Machine Learning for Zombie Hunting.
Firms' Failures and Financial Constraints.

Falco J. Bargagli-Stoffi*   Massimo Riccaboni†   Armando Rungi‡

This version: May 2020.

Abstract

In this contribution, we exploit machine learning techniques to predict the risk of failure of firms. Then, we propose an empirical definition of *zombies* as firms that persist in a status of high risk, beyond the highest decile, after which we observe that the chances to transit to lower risk are minimal. We implement a Bayesian Additive Regression Tree with Missing Incorporated in Attributes (BART-MIA), which is specifically useful in our setting as we find that patterns of undisclosed accounts correlate with firms' failures. After training our algorithm on 304,906 firms active in Italy in the period 2008-2017, we show how it outperforms proxy models like the Z-scores and the Distance-to-Default, traditional econometric methods, and other widely used machine learning techniques. We document that *zombies* are on average 21% less productive, 76% smaller, and they increased in times of financial crisis. In general, we argue that our application helps in the design of evidence-based policies in the presence of market failures, for example optimal bankruptcy laws. We believe our framework can help to inform the design of support programs for highly distressed firms after the recent pandemic crisis.

Keywords: machine learning; Bayesian statistical learning; financial constraints; bankruptcy; zombie firms

JEL Codes: C53; C55; G32; G33; L21; L25

---

*Mail to: falco.bargaglistoffi@imtlucca.it. Laboratory for the Analysis of Complex Economic Systems, IMT School for Advanced Studies, piazza San Francesco 19 - 55100 Lucca, Italy and Faculty of Economics and Business, KU Leuven, Naamsestraat 69 - 3000 Leuven, Belgium.

†Mail to: massimo.riccaboni@imtlucca.it. Laboratory for the Analysis of Complex Economic Systems, IMT School for Advanced Studies, piazza San Francesco 19 - 55100 Lucca, Italy.

‡Corresponding author: Mail to: armando.rungi@imtlucca.it. Laboratory for the Analysis of Complex Economic Systems, IMT School for Advanced Studies, piazza San Francesco 19 - 55100 Lucca, Italy.

1 Introduction

In this contribution, we propose machine learning techniques as suitable tools to provide predictions of firms' failures that can be used for credit scoring. Thus, we define *zombies* as firms that persist in a high-risk status, i.e., they are located on the highest decile of probability distributions of failures. In fact, we document that this is the segment of the distribution after which the chances of transiting to a lower risk status are minimal.

The problem of spotting non-viable firms is relevant for both academics and practitioners, whether the reason is to assess the credit risk of a single firm or to detect the portion of an entire economy that is in trouble. Traditionally, the problem is studied through the lens of a financial company that needs assessing the healthiness of a firm from albeit limited information retrieved from financial accounts[1]. Hence, scholars and practitioners have been struggling for decades to spot a firm's viability after benchmark exercises on firm-level indicators of financial constraints, e.g., when estimating the Z-scores (Altman, 1968; Altman et al., 2000), the Distance-to-Default of a firm (Merton, 1974), or the investment-to-cash-flow sensitivity (Fazzari et al., 1988). Eventually, if a financial company keeps credit flowing to otherwise insolvent firms, it may be stuck in a *zombie-lending* relationship (Peek and Rosengren, 2005; Caballero et al., 2008). In this case, a financial company may not find it convenient to pull the plug off credit and let the firm go bankrupt because it prefers avoiding disclosure of a too high share of non-performing loans. Thus, a so-called *zombie firm* outlives the market only because it has access to external financial resources. Recent OECD studies (Andrews et al., 2017; McGowan et al., 2018; Banerjee and Hofmann, 2018; Andrews and Petroulakis, 2019) extend the notion of *zombie firms* to include all low-productivity firms that would typically exit in a competitive market. Crucially, they show that *zombies*

---

[1]The seminal reference is to a departure from the Modigliani and Miller (1958) theorem, according to which capital structures should not be relevant to a firm's value in the absence of market frictions, among others including bankruptcy costs. Hence, its ability to find external financial resources should entirely rely on the viability of its investment projects. However, since frictions cannot be expunged from financial markets, a company's capital structure is actually informative on the viability of a firm and its investment projects. See also the discussion by Rajan and Zingales (1995) for an international perspective.

represent a non-negligible share of modern economies, up to 10% of incumbent firms, while absorbing up to 15%, 19% and 28% of the capital stock in countries like Spain, Italy, and Greece, respectively. In fact, in some economic environments more than others, market selection processes may not work properly for a variety of reasons, and spotting non-viable firms can be especially useful for avoiding a misallocation of productive and financial resources.

Against this background, we argue that the empirical problem of assessing whether a firm is a *zombie* is strictly related to the more general problem of determining its credit risk and, in turn, a measure of credit risk can be usefully framed in terms of survival to bankruptcy. In the end, a *zombie* is a non-viable firm that can escape failure despite its extreme financial distress, i.e., despite having the highest credit risk. Taken from another perspective, the healthier firms are, the farther both from bankruptcy and *zombie* status.

For our purpose, we implement a Bayesian Additive Regression Tree with Missing Incorporated in Attributes (BART-MIA) (Kapelner and Bleich, 2015), which is robust to non-random patterns of missing financial accounts. The ability of BART-MIA to learn from missing values is especially useful when distressed firms have the possibility to conceal information, as we find that there is a positive correlation between a firm's failure and patterns of missing financial accounts in our data. Indeed, we provide evidence that most missing variables are often the ones that have been used as proxies for *zombies* or financial constraints.

To shed light on our hypotheses, we start by training the BART-MIA on a sample of 304,906 Italian firms in the period 2008-2017. From our point of view, Italy is a compelling case of a country that hosts a relevant share of inefficient firms that hamper the potential of the economy (Calligaris et al., 2018). In addition, Italy is one of the countries most severely hit by the recent pandemic crisis, after firms had to lock down and financial distress increased. In similar cases, the call between viable firms and *zombies* is tougher for a policymaker that wants to support the economy while avoiding waste of financial resources.

The underlying intuition is simple. Based on the experience of firms that already failed in previous periods, we derive predictions on the risk of failure of firms that operate in the

following periods. Every time we compare observed outcomes against predicted outcomes, the algorithm updates and reduces the prediction errors in the next periods after processing new in-sample information on financial accounts. We end up with a probabilistic measure of firms' likelihoods to fail that one can use to assess credit risk: the higher the value of the probabilistic measure, the less viable the firm is on the market. Our machine learning framework improves upon existing models of credit rating by exploiting an extensive battery of firm-level economic and financial indicators that potentially incorporate different pieces of information regarding both the core economic activity of the firm and its ability to meet financial obligations. Indeed, our baseline predictions explain up to 0.97 of Area-Under-the-Curve (AUC), and the Precision-Recall (PR) performance reaches 0.75.

Further, we show how BART-MIA outperforms previous credit scoring models (i.e., Z-score and Distance-to-Default models), standard econometric methods (i.e., a simple logistic regression), and also other machine learning techniques (i.e., Classification and Regression Tree, Random Forest, Super Learner). Crucially, we find that emerging patterns of missing financial accounts correlate with a firm's failure, possibly because managers more likely conceal accounts when under financial distress. The latter correlation is the main reason why BART-MIA is the best predictive algorithm in our setting since it exploits the information on missing values as yet another predictor of the outcome.

Interestingly, when we implement a rigorous LOGIT-LASSO (Ahrens et al., 2020; Belloni et al., 2016a) to check which indicators better contribute to predicting firms' failures on our training sample, we find that:

(i) no single indicator predicts better than the ensemble a firms' failure, because each taken alone ends up with a higher rate of *false positives* when the test wrongly indicates that a firm is at a high risk of failure, or a higher rate of *false negatives* when the test wrongly indicates that a firm is viable;

(ii) the ranking of indicators is different from one year to the other, as the algorithm usefully responds to updates after new out-of-sample information becomes available,

3

while catching unobserved peculiarities of the economic environment.

In fact, by construction, machine learning techniques work better than single binary indicators if one includes as much valuable in-sample information as possible, while updating every time there is new out-of-sample information, because prediction errors dynamically reduce after independent tests that minimize deviations between realized and predicted outcomes (Athey, 2018).

Finally, we define a *zombie firm* as a firm that persists in a status of high risk, which we conventionally set above the 9th decile of our predictions of failure for at least three years. The latter threshold is particularly meaningful, since we find that the chances to transit to a lower risk are minimal, whereas up to 64% firms get stuck in following periods. A probation period of three years allow discounting some short-term break-even strategies, for example by start-ups. Following our definitions, we find that in Italy the share of *zombies* falls in a range between 2.5% and 4.5% along the period of analysis. Interestingly, we document that *zombies*:

(i) are counter-cyclical, as their share increases in times of crisis and decreases in times of economic recovery;

(ii) are on average 21% less productive and 75% smaller if size is measured in terms of employment.

In conclusion, we argue that our framework should be of general interest to the policymaker, for example, when she has to design optimal bankruptcy laws[2]. Indeed, tracking a firm's risk of failure allows all stakeholders, not only creditors, to understand whether there is a chance for restructuring and, if not, to prevent incumbent albeit non-viable firms from wasting additional economic resources. Evidence-based methods are even more important after the recent pandemic crisis since we expect that financial support needs targeting

---

[2]See also the suggestions by the European Directive 2012/30/EU, and the recent Italian Law on business failures on October 19th, 2017, n. 155 that lays the legal foundations for early signaling of firm's crises to enhance targeted interventions.

firms that have a real chance to recover and stay on their feet in regular times, avoiding a misallocation of resources.

The remainder of the paper is organized as follows. In Section 2, we introduce our data, and we provide some preliminary evidence. In Section 3 we illustrate our empirical strategy and show baseline results. Section 4 shows how *zombie firms* are. Section 5 concludes with a discussion of potential applications and future work.

2   Data and preliminary evidence

We source financial accounts from the ORBIS database[3], compiled by the Bureau Van Dijk, for manufacturing firms active in Italy at least one year in the period 2008-2017. Notably, Italy is a compelling case to study firms' failures and *zombie firms*. It is a country where relatively inefficient firms hamper the economy's growth potential (Calligaris et al., 2016; Bugamelli et al., 2018), perpetuating geographic divergence (Rungi and Biancalani, 2019), and extensively studied by international organizations (McGowan et al., 2018; Andrews and Petroulakis, 2019).

For our purpose, we use two main variables that help to identify firms' failures: the status of a firm and the date in which it becomes inactive. Table 1 depicts our sample coverage by firm status in the period of analysis. We assume that a company failed in the first year when it is reported as being "Bankrupted", "Dissolved," or "In Liquidation" as from original data. Altogether, the share of exiting firms constitutes about 5.7% of the entire sample, which is a figure not too far from the average official 6.3% indicated by ISTAT, the national statistics

---

[3]ORBIS firm-level data (Orbis, 2020) have become a common source for global financial accounts. For a previous usage of this database, among others, see Gopinath et al. (2017) and Cravino and Levchenko (2016). Coverage of smaller firms and some financial accounts can change from country to country, following filing requirements by national business registries, as observed in validation exercises by Kalemli-Ozcan et al. (2015) and Gal (2013). In the case of Italy, the original information provider for Italian financial accounts is CERVED, a credit rating agency. The Bureau Van Dijk standardizes and translates original financial accounts to make them comparable across countries. Please note that ORBIS, differently from other platforms from the same Bureau Van Dijk (e.g., AIDA or AMADEUS), does not drop exiting firms in our period of analysis. It complements financial accounts with other information from different sources on ownership, management, and intellectual property rights, which we also use for predictions.

office, for the same period. Figure 1 maps the firms' failures by NUTS 3-digit regions in logarithmic scale. As expected, we find a concentration of failures in metropolitan regions but full representativeness of the entire Italian territory, as we detect failures in any region in our period of analysis.

Table 1: Firms by status

Status	Active	Bankrupted	Dissolved	In Liquidation	Total
Sample	287,586	1,533	8,540	7,221	304,906
Percentage	94.33%	0.50%	2.80%	2.37%	100%

Figure 1: Geographic Coverage
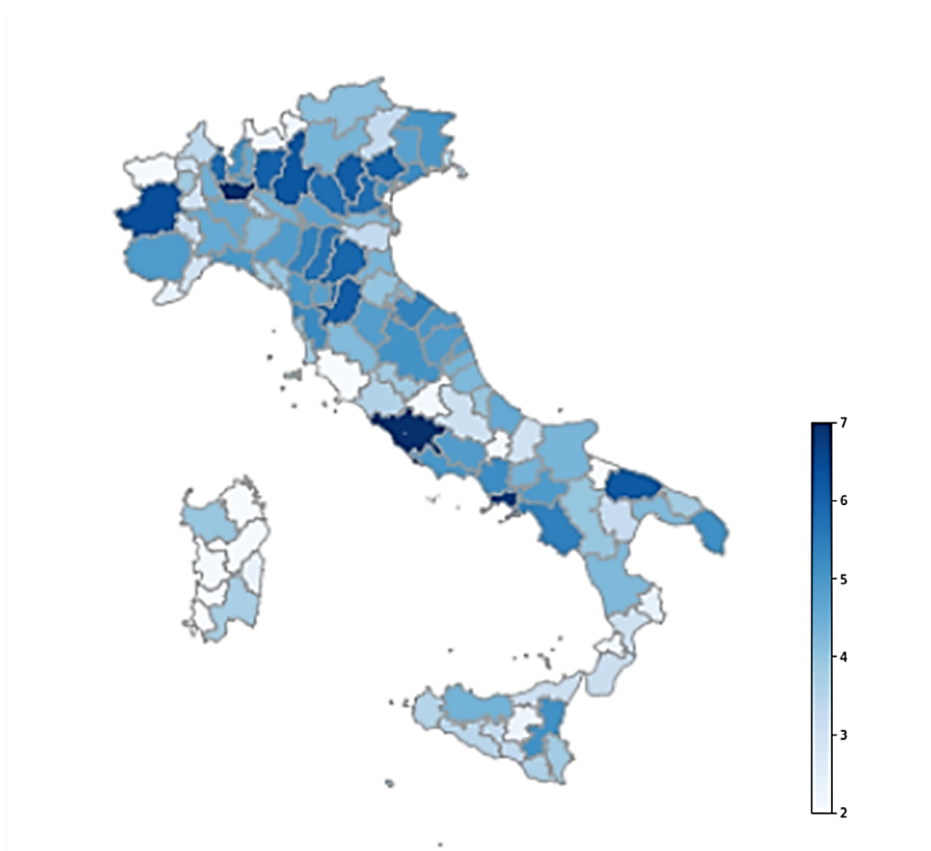
Note: Number of firms' failures reported in log scale for NUTS 3 regions. We assume a firm fails when it is reported as "Bankrupted", "Dissolved or "In liquidation". A greater number of failures is reported in metropolitan areas (Milan, 1,016; Rome, 1007; Naples, 840; Turin, 616). All regions are covered in our sample.

Additionally, we use a wide battery of economic and financial indicators to train our predictive models. The battery includes: i) original firm-level financial accounts; ii) indicators that have been adopted in previous literature to proxy firm-level financial constraints; iii) indicators that have been used so far to spot *zombie firms*; iv) indicators that are adopted as an alert on firms' crises in the most recent Italian bankruptcy law[4]. Any predictor we consider is described in detail in an Appendix Table A1.

Please note that many of the indicators that we pick as predictors have been used in very different frameworks to assess how much a firm is in trouble. From our machine learning perspective, they cannot be interpreted as drivers of failure. It is sufficient that they contribute with an albeit small piece of information on the healthiness of a firm. In our prediction framework, they can even border on multicollinearity, as for example in the case of different measures of efficiency, liquidity, and solvency ratios. As far as we are not interested in the identification of a causal contribution to a firm's failure, multicollinearity is not a problem for our predictions (Makridakis et al., 2008; Shmueli et al., 2010). On the contrary, we will discuss in Section 3.4 how, by construction, one cannot seal off the empirical contribution of one indicator from the entire set in a context of a pure prediction problem. For this reason, the set of predictors we use should be considered as a non-separable ensemble, and a discussion on the statistical significance of each predictor is not relevant for our scope.

In Appendix Figure A1, we visualize a map of missing values in our sample. Besides mandatory and basic information (volume of activity, profits, location, industry affiliation, ownership, and intellectual property rights), many other financial accounts show different patterns of missing values across firms. After we run a series of chi-squared tests in Appendix Table A2, we do find that there is a positive statistical association between the patterns we observe from the sample and the event of a firm's failure. In a nutshell, it is more likely

---

[4]A recent reform of the bankruptcy law (L. 155/2017 and DL. 14/2019) proposes an early warning system based on indicators identified by practitioners. The purpose is to spot firms that are in trouble before it is too late, preserving entrepreneurial abilities and finding a way out of the crisis. It delegated practitioners (specifically the national accountants' association: "Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili") to provide a list of indicators that could help assess the status of a firm's crisis.

that a firm will fail when a pattern of missing financial accounts is observed. The exercise performed in Table 2 clearly shows such correlations. We run a simple logistic regression taking as a dependent variable the observed failure of a firm. Then, we include a binary regressor equal to one if the predictor is missing at least once in the three years before the failure occurs. Fixed effects by NUTS 3-digit region and NACE 2-digit industries are included. We report results by each predictor by row in Table 2.

Table 2: Missing predictors and firms' failures

Missing predictor	Odds ratio	Std. Error	N. obs.	Pseudo R2
Interest Coverage Ratio	1.91***	(0.30)	304,906	0.026
Interest Benchmarking	1.29***	(0.20)	304,906	0.019
Value Added	2.37***	(0.37)	304,906	0.033
Z-score	2.55***	(0.38)	304,906	0.037
Total Factor Productivity	2.37***	(0.39)	304,906	0.034
Profitability	1.91***	(0.30)	304,906	0.027

Note: We report odds ratios after a logit specification where the dependent variable is a firm's failure and the binary regressor is equal to one when at least one missing value is detected in the latest three years. Fixed effects at the region and industry level. Errors are clustered by industry.

The correlations reported above are particularly relevant for the scope of our analyses. The main issue is sample selection when observations are selectively missing for some categories of firms. In this case, we have two potential sources of sample selection bias: i) firms in distress *vis à vis* firms not in distress, because the first can have an incentive to disclose less information than the latter; ii) smaller firms *vis à vis* bigger firms because the first are often exempted from a full financial report, according to Italian regulation [5]. In fact, the two sources can overlap, since smaller firms may also be the ones that suffer relatively more from financial distress. The application of a BART-MIA procedure, as described in Section

---

[5]Following Italian civil law, companies that do not quote financial activities on the stock exchange have the possibility to present more aggregate financial accounts when their size does not exceed simultaneously two of the following thresholds, for one or two consecutive periods: i) 4,400,000 euro of total assets; ii) 8,800,000 euro of operating revenues; iii) 50 employees. A limited financial statement always includes main items at the first or second digit of aggregation.

[3](), allows us considering both sources of sample selection when patterns of missing financial accounts emerge because the algorithm includes such patterns as a further predictor of firms' failures.

Interestingly, we find that most missing variables are also the ones that have been used in previous works as proxies for *zombies* or financial constraints. Take, for example, the case of the Interest Coverage Ratio (ICR), derived as a ratio between the earnings before interest and taxes (EBIT) and the interest expenses of a firm. When ICR is lower than one, Bank of England (2013) assumes that a firm is a *zombie* since it has problems in meeting financial obligations. In our sample, we find that about 19% of firms have an ICR smaller than one but, at the same time, there are 62.50% firms whose information on ICR is not present at all. Further, according to Bank of Korea (2013), a negative value-added is the most appropriate proxy for assessing a *zombie* status because it indicates that intermediate inputs have a higher market value than the firm's output. In the case of Italy, about 64.27% of companies do not report information on value-added in at least one period, while about 3% of them report a negative figure. Clearly, negative value added is a stricter condition than negative profits, as a firm can have no profits without destroying economic value. Indeed, firms' profitability is at the core of two very similar proxies of *zombies* adopted by Schivardi et al. (2017), when they compare firm-level profits with an external benchmark. Reproducing the same exercises, we find that about 3% of Italian firms are in trouble, while a large amount of the sample (62.50%) does not report any information on the predictor. Finally, both Caballero et al. (2008) and McGowan et al. (2018) run another benchmarking exercise comparing the interest paid by a firm to obtain external financial resources against the cost opportunity to invest in alternative safer assets. In our case, when we try to reproduce the same exercise with the yields on Italian state-issued bonds with a maturity of ten years, we find there is a high share (60.29%) of firms for which we do not have any information in at least one period when it was active.

Finally, we include as a predictor an estimate of Total Factor Productivity (TFP) follow-

ing the methodology proposed by Ackerberg et al. (2015), to take into account the simultaneity bias deriving from *ex-post* adjustments in the combination of factors of production. In this regard, firm-level TFPs allow us predictions based on the ability to transform inputs and sell the output on the market. In fact, the relationship between financial constraints and productivity is one of the most debated (see among others: Aghion et al. (2019); Ferrando and Ruggieri (2018)). The simple guess for our predictive models is that less productive firms are the ones that may encounter more difficulties in surviving to the market. At the same time, *zombie firms* have been often defined also in terms of (lack of) productivity (McGowan et al., 2018; Andrews et al., 2017; Andrews and Petroulakis, 2019).

3 Empirical strategy

Spotting a non-viable firm is difficult for obvious reasons. If financial accounts are bad, one could argue that it is just a matter of time before the becomes more competitive, given the right conditions. If financial accounts are good, one could argue that the worst has yet to come, because bad management choices will show up later. Trivially, only firms that are already bankrupt were certainly non-viable at some point. Still, an external observer will never know when it exactly happened because the manager of a firm in trouble has an incentive not to share private information.

In principle, an analyst would like to observe the entire horizon of the events to discount all possible scenarios and understand the value of a firm and its investment projects. But it is not possible, as this is the typical twin problem of a financial institution that faces uncertainty in the presence of informational asymmetries. On the one hand, the company has a distinct information advantage in its investment plans. On the other hand, both the financial institution and the company have limited power in predicting future economic shocks that may have either a positive or a negative impact.

In this context, we propose a machine learning framework that uses past information

about already failed firms to assess what the probability is that another firm in a similar shape will go bankrupt[6]. The wider the variety of past experiences we can rely on, the more precise the prediction on the healthiness, or lack thereof, of a firm (Kleinberg et al., 2015). Ultimately, ours is a perspective on the (lack of) resilience of a firm potentially using all the observables that could possibly hold a piece of information on its viability. We end up with a firm-level probabilistic measure bounded between 0 and 1 that tells us what is the chance that a firm exits from the market in the next period since other firms in a similar condition did. As we plot in Figure 2, the idea is that we can assess the distance of each firm from the highest financial distress, which we conventionally may assume in red above the 9th decile. In fact, following analyses will show that the highest decile is the segment of the risk distribution after which it is very difficult to transit back to lower distress.

Figure 2: Fictional distribution of failure's probability
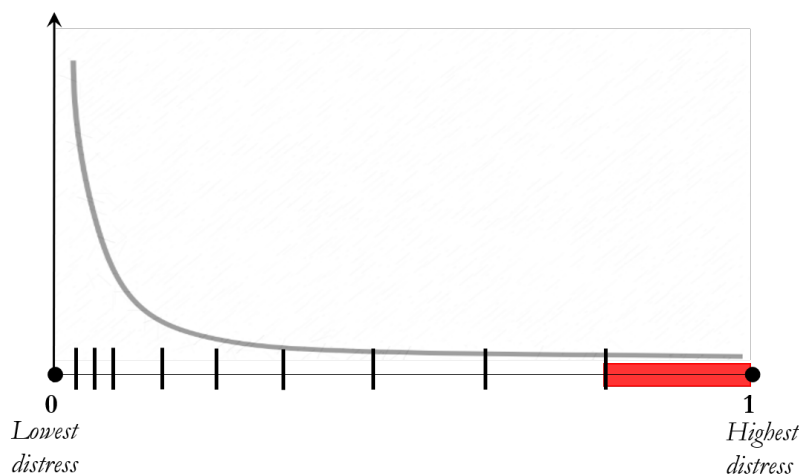
Importantly, we want to underline that it is not of our concern the identification of the drivers of firms' failures since ours is a straightforward prediction framework. Nonetheless, we do show later in the analysis how a list of best predictors of bankruptcy can change over

---

[6]Previous exercises to predict firms' failures from financial accounts have been tried in data science literature, although in the absence of a clear economic and financial framework. In this regard, Bargagli-Stoffi et al. (2020b) provide an extensive overview of the most recent data science literature related to the usage of machine learning for firm dynamics. Previous exercises using data similar to ours include Moscatelli et al. (2019) for Italy, and Weinblat (2018).

time, given the circumstances in which we run predictions, every time one updates with new out-of-sample information.

3.1   Predictions of failures

Let us consider a generic predictive model in the form:

$$f(\mathbf{X}_i) = Pr(Y_i = 1 \mid \mathbf{X}_i = x) \tag{1}$$

where Y_i is the binary realization of the outcome assuming value 1 if the ith firm exits from the market and zero otherwise, while $\mathbf{X}_i$ is the P-dimensional vector of firm-level predictors, where P is the number of predictors included in the model. The functional form that links predictors to outcomes is determined by the generic supervised machine learning technique used to predict out-of-sample information. Briefly, the generic algorithm will pick the best in-sample loss-minimizing function in the form:

$$\arg\min \sum_{i=1}^{N} L(f(x_i), y_i) \quad over \quad f(\cdot) \in F \qquad s.t. \qquad R\big(f(\cdot)\big) \leq c \tag{2}$$

where F is a function class from where to pick $f(\cdot)$, and $R\big(f(\cdot)\big)$ is the generic regularizer that summarizes the complexity of $f(\cdot)$. See also Mullainathan and Spiess (2017). In our case, the function $f(\cdot)$ is an element from the family of classification trees, or a combination thereof. The set of regularizers, R's, will change following the standards adopted by each method. Eventually, any algorithm shall take a loss function $L(f(x_i), y_i)$ as an input and look for the function that minimizes prediction losses[7]. Eventually, given new out-of-sample information, we derive a prediction for the failure of any firm based on the set of its up-to-

---

[7]For a general reference, see Breiman et al. (1984). Any tree $\mathcal{T}$ is built on *if-then* statements that split the dataset according to the observed values of predictors, allowing for non-linear relationships between the predictors and the outcomes. Hence, the generic algorithm for the construction of a tree, $\mathcal{T}$, is based on a top-down approach that recursively splits the main sample into non-overlapping sub-samples (i.e. the nodes and the leaves). Then, the tree is pruned iteratively with the regularizer R to improve its predictive ability and avoid data overfitting, when trees become too deep along many layers.

date financial accounts, both in the case of incumbent firms that already operated in previous periods and in the case of firms entering into the market for the first time. From another perspective, we interpret this probability range as a degree of risk by an investor that does not have any other information than what is included in the actual financial accounts.

In the following analyses, we consider as a baseline the BART - Bayesian Additive Regression Tree (Chipman et al., 2010). More specifically, we adopt a variant (BART-MIA) that is robust to non-random patterns of missing values in financial accounts to predict firms' failures (Kapelner and Bleich, 2015). To assess the performance of the model, we implement a standard two-fold cross-validation procedure that assigns 90% of the observations to the training sample and 10% to the test sample (Devijver and Kittler, 1982).

In general, the BART is an ensemble method based on the aggregation of different independent trees and three regularizing priors designed to prevent overfitting. The BART-MIA variant we adopt can be expressed as:

$$Y_i = f(\mathbf{X}_i) + \epsilon_i \approx \mathcal{T}_1(\mathbf{X}_i) + ... + \mathcal{T}_q(\mathbf{X}_i) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \qquad (3)$$

where the q distinct binary trees are denoted by $\mathcal{T}$, and each single tree comes with an entire structure made of nodes and leaves. The Bayesian component of the BART includes three priors: (i) the prior on the probability that a node will split at depth k is $\beta(1 + k)^{-\eta}$, where $\beta \in (0, 1), \eta \in [0, \infty)$, and the hyper-parameters are chosen to be $\eta = 2$ and $\beta = 0.95$; (ii) the prior on the probability distribution in the leaves is a normal distribution with zero mean: $\mathcal{N}(0, \sigma_q^2)$, where $\sigma_q = \sigma_0/\sqrt{q}$ and σ_0 can be used to calibrate the plausible range of the regression function; (iii) the prior on the error variance is $\sigma^2 \sim Inv - Gamma(v/2, v\lambda/2)$ where λ is determined from the data in a way that the BART will improve 90% of the times the root mean squared error (RMSE) of an OLS model.

BART-MIA extends the original BART algorithm by incorporating additional information coming from patterns of missing values (Kapelner and Bleich, 2015). This is done by

introducing, in each binary-tree component of the BART algorithm, the possibility to split on a missingness feature. As shown by Twala et al. (2008), this splitting-rule allows trees to better capture the direct influence of missing values as a further predictor of the response variable. Indeed, we observed in Section 2 how patterns of missing financial information are not random, or at least not completely random, with respect to firms' failures. If firms are financially distressed and/or if firms are smaller, it is more likely that they will be missing in any firm-level dataset. In this case, naïvely trimming the missing observations would introduce a sample selection bias, i.e., it would more likely exclude some categories of firms whose probability to fail is higher. Standard imputation techniques (e.g., conditional median imputation, Bayesian imputation) cannot solve the sample selection problem, as far as they are based on the use of in-sample information. Usefully, BART-MIA creates a new label when the values of a specific firm-level financial account are missing, and it tests the new label as an additional predictor for a firm's failure.

3.2 Validation against other machine learning techniques

We compare the quality of the prediction of BART-MIA with the Conditional Inference Tree (Hothorn et al., 2006), the Random Forest (Breiman, 2001), and the Super Learner (Van der Laan et al., 2007). The Conditional Inference Tree (Hothorn et al., 2006) that we apply is a simple variant of the Classification and Regression Tree (CART) algorithm (Breiman et al., 1984) based on a significance test procedure that avoids a bias towards variables with many possible splits (see Odén et al. (1975); Loh (2002); Hothorn et al. (2006)). The Random Forest is an ensemble method that aggregates different trees to get a stronger predictive power, and each tree is constructed by randomly picking different variables among all the possible predictors and randomly selecting a subset of the total number of observations (see also Breiman (2001)). The Super Learner (Van der Laan et al., 2007) is based on a weighted combination of other algorithms Van der Laan et al. (2007). For our purposes, we build the Super Learner as a convex combination of a logistic regression model, a Conditional

14

Interference Tree and a Random Forest.

Table 3: Models' horse race: out-of-sample prediction accuracy

Model	PR	AUC	F1-score	BACC	R^2	Train Obs	Test Obs
Logit	0.3576	0.8896	0.2098	0.8433	0.0829	83,537	9,282
Ctree	0.3568	0.8889	0.2000	0.7804	0.0654	83,537	9,282
Random Forest	0.4262	0.9050	0.2257	0.8515	0.0922	83,537	9,282
Super Learner	0.4311	0.9073	0.2232	0.8666	0.0945	83,537	9,282
BART-MIA	0.7484	0.9667	0.6328	0.8993	0.4038	83,537	9,282

In Table 3 and Figure 3, we show how BART-MIA always outperforms other state-of-the-art methods from the literature. It also has a higher predictive power than a classical logistic regression[8]. To compare the power of predictions across methods, in Table 3, we show five different performance measures commonly used for classification problems: Precision-Recall (PR); Area Under the receiver operating characteristic Curve (AUC); F1-Score; Balanced ACCuracy (BACC); adjusted R^2. Both AUC and PR vary between 0 and 1, where 0 indicates complete misclassification and 1 indicates perfect prediction. The AUC (Hanley and McNeil, 1982) is a general measure of predictive power that tells us how much we are able to classify failures *vis à vis* non-failures, hence with an accent on the *false discovery rate* (FDR). PR is particularly useful for our scope, because it accounts for both the overall share of true failures that we are able to predict in the data (i.e., the *sensitivity/recall* of predictions), as well as for the proportion of predicted failures that reveal to be true failures (i.e.,the *precision* of predictions). In fact, an evaluation of the *sensitivity/recall* alone could be misleading in a zero-inflated setting like ours, where the number of non-failures systematically exceeds the number of failures (for more details, see Saito and Rehmsmeier (2015); Fawcett (2006)). The F1-score (Van Rijsbergen, 1979) and the BACC (Brodersen et al., 2010) are used for cases of

---

[8]Please note how BART-MIA could always use a larger training sample, including firms that report missing values on predictors. This is an additional advantage of our baseline methodology. However, for sake of comparison, all methodologies in Table 3 have been trained on the same subset of roughly 84,000 observations for the training and 9,000 for the testing, without considering missing values.

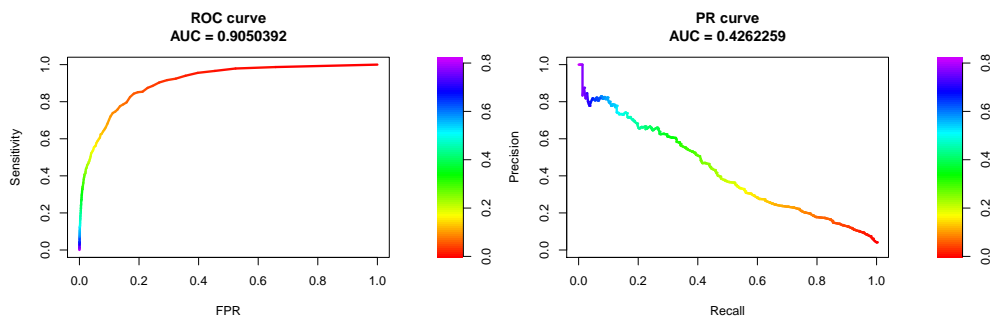unbalanced data. The first is based on a harmonic mean of *precision* and *recall*, whereas the second is a simple average between the rate of true positives and the rate of true negatives from our predictions.

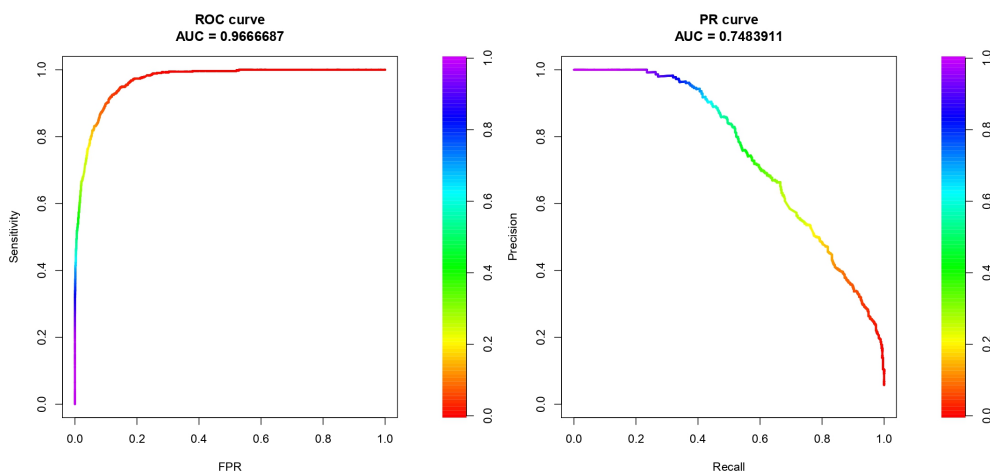Figure 3: Out-of-sample Goodness-of-Fit
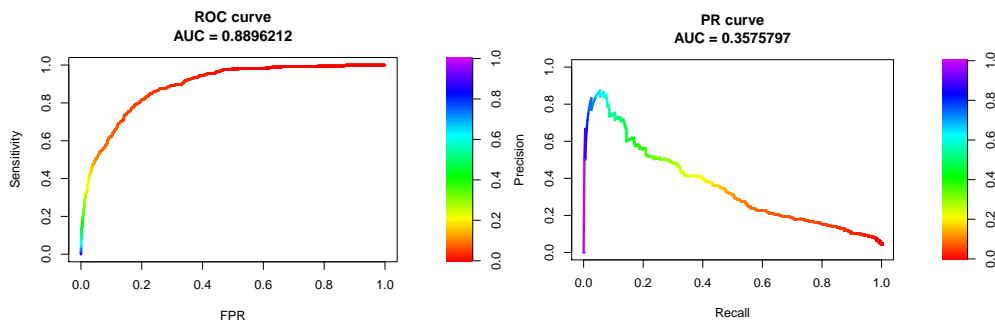
Panel (A): ROC Logit

Panel (B): PR Logit

Panel (C): ROC Super Learner

Panel (D): PR Super Learner

Panel (E): ROC BART-MIA

Panel (F): PR BART-MIA

16

Predictions from BART-MIA show a PR equal to 0.75 and an AUC equal to 0.97. In this context, the second best predictor is the Super Learner, which however stops at 0.43 and 0.91 in terms of PR and AUC, respectively.[9]. Interestingly, the greater improvement comes from a higher precision. This is shown by the larger gaps of the BART-MIA with respect to the performance measures that leverage on precision (namely, PR and F1-score). The latter is central to our purpose, as it means that BART-MIA has a higher predictive power in assessing which firms will fail one year from predictions.

3.3 Validation against proxy models of credit scoring

So far, we compared predictions of failures from different econometric and machine learning techniques, and we showed how BART-MIA displayed a higher predictive ability. Here, we compare our baseline predictions with widely known proxies for firm-level credit scores: the *Z-scores* (Altman, 1968) and the *Distance-to-Default* (Merton, 1974).

In the case of Z-scores, a selection of financial ratios (profitability, leverage, liquidity, solvency, and volumes of activity) is plugged in an equation with some weights to proxy their relative importance. The weights are taken from literature and obtained from previous scholarly estimates on the relative importance of these indicators in assessing a firm's distress. Hence, one gets a threshold that, if crossed, indicates that there is a high probability of bankruptcy in the next future. Different from the Z-scores, the Distance-to-Default by Merton (1974) focuses specifically on the ability of a company to meet its financial obligations. The original intuition is that a company's equity can be modeled as a call option on its assets. Hence, to build it, one needs combining firm-level accounts (company's assets, debts, market value) and information retrieved on financial markets (risk-free interest rate, a standard deviation of stock returns). Eventually, one plugs the variables in an equation

---

[9]It is worth noticing that the Super Learner is performing better than the Random Forest with respect to the accuracy, but not with respect to the F1-score. This is due to the fact that the Super Learner algorithm (Van der Laan et al., 2007) is optimized in a way to find a convex combination of algorithms that minimizes the accuracy of the ensemble method. In this particular case, this strategy is not optimal as the output data are unbalanced.

that returns the value of a theoretically fair call option[10].

We estimate both the Z-scores and the Distance-to-Default, and we report their predictive power in Table 4. In line with the interpretation of Z-scores, we decided to publish a full range of thresholds across the deciles by increasing levels of firms' distress. Distance-to-Default predictions show both higher precision (0.2680 vs. 0.1613) and a lower false discovery rate (0.7320 vs. 0.8387) than the ones obtained by Z-scores. However, BART-MIA outperforms both models, since its precision and false discovery rates are 0.8278 and 0.1722, respectively, when we fix the sample on the same number of observations for which data are available to calculate proxy models.

Table 4: Goodness-of-fit of Distance-to-Default (DtD) and Z-scores

Decile	Precision DtD	FDR DtD	Precision Z-score	FDR Z-score
1st	0.2680	0.7320	0.1613	0.8387
2nd	0.2680	0.7320	0.1505	0.8495
3rd	0.2680	0.7320	0.1505	0.8495
4th	0.2258	0.7742	0.1371	0.8629
5th	0.2022	0.7978	0.1269	0.8731
6th	0.1759	0.8241	0.1185	0.8815
7th	0.1569	0.8431	0.1108	0.8892
8th	0.1467	0.8533	0.1036	0.8964
9th	0.1411	0.8589	0.0981	0.9019
10th	0.1313	0.8687	0.0969	0.9031

Note: Performance of predictions to compare with BART-MIA. On the same test set for which we have no missing values in Distance-to-Default and Z-scores, BART-MIA's precision is 0.8278 and false discovery rate (FDR) is 0.1722.

3.4   High dimension is not *'a curse'* for failures

In previous analyses, we explicitly claimed that we needed as much in-sample information as possible to reduce out-of-sample errors in predictions. On the other hand, there could

---

[10]After the insights of the distance-to-default model, Black and Scholes (1973) developed their widely known model based on the observation that hedging an option one could remove a systemic risk component.

be problems of overfitting when data dimensionality increases[11]. In this Section, we show what happens when we sift through firm-level data and try to extract a set of predictors with the highest ability to spot financial distress. The natural candidate tool to shrink the dimensionality of a matrix of predictors is the LOGIT-LASSO (Ahrens et al., 2019), whose functional form in a panel setting is the following:

$$\underset{\beta \in \mathbb{R}^p}{\arg\min} \ \frac{1}{2N} \sum_{i=1}^{N} \left(y_{i,t}(x_{i,t-1}^T \beta) - log(1 + e^{(x_{i,t-1}^T \beta)}) \right)^2 \quad \text{subject to } \|\beta\|_1 \leq k. \qquad (4)$$

where $y_{i,t}$ is a binary variable equal to one if a firm i failed at time t and zero otherwise. Any $x_{i,t-1}$ is a lagged predictor chosen in $\mathbb{R}^p$ at time $t-1$, whereas $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$ and $k > 0$. The constraint $\|\beta\|_1 \leq k$ limits the complexity of the model to avoid overfitting, and k is chosen, following Ahrens et al. (2019), as the value that maximises the Extended Bayesian Information Criteria (Chen and Chen, 2008). To account for the potential presence of heteroskedastic, non-Gaussian and cluster-dependent errors, we adopt the rigorous penalization introduced by Belloni et al. (2016a). Top ten highest-ranked predictors are presented in Table 5.

Even though number of predictors varies in time, up to a maximum of twenty-one features in 2015, we find a core set of predictors that are frequently selected. The stable set includes indicators of financial distress (*Liquidity Returns, Interest Coverage Ratio, Interest Benchmark, Financial Constraint*), as well as indicators related to the core economic activities of the firms (*Negative Value Added, Total Factor Productivity, Size-Age*). Apparently, firms that are controlled by parent companies (*Corporate Control*) are less likely to fail in every period, i.e., the predictor enters in the algorithm with a negative coefficient. The same happens for the *Dummy Trademarks* and the *Dummy Patents*, as it makes sense that intangible assets reduces the probability to exit from the market[12].

---

[11]The so-called *curse of dimensionality* is an obstacle in many cases, when one works with finite data samples and many variables. The seminal reference is to the work by Bellman (1961), who introduced the notion of dimensionality reduction.

[12]See Table A1 for a full description of predictors and their construction.

Table 5: Top 10 predictors for firms' failures - Results from a rigorous LOGIT-LASSO

Rank	2017	2016	2015	2014	2013	2012	2011	2010	2009
1	Liquidity Returns	Negative Value Added	Negative Value Added	Negative Value Added	Liquidity Returns	Negative Value Added	Negative Value Added	Negative Value Added	Negative Value Added
2	Negative Value Added	Liquidity Returns	Corporate Control	Liquidity Returns	Negative Value Added	Profitability	Liquidity Returns	Liquidity Returns	Liquidity Returns
3	Corporate Control	Corporate Control	Financial Constraint	Solvency Ratio	Solvency Ratio	Financial Constraint	Financial Constraint	Profitability	Financial Constraint
4	Interest Coverage Ratio	Financial Constraint	Interest Coverage Ratio	Profitability	Profitability	Corporate Control	Corporate Control	Financial Constraint	Profitability
5	Financial Constraint	Interest Coverage Ratio	Profitability	Financial Constraint	Corporate Control	Solvency Ratio	Solvency Ratio	Corporate Control	Corporate Control
6	Solvency Ratio	Size-Age	Solvency Ratio	Corporate Control	Financial Constraint	Interest Coverage Ratio	Size-age	Solvency Ratio	Solvency Ratio
7	Size-age	Solvency Ratio	Size-age	Size-age	Size-Age	Liquidity Returns	Interest Coverage Ratio	Region (NUTS 2)	Interest Coverage Ratio
8	Profitability	Profitability	Interest Benchmark	Interest Coverage Ratio	Interest Coverage Ratio	Size-age	Financial Misallocation	Dummy Trademarks	Dummy Trademarks
9	Interest Benchmark	Interest Benchmark	Liquidity Ratio	TFP	TFP	Dummy Patents	Dummy Trademarks	Interest Coverage Ratio	Size-age
10	Liquidity Ratio	Liquidity Ratio	Capital Intensity	Liquidity Ratio	Dummy Patents	Dummy Trademarks	Dummy Patents	Dummy Patents	Capital Intensity

Note: The rankings are obtained after the implementation of a rigorous LOGIT-LASSO (Ahrens et al., 2020; Belloni et al., 2016b) every year on the entire battery of predictors described in Figure A1. Only the first ten selections are reported. The procedure selects a different number of predictors every year, up to a maximum of 21.

Some of the top predictors we show have been used to measure either financial distress or *zombie lending*. However, the ranking changes over time, and we do not detect any meaningful pattern in these changes. In this context, we cannot rely on a single indicator (or a set thereof) to derive predictions of failures. If we did, we would have a higher rate of *false positives* (when our forecasts indicate that a firm is at risk of failure, but it is not) and a higher rate of *false negatives* (when the predictions wrongly suggest that a firm in trouble deserves credit).

In fact, at this stage, we cannot exclude that a different ranking over time is due to better use of the newly acquired in-sample information after new failures have been observed. Also, it might be the case that a change in rankings reflects some shift in the business environment in which companies operate. We presume that applications to different aggregates (countries, regions, industries) may return different rankings. In the spirit of a pure prediction problem, we conclude that it is better to keep the entire battery of predictors, as far as a dimensionality problem on the predictors does not arise, when we can train predictions on a large enough number of firm-level observed outcomes.

3.5   Robustness and sensitivity checks

The first robustness we implement is to check whether any imputation of missing values allows us to obtain better predictions after reverting to other techniques different from BART-MIA. We adopt a CART-based methodology (Buuren and Groothuis-Oudshoorn, 2010), in the fashion of White et al. (2018), to fill in the missing financial accounts. We do not find any sensitive change in the ranking observed in Table 3. More in general, any finer tuning on the hyper-parameters does not seem to overturn the gap in performance between the BART-MIA and the other techniques.

A second check is done with respect to our choice of predictors. We want to be sure that nothing relevant is missing from our battery of predictors described in Table A1. Our sensitivity check is done in two steps by simulating synthetic predictors. In a first step,

we generate a predictor whose correlation with the outcome of failure is as high as one of the best-observed predictors (0.5), but it is not significantly correlated with any of the other predictors (for more details on this methodology, see Bargagli-Stoffi et al. (2020a)). Hence, we check that its inclusion in a new augmented model affects unit-level predictions only in 0.88% of cases (namely, just 0.88% of the predictions of the augmented model are significantly different from the predictions of the original in (3)). Then, we repeat the exercise by simulating a second synthetic predictor that is assumed correlated with both the outcome (0.5) and a predictor (0.3). In this case, unit-level predictions change only in 0.90% of cases. We conclude that the original choice of predictors already captures the bulk of the signal in the data[13].

A third issue we want to address is the zero-inflated nature of our outcome variable. Any data set will structurally include a greater majority of incumbent firms that rarely exit from the market, i.e., the zeros. In contrast, the share of firms' failures is usually of a one-digit magnitude. For example, in the case of Italy, as from Table 1, only 5.7% of firms failed in our period of analysis, in line with official figures from the Italian national statistics bureau. To clarify the nature of the problem, if we applied a methodology that predicted all the firms to fail in our data, we would still get an accuracy measure for correctly classified observations over total observations of 94.3%. To deal with similar unbalanced outcomes, one could either under-sample the label that accounts for the largest majority of observations, or over-sample observations for the other label, as suggested in other applications (Gruszczyński, 2019; Zhou, 2013). In our case, we argue that any re-sampling method is sub-optimal because we know from Section 2, Table 2, and Appendix Table A2 that patterns of missing values are positively correlated with the outcome. For example, any under-sampling may induce a higher association between the missing observations and the failure of firms. In our framework, we argue that it is better to look at alternative goodness-of-fit measures that

---

[13]Figure A2 provides a graphical intuition. It shows the standardized differences for the unit-level predictions of the augmented and the original model for 50 randomly selected observations. The standardized differences in the predictions do not deviate significantly, indicating no statistical difference between the predictions of the two models.

take into account unbalanced outcomes while keeping the sample intact. See Table 3 where we report PR, F1-score, and BACC, and we comment that BART-MIA predictions still show better accuracy.

4   A case for *zombie firms*

Originally, the notion of *zombie firms* relates to the phenomenon of *zombie lending*, when banks keep credit flowing to otherwise insolvent borrowers. In some cases, *zombie lending* can be a deliberate strategy to avoid a bank's budget restructuring, while only apparently complying with capital standards set by the financial regulators (Bonfim et al., 2020), e.g. the case of Japanese banks in the 1990s (Peek and Rosengren, 2005; Caballero et al., 2008). More recently, Schivardi et al. (2017) study Italian firms during the financial crisis in 2008 finding that credit misallocation after *zombie lending* increases the failure rate of otherwise healthy firms while reducing the failure rate of non-viable firms. The reason is that under-capitalized banks can decide to cut credit to more viable projects to avoid public disclosure of non-performing loans in their portfolio. Paradoxically, in times of financial crises, highly distressed firms may appear resilient thanks to continuous access to financial resources. From a more general perspective, recent OECD studies (Andrews et al., 2017; McGowan et al., 2018) show that when *zombie firms* delay their exit or restructuring, they drag down aggregate productivity by stifling reallocation of resources towards healthier firms, and by deterring the entry of potentially more innovative and younger firms.

But how do we spot a *zombie firm*? There is no consensus on the exact meaning of the *zombie* attribute beyond its evocative power. So far, scholars just adopted different working thresholds based on a proxy assessment of one or more available financial indicators in the absence of more precise guidance from theory. Ideally, one should consider firms' competitiveness and financial constraints in a dynamic perspective, having in mind the entire horizon of future events. After taking into account all future threats and opportunities, and

how they mirror on the flows of firm-level profits and net financial payments, one could obtain the theoretical equivalent of *zombie firms*. In the absence of the latter, an empirical identification is left to the creativity of scholars and practitioners. Caballero et al. (2008) define *zombies* as firms that receive subsidies under the form of bank credit after observing how interest payments compared to an estimated benchmark of debt structure and market interest rates. McGowan et al. (2018) assume that *zombies* are old firms that have persistent problems meeting their interest payments, although they center a policy discussion on the macroeconomic impact of low-productivity firms. Bank of Korea (2013) looks explicitly at when the interest coverage ratio is lower than one over three years. Bank of England (2013) overlooks financial management and considers firms that have both negative profits and negative value-added, hence focusing on the core activity of a company.

According to us, the direction of previous works is clear: one wants to deduce from actual financial accounts the future 'viability' of companies. If they do not seem in good shape now, the firm will likely be in trouble in the next future. Taken from this perspective, the empirical classification of *zombie firms* is a perfect case study for the application of machine learning techniques to firm-level data. It is a call for using in-sample information to predict an event that is out-of-sample.

Empowered by the previous intuition, we propose an identification of *zombies* starting from predictions of failures obtained in Section 3. We propose to classify as *zombies* those firms that are located on the right tail of the risk distribution, and for which the chances to recover from financial distress are minimal.

In notation, let us start considering the deciles along the predictions $f(\mathbf{X}_{i,t-1})$, where each $q_{j,t}$ is the threshold for the jth decile of failing probability at time t:

$$Q_{j,t} = \begin{cases} 1 & \text{if } f(\mathbf{X}_{i,t-1}) \geq q_{j,t} \ \cap \ Y_{i,t} \neq 1, \\ 0 & \text{otherwise.} \end{cases}$$

where $Y_{i,t} \neq 1$ indicates that the ith firm did not fail yet at time t.

Table 6: Transitions across deciles of risk

$t/t+1$	9th decile $t+1$	8th decile $t+1$	7th decile $t+1$	6th decile $t+1$	Below 6th decile $t+1$	Total $t+1$
9th decile t	0.64	0.24	0.07	0.02	0.03	1.00
8th decile t	0.22	0.39	0.21	0.06	0.12	1.00
7th decile t	0.08	0.20	0.28	0.23	0.21	1.00
6th decile t	0.04	0.08	0.23	0.30	0.35	1.00
Below 6th decile t	0.01	0.02	0.03	0.06	0.88	1.00

Figure 4: Transitions across deciles of risk: a visualization
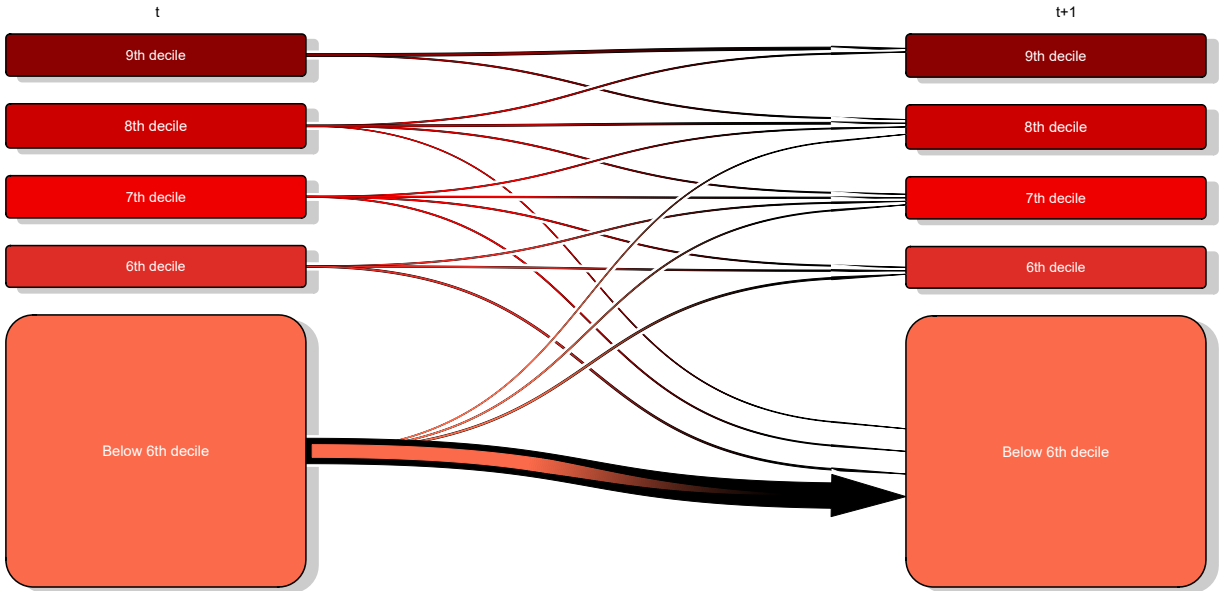
In Table 6, we report a transition matrix for the firms that did not fail, based on elaborations over the entire period of analyses in 2008-2017. In Figure 4, we also visualize a Sankey plot using the same information (Kennedy and Sankey, 1898). We observe that firms that our predictions locate beyond the 9th decile in a representative year t do not have a high chance to improve in $t+1$. See also the constant trends we report in Appendix Figure A3. In fact, the greatest majority of them (64%) gets stuck in the same highest-risk category,

24% transit to the 8th decile, and only 12% are able to recover and reach a more reasonable level of financial distress. Interestingly, the 9th decile is quite difficult to reach from the bottom of the distribution, as only 8%, 4% and 1% of firms that are below the 8th, 7th and 6th decile, respectively, are observed to transit to a situation of highest distress. Moreover, about 88% of company stay below the 6th decile in the entire period of analysis.

In general, we can say that, based on the information we observe, a firm in good shape does not easily shift into financial distress, but when it does it is difficult to get out of it. Against this evidence, it makes sense for us to set a working threshold at the 9th decile, as we realistically would encompass most difficult situations of the business life. Still, we want to allow firms a period of probation to discount some firms' break-even strategies in the short run, as for example in the case of newly-born firms and start-ups. For all the above reasons, we will define *zombie firms* those firms that persist for at least three years beyond the 9th decile of the risk distributions that we estimate based on the algorithm proposed in Section 3.

4.1 *Zombies* in Italy

Following the above definition, we report the share of *zombies* in Italy in Figure 5, and we compare them with the GDP growth rates observed in our period of analysis. Interestingly, we find that a range between 2.6% and 4.5% of manufacturing firms are teetering on bankruptcy. The share of *zombies* is higher immediately after the country went through a financial crisis in 2011, and it decreases in recent years from 2014 onward. In fact, the presence of *zombies* seems to be in a relationship with the economic cycle. The latter is a stylized fact that is worth further analyses, beyond the scope of this paper. We presume that, in times of crisis, many firms can be pushed on the verge of bankruptcy.

Finally, in Table 7 we briefly show what makes Italian *zombies* different from the other companies. We find that they have on average a lower productivity, 21.3% and 44.4% if measured as TFP or labor productivity, respectively. Moreover, they are on average consis-

tently smaller: they sell 120% less and employ 76% less than the otherwise representative healthy company.

Figure 5: *Zombie firms* and the economic cycle
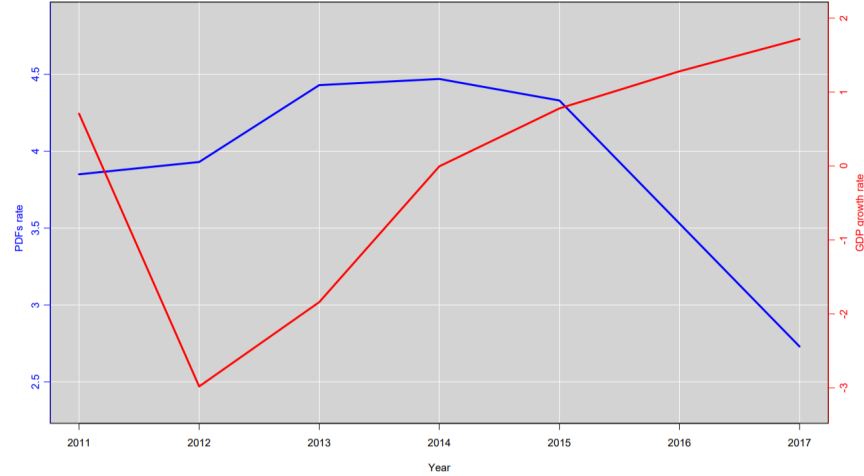
Note: The share of *zombies* measured on the left axis is compared against the nominal GDP growth rates on the right axis, sourced from the World Bank, in the period 2011-2017. *Zombie firms* are firms located on the right tail (9th decile) of the predicted risk distributions for at least three consecutive years.

Table 7: Productivity and size premia for *zombies* vs. healthy firms

Indicator (in logs)	Coeff.	Std. Error	N. obs.	Adj. R squared
Total Factor Productivity	-.213***	(.049)	568,288	.257
Labor Productivity	-.444***	(.038)	577,485	.139
Sales	-1.198***	(.055)	595,609	.189
Employees	-.760***	(.034)	578,532	.112

Note: We report coefficients of binary regressions where the dependent variable is an indicator of firm's productivity or size equal to one if the firm is classified as a *zombie* in at least one year of our sample. *Zombie firms* are firms located on the right tail (9th decile) of the predicted risk distributions for at least three consecutive years. Fixed effects at the region and industry level. Errors are clustered by region and industry.

5 Conclusions

In this contribution, we show how statistical learning can derive non-trivial information on a battery of financial indicators, and we successfully classify firms in risk categories after training on past failures. Our preferred algorithm is the BART-MIA, which outperforms other well-known econometric and machine learning techniques, especially in the presence of missing patterns of financial accounts.

Thanks to the machine learning framework, we can reduce prediction errors also against otherwise traditional tools for credit scoring, like the Z-scores and the Distance-to-Default. Therefore, we propose to classify as *zombie firms* the ones that persist in high-risk status because they locate on the right tail of our predictions for at least three years, beyond the 9th decile of risk, where we find that the chances to recover to smaller distress are minimal.

Previous references suggest how the identification of *zombies* may be crucial for financial institutions to avoid a waste of credit resources in insolvent companies when the latter outlive the market only thanks to some mechanisms of adverse selection generated by imperfect financial markets. Yet, we believe that our exercise can be useful from a more general perspective, to spot a share of an economy that is in trouble. The issue is all the more critical since recent studies discuss how *zombies* could hamper the growth potential of many countries, as they stifle the reallocation of productive resources in modern economies (Andrews et al., 2017; Banerjee and Hofmann, 2018; Andrews and Petroulakis, 2019).

In fact, we find that Italian *zombies* register lower levels of total factor productivity, while they are to find among smaller firms. Interestingly, we observe a possible relationship with the economic cycle that would be worth further investigation in the next future. The share of *zombies* is higher in the aftermath of twin financial crises in 2008 and 2011, but then it decreases in the latest years when recovery was underway. Unfortunately, as things were starting to get slowly better, the recent pandemic crisis has most probably pushed a big bunch of firms in a danger zone of high financial distress.

We cannot make predictions for a post-pandemic scenario yet. Still, we presume that the

problem of separating the companies that can stay on their feet alone from the ones that conceal their insolvency will be all the more pressing in the next future when the policymakers start to implement financial support programs. The challenge is to avoid a misallocation of resources and curb a much-needed economic recovery.

References

Ackerberg, D. A., Caves, K., Frazer, G., 2015. Identification properties of recent production function estimators. Econometrica 83 (6), 2411–2451.

Aghion, P., Bergeaud, A., Cette, G., Lecat, R., Maghin, H., 2019. Coase lecture - the inverted-u relationship between credit access and productivity growth. Economica 86 (341), 1–31.

Ahrens, A., Hansen, C. B., Schaffer, M. E., 2019. lassopack: Model selection and prediction with regularized regression in stata. arXiv preprint arXiv:1901.05397.

Ahrens, A., Hansen, C. B., Schaffer, M. E., 2020. lassopack: Model selection and prediction with regularized regression in stata. The Stata Journal 20 (1), 176–235.

Altman, E. I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance 23 (4), 589–609.

Altman, E. I., et al., 2000. Predicting financial distress of companies: revisiting the z-score and zeta models. Stern School of Business, New York University, 9–12.

Andrews, D., McGowan, M. A., Millot, V., et al., 2017. Confronting the zombies: Policies for productivity revival. Tech. rep., OECD Publishing.

Andrews, D., Petroulakis, F., Feb. 2019. Breaking the shackles: Zombie firms, weak banks and depressed restructuring in Europe. Working Paper Series 2240, European Central Bank.

Athey, S., 2018. The impact of machine learning on economics. In: The Economics of Artificial Intelligence: An Agenda. University of Chicago Press.

Banerjee, R. N., Hofmann, B., 2018. The rise of zombie firms: causes and consequences. BIS Quarterly Review (September), 67–78.

Bank of England, 2013. Inflation report. Tech. rep., Bank of England, London.

Bank of Korea, 2013. Financial stability report. Tech. rep., Bank of Korea, Seoul.

Bargagli-Stoffi, F. J., De Beckker, K., De Witte, K., Maldonado, J. E., 2020a. Assessing

sensitivity of predictions. a novel toolbox for machine learning with an application on financial literacy. Working paper.

Bargagli-Stoffi, F. J., Niederreiter, J., Riccaboni, M., 2020b. Supervised learning for the prediction of firm dynamics. Mimeo.

Bellman, R. E., 1961. Adaptive Control Processes. Princeton University Press.

Belloni, A., Chernozhukov, V., Hansen, C., Kozbur, D., 2016a. Inference in high-dimensional panel models with an application to gun control. Journal of Business & Economic Statistics 34 (4), 590–605.

Belloni, A., Chernozhukov, V., Wei, Y., 2016b. Post-selection inference for generalized linear models with many controls. Journal of Business & Economic Statistics 34 (4), 606–619.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of political economy 81 (3), 637–654.

Bonfim, D., Cerqueiro, G., Degryse, H., Ongena, S., 2020. On-site inspecting zombie lending. Available at SSRN 3530574.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J., 1984. Classification and regression trees. Belmont, CA: Wadsworth & Brooks.

Brodersen, K. H., Ong, C. S., Stephan, K. E., Buhmann, J. M., 2010. The balanced accuracy and its posterior distribution. In: 2010 20th International Conference on Pattern Recognition. IEEE, pp. 3121–3124.

Bugamelli, M., Lotti, F., Amici, M., Ciapanna, E., Colonna, F., DAmuri, F., Giacomelli, S., Linarello, A., Manaresi, F., Palumbo, G., Sc, F., Jan. 2018. Productivity growth in Italy: a tale of a slow-motion change. Questioni di Economia e Finanza (Occasional Papers) 422, Bank of Italy, Economic Research and International Relations Area.

Buuren, S. v., Groothuis-Oudshoorn, K., 2010. mice: Multivariate imputation by chained equations in r. Journal of statistical software, 1–68.

Caballero, R. J., Hoshi, T., Kashyap, A. K., 2008. Zombie lending and depressed restructuring in japan. American Economic Review 98 (5), 1943–77.

Calligaris, S., Del Gatto, M., Hassan, F., Ottaviano, G. I., Schivardi, F., et al., 2016. Italys productivity conundrum. a study on resource misallocation in italy. Tech. rep., Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Calligaris, S., Del Gatto, M., Hassan, F., Ottaviano, G. I. P., Schivardi, F., 09 2018. The productivity puzzle and misallocation: an Italian perspective. Economic Policy 33 (96), 635–684.

Chen, J., Chen, Z., 2008. Extended bayesian information criteria for model selection with large model spaces. Biometrika 95 (3), 759–771.

Chipman, H. A., George, E. I., McCulloch, R. E., et al., 2010. Bart: Bayesian additive regression trees. The Annals of Applied Statistics 4 (1), 266–298.

Cravino, J., Levchenko, A. A., 11 2016. Multinational Firms and International Business Cycle Transmission*. The Quarterly Journal of Economics 132 (2), 921–962.

Devijver, P. A., Kittler, J., 1982. Pattern recognition: A statistical approach. Prentice hall.

Fawcett, T., 2006. An introduction to roc analysis. Pattern Recognition Letters 27 (8), 861–874.

Fazzari, S. M., Hubbard, R. G., Petersen, B. C., 1988. Financing Constraints and Corporate Investment. Brookings Papers on Economic Activity 19 (1), 141–206.

Ferrando, A., Ruggieri, A., 2018. Financial constraints and productivity: Evidence from euro area companies. International Journal of Finance & Economics 23 (3), 257–282.

Gal, P. N., 2013. Measuring total factor productivity at the firm level using oecd-orbis.

Gopinath, G., Kalemli-Özcan, Ş., Karabarbounis, L., Villegas-Sanchez, C., 2017. Capital allocation and productivity in south europe. The Quarterly Journal of Economics 132 (4), 1915–1967.

Gruszczyński, M., 2019. On unbalanced sampling in bankruptcy prediction. International Journal of Financial Studies 7 (2), 28.

Hadlock, C. J., Pierce, J. R., 2010. New evidence on measuring financial constraints: Moving beyond the kz index. The Review of Financial Studies 23 (5), 1909–1940.

Hanley, J. A., McNeil, B. J., 1982. The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology 143 (1), 29–36.

Hothorn, T., Hornik, K., Zeileis, A., 2006. Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical statistics 15 (3), 651–674.

Kalemli-Ozcan, S., Sorensen, B., Villegas-Sanchez, C., Volosovych, V., Yesiltas, S., 2015. How to construct nationally representative firm level data from the orbis global database. Tech. rep., National Bureau of Economic Research.

Kapelner, A., Bleich, J., 2015. Prediction with missing data via bayesian additive regression trees. Canadian Journal of Statistics 43 (2), 224–239.

Kennedy, A. B. W., Sankey, H. R., 1898. The thermal efficiency of steam engines. Minutes 134, Proceedings of the Institution of Civil Engineers.

Kleinberg, J., Ludwig, J., Mullainathan, S., Obermeyer, Z., 2015. Prediction policy problems.

American Economic Review 105 (5), 491–95.

Loh, W.-Y., 2002. Regression tress with unbiased variable selection and interaction detection. Statistica sinica, 361–386.

Makridakis, S., Wheelwright, S. C., Hyndman, R. J., 2008. Forecasting methods and applications. John wiley & sons.

McGowan, M. A., Andrews, D., Millot, V., 2018. The walking dead? zombie firms and productivity performance in oecd countries. Economic Policy 33 (96), 685–736.

Merton, R. C., 1974. On the pricing of corporate debt: The risk structure of interest rates. The Journal of finance 29 (2), 449–470.

Modigliani, F., Miller, M. H., 1958. The cost of capital, corporation finance and the theory of investment. The American Economic Review 48 (3), 261–297.

Moscatelli, M., Narizzano, S., Parlapiano, F., Viggiano, G., et al., 2019. Corporate default forecasting with machine learning. Tech. rep., Bank of Italy, Economic Research and International Relations Area.

Mullainathan, S., Spiess, J., 2017. Machine learning: an applied econometric approach. Journal of Economic Perspectives 31 (2), 87–106.

Nickell, S., Nicolitsas, D., 1999. How does financial pressure affect firms? European Economic Review 43 (8), 1435–1456.

Odén, A., Wedel, H., et al., 1975. Arguments for fisher's permutation test. The Annals of Statistics 3 (2), 518–520.

Orbis, 2020. Orbis company information across the globe. Bureau van diijk, A Moody's Analytics Company.
URL https://orbis.bvdinfo.com/

Peek, J., Rosengren, E. S., 2005. Unnatural selection: Perverse incentives and the misallocation of credit in japan. American Economic Review 95 (4), 1144–1166.

Rajan, R. G., Zingales, L., December 1995. What Do We Know about Capital Structure? Some Evidence from International Data. Journal of Finance 50 (5), 1421–1460.

Rungi, A., Biancalani, F., 2019. Heterogeneous firms and the north–south divide in italy. Italian Economic Journal, 1–23.

Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. PloS one 10 (3).

Schivardi, F., Sette, E., Tabellini, G., 2017. Credit misallocation during the european financial crisis. Working Paper Series 11901, CEPR.

Shmueli, G., et al., 2010. To explain or to predict? Statistical science 25 (3), 289–310.

Twala, B., Jones, M., Hand, D. J., 2008. Good methods for coping with missing data in decision trees. Pattern Recognition Letters 29 (7), 950–956.

Van der Laan, M. J., Polley, E. C., Hubbard, A. E., 2007. Super learner. Statistical Applications in Genetics and Molecular Biology 6 (1).

Van Rijsbergen, C. J., 1979. Information retrieval. London: Butterworths.

Weinblat, J., 2018. Forecasting european high-growth firms-a random forest approach. Journal of Industry, Competition and Trade 18 (3), 253–294.

White, T. K., Reiter, J. P., Petrin, A., 2018. Imputation in us manufacturing data and its implications for productivity dispersion. Review of Economics and Statistics 100 (3), 502–509.

Zhou, L., 2013. Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. Knowledge-Based Systems 41, 16–25.

Appendix: Tables and Graphs

Table A1: Panel (A): List of predictors for firms' failures.

Variables	Description
Operating Revenues, Material Costs, Costs of Employees, Added Value, Taxation, Tax and Pensions' Payables, Financial Revenues, Financial Expenses, Interest Payments, Number of Employees, Net Income, Cash Flow, EBITDA (Earnings before interest, Taxation, Depreciation and Amortization), Total Assets, Fixed Assets, Intangible Fixed Assets, Current Assets, Shareholders' Funds, Retained Earnings, Long-Term Debt, Loans, Current Liabilities	Original financial accounts expressed in euro.
Corporate Control	A binary variable equal to one if a firm belongs to a corporate group.
Number of Patents	The portfolio of patents granted to a firm by patent offices (*Dummy Patents* equal to 0 if the firm issued no patents, and 1 otherwise).
Number of Trademarks	The total number of trademarks issued to the firm by national or international trademark offices (*Dummy Trademarks* equal to 0 if the firm issued no trademarks, and 1 otherwise).
Consolidated Accounts	A binary variable equal to one if the firm consolidates accounts of its subsidiaries
NACE rev. 2	A 4-digit industry affiliation following European classification NACE rev. 2.
NUTS 2 regions	The region in which the company is located.
Productive Capacity	it is an indicator of investment in productive capacity computed as $\frac{Fixed\ Assets_t}{Fixed\ Assets_{t-1}+Depreciation_{t-1}}$.
Capital Intensity	Fixed Assets/Number of employees.
Labour Productivity	It is a ratio of added value over the number of employees.
Interest Benchmarking	It is a *zombie* proxy proposed by Caballero et al. (2008) and calculated as $R^* = rs_{t-1}BS_{i,t-1} + (\frac{1}{5}\sum_{j=1}^{5} rl_{t-j})BL_{i,t-1} + rcb_{5y,t} \cdot Bonds_{i,t-1}$, where $BS_{i,t-1}$ are short-term bank loans, $BL_{i,t-1}$ are long-term bank loans, rs_{t-1} are the average short-term prime rate in year t, rl_{t-j} is the average long-term prime rate in year t, *Bonds* are the total outstanding bonds, $rcb_{5y,t}$ is the minimum observed rate on any convertible corporate bond issued over the previous five years.

## Table A1: Panel (B): List of predictors for firms' failures.

Variables	Description
Interest Coverage Ratio (ICR)	It is calculated as EBIT/Interest Expenses. When it is less than one, Bank of Korea (2013) and McGowan et al. (2018) assume a firm is a *zombie*.
Financial Misallocation	It is a binary indicator adopted by Schivardi et al. (2017) for catching *zombie lending*, based on both $ROA \frac{\frac{1}{3}\sum_{t=1}^{3} EBITDA_t}{Total\ Assets} < prime$ and $Leverage = \frac{Financial\ Debt}{Total\ Assets} > \tilde{L}$, where *prime* is the measure of the cost of capital for firms with a Z-score equal to 1 or 2, and where $\tilde{L}$ is the median value of leverage in the current year for firms that exited in two following years.
TFP	It is the Total Factor Productivity of a firm computed as in Ackerberg et al. (2015).
Financial Constraints	It is a proxy of financial constraints as in Nickell and Nicolitsas (1999), calculated as a ratio between interest payments and cash flow
Enterprise Value (listed companies only)	It is a synthetic value calculated considering other 10 comparable listed companies in terms of Market Capitalization, Minority Interest, Preferred shares, Long Term debt, Loans, Other short term debt, Cash.
Negative Added Value	It is a binary variable adopted by Bank of Korea (2013) for zombie firms, equal to one when Added Value is negative, i.e. when the value of sold output is less than purchases of intermediate inputs.
Size-Age	It is a synthetic indicator proposed by Hadlock and Pierce (2010), equal to $-0.737 * \log(totalassets) + 0.043 * \log(totalassets))^2 - 0.040 * age$.
Profitability	Calculated as $EBITDA/Total\ Assets$, and adopted by Schivardi et al. (2017) as a control for *zombie lending*
Financial Sustainability	It is a ratio calculated as Financial Expenses over Operating Revenues.
Capital Adequacy Ratio	It is a ratio of Shareholders' Funds over Short and Long Term Debts.
Liquidity Ratio	(Current/Assets - Stocks)/Current/Liabilities
Solvency Ratio	(Shareholders funds / (Non current liabilities + Current liabilities)) * 100
Liquidity Returns	It is the ratio of cash flow over total asset
Tax and Pension Payables	It is the ratio of the sum of tax and pension payables over total assets.

Table A2: Missing predictors and firms' failures - Chi-square tests

	Firm's failure				Test Statistic
	0		1		
	$N = 287587$		$N = 17319$		
Interest Benchmarking : 0	38%	(110524)	61%	(10530)	χ_1^2=3414.25, P<0.001
Interest Benchmarking : 1	62%	(177063)	39%	(6789)	
Interest Coverage Ratio : 0	37%	(105907)	49%	(8422)	χ_1^2=970.93, P<0.001
Interest Coverage Ratio : 1	63%	(181680)	51%	(8897)	
Negative Value Added : 0	34%	(98014)	63%	(10915)	χ_1^2=5958.81, P<0.001
Negative Value Added : 1	66%	(189573)	37%	(6404)	
Financial Constraint : 0	37%	(105904)	49%	(8419)	χ_1^2=968.27, P<0.001
Financial Constraint : 1	63%	(181683)	51%	(8900)	
Financial Misallocation : 0	39%	(112560)	54%	(9276)	χ_1^2=1415.82, P<0.001
Financial Misallocation : 1	61%	(175027)	46%	(8043)	
Total Factor Productivity : 0	36%	(104345)	38%	(6600)	χ_1^2=23.52, P<0.001
Total Factor Productivity : 1	64%	(183242)	62%	(10719)	
Solvency Ratio : 0	41%	(118851)	63%	(10897)	χ_1^2=3115.5, P<0.001
Solvency Ratio : 1	59%	(168736)	37%	(6422)	
Liquidity Ratio : 0	42%	(119357)	72%	(12543)	χ_1^2=6362.72, P<0.001
Liquidity Ratio : 1	58%	(168230)	28%	(4776)	
Size-Age : 0	42%	(120260)	75%	(12989)	χ_1^2=7310.19, P<0.001
Size-Age : 1	58%	(167327)	25%	(4330)	
Liquidity Returns : 0	39%	(112561)	54%	(9277)	χ_1^2=1416.88, P<0.001
Liquidity Returns : 1	61%	(175026)	46%	(8042)	
Labour Productivity : 0	34%	(97253)	36%	(6221)	χ_1^2=32.23, P<0.001
Labour Productivity : 1	66%	(190334)	64%	(11098)	
Profitability : 0	37%	(105907)	49%	(8422)	χ_1^2=970.93, P<0.001
Profitability : 1	63%	(181680)	51%	(8897)	
Financial Sustainability : 0	41%	(117294)	64%	(11119)	χ_1^2=3673.94, P<0.001
Financial Sustainability : 1	59%	(170293)	36%	(6200)	
Capital Intensity : 0	36%	(104122)	42%	(7325)	χ_1^2=261.17, P<0.001
Capital Intensity : 1	64%	(183465)	58%	(9994)	

Note: Chi-square tests for the null hypothesis that missing predictors do not correlate with the event of failure. Number of observations in parentheses.

Figure A1: Patterns of missing values from a random subsample of 10,000 observations
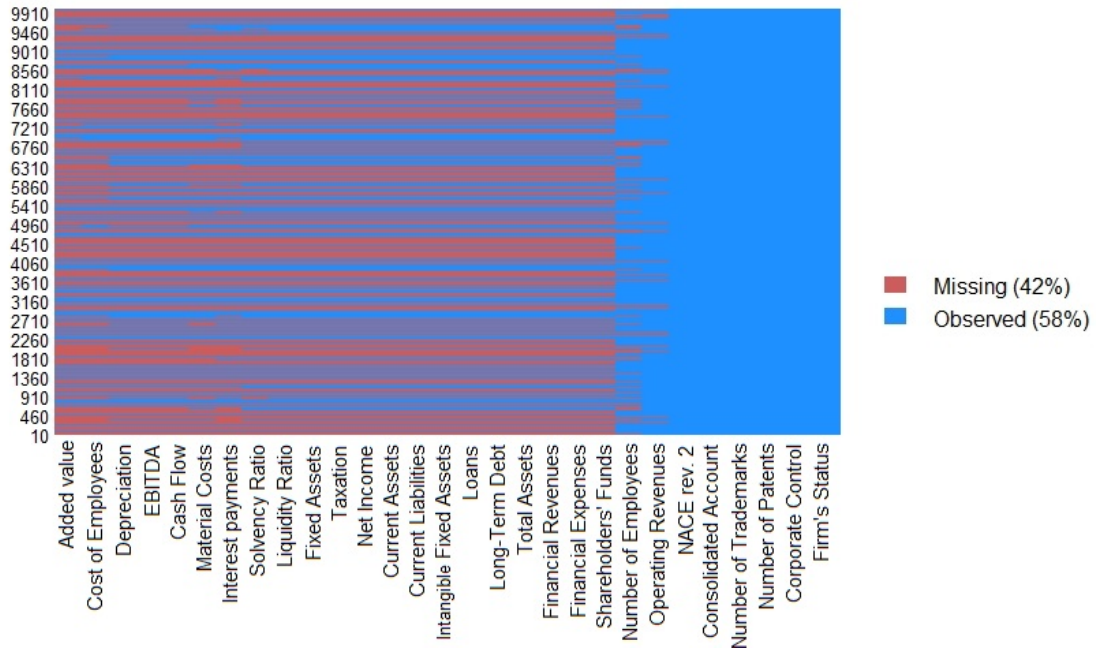
Panel (A): Missing values of original financial accounts
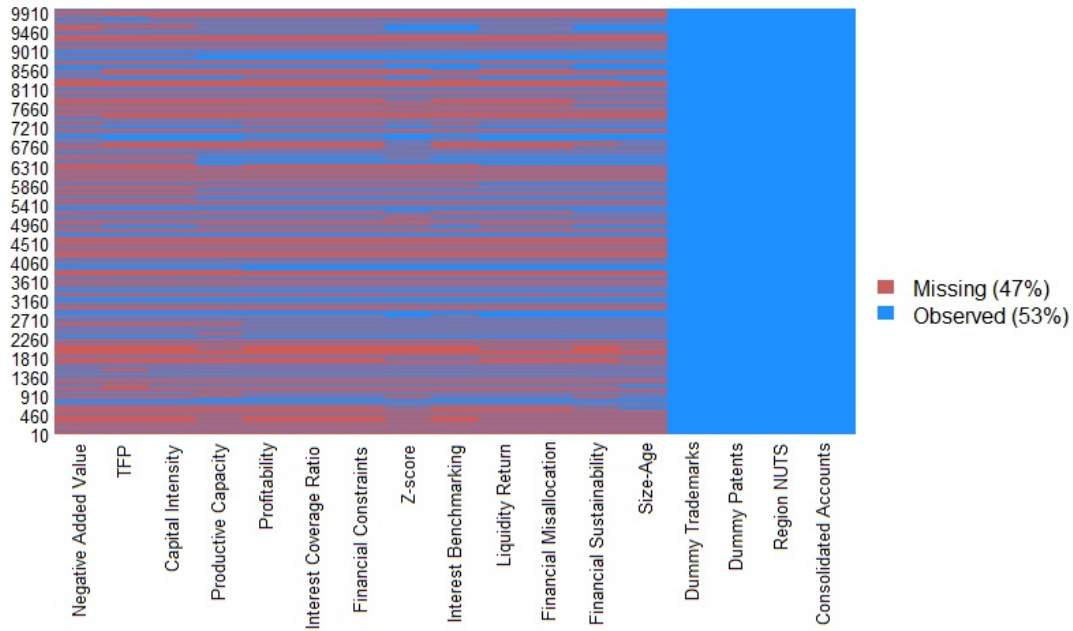
Panel (B): Missing values of indicators based on original financial accounts.

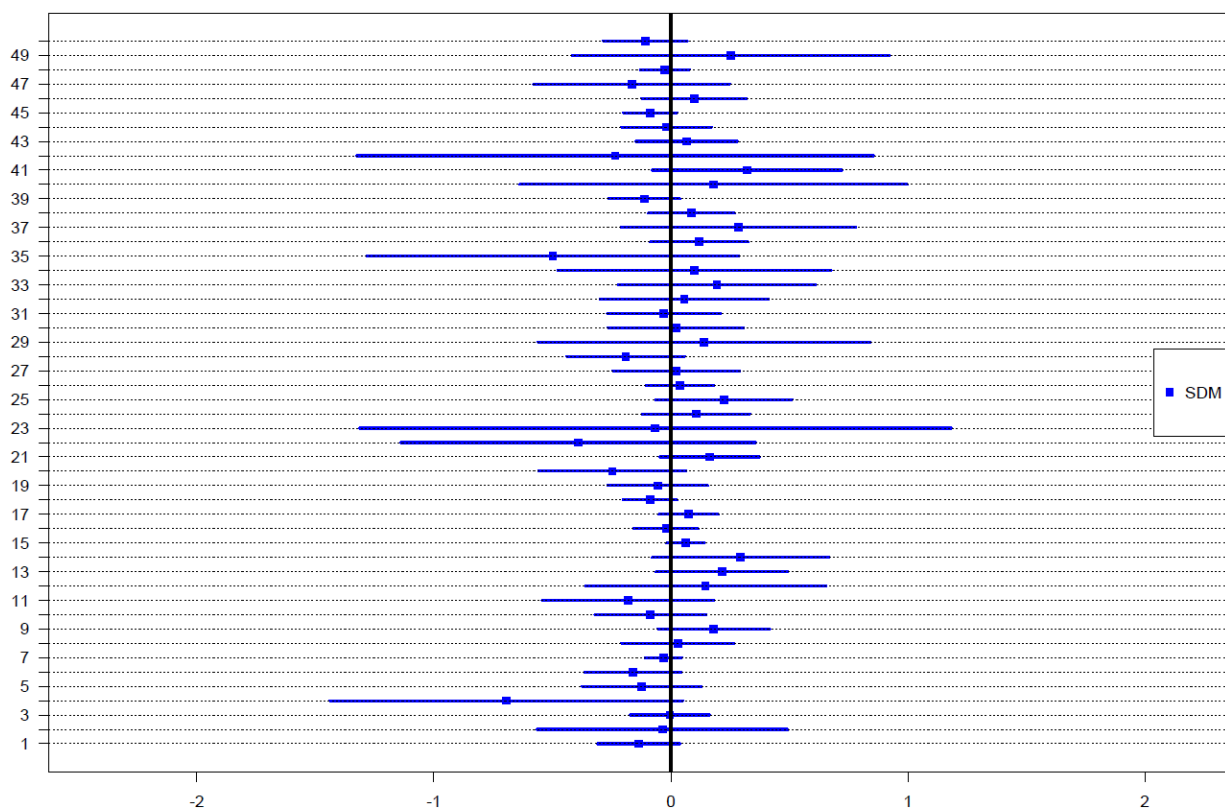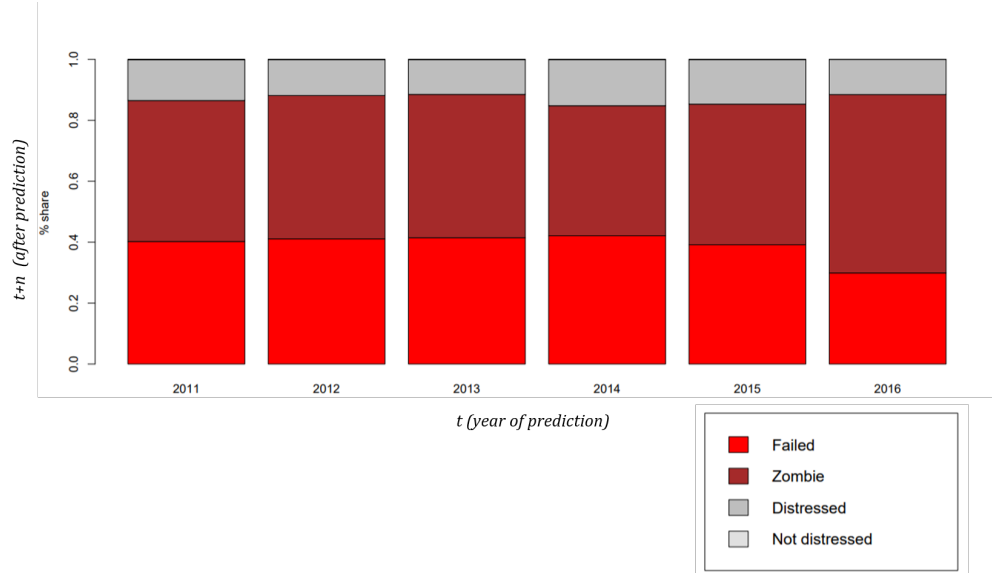Figure A2: Standardized Differences in the Predictions of Augmented and Original BART

Figure A3: Transitions after predictions of a *zombie* status

Note: The bars of the graph indicate the transition of *zombie* firms in the years following predictions: i) to failure (bright red); ii) to a relatively lower distress (dark grey); iii) to a permanence in a *zombie* status. Negligible and non-visible is the share of firms that transit to an area of no distress (bright grey), below the median prediction. Please note that we cannot report year 2017, as we cannot compare with actual observations in following years.