Corporate bankruptcy prediction: Can KMV-Merton model add value to support vector machines forecasts?

Francisco Loureiro

Dissertation written under the supervision of Nuno Silva

Dissertation submitted in partial fulfilment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, January 2020.

Abstract

Title: Corporate bankruptcy prediction: Can KMV-Merton model add value to support vector machine forecasts?

Author: Francisco Soares da Cruz Loureiro

Keywords: Bankruptcy prediction; Machine Learning; Structural credit models; Support Vector Machines.

This dissertation aims to assess if the output from the KMV-Merton model, the so-called distance to default, can contribute to the support vector machines model with the ultimate goal of better forecasting the bankruptcy of a company. The considered dataset covers 248 non-financial U.S. companies between 2000 and 2018. It was found evidence that the distance to default contributes, within a given range of variables considered, to a better F1-Score using both cross-validation and percentage ratio split. Additionally, the results show that the distance to default is a better predictor than a simpler market-based variable such as the debt-to-equity ratio. This suggests that the Merton-model setup *per se* is useful for default prediction. As expected, taking the F1-Score as a reference, the results also indicate that using company information a year prior to default provides better results than using data two years prior to default. Lastly, given the dataset used and the assumptions stated, this study is not conclusive regarding which out-of-sample evaluation method offers better results, the percentage ratio split, or the stratified K-fold cross-validation.

Resumo

Título: Previsão de falência de empresas: Pode o modelo KMV-Merton adicionar valor ás previsões da máquina de vetor suporte?

Autor: Francisco Soares da Cruz Loureiro

Palavras-chave: Previsão de falência; Machine Learning; Modelos estruturais de crédito; Máquinas de Vetor de Suporte.

Esta dissertação tem como objetivo avaliar se o resultado do modelo KMV-Merton, a conhecida distância ao incumprimento, pode contribuir para o modelo de máquinas de vetor de suporte com o objetivo final de prever melhor a falência de empresas. O conjunto de dados considerado abrange 248 empresas não financeiras dos E.U.A entre 2000 e 2018. Encontra-se evidência que a distância ao incumprimento contribui, dentro de um determinado grupo de variáveis, para um melhor F1-Score utilizando tanto a validação cruzada como a divisão percentual. Além disso, os resultados mostram que a distância ao incumprimento é um melhor previsor comparativamente a uma variável de mercado mais simples tal como a dívida sobre o valor de mercado do capital próprio. Isso sugere que a configuração do modelo Merton por si só é útil para a previsão de falência. Como esperado, considerando o F1-Score como referência, os resultados também indicam que o uso de informações da empresa um ano antes da falência fornece melhores resultados do que o uso de dados dois anos antes da falência. Por fim, dado o conjunto de dados usados e as premissas assumidas, este estudo não é conclusivo em relação a qual método de avaliação *out-of-sample* oferece melhores resultados, a divisão percentual ou a validação cruzada.

Acknowledgments

This dissertation represents the end of a long road made of highs and lows but certainly of very few regrets. I had the pleasure to learn and to work within an environment of companionship and excellence, something I will carry with me for the rest of my life.

First of all, I have to deeply thank my supervisor, Nuno Silva, for being an authentic lighthouse in the development of this thesis. The constant support and motivation provided were crucial to close this chapter of life.

Second, I have to thanks my brother, Tiago Loureiro, for being someone I will always look up to. Without his presence, undoubtedly, I would not be who I am today.

Third, I would like to thank my father and my mother for always being present on my best and worst moments, something I will never be able to repay them.

Lastly, I want to thanks my colleagues and professors over the years, that in one way or another, shaped me to be who I am, hopefully, a better citizen.

Table of Contents

List of Figures

List of Tables

List of Appendix

Chapter 1

1 Introduction

1.1 Context and motivation

Credit risk became a very familiar concept in the last few years, primarily as a result of the 2007 crisis. One can define it as the risk of one party failing to pay the other party a previously agreed amount at a given date. This risk represents a severe threat, particularly to banks, as their business model consists of lending money to other parties. Segregating good counterparties from bad counterparties is thus essential for bank profitability and solvency. In addition, from Basel II onwards, banks may, if authorized by national supervisors, use internal rating models in order to estimate the probability of default of their clients. This means that good default prediction models are important not only to avoid credit losses but also to perform regulatory capital calculations.

Though increasingly important, the development of models to predict corporate bankruptcy has been a critical topic in finance for both practitioners and academics throughout the last century. The literature goes back to the beginnings of the 1930s. However, until the 1960s, the literature was mainly focused on univariate analysis. In 1968, Altman presented a multivariate model that is still very relevant as a criterion to successfully assess the credit risk of a corporation. The author proposed a five-factor model to predict the bankruptcy of manufacturing companies. This model became widely known as the Altman Z-score. In the subsequent years, other discriminant models were developed differing among each other, mostly on the number of factors considered. In 1980, Ohlson pioneered the use of logit-based models in the bankruptcy prediction field. The author was motivated by fragilities recognized in multivariate analysis, such as the lack of interpretability, for instance, of the Altman Z-score. Currently these are still the most used default prediction models.

In parallel to the above referred papers, a new class of models, called structural credit risk models, emerged from Black and Scholes (1973) option pricing theory. Differently from previous models, which were focused only on default prediction, these models aim to provide a way of relating the credit risk of a firm and its capital structure. Robert Merton's groundbreaking paper released in 1974 was the first to make use of the Black and Scholes theory. The author perceived that the equity of a firm could be seen as a call option on its assets with strike price equal to the face value of debt. Consequently, when the asset value falls below this threshold, the call option is not exercised, and the firm is handed over to its debtholders. The default probability can thus be seen as the probability of the asset value falling below nominal debt value. Building on the Black-Scholes-Merton framework, the KMV corporation, developed a model based on Merton´s 1974 paper in the late 1980s. This was entitled as KMV-Merton model. This model brought two major enhancements over the Merton model. First, Merton model assumes that the whole debt of a firm is exclusively constituted by a single zero-coupon bond. In reality, a firm's debt structure is not as straightforward as that. In addition, it is possible that a firm continues operating despite a negative net worth value. According to KMV, an effective approach is considering that the value of debt, which is the strike price of the option, equals the sum of current liabilities and half of the long-term liabilities. Secondly, the KMV-Merton model does not use a normal distribution when assessing the probability of default. Instead, KMV uses a proprietary empirical distribution of default rates. These two enhancements have been referred to improve Merton's model capacity to predict default. Still, as structural models are calibrated using stock markets data, the use of these models continues mostly constrained to publicly traded firms. This important constrain has prevented the use of these models in the retail banking sector.

In the late 1980s, machine learning methods started being applied in the bankruptcy prediction field. Machine learning is a general term that encompasses a large number of techniques ranging from neural networks to random forests and support vector machines. Several studies, such as Huang et al. (2004), have demonstrated that these methods have better predictive capacity than more traditional statistical methods. As the authors state, the major difference is that traditional statistical methods impose structures to models, for instance, linearity in regression analysis. On the other hand, machine learning methods allow the model to learn the specific data structure without impositions. Support vector machines, which was introduced by Vapnik (1998), is one of the most recent machine learning techniques. The aim of this model is finding a hyperplane capable of maximizing the distance between two decision classes. This principle can thus be applied in order to separate bankrupt from non-bankrupt companies.

1.2 Goals and document structure

The main objective of this dissertation is to check whether one of the major outputs of the Merton model, the distance to default, can add value to the SVM framework. With this purpose, accounting and market data are gathered both from bankrupted firms and non-bankrupted firms during the period that ranges between 2000 and 2018. This dissertation also covers three additional questions:

- Is the distance to default a better contributor to the SVM model as opposed to a simpler market-related variable, such as debt-to-market equity ratio?
- Does cross-validation provide better results as opposed to the traditional 80:20 split?
- What are the differences in the prediction results using instances[1] of companies who went bankrupt a year prior to the default event compared to using instances of the same companies but two-year prior?

Regarding the document structure, the next chapter discusses the main literature on bankruptcy prediction giving a historical perspective on the evolution of the field. In Chapter 3, it will be explained and debated the models which are going to be employed. In Chapter 4, an explanation of how the used datasets were constructed is provided and some statistical analyses are presented. Chapter 5 describes how the models will be estimated and explains the several metrics used in the assessment. Chapter 6 presents and analyzes the obtained results as well as answering the research questions. Finally, Chapter 7 concludes this report.

[1] Hereafter, each company data regarding a certain year will be labeled as an instance.

Chapter 2

2 Literature Review

2.1 Initial Studies on credit risk prediction

The initial studies on bankruptcy prediction were univariate analysis, which aimed to ascertain which ratio could better forecast the future financial position of a firm. One of these frontline studies was conducted by Merwin (1942). The author decided to analyze small manufacturers. In order to do so, Merwin collected data from one thousand companies, whose assets in 1926 amounted to less than $250.000 from five different industries. Using data from 1926 till 1936, the author compared the mean ratios of non-bankrupt firms with those who filed for bankruptcy. This study culminated in two main conclusions. First, the financial characteristics of companies that eventually went bankrupt start to differ from the most successful ones four to five years before the bankruptcy event occurs. Second, three ratios were found to be particularly powerful indicators of possible business failure: net working capital to total assets, current assets to current liabilities and net worth to total debt.

Following the same line of thought, Beaver (1966) selected seventy-nine failed firms and seventy-nine non-failed firms from thirty-eight different industries and gathered data from 1954 to 1964. The author computed thirty ratios from the financial statements and proceeded to examine and test the individual ratio predictive ability to correctly classify non-bankrupt and bankrupt companies. From this study, it was concluded that failed firms tend to incur in more debt compared to the non-failed ones. The author also found that the two ratios with better predictability were net income to total debt and net income to sales. Beaver´s paper section regarding suggestions for future research provided great insights for what would happen next as the author suggested a multi-ratio analysis. This analysis, instead of testing the predictive ability of each ratio as its own, would consider several different ratios together.

In accordance with the abovementioned research, Altman (1968) develop a multivariate study using a sample of thirty-three non-failed manufacturing firms and thirty-three failed manufacturing firms. The author prosecutes a multiple discriminant analysis in order to predict the bankruptcy of a firm. Altman starts from a list of twenty-two different ratios and ends up selecting the five with the higher predictive ability of a bankruptcy event. These five ratios are:

working capital to total assets, retained earnings to total assets, earnings before interest and taxes to total assets, market value of equity to book value of total debt, and sales to total assets. The output of this model is known as the Altman Z-score. Firms with a Z-score superior to 2.99 are predicted as safe from bankruptcy while those who present a value below 1.81 are classified as vulnerable. The author labels the zone between these two areas as the "gray area," and no conclusions can be taken while in this range. The Altman Z-score displayed a 95% accuracy in predicting the default of a certain firm one year prior to the event. Nonetheless, the predictive ability decreases as the number of years to the bankruptcy event increases.

Ohlson (1980) recognizes some fragilities on multivariate discriminant analyses leading him to build a conditional logic model. The vulnerabilities identified are mainly concerned with statistical requirements imposed by these models, difficult interpretation of the output scores, and the fact that failed and non-failed firms are matched by criteria such as size and industry, which the author refers to as "somewhat arbitrary". Ohlson´s score is the result of nine independent variables that he argued to have predictability power but gave no theoretical justification for its selection. The period selected by the author in his analysis ranged between 1970 and 1976 in which he observed one hundred and five failed firms and two thousand non-failed firms that have been trading on US stock exchange for a minimum of three years. He proceeded with the estimation of three different models applying a logistic regression with different cut-off points. The models aimed to predict bankruptcy within one year, within two years and between one and two years, respectively. The results indicate that the size, the financial structure of the firm, and current liquidity ratio are vital variables in order to ascertain a possible bankruptcy event.

Although the majority of multivariate analysis performed well, certain criticisms can be made. Altman and Saunders (1996) pinpoint three main concerns. First, these models are based on book value accounting data, which neglects the continuous nature of the borrower's conditions. Second, modeling real-world conditions assuming linear relations is most likely a mistake. Finally, these credit-scoring bankruptcy prediction models are barely linked to a theoretical model.

2.2 The Black-Scholes-Merton model and its applications

A more prominent alternative emerged with the work of Black and Scholes (1973) and Merton (1974). The Black and Scholes model is an option pricing model based on the premise that it should not be possible to make profits by creating portfolios of either long or short positions in options and their underlying stock. According to this model and in order to calculate the price of an option, the required inputs are - the strike price, the current stock price, the time to expiration, and the volatility. Merton (1974) makes use of this framework, assuming that the capital structure of the firm is constituted by equity and by a zero-coupon bond. The asset value is, within this framework, considered to be the sum of debt and equity. It also assumes that asset value follows a geometric Brownian motion process. Intuitively, the idea is that the asset value in the next moment in time is similar to the previous asset value plus some independent random change. Under this model, Merton reckons that debtholders are shorting a put option on the assets of the firm. Using the put-call parity, equity can be seen as a call option on the firm´s assets with strike price equal to the face value of debt. If the firm´s asset value is less than the face value of debt at maturity, equity holders will deliver the firm to the bondholders. Contrarily, if the firm´s asset value at maturity is higher than the face value of debt, equity holders will not default.

Black and Cox (1976) extended this view. The authors considered in their model that default may occur whenever the asset value falls below a specific threshold even before debt maturity. The explanation lies in the fact that bondholders, within this framework, have the right to exercise a safety covenant, allowing them to liquidate the firm.

Agarwal and Taffler (2008) point out the benefits of using the Black-Scholes-Merton framework as opposed to the accounting ratio-based models. First, these models have a reasonable theoretical framework behind them. Second, and according to the efficient market theory, market prices should manifest all the information contained in and out of the accounting statements. Third, market-related variables are not likely to be influenced by firm accounting policies. Fourth, stock prices are expected to express future expected cashflows hence more pertinent in order to make predictions. Lastly, the output is neither time or sample dependent.

Although this model was a huge breakthrough, many of the assumptions are disregarded in practical implementation. KMV-Merton model, firstly developed by KMV corporation and later acquired by Moody´s in the late 1980s, is built on the application of financial derivatives theory based on Merton´s (1974) framework. The purpose of the model is to provide an assessment of how likely a company is to default. According to Crosbie et al. (2003), the market value of the firm´s assets, asset risk, and leverage are the three main root factors that are vital to ascertain the probability of default. According to the authors, the risk of default will be higher whenever the market value of assets approximates to the level of the book value of liabilities. This means that the probability of default will depend positively on the asset volatility and the value of liabilities and negatively on the market value of corporate assets. A default event will ultimately occur when the market value of assets is not enough to repay the liabilities.

In order to apply the KMV-Merton model, two unknowns must be quantified, which are the firm´s asset value and asset volatility. In order to calculate these and under Merton´s framework, one possible approach is solving a system of two nonlinear simultaneous equations. A second alternative consists of using an iterative approach that imposes constant asset volatility during the estimation process. Any increases in equity volatility are then attributed to leverage variations. Hence, in periods of analysis where leverage drastically changes, the authors recommend the use of the latter.

2.3 Machine learning

Murphy et al. (2019) define machine learning as:

> *"…an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment."*

Over the past decades, machine learning has gained a notorious interest in a range of different fields. As Mitchell et al. (2015) state, the use of machine learning has been adopted in a variety of important areas such as in health care, manufacturing, education, marketing, financial modeling and policing. However, it was during the 1990s that machine learning first major real-world application was presented to us as the spam filter emerged.

According to Géron (2019), machine learning tasks are typically classified into three broad categories: supervised learning, unsupervised learning, and reinforcement learning. The first relies on data that is labeled. The second is commonly used for problems involving clustering tasks, whereas the data is unlabeled. Last but not least, reinforcement learning, which can be described as a learning tool in which an agent observes the surrounding environment, select and perform actions. Consequently, rewards or penalties are given. Hence, the agent needs to choose the best strategy in order to maximize rewards. Within each of these categories, there are several machine learning techniques that can be used depending on what problem one is facing. The most acknowledged models within machine learning are K-nearest neighbors, support vector machines, decision trees, random forests, and lastly, neural networks, which are all integrated into the category of supervised learning algorithms.

One of the first machine learning techniques applied to predict bankruptcy was neural networks, which was inspired by the way human brain functions. Neural networks are an attempt to simulate the human´s biological neural networks. The most basic form of these models can be summarized in three steps. First, every input value is multiplied by a certain weight. Second, the weighted input values are summed with a bias term. Finally, the obtained result from the second step is passed through an activation function. Odom et al. (1990) decided to compare the predictive ability of neural networks and the multivariate discriminant analysis in predicting bankruptcy. The sample of companies selected was composed of sixty-five firms that went bankrupt and sixty-four non-bankrupt firms from 1975 to 1982. The authors considered as input values the same variables that Altman considered in his work in 1968. This study's results demonstrated that neural networks performed better than the multivariate discriminant analysis in all three training sample proportions considered. More recently, Zhang et al. (1999) compared neural networks to logistic regressions. Typically, when testing models, it is common to divide data into a training sample and a test sample. The first is used so the model can learn as opposed to the second, which is used to evaluate how well can the model predict. The authors considered that by doing the traditional percentage division on the data, they would be introducing bias in model selection and evaluation. In other words, the features of the test sample may be significantly different from those in the training sample. This process is thus not recommended for small samples. Instead, as an alternative, the authors decide to use a method called cross-validation, which consists of partitioning the data in several subsets. These subsets are then used to train and test the data in multiple ways. Two cross-validations methods were considered. The first method, which was named as "small test set" consisted of dividing the data into five equal parts, called folds, and make use of four of those to train the model. The testing is conducted on the remaining fold. Subsequently, the authors consecutively perform this task until every one of the five folds has been used as a testing sample. In the end, the authors do an average of the results obtained. The second method, called "large test set" consists of using the whole dataset to train and then test on each of the five cases. Six variables were selected, the five from Altman paper from 1968 plus the ratio of current assets to current liabilities. Their justification was grounded on the belief that this ratio has a direct influence on the probability of a firm entering a default situation. The results demonstrated clear superiority of neural networks in contrast to the logistic regression method. The accuracy of neural networks for the small test set and large test set was respectively 80.46 % and 86.64% compared to the 78.18% and 78.65% achieved by the logistic regression method.

Within the various machine learning techniques, decision trees learning algorithms are one of the most popular machine learning techniques mainly because of its intuitive understanding. Decision trees can be one of two types, classification or regression trees. The first is used when one is interested in predicting whose class a particular instance belongs. The second is used when the target variable, which one wants to predict, is a real number. Within this model framework, instances are categorized through a tree. The instances aimed to be evaluated, starting at the root node, will face, throughout the tree, at every node, a specific test regarding a particular attribute. Each branch resulting from a node outputs a possible value for that attribute. The final classification of a specific instance is accomplished when a leaf is reached with no more branches. Gepp et al. (2015) aimed at predicting financial distress using decision trees algorithm and logistic regression. The results demonstrated a clear superiority, regarding the accuracy metric, of the decision tree model over the logistic regression

Another widely known, although relatively recent, machine learning technique is support vector machines. This technique is mostly applied in binary classification tasks, and the main principle behind it is constructing a hyperplane that maximizes the distance between elements of different data classes. Min et al. (2005) applied this method to a sample of 1888 companies in which 944 went bankrupt. The authors employed a grid-search technique using 5-fold cross-validation in order to obtain the optimal parameters linked to a kernel function. Even though the authors recognized the existence of three main kernel functions, the decision was to employ the radial basis function based on previous studies that found that the latter provide overall better results. In order to make an out of sample analysis, the authors divided the data in a traditional 80:20, which means that they made use of 80 % of the data for training and the remaining 20% for testing. The support vector machine model proposed by the authors was then compared to other methodologies such as neural networks and logistic regression. The results revealed a clear superiority of the support vector machines method with an accuracy of 88.01% in the training sample and 83.07% in the test sample.

Chapter 3

3 Models

3.1 KMV-Merton Model

The KMV-Merton model emerged in the late 1980s, grounded on Merton´s (1974) seminal bond pricing model. The model rapidly became famous for its capacity to use forward-looking market information for default prediction. In this section, I will start by presenting Merton's model, and then I will explain what is different in Merton-KMV model.

Consider a firm with a pre-specified liquidation date T financed by equity and a single class of zero-coupon bond that must be paid at time T. Following Merton's (1974), one can recognize equity as a call option on the firm´s assets with strike price equal to the face value of debt. Now assume that the market value of the firm´s assets follows a stochastic process known as the geometric Brownian motion:

$$dV_A = \mu V_A dt + \sigma_A V_A dz \qquad (1),$$

where:

dV_A is the change in asset value,

μ, σ_A are respectively, the firm´s asset drift rate and asset volatility,

dz is a Wiener process.

Within this framework, the market value of the assets evolves stochastically due to a predictable component, the drift rate, as well as due to the occurrence of unexpected shocks, whose size is determined by the volatility term.

Solving equation (1), one can determine the asset value at maturity as:

$$lnV_A^T = ln\, V_A + (\mu - \frac{\sigma_A^2}{2})T + \sigma_A \sqrt{T}\varepsilon \qquad (2),$$

where ε is a standard normal random variable.

From the above equation, it is clear that the market value of assets is log-normally distributed.

In this model, the probability of default is the probability that the call option ends up out of the money. Hence, the probability of default can be written as:

$$Pr_{default} = \Pr(V_A^T \leq X_T \mid V_A^0 = V_A) = \Pr\,(lnV_A^T \leq lnX_T \mid V_A^0 = V_A) \qquad (3),$$

where X_T is the firm nominal debt.
Substituting equation (2) above on equation (3):

$$Pr_{default} = \Pr\,(ln\, V_A + (\mu - \frac{\sigma_A^2}{2})T + \sigma_A\sqrt{T} \leq X_t) \qquad (4).$$

Rearranging one obtains:

$$Pr_{default} = \Pr\,(-\frac{ln\frac{V_A}{X_t} + (\mu - \frac{\sigma_A^2}{2})T}{\sigma_A\sqrt{T}} \geq \varepsilon) \qquad (5),$$

where the symmetric of the term within brackets is usually called the distance to default. Since the error follows a Normal distribution, one can write:

$$Pr_{default} = \mathrm{N}\,(-\,Distance - to - default) \qquad (6).$$

Distance to default can be understood as the number of standard deviations the asset value is expected to be away from the default barrier at time T. For instance, considering that the asset value today equals 200, the drift rate is 20% and the asset volatility is 25% and T equals 1, the value of the asset at maturity can be obtained by:

$$V_A^T = V_A \times\, e^{(\mu - \frac{\sigma_A^2}{2})T} + \sigma_A\sqrt{T}\varepsilon \qquad (7),$$

Hence, in this specific scenario, the expected asset value at maturity equals 236.8. If one calculates the expected value of ln $\left(\frac{V_A^T}{X}\right)$, considering X =100, the value obtained equals 86.2%. The last step consists of standardizing. To do so, one has to divide by asset volatility. The result obtained is roughly equal to 3.4, meaning that the expected value of the assets at maturity is 3.4 standard deviations away from the default barrier.

In a practical implementation of the Merton model, one needs to estimate the market value of assets and the volatility of asset returns. Fortunately given the parallel with the Black and Scholes (1973) model, one can use its European call option formula to state equity value:

$$E = V_A N(d_1) - e^{-r(T-t)} X N(d_2) \qquad (8),$$

where

E is the market value of the equity,

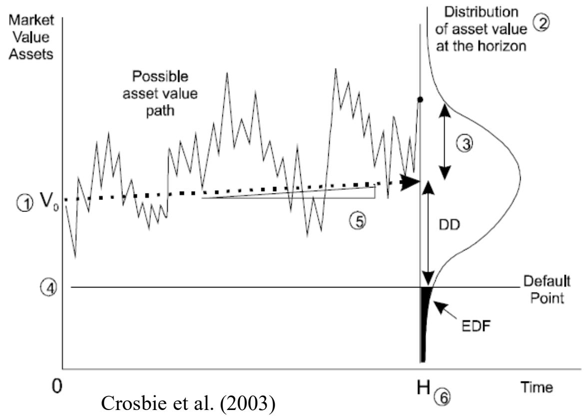
$d_1 = \frac{\ln\left(\frac{V_A}{X}\right) + (r + \frac{\sigma_A^2}{2})(T-t)}{\sigma\sqrt{T-t}}$ and,

$d_2 = d_1 - \sigma_A \sqrt{T-t}$.

The d_1 formula is very similar to the distance to default. However, instead of the expected asset return, μ, one has r, which represents the risk-free rate. For all traded firms for which equity value is known, this formula can be used to extract the implied market value of assets and return volatility.

KMV-Merton has two main advantages over Merton's model. First, Merton oversimplifies the capital structure of a company by considering that a company solely has a zero-coupon bond as liabilities. Hence, a company defaults if, at maturity, the value of the assets is below the nominal value of debt. Instead, KMV-Merton, considers that a company only defaults if the market value of the assets is below a certain barrier at maturity, known as the default point. Grounded on empirical studies, KMV found that companies usually default when their asset value at a pre-specified maturity lies between current liabilities and long-term liabilities. Second, KMV obtains a non-parametric relationship between the distance to default and the probability of defaulting by constructing their own distribution based on historical data. While in Merton's model, the distance to default is evaluated on the Normal distribution, in the KMV-Merton model it is evaluated on this proprietary distribution. As this proprietary distribution encompasses many historical downturns, it seems to incorporate scenarios of significant changes in a company´s market value. As result, differently from the Normal distribution used in Merton's model, the KMV distribution gives non-negligible probabilities of default for relatively high distances to default. In addition, the KMV distribution does not lead to abnormally high default probabilities when the distance to default is very low.

Figure 3.1 summarizes the six variables that are responsible for the determination of the probability of default over a specific horizon. Those are the current asset value (1), the distribution of the asset value at the horizon (2), the volatility of asset returns (3), the default point (4), the expected growth rate in the asset value (5) and finally the horizon itself (6). As one can observe, the probability of default is the probability that the asset value at maturity is below the default point, which is constant within this framework



Crosbie et al. (2003)

Figure 3-1 – Key variables in determining the probability of default

3.2 Support vector machines (SVM)

Support vector machines is a machine learning technique that falls into the category of supervised learning models. This means that it relies on labeled input data in order to learn a function that gives estimates of the output of an unlabeled data point. This method is quite flexible since it can be used for both regression and classification. For the sake of this paper, I will be concerned with the explanation of SVM regarding classification[2].

Figure 3.1 is useful to explain the main idea behind SVM (Géron (2019)). In this figure, the objective is to separate two classes, both flower species, Iris-Versicolor and Iris-Setosa. However, one can realize that the left graph decision boundaries are not sufficiently reliable for two reasons. First, the dashed green line does not even correctly separate the classes in the training set. Second, the purple and red lines, although properly separate the two classes, most probably will not predict new instances with the desirable accuracy. On the other hand, the right graph shows what SVM does. It identifies a hyperplane that maximizes the margin between the

[2] The spam filter is an example of it. The model learns from past emails and will classify new instances either as spam or no spam.

two classes. As one can perceive, support vectors, which are the data-points that lie closest to the decision surface, are used to define the hyperplane. Thus, all other points are not employed in order to define the boundary between the two classes. In other words, support vectors are the elements of the training data, which, if eliminated, would cause a change in the position of the SVM hyperplane.

Hence, SVM can be used to find the optimal hyperplane that separates the companies in our dataset into two groups, bankrupt and non-bankrupt.

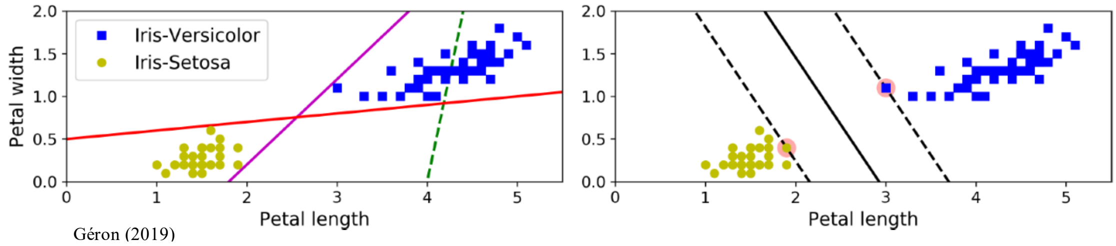


Figure 3-2 - Géron 2019 - example SVM

3.2.1 Linearly Separable Data - Binary Classification

According to Min et al. (2005), the SVM algorithm can be described along the following lines. Taking as input vectors $x_i = (x_i^1, \dots, x_i^n)^T$ and as target labels $y_i \in \{-1, +1\}$, one can formulate the support vector machine classifier as following:

$$\begin{cases} \boldsymbol{w}^{\boldsymbol{T}} \cdot x_i + b \geq 1, if \quad y_i = +1 \\ \boldsymbol{w}^{\boldsymbol{T}} \cdot x_i + b \leq 1, if \quad y_i = -1 \end{cases} \qquad (9),$$

where $\boldsymbol{w}$ represents the weight vector, and b represents the bias term. The dimension of $\boldsymbol{w}$ will be the same as the number of features used in order to classify our analyzed firms as bankrupt or non-bankrupt.

Alternatively, the SVM classifier can be presented as:

$$y_i[\boldsymbol{w}^{\boldsymbol{T}} \cdot x_i + b] \geq 1, \qquad i = 1, \dots, N \qquad (10),$$

When $[\boldsymbol{w}^T \cdot x_i + b]$ *is negative*, $\Rightarrow y_i$ *is negative* $\Rightarrow y_i[\boldsymbol{w}^T \cdot x_i + b]$ *is positive*.

The dashed lines in the right panel of Figure 0.3, which are constructed solely based on the support vectors, specifically satisfy the following formulation:

$$[\boldsymbol{w}^T \cdot x_i + b] = +1, \quad i = 1, \dots, N \Rightarrow Dashed\ line\ on\ the\ right \quad (11),$$
$$[\boldsymbol{w}^T \cdot x_i + b] = -1, \quad i = 1, \dots, N \Rightarrow Dashed\ line\ on\ the\ left \quad (12),$$

When applying the SVM, the weights are chosen in order to maximize the margin, which represents the distance between the two dashed lines. It can be demonstrated that the margin width is equal to:

$$width = (x^+ - x^-) \cdot \frac{\vec{w}}{\|w\|} = \frac{1-b-(-b-1)}{\|w\|} = \frac{2}{\|w\|} \quad (13),$$

In order to maximize the width, one has to minimize $\| w \|$. For mathematical convenience and equivalently one can simply minimize $\frac{1}{2} \| w \|^2$.

$$\min\left(\frac{1}{2} \| w \|^2\right) \quad (14),$$

Constrained to:

$$y_i[\boldsymbol{w}^{\boldsymbol{T}} \cdot x_i + b] = 1, \quad i = 1, \dots, N \quad (15),$$

In order to solve this minimization problem, one can employ the Lagrange multiplier method.

$$L_P = \frac{1}{2} \| w \|^2 - \sum_{i=1}^{L} \alpha_i \left[y_i(\boldsymbol{w}^{\boldsymbol{T}} \cdot x_i + b) - 1\right], \quad i = 1, \dots, N$$
$$= \frac{1}{2} \| w \|^2 - \sum_{i=1}^{L} \alpha_i y_i \left(\boldsymbol{w}^{\boldsymbol{T}} \cdot x_i + b\right) + \sum_{i=1}^{L} \alpha_i, \quad i = 1, \dots, N \quad (16),$$

The next step consists of finding the values of $\boldsymbol{w}$ and b, which minimize the above equation:

$$\frac{dL_P}{d\mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{L} \alpha_i y_i x_i, \quad i = 1, \dots, N \quad (17),$$

$$\frac{dL_P}{db} = 0 \Rightarrow \sum_{i=1}^{L} \alpha_i y_i = 0, \quad i = 1, \dots, N \qquad (18),$$

By substituting (17) and (18) in (16), one can write:

$$L_D = \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{L} \alpha_i \alpha_j y_i y_j x_i \cdot x_j,\ \alpha_i \geq 0\ \forall_i,\ \sum_{i=1}^{L} \alpha_i y_i = 0 \qquad (19),$$

L_D is known as the dual form of the primary form, L_P. In order to solve equation (22), one has to identify the vector α, which maximizes the abovementioned function. Thus, one can make use of a quadratic programming solver. Thereafter, once the value of α is obtained, one can calculate the $\boldsymbol{w}$**.** Ultimately, the bias term, b, can be computed as follows:

$$b = y_S - \sum_{m \in S} \alpha_m y_m x_m \cdot x_S \qquad (20),$$

where S is used to denote the indices of the support vectors.

It is now possible to classify a new instance of data. In order to do so, and given an unknown point, u, one has to determine the sign of the following equation:

$$\boldsymbol{w^T} \cdot u + b \Leftrightarrow \left(\sum_{i=1}^{L} \alpha_i y_i x_i \cdot u\right) + b,\ i = 1, \dots, N \qquad (21),$$

3.2.2 Soft Margin Classification – Binary Classification

In real-world situations, the majority of the problems endorsed by academics and practitioners do not have a linear structure. For the purpose of extending the SVM methodology to process data that is not fully linearly separable, the constraints imposed by equation (10) are relaxed. Hence, a new term, ε_i, is introduced in the equation. Another reason to introduce this new term besides the one abovementioned is the fact that the SVM technique is susceptible to be affected by outliers. This new formulation is known as soft margin classification since, within this scenario, the strict imposition of instances to be off the boundaries is disregarded. Therefore, one can present it as:

$$\begin{cases} \mathbf{w}^{\mathrm{T}} \cdot \mathrm{x_i} + \mathrm{b} \geq 1 - \varepsilon_{\mathrm{i}}, \text{ for } \mathrm{y_i} = +1 \\ \mathbf{w}^{\mathrm{T}} \cdot \mathrm{x_i} + \mathrm{b} \leq 1 + \varepsilon_{\mathrm{i}}, \text{ for } \mathrm{y_i} = -1 \\ \varepsilon_{\mathrm{i}} \geq 0 \end{cases} \quad (22),$$

where ε_i is a positive slack variable, which allows observations to end up on the wrong side of the margin.

Hence, the minimization problem can be stated as:

$$min\left(\frac{1}{2} \parallel \mathrm{w} \parallel^2\right) + \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{L}} \varepsilon_{\mathrm{i}} \quad (23),$$

s.t.

$$\begin{cases} \mathrm{y_i}[\mathbf{w}^{\mathrm{T}} \cdot \mathrm{x_i} + \mathrm{b}] = 1 - \varepsilon_{\mathrm{i}}, \quad \mathrm{i} = 1, \ldots, \mathrm{N} \\ \varepsilon_{\mathrm{i}} \geq 0, \quad \mathrm{i} = 1, \ldots, \mathrm{N} \end{cases} \quad (24),$$

The primal and dual form of the Lagrange multiplier can be written respectively as:

$$\mathrm{L_P} = \frac{1}{2} \parallel \mathrm{w} \parallel^2 + \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{L}} \varepsilon_{\mathrm{i}} - \sum_{\mathrm{i}=1}^{\mathrm{L}} \alpha_{\mathrm{i}} \mathrm{y_i} \left(\mathbf{w}^{\mathrm{T}} \cdot \mathrm{x_i} + \mathrm{b}\right) + \sum_{\mathrm{i}=1}^{\mathrm{L}} \alpha_{\mathrm{i}} \quad (25),$$

$$\mathrm{L_D} = \sum_{\mathrm{i}=1}^{\mathrm{L}} \alpha_{\mathrm{i}} - \frac{1}{2}\sum_{\mathrm{i,j}=1}^{\mathrm{L}} \alpha_{\mathrm{i}}\alpha_{\mathrm{j}}\mathrm{y_i}\mathrm{y_j}\mathrm{x_i} \cdot \mathrm{x_j},\ 0 \leq \alpha_{\mathrm{i}} \leq \mathrm{C}\ \forall_{\mathrm{i}}, \quad \sum_{\mathrm{i}=1}^{\mathrm{L}} \alpha_{\mathrm{i}}\mathrm{y_i} = 0 \quad (26),$$

The final result will be equal to (21) but now the α_i are determined by (26).

This new term C can be understood as a hyperparameter that aims to keep the margin as large as possible but, at the same time, limiting margin violations. A smaller value of C conduces to a wider margin, but more violations will be incurred. As opposed, a higher C will lead to fewer margin violations. Nonetheless, the margin will be smaller.

3.2.3 Non-Linearly Separable Data - Binary Classification

Linear SVM classifiers are usually referred to be very efficient. However, it occurs that many datasets are not linearly separable. For instance, the data presented in the left graph in Figure 3.2 is not linearly separable. Nonetheless, one can solve the problem by adding further dimensions where it is possible to separate the data linearly. The right graph achieves the goal by inserting a new feature, $x_2 = (x_1)^2$.

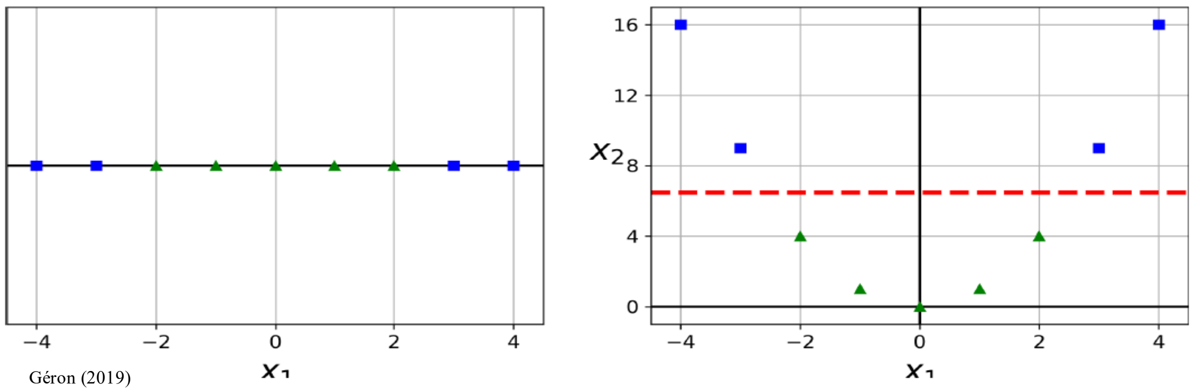


Figure 3-3 – Géron 2019 – Linearly separate non-linear data

Although adding polynomial features is a simple exercise, a very high polynomial degree will make the model slow. However, there is a mathematical methodology to solve this problem. This is known as the kernel trick. This method gives similar results as if one added many features but without actually having to add them. Thus, if one can´t define a hyperplane by linear equations, the data should be mapped into a higher dimensional space by making use of some nonlinear mapping function ϕ. Following Gerón (2019) example, if one wants to apply a 2nd-degree polynomial transformation to a two-dimensional set, it will follow as:

$$\phi(\mathbf{x}) = \phi\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} x_1^2 \\ \sqrt{2}x1x2 \\ x_2^2 \end{pmatrix} \quad (27),$$

This new vector is now in a three-dimensional space. In case one wants to apply this transformation to two-dimensional vectors and compute the dot product, this can be presented as:

$$\phi(\mathbf{a})^T\phi(\mathbf{b}) = \begin{pmatrix} a_1^2 \\ \sqrt{2}a1a2 \\ a_2^2 \end{pmatrix}^T \begin{pmatrix} b_1^2 \\ \sqrt{2}b1b2 \\ b_2^2 \end{pmatrix} = (\mathbf{a}^T\mathbf{b})^2 \quad (28),$$

A kernel is a function which is able to calculate the dot product $\phi(\boldsymbol{a})^T\phi(\boldsymbol{b})$ by using the original $\boldsymbol{a}$ and $\boldsymbol{b}$ vectors without needing to know the transformation ϕ. Géron (2019) states Mercer´s theorem which says that under several conditions known as Mercer´s conditions, there is a function ϕ that maps $\boldsymbol{a}$ and $\boldsymbol{b}$ into a different space such that $K(\boldsymbol{a},\boldsymbol{b}) = \phi(\boldsymbol{a})^T\phi(\boldsymbol{b})$. Consequently, even without knowing what is the ϕ one can make use of the kernel function.

Therefore, the new formulation is presented as:

$$y_i[\mathbf{w}^T \cdot \phi(\mathbf{x_i}) + b] = 1 - \varepsilon_i, \quad i = 1, \dots, N \qquad (29),$$

Moreover, the dual form of the Lagrange multiplier can now be written as:

$$L_D = \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{L} \alpha_i\alpha_j y_i y_j K(\mathbf{x_i}, \mathbf{x_j}), \quad 0 \le \alpha_i \le C \;, \sum_{i=1}^{L} \alpha_i y_i \qquad (30),$$

Same as before, for a new instance, one has to obtain the sign of:

$$\mathbf{w}^T \cdot K(\mathbf{x_i}, \mathbf{x_j}) + b, \quad \forall_i \;, \forall_j \qquad (31),$$

Several kernel functions may be used. According to Géron (2019), the most widely employed are the followings:

Linear: $K(\boldsymbol{a},\boldsymbol{b}) = \boldsymbol{a}^T\boldsymbol{b}$

Polynomial: $K(\boldsymbol{a},\boldsymbol{b}) = (\gamma\boldsymbol{a}^T\boldsymbol{b} + \boldsymbol{r})^d$

Radial basis: $K(\boldsymbol{a},\boldsymbol{b}) = e^{(-\gamma\|\boldsymbol{a}-\boldsymbol{b}\|^2)}$

Figure 3.3 is relevant in order to understand the gamma factor in the radial basis function kernel. As one can perceive, from all the figures, the top left figure represents a situation where the data is underfitting the most, meaning that the model may not be capturing the main trends. Therefore, one may want to increase the gamma value. Contrarily, if the model is overfitting, meaning that it may not generalize well for new instances, as in the bottom left figure, one should reduce its value. Therefore, gamma acts as a regularization parameter.

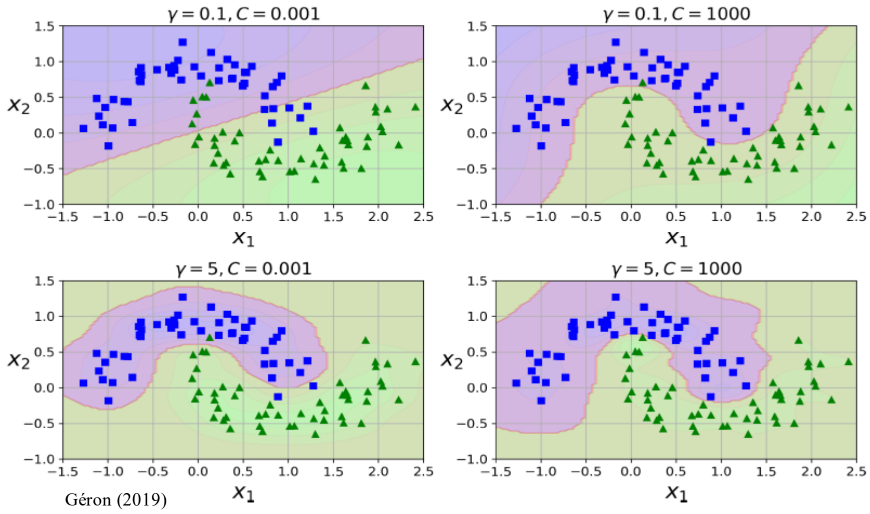


Figure 3-4 – Géron 2019 – Example of gamma and C factors

In summary, a low gamma will lead to less accuracy or, in other words, higher variance, although the results will be less biased. Therefore, choosing the optimum parameters is a vital task in this process. The relations abovementioned are summarized in Table 3.1.

	High Gamma	Low Gamma	High C	Low C
Variance	Low	High	High	Low
Bias	High	Low	Low	High

Table 3-1 – Summary table – C and gamma factors

Chapter 4

4 Data and Setup

4.1 Datasets

The main objective of this dissertation is to assess if the KMV-Merton main output, the distance to default, can be a valuable default predictor within the SVM model. In order to do so, two distinct datasets are constructed. The fact that the KMV-Merton Model can only be applied to listed companies restricted significantly this study as compared to a typical application of SVM, which relies only on accounting data. Hence, the first goal is to retrieve a data set that can be used with the purpose of applying the KMV-Merton model successfully.

The data was retrieved from CRSP, and CRSP/Compustat Merged, provided by Wharton Research Data Services, from 2000 to 2018. The considered databases were crucial to obtain both accounting and market data and to identify the companies which ultimately defaulted. Companies whose field "Research Company Reason for Deletion" presented the values 2 (Bankruptcy) or 3 (Liquidation) were considered as bankrupt in the year after the last available accounting information. Moreover, only companies present in at least two years were considered so that one could predict the bankruptcy event one and two years prior. Additionally, only companies that had complete data regarding short and long-term liabilities were considered. Regarding the field "Global Industry Code", all the companies whose code was equal to 40 (i.e. "Financials"), were eliminated. Lastly, for the sake of uniformity, only companies whose accountability data was referred to the last day of the year were considered. The final number of bankrupt companies was 124, contrastingly to the 3053 non-bankrupt companies. This is a highly imbalanced panel for two reasons. [3] First, the number of non-bankrupt firms is significantly higher than the number of bankrupt firms. Second, even if it was the same, the number of non-bankrupt instances would be significantly higher than the number of bankrupt instances because in most time moments bankrupt firms appear as non-bankrupt. In what concerns bankrupt companies, for the majority of the analysis, solely the data regarding

[3] Imagining the data contained 3053 non-bankrupt companies and 124 bankrupt firms. This would mean that it would have 3053 multiplied by 18 years (54954) instances of non-bankrupt companies and only 124 instances of bankrupt companies.

the year prior default or the two-years prior default will be used. In furtherance of proportionate a better, however, far from perfect balance in terms of bankrupt and non-bankrupt instances, 124 companies were randomly selected from the 3053 non-bankrupt companies. In the end, 248 companies were considered.

As one can observe in Figure 4.1, the years of 2002, 2008, and 2009 were the years with more bankruptcies. Though the recent financial crisis had severe consequences in global economic activity, which led to a significant increase in the number of corporate defaults in 2008 and 2009, it is somehow surprising that the overall number of defaults in the sample is not very different from the one in 2002.

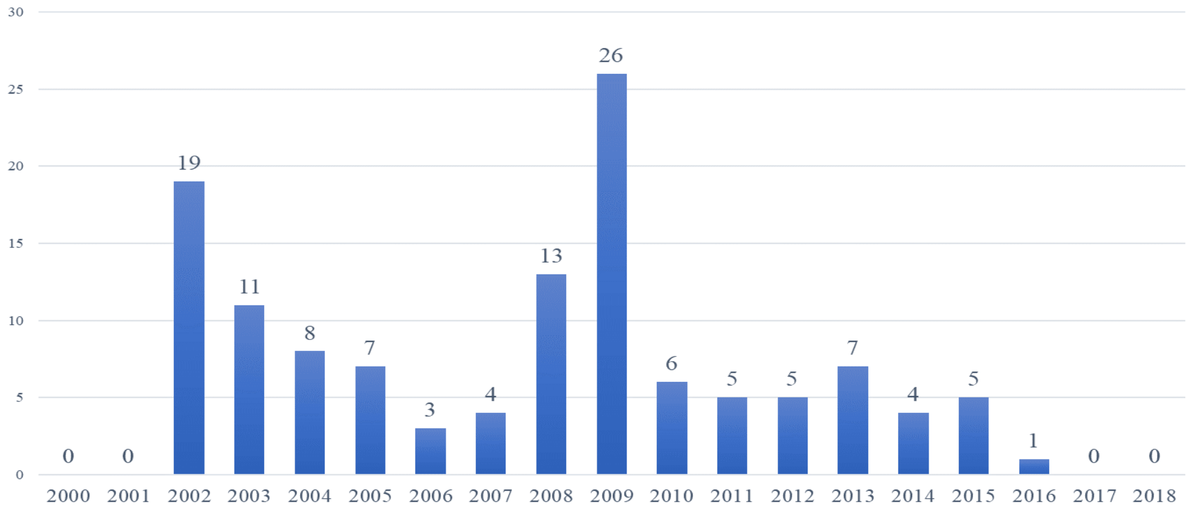


Figure 4-1- Number of bankruptcies along the years

The three most common sectors in the considered dataset are: “Health Care”, “Information and Technology”, and “Industrials” (Figure 4.2). Furthermore, one can notice that the three sectors displaying the highest number of bankruptcies are : “Consumer Discretionary”, “Health Care” and “Information and Technology”.

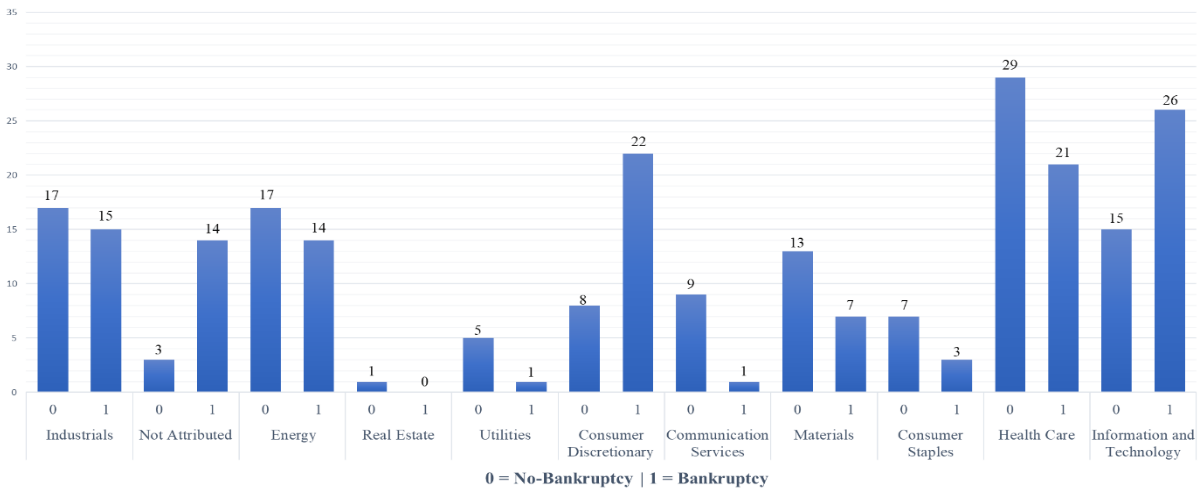


Figure 4-2 – Number of bankrupt and non-bankrupt companies by sectors (Global Industry Classification Sector)

Concerning the application of SVM, a second dataset was constructed with the same companies and timeframe as the one used for the Merton Model. Following Min et al. (2005) rationale, several financial ratios are going to be used in order to predict the bankruptcy of a company. Whenever a ratio could not be computed for a particular company, its value was considered to be equal to the sector average. The ratios selected were based on Gissel et al. (2007). The authors reviewed the vast majority of bankruptcy prediction literature and provided a list of 42 ratios that were considered in five or more of the studies. From these, eleven ratios were selected. In order to obtain those, the needed variables were collected from CRSP/Compustat Merged dataset, and afterward, the ratios were calculated as in Table 4.1. For the sake of inferencing, whether the distance to default from KMV-Merton can add value *vis-à-vis* a simple market-related ratio, the debt-to-market equity was added to the dataset.

Variables	**Type**
Current Assets / Total Assets (1)	Liquidity
Current Assets / Current Liabilities (2)	Liquidity
Cash / Total Asset (3)	Liquidity
Working Capital / Total Assets (4)	Liquidity
Ln (Total Assets) (5)	Size
Debt / Total Assets (6)	Leverage
Debt[4]/Market Equity (7)	Leverage
Sales / Total Assets (8)	Efficiency
EBIT / Total Assets (9)	Profitability
Earnings [5]/ Total Assets (10)	Profitability
Earnings / Stockholder Equity (11)	Profitability
Retained Earnings / Total Assets (12)	Profitability

Table 4-1 – Selected variables - SVM model

From the primary dataset, three sub-datasets were constructed. The first one, named df1, considers companies that defaulted in one-year time. The second dataset, df2, considers companies that defaulted in a two-years horizon. The third dataset, df3, contains the data regarding the companies which did not default.

[4] Total Debt
[5] Income Before Extraordinary Items

Figures 4.3-4.5 provide summary statistics regarding the ratios which are going to be employed in the SVM model. The variables (9), (10), (11), and (12), all profitability indicators, display lower average values for companies who went bankrupt as opposed to non-bankrupt companies. Regarding variable (7), a leverage indicator, bankrupt companies exhibit a higher average value and lower standard deviation. Concerning the ratio sales-to-total assets (8), non-bankrupt companies demonstrate higher average value. In what concerns the liquidity ratios, unexpectedly, bankrupt companies display higher average values in variables (1) and (2). This suggests that risky firms, knowing that they may not be able to borrow when faced by a negative shock, prefer to hold more liquid assets than more solid firms. As one would expect, bankrupt companies' profitability indicators, a year prior to default, display average lower values when compared to the dataset which considers all the instances of bankrupt companies, Figure 4.4. Although liquidity indicators variables (1), (2) and (3) exhibit higher average values for bankrupt companies a year prior to default comparatively to those presented in Figure 4.4, the percentile 25 for every single liquidity indicator is lower. Regarding leverage indicators, bankrupt companies a year prior default demonstrate higher average values when compared to the dataset which considers all instances of the bankrupt companies.

	Variables for non-bankrupt companies					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Average	0.44	3.13	0.16	0.25	6.22	0.24
Standard deviation	0.28	3.30	0.20	0.30	2.37	0.27
Skewness	0.24	4.39	1.93	-2.51	0.05	3.21
Excess Kurtosis	-1.05	33.89	3.64	41.95	-0.66	28.08
Jarque-Bera	76	70173	1609	102187	25	47482
Percentile 25	0.17	1.33	0.02	0.04	4.51	0.00
Median	0.42	2.31	0.08	0.21	6.26	0.19
Percentile 75	0.66	3.65	0.23	0.43	7.83	0.37
	(7)	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
Average	0.77	0.93	0.10	0.04	0.46	-1.19
Standard deviation	3.59	1.34	1.22	1.26	7.00	4.40
Skewness	16.15	6.30	9.04	8.02	3.78	-5.89
Excess Kurtosis	334.80	59.22	102.76	93.09	97.25	44.89
Jarque-Bera	6476937	209882	623233	510837	544702	123339
Percentile 25	0.00	0.29	-0.03	-0.06	-0.09	-0.59
Median	0.13	0.61	0.06	0.03	0.07	0.00
Percentile 75	0.53	1.10	0.12	0.08	0.16	0.26

Figure 4-3 – Summary statistics – Non-bankrupt companies

	Variables for bankrupt companies					
	(1)	(2)	(3)	(4)	(5)	(6)
Average	0.47	5.38	0.16	0.23	5.17	0.31
Standard deviation	0.30	23.06	0.21	0.41	1.59	0.51
Skewness	0.24	20.71	1.72	-2.43	-0.07	7.44
Excess Kurtosis	-1.13	487.54	2.32	23.07	-0.04	94.26
Jarque-Bera	42	6723517	484	15611	1	255727
Percentile 25	0.21	1.02	0.02	0.00	4.18	0.01
Median	0.44	1.79	0.07	0.16	5.15	0.17
Percentile 75	0.71	3.98	0.22	0.46	6.32	0.46
	(7)	(8)	(9)	(10)	(11)	(12)
Average	4.12	0.81	-0.12	-0.20	0.19	-1.83
Standard deviation	29.56	0.81	0.47	0.57	7.89	6.92
Skewness	14.10	1.82	-4.24	-4.01	22.22	-11.24
Excess Kurtosis	222.26	5.31	41.63	29.51	543.06	172.43
Jarque-Bera	1409613	1164	50685	26259	8337496	849123
Percentile 25	0.01	0.20	-0.17	-0.27	-0.40	-1.07
Median	0.20	0.59	-0.01	-0.06	-0.06	-0.27
Percentile 75	1.18	1.20	0.06	0.02	0.11	0.03

Figure 4-4 – Summary statistics – bankrupt companies

	Variables for bankrupt companies (Year prior bankruptcy)					
	(1)	(2)	(3)	(4)	(5)	(6)
Average	0.49	10.03	0.17	0.11	4.87	0.47
Standard deviation	0.33	51.33	0.24	0.66	1.73	0.90
Skewness	0.19	10.08	1.70	-2.76	-0.10	6.00
Excess Kurtosis	-1.35	106.91	1.96	14.57	-0.02	46.45
Jarque-Bera	10	61149	80	1255	0	11891
Percentile 25	0.20	0.81	0.01	-0.06	3.74	0.00
Median	0.44	1.47	0.07	0.11	4.85	0.25
Percentile 75	0.82	3.63	0.21	0.43	6.05	0.58
	(7)	(8)	(9)	(10)	(11)	(12)
Average	16.03	0.79	-0.30	-0.47	1.54	-4.59
Standard deviation	67.00	0.95	0.64	0.82	18.09	14.13
Skewness	6.15	2.38	-4.57	-3.75	9.99	-6.35
Excess Kurtosis	40.81	8.29	30.92	18.09	106.26	48.03
Jarque-Bera	9389	473	5372	1982	60398	12754
Percentile 25	0.00	0.11	-0.35	-0.64	-1.09	-3.34
Median	0.73	0.45	-0.07	-0.19	-0.12	-0.65
Percentile 75	3.83	1.26	0.01	-0.02	0.11	0.00

Figure 4-5– Summary statistics – bankrupt companies a year prior default

4.2 Data tools

With the purpose of applying the support vector machines technique, and to facilitate the calculation of the distance to default of the 248 companies, Python, which is a programming language, will be utilized. Python offers several libraries that provide built-in functions to ease the work of data scientists.

In this paper, the main libraries which are going to be employed are:

- Imblearn
- Matplotlib
- Numpy
- Pandas
- Scipy
- Sklearn

Chapter 5

5 Model Estimation

5.1 KMV- Merton Model

In order to calculate the distance to default, there are two unknowns that must be calculated: the asset value and asset volatility. Several approaches have been proposed in the literature. In this dissertation, I follow Loeffler et al. (2011) where these unknowns are calibrated by exploiting an iterative approach. Compared to the system of equations method, this approach has been shown to lead to more stable results and to be more consistent with the model assumptions as asset volatility is kept constant within every year.
I will start by writing the Black-Scholes-Merton formula in a slightly different way:

$$V_A^T = \frac{V_E{}^T + e^{-rT}X^{KMV}N(d_2)}{N(d_1)} \quad (32),$$

where X^{KMV}= Total Liabilities of the firm.

The objective of the method is to compute a time series of V_A^T and a single asset volatility parameter based on a time-series of equity values. The formula above will be used with that purpose for all the trading days of the past year. This means that roughly 260 equations will be obtained. The method works as follows. First, one calculates the market value of assets using equation (32) and assuming a reasonable starting value for asset volatility.[6] Once a time-series of V_A^T is obtained one can compute asset volatility as the standard deviation of the logarithmic asset returns. This asset volatility is then used in a second iteration in order to obtain a new time-series for V_A^T. This process will continue successively until the procedure converges. The process stops whenever the sum of squared differences between consecutive asset values fall below 10^{-4} .

[6] This method is considered to work well for reasonable asset volatility proxies. However, Loeffler et al. (2011) suggest to compute the asset volatility as the standard deviation of the logarithmic asset returns, with assets computed as the sum of the market value of equity and book value of the total liabilities.

In the classic Merton model, the default point is just corporate nominal liabilities. In KMV-Merton model, however, one has to define what is the default point. According to Crosbie et al. (2003), the default point will generally lie somewhere in between short-term liabilities and total liabilities. As a widely used approach, it will be considered to be as:

$$Default\ Point = current\ liabilities + \frac{1}{2}\ long\ term\ liabilities\ .$$

The drift rate, μ, in the physical distance to default formula, will be calculated in the same way Loeffler (2011) did:

$$Drift\ Rate = ln\left(1 + \left(risk - free + Beta \times (Excess\ Market\ Return)\right)\right),$$

where the annual risk-free (US treasury 10-year bond rate) and the excess market return were obtained from the Damoranan website [7] while the asset´s betas were calculated by regressing the excess return of the asset value on the excess return of the S&P 500.

5.2 Support Vector Machine

With the purpose of evaluating the SVM model regarding the prediction of bankruptcy events, two different out-of-sample evaluation methods will be employed. The first method, which for future reference, is labeled as "percentage ratio split", consists of dividing the dataset, accordingly to a percentage ratio, into a training sample and a test sample. The latter is used to evaluate the model, which was created using the first sample. Therefore, as in Min et al. (2005), the data will be split in an 80:20 ratio, which means 80% of the data will be used to train, and the remaining 20% will be utilized for testing. The second considered approach is known as stratified K-fold cross-validation and it is referred hereafter as "cross validation". This method consists of dividing the data in K folds of equal size. However, it creates these by preserving the percentage of samples for each class: bankrupt and non-bankrupt. Then, one successively tests each subset using the remaining folds as the training environment. Figure 5.1 exemplifies a 5-Fold cross-validation.

[7] http://pages.stern.nyu.edu/~adamodar/

1st Estimation	Train subset				Test subset
2nd Estimation	Train subset			Test subset	Train subset
3rd Estimation	Train subset		Test subset	Train subset	
4th Estimation	Train subset	Test subset	Train subset		
5th Estimation	Test subset	Train subset			

Figure 5-1 – 5-Fold cross-validation - example

An identified problem in the used dataset is concerned with the imbalancement of the data, meaning that the number of instances regarding a bankrupt incident is lower when compared with the opposite event. Therefore, the SVM model will struggle to learn, in both out-of-sample methods employed, since it is not being provided with enough relevant data.

Chawla et al. (2002) proposed a methodology called Synthetic Minority Over-sampling (SMOTE) that helps tackle this problem. SMOTE is an approach in which the minority class is oversampled by generating synthetic examples. For each minority sample, one can create new synthetic examples at some distance from them but towards one of their neighbors. This methodology makes use of another supervised machine learning classification technique known as K-nearest neighbors. Imagine one wants to classify a given point in one of two classes. The method works in two steps. First, the K-nearest points to the one mentioned are obtained. Second, one assesses the majority of votes of its K-neighbors. For the sake of finding the K-nearest points, one can calculate the Euclidian distance between the point of interest and all the others.

Another common requirement when working with machine learning models is features standardization. This is important in the sense that some variables may have different magnitudes. Consequently, the model will wrongly give more importance to those. Therefore, the variables are standardized by subtracting the mean and diving by the standard deviation.

As explained in chapter 3, several kernels may be utilized when applying SVM. However, similar to Min et al. (2005), in this dissertation, solely the radial basis function will be employed. The relation is expected not to be linear, so the linear kernel was disregarded. The polynomial kernel was not considered because it is usually referred as more time consuming without leading to better results. Also, variables C and γ may be optimized. To do so, and as in Min et al. (2005), one can apply a technique called grid-search, which makes use of the K-fold cross-validation method previously explained. Hence, several possibilities of the vector $(C\ ,\gamma)$ will be tested, and the one which achieves higher accuracy in the cross-validation will be selected.

In summary, one will follow the subsequent steps for both percentage ratio split and stratified K-fold cross-validation methods. First, SMOTE is applied in order to attain a better data balance. Second, the variables are then standardized following the procedure above explained. Third, the vector (C , γ) is optimized by applying the grid-search technique. Finally, and after the preprocessing of the data, SVM is employed.

5.3 Metrics

Table 5.1 represents a confusion matrix, which is frequently used to describe the performance of a given classifier. Whenever one predicts that a company goes bankrupt (1) and the company actually goes bankrupt (1), one labels it as a "True Positive" (TP). In case one predicts that the company will not go bankrupt (0), but it actually goes bankrupt (1), it is called a "False Negative" (FN). In the event, a company is predicted to go bankrupt (1), but it doesn´t (0), one labels it as a "False Positive" (FP). The last scenario regards a case where one forecasts that a company will not go bankrupt (0) and in fact, it does not (0). This is labeled as a "True Negative" (TN).

		Actual Values	
Prediction		Positive (1)	Negative (0)
	Positive (1)	True Positive (TP)	False Positive (FP)
	Negative (0)	False Negative (FN)	True Negative (TN)

Table 5-1 – Confusion matrix - example

Several performance metrics can be computed based on this matrix. These are now presented.

Accuracy, shown in equation (33), is a metric that reveals how often a classifier is correct.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{FN} + \text{TN})} \qquad (33).$$

Accuracy is a very important indicator. However, one should not merely consider this indicator, and the reason can be easily illustrated. For instance, in a case where the dataset is imbalanced, meaning that it has substantially more instances of one class compared to the other if one predicts that the instances solely belong to the majority class, then a very high accuracy is accomplished. However, the model never predicts well the minority class.

Misclassification Rate, represented in equation (34) tells how often a classifier makes the wrong decision. Nevertheless, it shares the same mentioned weakness identified in the accuracy metric.

$$\text{Misclassification Rate} = \frac{(\text{FP} + \text{FN})}{(\text{TP} + \text{FP} + \text{FN} + \text{TN})} = 1 - \text{Accuracy} \qquad (34).$$

The true positive rate, also called sensivity or recall, tells how often one predicts that the event is positive among all the positive events, and in fact, it is. This is presented in equation (35).

$$\text{True Positive Rate} \mid \text{Sensivity} \mid \text{Recall} = \frac{\text{TP}}{(\text{FN} + \text{TP})} \qquad (35).$$

The false positive rate, exhibited in equation (36), measures among the negative instances, the percentage incorrectly classified as positive.

$$\text{False Positive Rate} = \frac{\text{FP}}{(\text{TN} + \text{FP})} \qquad (36),$$

The true negative rate, also called specificity, shown in equation (37), enlights one regarding the percentage of cases where one predicts that the event is negative from all the negative instances:

$$\text{True Negative Rate} \mid \text{Specificity} = \frac{\text{TN}}{(\text{FP} + \text{TN})} \qquad (37).$$

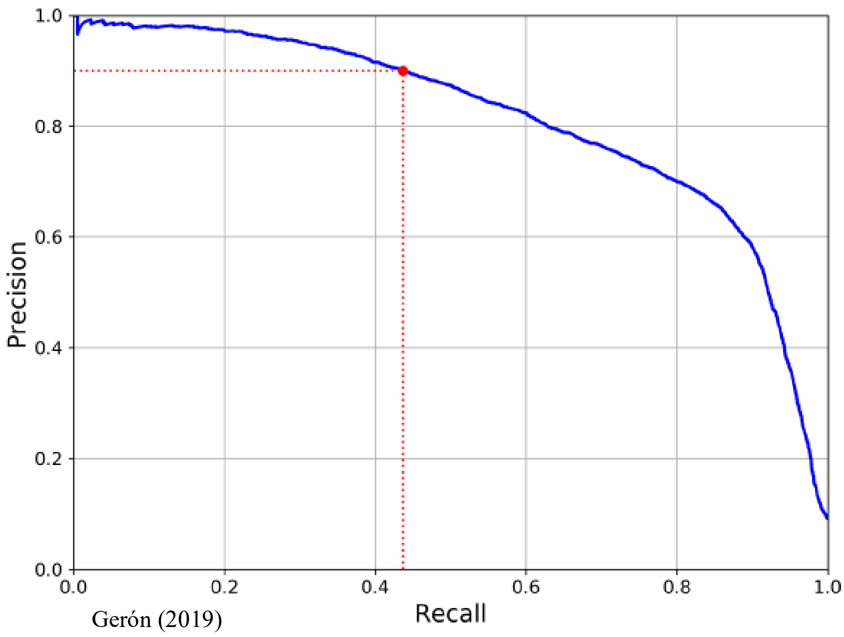
Precision, as in equation (38), measures among all the predicted positive records, those which are in fact, positive:

$$\text{Precision} = \frac{\text{TP}}{(\text{FP} + \text{TP})} \quad (38).$$

The F-score, shown in equation (39), measures the model overall accuracy. It calculates the harmonic mean of precision and recall, giving them the same weight. Hence, it takes into consideration both false positives and false negatives.

$$\text{F} - \text{score} = 2 \times \frac{precision \times recall}{\text{precision} + \text{recall}} \quad (39).$$

Last but not least, the Precision-Recall (PR) curve, exemplified in Figure 5.2, assesses the performance of the employed model as the threshold changes by mapping the tradeoff between precision and recall. In cases where data is highly imbalanced, PR curves are preferable when compared with other binary classification tools (Boyd et al. (2013)). A common practice relies on computing the area under the curve (PR AUC) and use it as a performance measure[8].



Gerón (2019)

Figure 5-2 – Precision-Recall curve - example

[8] The most acknowledge diagnostic tool in binary classification is the Receiver Operating Characteristic (ROC) curve, which plots the true positive rates and the false positive rate. However, it is not the most adequate measure when dealing with highly imbalanced datasets.

Chapter 6

6 Results

6.1 KMV-Merton model

Regarding the obtained variables from the KMV-Merton model, Table 6.1, summarizes the evolution of the median values throughout the years. As one would expect, the years of 2008 and 2009 demonstrated a decrease in the one-year median distance to default comparatively to the previous years. This is certainly a consequence of the subprime financial crisis. These years were characterized by high asset volatility. Also, the years of 2001 and 2002 displayed a low one-year median distance to default. These years are encompassed in a period acknowledged as the "dot-com crash", which is known for the plummet of the stock prices as part of a correction that began since the beginning of the millennium. Similarly, this period portrayed high asset volatility.

	Median					
	One-year distance to default	**Total Liabilities/Total Assets**	**Drift rate**	**Asset volatility**	**Equity Volatility**	**Bankruptcies**
2000	2.67	48%	0.05	43%	82%	0
2001	3.03	51%	0.05	51%	84%	0
2002	3.36	51%	0.04	39%	72%	19
2003	5.17	48%	0.05	31%	55%	11
2004	5.68	43%	0.05	32%	47%	8
2005	6.37	52%	0.05	28%	42%	7
2006	6.05	50%	0.05	28%	42%	3
2007	5.51	46%	0.05	31%	46%	4
2008	2.48	46%	0.05	48%	85%	13
2009	3.58	49%	0.05	39%	72%	26
2010	5.67	47%	0.05	27%	47%	6
2011	5.34	51%	0.05	32%	53%	5
2012	6.19	50%	0.03	27%	46%	5
2013	6.71	43%	0.04	26%	38%	7
2014	6.25	47%	0.03	27%	43%	4
2015	6.47	48%	0.03	31%	41%	5
2016	5.61	50%	0.03	28%	49%	1
2017	5.55	50%	0.03	25%	39%	0
2018	4.96	53%	0.04	28%	47%	0

Table 6-1 – KMV Merton model results by year

Furthermore, and by analyzing Table 6.2, one can observe that "Health Care", "Consumer Discretionary", and "Information and technology" sectors displayed a lower one-year median distance to default, higher equity volatility, and were the sectors with most bankruptcies.

	Median					
	One-year distance to default	Total Liabilities/Total Assets	Drift rate	Asset volatility	Equity Volatility	Bankruptcies
Industrials	4.92	58%	0.05	25%	51%	15
Information and Technology	3.24	38%	0.05	44%	63%	26
Communication Services	5.19	59%	0.05	24%	53%	1
Real Estate	11.62	47%	0.06	17%	22%	0
Health Care	3.89	29%	0.04	56%	65%	21
Energy	5.75	59%	0.04	28%	49%	14
Consumer Staples	4.91	43%	0.04	34%	52%	3
Consumer Discretionary	4.34	52%	0.05	29%	54%	22
Materials	6.65	49%	0.04	23%	47%	7
Utilities	13.81	67%	0.07	7%	24%	1
Not attributed	18.53	4%	0.03	15%	34%	14

Table 6-2 – KMV Merton model results by sector

Moreover, one can observe from Figure 6.1 that the one-year median distance to default varies along the years in some sectors more than others. The sectors with higher standard deviation through the years are the "Real Estate", "Utilities" and "Communication Services". Also, one can observe a vigorous fall of the one-year median distance to default in the years of 2008 and 2009 when compared with the previous years.

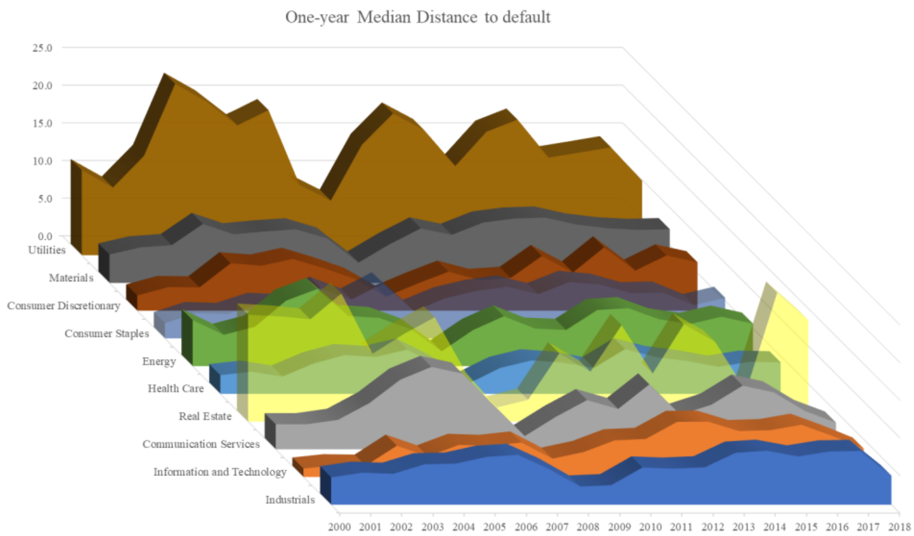


Figure 6-1 – One-year median distance to default by year and by sector

Table 6.3 presents the main outputs from the KMV Merton model by dataset. Based on this table one can observe that the set of companies one-year prior to the event, df1, displayed a lower one-year median distance to default, higher median equity volatility, and were expressively more levered comparatively to the df2 dataset. Additionally, and as expected, one can notice that df3, when compared with the other two datasets, presents lower median asset volatility, higher one-year median distance to default, lower median leverage ratio, and lower median equity volatility.

	Median				
	Asset volatility	Equity volatility	Drift rate	One-year distance to default	Total Liabilities/Total Assets
df1	38.9%	91.1%	0.04	1.93	69.6%
df2	39.7%	76.8%	0.05	2.88	49.5%
df3	29.9%	46.4%	0.04	5.70	48.6%

Table 6-3 – KMV Merton Model results for df1, df2 and df3

6.2 Support Vector Machines

This section presents all results from the application of SVM to my dataset. The variables considered are all the ratios presented in Table 4.1, including the distance to default obtained from the KMV-Merton model. Four different cases are considered. The first and second cases are concerned with the application of the SVM model using the percentage ratio split while the third and the fourth are concerned with the application of the stratified K-fold cross-validation technique. The first differs from the second, and the third differs from the fourth in the sense that the first and the third cases make use of instances of the companies a year prior to default while the second and fourth cases use the instances of companies two-years prior default. Each of the considered cases incorporates the same sample of the non-bankrupt companies data, df3. Within each of these cases, my main question is to understand whether the distance to default is able to add value as compared to a simpler market-related variable such as debt-to-market equity. In order to answer this question, one has to choose a performance measure among the ones presented in section 5.3. It is relatively obvious that entering a deal with a counter-party that will default is worse than not entering a deal with a healthy counter-party. However, this rationale will lead banks to give no credit. Although attaining good accuracy is imperative, as the datasets considered are imbalanced, one should also focus the attention on other metrics. Both recall and precision seem to be especially important as one wants to achieve a high number of true positives. However, if the model always predicts that a company will fail, it will display a high recall. This occurs because recall is concerned with the percentage of bankrupt events that were correctly identified from all the actual cases. However, banks that use this model would never lend money. Therefore precision, which evaluates the percentage of correctly identified bankruptcies among all the predicted ones, should also be considered. F1-Score takes both recall and precision into consideration. Hence, F1-Score is going to be assumed as the core metric in the subsequent analysis as it considers both precision and recall.

6.2.1 Percentage Ratio Split

Case 1:

In this first case, df1 and df3 are concatenated. This new dataset contains 1508 instances, 124 from df1 and 1384 from df3. As the dataset was split following an 80:20 ratio, the test environment contains only 302 instances.

The following parameters are employed in the grid search methodology:

$$(Kernel, C, \gamma) = (RBF^{9}, 10{,}1)$$

Table 6.4 presents the obtained confusion matrix and Table 6.5 presents the main measures presented in section 5.3.

	Predict No Bankruptcy	**Predict Bankruptcy**
Actual No Bankruptcy	262	15
Actual Bankruptcy	16	9

Table 6-4 - Confusion Matrix - All variables – Case 1

As one can observe from Table 6.5, the model presented an accuracy of nearly 90%. However, precision and recall metrics are not as attractive as one would want.

Metrics	
Accuracy	89.70%
Precision	37.50%
Recall	36.00%
Misclassification Rate	10.30%
F1 Score	36.70%

Table 6-5 – Metrics – All variables – Case 1

[9] Radial Basis Function

The area under the PR curve (Figure 6.2) displays a value of 0.42.

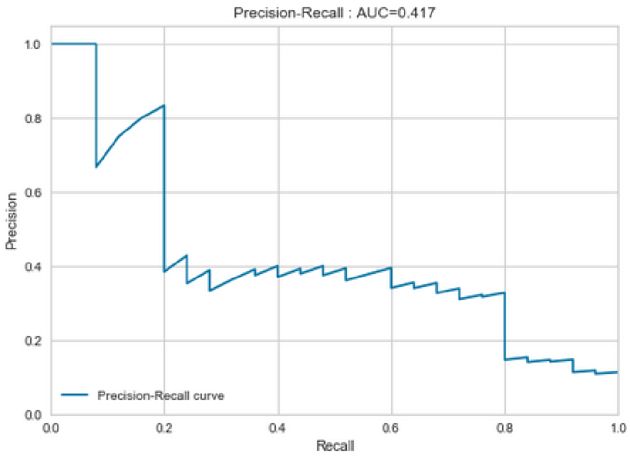


Figure 6-2 – PR Curve – All variables – Case 1

Table 6.6 presents the results obtained in different scenarios by changing the variables employed. Hereafter the tables are assumed to be read from left to right and starting in "Panel A" to ending in "Panel B". For instance, if one refers to the first column, one means "Panel A" first column. As well as if one refers to the last column, it is referring to "Panel B" last column. Table 6.6 first column displays a scenario where all the variables are utilized to predict bankruptcy (baseline model), similarly to what was presented above. The following two columns represent a situation where only a single variable is used (either debt/market equity or the distance to default). The remaining columns were constructed by removing the variable indicated from the baseline model.

From the abovementioned table one can take two important conclusions. First, market-related variables clearly add value. In the scenario where both market-related variables, distance to default, and debt-to-market equity are removed, the model seems to underperform, regarding all the considered metrics, compared to both the scenarios where individually the variables are excluded. Second, based on the two columns before the last, one can conclude that the distance to default adds more value to the baseline model than the debt-to-market equity. In the scenario where the distance-to-default is removed from the model, the F1-score decreases 3 percentual points *vis-à-vis* the baseline model. In addition, recall decreases by 4 percentual points, precision falls by almost 1.2 percentual points and the PR AUC falls by 3 percentual points. This contrasts with the one obtained when debt-to-market equity is excluded. In this case, the F1-score actually increased 4 percentual points, while the PR AUC persisted roughly the same.

This suggests that the debt-to-market equity is not a close substitute of the distance to default and that the model itself adds value. In line with this result, the second and third columns also show that the distance to default *per se* leads to an F1-score significantly higher than the debt-to-market equity ratio.

Overall the best-performing model, both in terms of the F1-score and the PR AUC, is the one obtained when liquidity variables are removed. This might be explained as 4 liquidity indicators are being employed, and probably the marginal contribution of each one of these is just creating “noise”. Additionally, as seen in chapter 4, companies that default in a year-time display higher values in what concerns liquidity ratios. In this particular case, the best scenario, following the initial assumption, is when working capital to total assets is removed as the F1-score presents the highest value.

				All variables minus				
Panel A	All variables	Debt/Market Equity	Distance to Default	Current Assets / Total Assets	Current Assets / Current Liabilities	EBIT / Total Assets	Working Capital / Total Assets	Debt / Total Assets
Accuracy	90%	85%	77%	91%	91%	89%	91%	90%
Recall	36%	36%	68%	40%	44%	32%	52%	40%
Precision	38%	24%	22%	48%	44%	35%	48%	40%
F1-Score	37%	29%	33%	43%	44%	33%	50%	40%
PR AUC	42%	30%	24%	46%	40%	42%	48%	44%
Kernel	RBF	RBF	RBF	RBF	RBF	RBF	RBF	RBF
C	10	100	1	100	10	100	1 000	100
gamma	1	1	10	1	1	1	1	1

	All variables minus								
Panel B	Cash / Total Asset	Sales / Total Assets	Ln (Total Assets)	Retained Earnings / Total Assets	Earnings / Stockholder Equity	Earnings / Total Assets	Debt/Market Equity	Distance to Default	Distance to Default & Debt/Market Equity
Accuracy	91%	90%	89%	90%	90%	90%	90%	90%	89%
Recall	40%	40%	44%	32%	40%	36%	40%	32%	24%
Precision	43%	40%	35%	36%	40%	39%	42%	36%	29%
F1-Score	42%	40%	39%	34%	40%	37%	41%	34%	26%
PR AUC	43%	43%	34%	35%	42%	44%	42%	39%	32%
Kernel	RBF	RBF	RBF	RBF	RBF	RBF	RBF	RBF	RBF
C	10	100	10	10	100	100	100	10	1000
gamma	1	1	1	1	1	1	1	1	1

Table 6-6 – SVM results – Case 1

Case 2:

This case represents the situation where df2 is concatenated with df3. In other words, the data employed in this case regards those of non-bankrupt companies and the instances of companies two years prior to the default. The data similarly to case 1, was divided following an 80:20 ratio. The grid search methodology outputted the following vector:

$$(Kernel, C, \gamma) = (RBF, 100,1)$$

The confusion matrix, metrics, and PR curve are presented, respectively, in Figures 6.3, 6.4, and 6.5 The baseline model exhibits a high accuracy, but not so good precision and recall. As one can observe from the confusion matrix, from the 25 bankruptcy events presented in the test data, only 10 were correctly identified. In this case, the PR AUC equals approximately 0.4.

	Predict No Bankruptcy	**Predict Bankruptcy**
Actual No Bankruptcy	265	12
Actual Bankruptcy	15	10

Figure 6-3 - Confusion Matrix - All variables – Case 2

Metrics	
Accuracy	91.10%
Precision	45.50%
Recall	40.00%
Misclassification Rate	8.90%
F1 Score	42.60%

Figure 6-4– Metrics – All variables – Case 2

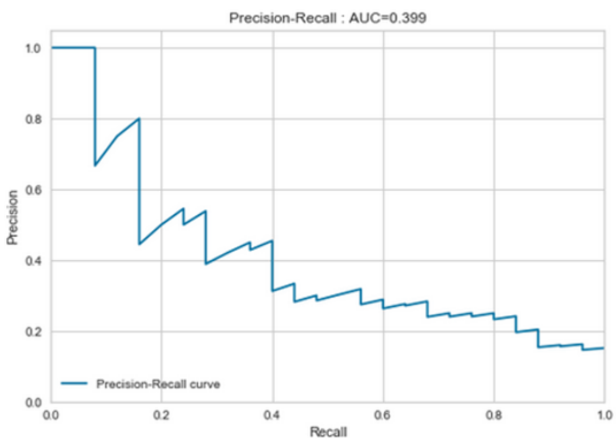


Figure 6-5 – PR Curve – All variables – Case 2

Overall the results obtained are very similar to the ones presented in case 1. Similar to case 1, the model seems to underperform, regarding all the considered metrics, compared to both the cases where individually the variables debt-to-market equity and distance to default were excluded. Therefore, adding a market-related variable seems to be pertinent. Moreover, distance to default seems again to be a better contributor when compared to debt-to-market equity by assessing the differences between the F1-score between the last three columns. When the distance to default is removed from the baseline model, the F1-score value decreases nearly 8 percentual points, while only 5 percentual points when the debt-to-market equity ratio is excluded. In addition, the PR AUC decreased in the scenario where debt-to-market equity is excluded as opposing to the scenario where the distance to default variable is removed. Again, when employed solely, the distance to default presents a better performance than the debt-to-market equity ratio as it is clear from the higher F1-score.

In general, the model that performs better according to the F1-score metric seems to be when retained earnings-to-total assets is removed. The explanation might be again related to the fact that 4 profitability ratios are being used, and this variable might be just adding “noise”. However, another possible explanation might be related to the fact that two-years prior to default, this ratio does not differ that much from the non-bankrupt dataset. The average value for this ratio for the df1 is -4.59, for the df2 is -2.11 and for the df3 is -1.19, which is corroborative with the given explanation.

By analyzing the baseline model and when compared to the first case, one observes that the F1-Score and accuracy metrics displayed higher values despite presenting a lower PR AUC. However, if one considers all the sub scenarios presented, then, generally, using the one-year prior to default instances is preferable than using the two-years prior to default instances. In 9 of the 17 scenarios considered, the F1-score displays a higher value for case 1. Furthermore, one can observe that in 12 of the 17 considered scenarios, the PR AUC for case 1 outperforms case 2.

				All variables minus				
Panel A	All variables	Debt/Market Equity	Distance to Default	Current Assets / Total Assets	Current Assets / Current Liabilities	EBIT / Total Assets	Working Capital / Total Assets	Debt / Total Assets
Accuracy	91%	83%	69%	91%	91%	90%	91%	91%
Recall	40%	20%	68%	40%	48%	40%	36%	40%
Precision	45%	14%	16%	43%	44%	42%	45%	43%
F1-Score	43%	17%	26%	42%	46%	41%	40%	42%
PR AUC	40%	15%	15%	49%	47%	45%	45%	39%
Kernel	RBF	RBF	RBF	RBF	RBF	RBF	RBF	RBF
C	100	1	10	10	100	10	100	100
gamma	1	1	10	1	1	1	1	1

	All variables minus								
Panel B	Cash / Total Asset	Sales / Total Assets	Ln (Total Assets)	Retained Earnings / Total Assets	Earnings / Stockholder Equity	Earnings / Total Assets	Debt/Market Equity	Distance to Default	Distance to Default & Debt/Market Equity
Accuracy	90%	91%	86%	91%	90%	90%	90%	90%	88%
Recall	28%	40%	44%	52%	40%	32%	36%	32%	24%
Precision	35%	43%	28%	46%	40%	40%	39%	38%	27%
F1-Score	31%	42%	34%	49%	40%	36%	37%	35%	26%
PR AUC	32%	35%	29%	48%	42%	34%	36%	40%	27%
Kernel	RBF	RBF	RBF	RBF	RBF	RBF	RBF	RBF	RBF
C	1 000	1 000	100	100	10	100	1 000	1 000	100
gamma	1	1	1	1	1	1	1	1	1

Table 6-7– SVM results – Case 2

6.2.2 Cross-Validation

Case 3:

Similarly to case 1, df1 will be concatenated with df3. However, instead of the percentage ratio split, the stratified 5-fold cross-validation method will be employed. As explained in section 5.2, for each fold, a grid search optimization will be applied. Six different specifications are considered. These are: all the variables, all variables minus distance to default, all variables minus debt-to-market equity, all variables minus distance to default and debt-to-market equity, only distance to default and finally, only debt-to-market equity.

Table 6.8 presents the various model specifications abovementioned. Similar to cases 1 and 2, the exclusion of both the distance to default and the debt-to-market equity leads to a reduction in all performance indicators. The reduction, in this case, is however significantly lower than the ones shown in the previous cases. Moreover, distance to default seems, once again, to be preferable when compared to debt-to-market equity by analyzing the F1-score of the last three columns.

Similarly to cases 1 and 2, the distance to default is shown to be a better predictor than the debt-to-equity ratio when assessed against the baseline model. For instance, the average F1-score declined 6 percentual points when the distance to default is removed. Contrarily, when debt-to-market equity is removed, the F1-score actually increased. This indicates that the distance to default variable contribution seems to outperform the debt-to-market equity variable when inserted within a given framework of variables. Finally, and differently from cases 1 and 2, it is not clear in this case that the distance to default is a better predictor than the debt-to-market equity ratio when solely employed. Table 6.8 shows that the distance to default, when solely employed and compared to the debt-to-market equity variable, demonstrates higher PR AUC, but lower F1-score. Hence, debt-to-market equity seems to be preferable individually.

				All variables minus		
Metrics	All variables	Debt/Market Equity	Distance to Default	Debt/Market Equity	Distance to Default	Distance to Default & Debt/Market
Accuracy	90%	87%	73%	90%	88%	89%
Recall	44%	35%	72%	50%	40%	43%
Precision	38%	29%	19%	40%	31%	35%
F1-Score	41%	32%	30%	44%	35%	39%
PR AUC	40%	26%	29%	37%	33%	36%

Table 6-8 – SVM results – Case 3 – Average of the K-folds

Case 4:

Last but not least, case 4, similarly to case 2, concatenates df2 with df3 but instead makes use of the stratified 5-fold cross-validation technique.

Identically to the already analyzed cases, by removing both market-related variables, the F1-score decreased. In this specific case by 8 percentual points. Opposingly to the already considered cases, by comparing the F1-score of the last three columns, one can realize that adding distance to default is not as attractive as adding the debt-to-market equity variable. However, one may still conclude that adding a market-related variable seems to be relevant. Once again, by excluding the variable distance to default, both recall and F1-score metrics decrease, comparatively to the baseline model. In this case, debt-to-market equity seems to outperform distance to default. When debt-to-market equity variable is removed, the F1-score decreased 3 percentual points while when distance to default was excluded, it only fell by 2 percentual points. Finally, and differently from case 3, the distance to default seems to outperform the debt-to-market equity ratio when solely employed. Table 6.9 demonstrates that the distance to default comparatively to the debt-to-market equity exhibits a similar PR AUC but a higher F1-score when individually considered.

Lastly, by comparing both cases 3 and 4, one can observe that case 3, for all the considered scenarios, displayed a higher F1-score and a higher PR AUC. Hence, using the instances a year prior to default seems to be preferable.

				All variables minus		
Metrics	All variables	Debt/Market Equity	Distance to Default	Debt/Market Equity	Distance to Default	Distance to Default & Debt/Market
Accuracy	89%	84%	63%	89%	90%	91%
Recall	32%	21%	73%	29%	28%	22%
Precision	33%	17%	15%	35%	43%	42%
F1-Score	32%	18%	25%	29%	30%	24%
PR AUC	30%	18%	18%	30%	32%	28%

Table 6-9 – SVM results – Case 4 – Average of the K-folds

Heretofore, the discussion on which out-of-sample method performed better has not been emphasized. Figures 6.6 and 6.7 are informative regarding this particular question. The first mentioned figure regards the comparison of both methodologies in the cases which consider both df1 and df3. As one can perceive, in 5 of the 6 scenarios, the stratified K-fold cross-validation seems to achieve better results in what concerns the F1-score. However, the percentage ratio split method demonstrates higher values in 4 of the 6 scenarios regarding the PR AUC. The accuracy results are very similar in both methodologies. Nevertheless, as the F1-

score was considered the key metric in the analyzes, it seems that the performance of the stratified K-fold cross-validation employed was superior.

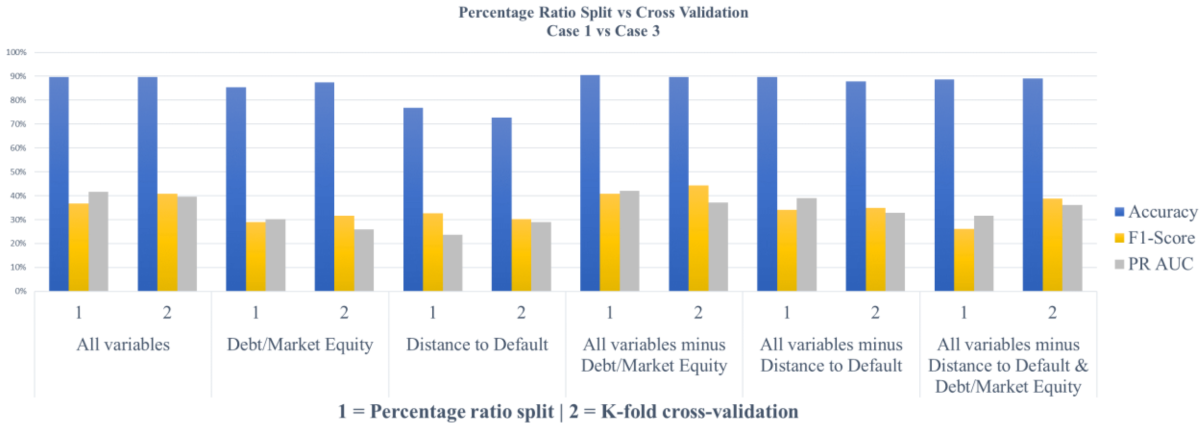


Figure 6-6 – Cross Validation - Percentage Ratio split – Case 1 vs Case 3

Regarding cases 2 and 4, presented in Figure 6.7, one can observe that contrarily to the analyzed situation above, the stratified K-fold cross-validation seems not to be preferable when compared to the percentage ratio split method concerning the F1-score as in 5 of the 6 scenarios it displays lower values. In what respects the PR AUC, in 3 scenarios, one of the methodologies is superior, while in the remaining 3 scenarios, the other out-of-sample method performs better. All things considered, the percentage ratio split method, accordingly to the initial assumption, exhibits better results.

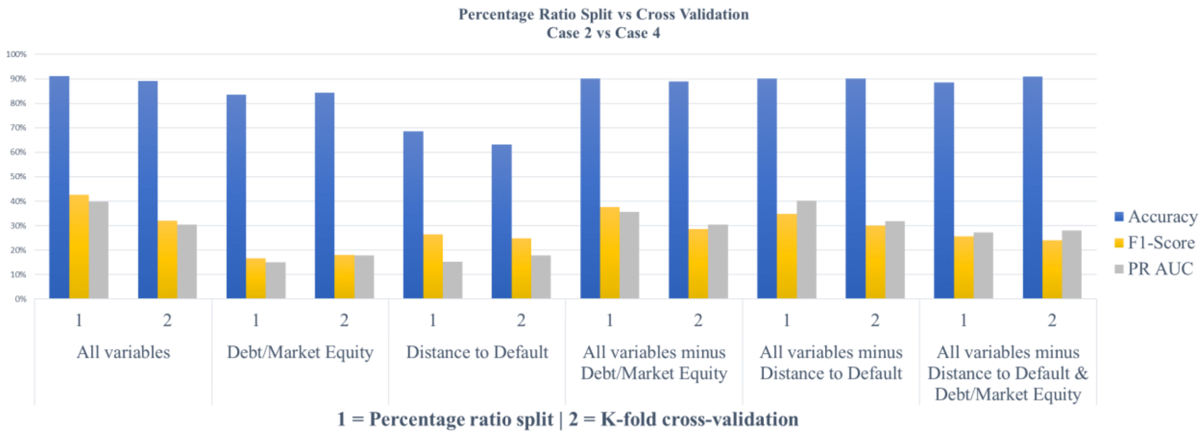


Figure 6-7– Cross Validation - Percentage Ratio split – Case 2 vs Case 4

Table 6.10 summarizes the abovementioned presented results. Four conclusions are worth noting. First, the distance to default has shown to be an important variable within the given framework variables, since, in all the four cases, it contributed to overall better results.

Second, distance to default, when assessed against the debt-to-market equity variable, in both case 1 and case 2, performed better individually, contributed more to the framework which considered all the variables and was also preferable when assessed against the case where both market-related variables were excluded. Regarding case 3, the distance to default outperformed the debt-to-market equity variable when assessed against the baseline model. Moreover, in the scenario where both market-related variables were excluded, distance to default was again superior as it contributed to a higher F1-score. However, individually, the debt-to-market equity displayed a higher F1-Score. Oppositely, case 4 demonstrates that individually the distance to default performs better despite contributing less when assessed against the baseline model. In the scenario where both market-related variables were excluded, debt-to-market equity was also preferable when compared to the distance to default variable as it enhanced more the F1-score. All things considered, distance to default seems to be more valuable than the debt-to-market equity when utilized in the SVM model.

Third, cross-validation delivered better results when one considers instances one-year prior default. However, when one considers instances two-years prior default, the 80:20 split outperforms. Hence, by considering both cases, no conclusions can be taken about whether one is preferable to the other.

Lastly, the results obtained using the one-year prior to default instances, are generally superior to those which make use of the two-years prior to default instances.

Distance to default		**Cross-Validation > 80:20 Split**	
Improved	**Did not Improve**	**Yes**	**No**
4	0	1	1

One-year prior default > Two-years prior default		**Distance to default > Debt/Market Equity**	
Yes	**No**	**Yes**	**No**
2	0	9	3

Table 6-10 – Summary results

Description: First, in all the four considered cases, distance to default improved the overall performance of the metrics. Second, K-fold cross-validation displayed better results comparatively to the 80:20 split while using the instances a year prior to default. The opposite happens when using the instances two-years prior to default. Third, using the one-year prior default instances in both out-of-sample methods demonstrated to be preferable. Lastly, distance to default was considered to be in 9 out of the 12 analyzed cases to be a better contributor when assessed against debt-to-market equity. For each of the four cases, three different components were analyzed: individually, within a given set of variables and by removing both market-related variables.

Chapter 7

7 Conclusion

This dissertation aims to contribute to the already extense literature on bankruptcy prediction. With that purpose, four questions were considered. The first and main question of this empirical study is to assess whether the output from the KMV-Merton model, the distance to default variable, can contribute to the SVM model with the ultimate goal of better forecasting the bankruptcy of a company. Secondly, I am interested in understanding whether the distance to default is preferable to the use of a more straightforward market-related variable such as debt-to-market equity. In order to answer these two questions, four model specifications were considered: excluding the variable distance to default from the baseline model, excluding the variable debt-to-market equity from the baseline model, excluding both variables from the baseline model and the baseline model itself. By comparing the main performance metrics in each of these cases one can have an idea of the marginal contribution of the distance to default and debt-to-market equity. Third, this dissertation assesses the performance of two particular out-of-sample evaluation methods, notably, the percentage ratio split and the stratified K-fold cross-validation technique. The fourth and last proposed question regards whether using the instances one-year prior to default is preferable to the use of the instances of the companies two-years prior to the default event. In order to assess that, four different cases were constructed. The first case, which was evaluated following a percentage ratio split, used the instances of companies a year prior-to default. The second case, which also followed the same out-of-sample methodology, used the instances two-years prior default. The third and the fourth cases followed the same rationale but instead employed the stratified K-fold cross-validation technique.

The dataset used to answer these questions considered 248 non-financial companies from the United States and covers the period between 2000 and 2018. This dataset was mainly composed of companies from the "Health Care", "Information and Technology", and "Industrials" sectors.

The obtained results were overall very elucidatory. It was found evidence that when applying the SVM model, the distance to default variable is relevant when considered within a given framework of variables. Also, the distance to default variable is a better contributor when compared to the more simplistic market-related variable. Moreover, in three of the four cases considered, adding a market-related variable improved both the F1-score and the PR AUC. These conclusions are of major relevance, enriching the current literature on bankruptcy prediction associated with the use of the SVM.

Regarding the third abovementioned question, this study is not conclusive in what concerns which out-of-sample evaluation method offers better results; the percentage ratio split, or the stratified K-fold cross-validation. Notwithstanding, as Zhang et al. (1999) stated, by considering the traditional percentage division on the data, one may probably be introducing bias in the model as the features of the test sample will possibly significantly differ from those in the training sample.

Lastly, one can conclude that the cases, which considered the instances of companies a year prior to default, seem to exhibit overall better results in what regards the F1-score when compared with the cases which made use of the instances of the companies' two-years prior to default.

Notwithstanding, this study has some limitations. First, one has assumed, in order to answer the questions, a given set of variables. However, if other different variables were chosen, the results could slightly diverge from the ones obtained. Second, the dataset only considers US firms, so one can´t make extrapolations regarding other geographical areas. Third, as in chapter 4, one has assumed that a company is bankrupt in the year after the last available accounting information, which in a few cases, may not be accurate. Finally, the last identified limitation regards the using of the SMOTE. Since different results are obtained depending on the synthetic examples outputted by this technique, one possible solution to minimize this drawback could be achieved by testing the model several times and then average the results.

8 Appendix

	Skewness	Kurtosis
One-year distance to default	13.54	205.34
Total Liabilities/Total Assets	8.48	165.33
Drift rate	12.42	199.78
Asset volatility	18.58	434.79
Equity Volatility	3.40	30.55

Appendix 1– Statistics (Skewness/Kurtosis) - all the data

Case 3

	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	
Metrics	All variables					Average
Accuracy	0.89	0.89	0.90	0.91	0.89	0.90
Recall	0.44	0.36	0.36	0.68	0.38	0.44
Precision	0.37	0.36	0.38	0.47	0.33	0.38
F1-Score	0.40	0.36	0.37	0.56	0.35	0.41
PR AUC	0.32	0.39	0.35	0.55	0.37	0.40
Kernel	RBF	RBF	RBF	RBF	RBF	
C	1000.00	10.00	10.00	1000.00	1000.00	
gamma	0.10	1.00	1.00	0.10	1.00	

Appendix 2– Case 3 – K-Fold estimation – All variables

	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	
Metrics	All variables minus Distance to Default					Average
Accuracy	0.89	0.87	0.86	0.88	0.88	0.88
Recall	0.44	0.28	0.36	0.60	0.33	0.40
Precision	0.38	0.26	0.26	0.36	0.30	0.31
F1-Score	0.41	0.27	0.31	0.45	0.31	0.35
PR AUC	0.31	0.31	0.32	0.45	0.26	0.33
Kernel	RBF	RBF	RBF	RBF	RBF	
C	1000.00	10.00	1000.00	1000.00	1000.00	
gamma	0.10	1.00	0.10	0.10	0.10	

Appendix 3- Case 3 – K-Fold estimation – All variables minus Distance to default

Metrics	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	Average
	All variables minus Debt/Market Equity					
Accuracy	0.91	0.90	0.85	0.91	0.90	0.90
Recall	0.52	0.40	0.40	0.68	0.50	0.50
Precision	0.46	0.40	0.26	0.49	0.40	0.40
F1-Score	0.49	0.40	0.31	0.57	0.44	0.44
PR AUC	0.45	0.33	0.22	0.49	0.37	0.37
Kernel	RBF	RBF	RBF	RBF	RBF	
C	10.00	1000.00	100.00	10.00	1000.00	
gamma	1.00	1.00	0.10	1.00	0.10	

Appendix 4 – Case 3 – K-Fold estimation – All variables minus debt-to-market equity

Metrics	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	Average
	Debt/Market Equity					
Accuracy	0.89	0.88	0.89	0.86	0.84	0.87
Recall	0.24	0.36	0.40	0.44	0.33	0.35
Precision	0.32	0.30	0.37	0.28	0.21	0.29
F1-Score	0.27	0.33	0.38	0.34	0.25	0.32
PR AUC	0.17	0.22	0.37	0.31	0.21	0.26
Kernel	RBF	RBF	RBF	RBF	RBF	
C	1.00	1000.00	10.00	100.00	1.00	
gamma	1.00	10.00	1.00	10.00	10.00	

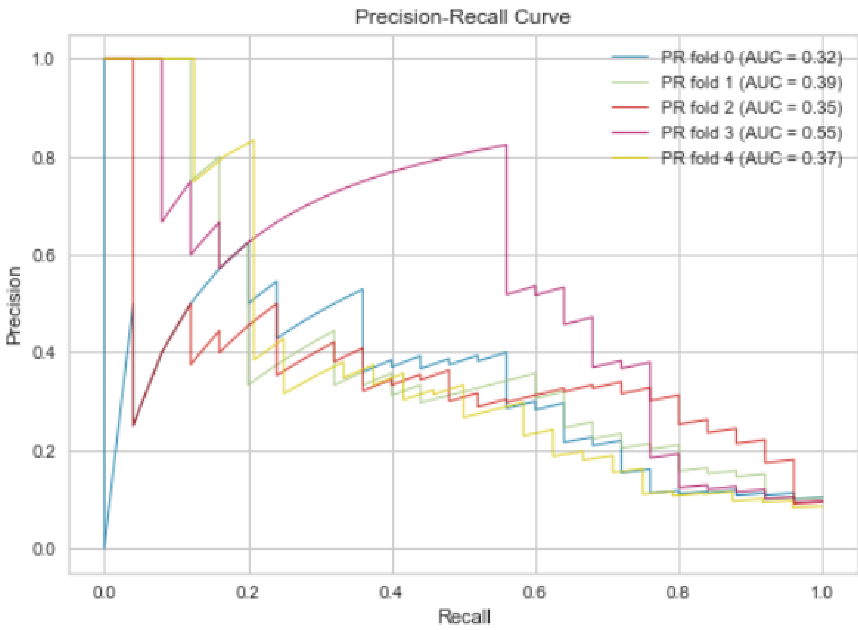
Appendix 5 – Case 3 – K-Fold estimation – debt-to-market equity

Metrics	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	Average
	Distance to Default					
Accuracy	0.72	0.76	0.70	0.70	0.75	0.73
Recall	0.72	0.68	0.84	0.68	0.67	0.72
Precision	0.19	0.21	0.20	0.17	0.19	0.19
F1-Score	0.30	0.32	0.32	0.27	0.30	0.30
PR AUC	0.35	0.19	0.34	0.36	0.21	0.29
Kernel	RBF	RBF	RBF	RBF	RBF	
C	10.00	10.00	1000.00	1000.00	1000.00	
gamma	10.00	10.00	1.00	1.00	1.00	

Appendix 6 – Case 3 – K-Fold estimation – Distance to default

Metrics	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	Average
	All variables minus Distance to Default & Debt/Market equity					
Accuracy	0.90	0.89	0.87	0.90	0.89	0.89
Recall	0.44	0.36	0.40	0.68	0.29	0.43
Precision	0.41	0.35	0.30	0.43	0.29	0.35
F1-Score	0.42	0.35	0.34	0.52	0.29	0.39
PR AUC	0.41	0.32	0.29	0.46	0.33	0.36
Kernel	RBF	RBF	RBF	RBF	RBF	
C	100.00	10.00	1000.00	1000.00	1000.00	
gamma	1.00	1.00	0.10	0.10	0.10	

Appendix 7 – Case 3 – K-Fold estimation – All variables minus debt-to-market equity & distance to default



Appendix 8 – Case 3- K-Fold Estimation – PR for each fold

Case 4

Metrics	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	Average
	All variables					
Accuracy	0.89	0.89	0.87	0.88	0.91	0.89
Recall	0.44	0.24	0.24	0.28	0.38	0.32
Precision	0.38	0.32	0.23	0.28	0.45	0.33
F1-Score	0.41	0.27	0.24	0.28	0.41	0.32
PR AUC	0.40	0.24	0.28	0.25	0.34	0.30
Kernel	RBF	RBF	RBF	RBF	RBF	
C	10.00	10.00	10.00	1000.00	10.00	
gamma	1.00	1.00	1.00	0.10	1.00	

Appendix 9 – Case 4 – K-Fold estimation – All variables

Metrics	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	Average
	All variables minus Distance to Default					
Accuracy	0.89	0.89	0.92	0.88	0.92	0.90
Recall	0.48	0.20	0.08	0.36	0.29	0.28
Precision	0.36	0.29	0.67	0.32	0.50	0.43
F1-Score	0.41	0.24	0.14	0.34	0.37	0.30
PR AUC	0.42	0.19	0.25	0.34	0.39	0.32
Kernel	RBF	RBF	RBF	RBF	RBF	
C	10.00	10.00	10.00	1.00	10.00	
gamma	1.00	1.00	10.00	1.00	1.00	

Appendix 10 - Case 4 – K-Fold estimation – All variables minus Distance to default

Metrics	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	Average
	All variables minus Debt/Market Equity					
Accuracy	0.88	0.87	0.92	0.87	0.91	0.89
Recall	0.52	0.20	0.08	0.32	0.33	0.29
Precision	0.34	0.21	0.50	0.27	0.44	0.35
F1-Score	0.41	0.20	0.14	0.29	0.38	0.29
PR AUC	0.42	0.23	0.29	0.26	0.31	0.30
Kernel	RBF	RBF	RBF	RBF	RBF	
C	10.00	10.00	10.00	1000.00	100.00	
gamma	1.00	1.00	10.00	0.10	1.00	

Appendix 11– Case 4 – K-Fold estimation – All variables minus debt-to-market equity

Metrics	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	Average
	Debt/Market Equity					
Accuracy	0.86	0.83	0.82	0.80	0.90	0.84
Recall	0.16	0.08	0.24	0.32	0.25	0.21
Precision	0.15	0.06	0.15	0.16	0.32	0.17
F1-Score	0.16	0.07	0.18	0.21	0.28	0.18
PR AUC	0.18	0.14	0.11	0.16	0.24	0.18
Kernel	RBF	RBF	RBF	RBF	RBF	
C	1000.00	0.10	1.00	0.10	1000.00	
gamma	0.10	10.00	10.00	10.00	0.10	

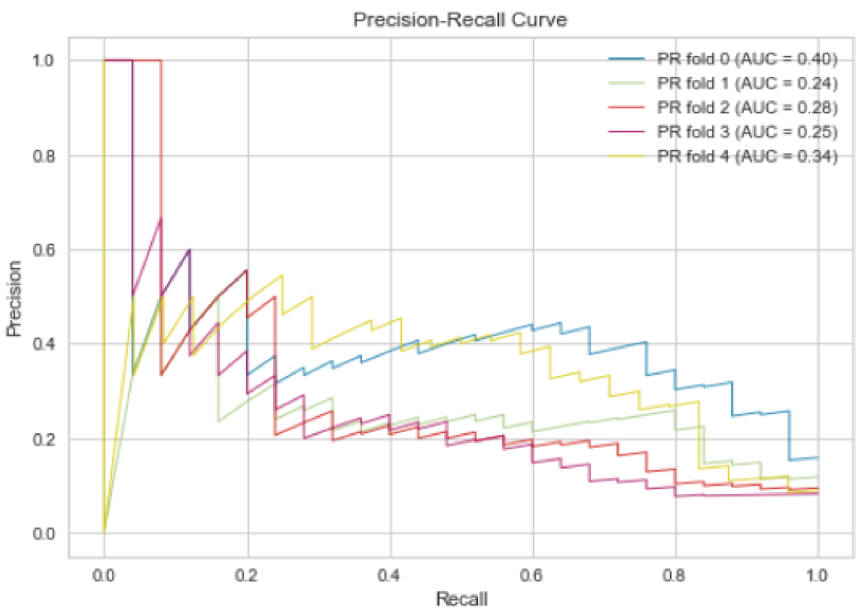
Appendix 12 – Case 4 – K-Fold estimation – debt-to-market equity

Metrics	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	
	Distance to Default					Average
Accuracy	0.61	0.60	0.69	0.61	0.64	0.63
Recall	0.84	0.72	0.68	0.60	0.83	0.73
Precision	0.16	0.14	0.17	0.12	0.16	0.15
F1-Score	0.26	0.23	0.27	0.20	0.27	0.25
PR AUC	0.33	0.14	0.15	0.13	0.15	0.18
Kernel	RBF	RBF	RBF	RBF	RBF	
C	1.00	100.00	100.00	1.00	10.00	
gamma	10.00	10.00	10.00	10.00	10.00	

Appendix 13 – Case 4 – K-Fold estimation – Distance to default

Metrics	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	
	Distance to Default & Debt/Market equity					Average
Accuracy	0.88	0.91	0.92	0.92	0.90	0.91
Recall	0.56	0.08	0.08	0.12	0.25	0.22
Precision	0.35	0.40	0.50	0.50	0.35	0.42
F1-Score	0.43	0.13	0.14	0.19	0.29	0.24
PR AUC	0.38	0.23	0.28	0.25	0.27	0.28
Kernel	RBF	RBF	RBF	RBF	RBF	
C	10.00	10.00	10.00	10.00	10.00	
gamma	1.00	10.00	10.00	10.00	10.00	

Appendix 14 – Case 4 – K-Fold estimation – All variables minus debt-to-market equity & distance to default



Appendix 15 – Case 4- K-Fold Estimation – PR for each fold

9 References

Agarwal, Vineet, and Richard J. Taffler. "Comparing the Performance of Market-Based and Accounting-Based Bankruptcy Prediction Models." *SSRN Electronic Journal*, 2006, doi:10.2139/ssrn.968252.

Aguiar-Díaz, Inmaculada, and María Victoria Ruiz-Mallorquí. "Causes and Resolution of Bankruptcy: The Efficiency of the Law." *The Spanish Review of Financial Economics*, vol. 13, no. 2, 2015, pp. 71–80., doi:10.1016/j.srfe.2015.04.001.

Alaminos, David, et al. "A Global Model for Bankruptcy Prediction." *Plos One*, vol. 11, no. 11, 2016, doi:10.1371/journal.pone.0166693.

Altman, Edward I, and Anthony Saunders. "Credit Risk Measurement: Developments over the Last 20 Years." *Journal of Banking & Finance*, vol. 21, no. 11-12, 1997, pp. 1721–1742., doi:10.1016/s0378-4266(97)00036-8.

Altman, Edward I. "Financial Ratios, Discriminant Analysis And The Prediction Of Corporate Bankruptcy." *The Journal of Finance*, vol. 23, no. 4, 1968, pp. 589–609., doi:10.1111/j.1540-6261.1968.tb00843.x.

Beaver, William H. "Financial Ratios As Predictors of Failure." *Journal of Accounting Research*, vol. 4, 1966, p. 71., doi:10.2307/2490171.

Bharath, Sreedhar T., and Tyler Shumway. "Forecasting Default with the KMV-Merton Model." *SSRN Electronic Journal*, 2004, doi:10.2139/ssrn.637342.

Black, Fischer, and John C. Cox. "Valuing Corporate Securities: Some Effects Of Bond Indenture Provisions." *The Journal of Finance*, vol. 31, no. 2, 1976, pp. 351–367., doi:10.1111/j.1540-6261.1976.tb01891.x.

Black, Fischer, and Myron Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy*, vol. 81, no. 3, 1973, pp. 637–654., doi:10.1086/260062.

Boyd, Kendrick, et al. "Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals." *Advanced Information Systems Engineering Lecture Notes in Computer Science*, 2013, pp. 451–466., doi:10.1007/978-3-642-40994-3_29.

Crosbie, Peter J., and Jeffrey R. Bohn. Modeling Default Risk. 2003.

Fletcher, Tristan. "Support Vector Machines Explained." *Academia.edu*, 2008, www.academia.edu/22709227/Support_Vector_Machines_Explained.

Gepp, Adrian, and Kuldeep Kumar. "Predicting Financial Distress: A Comparison of Survival Analysis and Decision Tree Techniques." *Procedia Computer Science*, vol. 54, 2015, pp. 396–404., doi:10.1016/j.procs.2015.06.046.

Géron Aurélien. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc, 2019.

References

Gissel, Jodi, et al. “A Review of Bankruptcy Prediction Studies: 1930-Present.” *Journal of Financial Education*, vol. 33, no. Winter, 2007, pp. 1–42.

Huang, Cheng-Lung, et al. “Credit Scoring with a Data Mining Approach Based on Support Vector Machines.” *Expert Systems with Applications*, vol. 33, no. 4, 2007, pp. 847–856., doi:10.1016/j.eswa.2006.07.007.

Huang, Zan, et al. “Credit Rating Analysis with Support Vector Machines and Neural Networks: a Market Comparative Study.” *Decision Support Systems*, vol. 37, no. 4, 2004, pp. 543–558., doi:10.1016/s0167-9236(03)00086-1.

Jordan, M. I., and T. M. Mitchell. “Machine Learning: Trends, Perspectives, and Prospects.” *Science*, American Association for the Advancement of Science, 17 July 2015, science.sciencemag.org/content/349/6245/255.full.

Löffler, Gunter, and Peter N. Posch. *Credit Risk Modeling Using Excel and VBA*. Wiley Publishing, 2011.

Hillegeist, Stephen A., et al. “Assessing the Probability of Bankruptcy.” *SSRN Electronic Journal*, 2002, doi:10.2139/ssrn.307479.

Merton, Robert C. “On the Pricing of Corporate Debt: The Risk Structure of Interest Rates.” *The Journal of Finance*, vol. 29, no. 2, 1974, p. 449., doi:10.2307/2978814.

Merwin, Charles Lewis. *Financing Small Corporations in Five Manufacturing Industries, 1926-36. A Dissertation in Economics. Financing Small Corporations in Five Manufacturing Industries,* 1942

Min, J, and Y Lee. “Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters.” *Expert Systems with Applications*, vol. 28, no. 4, 2005, pp. 603–614., doi:10.1016/j.eswa.2004.12.008.

Naqa, Issam El, and Martin J. Murphy. “What Is Machine Learning?” *Machine Learning in Radiation Oncology*, 2015, pp. 3–11., doi:10.1007/978-3-319-18305-3_1.

Odom, M.d., and R. Sharda. “A Neural Network Model for Bankruptcy Prediction.” *1990 IJCNN International Joint Conference on Neural Networks*, 1990, doi:10.1109/ijcnn.1990.137710.

Ohlson, James A. “Financial Ratios and the Probabilistic Prediction of Bankruptcy.” *Journal of Accounting Research*, vol. 18, no. 1, 1980, p. 109., doi:10.2307/2490395.

Sundaresan, Suresh. “A Review of Merton’s Model of the Firm’s Capital Structure with Its Wide Applications.” *Annual Review of Financial Economics*, vol. 5, no. 1, 2013, pp. 21–41., doi:10.1146/annurev-financial-110112-120923.

Vapnik, Vladimir Naumovich. *Statistical Learning Theory*. Wiley, 1998.

References

Zhang, Guoqiang, et al. “Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis.” *European Journal of Operational Research*, vol. 116, no. 1, 1999, pp. 16–32., doi:10.1016/s0377-2217(98)00051-4.