Are you a Zombie? A Supervised Learning Method to Classify Unviable Firms and Identify the Determinants*

Angela De Martiis† Thomas L.A. Heil ‡ Franziska J. Peter§

May 8, 2021

Abstract

We examine the determinants of zombie companies using a comprehensive firm-level dataset of public corporations from Europe and the United States. We show that US zombie companies differ from their European peers on a modest number of firm-specific and industry-specific factors, but follow a similar pattern. Using decision trees, we document that income and leverage-related variables are among the main drivers classifying zombie companies in Europe and in the US. Shareholders' interests are however relevant to separate zombie from non-zombie corporations in the US. We observe a frequent mislabeling of zombie firms into other unviable types of firms. To account for this, we also examine the determinants of distressed firms and compare them to the zombie. We find that zombie and distressed are not comparable types of companies, rather companies at a different stage of their financial unviability. We also document that zombification is especially a European phenomenon, while distressed-type of firms are mostly populating the US economy. We find no major differences in terms of zombie company-specific determinants before and after the global financial crisis.

JEL codes: C55, C63, D22, E44, G32, G33
Keywords: zombie firms, financial constraint, decision trees, distressed firms

*We thank Philip Valta, Marc Brunner, and participants at the 2020 IFM Brown Bag seminar for useful comments and discussions.

†Corresponding author. University of Bern, Institute for Financial Management, Engehaldenstrasse 4, 3012 Bern, Switzerland. angela.demartiis@ifm.unibe.ch.

‡Zeppelin University, Chair of Empirical Finance and Econometrics, Am Seemooser Horn 20, 88045 Friedrichshafen, Germany. thomas.heil@zu.de.

§Zeppelin University, Chair of Empirical Finance and Econometrics, Am Seemooser Horn 20, 88045 Friedrichshafen, Germany. franziska.peter@zu.de.

1 Introduction

Long after the global financial crisis, the phenomenon of non-viable (i.e. zombie) firms remains a concern and stirs up growing debates among scholars and policymakers. Early studies define zombie firms as insolvent companies with little hope of recovery, but avoiding failure thanks to the support from their banks (Hoshi 2006).

Existing research investigates the reasons why these firms remain alive and concentrates on the consequences of what became a widespread phenomenon, while the media community narrates about the rise of the zombie that, supported by state-backed credit, spend their cash servicing debt instead of investing it.[1] There is evidence that the share of zombie firms has trended up since the late 1980s (Banerjee and Hofmann 2018; McGowan, Andrews, and Millot 2018) and that they appear to be linked to weakly capitalized banks (Caballero, Hoshi, and Kashyap 2008; Giannetti and Simonov 2013; Schivardi, Sette, and Tabellini 2017; Storz, Koetter, and Setzer 2017; Andrews and Petroulakis 2017; Acharya, Eisert, Eufinger, and Hirsch 2019; Acharya, Borchert, Jager, and Steffen 2020).

Yet, we are still far from being able to understand the characteristics and driving factors of zombie firms. There are some evident drivers, such as the size of the company and the industry (Hoshi 2006), but the literature lacks a thorough empirical investigation of the characteristics of such companies as well as the potential similarities and differences among zombie companies across countries and time.

Using two detailed and comprehensive sets of publicly listed companies' firm-level datasets from Compustat North America and Compustat Global Fundamentals Annual, we perform an immediate geographical inspection of the share of zombie companies. Figure 1 illustrates that the zombie phenomenon is not limited to a small subset of countries. Despite this fact, previous research has mostly focused on Japan or on a sample of specific countries (Caballero, Hoshi, and Kashyap 2008; McGowan, Andrews, and Millot 2018; Banerjee and Hofmann 2018; Banerjee and Hofmann 2020; Acharya, Crosignani, Eisert, and Eufinger 2020). In this study, we cover 32 European countries plus the United States on a time frame of over two decades that allows us to observe several business cycle developments.

Our empirical analysis delves deeper into the firm-specific characteristics of non-viable firms. We start by identifying the companies that are considered zombie. This represents a crucial point, given that the literature lacks a disciplined approach towards identifying such unviable firms, often mislabeling them. The main measure follows Banerjee and Hofmann (2020), while other measures are calculated as robustness.

To understand the country-specific firm-level characteristics of zombie firms, we employ a supervised learning method: the classification trees. When applying decision trees to our high dimensional dataset containing balance sheet, accounting, and fundamental variables, we refrain from making any a priori assumptions and instead let the data and algorithms select the main drivers. We repeat our analysis for different time periods in order to examine whether, and to which extent, firm-specific drivers vary throughout

[1] Financial Times article of January 2013 on *"Companies: the rise of the zombie"*, available at: `https://www.ft.com/content/7c93d87a-58f1-11e2-99e6-00144feab49a`.

time. In addition, we repeat the same binary structure to carefully review and analyze the characteristics of distressed and non-distressed companies and compare the latter to zombie firms. To the best of our knowledge, this is the first study that employs such a refined machine learning method to explore the firm-specific characteristics and behavior of zombie firms across countries and time.

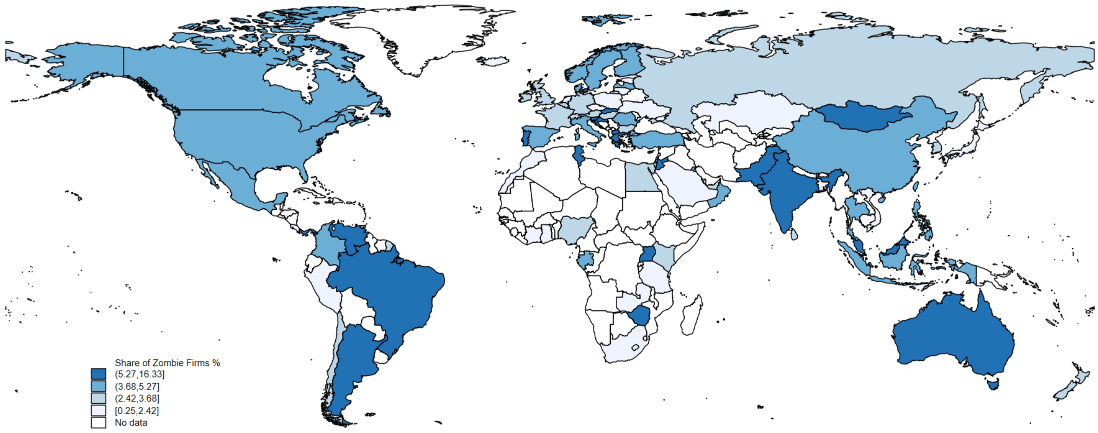


Figure 1: Global Zombie Shares. The map shows the presence of zombie companies by country and visualizes the share of zombie firms in the world. The map is scaled in different shades of blue according to the severity of the phenomenon. The countries for which we have no data are those displayed in white color. The countries that register the highest share of zombie, in dark blue, are Portugal, Greece, Cyprus, Croatia, Macedonia, Slovenia and Slovakia in Europe; Venezuela, Brazil and Argentina in Latin America; Jordan, Pakistan, India, Mongolia, Malaysia and Australia in Asia; Tunisia, Uganda and Zimbabwe in Africa. Source: Authors' projections on firm-level data from Compustat Global and Compustat North America.

In a second step, we implement a set of multi-classification trees that allow us to examine the systematic differences among zombie, non-zombie, healthy, recovered zombie, and other distressed-type of companies.

In particular, we examine the firm-specific characteristics of zombie firms in Europe and in the United States and, differently from Hoshi (2006), we document that zombie firms are companies with healthy periods in-between financially unsound years and, from the data processing, we can observe that they are likely to recover. The results further show that, US zombie firms differ from their European peers on a modest number of firm-specific and industry-specific factors, but follow a similar pattern. In this regard, we document that income and leverage-related variables are among the main drivers classifying zombie companies in Europe and in the United States. However, for US corporations shareholder's equity is a relevant driver that categorizes zombie versus non-zombies. Contrary to classic statistical methods, a decision tree algorithm allows us to not only detect the most important variables categorizing zombie versus non-zombie, but also extract which parts of a firm's income or debt are likely to predict the zombie status.

The results indicate that these specific characteristics remain relevant before and after the global financial crisis. Preliminary evidence suggest a frequent mislabeling of zombie firms into other distressed-type of firms. To account for this, we examine the firm-specific characteristics of distressed versus non-distressed firms and compare them to zombie-like firms to understand whether the two categories should be treated differently. This test allows us to document that zombie and distressed firms are often not comparable types of companies, rather firms at different stages of their financial unviability. The latter finding can yield relevant policy implications, given that zombie firms are often improperly treated as distressed companies.

In addition, the classification trees suggest that "zombification" is especially relevant among the European economies, where zombie companies are more prevalent, followed by the healthy. To the contrary, distressed-type of firms are mostly populating the US market where the two major classes of firms are the distressed and the healthy.

This study contributes to four strands of literature. The first, relates to the literature examining the so-called zombie firms, a phenomenon that was first investigated in reference to the Japanese banking crisis of the 1990s. With this respect, Caballero, Hoshi, and Kashyap (2008) explore the zombie lending behavior, a process in which large Japanese banks often engaged in sham loan restructurings in order to keep the credit flowing to otherwise insolvent borrowers. As a result, an increase in zombie firms generated a depression of the investments and of the employment growth of healthy companies, and distortions in the creation of jobs and productivity. Peek and Rosengren (2005) provide evidence that troubled Japanese banks allocated credit to highly indebted borrowers to avoid realizing the losses on their balance sheets. More recently, McGowan, Andrews, and Millot (2018) document an increase in the share of zombie companies also in the OECD economies, between 2003 and 2013. Schivardi, Sette, and Tabellini (2017) confirm that Italian zombie firms obtained credit from undercapitalized banks. The latter authors highlight the identification challenges that come with the analysis of the zombie phenomenon (Schivardi, Sette, and Tabellini 2020). Hoshi (2006) identifies zombie firms in Japan and investigates some of their main characteristics in a set of probit regressions. Within these studies, we contribute by examining zombie firms in Europe and in the United States and by identifying the characteristics of such non-viable firms with respect to non-zombie, healthy, recovered zombie, and other distressed-type of companies.

The second strand, relates to the literature on zombie firms and machine learning. Within this literature we are, to the best of our knowledge, the first to examine the characteristics of zombie firms across countries and time using a supervised learning algorithm that allows us to better classify the zombie and separate them from the non-zombie and other firms' categories, such as the distressed, the recovered, and the healthy. The large amount of data proves impractical to analyze the characteristics of non-viable firms via classic statistical models. We instead exploit an algorithmic modeling, precisely classification trees-like algorithm, to find the important drivers out of a broad range of explanatory variables. The algorithm behind a decision tree searches through the whole range of explanatory variables and subsequently finds the variables that better classify zombie versus non-zombie, or in a multi-classification tree setting, zombie versus other categories of firms. Within this strand, the only other paper using machine learning to predict firm failure is that of Bargagli Stoffi, Riccaboni, and Rungi (2020), which contrary to our study proposes an alternative zombie definition using data on Italian firms.

The third strand, draws upon the corporate finance literature examining firms' financial distress. Within this sizable literature, at the intersection between finance and accounting, Altman (1968) and Ohlson (1980) are among the first to have explored the concept of firm financial distress by proposing measurement techniques still widely used and recently complemented by market-value based measures, such as the one of Campbell, Hilscher, and Szilagyi (2008). We contribute to this literature by examining the firm-level characteristics of distressed firms in Europe and in the United States and comparing them to zombie companies. In the literature, zombie firms are often treated as distressed-type of firms. We thus exploit a classification tree algorithm to observe whether distressed companies can be compared to zombie firms or whether relevant differences in terms of firm-specific characteristics emerge. In performing this exercise, we confirm some of the known factors characterizing distressed companies, but we also shed light on unexplored differences between zombies and distressed that add to the empirical finance literature (Chan and Chen 1991; Campbell, Hilscher, and Szilagyi 2008; Skinner and Soltes 2011; Eisdorfer, A. Goyal, and Zhdanov 2018; Kahle and Stulz 2017), by documenting, among other factors, that US zombie firms are often characterized by shareholders' interests.

The last strand, relates to the literature examining regulatory and bankruptcy frameworks across countries. By analyzing the characteristics of zombie and non-zombie across countries, the existence of differing regulatory regimes plays a crucial role as evergreening incentives are stronger in countries with weak insolvency frameworks (McGowan, Andrews, and Millot 2017). Within the European countries, a more harmonized insolvency framework has the potential to make the long-term existence of zombie firms less of a concern. Existing studies account for differences in bankruptcy codes when examining defaults (Favara, Morellec, Schroth, and Valta 2017) and levered firms (Acharya and Subramanian 2009), and also differences between civil law and common law countries (Djankov, McLiesh, and Shleifer 2007). We add to this literature by highlighting the role played by differing regulatory frameworks in explaining cross-country differences among viable and non-viable firms. In this regard, we argue that the firm-specific differences between US and European zombie firms can be explained by unalike insolvency laws, where stringent insolvency regimes are more efficient at rehabilitating viable firms and liquidating non-viable ones (McGowan, Andrews, and Millot 2017).

The remainder of the paper proceeds as follows. Section 2 describes the data, defines alternative measures for zombie companies, discusses the rationale behind their existence, and shows empirical evidence for their prevalence. Section 3 presents the analysis of the zombie characteristics using decision trees, it outlines the methodology (3.1), presents the empirical analysis of zombie versus non-zombie decision trees (3.2), the classification trees of distressed versus non-distressed companies (3.3), followed by the multi-classification tree setting where the firm-level characteristics of zombie, distressed, recovered, and healthy firms are analyzed (3.4), and finally provides a benchmark analysis using logistic regressions (3.5). Section 4 concludes.

2 Data and Descriptive Statistics

2.1 Data

We use firm-level data from Compustat Global and Compustat North America Fundamentals Annual. The first database provides financial and market data about active and inactive public companies from more than 80 countries, including coverage of over 96% of European market capitalization with annual data history that goes back to 1978. The second database covers publicly listed companies from the United States and Canada. The rich data allows us to gather financial, balance sheet, and market data information covering several business cycle expansions and contractions in economic activity from the late 1980s to 2018. In addition, we add comprehensive stock price data from Thomson Reuters Datastream to Compustat Global dataset using the international securities identification number, ISIN codes, of each company from 1990 to 2018.

In terms of data pre-processing, we restrict both datasets to the years 1996-2018 and delete the observations with missing company unique identifier, the gvkey, and missing information on the fiscal year, fyear. We remove all gvkey-fyear duplicates and drop all year-company combinations that have less than 99% observations.[2] Moreover, we drop all variables that display missing values for more than 65% of their observations. Moreover, we restrict the datasets to contain only variables that appear in the Global and North America version of Compustat. Therefore, the procedure performs a first selection of variables leaving us with approximately 65 variables in both datasets.

Following previous studies on zombie firms, we exclude all companies belonging to the utilities, financial, insurance, and banking industries.[3] Additionally, we winsorize each variable at the 5% and 95% percentile and drop all observations below and above this threshold to reduce the effect of outliers. Ultimately, we impute the remaining missing values and find that K-Nearest-Neighbors imputation produces promising results in terms of predictive power.[4] This data preparation process yields a well-stocked dataset consisting of approximately 15000 observations per year for 32 European countries and 6000 observation per year for the United States.

With respect to firm-specific characteristics, we additionally compute a set of performance measures that are commonly used in the empirical finance literature to capture firm size, asset tangibility, profitability, risk, liquidity, market value, and growth opportunities. We use the latter information to understand whether, and to which extent, these measures contribute to the classification of a company as a zombie. In Appendix A we report the description and definition of the variables used.[5]

[2] We believe these are dead companies still listed.

[3] For the industry classification, we consider the GIC group which is based on the global industry classification standard (GICS), captured in Compustat by the variable *ggroup*, and developed by the S&P Dow Jones Indices and the MSCI.

[4] Additionally, we test mean-value imputation, however, this simple procedure produces to general results and might bias the dataset to much.

[5] In terms of firm-specific information, we follow Fama and French (2001), Frank and V. K. Goyal (2003), Myers (2001), Baker and Wurgler (2002), Chan and Chen (1991), Kahle and Stulz (2017), Jong, Kabir, and Nguyen (2008), Kayhan and Titman (2007), and Fan, Titman, and Twite (2012).

2.2 Zombie Measures

The existing literature provides different approaches to define a zombie company. Each of them has their own limitations, advantages, and disadvantages.

Caballero, Hoshi, and Kashyap (2008) and Hoshi (2006) identify Japanese companies as zombie whenever they receive subsidized credit, i.e. loans at advantageous interest rates, at rates below those of the most creditworthy companies. Fukuda and Nakamura (2011) add two criteria, profitability and evergreen lending, to avoid type one and two errors, while more recently McGowan, Andrews, and Millot (2018) adopt a measure based on the interest coverage ratio, an accounting measure that captures the persistent lack of profitability in mature companies. Banerjee and Hofmann (2018) add to the latter a measure of market expectations about the company's future profit potential, the Tobin's q. Acharya, Crosignani, Eisert, and Eufinger (2020) and Acharya, Eisert, Eufinger, and Hirsch (2019) use two criteria based on the interest coverage ratio and leverage of the company, plus the subsidized credit received by the non-viable firms.[6]

Our main measure follows Banerjee and Hofmann (2020) who classify a firm as a zombie whenever its ICR_{it} is less than one for at least three consecutive years and its Tobin's q is below median within a sector in a given year. The latter authors often restrict the interest coverage ratio to two years, we instead remain with the three years window as a more reasonable time to develop the *zombie* status, but in accordance we drop the age limit of at least 10 years old as the majority of the companies are mature.[7]

The interest coverage ratio, a measure based on the financial operating characteristics of a company, for firm i, in year t, is computed as $EBIT_{it}/IE_{it}$, where $EBIT_{it}$ is earnings before interest and taxes, and IE_{it} denotes the interest expense, for each firm i at time t.[8] The final measure is a binary variable. As robustness, we also compute two

[6]If we look at the vast body of literature on financial distress, we find that different measurement techniques are provided. From the traditional accounting-based measures of Altman (1968) and Ohlson (1980), to the more recent market value-based measures of Campbell, Hilscher, and Szilagyi (2008). We are interested in measuring zombie firms, which are measured differently to distressed firms. For the latter type of companies we adopt the Altman (1968) Z-score.

[7]To compute the Tobin's q we use the market-to-book ratio, the most common proxy for finance average q (Erickson and Whited 2006), plus an additional Tobin's q measure computed by adding to the book value of total assets the market value of equity and subtracting the book value of common equity divided by the book value of total assets. To capture firms' investment opportunities one would ideally account for intangible capital, especially when examining zombie companies with respect to their healthy peers. The Total q developed by Peters and Taylor (2017) would fit this purpose. The latter measure is however especially trained for US data, and Compustat North America data items, on US companies while not for European companies captured via Compustat Global.

[8]As per Compustat data, we use the variable *xint* as a measure of interest expense which represents the company's gross periodic expense in securing long- and short-term debt. From the established studies, both interest expense and interest payments (interest paid) are used to compute the ICR_{it} (McGowan, Andrews, and Millot 2018; Banerjee and Hofmann 2018; Acharya, Crosignani, Eisert, and Eufinger 2020). We recall that the variable *xint*, as from Compustat Data Guide, includes also items such as: amortization of debt discount or premium, debt issuance expense (such as, underwriting fees, brokerage costs, advertising costs), discount on the sale of receivables of a finance subsidiary, factoring charges, finance charges, interest expense on both long and short-term debt, interest on customer advances, other financial expenses, and retail companies' interest expense.

other measures (McGowan, Andrews, and Millot 2018; Acharya, Crosignani, Eisert, and Eufinger 2020).

2.3 Rationale and Zombie Prevalence

Given the rich data set, we can monitor the ups and downs in the recovery process of zombie companies. In doing this exercise, it becomes clear that in order to fully understand the zombie phenomenon it is crucial to analyze the recovered zombie, their healthy peers, and compare the zombie to other non-viable types of companies. Zombie firms are often treated as financially distressed companies, up to the point where the two are used interchangeably to denote one or the other. We, however, document that zombie firms often differ from distressed companies, and should thus be treated differently.

In this setting, the healthy firms serve as our baseline group and are identified as those companies that are never zombie nor distressed throughout the entire period of observation, while distressed firms are, according to existing definitions, companies close to default (Altman 1968; Ohlson 1980; Gordon 1971). To measure distressed-type of firms we use the Altman Z-score (Altman 1968). The recovered are instead those that leave the zombie status at least once. We capture the recovered by counting the number of zombie spells in our firm sample. This allows us to observe whether zombie firms recover, and how often.

Table 1 reports descriptive statistics on a set of performance measures for our sample of companies in Europe and in the United States. Zombie firms fall behind their healthy peers on a number of characteristics, they show differences and similarities to the distressed, while the recovered zombie prove growth potentials.

	Healthy		**Distressed**		**Zombie**		**Recovered**	
	EU	**US**	**EU**	**US**	**EU**	**US**	**EU**	**US**
Leverage	0.469	0.471	0.633	0.696	0.631	0.742	0.537	0.535
Net Leverage	0.048	0.101	0.267	0.297	0.222	0.286	0.145	0.162
Asset Tangibility	0.281	0.243	0.375	0.270	0.290	0.224	0.310	0.261
Cash ST Investments	0.115	0.076	0.063	0.064	0.064	0.077	0.083	0.075
Operating Profit	0.104	0.136	0.049	-0.027	-0.015	-0.049	0.047	0.081
Capex	0.038	0.045	0.028	0.032	0.015	0.021	0.028	0.039
Ebit ICR	7.417	5.228	1.115	-1.234	-2.613	-2.524	1.064	1.389
Δ Total Assets	0.069	0.086	0.021	-0.030	-0.039	-0.062	0.020	0.040
Size(Log Tot. Assets)	7.262	4.600	7.723	3.515	6.429	3.426	6.773	4.222

Table 1: Healthy, Distressed, Zombie, and Recovered Firms. This table presents descriptive statistics on our sample of companies in Europe and the US. We report median values of leverage, net leverage, asset tangibility, cash and short-term investment, operating profit, Capex ratio, EBIT interest coverage ratio, change in total assets, and size as log of total assets. The healthy are those that are never zombie. Zombie takes the value of 1 if its ICR is below 1 for at least 3 consecutive years and the Tobin's q below median. The distressed have a Z-score below 1.81. The recovered are those that exit the zombie status at least once. Source: Authors' projections on Compustat data.

There are several examples of well-known public companies that have been in dire straits for several years and would have not being able to survive for long without financial support.[9] The existing literature refers to these occurrences as zombie companies, i.e. business entities that are unable to cover their debt servicing costs from their current profits over an extended period of time. What are zombie companies, how can we measure them, and what are the existing channels potentially explaining their existence? In this section, we take a global perspective and explore their trend (Figure 2), the drivers, existing channels, and the regulatory framework.

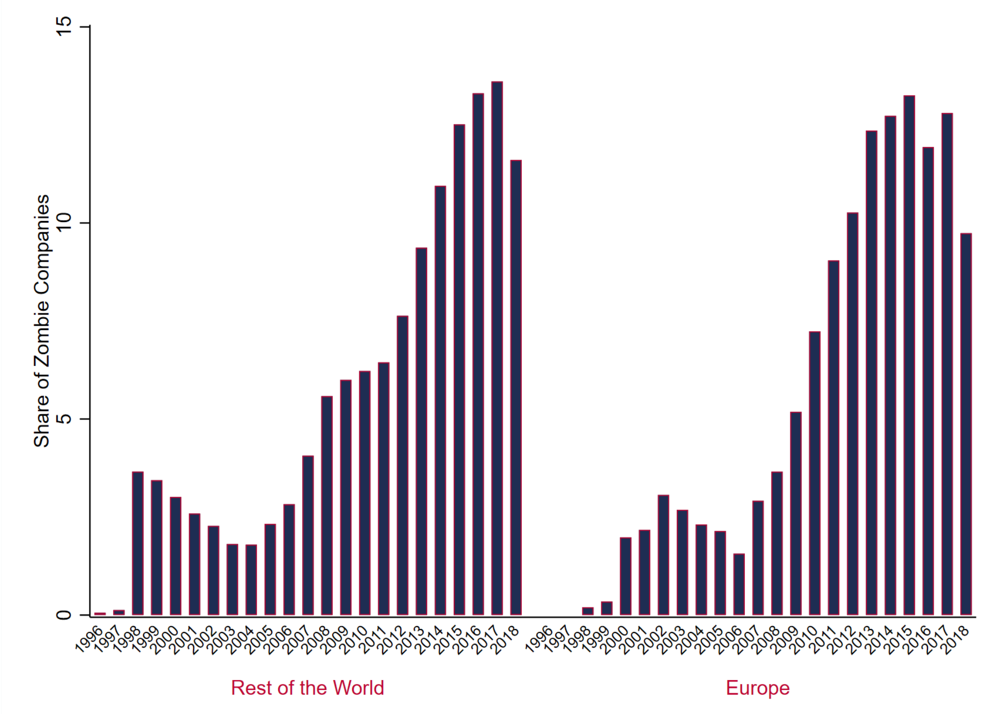


Figure 2: Zombie Trend in Europe and in the Rest of the World. This figure shows the share of zombie companies in Europe to the right and in the rest of the world to the left. The rest of the world includes Asia and Latin America and excludes the United States and Canada that are plotted separately in Figure A1. The plotted time-frame of analysis considers the years from 1996 to 2018. Source: Authors' projections on Compustat Global data.

We can observe a rise in zombie shares over the last 20 years. Figure 2 shows that the share of publicly quoted zombie across Europe and the rest of the world has gone up, from close to zero in the 1990s to roughly 15% in recent years. Interestingly, the phenomenon

[9]A recent example is the case of JCPenny, an American department store chain that raised $400 million in debt through the financial assistance process, extended to thousands of other companies, of the Federal Reserve. This case appeared on February 5 2020 at: `https://www.ft.com/content/1d87c9ec-4762-11ea-aeb3-955839e06441`. Nonetheless, in May 2020 JCPenny filed for bankruptcy protection under Chapter 11. In Europe, Stefanel S.p.A. and Feltrinelli S.p.A., Italian manufacturing companies, were classified zombie-like firms (Acharya, Eisert, Eufinger, and Hirsch 2019).

appeared during the late 1990s early 2000s, historically over the dot-com bubble, both in Europe as in the rest of the world. It however spiked up during the global financial crisis, especially evident in Europe, to then decrease slightly during recent years.

The term *zombie* firms appeared in reference to the US Savings and Loans crisis (S&Ls) of the 1980s and 1990s (Kane 1989) and the Japanese banking crisis of the 1990s. In the latter historical event, Caballero, Hoshi, and Kashyap (2008) document the phenomenon of forbearance lending, a situation in which large banks kept the credit flowing to otherwise insolvent borrowers, also called zombie firms. Recent studies confirm the link between weakly capitalized banks and zombie firms (Acharya, Eisert, Eufinger, and Hirsch 2019; Schivardi, Sette, and Tabellini 2017; Giannetti and Simonov 2013), while others document the increasing trend, above observed, and examine whether other channels, such as the level of interest rates, could be another factor explaining the incidence of zombification (Banerjee and Hofmann 2018).

On the one hand, a common belief shares the idea that Europe might be a repeat of Japan's experience where weak banks were not sufficiently recapitalized and did not foreclose on zombie borrowers to avoid reporting the losses (Andrews and Petroulakis 2017). On the other hand, banks may also lend to zombie firms because of strong relationships. At the same time, in addition to *wrong* bank lending behaviors and excessive levels of corporate debt, today's zombie conundrum is also characterized by an environment in which unconventional monetary policy measures were adopted by central banks in response to the global financial crisis.[10] Two channels are considered by the literature.

The zombie that are flooding most of the economies are often regarded as troubled companies that under normal economic conditions would exit the market and be replaced by new entrants. Let us consider an economy with and without zombie-like firms.

A world without zombie firms consists of businesses that are well-established in an industry and would-be entrants that could eventually enter the market in the future. In the event of a common shock, and in a normal competitive setting, the least performing companies exit the market. In the case of a permanent shock, the economy would adjust to the new equilibrium but a lower number of companies would exist. In an economy populated by zombie, the entry and exit dynamics are instead different as such companies are allowed to remain in the market for longer because of the financial support they receive from creditors, either their banking counterpart or the government (Satu, Vanhala, and Verín 2020). In the latter scenario, there exists a congested market where zombie and healthy firms have to compete under unequal conditions (Hoshi 2006; Caballero, Hoshi, and Kashyap 2008).

[10] In this regard, Acharya, Eisert, Eufinger, and Hirsch (2019) investigate the European Central Bank's announcement of the Outright Monetary Transactions (OMT) program, an unlimited short-term sovereign bond purchases program launched to limit the 2012 European sovereign debt crisis, documents the banks' lending capacity before and after the OMT announcement, and show that post-OMT about 8% of the loans were still granted to zombie firms. It is important to recall that the ECB did not buy any bonds, the pure announcement that it could potentially buy an unlimited amount of bonds was sufficient to please the sovereign bond market. In terms of financing conditions, De Martiis (2020) investigates the effects of organized crime, a well-established business counterpart, on unproductive-type of firms.

Within this theoretical framework, we articulate the potential channels at work in order to unfold the rationale behind the ongoing presence of such firms. The widespread existence of non-viable (i.e. zombie) firms is a credit misallocation issue, where credit is allocated to companies that are not economically viable, thus keeping them afloat for longer. The literature has put forward two potential explanations about occurring channels: *(i)* the banking channel and *(ii)* the monetary policy channel. The first established channel, considers a banking system that is not adequately recapitalized, where banks with equity shortfalls engage in evergreening to avoid loan loss recognition (Caballero, Hoshi, and Kashyap 2008; Acharya, Eisert, Eufinger, and Hirsch 2019; Schivardi, Sette, and Tabellini 2017). The second channel, would instead suggest that indebted firms could become more viable at lower interest rates thus reducing the pressure on zombie firms to exit the market or restructure (Banerjee and Hofmann 2018). With respect to the latter channel, the literature has however not found a clear link between the interest rates and the incidence of zombie firms. In terms of resources misallocation, the existing literature also shows that a decline in the real interest rate increases the dispersion of the return to capital and generates lower productivity growth as capital inflows are directed to unproductive companies (Gopinath, Kalemli-Ozcan, Karabarbounis, and Villegas-Sanchez 2017), that a large share of firms are still alive despite low productivity levels (Calligaris et al. 2016), and that companies receiving government subsidies are less likely to die (Satu, Vanhala, and Verín 2020).

The microeconomic setting should also be considered. To account for the persistence of zombie companies, differing regulatory frameworks are also part of the equation. On the one hand, there is a set of countries with efficient regimes that allow to prevent and solve insolvencies, while on the other hand the majority are progressing slowly (McGowan, Andrews, and Millot 2017). In terms of differing legal systems, Haselmann and Wachtel (2010) show that banks operating in a well-functioning legal environment lend relatively more to small and medium-sized enterprises, while in an unsound legal system they tend to lend more to large enterprises and governments. Accounting for differences in the bankruptcy codes, Favara, Schroth, and Valta (2012) show that the prospect of strategic default on the firm's debt affects the firm's equity beta and this effect decreases in countries where debt contracts cannot be easily renegotiated. Efficient reorganization and liquidation procedures are also crucial in the design of financial contracts and firm investment (Rodano, Serrano-Velarde, and Tarantino 2016). Djankov, McLiesh, and Shleifer (2007) highlight differences between common law and civil law countries in terms of creditor rights and public registries. In addition, civil law countries, like France and Germany, have developed a high level of protection for creditors in the form of controls over the management of debtor firms, while common law countries, like the UK and USA, have reached a high degree of protection in relation to secured creditors' contractual rights over firms' assets (Deakin, Mollica, and Sarkar 2017). How different countries deal with unviable firms varies over time and depends on unequal contract laws, securities laws, criminal laws, the availability of extrajudicial options, the institutional development of a country (i.e. courts, creditors, banks, and government), and the diversity of claims and the degree of information asymmetries (Claessens, Djankov, and Mody 2001).

The existing bankruptcy and restructuring frameworks might however not be fully used to target zombie-like firms, given that the latter are often not comparable to distressed-type of firms, as they are companies with healthy periods in-between unsound years and are likely to recover from the zombie status.

The geographical concentration of zombie firms in the United States, Figure 3, and in Europe, Figure 4, highlights firm-specific differences, but also potential regulatory and governmental divergences (McGowan, Andrews, and Millot 2017). With respect to the latter, we also observe how industry-specific factors play a role across countries (Figures A2).

Examining the US dataset, Figure 3 registers the highest shares of zombie companies, in dark blue, in the US states of Montana, Idaho, Wyoming, Utah, Colorado, Nevada, New Mexico, Texas, Florida, West Virginia, New Jersey, and Rhode Island.

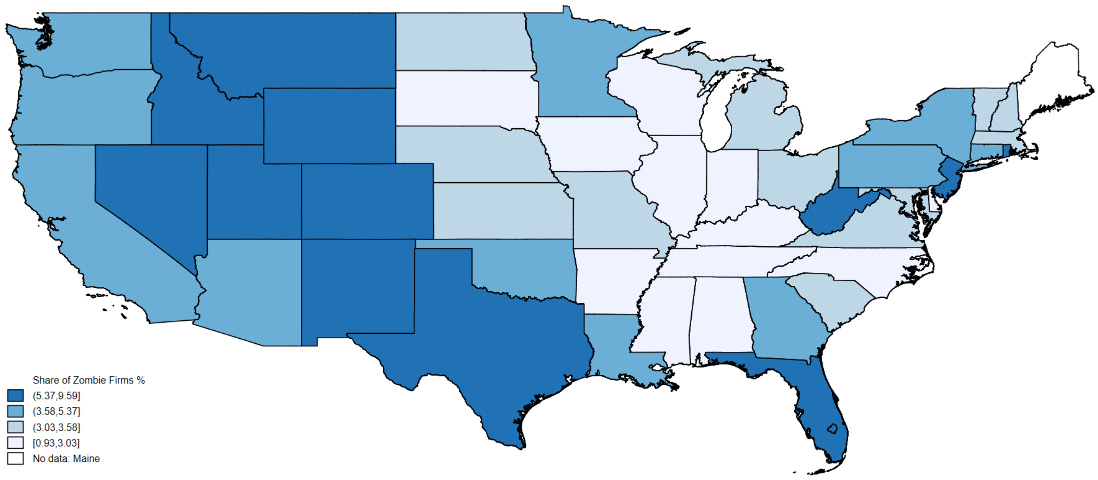


Figure 3: Zombie Shares in the United States. The map shows the presence of zombie companies by state. The map is scaled in different shades of blue according to the severity of the phenomenon. In dark blue are those states with the highest zombie shares. The only state for which we have no data at disposal is the state of Maine, in white color. We exclude from the map Alaska, Hawaii, Puerto Rico, the Virgin Islands, and all minor islands. Source: Authors' projections on Compustat North America data.

From Compustat Global dataset, Figure 4 plots the zombie shares for our sample of European countries. The map documents that the countries with the highest share of zombie firms are Portugal, Greece, Cyprus, Croatia, Macedonia, Slovenia, and Slovakia.

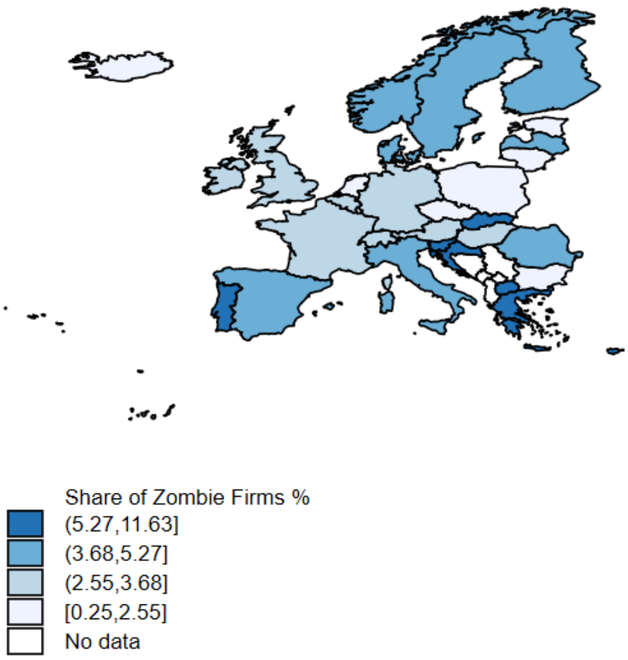


Figure 4: Zombie Shares in Europe. The map shows the presence of zombie companies by country. The map is scaled in different shades of blue according to the severity of the phenomenon. In dark blue are those countries with the highest zombie shares. The countries for which we have no data are Albania, Serbia, Montenegro, Kosovo and Bosnia and Herzegovina, in white color. Source: Authors' projections on Compustat Global data.

3 Firm-Specific Determinants

3.1 Decision Trees

The high dimensions of our dataset renders it cumbersome, if not impossible, to analyze the firm-specific characteristics of financially unviable companies based on classic statistical models. Such an approach would imply to make a priori assumptions and rely on a subset of the available variables. In line with Breiman (2001), we instead use an algorithmic modeling to find the important variables. In consideration of the large amount of input variables at disposal, we consider decision trees an appealing and intuitive approach to identify the most important firm-specific drivers out of a broad range of possible explanatory variables. The algorithm underlying the decision tree searches through the whole range of explanatory variables and subsequently, i.e., at each iteration, finds the variable that can better classify zombie and non-zombie. Moreover, at each iteration the algorithm searches for the best split - input variable combination that reduces the loss function the most. The advantage of a decision tree is its simplicity in combination with outstanding interpretability through elegant visualization. In contrast to classic statistic knowledge-based models, i.e., a Logit model, the tree finds the firm-specific characteristics directly from the data without the need for assumptions. Therefore, decision trees provide a novel perspective on the characteristics of zombie, distressed, recovered zombie, and healthy companies.

The idea of decision trees is to subsequently split the input space X into rectangular segments and provide a decision at each one of those rectangles. Accordingly, in each section, the outcome variable y is modeled with a different constant, e.g., the mean, in regression problems or majority vote in classification problems. The algorithm makes one binary split only for a single input feature at each iteration. After each iteration, the tree repeats the procedure in the new sub-samples.

In order to construct the input space regions, we follow the popular CART algorithm, which finds exclusive, non-overlapping regions $R_1, \ldots, R_j$ with a rectangular shape. Consider a sample of input and output (y, X), where y is a discrete variable with classes K and $X = (x_1, x_2, \ldots, x_p)$ includes the input variables. We require the algorithm to automatically find the best input variable and split point s at each iteration. The proportion of the response variable y for each region R_j, is thus given by:

$$\hat{p}_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = k), \tag{1}$$

where I is the indicator function. A standard loss function of the CART algorithm is the Cross-entropy, given by:

$$L(p) = -\sum_{k=1}^{K} \hat{p}_{mk} log(\hat{p}_{mk}), \tag{2}$$

where p_k is the probability of class k and the impurity reaches its minimum if all observations are classified correctly. However, a direct, contemporaneous computation of the regions by minimizing the loss function is not feasible as the input space can be split in infinite combinations of sub-rectangles. Therefore, we start with a top-down approach of binary splitting. Assume a first splitting variable l and a splitting point s, we choose the first pair of regions as:

$$R_1(l,s) = \{X|X_l \leq s\} \text{ and } R_2(l,s) = \{X|X_l > s\}. \tag{3}$$

Ultimately, we find the splitting variable l and split point s by solving $arg\max_k \hat{p}_{lk}$. After partitioning the input space in two regions, based on the best splitting variable and split point, the process is repeated within each region. High interpretability provides a prime tool to determine important factors characterizing zombie firms.[11]

In a second step, we augment our algorithm and provide multivariate decision trees to delve deeper into the factors that are more conducive to a company being a zombie, distressed, healthy, or recovered. In performing this exercise, we account for the fact that zombie companies are mostly growing companies with unsound periods.

We develop a set of decision trees for both geographical areas, the United States and Europe, and we account for time-varying differences to understand whether firm-specific drivers change in response to economic downturns. To do this, we estimate the trees prior to the global financial crisis and afterwards, i.e. in 2007 and 2016.

The decision trees provide the variable name and split point, and the % of observations used by the algorithm at each node. The entropy provides a measure of the node purity and the values show the % of non-zombie (left) and zombie (right) after the split. Accordingly, nodes with a deeper color are more pure and show how well the explanatory variable separates the possible categories. I.e., values of $[0.6, 0.4]$ represent a sample with 60% of non-zombie and 40% of zombie after the split. Nodes with a blue color indicate a majority of zombie and vice versa in orange. A white node shows an indecisive split.

3.2 Empirical Evidence: Zombie Firms

3.2.1 Europe

We analyze the firm-specific characteristics of zombie versus non-zombie publicly listed companies for a sample of 32 European countries.[12] To account for time-varying differences, Figure 5 shows the decision tree results before the global financial crisis in 2007, while Figure 6 observes what characterizes a zombie company in 2016, a non-crisis period.

At first, we document similarities between years before and after the global financial crisis, underlining that regardless of economic downturns specific drivers persist. Among the European countries two aspects stand out: *(i)* income-related variables are the main firm-specific characteristics, followed by *(ii)* operating expenses, liabilities, and stock

[11] For an in-depth review on the topic we refer to Hastie, Tibshirani, and Friedman (2009), or to the preliminary book by Breiman, Friedman, Stone, and Olshen (1984).

[12] Our European sample, from Compustat Global, includes: Belgium, Bulgaria, Switzerland, Cyprus, Czech Republic, Germany, Denmark, Spain, Estonia, Finland, France, Great Britain, Greece, Croatia, Hungary, Ireland, Italy, Lithuania, Luxembourg, Latvia, Macedonia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Sweden, Austria.

related variables. In terms of income-related variables, pretax income (*pi*) allows for the most crucial binary split to classify zombie companies in Europe, indicating that for low values of operating income the tree predicts that the company is likely a zombie, vice versa for higher values thus leading to orange-colored nodes. Among other income-specific variables *nopi*, is also recurrent in the higher splits. The decision tree algorithm allows us, contrary to classic statistical models, to detect and show which parts of a firm's income likely determine the zombie status. In terms of debt-related variables, long-term debt due in one year, *dd1*, is instead the most decisive driver. In this regard, we recall that in order to avoid any potential simultaneity in our results we exclude from the algorithm the variables related to the zombie definition.

In addition, the algorithm highlights how companies with low operating income values, high total liabilities, and high levels of common stock are classified as zombie, likewise before and after crisis years. We thus confirm that the zombie phenomenon in Europe is well described by the presence of overly indebted firms, where a combination of low income, low returns, and a high debt ratio is indeed a typical feature. In this regard, Hoshi (2006) documents that Japanese zombie companies are smaller firms, less profitable, more indebted, more likely to be found in non-manufacturing industries, and often located outside large metropolitan areas. With respect to geographic-specific factors, we refer to Section 3.5 where a fixed-effects logistic regression model that incorporates country fixed effects is presented. In our case, no specific industries are returned by the algorithm to categorize zombie firms in Europe, and interestingly firm size is not selected among the relevant primary characteristics.

Another interesting result suggests that before the crisis total assets *at* did not play a major role in determining a zombie, while in after the crisis the importance rises. The latter result is mostly explained by the crowding out of healthy firms' investments during crisis periods (Banerjee and Hofmann 2018).

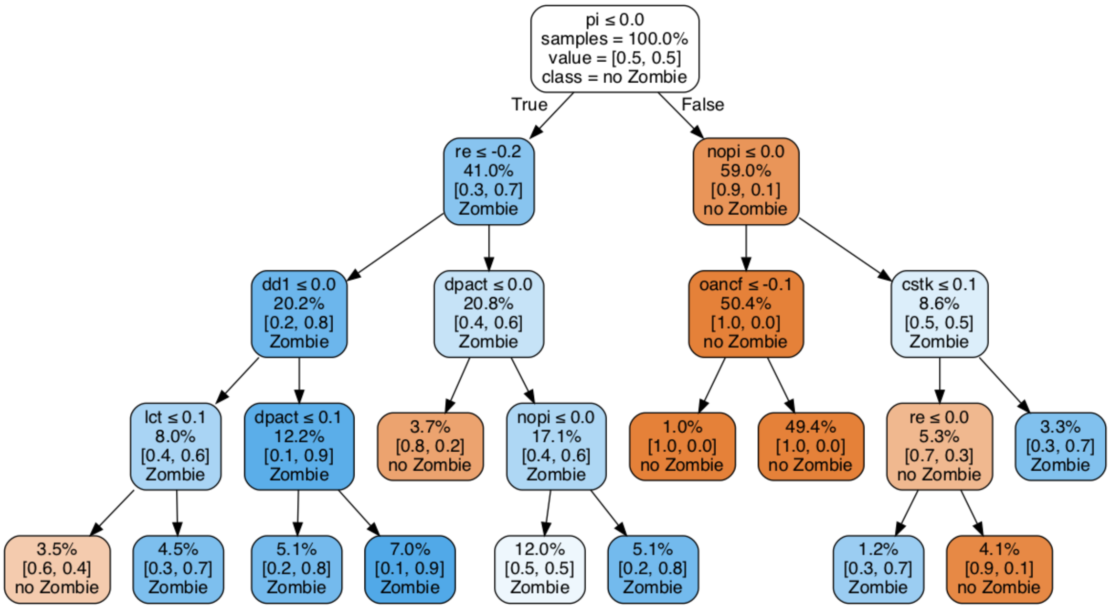


Figure 5: Zombie versus Non-Zombie, Europe 2007. This figure shows the decision tree for Europe 2007. Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is provided at the top of each node. The purity of the nodes is given by a higher entropy and a darker color. We measure zombie following Banerjee and Hofmann (2020).

Legend: *pi* Pretax Income, *re* Retained Expenses, *nopi* Non-Operating Income, *dd*1 Long-Term Debt due in 1 Year, *dpact* Depreciation, Depletion and Amortization, *oancf* Operating Activities Net Cash Flow, *cstk* Common Stocks, *lct* Current liabilities total.

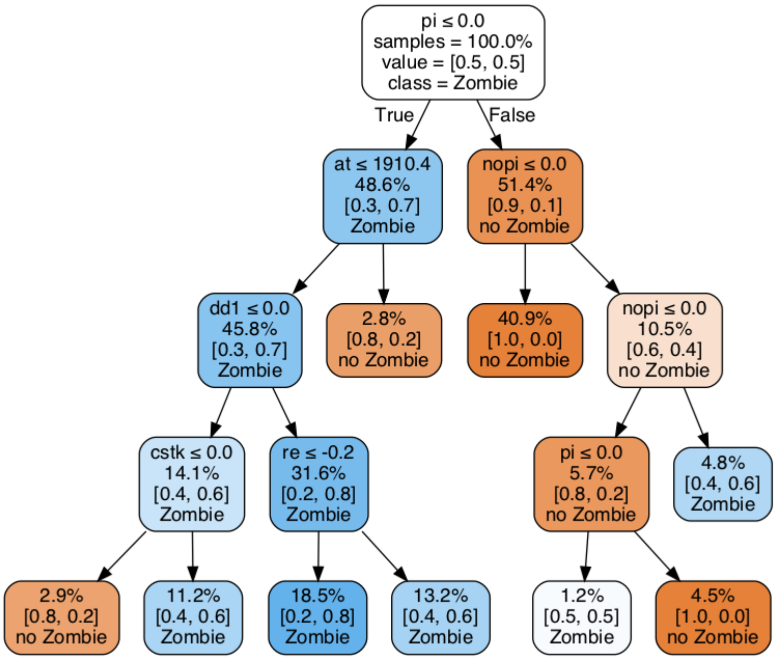


Figure 6: Zombie versus Non-Zombie, Europe 2016. This figure shows the decision tree for Europe 2016. Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is provided at the top of each node. The purity of the nodes is given by a higher entropy and a darker color. We measure zombie following Banerjee and Hofmann (2020).

Legend: *pi* Pretax Income, *at* Total Assets, *nopi* Non-Operating Income, *dd*1 Long-Term Debt due in 1 Year, *cstk* Common Stock, *re* Retained Earnings.

3.2.2 United States

We implement classification trees on US firm-level data to understand the main characteristics of zombie companies in the United States, a less investigated country sample by existing studies. Figure 7 and 8 report the results. First, we find no major differences between the two time frames observed, crisis and healthy periods respectively. Second, the trees predict that income-related variables are the main drivers classifying zombie firms and separating them from the non-zombie. Third, the root node returns the variable pretax income, *pi*, as the most important feature during crisis periods and total assets, *at*, during healthy years. In addition, specific industries are not influential in separating the class of zombie firms from the non-zombie. In both time frames, variables related to the firm operating activities seem to be more predictive. Specifically, operating activities net cash flow, *oancf*, relates to the firm's day-to-day activities of producing and selling and it is composed of operating cash flow, capital spending, and change in net working capital. Put it differently, such financial information tells us whether the firm's business operations are sufficient to cover its everyday cash flows. Debt-related variables, like debt in current liabilities, *dlc*, also display a higher entropy, meaning that the algorithm is decisive in its prediction.

During crisis times (Figure 7), the algorithm indicates that for higher values of earnings before taxes and operating activities a company is likely a non-zombie, while for negative values of pretax income and total assets values below $75.8 million it is likely a zombie.[13] This validates the importance of cash flow, in terms of inflows and outflows of money in and out of the business, and the total assets in terms of firm's liquidity as the more liquid a business is, the less likely it is to experience difficulties repaying its debts. Also, if a substantial portion of the firm's assets is funded by debt, the company has more liabilities than assets and is more likely to incur financial difficulties. The algorithm further predicts that with positive values of pretax income, of total income taxes, *txt*, and depreciation and amortization, *dp*, a company is likely a non-zombie.

Contrary to standard statistical models, the trees show which parts of a firm's income are likely to predict the zombie status. In this regard, during crisis periods total non-operating income, *nopi*, and income before extraordinary items, *ib*, are relevant income-related features separating zombie versus non-zombie.

During healthy times (Figure 8), total assets represents the most important binary split to separate the zombie from the non-zombie and the algorithm suggests that for higher values of total assets, above $387.2 million, and of operating activities net cash flow the firm is likely a non-zombie. Vice-versa, for smaller values of total assets, we follow the tree to the left, and with negative pretax income the company is likely a zombie. In normal economic times, we also document that specific parts of a firm's investment and operating activities are likely to categorize zombie versus non-zombie.

In particular, capital expenditures, *capx*, and accounts receivable, *recch*, are part of the investment activities and operating activities of a firm's cash flow, respectively. For values of capital expenditures below zero a company is likely a zombie, in accordance with (Asquith, Gertner, and Scharfstein 1994), thus confirming the fact that capital expenditures drop when firms are in financial difficulties. The latter result might reflect the firm's response to a lack of good investment opportunities.

[13] As from Compustat Guide, total assets *(at)* is the sum of other assets, total current assets, total property, plant, and equipment, intangible assets, and investments and advances.

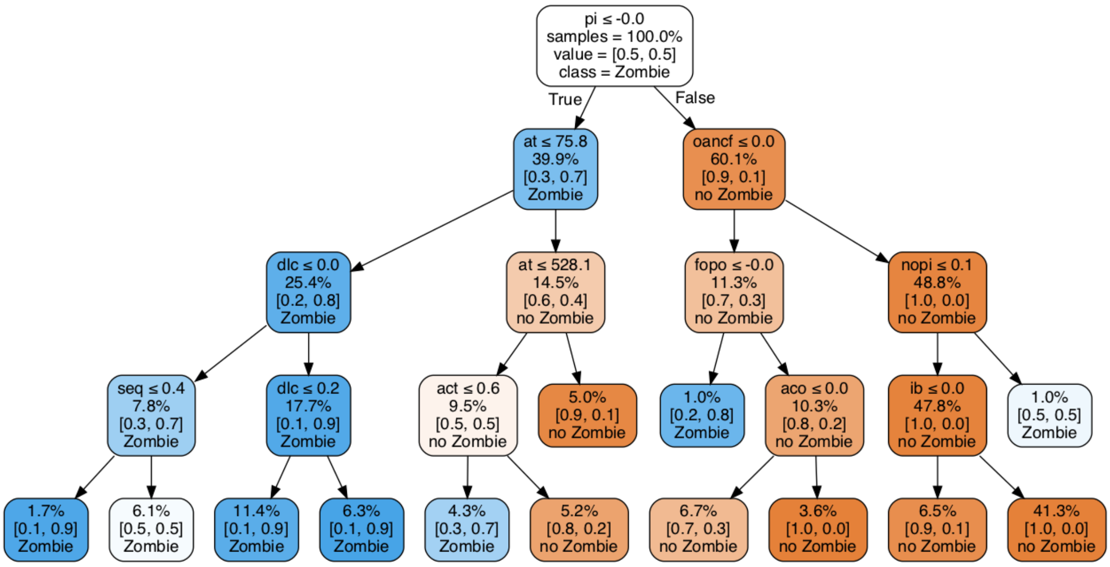


Figure 7: Zombie versus Non-Zombie, United States 2007. This figure shows the decision tree for the US in 2007. Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is provided at the top of each node. The purity of the nodes is given by a higher entropy and a darker color. We measure zombie firms following (Banerjee and Hofmann 2020).

Legend: *pi* Pretax Income, *at* Total Assets, *oancf* Operating Activities Net Cash Flow, *nopi* Non-Operating Income, *fopo* Funds from Other Operations, *aco* Other Current Assets, *ib* Income Before Extraordinary Items, *act* Total Current Assets, *dlc* Debt in Current Liabilities, *seq* Shareholder's Equity.

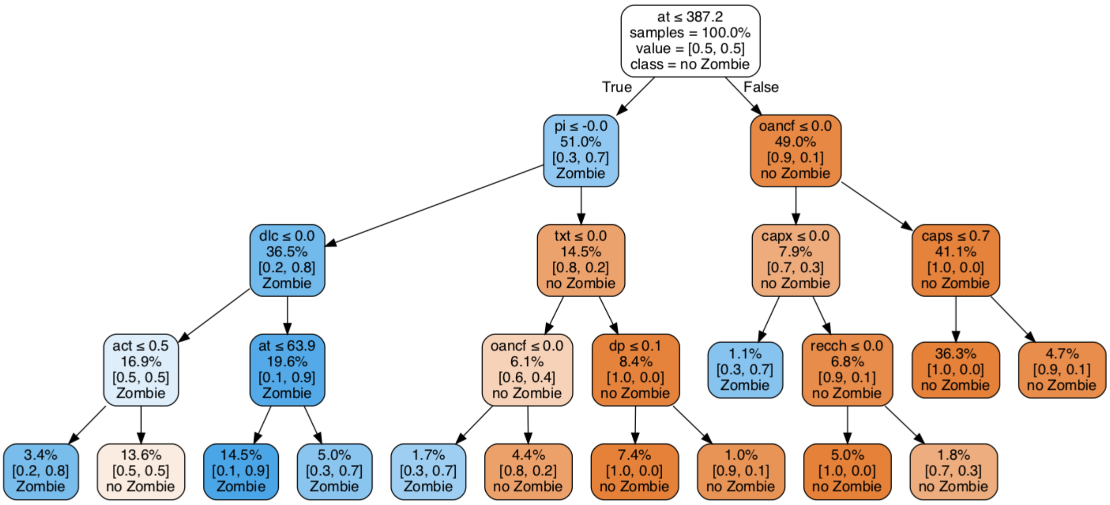


Figure 8: Zombie versus Non-Zombie, United States 2016. This figure shows the decision tree for the US in 2016. Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is provided at the top of each node. The purity of the nodes is given by a higher entropy and a darker color. We measure zombie firms following (Banerjee and Hofmann 2020).

Legend: *at* Total Assets, *pi* Pretax Income, *oancf* Operating Activities Net Cash Flow, *caps* Capital Surplus/Share Premium, *capx* Capital Expenditures, *recch* Accounts Receivable, *dp* Amortization and Depreciation, *txt* Total Income Taxes, *dlc* Debt in Current Liabilities, *act* Total Current Assets.

3.3 Empirical Evidence: Distressed Firms

3.3.1 Europe

Figure 9 provides the results of the characteristics of distressed versus non-distressed publicly listed companies in 32 European countries before the global financial crisis, while Figure 10 analyzes a financially healthy time period, 2016. The main findings are captured by two aspects: *(i)* shareholder's equity and *(ii)* income. Both types of variables show similarities between the two years. Shareholder's equity, *seq*, is the key variable that is returned by both decision trees in their root node, the node in white color, which represents the most important split to classify a company as distressed or non-distressed. Therefore, companies with lower values of shareholder's equity are more likely to be distressed. With respect to income, we find multiple variables that are useful in the classification of a distressed company. Moreover, companies with lower values of income and sales are more likely distressed. Especially pretax income, *pi*, income before extraordinary items, *ib*, and sales/turnover, *sale*, are predictive. Interestingly, the exact opposite of income, expenses and investments also matters. Hence, we find accrued expenses, *xacc*, and investments and advances, *ivao*, to define distressed companies likewise. In fact, this is in contrast to the tree structure of zombie companies.

Leverage-related variables are likewise selected to classify financially distressed firms and zombie firms. In the latter case, the liabilities total, *lt*, is another very decisive and recurring explanatory variable returned by the classification tree algorithm to capture zombie companies in Europe before and after the global financial crisis. Therefore, both distressed and zombie have a debt-level component in their financial structure that makes them similar, but at the same time income-specific items are especially categorizing zombie versus non-zombie. From the previous section, we recall that other income-related variables such as pretax income, *pi*, retained earnings, *re*, and income before extraordinary items, *ib*, are often recurring in the higher splits.

These findings show that both zombie and distressed firms have accrued debts weighting down on their operating activities. At the same time, given their level of operating income, we observe from the data processing that zombie companies are likely to recover rather than dying. Distressed companies appear instead at a different stage of their unviability, as also suggested from Table 1, making them more likely to default or enter bankruptcy in order to protect their assets from creditors.

Among other explanatory variables that are relevant to classify distressed versus non-distressed companies, common stock, *cstk*, and stock price close return, *prcc_ return*, are often identified after crisis years, Figure 9, as well as during healthy periods, Figure 10, followed by the level of cash, *ch*, especially in 2016.

On the one hand, the corporate finance literature documents that the return on assets is an important financial ratio predicting corporate distress, as it captures an ongoing underperformance due to operating decisions or external forces, and shows that some of the most salient characteristics of distressed companies are low market value, high leverage, cash flow problems, and prices sensitive to negative conditions (Chan and Chen 1991). On the other hand, there is no evidence on the specific differences or similarities between zombie and distressed. Our findings thus indicate that there are some specific parts of the income of a company, such as pretax income, non-operating income, and income before extraordinary items, that distinguish zombie from non-zombie firms and that can be used as a diagnostic tool to better categorize zombie companies in Europe.

Specific parts of the debt of a company, such as liabilities total and book leverage, can instead classify both the distress stage and the zombie stage. In this regard, a set of descriptive statistics (Table 1) further documents that distressed European firms have higher leverage, net leverage, asset tangibility, and operating profit than zombie firms.

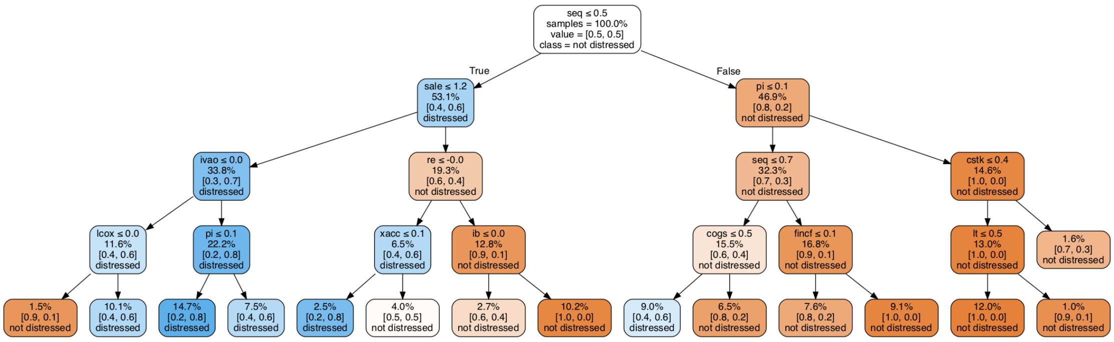


Figure 9: Distressed versus Non-Distressed, Europe 2007. This figure shows the decision tree for Europe 2007. Distressed are measured with the Z-score (Altman 1968). Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is at the top of each node. The nodes purity is given by higher entropy and darker color.

Legend:*pi* Pretax Income, *ivao* Other Investments and Advances, *lcox* Other Current Liabilities, *xacc* Accrued Expenses, *ib* Income before Extraordinary Items, *cogs* Cost of Goods sold, *cstk* Common Stocks, *lt* Total Liabilities

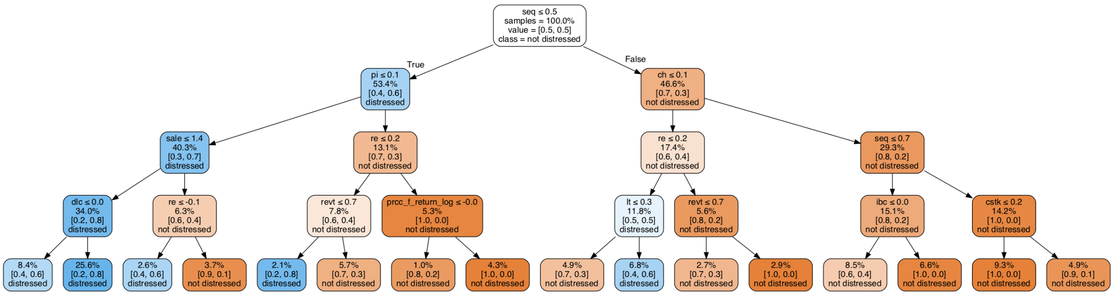


Figure 10: Distressed versus Non-Distressed, Europe 2016. This figure shows the decision tree for Europe 2016. Distressed are measured with the Z-score (Altman 1968). Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is at the top of each node. The nodes purity is given by higher entropy and darker color.

Legend: *seq* Shareholders Equity, *pi* Pretax Income, *ch* Cash, *sale* Sale/Turnover(net), *re* Retained Expenses, *dlc* Debt in Current Liabilities, *revt* Total Revenue, $prcc_return$ Stock Price Return, *lt* Total Liabilities, *ibc* Income before Extraordinary Items, *cstk* Common Stock.

3.3.2 United States

Figure 11 shows the results of the characteristics of distressed versus non-distressed companies in the United States during the crisis years, while Figure 12 documents the main drivers during a period of time that we consider financially healthy. The main features relate to the shareholders' interest in the company. First, we find that in both time periods retained earnings, *re*, is the most important binary split separating distressed versus non-distressed companies. Second, during crisis periods debt and income-related variables are relevant predictors. Third, during healthy times shareholder's equity separates the distressed from the non-distressed. Fourth, similarly to the zombie binary trees, industries are not relevant in classifying distressed versus non-distressed firms.

During crisis times (Figure 11), the algorithm indicates that for positive values of retained earnings, *re*, and of income before extraordinary items, *ib*, a company is likely non-distressed, while for negative values of retained earnings and total liabilities, *lt*, it is likely in the distress status. This confirms the importance of income and debt-related characteristics during economic downturns as firm-specific factors that categorize distressed-types of firms. Of the nodes with a darker color, the most decisive, the algorithm predicts that higher total liabilities, *lt*, and negative return values, *prcc_f*, a company is further categorized as distressed.[14]

The tree for the healthy years (Figure 12), presents different peculiarities. More precisely, retained earnings represents the most important binary split to separate the distressed from the non-distressed, followed by shareholder's equity, *seq*, as second layer of separation. The algorithm predicts that for positive values of retained earnings, above $-1.1 million of the split point, and of shareholder's equity the firm is likely non-distressed. Vice-versa, for negative values of retained earnings, we follow the tree to the left, and with shareholder's equity below the value of 0.7 the company is classified as distressed.

In normal economic times, we find that capital surplus/share premium, *caps*, and the return, *prcc_f*, are also highly predictive of the distress status of US firms.

The results further document that variables related to shareholder's equity highlight the main differences between distressed and zombie, where the latter are characterized by structural and performance based factors, while the second by shareholders interests. In this regard, Kahle and Stulz (2017) document how US public corporations changed over the last forty years, they have been paying out a higher share of net income to shareholders in recent years than in the past, and they differ largely with institutions holding the largest shares.

[14] The return is calculated using Compustat variable stock price close fiscal, *prcc_f*, approximated with the logarithm $log(\frac{T}{t-1})$.

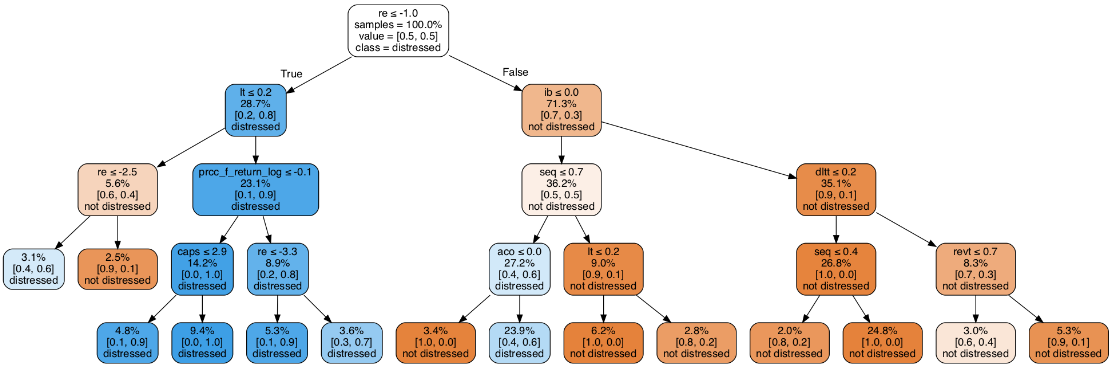


Figure 11: Distressed versus Non-Distressed, United States 2007. This figure shows the decision tree for the US in 2007. Distressed are measured with the Z-score (Altman 1968), while zombie firms following (Banerjee and Hofmann 2020). Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is at the top of each node. The nodes purity is given by higher entropy and darker color.

Legend: *re* Retained Earnings, *ib* Income Before Extraordinary Items, *seq* Shareholder's Equity, *dltt* Total Long-term Debt, *revt* Total Revenue, *lt* Liabilities Total, *prcc_f_return_log* Stock price close fiscal approximated with the logarithm, *caps* Capital Surplus/Share Premium, *aco* Other Current Assets.

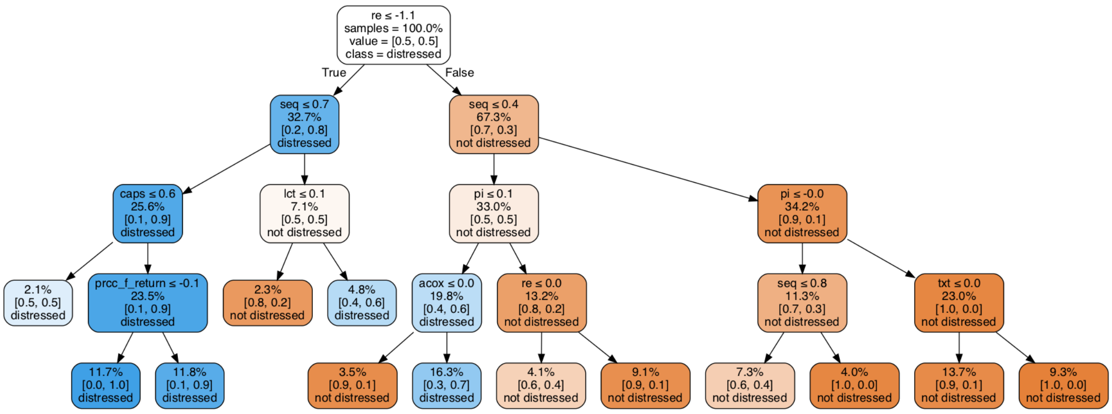


Figure 12: Distressed versus Non-Distressed, United States 2016. This figure shows the decision tree for the US in 2016. Distressed are measured with the Z-score (Altman 1968), while zombie firms following (Banerjee and Hofmann 2020). Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is at the top of each node. The nodes purity is given by higher entropy and darker color.

Legend: *re* Retained Earnings, *seq* Shareholder's Equity, *pi* Pretax Income, *txt* Total Income Taxes, *acox* Other Current Assets excluding prepaid expenses, *prcc_f_return_log* Stock price close fiscal approximated with the logarithm, *caps* Capital Surplus/Share Premium, *lct* Total Current Liabilities.

3.4 Multi-Class Analysis: Zombie, Distressed, Recovered, and Healthy Firms

The multi-class trees follow the same structure of the binary. The importance of the nodes and items they inherit are the same, and at each iteration the algorithm finds a variable that separates one category from the others. The impurity and the item value, however, inherit four values. From left to right, the values describe the proportion of healthy, distressed, zombie, and recovered companies within the node and are crucial to clarify how well the explaining variable separates the firms' categories. The categories of healthy, distressed, zombie, and recovered are represented in orange, green, blue, and purple color, respectively. A dark blue colored node contains mostly zombie. In contrast to the binary trees, the multi-class has a horizontal structure where the if-else reasoning changes. We follow the tree upwards if the statement is *True* and downwards otherwise.[15]

The interpretation is as follows: the predicted category has, after the initial split, the most significant proportion in the new sub-input space, therefore if the node predicts a healthy company the algorithm finds the most significant decrease in the loss function by separating the *good* companies from the others.[16]

3.4.1 Europe

In Figure 13 we document the pre-crisis results, while in Figure 14 the post-crisis.

Of the main findings, similarly to the binary trees, in both time periods the root node returns pretax income, *pi*, underlining the importance of income-related characteristics for the first split. An outcome that yields relevant information. Smaller values of pretax income lead to zombie firms, blue node, while larger values to healthy companies, orange node. During the pre-crisis, a firm with low pretax income, low retained earnings, *re*, and long-term debt due in one year, *dd1* is likely a zombie. Vice-versa with positive values of pretax income, low income before extraordinary items, *ibc*, and low non-operating income, *nopi*, the company is in the distress status. With higher values of income operating activities, higher non-operating income, and sale/turnover, *sale*, the firm is healthy. We however underline that most of the higher splits are inconclusive as the color of the nodes is light, indicating the indecisiveness of the algorithm in separating the categories. Nevertheless, lower splits display a darker color and are decisive in categorizing healthy, distressed, and zombie firms; the recovered are instead the least represented. Pre-crisis, zombie and healthy firms are prevalent.

In the post-crisis, the pattern is similar but the algorithm is irresolute in categorizing especially the healthy, while quite decisive for the zombie class. Negative values of pretax income and low total assets predict zombie and distressed firms, however total liabilities, *lt*, other current assets, *acox*, property, plant, and equipment, *ppent*, and inventories of finished goods, *invfg*, are predictive of zombie firms. The latter are prevalent also in the post-crisis period.

[15] The fact that the root node identifies as first class the distressed or zombie is a random allocation.

[16] Multi-output decision trees provide an argumentative separation of the categories by majority votes.

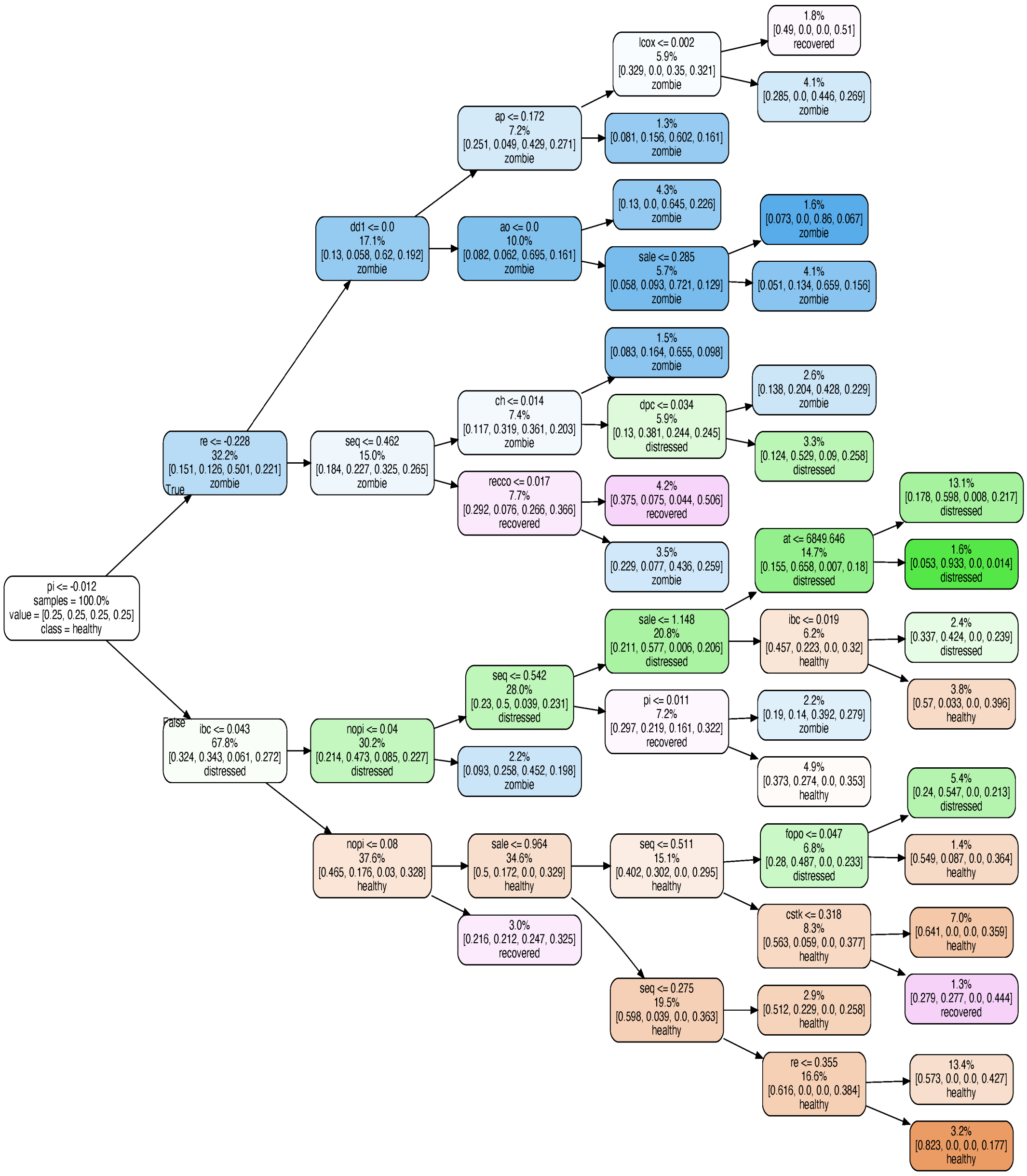


Figure 13: Zombie, Distressed, Recovered, and Healthy, Europe 2007. This figure shows the multi-classification tree for Europe 2007. Distressed are measured with the Z-score (Altman 1968), while the zombie follow (Banerjee and Hofmann 2020). Healthy are those that were never zombie nor distressed. Recovered are those that exited the zombie status. Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is provided at the top of each node. The nodes purity is given by a higher entropy and a darker color.

Legend: *lt* Total Liabilities, *txt* Total Income Taxes, *pi* Pretax Income, *re* Retained Expenses, *dd*1 Long-Term Debt due in one Year, *ibc* Income before Extraordinary Items, *nopi* Non-Operating Income, *seq* Shareholder's Equity, *ap* Accounts Payable , *sale* Sales/Turnover(net), *at* Total Assets, *cstk* Common Stocks, *invrm* Raw Materials/Inventory, *lcox* Other current Liabilities, *rect* Total Accounts Receivable, *revt* Total Revenue, *fopo* Other Funds from Operations, *recco* Accounts Receivables, *dpc* Amortization and Depreciation, *ch* Cash and Due from Banks, *ao* Other Assets.

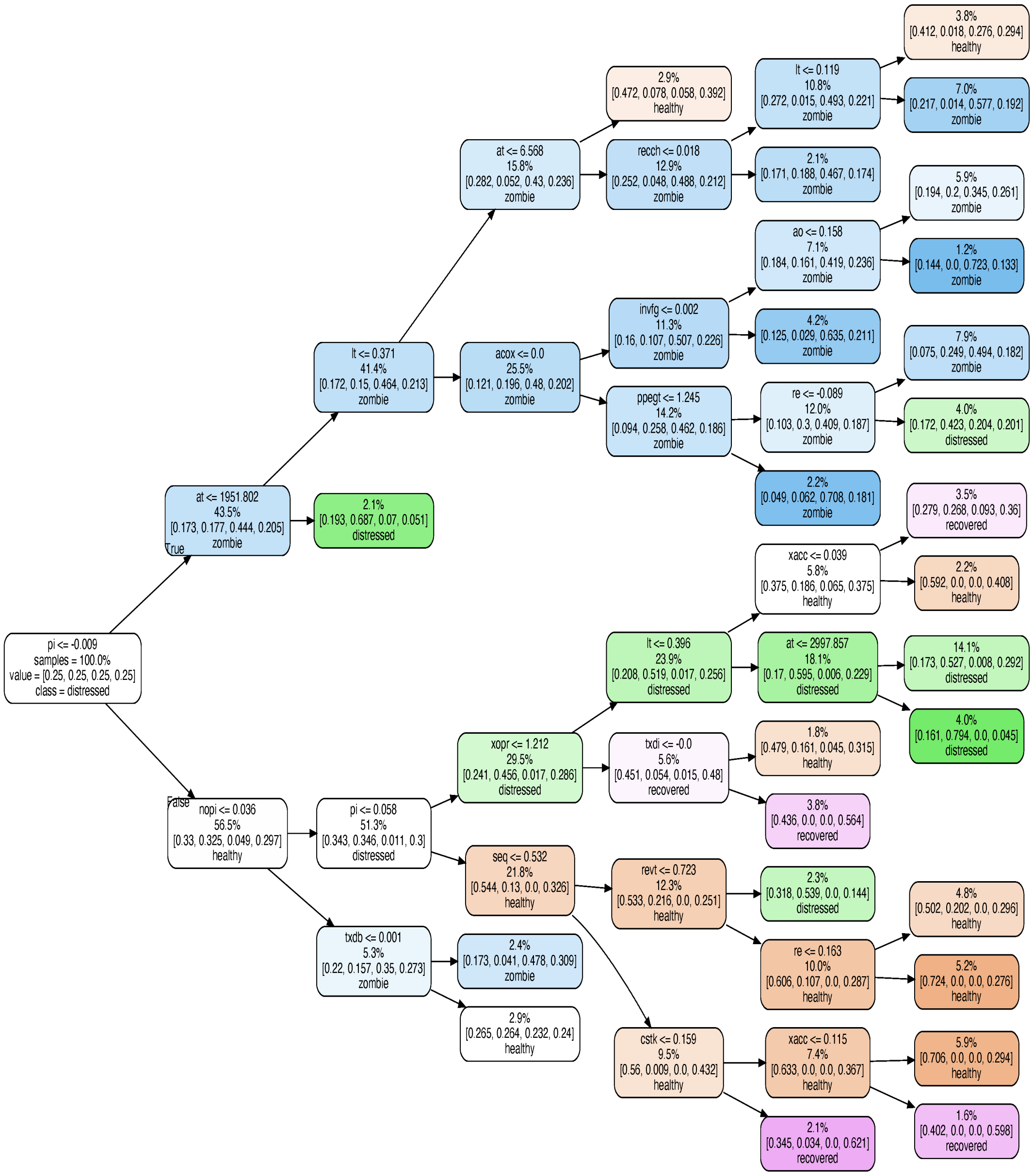


Figure 14: Zombie, Distressed, Recovered, and Healthy, Europe 2016. This figure shows the multi-classification tree for Europe, 2016. The main measure is used to identify the zombie. Distressed are measured with the Z-score (Altman 1968). Healthy are those with an interest coverage ratio above 1. Recovered are those that exited the zombie status. Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is provided at the top of each node. The nodes purity is given by a higher entropy and by a darker color.

Legend: *pi* Pretax Income, *nopi* Non-Operating Income, *at* Total Assets, *txdb* Deferred Taxes Balance Sheet, *lt* Total Liabilities, *recch* Decrease (increase) Accounts Receivable/Debtors, *acox* Other Current Assets, *xopr* Operating Expense, *seq* Shareholder Equity, *invfg* Finished Goods Inventories/Stocks, *ppegt* Total Gross Property Plant and Equipment, *txdi* Deferred Taxes, Tax Credit, *revt* Total Revenue, *cstk* Common Stocks, *ao* Other Assets, *xacc* Accrued Expenses.

3.4.2 United States

Figure 15 documents the results of the multi-class tree pre-crisis, while Figure 16 shows the main characteristics of US companies post-crisis.

The pre-crisis multi-class tree (Figure 15), provides us in a snapshot a prediction of the most important firm-specific features characterizing and separating zombie, distressed, recovered zombie, and healthy corporations in the United States. Of the main findings, we document that total assets, *at*, is the most important variable returned by the algorithm in the root node. Next, income before extraordinary items, *ib*, represents the second layer of separation between healthy and zombie firms, followed by debt-related features from the balance sheet of the company, total liabilities, *lt*, and debt in current liabilities, *dlc*. In particular, the algorithm predicts that with total assets values below $105 million and negative income before extraordinary items the company is likely a zombie, but with debt in current liabilities below or equal to zero the zombie firm is likely to recover. Vice-versa, with total assets values above the split point threshold, positive values of income before extraordinary items, and low values of total liabilities a company is categorized as healthy. Net income also separates healthy form distressed-type companies, followed by assets-related features, other current assets, *aco*, and total assets. If a company has positive net income, but very low values of assets and retained earnings it is likely a distressed. At the same time, income and debt variables distinguish zombie from recovered zombie, followed by operating activities net cash flow, *oancf*, and sale/turnover, *sale*, characteristics.

The post-crisis multi-class tree (Figure 16), shows a similar pattern. More precisely, the algorithm returns the variable total assets as the most important split point, the white node, and income-related variables, pretax income, *pi*, and income before extraordinary items, *ib*, as second most important separation variables between zombie and healthy firms, followed again by debt-related features, total liabilities and debt in current liabilities. The tree further highlights a prevalence of healthy companies during post-crisis periods, followed by zombie, while to a lesser extent distressed-type of firms. In addition, we document that income and debt-related variables, pretax income, income taxes, and debt in current liabilities, separate and categorize zombie versus recovered zombie firms, while income and shareholder's equity separate the healthy from the distressed.

The multi-classification trees allow us to not only put the zombie companies under scrutiny, but to compare them with other classes of companies so to underline relevant patterns in terms of firm-specific differences and similarities. In this respect, the main findings seem to suggest that zombie companies are at a different stage of their financial unviability in comparison to distressed firms that show the typical characteristics of firms close to default.

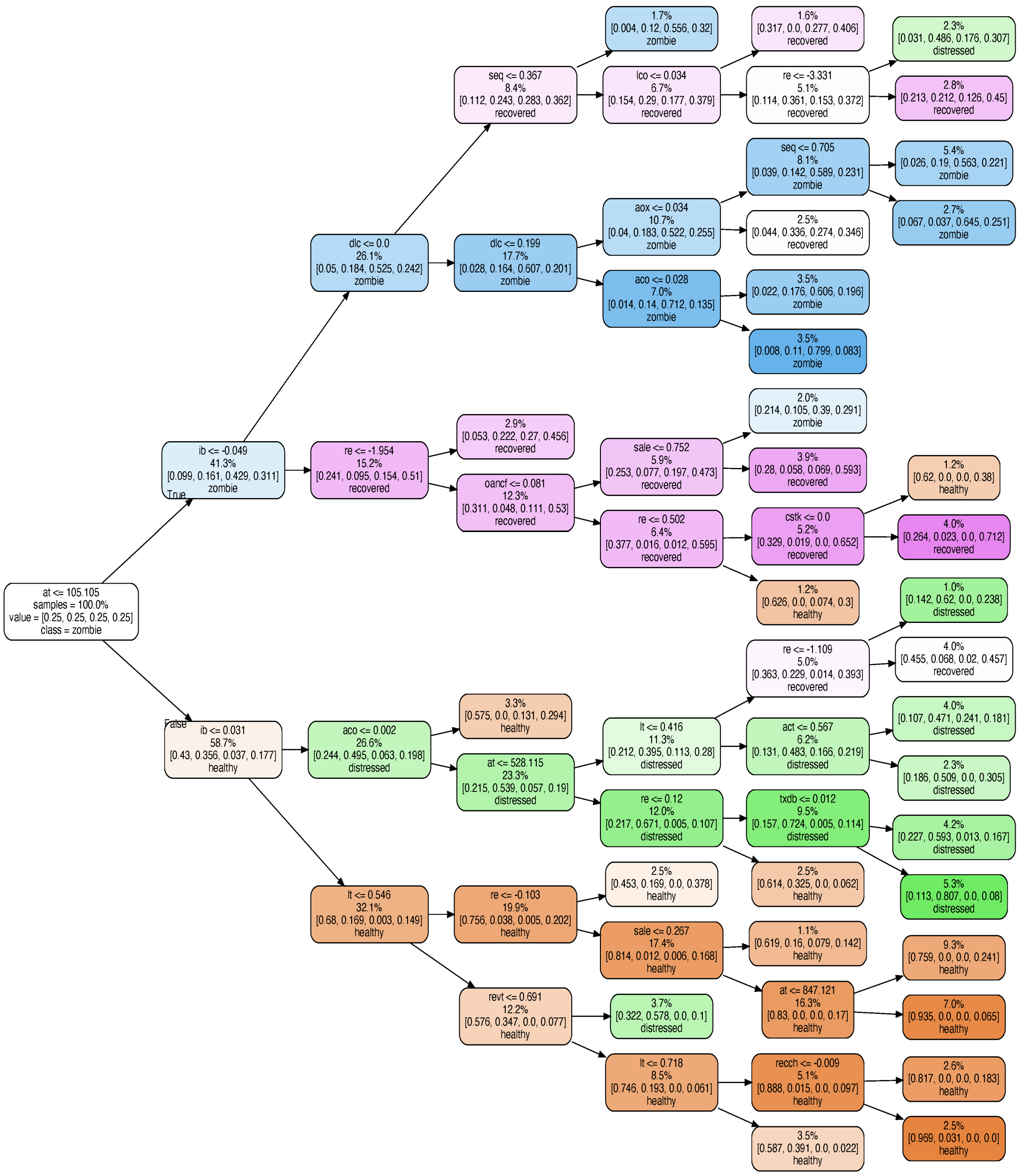


Figure 15: Zombie, Distressed, Recovered, and Healthy, United States 2007. This figure shows the multi-class decision tree for the US 2007. Distressed are measured with the Z-score (Altman 1968), while the zombie firms following (Banerjee and Hofmann 2020). Healthy are those that were never zombie nor distressed. Recovered are those that exited the zombie status. Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is provided at the top of each node. The nodes purity is given by a higher entropy and a darker color.

Legend: *at* Total Assets, *ib* Income before Extraordinary Items, *lt* Total Liabilities, *dlc* Debt in Current Liabilities, *revt* Total Revenue, *recch* Accounts Receivable, *sale* Sale/Turnover, *re* Retained Earnings, *aco* Other Current Assets, *txdb* Deferred Taxes, *cstk* Common Stock, *oancf* Operating Activities Net Cash Flow, *aox* Other Assets excluding Deferred Changes, *seq* Shareholder's Equity, *lco* Other Current Liabilities.

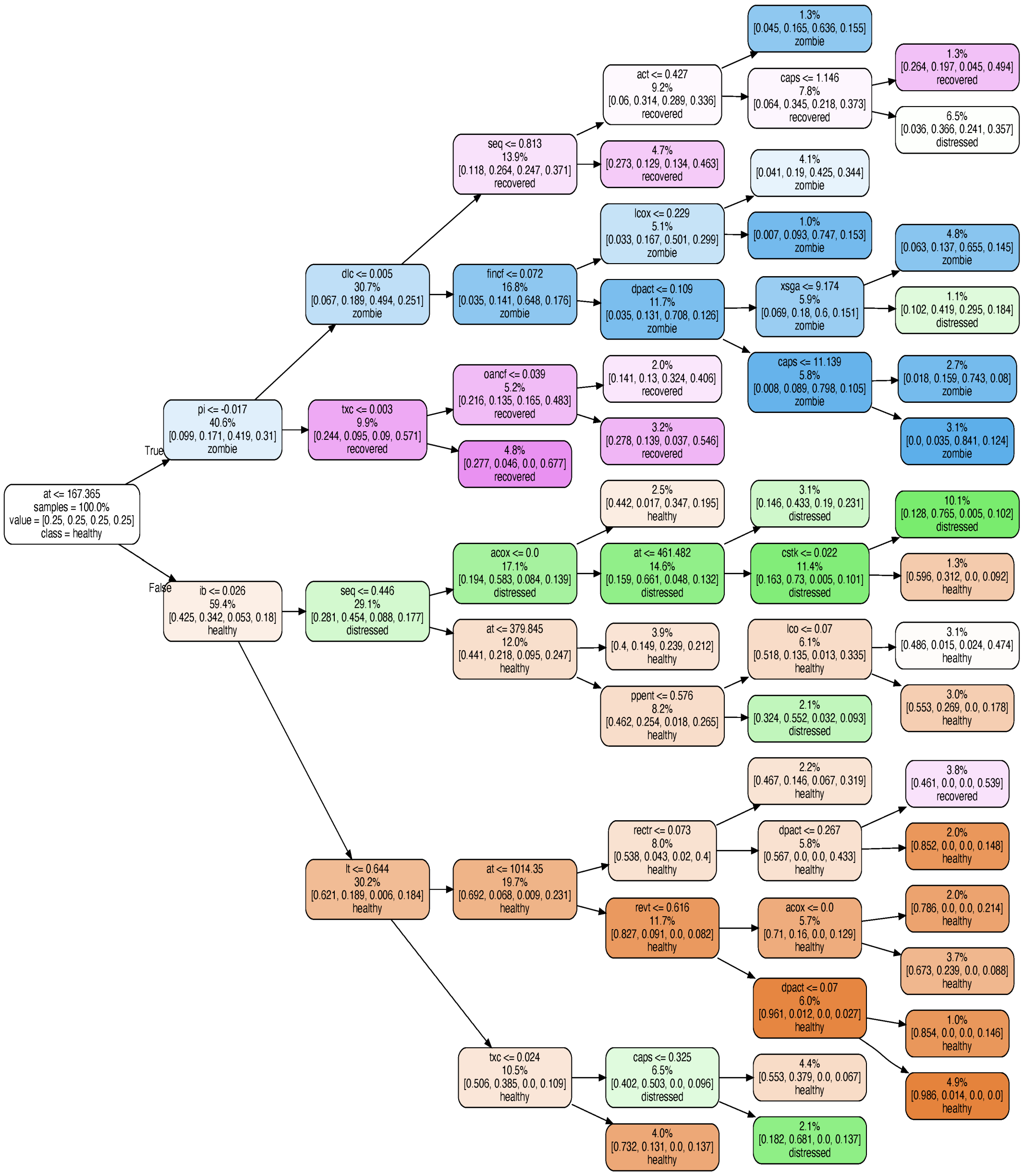


Figure 16: Zombie, Distressed, Recovered, and Healthy, United States 2016. This figure shows the multi-class decision tree for the US 2016. Distressed are measured with the Z-score (Altman 1968), while the zombie firms following (Banerjee and Hofmann 2020). Healthy are those that were never zombie nor distressed. Recovered are those that exited the zombie status. Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is provided at the top of each node. The nodes purity is given by a higher entropy and a darker color.

Legend: *at* Total Assets, *ib* Income Before Extraordinary Items, *pi* Pretax Income, *lt* Total Liabilities, *dlc* Debt in Current Liabilities, *revt* Total Revenue, *recch* Accounts Receivable, *sale* Sale/Turnover, *re* Retained Earnings, *txc* Current Income Taxes, *cstk* Common Stock, *oancf* Operating Activities Net Cash Flow, *acox* Other Assets excluding Deferred Changes, *seq* Shareholder's Equity, *lco* Other Current Liabilities, *caps* Capital Surplus/Share Premium, *dpact* Depreciation and Amortization, *rectr* Trade Accounts Receivable, *ppent* Total Property, Plant, and Equipment, *xsga* Selling, General, and Administrative Expenses, *fincf* Financing Activities Net Cash Flow.

3.5 Benchmark Analysis: Logistic Models

We analyze the performance of the firm-specific variables detected by the decision tree algorithm using logistic regression models. Table 2 and 3 show the results for the European country sample, while Table 4 and 5 report those for the US.

We perform separate logistic regressions for zombie and distressed companies for the years 2007 and 2016. We include dummy variables for the industries and countries/states to account for within countries/states differences. Ultimately, we provide 16 different logistic regression models that show the performance of the trees' extracted explanatory variables and the firm-specific differences and similarities. In both datasets, Europe and the United States, the industry and the location have a relevant effect.

We find that most variables chosen by the decision tree algorithm provide significant estimates also in the logistic models. The sign of the coefficient estimates is also mostly in line with the trees. More specifically, we document that income-related variables, such as non-operating income and pretax income, are a crucial indicators for zombie companies, where an increase in income decreases the probability of a company being a zombie. Income variables distinguish zombie from distressed, as income variables do not significantly impact distressed companies. The tree selects different variables for zombie and distressed companies. We further observe that the return on assets and the stock price provide significant information to categorize zombie companies but not distressed-types of firms in the US post-crisis. Besides the differences some similarities emerge, especially with respect to liabilities and debt. The higher significance level of total debt in liabilities for distressed companies is in line with our argumentation that distressed firms, in contrast to zombie, are in a more advanced stage of their financial unviability. In comparison to the income-driven European companies, for the US we find that stock market returns significantly categorize and determine distressed companies. Accordingly, lower stock prices predict distressed companies. We do not find this pattern for zombie companies. Similarly to Europe, industry-specific factors do play a relevant role also in the United States. Differences among US states, (Tables 4 and 5), emerge as well. At the same time, for US corporations, income and earnings variables are equally important for zombie and distressed firms categorization. Comparing the results between the 2007 and 2016, differences in the significance level of income variables, i.e. non-operating income is significant for the year 2007 but not for 2016, emerge for Europe. A similar pattern is found for the distressed in Europe and is equally true in the US. Some variables thus provide consistent explanatory power for zombie and distressed, however their influence changes before and after the financial crisis.

Logistic regressions, combined with decision trees, provide novel insights into zombie and distressed companies' firm-specific characteristics. The regression approach further shows that the decision tree finds essential variables from a sizeable input space and correctly splits the trees similarly to the coefficient estimates sign. Decision trees thus provide an excellent sorting tool for detecting and categorizing zombie and other non-viable type of companies.

Table 2: Logit Europe 2007

This table provides the results for four logistic regressions. In this table, we show the results for Europe in 2007, including country and industry dummies, and it allows to directly observe differences between zombie and distressed companies.

Country			Industry		
Variables	*Zombie*	*Distressed*	Variables	*Zombie*	*Distressed*
intercept	-1.769***	-1.228***	*intercept*	-1.75***	-1.37***
pi	-0.013***	0.000008	*pi*	-0.01***	0.00003
re	-0.0013***	-0.00003	*re*	-0.001**	-0.00003
nopi	0.012***		*nopi*	0.01***	
*dd*1	-0.007***		*dd*1	-0.006***	
dpact	-0.00061***		*dpact*	-0.0005	
cstk	-0.00032***	0.000095	*cstk*	-0.0007	0.00003
seq		-0.0000036***	*seq*		-0.0000005
sale		-0.0000008	*sale*		-0.0000005
ivao		-0.000002	*ivao*		-0.000002
oancf	-0.004**		*oancf*	-0.001	
Belgium	+	+	*Energy*	–*	–*
Bulgaria	–	–	*Materials*	–	–
Switzerland	–	–	*Capital Goods*	–	–
Cyprus	+	+***	*Commercials*	–	–
Czech Republic	–	–	*Transportation*	+	+
Germany	+	+	*Software*	–	–
Denmark	–	–	*Technology*	+	–
Spain	–	+	*IT Tech*	+	$^{-}$
Estonia	+	+	*Telecom Service*	+	–
Finland	+	–.	*Entertainment*	–	+
France	–	+	*Real Estate Invest*	+	+
Great Britain	–	$^{-***}$	*Automotive*	–	+
Greece	–	+***	*Consumer Durables*	+	+
Croatia	–	–	*Hotels*	–	+
Hungary	+	–	*Media*	+	+
Ireland	–	–	*Food, Beverage, Tobacco*	+	+
Italy	+	+*	*Household*	+	+
Lithuania	–	+	*Healthcare*	–	–
Luxembourg	–	–	*Pharmaceuticals*	+	–*
Latvia	+	–			
Macedonia	–	–			
Malta	–	+			
Netherlands	–	–			
Norway	–	–			
Poland	–	–***			
Portugal	–	+**			
Romania	–	–*			
Serbia	+	–			
Slovakia	–	+			
Slovenia	+	+			
Sweden	–	–			
Signif. codes:	0 : '***'0.001 : '**'	0.01 : '*'0.05 : '+'	< 0.05 : blank		

Table 3: Logit Europe 2016

This table provides the results for four logistic regressions. In this table, we show the results for Europe in 2016, including country and industry dummies, and it allows to directly observe differences between zombie and distressed companies.

Country			**Industry**		
Variables	*Zombie*	*Distressed*	Variables	*Zombie*	*Distressed*
intercept	-1.73**	-1.23***	*intercept*	-1.66***	-1.023***
cstk	-0.003***	0.0009	*cstk*	-0.0012**	-0.00024***
pi	-0.001***	0.00008	*pi*	-0.001***	0.0001
*dd*1	-0.003***		*dd*1	-0.002**	
nopi	0.00007		*nopi*	0.00006	
re	-0.007***	-0.00003	*re*	-0.0006***	-0.00002
seq		0.0000008	*seq*		-0.000054*
sale		-0.000008	*sale*		-0.000004
ivao			*ivao*		
ch		-0.0002*	*ch*		-0.0002*
ibc		-0.0002	*ibc*		0.00009
lt		0.0001***	*lt*		0.00009***
Belgium	+	+	*Energy*	+	−
Bulgaria	+	−	*Materials*	+	+
Switzerland	+	−	*Capital Goods*	+	+
Cyprus	+	+***	*Commercials*	−	+
Czech Republic	−	−	*Transportation*	−	+*
Germany	+	+	*Software*	+	−
Denmark	+	−	*Technology*	+	−*
Spain	+	+	*IT Tech*	+	+
Estonia	−	+	*Telecom Service*	−	+
Finland	−	−	*Entertainment*	+	+*
France	+	+	*Real Estate Invest*	+	+***
Great Britain	−	−***	*Automotive*	+	−
Greece	+*	+***	*Consumer Durables*	$^{+}$	+
Croatia	+**	−	*Hotels*	+	+**
Hungary	+	−	*Food, Beverage, Tobacco*	$^{+}$	+*
Ireland	+	−*	*Healthcare*	$^{+}$	−
Italy	+*	+*	*Pharmaceuticals*	$^{+}$	+
Lithuania	−	+			
Luxembourg	−	−			
Latvia	+	−			
Macedonia	+	−			
Malta	+	+			
Netherlands	+	−			
Norway	+	−			
Poland	−*	−***			
Portugal	+	+**			
Romania	+	−*			
Serbia	−	−			
Slovakia	+	+			
Slovenia	−	+			
Sweden	+	−			
Signif. codes:	0 : '***'0.001 : '**'	0.01 : '*'0.05 : '+'	< 0.05 : blank		

Table 4: Logit United States 2007

This table provides the results for four logistic regressions. In this table, we show the results for the United States in 2007, we include state and industry dummies, and it allows to directly observe differences between zombie and distressed companies.

Country			**Industry**		
Variables	*Zombie*	*Distressed*	Variables	*Zombie*	*Distressed*
intercept	-14.74	14.4	*intercept*	-3.1**	-1.54***
pi	-0.022***		*pi*	-0.02***	
oancf	-0.00004		*oancf*	-0.00004	
dlc	-0.0088***		*dlc*	-0.008***	
fopo	-0.04***		*fopo*	-0.04***	
nopi	-0.003***		*nopi*	-0.04***	
ib	0.007	-0.0007**	*ib*	0.007	-0.0009***
re		-0.0006***	*re*		0.0005***
lt		0.00003	*lt*		0.00003
seq		0.000006	*seq*		0.00005
dltt		-0.000004	*dltt*		0.0001*
Alaska	−	+	*Energy*	+*	+
Alabama	+	+	*Materials*	+	+
Arkansas	+	+	*Capital Goods*	+	+
Arizona	+	+	*Commercials*	+*	+
California	+	+	*Transportation*	+	+*
Colorado	+	+	*Software*	+	+*
Connecticut	+	+	*ITTech*	+	+
District of Columbia	+	+	*Telecom Services*	+**	+***
Delaware	+	+	*Entertainment*	+*	+***
Florida	+	+	*Real Estate Invest*	+	−
Georgia	+	+	*Automotive*	+*	+
Hawaii	+	+	*Consumer Durables*	+	+
Iowa	+	+	*Hotels*	+	+*
Idaho	+	+	*Media*	+	+***
Illinois	+	+	*Food, Beverage, Tobacco*	+	+
Indiana	+	+	*Households*	+*	+
Kansas	+	+	*Healthcare*	+*	+*
Kentucky	+	+	*Pharmaceuticals*	+	+***
Louisiana	+	+	*Real Estate*	+	−
Maine	+	+			
Maryland	+	+			
Michigan	+	+			
Minnesota	+	−			
Missouri	+	+			
Mississippi	+	+			
Montana	+	+			
North Carolina	+	+			
North Dakota	+	+			
Nebraska	+	+			
New Hampshire	+	+			
New Jersey	+	−			
New Mexico	+	+			
Nevada	+	+			
New York	+	+			
Ohio	+	+			
Oklahoma	+	+			
Oregon	+	+			
Pennsylvania	+	+			
Puerto Rico	−	−			
Rode Island	+	+			
South Carolina	+	+			
South Dakota	−	+			
Tennessee	+	+			
Texas	+	+			
Utah	+	+			
Virginia	+	+			
Vermont	−	+			
Washington	+	+			
Wisconsin	+	+			
Wyoming	−	+			
Signif. codes:	0 : '***'0.001 : '**'	0.01 : '*'0.05 : '+'	< 0.05 : blank		

Table 5: Logit United States 2016

This table provides the results for four logistic regressions. In this table, we show the results for the United States in 2016, we include state and industry dummies, and this allow to directly observe differences between zombie and distressed companies.

	Country			**Industry**	
Variables	*Zombie*	*Distressed*	Variables	*Zombie*	*Distressed*
intercept	17.67	-14.4	*intercept*	-1.1***	-0.087***
pi	-0.003***	-0.99***	*pi*	-0.004***	-0.001***
oancf	-0.0002		*oancf*	-0.00007	
dlc	-0.003*		*dlc*	-0.003**	
txt	0.0001		*txt*	0.00009	
capx	0.0001*		*capx*	0.00000005	
caps	-0.002***	-0.045***	*caps*	-0.002***	0.0002***
re		-0.0006***	*re*		-0.0002***
ib		-0.0007**	*ib*		-0.035**
seq		0.000006	*seq*		-0.0001**
dltt		0.000004	*dltt*		-0.035**
lct		0.0002***	*lct*		0.0002***
prcc return		-0.03***	*prcc return*		-0.2***
acox		0.0007	*acox*		-0.001
Alaska	−	+	*Energy*	+	$+^{*}$
Alabama	−	+	*Materials*	−	$+^{*}$
Arkansas	−	+	*Capital Goods*		+
Arizona	−	+	*Commercials*	+	+
California	−	+	*Transportation*	−	+
Colorado	−	+	*Software*	+	$+^{*}$
Connecticut	−	+	*Technology*	−	+
District of Columbia	−	+	*IT Tech*	+	−
Delaware	$-^{**}$	+	*Telecom Service*	+	$+^{***}$
Florida	−	+	*Entertainment*	+	$+^{*}$
Georgia	−	+	*Real Estate Invest*	+	$-^{***}$
Hawaii	−	+	*Consumer Durables*	−	−
Iowa	−	+	*Hotels*	−	+
Idaho	−	+	*Food, Beverage, Tobacco*	+	+
Illinois	$-^{*}$	+	*Household*	+	+
Indiana	−	+	*Healthcare*	+	$+^{**}$
Kansas	$-^{+}$	+	*Pharmaceuticals*	+	$+^{***}$
Kentucky	−	+			
Louisiana	−	+			
Maine	−	+			
Maryland	$-^{*}$	+			
Michigan	−	+			
Minnesota	−	+			
Missouri	−	+			
Mississippi	−	+			
Montana	−	+			
North Carolina	−	+			
North Dakota	−	−			
Nebraska	−	+			
New Hampshire	−	+			
New Jersey	−	+			
New Mexico	−	+			
Nevada	−	+			
New York	−	+			
Ohio	−	+			
Oklahoma	−	+			
Oregon	−	+			
Pennsylvania	−	+			
Puerto Rico	−	−			
Rode Island	−	+			
South Carolina	−	+			
South Dakota	−	+			
Tennessee	−	+			
Texas	−	+			
Utah	−	+			
Virginia	−	+			
Vermont	−	+			
Washington	−	+			
Wisconsin	−	+			
Wyoming	−	+			
Signif. codes:	0 : '***'0.001 : '**'	0.01 : '*'0.05 : '+'	$<$ 0.05 : blank		

4 Conclusion

The zombie phenomenon is not a myth, rather a reality affecting several countries globally since the late 1990s. The way this phenomenon manifest itself however can differ from one geographical area to another, given firm-specific differences, industry-specific factors, and diverging regulatory frameworks. This study proposes a supervised learning method, based on decision trees, to examine the characteristics of non-viable (i.e. zombie) companies in Europe and in the United States. The high dimensional dataset allows us to observe the incidence of zombification from an international perspective over several economic cycles. The aim is to better categorize zombie firms, separate them from other classes of companies, identify differences and similarities among zombie, distressed, healthy, and recovered zombies, and understand whether and to which extent firm and country-specific characteristics change during pre-crisis and post-crisis periods.

We use two exhaustive firm-level datasets, Compustat Global and Compustat North America, on a sample of publicly listed companies from the United States and 32 European countries over a period of twenty-two years. Such datasets allow us to feed a well-trained supervised learning algorithm that returns us a series of valuable information on zombie and non-zombie companies. We refrain from any a priori assumptions and instead use an algorithmic modeling to find the main characteristics. This method, gives us the privilege to put the zombie under the magnifying glass, monitor them over time and across countries, and add other classes of firms, viable and non-viable, such as the distressed, the healthy, and the recovered zombie, into a multi-class tree-like model. We can think of the decision trees as an efficient sorting mechanism.

The results show that, US zombie firms differ from their European peers on a modest number of firm-specific and industry-specific factors, but follow a similar pattern. Income and leverage-related variables are among the main drivers classifying zombie companies in Europe and in the United States. However, for US corporations shareholder's equity is a relevant driver that categorizes zombie versus non-zombies. Zombie firms are often misclassified as financially distressed companies, making their identification a challenging task that lacks a disciplined approach. To account for this, we further examine distressed-type of firms and compare them to the zombie using a binary decision tree setting. The findings show that zombie and distressed are often not comparable types of companies, rather companies at different financial stages. Differently to existing studies, we detect specific persisting characteristics. The decision trees suggest that both distressed and zombie firms in Europe have likewise a high debt-level component in their financial structure, but income-specific items especially categorize zombie, while leverage-related variables classify the distressed. In the United States, shareholder's equity categorizes and distinguishes distressed versus zombie firms. We find no major differences before and after the global financial crisis.

We further complement the classification trees with a series of logistic regressions on the various classes of identified companies and confirm the main findings. The results remain robust also to alternative zombie measure.

This study uses decision trees to categorize zombie companies, separate them from other firms' categories, and identify the most important driving features of viable versus non-viable firms in Europe and in the United States. We identify specific differences in the characteristics of zombie firms in Europe versus zombie in the United States, pointing to intrinsic differences in terms of corporate policies, industries, and regulatory frameworks. Contrary to existing studies using standard methods, classification trees can serve as an efficient sorting tool able to detect and distinguish various classes of companies, and to observe salient firm-specific characteristics that can help inform policy relevant interventions in coping with differing non-viable firms across countries and reducing the scope for zombie lending. Further research sees an extension of the method, using additional machine learning methods, to incorporate country-specific factors, insolvency frameworks, and information about private companies.

References

Acharya, Viral V., Lea Borchert, Maximilian Jager, and Sascha Steffen (2020). "Kicking the Can Down the Road: Government Interventions in the European Banking Sector". In: *National Bureau of Economic Research* 27537.

Acharya, Viral V., Matteo Crosignani, Tim Eisert, and Christian Eufinger (2020). "Zombie Credit and (Dis-)Inflation: Evidence from Europe". In: *National Bureau of Economic Research* 27158.

Acharya, Viral V., Tim Eisert, Christian Eufinger, and Christian Hirsch (2019). "Whatever it Takes: The Real Effects of Unconventional Monetary Policy". In: *The Reveiw of Financial Studies* 32.9, pp. 3366–3411.

Acharya, Viral V. and Krishnamurthy V Subramanian (2009). "Bankruptcy codes and innovation". In: *The Review of Financial Studies* 22.12, pp. 4949–4988.

Altman, Edward I. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy". In: *The Journal of Finance* 23.4, pp. 589–609.

Andrews, Dan and Filippos Petroulakis (2017). "Breaking the Shackles: Zombie Firms, Weak Banks and Depressed Restructuring in Europe". In: *OECD Working Papers* 1433.

Asquith, Paul, Robert Gertner, and David Scharfstein (1994). "Anatomy of Financial Distress: An Examination of Junk-Bond Issuers". In: *The Quarterly Journal of Economics* 109.3, pp. 625–658.

Baker, Malcolm and Jeffrey Wurgler (2002). "Market Timing and Capital Structure". In: *The Journal of Finance* 57.1, pp. 1–32.

Banerjee, Ryan and Boris Hofmann (2018). "The rise of zombie firms: causes and consequences". In: *BIS Quarterly Review* September.

— (2020). "Corporate zombies : Anatomy and life cycle". In: *BIS Working Papers* 882.

Bargagli Stoffi, Falco, Massimo Riccaboni, and Armando Rungi (2020). "Machine Learning for Zombie Hunting. Firms' Failures and Financial Constraints".

Breiman, Leo (2001). "Random forests". In: *Machine learning* 45.1, pp. 5–32.

Breiman, Leo, Jerome Friedman, Charles J Stone, and Richard A Olshen (1984). *Classification and regression trees.* CRC press. ISBN: 0412048418.

Caballero, Ricardo J., Takeo Hoshi, and Anil K. Kashyap (2008). "Zombie lending and depressed restructuring in Japan". In: *American Economic Review* 98.5, pp. 1943–1977.

Calligaris, Sara et al. (2016). "Italy's Productivity Conundrum. A Study on Resource Misallocation".

Campbell, John Y., Jens Hilscher, and Jan Szilagyi (2008). "In search of distress risk". In: *Journal of Finance* 63.6, pp. 2899–2939.

Chan, K C and Nai-Fu Chen (1991). "Structural and Return Characteristics of Small and Large Firms". In: *The Journal of Finance* 46.4, pp. 1467–1484.

Claessens, Stijn, Simeon Djankov, and Ashoka Mody (2001). "Resolution of Financial Distress. An International Perspective on the Design of Bankruptcy Laws". In: *WBI Development Studies* 23243.

De Martiis, Angela (2020). "The Effects of Organized Crime on Distressed Firms".

Deakin, Simon, Viviana Mollica, and Prabirjit Sarkar (2017). "Varieties of creditor protection: Insolvency law reform and credit expansion in developed market economies". In: *Socio-Economic Review* 15.2, pp. 359–384.

Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer (2007). "Private credit in 129 countries". In: *Journal of Financial Economics* 84.2, pp. 299–329.

Eisdorfer, Assaf, Amit Goyal, and Alexei Zhdanov (2018). "Distress Anomaly and Shareholder Risk: International Evidence". In: *Financial Management* 47.3, pp. 553–581.

Erickson, Timothy and Toni M. Whited (2006). "On the accuracy of different measures of q". In: *Financial Management* 35.3, pp. 5–33.

Fama, Eugene F. and Kenneth R. French (2001). "Disappearing dividends: Changing firm characteristics or lower propensity to pay?" In: *Journal of Financial Economics* 60.1, pp. 3–43.

Fan, Joseph P H, Sheridan Titman, and Garry Twite (2012). "An International Comparison of Capital Structure and Debt Maturity Choices". In: *Journal of Financial and Quantitative Analysis* 47.1, pp. 23–56.

Favara, Giovanni, Erwan Morellec, Enrique Schroth, and Philip Valta (2017). "Debt enforcement, investment, and risk taking across countries". In: *Journal of Financial Economics* 123.1, pp. 22–41.

Favara, Giovanni, Enrique Schroth, and Philip Valta (2012). "Strategic Default and Equity Risk Across Countries". In: *Journal of Finance* 67.6, pp. 2051–2095.

Frank, Murray Z and Vidhan K Goyal (2003). "Testing the pecking order theory of capital structure". In: *Journal of Financial Economics* 67.2, pp. 217–248.

Fukuda, Shin ichi and Jun ichi Nakamura (2011). "Why did 'zombie' firms recover in Japan?" In: *World Economy* 34.7, pp. 1124–1137.

Giannetti, Mariassunta and Andrei Simonov (2013). "On the real effects of bank bailouts: Micro-evidence from Japan". In: *American Economic Journal: Macroeconomics* 5.1, pp. 135–167.

Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez (2017). "Capital Allocation and Productivity in South Europe". In: *The Quarterly Journal of Economics* 132.4, pp. 1915–1967.

Gordon, M. J. (1971). "Towards a Theory of Financial Distress". In: *The Journal of Finance* 26.2, pp. 347–356.

Haselmann, Rainer and Paul Wachtel (2010). "Institutions and bank behavior: Legal environment, legal perception, and the composition of bank lending". In: *Journal of Money, Credit and Banking* 42.5, pp. 965–984.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media. ISBN: 0387848584.

Hoshi, Takeo (2006). "Economics of the living dead". In: *Japanese Economic Review* 57.1, pp. 30–49.

Jong, Abe de, Rezaul Kabir, and Thuy Thu Nguyen (2008). "Capital structure around the world: The roles of firm- and country-specific determinants". In: *Journal of Banking & Finance* 32.9, pp. 1954–1969.

Kahle, Kathleen M. and René M. Stulz (2017). "Is the US public corporation in trouble?" In: *Journal of Economic Perspectives* 31.3, pp. 67–88.

Kane, Edward J (1989). "The High Cost of Incompletely Funding the FSLIC Shortage of Explicit Capital". In: *Journal of Economic Perspectives* 3.4, pp. 31–47.

Kayhan, Ayla and Sheridan Titman (2007). "Firms' histories and their capital structures". In: *Journal of Financial Economics* 83.1, pp. 1–32.

McGowan, Müge Adalet, Dan Andrews, and Valentine Millot (2017). "Insolvency Regimes, Zombie Firms and Capital Reallocation". In: *OECD Working Papers* 1399.

— (2018). "The walking dead? Zombie firms and productivity performance in OECD countries". In: *Economic Policy* 33.96, pp. 685–736.

Myers, Stewart C. (2001). "Capital Structure". In: *The Journal of Economic Perspectives* 15.2, pp. 81–102.

Ohlson, James A. (1980). "Financial Ratios and the Probabilistic Prediction of Bankruptcy". In: *Journal of Accounting Research* 18.1, p. 109.

Peek, Joe and Eric S. Rosengren (2005). "Unnatural selection: Perverse incentives and the misallocation of credit in Japan". In: *American Economic Review* 95.4, pp. 1144–1166.

Peters, Ryan H. and Lucian A. Taylor (2017). "Intangible capital and the investment-q relation". In: *Journal of Financial Economics* 123.2, pp. 251–272.

Rodano, Giacomo, Nicolas Serrano-Velarde, and Emanuele Tarantino (2016). "Bankruptcy law and bank financing". In: *Journal of Financial Economics* 120.2, pp. 363–382.

Satu, Nurmi, Juuso Vanhala, and Matti Verín (2020). "The life and death of zombies - evidence from government subsidies to firms". In: *Bank of Finland Discussion Papers* 8.

Schivardi, Fabiano, Enrico Sette, and Guido Tabellini (2017). "Credit Misallocation During the European Financial Crisis". In: *BIS Working Papers* 669.

Schivardi, Fabiano, Enrico Sette, and Guido Tabellini (2020). "Identifying the Real Effects of Zombie Lending". In: *The Review of Corporate Finance Studies* 9.3, pp. 569–592.

Skinner, Douglas J. and Eugene Soltes (2011). "What do dividends tell us about earnings quality?" In: *Review of Accounting Studies* 16.1, pp. 1–28.

Storz, Manuela, Michael Koetter, and Ralph Setzer (2017). "Do we want these two to tango? On zombie firms and stressed banks in Europe". In: *European Central Bank Working Paper Series* 2104.

A Appendix

A.1 Variables Description

Abbreviation	Description
oiadp	Operating Income After Depreciation
pi	Pretax Income
cstk	Common Stock
roa	Return on Assets
roe	Return on Equity
epsincon	Earnings per Share including Extraordinary Items
epsexcon	Earnings per Share excluding Extraordinary Items
re	Retained Earning
txt	Total Income Taxes
txditc	Deferred Income Taxes
cshtr	Common Shares Traded
prch	Share Price High
prcl	Share Price Low
prcc	Share Price Close
ebitda	Earnings Before Interest Taxes Depreciation & Amortization
ivncf	Investing Activities Net Cash Flow
lt	Total Liabilities
lco	Other Current Liabilities
csho	Common Shares Outstanding
nopi	Nonoperating Income
cshpria	Common Shares for Basic Earnings Per Share
dpact	Depreciation, Depletion and Amortization
dpc	Depreciation and Amortization
dlc	Total Debt in Current Liabilities
ero	Other Equity Reserves
dvpsx	Dividends Per Share Ex-date
dvpsp	Dividends Per Share Pay-date
wcap	Working Capital
at	Total Assets
oancf	Operating Activities Net Cash Flow
np	Notes Payable Short-Term Borrowing
ib	Income Before Extraordinary Items
xacc	Accrued Expenses
seq	Stockholders Equity
sale	Annual Sales
ch	Cash and Due from Banks
ivao	Investment and Advances - other
dd1	Long-Term Debt due in 1 year
ibc	Income before Extra Items
dltt	LT Debt -Total
capx	Capital Expenditures - Annual

caps	Capital Surplus
acox	Current Assets

Table A1: Binary and Multi-Class Trees Variables List. The Table reports a summary of the firm-level variables returned by the classification trees with their respective Compustat item name and description.

Variable	**Definition**
Book Leverage =	*Total Liabilities / Total Assets*
Net Book Leverage =	$(Total\ Debt - Cash\ \&\ ST\ Investments)\ /\ Total\ Assets$
Market Value =	$Nr.\ Common\ Shares\ Outstanding \times Share\ Price$
Market Leverage =	$Total\ Debt\ /\ (Total\ Debt + Preferred\ Stock\ at\ Book\ Value + Common\ Equity\ at\ Market\ Value)$
Asset Tangibility =	*Net PP&E / Total Assets*
Cash & ST Investment Ratio =	*Cash & ST Investments / Total Assets*
Return on Equity (ROE) =	*Net Income / Total Shareholders Equity at Book Value*
Profit Margin =	*Net Income / Sales*
Capex Ratio =	*Capital Expenditures / Total Assets*
Dividend Yield =	$Dividends\ per\ Common\ Share\ /\ Price\ per\ Common\ Share_{t-1}$
Total Payout Ratio =	$(Dividends + Repurchases)\ /\ Net\ Income$
Δ Total Assets =	$(Total\ Assets_t - Total\ Assets_{t-1})\ /\ Total\ Assets_{t-1}$
Return on Assets (ROA) =	*Operating Income after Depreciation / Total Assets*
Size =	$Log(Total\ Assets)$

Table A2: Variables Construction. The table reports a list of profitability ratios used as additional performance measures. Data from Compustat North America and Compustat Global.

A.2 Descriptive Statistics

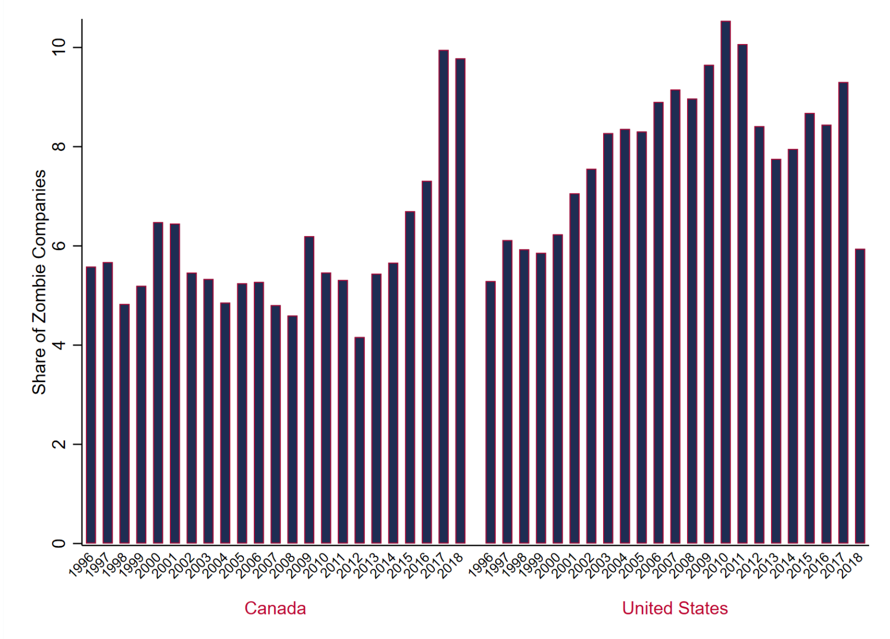


Figure A1: Zombie Shares, United States and Canada. The figure shows the share of zombie in the US and Canada from 1996 to 2018. Zombie firms are measured with the main definition.

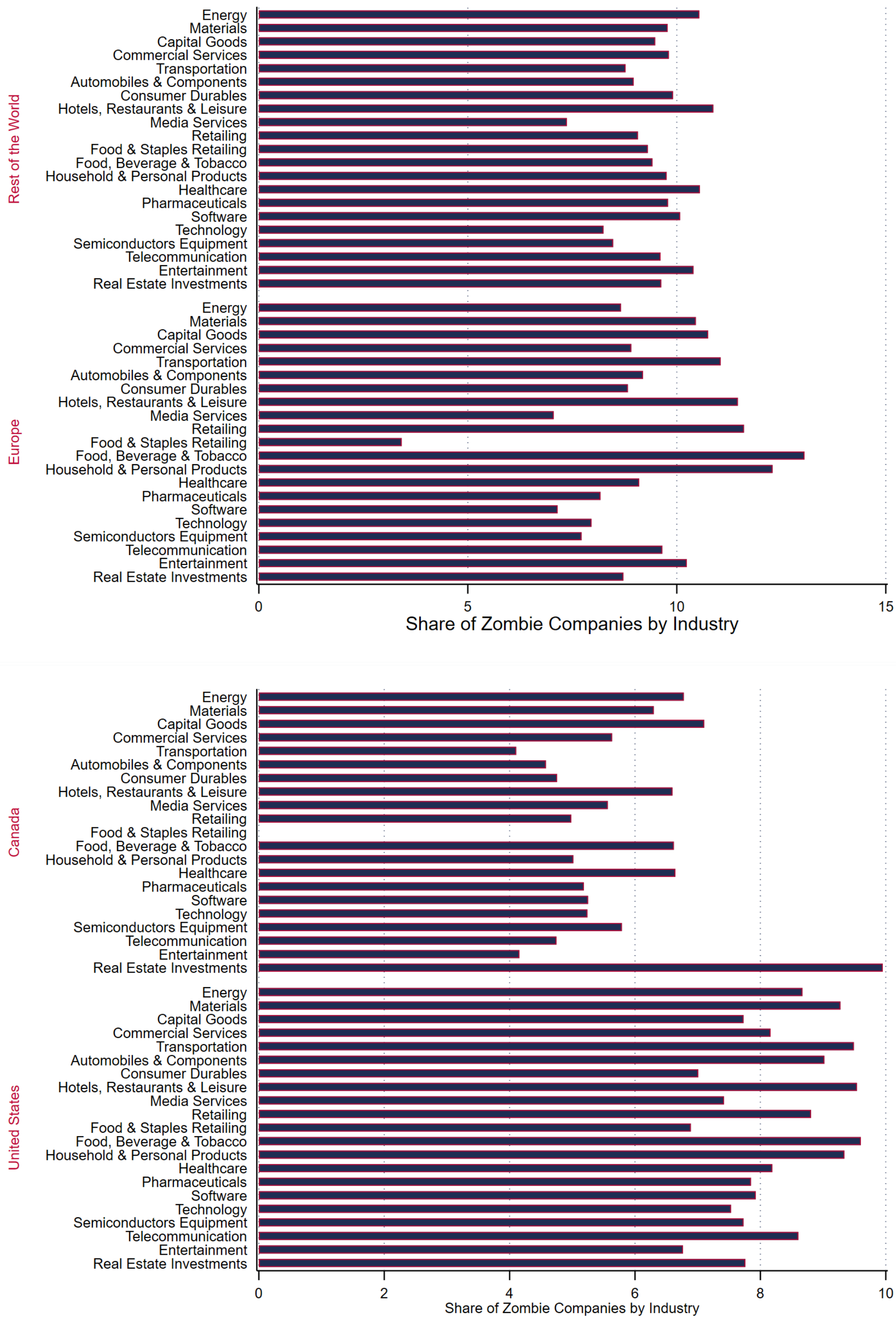


Figure A2: Zombie Shares by Industry. The upper chart shows the zombie shares by industry, GIC group, in Europe and in the Rest of the World, the lower chart in the United States and Canada. Zombie firms are measured with the main definition.

A.3 Measuring Zombie

The strand of literature focusing on zombie companies provides different approaches to identifying a company as a zombie, each one of them with its own advantages and drawbacks. The zombie definition itself, together with data limitations, explains the existing measurement challenges.

One of the first measures originates from the study of Caballero, Hoshi, and Kashyap (2008) on Japanese companies during the 1990s banking crisis. The authors classify a company as a zombie whenever it receives subsidized credit at an interest rate that is below the one applied to the most creditworthy companies. The actual interest payments made by the companies are then compared to an estimated benchmark, R^*, based on the firm's debt structure and market prime rate. The minimum required interest payment for each firm i in year t, R^*_{it}, is defined as:

$$R^*_{it} = rs_{t-1}BS_{it-1} + \Big(\frac{1}{5}\sum_{j=1}^{5} rl_{t-j}\Big)BL_{it-1} + rcb_{\text{min over last 5 years, t}} \times Bonds_{it-1}. \qquad (1)$$

where BS_{it}, BL_{it}, and $Bonds_{it}$ represent short-term loans (less than one year), long-term bank loans (more than one year), and total bonds outstanding (including convertible bonds and warrant-attached bonds), respectively, for firm i at end of year t; while rs_t, rl_t, and $rcb_{\text{min over last 5 years, t}}$ represent the average short-term prime rate in year t, the average long-term prime rate in year t, and the minimum coupon rate on any convertible corporate bond issued in the last five years before t.

Given that we are interested in examining the characteristics of zombie firms from an international perspective, and given the data constraints, replicating such measure does not fit the study.[17] As noted in Banerjee and Hofmann (2018), by employing this measure one would encounter three potential limitations: *(i)* identifying with precision the subsidized credit granted by the banks to the companies would be a challenge, *(ii)* banks may grant subsidised credit for other reasons, such as long-standing relationships, and *(iii)* when interest rates are very low for longer periods, subsidized lending rates would have to be near zero or even negative.

For these reasons, the more recent zombie literature often adopts a definition that relies on the accounting information of such firms to capture their unproductive nature, their age, and in some cases their future growth potential. The most widely used measure evolves around the interest coverage ratio, initially used in the study of McGowan, Andrews, and Millot (2018), and subsequently used in other academic studies and central banks' reports. The interest coverage ratio is a measure that goes beyond the debt composition of the company and looks at the operating income and at the persistency of the condition of *distress*. A company is considered a zombie whenever its $ICR_{it} < 1\ for\ 3\ consecutive\ years\ and\ age \geq 10\ years$. In Banerjee and Hofmann (2018) the latter measure is complemented with an additional factor, the company's future growth potential, captured with the Tobin's q. In Banerjee and Hofmann (2020) the age factor is dismissed. In Acharya, Crosignani, Eisert, and Eufinger (2020) a company is considered a zombie if it meets two criteria: (i) the firm's ICR is below the median and its leverage ratio is above the median, (ii) the share of interest expenses relative to the sum of its outstanding loans, credit, and bonds in a given year is below the interest paid by the most creditworthy firms. The latter criterion follows the subsidised credit measure of Caballero, Hoshi, and Kashyap (2008). In our study, we follow Banerjee and Hofmann (2020).

[17] From Caballero, Hoshi, and Kashyap (2008), we recall that the authors do not know the exact interest rates on specific loans, bonds, or commercial paper, nor the exact maturities of any of these obligations. Overall, the subsidised credit definition fits with the investigation of the zombie lending channel (Giannetti and Simonov 2013; Acharya, Eisert, Eufinger, and Hirsch 2019).

A.4 Decision Tree Example

Let us assume that we want to predict if a person pays back a credit and we have the age and account balance of that person. The tree searches the variable age and account balance for a split point that helps to separate the persons who repay the credit from those who do not. As an arbitrary example, the tree may split the variable account balance at \$1905, meaning that the tree separates all persons with an account balance lower than \$1905 from those with a higher account balance. Therefore, the input space X is separated. This searches for the best variable and iterates the split point until a stopping criteria is fulfilled. Note that in the next step the tree may separate out the input space of persons with less than \$1905 and an age higher than 54. Below, we show how this artificial example would translate into a simple decision tree setting:

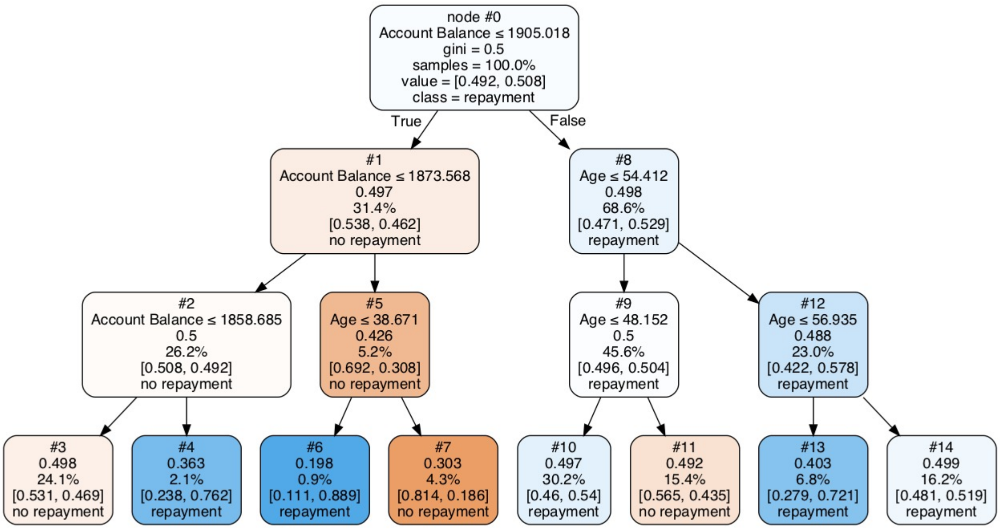


Figure A3: Credit Repayment Example. This figure provides a simple example of a credit repayment process based on artificial data. The objective is to show the mechanism behind a decision tree algorithm. Source: Authors' own estimations.

A.5 Additional Results

Figure A4 shows the firm-specific characteristics of zombie vs. non-zombie in Europe using an alternative zombie definition that follows McGowan, Andrews, and Millot (2018). Data relates to the crisis years, 2007, in the upper tree, and healthy years, 2016, in the lower tree.

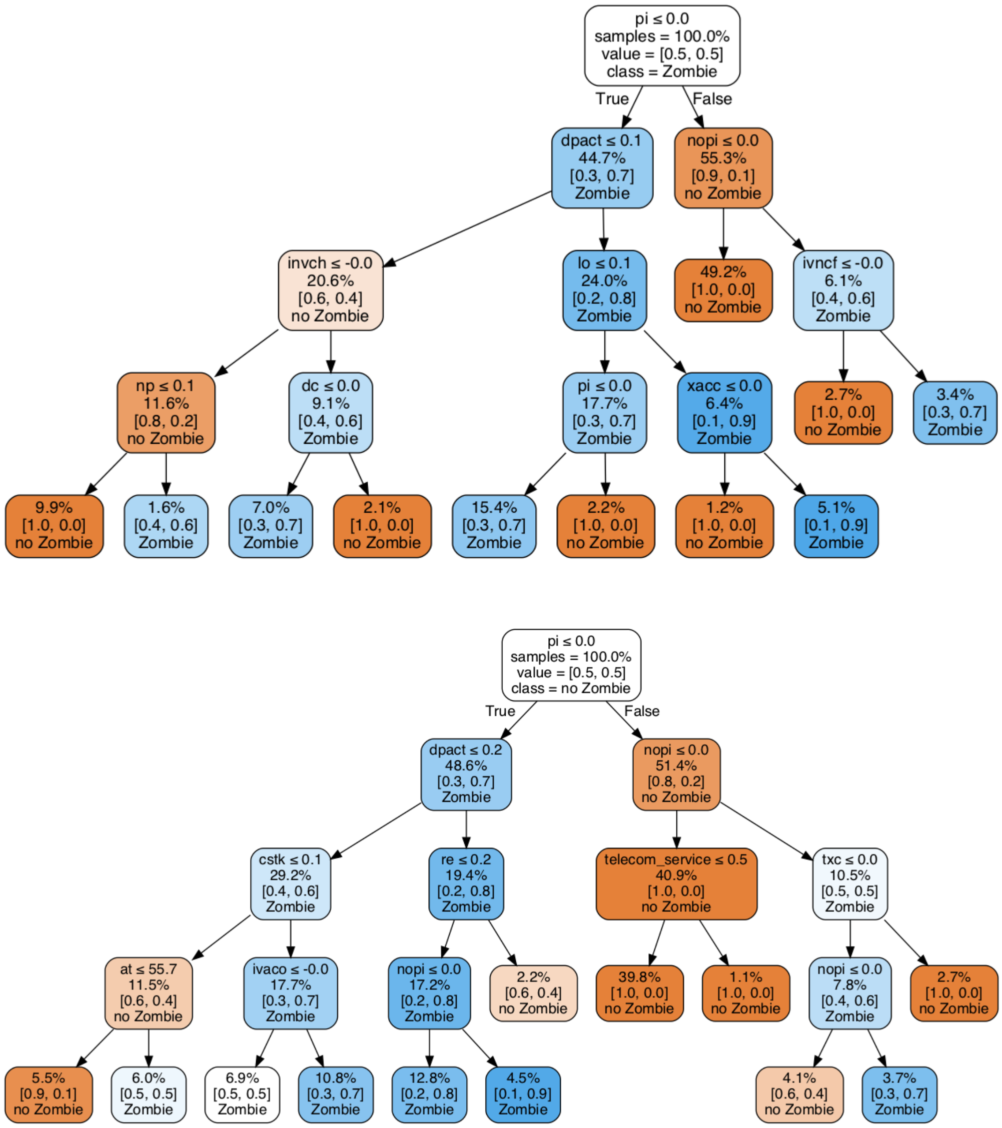
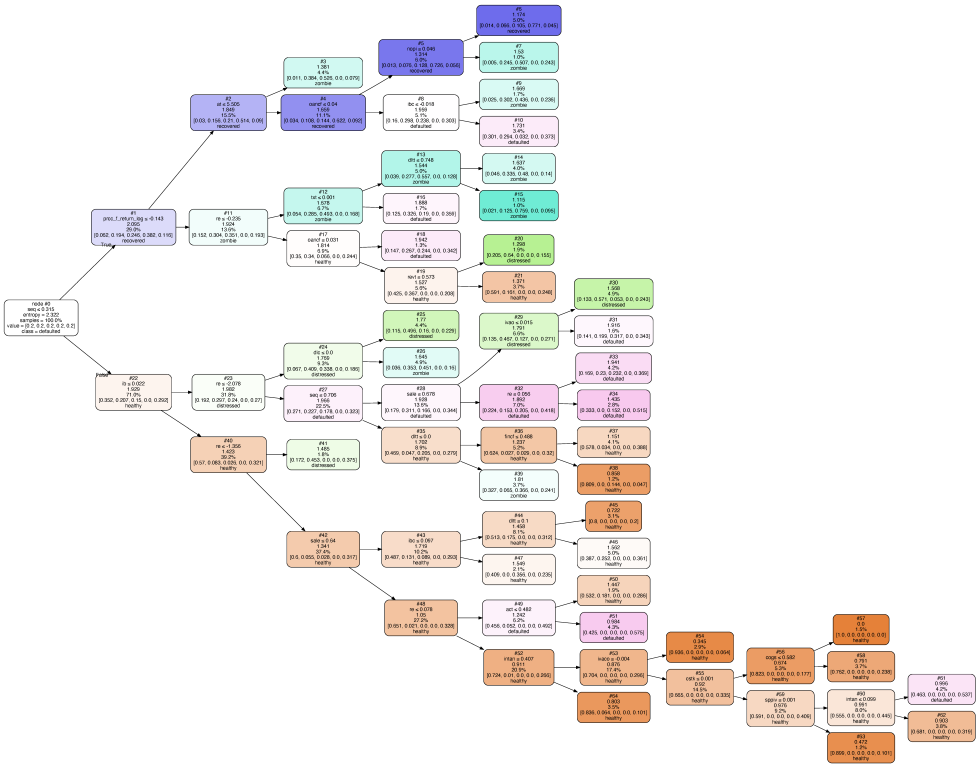


Figure A4: Zombie versus Non-Zombie, Europe. Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is at the top of each node. The nodes purity is given by higher entropy and darker color.

Legend: *oiadp* Operating income after depreciation, *cme* Common market equity, *pi* Pretax income, *cstk* Common stock, *bl* Book leverage, *nopi* Nonoperating income, *txt* Total income taxes, *roa* Return on assets, *dpact* Depreciation, depletion and amortization, *fincf* Financing activities net cash flow.

Figure A5 below shows the firm-specific characteristics of zombie vs. non-zombie in the US using an alternative zombie definition that follows McGowan, Andrews, and Millot (2018). Data relates to the crisis years, 2007, in the upper tree, and healthy years, 2016, in the lower tree.



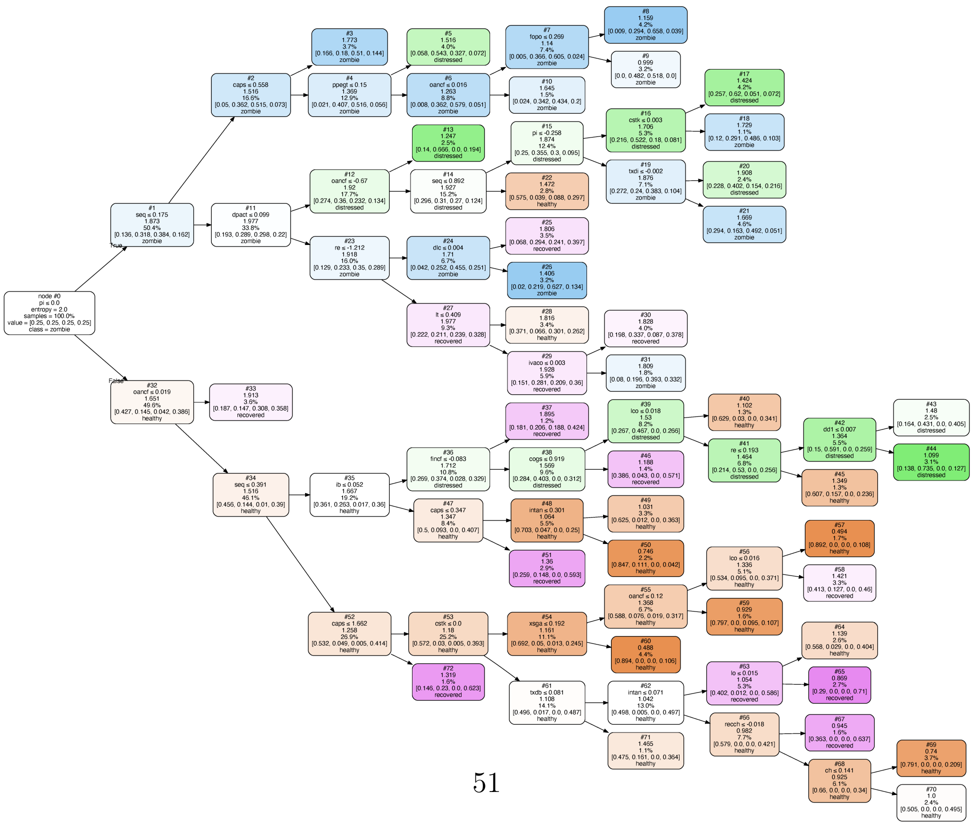


Figure A5: Zombie versus Non-Zombie, United States. Higher splits provide higher importance for the decision. The decision iteration of the CART algorithm is at the top of each node. The nodes purity is given by higher entropy and darker color.

Legend: *oiadp* Operating income after depreciation, *cme* Common market equity, *pi* Pretax income, *cstk* Common stock, *roa* Return on assets, *nopi* Nonoperating income, *dpact* Depreciation, depletion and amortization, *fincf* Financing activities net cash flow.