Brazilian jurisdiction can provide additional elements to the debate.

As discussed before, Brazilian regulation framework applies an incurred loss model to credits past due. Accordingly, credits under such condition face a loss-given-default of 100%, so the loan loss provision rapidly evolves from 0% to 100% in just 180 days. As a result, independently of the type of credit under consideration[7], 180-day past due credits are 100%

[7] There is a single exception to this prescription: long-term credits with at least a 36-month long due date are eligible to count the delinquency in double, taking 360 days to reach the 100% provision. The exception ends when the due date is less than 36 months.

provisioned. Still according to regulation, past due credits can remain for another 180-day period in the books. After this maximum 180 to 360-day period, they have to be written-off.

Though this prescription may be adequate for some kinds of credit, it is certainly overstated for others. The fast-growing provisions create a mismatch between the accounting value and the economic value of the credit portfolio and stretching the period to reach a full provisioning, though irregular, can present a solution to coordinate accounting and economic values.

From the accounting perspective, provisioning means a loss estimate. Be it derived from an expected loss or an incurred loss rationale, reported provision is not definitive until it turns into write-off. Thus, there is a time-lapse between provisioning and writing-off, which should correspond to the gradual deterioration of the credit. In accordance to regulation, the time-lapse is a maximum period of 360 days. Following this reasoning, an increase in LLP should translate into a write-off one year later. In case an additional lag is observed, it is arguable that financial institutions lengthen the deterioration period, possibly to accommodate accounting and economic mismatches.

To investigate this hypothesis, we resort to a dynamic analysis of a vector autoregressive (VAR) model, through an impulse response function method. It permits evaluation of the impulse on key variables caused by shocks (or innovations) provoked by residual variables over time (Sims, 1980).

The variables used in the analysis are the loan loss provision scaled by the loan portfolio (PROV), the write-offs also scaled by the loan portfolio (WOFF), and the rate of change of the credit portfolio (CRED)[8]. All of the variables are monthly time series that correspond to the aggregate of the Brazilian financial system spanning from 1/2006 to 9/2019. The choice of the

[8] All data is provided by the CBB financial data public repository (If.Data).

VAR lag order was determined using the Akaike information criterion (AIC), the Schwarz information criterion (SC) and the Hannan–Quinn information criterion (HQ) (Table A.5, Appendix A). Based on AIC, SC and HQ, the VAR lag order is 1, with constant. The stability test for the VAR is shown through Fig. A.1 in Appendix A. Fig. 1 shows the results.

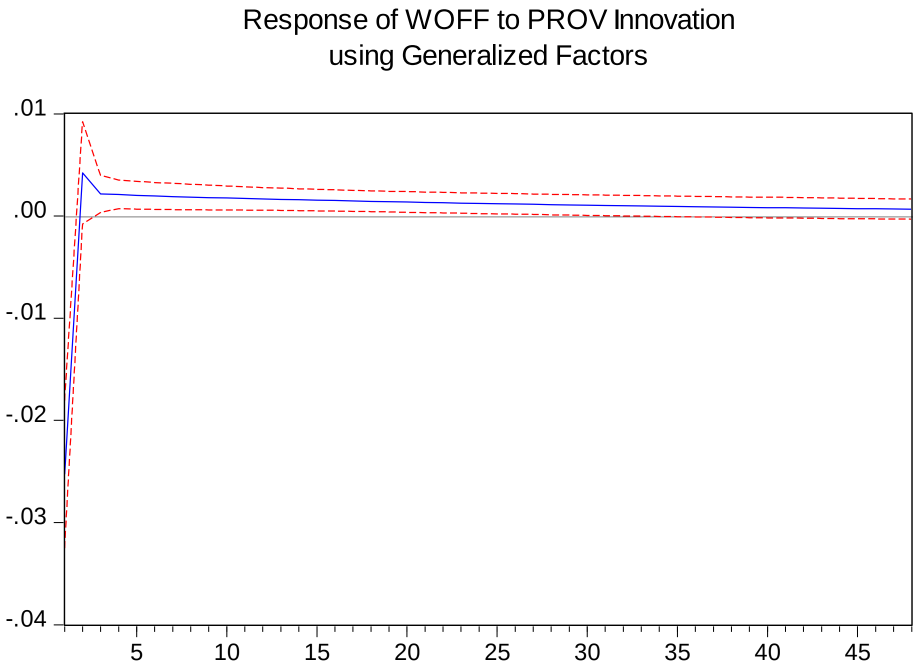

Fig. 1 – Impulse-response function of writes-offs (WOFF) due to a shock in provisions (PROV)

From Fig. 1, one can notice that an unexpected positive shock on provisions (PROV) causes an increase in write-offs (WOFF) around three months after the innovation. It agrees with loan loss provisioning regulation instructions, in which the percentages of the amount due to be constituted as provisions speed up from 10% ("D"; up to 90 days past due) to 100% ("H", more than 180 days past due) in just a quarter. The effect on WOFF of a positive shock in PROV lingers for as long as 36 to 38 months, much longer than the 12 months regulation permits[9].

The VAR analysis through the impulse response function provides evidence that the financial system lengthens the provisioning period in more than two years. Therefore, any training set

[9] Even considering the double counting exception, the maximum period a deteriorated credit should remain in the books is no longer than 18 months.

built upon bank's credit classification experience for provisioning purposes embeds the efforts to lengthen the deterioration period and compromises the quality of the trained algorithm. Whether the lengthening efforts represent an attempt to overcome an accounting-economic mismatch remains an open question. However, this finding casts light on the bad performance of the self-supervised approach.

4. Concluding remarks

This study investigates the need for credit supervision as conducted by CBB. To the extent the revised literature informs, it is the first time a natural experiment, such as a credit on-site examination, converts into a case study to compare the performance of CBB on-site supervision with a hypothetical self-supervision where banks themselves assess their loan portfolios without external intervention. In particular, the results of a real case on-site credit portfolio examination are used to compare the performance of two different machine learning sampling approaches: the first one based on good and bad ratings informed by banks, and the second one based on past on-site loan portfolio examinations conducted by CBB's banking supervision.

Overall efficiency of CBB's supervisory approach is higher and the number of bad borrowers not identified by the self-supervised approach, the type-2 error, is nearly twice as big as CBB's approach. On-site supervision approach is capable to identify 40 out of 59 bad borrowers, which corresponds to a true positive rate of 0.68. On the other hand, the self-supervision approach only catches 22 out of 59 bad borrowers, which means a true positive rate of 0.37. From the type-2 error standpoint, on-site supervision approach failed to identify 19 bad borrowers, leading to a false negative rate of 0.32, while self-supervised approach failed to identify 37 bad borrowers, a type-2 error rate of 0.63.

The consistently higher performance of CBB's supervisory approach in relation to the self-supervision approach makes a case for the necessity of on-site credit portfolio examination, as conducted by CBB. However, one cannot avoid asking: How is it possible that experts in credit management perform so poorly when classifying borrowers according to their credit risk?

To solve this puzzle, we extend the study using a dynamic analysis of a vector autoregressive (VAR) model, through an impulse response function method, to investigate whether the loan loss provisioning meets other objectives, rather than credit risk purposes. In particular, the mismatch that can arise from LLP accounting regulation and the economic deterioration of the credit.

The VAR analysis through the impulse response function provides evidence that the financial system lengthens the writing off period in more than two years. As a result, credit risk classification evolves slowly towards bad ratings, which makes provisions, on average, lower than expected. Training sets built upon this data are compromised and so are the models they provide.

The contribution of this study is threefold. It innovates by comparing on-site supervision and self-supervision performances against a common ground represented by a natural experiment turned into real case study. Regarding the methodology, it uses recently available machine learning algorithms to develop sampling models based on on-site credit supervision experience, as well as banks' experience to establish the comparison. Finally, it asserts the necessity of on-site credit supervision conducted by an independent external agent, such as the Central Bank, and suggests that the poor performance of the self-supervision approach derives from the use of the loan loss provision to overcome an accounting-economic mismatch.

References

Agenor, P. R.; Zilberman, R. (2015). Loan loss provisioning rules, procyclicality, and financial volatility. Journal of Banking & Finance, 61, 301-315.

Alali, F.; Jaggi, B. (2011). Earnings versus capital ratios management: Role of bank types and SFAS 114. Review of Quantitative Finance and Accounting, 36(1), 105-132.

Anandarajan, A.; Hasan, I.; McCarthy, C. (2007). Use of loan loss provisions for capital, earnings management and signaling by Australian banks. Accounting & Finance, 47(3), 357-379.

Antunes, J. A. P.; De Moraes, C.O.; Rodrigues, A. (2017) How financial intermediation impacts on financial stability? Applied Economics Letters, v. 26, p. 1-5.

Antunes, J. A. P.; De Moraes, C.O.; Rodrigues, A. (2018). Financial Intermediation Analysis from Financial Flows. Journal of Economic Studies (Bradford).

Barth, J., Caprio. G., Levine, R. (2004). Bank Regulation and Supervision: What works best? Journal of Financial Intermediation, 13(2), 205-248.

Barth, J.R.; Caprio, G.; Levine, R. (2008). Bank Regulations are changing: For Better or Worse? World Bank Policy Research Working Paper 4646.

Barth, J.; Lin, C.; Ma, Y.; Seade, J.; Song, F. (2013). Do bank regulation, supervision and monitoring enhance or impede bank efficiency? Journal of Banking & Finance, 2013, vol. 37, issue 8, 2879-2892.

Bhattacharya, S., Plank, M., Strobl, G. and Zechner, J. (2002) Bank regulation with random audits. Journal of Economic Dynamics and Control 26, 1301-1321.

BCBS - Basel Committee on Banking Supervision (2010). Basel III: A global regulatory framework formore resilient banks and banking systems, Bank for International Settlement.

Bernanke, B. (2010). Monetary Policy and the Housing Bubble. Speech on January 3, 2010 at the Annual Meeting of the American Economic Association, Atlanta, Georgia.

Blanchard, O. (2009). The State of Macro. Annual Review of Economics, Annual Reviews, vol. 1(1), pages 209-228, 05.

Borio, C.; Furfine, C.; Lowe, P. (2001). Procyclicality of the financial system and financial stability: Issues and policy options. BIS Papers Nº.1, 1-57.

Breiman, L. (2001). Random Forests. Machine Learning, 45, 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and Regression Trees. Chapman & Hall/CRC.

Bryson (1961). A gradient method for optimizing multi-stage allocation processes. Proceedings of the Harvard Univ. Symposium on digital computers and their applications.

De Chiara, A.; Livio, L.; Ponce, J. (2018). Flexible and mandatory banking supervision. Journal of Financial Stability, 34, 86-104

Delis, M.; Staikouras, P. (2009). Supervisory Effectiveness and Bank Risk. Review of Finance, 15(3). doi: 10.1093/rof/rfq035

El Sood, H. A. (2012). Loan loss provisions and income smoothing in US banks pre and post the financial crisis. International Review of Financial Analysis, 25, 64-72.

FSB (2017). Financial Stability Implications from FinTech: Supervisory and Regulatory Issues that Merit Authorities' Attention. Financial Stability Board.

FSB (2017_A). Artificial intelligence and machine-learning in financial services. Market developments and financial stability implications. Financial Stability Board

FSI (2018). Innovative technology in financial supervision (suptech) – the experience of early users. Financial Stability Institute. FSI Insights on policy implementation Nº 9.

Group of Thirty. (2008). The structure of Financial Supervision Approaches and Challenges in a Global Marketplace. Washington, DC.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). The elements of statistical learning: Data mining, inference, and prediction. Springer series in statistics. Springer, New York, 2nd edition.

Hellwig, M. (2014). Financial Stability, Monetary Policy, Banking Supervision, and Central Banking. Max Planck Institute for Research on Collective Goods

Hinton, J; Osindero, S; Teh, Y. (2006). A Fast Learning Algorithm for Deep Belief Nets. Neural Computation 18, 1527–1554.

Kelley, H. (1960). Gradient theory of optimal flight paths. ARS Journal. **30** (10): 947–954. doi:10.2514/8.5282

Kilic, E.; Lobo, G. J.; Ranasinghe, T.; Sivaramakrishnan, K. (2012). The impact of SFAS 133 on income smoothing by banks through loan loss provisions. The Accounting Review, 88(1), 233-260.

Krauss, C.; Do, X. A.; Huck, N. (2016). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. FAU Discussion Papers in Economics, No. 03/2016

Laeven, L.; Majnoni, G. (2003). Loan loss provisioning and economic slowdowns: Too much, too late? Journal of Financial Intermediation, 12(2), 178e197.

Leventis, S.; Dimitropoulos, P. E.; Anandarajan, A. (2011). Loan loss provisions, earnings management and capital management under IFRS: The case of EU commercial banks. Journal of Financial Services Research, 40(1/2), 103-122.

Lobo, G. J.; Yang, D. H. (2001). Bank managers' heterogeneous decisions on discretionary loan loss provisions. Review of Quantitative Finance and Accounting, 16(3), 223-250.

Marton, J.; Runesson, E. (2017). The predictive ability of loan loss provisions in banks e Effects of accounting standards, enforcement and incentives. The British Accounting Review, 49(2), 162-180.

Masciandaro, D.; Quintyn, M. (2015). The Governance of financial supervision: recent developments. Journal of Economic Surveys. Vol. 00, No. 0, pp. 1–25.

NMC (1999). National Monetary Council Resolution 2682. https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/44961/Res_2682_v2_P.pdf

Ozili, P. K. (2015). Loan loss provisioning, income smoothing, signalling, capital management and procyclicality: Does IFRS matter? Empirical evidence from Nigeria. Mediterranean Journal of Social Science, 6(2), 224-232.

Ozili, P.; Outa, E. (2017). Bank loan loss provisions research: A review. Borsa Istanbul Review, 17-3, 144-163.

Perez, D.; Salas-Fumas, V.; Saurina, J. (2008). Earnings and capital management in alternative loan loss provision regulatory regimes. European Accounting Review, 17(3), 423e445.

Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain. Psychological Review, 65(6), 386–408. doi:10.1037/h0042519.

Samuel, A. (1959). Some Studies in Machine-learning Using the Game of Checkers. IBM Journal, 211-229.

Schapire, R. (1990). The Strength of Weak Learnability. Machine Learning, 5, 197-227.

Schmidhuber, J. (2015). Deep Learning. Scholarpedia, 10(11). doi:10.4249/scholarpedia.32832

Sims, C. (1980). Macroeconomics and reality. Econometrica, 48(1), 1-48.

Stefanadis, C. (2003). Self-regulation, innovation, and the financial industry. Journal of Regulatory Economics, 23:1, 5-25,

Tressel T.; Verdier, T. (1990). Optimal Prudential Regulation of Banks and the Political Economy of Supervision. IMF working paper WP/14/90.

Werbos, P. J. (1975). Beyond Regression: New tools for prediction and analysis in the behavioral sciences. Thesis presented to The Committee of Applied Mathematics.

White, W. R. (2006). Is Price Stability Enough? BIS Working Paper No. 205.

Wong, E.; Fong, T.; Choi, H. (2011). Procyclicality of loan-loss provisioning and systemic risk in the Hong Kong banking system. Hong Kong Monetary Authority Quarterly Bulletin. Occasional Paper No. 1. Available at: http://www.hkma.gov.hk/media/eng/publication-and-research/research/

Appendix A

Table A.1: Variables descriptions

Variable	Description	Format
BCP_0	Borrower's cash performance index: realized cash flows to expected cash flows ratio. Reflects the amount paid by the borrower in relation to the contractual payment forecast. Considers a six-month period previously to the starting date and does not net off rollovers from the amount realized.	From 0 to 1, with 4 decimal places
BCP_1	Borrower's cash performance of first order index: net realized cash flows to expected cash flows ratio. Reflects the net amount paid by the borrower in relation to the contractual payment forecast. Considers a six-month period previously to the starting date and nets off rollovers from the amount realized.	From 0 to 1, with 4 decimal places
BLP	Borrower's liquidity performance index: net realized cash flows to the amount due ratio. Reflects the net amount paid by the borrower in relation to the credit balance. Considers a six-month period previously to the starting date and nets off rollovers from the amount realized.	From 0 to 1, with 4 decimal places
Grace_period	Credit balance under grace period	Dummy (1 if yes; 0 otherwise)
Credit_due_360	Amount of credit due up to 360 days to the credit balance ratio.	From 0 to 1, with 4 decimal places
Credit_due_720	Amount of credit due up to 720 days to the credit balance ratio.	From 0 to 1, with 4 decimal places
Credit_due_1440	Amount of credit due up to 1440 days to the credit balance ratio.	From 0 to 1, with 4 decimal places
Credit_due_1800	Amount of credit due as of 1800 days from starting date to the credit balance ratio.	From 0 to 1, with 4 decimal places
Credit_past_due_Bank	Amount of credit past due to the credit balance ratio in the bank.	From 0 to 1, with 4 decimal places
Credit_past_due_Bank_1	Amount of credit 90 days past due to the credit balance ratio in the bank.	From 0 to 1, with 4 decimal places
Credit_past_due_Bank_2	Amount of credit 180 days past due to the credit balance ratio in the bank.	From 0 to 1, with 4 decimal places
Credit_past_due_FS	Amount of credit past due to the credit portfolio ratio in the financial system (except for the bank).	From 0 to 1, with 4 decimal places
Credit_past_due_FS_1	Amount of credit 90 days past due to the credit portfolio ratio in the financial system (except for the bank).	From 0 to 1, with 4 decimal places
Credit_past_due_FS_2	Amount of credit 180 days past due to the credit portfolio ratio in	From 0 to 1, with 4 decimal places

	the financial system (except for the bank).	
Credit_balance_1	Credit balance rate of change between $MONTH_{-1}$ and $MONTH_0$ (starting date).	From -1 to $+\infty$, with 4 decimal places
Credit_balance_2	Credit balance rate of change between $MONTH_{-2}$ and $MONTH_{-1}$.	From -1 to $+\infty$, with 4 decimal places
Credit_balance_3	Credit balance rate of change between $MONTH_{-3}$ and $MONTH_{-2}$.	From -1 to $+\infty$, with 4 decimal places
Credit_balance_4	Credit balance rate of change between $MONTH_{-4}$ and $MONTH_{-3}$.	From -1 to $+\infty$, with 4 decimal places
Credit_balance_5	Credit balance rate of change between $MONTH_{-5}$ and $MONTH_{-4}$.	From -1 to $+\infty$, with 4 decimal places
Credit_balance_growth	Credit balance rate of change between $MONTH_{-5}$ and $MONTH_0$. (starting date).	From -1 to $+\infty$, with 4 decimal places
Revolving_credit_growth	Revolving credit balance rate of change between $MONTH_{-5}$ and $MONTH_0$. (starting date).	From -1 to $+\infty$, with 4 decimal places
Provisions_Bank	Credit provisions to credit portfolio ratio in the bank.	From 0 to 1, with 4 decimal places
Provisions_FS	Credit provisions to credit portfolio ratio in the financial system (except for the bank).	From 0 to 1, with 4 decimal places
Revolving_credit	Revolving credit to credit portfolio ratio.	From 0 to 1, with 4 decimal places
Loan	Loans to credit portfolio ratio.	From 0 to 1, with 4 decimal places
Write-offs_FS	Write-offs to credit portfolio ratio in the financial system.	From 0 to 1, with 4 decimal places

Table A.2: Machine learning algorithm settings

Machine learning algorithm	**Settings**
Random Forest	Number of trees: 300. Further settings as default.
Extreme-gradient-Boosted Trees	Booster: gbtree; eta: 0.3; Gamma: 0; Max_depth: 6; Lambda: 1; Alpha: 0. Further settings as default;
Artificial Neural Network	Activation function: relu; Loss-function: binary crossentropy; Optimizer: adam; Dropout: 0.1; Batch: 16; Epochs: 10; Layers: 2; Units: 32. Further settings as default.

Table A.3: Confusion matrix

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Where,

True Positive (TP): Correctly predicted positive sample

True Negative (TN): Correctly predicted negative sample

False Positive (FP, also type-1 error): Negative sample mistakenly predicted as positive

False Negative (FN, also type-2 error): Positive sample mistakenly predicted as negative

Table A.4: Efficiency metrics description

Efficiency Metrics	**Description**	**Formulae**

Accuracy	Number of correct predictions to total number of samples ratio.	$\frac{True\ Positives + True\ Negatives}{Total\ number\ of\ samples}$
Precision	Proportion of correct positive predictions, in relation to the total of positive predictions.	$\frac{True\ Positives}{False\ Positives + True\ Positives}$
F1 Score	Harmonic mean between precision and recall.	$2 * \frac{1}{\frac{1}{precision} + \frac{1}{recall}}$
True Positive rate (Recall or Sensitivity)	Proportion of correct positive predictions, in relation to all relevant samples, i.e., all positive samples.	$\frac{True\ Positives}{True\ Positives + False\ Negatives}$
False Positive rate (Specificity)	Proportion of negative samples mistakenly predicted as positive, in relation to all negative samples	$\frac{False\ Positives}{False\ Positives + True\ Negatives}$
False Negative Rate	Proportion of positive samples mistakenly predicted as negative, in relation to all relevant samples, i.e., all positive samples.	$\frac{False\ Negative}{False\ Negative + True\ Positive}$

Table A.5: VAR lag order

Lag	With Constant		
	AIC	SC	HQ
0	-14.13	-14.07	-14.10
1	-18.22*	-17.99*	-18.13*
2	-18.17	-17.76	-18.00
3	-18.16	-17.57	-17.92
4	-18.09	-17.33	-17.78
5	-18.05	-17.12	-17.67
6	-18.00	-16.89	-17.55
7	-17.94	-16.65	-17.41
8	-17.86	-16.40	-17.27

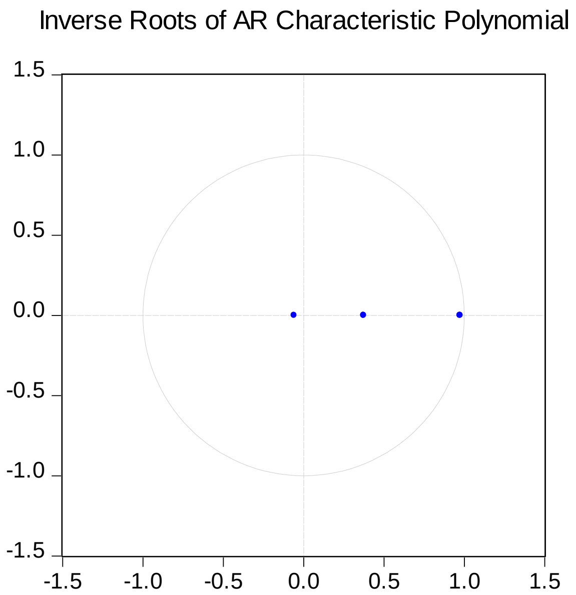

Fig. A.1 VAR Stability