Bridging the Gap Within the Default Probability Discipline: The Default Model of Synthesis

Master's Thesis

Copenhagen Business School

15th of May 2020

Authors

Aske-Emil Brandt

S103135

M.Sc. EBA. Finance & Accounting

Kristoffer Visbjerg

S103275

M.Sc. EBA. Applied Economics and Finance

Supervisor

Clara Dyhrberg

Credit Risk Modelling, Nordea

STU: 212,467 | Page Count: 111 | Standard Pages: 93.5

(This page intentionally left blank)

Abstract

The discipline of predicting company defaults is of great economic significance both due to the consequences for affected businesses and individuals, and due to the implications for investment and lending activity. This thesis investigates the discipline through an *academic*, *practical* and *methodological* lens, with the aim of contributing to the improvement of the predictive techniques available. The thesis evaluates three hypotheses deducted from the pertinent theory. In the first hypothesis, it examines *to what extent* the classic academic models of default probability have discriminative power. In the second and third hypothesis, it considers *through what measures* the practical model of credit risk can be improved. In the analysis of *to what extent the classic academic models of default probability have discriminative power*, the thesis applies the original frameworks of Altman, Ohlson, Merton and their respective re-estimations on a modern portfolio. The portfolio distinguishes itself from the related literature with a scope simultaneously covering a wide range of OECD countries and multiple industries. This analysis concludes that the classic academic approaches to credit risk do have significant discriminative power in a contemporary setting. The subsequent analysis of *through what measures the practical model of credit risk can be improved* is split into two parts. The first is academically driven and expand upon the practical model by combining the accounting-based and market-based paradigms of the field. This framework is labelled the "Default Model of Synthesis", and the thesis finds that it is superior to the model relied upon by practitioners. The second is methodologically driven and transcends a machine learning algorithm into the sphere of default probability. The research concludes that the practical model can be improved significantly by applying a random forest methodology. This finding also serves as a platform for discussing *the implications of machine learning in the practical discipline* of default probability. The discussion points toward interpretability and "institutional stickiness" as non-exclusive explanations for the neglection of machine learning amongst practitioners. Ultimately, the research contributes to the field of default probability both practically and academically. First, it conceptualizes a framework that relies on market and accounting theory in synthesis. Second, it extends the practical model to default probability through the application of a random forest algorithm.

Table of Contents

ABSTRACT **2**

LIST OF ABBREVIATIONS: FINANCIAL RATIOS **7**

1. INTRODUCTION **9**

1.1 CONTRIBUTION TO THE LITERATURE: WHY THIS THESIS IS INTERESTING? 10
1.2 HOW THE THESIS WILL ANSWER THE RESEARCH QUESTION 11
1.3 DELIMITATION AND HOW WE SEE PROBABILITY OF DEFAULT 12

2. PHILOSOPHY OF SCIENCE **14**

2.1 CRITICAL REALISM 14
2.2 MODE OF REASONING AND APPROPRIATENESS OF CRITICAL REALISM 16

3. THEORETICAL FRAMEWORK: A REVIEW OF DEFAULT PROBABILITY IN THEORY, ACADEMIA AND PRACTICE **18**

3.1 AN ACCOUNTING-BASED APPROACH TO DEFAULT PROBABILITY 18
3.1.1 The Accounting Theory of Credit Risk 18
3.1.2 Literature Review of the Accounting-based Approach 19
3.2 A MARKET BASED APPROACH TO CREDIT RISK 21
3.2.1 The Structural Theory of Credit Risk 21
3.2.2 Literature Review of the Structural Approach 22
3.3 THE LITERATURE ON PROBABILITY OF DEFAULT: A CLASH BETWEEN PARADIGMS 23
3.4 THE PRACTICAL APPROACH TO PROBABILITY OF DEFAULT 25
3.5 MACHINE LEARNING 26
3.5.1 Machine Learning Theory 26
3.5.2 Machine Learning in the Credit Risk Discipline 27
3.6 GAPS IN THE LITERATURE 28
3.7 FORMULATION OF HYPOTHESES 29

4. METHODOLOGICAL FRAMEWORK **32**

4.1 THE METHODOLOGY OF ALTMAN: MULTIVARIATE DISCRIMINANT ANALYSIS AND LINEAR REGRESSION 32
4.2 THE METHODOLOGY OF OHLSON: LOGIT MODEL 35
4.3 THE METHODOLOGY OF MERTON: THE ORIGINAL AND THE NAÏVE 37
4.4 THE PRACTICAL MODELLING METHODOLOGY 40
4.5 MACHINE LEARNING METHODOLOGY 42
4.5.1 Machine Learning Methodology: Classification Trees 42
4.5.2 Machine Learning Methodology: Random Forest 44
4.6 METRICS FOR COMPARING MODELS 45
4.6.1 Receiver Operating Characteristic Curves 46
4.6.2 Somers' D 49
4.7 TRAINING AND VALIDATION 50

5. DATA **51**

5.1 DATA COLLECTION DESIGN 51
5.2 DATA CLEANSING PROCEDURES 52
5.3 CALCULATION OF INPUT PARAMETERS 54
5.3.1 Accounting-based models 54
5.3.2 Structural Models 56

5.4 Final Descriptive Statistics 57

6. RESULTS 58

6.1 The Academic Approach to Default Probability 58
6.1.1 Original Framework of Altman 58
6.1.2 Re-estimation of the Altman Model 61
6.1.3 Original Framework of Ohlson 64
6.1.4 Re-estimation of the Ohlson Model 67
6.1.5 The Original Framework of Merton 69
6.1.6 The Naïve Approach to the Merton Framework 71
6.1.7 Overview and Evaluation of Hypothesis 1 73
6.2 The Practical Approach to Default Probability 76
6.2.1 The Practitioners' Default Probability Model 76
6.2.2 The Default Model of Synthesis 85
6.2.3 Overview and Evaluation of Hypothesis 2 90
6.3 The Machine Learning Approach to Default Probability 91
6.3.1 A Probability of Default Model using Classification Trees 92
6.3.2 Random Forest: Extending the Classification Tree Framework 95
6.3.3 Overview and Evaluation of Hypothesis 3 99

7. DISCUSSION 101

7.1 A Reflection on the Neglection of Machine Learning in Practice 101
7.2 A Question of Interpretability 102
7.3 Institutional Stickiness: A Possible Explanation for the Precedence of Logistic Regression 103

8. REFLECTIONS ON VALIDITY AND RELIABILITY 105

8.1 Measurement Validity, Reliability and Replicability 105
8.2 Internal Validity 106
8.3 External Validity 106

9. CONCLUSION 108

9.1 Contribution 110

10. AVENUES OF FURTHER RESEARCH 111

11. BIBLIOGRAPHY 112

12. APPENDICES 117

12.1 Appendix: Linearity Plots 117
12.2 Appendix: Correlation Matrix from R-software 118
12.3 Appendix: Code Book 119

Table of Illustrations

1. INTRODUCTION 9

TABLE 1.1: HYPOTHESES FORMULATED FROM PERTINENT LITERATURE AND THEORY....................11

2. PHILOSOPHY OF SCIENCE...14

FIGURE 1.1: THE THREE REALMS OF CRITICAL REALISM...16

3. THEORETICAL FRAMEWORK: A REVIEW OF DEFAULT PROBABILITY IN THEORY, ACADEMIA AND PRACTICE 18

FIGURE 3.1: PROBABILITY OF DEFAULT MODEL: ACCOUNTING THEORY 19
FIGURE 3.2: PROBABILITY OF DEFAULT MODEL: MARKET THEORY...22
TABLE 3.1: OVERVIEW OF PERTINENT THEORY AND LITERATURE FOR THE RESEARCH...................28
TABLE 3.2: FIRST HYPOTHESIS FORMULATED FROM PERTINENT LITERATURE AND THEORY..........29
TABLE 3.3: SECOND HYPOTHESIS FORMULATED FROM PERTINENT LITERATURE AND THEORY 30
TABLE 3.4: THIRD HYPOTHESIS FORMULATED FROM PERTINENT LITERATURE AND THEORY.........30
TABLE 3.5: HYPOTHESES FORMULATED FROM PERTINENT LITERATURE AND THEORY....................31
FIGURE 3.3: HYPOTHESES ANSWERING THE RESEARCH QUESTION..31
FIGURE 3.3: HYPOTHESES ANSWERING THE RESEARCH QUESTION..31

4. METHODOLOGICAL FRAMEWORK 32

TABLE 4.1: EDWARD ALTMAN (1968) – DIRECTION OF FINANCIAL RATIOS ON PROBABILITY OF DEFAULT..35
TABLE 4.2: JAMES OHLSON (1980) – DIRECTION OF FINANCIAL RATIOS ON PROBABILITY OF DEFAULT..36
FIGURE 4.1: ILLUSTRATION OF DEFAULT PROBABILITY IN THE MERTON FRAMEWORK....................38
FIGURE 4.3: THEORETICAL RECEIVER OPERATING CHARACTERISTIC-CURVE...................................46
TABLE 4.1: THEORETICAL CONFUSION MATRIX 47
FIGURE 4.4: GENERIC DENSITY PLOT WITH CUT-OFF VALUE..48

5. DATA 51

FIGURE 5.1: DISTRIBUTION OF COMPANIES ACROSS COUNTRIES...52
FIGURE 5.2: DATA CLEANSING PROCESS 53
FIGURE 5.3: DISTRIBUTION OF COMPANIES ACROSS TIME..54

6. RESULTS..58

FIGURE 6.1: ROC-CURVE FOR ORIGINAL ALTMAN MODEL..59
FIGURE 6.2: CONFUSION MATRIX AND DENSITY PLOT IN ORIGINAL ALTMAN MODEL.....................60
TABLE 6.1: RE-ESTIMATED ALTMAN MODEL...62
FIGURE 6.3: ROC-CURVE IN THE RE-ESTIMATED ALTMAN MODEL.......................................63
FIGURE 6.4: CONFUSION MATRIX AND DENSITY PLOT IN RE-ESTIMATED ALTMAN MODEL............64
FIGURE 6.5: ROC-CURVE IN ORIGINAL OHLSON MODEL..65
FIGURE 6.6.: CONFUSION MATRIX AND DENSITY PLOT IN ORIGINAL OHLSON MODEL....................66
TABLE 6.2: RE-ESTIMATED OHLSON MODEL...67
FIGURE 6.7: ROC-CURVE IN RE-ESTIMATED OHLSON MODEL 68
FIGURE 6.8: CONFUSION MATRIX AND DENSITY PLOT IN RE-ESTIMATED OHLSON MODEL............69
FIGURE 6.9: ROC-CURVE IN ORIGINAL MERTON MODEL 70
FIGURE 6.10: CONFUSION MATRIX AND DENSITY PLOT IN ORIGINAL MERTON MODEL..................71
FIGURE 6.11: ROC-CURVE IN NAÏVE MERTON MODEL..72

FIGURE 6.12: CONFUSION MATRIX AND DENSITY PLOT IN NAÏVE MERTON MODEL........................73
TABLE 6.3: OVERVIEW OF MODELS........74
TABLE 6.4: VERIFICATION OF HYPOTHESIS 1........76
TABLE 6.5: PREDICTOR CUT-OFFS AND DIRECTION........78
FIGURE 6.13: TRANSFORMATION PLOTS FOR MV/BD AND NI/TA........79
FIGURE 6.14: LINEARITY PLOTS FOR NICHANGE, SA/TA AND TL/EBITDA........80
TABLE 6.6: CORRELATION MATRIX - PRACTICAL MODEL OF DEFAULT PROBABILITY........82
TABLE 6.7: PRACTICAL MODEL OF DEFAULT........83
FIGURE 6.15: ROC-CURVE IN PRACTICAL MODEL........84
FIGURE 6.16: CONFUSION MATRIX AND DENSITY PLOT IN PRACTICAL MODEL........85
FIGURE 6.17: PROBABILITY OF DEFAULT MODEL: THE DEFAULT MODEL OF SYNTHESIS........86
FIGURE 6.18: TRANSFORMATION PLOTS FOR NAÏVEDD........87
TABLE 6.8: DEFAULT MODEL OF SYNTHESIS........88
FIGURE 6.19: ROC-CURVE IN DEFAULT MODEL OF SYNTHESIS........89
FIGURE 6.20: CONFUSION MATRIX AND DENSITY PLOT IN DEFAULT MODEL OF SYNTHESIS........89
TABLE 6.9: VERIFICATION OF HYPOTHESIS 2........90
TABLE 6.10: OVERVIEW OF MODELS........91
FIGURE 6.21: CLASSIFICATION TREE DIAGRAM........92
FIGURE 6.22: ROC-CURVE IN CLASSIFICATION TREE MODEL........94
FIGURE 6.23: CLASSIFICATION TREE MODEL – TOP-LEFT CUT-OFF POINT........95
FIGURE 6.24: IDENTIFICATION OF MINIMUM ERROR RATE........96
FIGURE 6.25: RANDOM FOREST MODEL – TOP-LEFT CUT-OFF POINT........97
FIGURE 6.26: ROC-CURVE IN RANDOM FOREST MODEL........97
FIGURE 6.27: RANDOM FOREST - VARIABLE IMPORTANCE........98
TABLE 6.11: VERIFICATION OF HYPOTHESIS 3........99
TABLE 6.12: OVERVIEW OF MODELS........100

7. DISCUSSION........101

FIGURE 7.1: INTERPRETABILITY AND COMPUTATIONAL BURDEN........103

9. CONCLUSION........108

TABLE 9.1: OVERVIEW OF VERIFIED HYPOTHESES........109
TABLE 9.2: OVERVIEW OF THESIS' CONTRIBUTIONS........110

List of Abbreviations: Financial Ratios

Financial Ratio	Calculation Description	Origin
WC/TA	"Working Capital" divided by "Total Assets"	Altman (1968)
RE/TA	"Retained Earnings" divided by "Total Assets"	Altman (1968)
EBIT/TA	"Earnings before Interest and Tax" divided by "Total Assets"	Altman (1968)
MV/DB	"Market Value of Equity" divided by "Book Value of Debt"	Altman (1968)
SA/TA	"Sales" divided by "Total Assets"	Altman (1968)
Size	Log of "Total Assets" divided by "Adj. GNP for OECD"	Ohlson (1980)
TL/TA	"Total Liabilities" divided by "Total Assets"	Ohlson (1980)
CL/CA	"Current Liabilities" divided by "Current Assets"	Ohlson (1980)
NI/TA	"Net Income" divided by "Total Assets"	Ohlson (1980)
FFO/TL	"EBITDA" minus "Change in NWC" divided by "Total Liabilities"	Ohlson (1980)
NIChange	"NI_t" minus "Ni_{t-1}" divided by "$\lvert NI_t \rvert$" plus "$\lvert Ni_{t-1} \rvert$"	Ohlson (1980)
LevDummy	"1" if "Total Liabilities" > "Total Assets", "0" otherwise	Ohlson (1980)
NIDummy	"1" if "Net Income" < 0 for the past two years, "0" otherwise	Ohlson (1980)
MertonDD	"Distance to Default" in Original Merton Framework	Merton (1974)
NaïveDD	"Distance to Default" in Naïve Merton Framework	Bharath & Shumway (2008)
TL/EBITDA	"Total Liab." divided by "Earngs. b. Interest, Tax, Dep. and Amort."	Plenborg et al. (2017)
FFO/CAPEX	"EBITDA" minus "Change in NWC" divided by "Capital Exp."	Plenborg et al. (2017)
FFO/CL	"EBITDA" minus "Change in NWC" divided by "Current Liabilities"	Plenborg et al. (2017)
WC/SA	"Working Capital" divided by "Revenue"	Plenborg et al. (2017)

(This page intentionally left blank)

1. Introduction

On the 27th of April 2020, the prudent Danish bookstore Arnold Busck filed for bankruptcy. This came in the wake of the events facilitated by the Coronavirus disease, which to a large extent have paused business activity throughout the globe. While this public health crisis has been a shock to the economy and inevitably will carry a recession with it, the challenge of predicting what companies that are at risk of going default remains. Why was it specifically an illustrious bookstore with a history of more than a century, which was the first large company in Denmark to become insolvent?

A corporate failure creates shockwaves that will affect multiple stakeholders; employees are discharged, suppliers loose contracts, shareholders' value disappears and business lenders will not be repaid. The negative consequences are widespread. On the other hand, bankruptcies are a natural occurrence in the business landscape. To limit the negative consequences and make optimal business decisions, a stakeholder need to understand why a default happens. For this reason, a quantifiable link needs to be established between the intangible external pressures, whether they come in the form of risk inherent in operations or force majeure events, and a firm's ability to withstand these. The stakeholder can obtain such quantifiable links through financial statements where the firm's leverage, solvency or liquidity can be assessed, or he can be of a belief that these are already priced by the market and examine the firm's stock price. These two sources of information might be powerful on their own, but if employed through a probability of default model then the basis of the decision is even more informed. Concerns like this are the center of attention within the discipline of default probability, which is an integral department in financial institutions with the role of guiding lending activity.

The practical discipline of credit risk is built on top of a vast academic field, which is deeply embedded in the corporate finance literature. Both the practical and academic side of default probability is heavily influenced by their methodological underpinnings, wherefore progressions in statistical analysis drive the development of the field both amongst scholars and amongst professionals. However, it is simultaneously a discipline which up until this point has been divided in two theoretical trenches facilitating a clash of paradigms. Furthermore, the field of credit risk has to a large extent been neglecting the advancements of machine learning algorithms. Like scholars before, this thesis will address how progression can be made within default probability, however, we shift the focus to relatively untouched lands. We are interested in both the conflict between the two strands of the literature and the methodological possibilities of machine learning. In this regard, we answer the following research question:

"To what extent does the classic academic approaches to default probability have discriminative power, and through which measures can the practical approach to credit risk be improved?"

In answering this, we will showcase the classic scholars of the discipline's relevance in a modern portfolio context. Furthermore, we will scrutinize the practical approach to default probability, and formalize how it can be improved upon by looking to both paradigms of the literature. Moreover, we will apply techniques from machine learning, in order to establish a platform for discussing the neglection of algorithms in practical credit risk. As such, the goal of the research is not to induct a new universal methodology, but to conceptualize a framework where the practical approach to probability of default can be improved.

1.1 Contribution to the Literature: Why This Thesis is Interesting?

In the existing literature there is research which considers both the testing of the existing default probability models and extensive alterations in order to develop new frameworks. This thesis contributes to the literature in five distinct ways, where two are deemed main contributions and three considered supporting additions.

First of all, the thesis develops a practical probability of default (PD) framework which combines the two academic paradigms within the credit risk literature. Second, it extends the practical model of credit risk through the application of a random forest algorithm. Supporting these two contributions, the thesis provides an extensive testing of the classic academic models within the discipline. Further, the thesis expands the portfolio scope in its research compared to the pertinent literature, as the focus for those is more limited in terms of both country and industries. Finally, it discusses the precedence of logistic regression in the practical field of default probability and provide two plausible explanations herefore.

The thesis at hand is interesting to the default probability discipline both academically and practically. First, it is relevant for the academia of credit risk as it reviews the main tenants within the field in a modern portfolio, as the scope has been expanded to consider all OECD countries across multiple sectors only constrained by the availability of data. Second, the thesis is relevant to the practical discipline of default probability through its conceptualization of a new framework which incorporates both the market and the accounting theory of the field. Further, the research is relevant to practitioners of PD as it extends the practical model through machine learning and additionally discusses the implications of this methodology.

1.2 How the Thesis Will Answer the Research Question

This thesis will be structured as follows. We start by establishing our philosophy of science position, as it is fundamental for how we undertake the academic investigation. Hereafter, we introduce the theoretical framework which the research of the thesis builds upon. This framework includes both the economic theory underlying the default probability field and a review of the pertinent literature. The theoretical framework concludes with the formulation of three hypotheses, which are set out to guide the answering of the research question.

Table 1.1: Hypotheses Formulated from Pertinent Literature and Theory

#	Hypothesis
1	*The classic academic approaches to default probability have discriminative power on a modern portfolio*
2	*The practical model outperforms the academic but is improved through a synthesis of market and accounting theory*
3	*The discriminative power of the practical model can be improved through the application of machine learning*

Following this, we outline for the methodological applications utilized in the thesis. Subsequently, we outline our data collection, operationalization processes, and the calculation of financial ratios and model inputs with a basis in the theoretical framework. This brings us to our results.

Our results will be testing the three hypotheses formulated in the theory chapter. The first entails an application of three classic academic models to the modern portfolio collected for this research, with the addition of a re-estimation for each respective model. The second consists of introducing the practical approach to default probability and furthermore expanding it through a combination of the paradigms in the academic literature. The third revolves around extending the practical model utilizing the methodology of machine learning algorithms. Then we turn to our discussion.

We discuss the striking neglection of machine learning in the practical field of default probability, reflecting on algorithms' ability to reach higher levels of predictive accuracy. Herefore, we bring two plausible and non-exclusive explanations of interpretability and "institutional stickiness", respectively. This discussion contemplates the findings of the thesis and sets the implications of the research into perspective. Next, we reflect on the validity of the findings we have made. Ultimately, we conclude the research, present our contributions, and suggest avenues for further academic investigation.

Carrying the thesis throughout is its research design, which in the case at hand is two-fold. It is overarchingly a theory testing design, where it is deductive and applies hypotheses to guide the academic endeavor. Yet, it also has an empirical underpinning as it considers the practical side of credit risk, which is conceptualized in order for it to be studied in the theoretical scope. It further aims to extend the practical discipline through rationalizations grounded in the supporting theory. Having outlined the structure and the research design of the thesis, we proceed to delimit the scope of the research.

1.3 Delimitation and How We See Probability of Default

Within the credit risk literature, a distinction is made between the different input factors for calculating a stakeholder's expected loss. These input factors are termed as exposure at default, loss given default and probability of default. Of these three elements, the research conducted in this thesis will only be devoted to probability of default. We view corporate defaults as a dichotomous event, entailing that a firm has either defaulted or is still operating. The discriminative power of models employing this as the definition has been shown to be equivalent to models employing laxer definitions (Engelmann & Rauhmeier, 2011). When referring to defaults we use bankruptcy, firm failure, insolvency and default interchangeably and do not consider whether a given firm is restructured and returns to an operating state after the event. Similarly, the terms of credit risk and probability of default is used to describe the likelihood of the default event for a given firm. Here, probability of default is given as a modelled score denoting class belonging that is restricted to the range between zero and one. Throughout the thesis the only forecast horizon for which the probability of default is modelled is a one year ahead horizon. This is in line with the length of the forecast horizon used in practice (ibid.).

We further make a distinction between statistical techniques and machine learning. In some parts of the machine learning literature, regression models such as the linear and the logistic is included (Baesens, 2014). We consider regression models statistical property and delimit the machine learning field to consider techniques such as support-vector machines (SVM), classification trees (CT) and neural networks (NN). Specifically, we draw the distinction between the two fields at the point of where the machine learning paradigm is born, following the distinction of Breiman (2001). As such, regression models exist outside the context of machine learning, wherefore it is considered a statistical model. Likewise, SVM, CT, NN and their likes do not exist outside this context, wherefore those are placed within machine learning. Along these lines, we do not differ between algorithms, data science, and deep learning, and consider those subfields within the branch of machine learning. When denoting the practical approach to default probability, we are referring to credit risk departments within banks, lenders and financial institutions, which share the characteristic of being supervised by legal authorities. In this regard, the practical approach and "practitioners" will be used interchangeably. When choosing the default probability frameworks of Altman, Ohlson and Merton as applications in our research, we are aware of the neglection of later and more modern scholars that have entered the academic discipline. However, those have been chosen deliberately, as we consider them both representative and fundamental for the field of credit risk today, both academically and practically.

Furthermore, we delimit ourselves from interpreting on the reasons for *why* specific factors drive company failure, as we are occupied with modelling defaults on a larger scale. As such, we will not engage in a discussion of the underlying mechanisms and specific financial ratios, as the overarching focus is on discriminative power. Likewise, we will differ from some parts of the literature by withstraining from trying to categorize companies in rating groups, similarly to credit rating agencies. We do however expand the scope compared to large parts of the academic literature, as we consider a broader range of companies. As such, we are not delimited to a specific industry sector, country or concise time period. Instead, we consider data available from publicly listed companies in the OECD countries from 2001 to 2019. It is chosen deliberately, in order to investigate default probability models in a wider and more comprehensive setting. In this regard we highlight two perspectives. First, we are not evaluating the models chosen from the literature in their original setting. Second, we are aware that the implications of PD-models are different for a broad portfolio, compared to a limited industry specific range of companies.

2. Philosophy of Science

Philosophy of science is a discipline within academic research, which in its essence underpins the endeavors of knowledge. As such, our philosophy of science position permeates the research of this thesis. Most fundamentally, the discipline considers the nature of reality, from which several subfields stem. These include the nature of truth, the availability of knowledge and the persistence of science. Those notions are captured by the two main concepts of philosophy of science, namely *ontology* and *epistemology*. Those two are interlinked and furthermore fundamental for how we approach the study of probability of default. Ontology is in many ways predominant of the two but also the most abstract. It considers the concept of being, or mere building blocks of reality. As such, it asks the question of what the world is made of, determining the reality of the world and its availability (Moses & Knutsen, 2012). The ontological conception to a large extent leads the epistemology, even though it does not decide it completely. Epistemological considerations encapsulate the study of knowledge. Thus, it assumes what we as researchers can know and in what ways, revolving around the question of what knowledge really is (Marsh & Furlong, 2010).

We identify three major strands within the discipline of philosophy of science, including *Naturalism, Constructivism*, and *Critical Realism*, where the two former are clearly most fundamental to the science itself, as it in many ways have driven it. Here, the three frameworks differ in their respective ontology and epistemology, and subsequently in their inherent methods for research. We as researchers consider ourselves as critical realists. In this regard, we consider the framework to be "worn as a skin rather than a sweater", as it underpins our academic research throughout (Marsh & Stoker, 2010). Thus, we are unable to compromise its theoretical presumptions, as it is a skin which cannot be taken off. For that reason, it should be introduced in relation to its counterparts.

2.1 Critical Realism

The critical realism (CR) framework will theoretically form our research conducted within the default probability discipline. It is a relatively young strand within philosophy of science and originates from Roy Bashkar (1975). In many ways, it places itself in between the two classical philosophies of science, naturalism and constructivism, as it draws upon the ontological and epistemological assumptions of both. Doing so, critical realism is an approach to research which considers reality to be explicit, whereas knowledge is both "limited and fallible" (Scott, 2005). Critical realism draws its ontology from the strand of naturalism, which considers reality as being existent, real, and independent of our perception of it. This constitutes the *realist* part of CR. For the research at hand, it means that companies exist, operate, report accounts and trade on markets independently from our research of them. Similarly, models of default probability are being build and utilized independently from our bare knowledge of them. This is in contrast to the ontology of constructivism, which perceive the world as being socially constructed, where reality only exists as we give it meaning (Moses & Knutsen, 2012). Yet, critical realism takes the notion of reality a step further than the naturalist framework, as it considers the world as being both deep and stratified denoted the "intransitive" dimension (Buch-Hansen, 2012).

Reality takes place in three distinctive levels, or domains. The first is the actual domain, where events are taking place. Here companies experience decreasing sales that lead them to defaulting in the ultimate case. Above is the empirical domain, where the actual domain is revealed and can be studied. It is in this domain that accounting books are published. Underlying both of those levels is the deep domain, in which structures determine the above levels. It is in the underlying domain that mechanisms such as consumer behavior shape what companies eventually default. For the thesis at hand, our research utilizes observable characteristics of the empirical domain to develop models that most accurately predict events in the actual domain, and exclusively in the actual domain. As such, we do not claim to be capable of modelling what mechanisms and structures that take place in the underlying domain, as a PD-model does not predict consumer behavior or the like. Thus, there may be interconnecting mechanisms in the deep domain, which is foreign to both the models developed and the research itself. Being true to the framework of critical realism therefore implies that we can never claim to have found complete or infallible truths within the ontological sphere (ibid).

Figure 1.1: The Three Realms of Critical Realism

Empirical Domain

Observable accounting and market information

Actual Domain

Where events are taking place: Company default

Deep Domain

Underlying mechanisms influencing the actual domain

Source: Personal collection

This serve as a segway into the epistemology of critical realism. Knowledge is never definitive, and truth is not obtained through relationships that remain unproven (Scott, 2005). As such, CR part ways with the epistemological part of naturalism. On the contrary, explanations of events are deducted from plausible justifications of events, and thus attempts to describe the world are inherently fallible. This is the *critical* component of CR. The epistemological notion of the framework is not however directly equal to that of constructivism, as the latter approach emphasizes context and social creation as underpinnings of its concept of knowledge (Buch-Hansen, 2012). Instead, the knowledge that research acquires is subject to the properties that the setting and data constitute. Wearing critical realism as a skin, our research does not claim to provide PD-models that are superior in all empirical settings, but rather models that uncover some of the mechanisms facilitating company defaults. Those models can be both convincing and probable given the data set applicated in the research, however they can never be considered complete.

2.2 Mode of Reasoning and Appropriateness of Critical Realism

As noted above, the academic endeavors made in critical realist research is inherently shaped by the philosophy of science position. For the thesis at hand, the ontological and epistemological assumptions of CR serve as a catalyst for contributing to the literature of default probability. First, we have retrieved accounting data and market data for a large portfolio of companies, which are observable information current in the empirical domain. We utilize this to develop different models of default probability, which are events taking place in the actual domain. The mechanisms taking place in the deep domain that influence the events of the actual domain, is outside the scope of our research. It could be argued that those structures are inherent in the significance of what financial ratios that impact defaulting companies the most, however, this avenue of reasoning is not the aim of the research.

Second, our mode of reasoning is that of abductive reasoning which revolves around retroduction. This method relates to the epistemological assumptions of critical realism and concerns finding the best possible explanation for a relationship given the information available. It aspires to acquire more complete knowledge of the reality through constantly questioning the application of techniques and the nature of data (Belfrage & Hauf, 2016). This proposition comes across in our research, as we search for a PD-model that can explain the defaults of the portfolio the better. Third, the CR strand of philosophy of science is denoted for allowing any range of research methods. This is an extension of its epistemology and the intrinsic quest for more complete knowledge. As such, the CR approach allow us as researchers to both apply a collection of statistical models, and subsequently include a political theorem that originates in institutionalism to discuss the status quo (Moses & Knutsen, 2012). Ultimately, it is central to severely underline that we do not claim to test or model all possible explanations for defaults, but rather test, develop, and discuss PD-models in the empirical vacuum that our portfolio constitutes.

3. Theoretical Framework: A Review of Default Probability in Theory, Academia and Practice

The theoretical framework of this thesis will be structured a presentation of the fundamental core of the research at hand. The section will be divided into seven parts that naturally build upon each other. The first introduces the accounting-ratio approach to estimating probability of default, as this approach laid the foundation for the academic discipline of default risk. Second, it will cover the structural approach, which introduces a market view to the probability estimation. For each of the two approaches, we will first present its theoretical underpinnings and secondly review the pertinent literature. The theoretical underpinnings will shed light on how the specific school is relevant for evaluating firm failure whereas the literature review will concern the methodological developments. The third part will outline how the underlying theoretical notions of these two paradigms clash with each other, as researchers within the field have failed to agree on one overarching approach. The fourth part will have a practical focus and cover the basis for approaching probability of default amongst practitioners. The fifth part will cover how inventions in machine learning and artificial intelligence algorithms have entered the probability of default discipline, as well as the theoretical foundation for practitioners' models being outperformed by machine learning. Lastly, a section will outlay the gaps in the pertinent literature, which will serve as a catalyst for a new framework to model credit risk, the "Default Model of Synthesis". The section culminates in the formulation of three hypotheses. These hypotheses come on the back of the review of the pertinent literature and theory and are formulated in order to guide the overarching research question.

3.1 An Accounting-based Approach to Default Probability

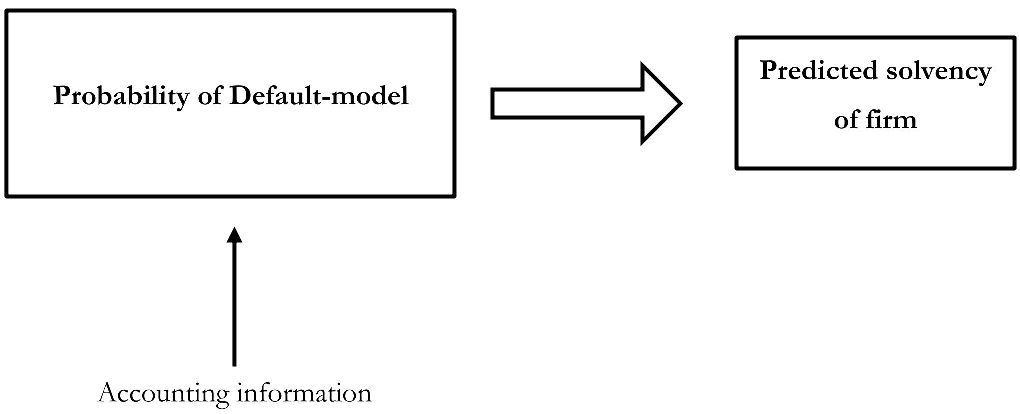
The cornerstone of the default probability discipline is accounting information, as it constitutes the foundation for both the practical and the academic field to build upon. Therefore, it is natural to introduce this position first.

3.1.1 The Accounting Theory of Credit Risk

At the very core of a company is liquidity. Liquidity denotes the given firm's ability to meet its liabilities. As such, if the company is short of liquidity, it will inevitably default (Plenborg et al., 2017). A default hurt potential shareholders, equity holders and engaged stakeholders of a company. However, those actors have limited knowledge on the individual company's liquidity and financial information, wherefore authorities have established accounting standards. Thus, the purpose of accounting is to minimize the information asymmetry between stakeholders. This indicates the relevance of looking to a company's accounting books when evaluating its probability of default.

Along the same line, Beaver defined a company as a reservoir of liquidity, where the operation includes in and out-flows of the reservoir. Ultimately, if the reservoir is emptied, the firm is defaulting (Beaver 1966). As such, the financial statements are witnessing the status of the reservoir. It is a clear illustration of what Plenborg et al. denote as "financial health" (2017). The financial statements of a company are testimonial to its financial health, wherefore the interest into whether the given company will go bankrupt should start here. In its essence, the financial statements contain indications of whether a company will be capable of meeting its obligations. It means that both the theoretical motive and the mere raison d'etre of accounting underpins the role of financial ratios in probability of default modelling. As financial ratios are computed from the raw statements of a given company, those are reflecting the company's financial health in relation to its liabilities. Thus, there is a theoretical foundation for basing models of credit risk on financial ratios. An illustration of the principle behind an accounting-based model is visualized below.

Figure 3.1: Probability of Default Model: Accounting Theory



Source: Personal collection

3.1.2 Literature Review of the Accounting-based Approach

William Beaver established the discipline of probability of default in an academic context, with a univariate approach (1966). According to Beaver himself, the single best ratio predicted default as accurate as multi-ratio models, which considered several accounting perspectives. The Beaver framework used a methodology where 79 failed and non-failed firms were matched on asset size and utilized to draw conclusions on what ratios could predict defaults. The study found that the *cash flow/total liabilities* ratio was the single most powerful predictor in terms of accurately predicting default in a period of five to one year before the distress or non-distress situation. While Beaver's findings were both simple and limited in its scope of work, it pathed an academic highway for scholars to follow and expand upon in the coming years.

Those following years were characterized by scholars utilizing a similar approach to Beaver, specifically with the methodology of multivariate discriminant analysis (MDA), which further allows to combine different financial ratios. These approaches were primarily dominated by the work of Altman, who by peers are considered one of the main scholars of the discipline, driven by two aspects. One, the simplicity and applicability of his model. Two, the several re-visitations of his framework. In his original work, 33 defaulting companies were matched with 33 solvent firms and observed over a 20-year period (1968). Like Beaver, Altman matched the companies of his study on asset size. Ultimately, his research concluded in a model where five accounting ratios multiplied by their respective coefficients gave a "Z-score", which popularized his name in the overall finance literature.

Although Altman received wide recognition for formalizing a simple and applicable approach to the study of default risk, his research left a wake of criticism to the MDA-approach and its statistical assumptions, which Altman violated. This criticism materialized into the application of new methodologies, specifically conditional profitability models. Within this strand of the ratio-focused literature, Ohlson developed a logistic regression model based on nine parameters including two dummy variables (1980). Ohlson's study included more than 2,100 firms, whereof only 105 were bankrupt cases. As such, Ohlson pioneered the literature in two major ways. First, he introduced a much larger sample size in the development of default risk models. Second, he moved away from a sample where the number of distressed and non-distressed firms were evenly distributed and matched on asset size. Both aspects were enabled by moving away from the MDA methodology. This development was levered upon by further scholars of the conditional profitability models through the 1980s (Hamer 1983; Keasy & Watson 1986).

3.2 A Market based Approach to Credit Risk

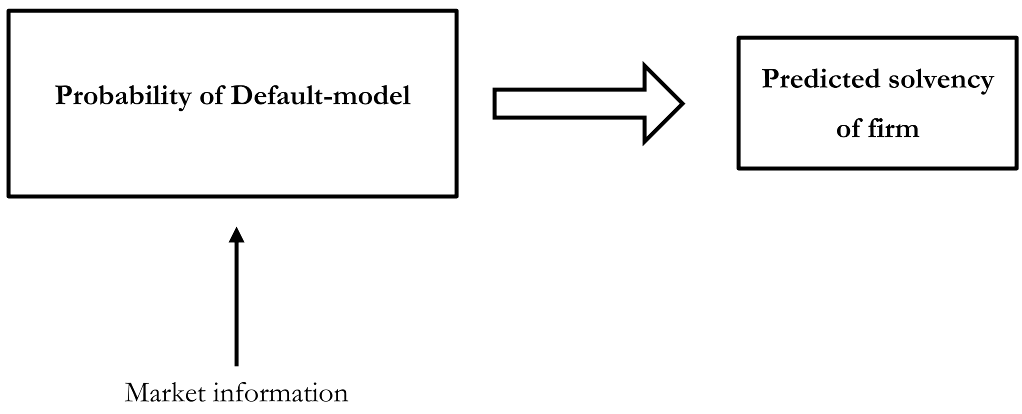
While the accounting ratio-focused approach to modelling credit risk constituted one way to model probability of default, a simultaneous market-focused literature on credit risk developed with a cornerstone in option-pricing.

3.2.1 The Structural Theory of Credit Risk

Efforts towards an increased theoretical underpinning of predicting probability of default began in the mid 1970's. This change in paradigm was inspired by the option pricing theories of Black and Scholes (1973) and the frameworks are often referred to as contingent claims models, market-based models or structural models. A general belief permeating the framework is that the market already reflects the information contained in the financial statement. Additionally, financial statements are backward looking describing historical figures whereas the market is inherently forward-looking and working through future expectations (Hillegeist et al., 2004).

A wide range of scholars have employed different varieties of these models including the first developed by Merton (1974). The original Merton model and the assumptions behind are derived from a firm's capital structure and the relative market development between the assets, equity and liabilities. Merton shows that the probability of default can be inferred from the market value and volatility of these two sides of the balance sheet (Merton, 1974). Following this framework, the market value of a firm's equity is identical to the price of a call option on the assets with a strike price of the face value of debt. This is because shareholders are residual claimants on the firm's assets after debtholders have been paid. When the option expires and debt repayment is due, the firm is insolvent if the market value of the assets is exceeded by the notional value of debt. The market value of equity will then be zero. As the shareholders have in effect sold the assets of the firm to the debtholders, they will let the option expire and default if the assets are worth less than the liabilities (Löffler & Posch, 2011). The primary assumptions behind the model are questionable in practice. The first underpinning of the model is that a firm's total debt can be viewed as a zero-coupon bullet bond. Secondly, a corporation can only default upon maturity of the bond (Merton, 1974). An illustration of the principle behind market-based default probability models is shown below.

Figure 3.2: Probability of Default Model: Market Theory



Source: Personal collection

3.2.2 Literature Review of the Structural Approach

As a response towards the naïve assumptions put forward by Merton a series of literature emerged. Geske (1977) relaxed the zero-coupon bond assumption and instead put forward a framework in which debt is viewed as a compound coupon paying bond. The bond is hence no longer viewed as being finite per se, rather to finance the ongoing coupon payments and to buy the next option, shareholders are required to issue new equity at each coupon date. The firm will default if shareholders decide against selling new equity as coupon payments are then not met. Shareholders will fail in raising new equity when the market value of equity after the coupon payment is less than the value of the payment (Geske, 1977). In this way, Geske still adheres to the timing assumption of Merton in the sense that default occurs at a fixed occasion, but instead of default occurring at expiry of the option it now occurs at the coupon date.

The timing assumption is also challenged by Black and Cox (1976) as they allow default to take place prior to maturity. In this framework debt is still viewed as a perpetual bond with coupon payments. However, in their work they introduce the concept of boundaries or default barriers. The lower boundary is to be thought of as a level of firm value at which default will take place. The default barrier may be given exogenously by the contract with the debtholders or decided endogenously by the shareholders as an optimal decision problem (Black & Cox, 1976). Following this, the shareholders will choose the default barrier as the point in time where outstanding debt is minimized. The reasoning behind is that from the shareholders perspective and given the current position of the firm, the value of the option is influenced by multiple scenarios in conjunction. These scenarios include the value at maturity if the firm has not defaulted before then, the value if the firm defaults at the upper or the lower boundary and the value of the residual claims.

Brockman and Turtle (2003) builds on the default barrier work of Black and Cox and argues that the standard view of equity presented by Merton as a path-independent option framework is wrong. Path-independency in this sense relates to the fact that the payoff of the option only depends on the underlying asset value at maturity, i.e. that default can only occur at this fixed point in time. Rather, equity should be viewed as a path-dependent down-and-out call option on the assets (Brockman & Turtle, 2003). Their option framework explicitly incorporates a barrier that causes the termination of the option when breached, i.e. directly leading to default. This implies that the specific rise or decline in asset value throughout the lifetime of the option affects the payoff and hence how default probability is theoretically underpinned. Brockman and Turtle are consequently modelling and empirically testing the theoretical notions put forward by Black and Cox.

Bharath and Shumway (2008) examined whether the structural framework provided by Merton was a sufficient way to forecast probability of default. Their research came on the back of the numerous criticisms related to assumption of the Merton framework. However, their way of addressing the assumptions behind the model differed from prior scholars in the field. In their paper they hypothesized that if the Original Merton Model was true, then it should be impossible to improve its forecasting abilities. They recognized that the functional form of Merton's framework could not be completely reduced to a combination of simple variables, but that improvements within the approach are possible. To test this, they developed a naïve version of the original model that reduced some of the methodological complexities while maintaining the same discriminative power (Bharath & Shumway, 2008). The results of their research indicated that the Merton model is an important but not complete model, and that the structural approach to default probability provides useful guidance for future predictive models.

3.3 The Literature on Probability of Default: A Clash Between Paradigms

A clash between the two paradigms of the accounting-based approach and the structural approach to probability of default exist in the academic literature. This clash is present with respect to both the theoretical underpinnings and empirical results of the two approaches tested against each other. An array of researchers compares their version of a structural model against the accounting-based models and find that their structural models are superior (Tudela et al., 2003; Hillegeist et al., 2004). Another strand of the literature finds that a simplified version of the structural models outperforms their deeper theoretically supported counterparts (Bharath & Shumway, 2008; Jackson & Wood, 2013). There also exist empirical tests of structural models where the accounting-based approaches are superior especially for shorter forecast horizons (Reisz & Perlich, 2007).

Researchers advocating for the structural approach being superior have several theoretical arguments. Here, Vassalou and Xing (2004) present arguments concerning the inputs used in accounting models. As the inputs stem from financial statements they are inherently backward-looking, describing either just a snapshot in time of the balance sheet or the historical figures of the income statement. This critique is mirrored and further expanded in Hillegeist et al. (2004). They add that financial statements by design is formulated on a going-concern basis, why the ability of accounting-based models to predict a future event such as bankruptcy is limited. As structural models are based on market data, they are aggregating a larger range of information than their counterparts. The academics also ascribe the inferiority of the accounting models to the lack of a volatility measure. They argue that a strength of the structural models not prevalent in accounting models is their ability to vary the weight given to leverage by employing a measure of volatility (Hillegeist et al., 2004).

Another range of scholars aim at refuting the above points of critique. They argue that the structural models cannot be superior due to their theoretical foundations, as the assumptions behind are heavily violated in practice (Duffie & Lando, 2001). Duffie and Lando argue that the models rely on the stock market to accurately reflect all the information included in the financial statement and suggest that this is not the case in practice. This underpins the findings of Reisz & Perlich (2007) that the accounting-based models are superior for shorter forecast horizons. They however add, that the natural forward-looking characteristic of financial markets is why structural models are outperforming for longer forecast horizons. As a contrast to the structural argument regarding the inclusion of a volatility measure, Reisz & Perlich implies that structural models tend to overestimate the probability of default for highly levered, highly volatile firms and vice versa for non-levered, non-volatile firms. Agarwal & Taffler (2008) point out that the main take-away from the empirical studies of structural models should not be a relative superiority in performance.

Rather, the studies show a poor performance of the accounting models chosen for comparison than a convincing performance of the structural models.

3.4 The Practical Approach to Probability of Default

The practical approach to probability of default is outlined by practitioner Evelyn Hayden from BAWAG PSK (Engelmann & Rauhmeier, 2011). She presents two major properties of the internal scoring models for corporate exposures. First, the practical approach to default probability relies predominantly on financial ratios. As such, the practitioner's approach originates from the accounting ratio strand of the literature. Second, the primary model utilized in the practical approach is the logistic regression model, which stems from Ohlson's contribution to the field. The logistic regression model enjoys precedence in practice, as it is easily interpretable, yields a direct probability of default, and enables an easy analysis of the potential explanatory variables (ibid).

It is from the practical side of PD emphasized that the binary classification outcome of default versus solvent is a powerful predicting tool. Furthermore, the Basel Committee on Banking Supervision underlines that a one-year horizon is a "common habit" in practice and is supported from the regulatory side (1999). When building the optimal logistic regression for modelling default probability, the practical approach starts by carefully investigating the statistical relationships between explanatory variables. Subsequently, it aims to end with a parsimonious model of only a few, but highly explanatory, regressors, decided through a backward elimination process.

Another contribution to the practical side of credit risk is that of the three big credit rating agencies, namely Standard & Poor's, Moody's, and Fitch Group. These have through their law-cemented dominance within credit rating developed the most sophisticated models of default probability. An example hereof is the Moody's methodology of 2000, which combines both the accounting ratio-paradigm and the market paradigm of the PD-literature (Sobehart & Stein, 2000). Along these lines, those institutions are considered to be extremely advanced in both using the market approach and incorporating machine learning models. As such, pushing the methodological boundary in order to constantly maintain the utmost advanced position within the discipline (Baldassarri & Chen, 2016). However, we as researchers do no not consider the methodologies of the "Big Three" representative of the practical approach to default probability. This is explained by them being a different beast compared to the average institution with an internal rating model. Furthermore, these credit rating agencies are intermediaries rather than lenders, making them inherently different in their approach to credit risk as they are not under legal supervision similarly to banks. Thus, we assume the propositions of Hayden in Engelmann and Rauhmeier (2011) to be representative of the practical approach to default probability.

3.5 Machine Learning

The difference between statistical models and machine learning are often given through the distinction between their respective purposes. Statistics are used to infer relationships between variables, whereas machine learning are used to reach the best predictions (Baesens 2014). If this distinction is taken for granted, it underlines the relevance of applying machine learning in probability of default modelling. In practice, credit risk is a discipline which revolves around predicting company defaults most accurately.

3.5.1 Machine Learning Theory

The theoretical foundation for machine learning is to some extent inspired by the functionality of the human mind. It contains the ability to learn from previously observed incidents and utilize the experience for future decisions. Like a child learning the harm of fire by putting its hand on the flame of a candle. This is what sets machine learning apart from statistics (Shalev-Schwartz & Ben-David, 2014). More specifically, methodologies of machine learnings are given the ability to learn and adapt accordingly. The process of learning from memorization carries over into inductive conclusions. Those conclusions are the inference part of machine learning. Furthermore, they are used to make generalizations, where a given model's previous encounters with observations can be utilized for extracting knowledge about the data at hand.

On the applicability of machine learning, two theoretical underpinnings are present. First, one of the major uses of machine learning is that of supervised learning. It is problems where a model is initially given the information on what result to look for, such as finding relationships resulting in company default. This application is easily comparable to that of building statistical models, as it involves explaining y given some information x. Second, within supervised machine learning is classification problems, which is covered by models that are yielding categorical responses. Likewise, this application is easily comparable to a statistical model such as the logistic one, as it has a binary outcome.

Another theoretical property of machine learning is that it is considered applicable when a task is too complex to program through traditional computer science (Shalev-Schwartz & Ben-David, 2014). This relates to the ability to learn from previous experiences and is utilized for instance in online search engines and recommendation systems. This segways into machine learning being capable of reaching higher predictive power than its statistical counterparts. By these means, the machine learning methodology fits well in the discipline of default probability as a portfolio can reach extremely large sizes with numerous interconnected relationships. Furthermore, those portfolios might have changing characteristics as the economy go through fluctuations, and the need for recognizing new patterns arise over time.

3.5.2 Machine Learning in the Credit Risk Discipline

Despite the fact that machine learning algorithms have a tradition with supervised classification problems that spans across several decades, the literature within probability of default-modelling is relatively sparse. Charitou et al. built specifically on Ohlson's logit model and expanded the framework by introducing the methodology of neural networks to UK-based firms (2004). Neural networks consist of inserting input factors, which are broken down to an arbitrary number of abstraction layers that each focus on a minor part of the data (Baesens 2014). The layers of abstraction connect through previous experiences and give a result based on the intended product of the model. With this application, the researchers reached an high accuracy on a hold-out sample. However, in their research they utilized a limited sample size where distressed and non-distressed firms were equally split. This methodology proposes clear limitations, wherefore the implications of the study are restricted, although it introduced a new approach in the credit risk literature. The latter is showcased by how the approach have been followed by other scholars such as Addo et al. (2018) and Petropoulos et al. (2020).

Another strand of machine learning that have gained some precedence within the literature of default probability is that of classification trees (Jackson & Wood, 2013). Like in the academic endeavor of Charitour et al., the tree models extend the theoretical framework of accounting ratios, such as the one of Ohlson (1980). A decision tree come in the visual form of a tree, where it from its root node, or starting point, make binary splits of the variables in order to segregate the classes, default and solvent, most efficiently. Ultimately, decision trees return a cut-off point, which gives the best separation between the two groups, wherefore its methodology overcomes one of the challenges for statistical models within default probability. One of the major decision tree contributions to the literature of PD is that of Bastos (2008). He achieved above 80% area under the receiver operating characteristic curve for some datasets. Similar results are gained by Chang et al., who developed credit risk assessment models for short-term defaults (2014). However, three perspectives are present for machine learning within the discipline of default probability. First, while data science approach to PD is present, it is nonetheless a minor part of the literature compared to the traditional accounting and market models. Second, it mainly builds upon the accounting-ratio tenant of the credit risk literature. Third, it shows to have won relatively little ground in the practical approach to PD, even though it is established to show more discriminative power.

3.6 Gaps in the Literature

We have outlined the most central tenants of both the literature and the theory underlying the field of default probability. Prior to identifying gaps in the literature and clarifying how the research at hand progresses, a summary of the pertinent theory and literature is presented. It should be underlined that these scholars highlighted by no means represent the field completely. Likewise, they only constitute a minority of the scholars reviewed in this thesis. However, we deem them most central to the discipline of PD as a whole, wherefore they create the foundation from which we will conduct our research.

Table 3.1: Overview of Pertinent Theory and Literature for the Research

Framework	Methodology	Theory	Country	Portfolio
Altman (1968)	*MDA (Linear)*	*Accounting*	*US*	*66*
Ohlson (1980)	*Logistic Regression*	*Accounting*	*US*	*2163*
Merton (1974)	*Option Pricing*	*Market*	*US*	*No Portfolio*
Bharath and Shumway (2008)	*Simple Option Pricing*	*Market*	*US*	*Not Disclosed**
Machine Learning PD	*Class. Tree/Neural*	*Mostly Accounting*	*US/UK*	*Generally small portfolios*
Practical Model (Hayden)	*Logistic Regression*	*Accounting*	*Local*	*Varying*

*Bharath and Shumway (2008) did not disclose their portfolio size. However, they did present that it consisted of 1,449. defaults.
Source: Personal collection

Based on both the theory and literature examined in this section, we identify two major gaps in the academic and practical approaches to default probability. These gaps present areas of the credit risk discipline that is relatively unexposed or awaits further uncovering.

First, the contrast between the two paradigms of the academic approach PD is continuously stark. As such, additions to the literature seem to fall into either the trench of accounting ratios or the one of structural models, where little research try to combine the two strands. The neglection of such a combination is the first gap. Second, there is gap related to the application of machine learning in the default probability discipline, both in an academic and practical context. Despite the superior classification properties of this distinct science, it is not granted the precedence its results demand. This is the second gap.

3.7 Formulation of Hypotheses

In order to guide the research question: *To what extent does the classic academic approaches to default probability have discriminative power, and through which measures can the practical approach to credit risk be improved?* We formulate three hypotheses to examine both the academic and the practical field of PD. Those hypotheses are grounded in both the pertinent literature and the theory surrounding this research, wherefore they are a natural extension hereof.

The first hypothesis addresses the initial part of the research question, which considers to what *extent the classic academic approaches to default probability have discriminative power*. This hypothesis is drawing upon both the prior literature review and the theory. Here we identify the frameworks of Altman, Ohlson and Merton as representative of the classic academic approaches within the academic field of PD. The above theoretical framework supports that there furthermore is a theoretical foundation for these approaches. However, as these models are somewhat ageing, we are motivated to evaluate these frameworks on our portfolio of companies, in order to investigate the theoretical classification strength in a modern empirical context. Doing so, we both test the frameworks of Altman, Ohlson and Merton in their original form and through re-estimations. This part of the research is academically motivated and lead us to formulate our first hypothesis: *The classic academic approaches to default probability have discriminative power on a modern portfolio.*

Table 3.2: First Hypothesis Formulated from Pertinent Literature and Theory

#	Hypothesis
1	*The classic academic approaches to default probability have discriminative power on a modern portfolio*

Subsequently, we approach the next part of the research question, which contemplates *through which measures the practical approach to credit risk can be improved.* In this regard, the thesis intends to build upon the practical knowledge that supports how the logistic regression model is the primary tool for practitioners. Here, it first aims to establish that the practical approach to PD-modelling is superior to the academic, as it is not operating within a predetermined framework. Having done so, we claim that practitioners can improve the discriminative power of their model through a combination of the two strands of the probability of default literature, namely the financial ratio approach and the market approach. Thus, this claim also builds upon the theoretical framework underlying these approaches. We examine the practical model in its traditional form in contrast to one combined with market theory, wherefore this part of the research is driven by the practical approach to default probability. In doing this, we address the first of the two gaps identified above. As such, we formulate our second hypothesis: *The practical model outperforms the academic but is improved through a synthesis of market and accounting theory.*

Table 3.3: Second Hypothesis Formulated from Pertinent Literature and Theory

#	Hypothesis
2	*The practical model outperforms the academic but is improved through a synthesis of market and accounting theory*

Ultimately, we investigate another perspective of the second part of the research question: *through which measures the practical approach to credit risk can be improved*. This part of the research comes on the back of the evaluation of the classic approaches to credit risk, and the subsequent study of the practical model of default probability. Here, we build our research on the theory suggesting that machine learning models can achieve superior results in comparison to traditional statistical models. We intend to place our research alongside the findings of the sparse machine learning applications in the PD-literature by employing a random forest algorithm. Overall, this part intends to improve the discriminative power of the practical model through an introduction of machine learning, wherefore it is methodologically driven. By these means, we confront the second of the two identified gaps in the literature, and formulate our third hypothesis: *The discriminative power of the practical model can be improved through the application of machine learning.*

Table 3.4: Third Hypothesis Formulated from Pertinent Literature and Theory

#	Hypothesis
3	*The discriminative power of the practical model can be improved through the application of machine learning*

As previously mentioned, the hypotheses are formulated in order to guide the research of the thesis. While the thesis progresses, the hypotheses will be evaluated and either rejected or verified. In total, we have formulated three hypotheses.

Table 3.5: Hypotheses Formulated from Pertinent Literature and Theory

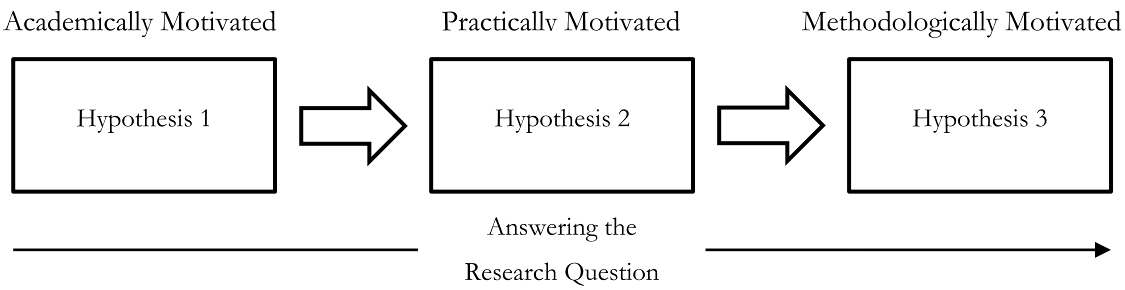
#	Hypothesis
1	*The classic academic approaches to default probability have discriminative power on a modern portfolio*
2	*The practical model outperforms the academic but is improved through a synthesis of market and accounting theory*
3	*The discriminative power of the practical model can be improved through the application of machine learning*

As stated above, the three hypotheses have each of their motivation. Where the first relates predominantly to the academic approach in the probability of default discipline. In the second hypothesis, the lens shifts to cover the practical approach to default probability.

Figure 3.3: Hypotheses Answering the Research Question

For the latter hypothesis, the motivation is primarily methodological. The motivations are summarized in the table below. The progression of the thesis is as such that after we have examined its three hypotheses, we have established a platform, on which we can discuss the further feasibility implications of using machine learning to model default probability in practice.

Figure 3.3: Hypotheses Answering the Research Question



Source: Personal collection

4. Methodological Framework

The methodology chapter of this thesis will be split into six parts and comes on the back of the review of the literature and theory section. The first three will present the methodologies of Altman, Ohlson, and Merton individually. These are the existing theoretical frameworks which the thesis will replicate and re-estimate on its dataset. The fourth part will dissect the methodology applied by practitioners to develop logistic regression PD models. The fifth section will outline for the machine learning methodology of classification trees and random forests, which we will transcend into the default probability field. The last section will outline different evaluation metrics that the thesis will utilize to compare its models, which are metrics drawn from both the literature and practice.

4.1 The Methodology of Altman: Multivariate Discriminant Analysis and Linear Regression

When Edward Altman initially presented his first model for predicting default of companies, his methodology was strongly dictated by the precedence of other scholars, such as William Beaver (1966). As such, he developed a five-factor model constituted by chosen business ratios, where coefficients were determined through multivariate discriminant analysis. Through arranging companies' features, i.e. their financial ratios, in vectors, the methodology assigns the features coefficients such that it segregates the two groups most efficiently. The two groups are here a group of non-default and a group of default companies respectively. The response variable of Altman's framework is denominated the *Altman Z-score*, which is a result that can be translated into a credit rating. Here, Altman suggest that a Z-score of less than 1.81 indicates that a given company is looking into default, whereas it alternatively is in the "safe zone" with a score above 2.99. A Z-score in between those thresholds indicates red flags for the company (Altman, 1968). The original model of Altman, which we will apply to our dataset is the following:

$$Z = 1.2\,X_1 + 1.4\,X_2 + 3.3\,X_3 + 0.6\,X_4 + 0.999\,X_5 \qquad (4.1)$$

While the MDA-approach shares properties with the linear regression, namely that it tries to fit a linear function to the dataset, it is further specifically useful for the purpose of separating two groups from each other. The MDA inherently tries to fit the coefficients that segregate the defaults from the solvent companies, however, it is also here that the approach meets its limitations. The MDA requires a balanced portfolio of an equal number of default and non-default companies. Altman adheres to this property of the methodology by taking a sample of 33 bankrupt industrial companies and match them with 33 solvent (ibid). Altman matched the companies on asset size, in order to fulfil another inheritance of the MDA. As the methodology reduces the dimensionality of two groups to one through fitting a linear function, it requires equal variance-covariance of the features in the two groups. Altman deliberately violates this assumption, which undermines the approach to some extent - a concern which have been raised by other scholars of the credit risk literature (Ohlson, 1980).

In order to overcome the problems inherited by the MDA-approach, we estimate a linear regression model on the response variable of default. As such, the Altman Z-score re-estimation conducted in this thesis will follow a different methodology than the original one. It is better suited with the theoretical assumptions of the two statistical models, as our portfolio is neither balanced between defaults and non-defaults or is exhibiting equal variance-covariance matrices between the groups. However, this linear model approach to estimating Altman's framework is supported by Engelman and Rauhmeier who classifies it as a linear regression model (2011). Thus, the re-estimated coefficients of Altman will be done following a regression of the type:

$$Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \tag{4.2}$$

Here, the Z-score also denotes credit score which can be translated into a probability of default. However, it will be on a different scale than Altman's original coefficients, as the re-estimated model regress a binary response variable of either 0 or 1 onto the financial ratios determined by Altman's framework. However, as the propositions for estimating a binary outcome with a linear regression, the response variable can take values outside of the 0 to 1 range.

This raises another problem, which is shared by both the re-estimated model and the original Altman model. The Z-score approach to probability of default aims to estimate a credit score, which can be translated into a credit rating, rather than predicting either default or solvent. As such, Altman diverts from the classic paradigm within the literature, and also the definition of PD-estimation adopted by the research at hand, which treats the dependent variable as a dichotomous, discrete and non-overlapping classification (Jackson & Wood, 2013; Agarwal & Taffler, 2008). Following several scholars of the pertinent literature we overcome this problem by translating the Z-score into a probability of default, through inserting it into the logistic cumulative distribution function of the form:

$$Probability\ of\ Company\ Default = \frac{1}{1+e^{-z}} \tag{4.3}$$

Following this we also turn around the coefficients of the model, following the procedure of other scholars such as Hillegeist et al. (2004). In Altman's original framework, a larger Z-score indicated a more solvent company position, whereas the value of 1 in our dataset indicates a defaulted company. Through this operation we organize the coefficients so they are more comparable across the two versions of the Altman framework, and to the other models developed by the research. Furthermore, we assume that the comparability between the two models is uncompromised, although they are estimated through different methodological approaches and result in Z-scores on different scales. First, we claim that it is rather a question of how the direction of the coefficients of the model impact the probability of default than it is the size of the specific coefficient. Second, we underline that the ultimate unit of comparison is one model's ability to classify defaulting companies successfully. Third, we highlight that the re-estimation model has been developed from the original Altman framework, building solely on the financial ratios as predictors presented herein (Altman 1968).

The five financial ratios in Altman's framework is chosen deliberately to capture different aspects of business operation. As such, Altman's variables are chosen among a larger pool of ratios, in which he examines the interrelationship between them (ibid). It is done in order to prevent that ratios capture similar effects. The ratios chosen by Altman were 1) *Working Capital/Total Assets*, 2) *Retained Earnings/Total Assets*, 3) *EBIT/Total Assets*, 4) *Market Value of Equity/Book Value of Liabilities, 5*) *Sales/Total Assets*, and represented the categories of liquidity, profitability, leverage, solvency and activity. Altman indicates a direction of each ratio, in terms of which effect it has on the probability of firm failure, where each is determined according to economic theory. For all five ratios it rules that it should be accompanied by a negative sign, such that an increase in that given ratio should decrease the probability of default, ceteris paribus. Exemplified through two companies that operate pari passu the one with a larger *retained earnings/total assets* ratio would be classified as less likely to default. By these means, the ratios of Altman including their effect on probability of default can be summarized in the following table.

Table 4.1: Edward Altman (1968) – Direction of Financial Ratios on Probability of Default

Increasing Effect	*Decreasing Effect*	*Divergent Effect*
	Working Capital / Total Assets	
	Retained Earnings / Total Assets	
	EBIT / Total Assets	
	Mkt. Value of Equity / Bk. Value of Liabilities	
	Sales / Total Assets	

Source: Personal collection

4.2 The Methodology of Ohlson: Logit Model

James Ohlson reacted in 1980 to Altman's original framework by estimating the probability of default with a logistic regression. Not only did Ohlson overcome the problems inherited by the MDA, he also employed a methodology that is ideal when dealing with a dichotomous classification problem (Ohlson, 1980). The logistic regression response variable takes the values of either 0 or 1, where a defaulted company should be designated the value of 1, and a solvent company should be predicted with 0. In the original framework of Ohlson, he built upon Altman even further, as he continued the Z-Score as the product of his own model and termed in an O-Score. However, the Z-score equivalent of his framework were to be entered into a logistic equation, equal to the one we applied in our re-estimation of Altman's model. Ohlson further expanded the literature of PD by introducing new ratios, including two dummy variables. Thus, he increased the comprehensiveness of the financial ratios going into a probability of default model. He estimated the coefficients of the variables with a logistic regression function. As such, the Ohlson model's statistical approach can be captured with the following:

$$\begin{aligned} O = &-1.32 - 0.407X_1 + 6.03X_2 - 1.43X_3 + 0.0757X_4 \\ &-2.37X_5 - 1.83X_6 + 0.285X_7 - 1.72X_8 - 0.521X_9 \end{aligned} \tag{4.4}$$

$$Probability\ of\ Company\ Default\ = \frac{1}{1 + e^{-O}} \tag{4.5}$$

The framework of Ohlson totals nine ratios, or company features, which are utilized for predicting corporate default. The nine ratios are: 1) *Size,* 2) *Total Liabilities/Total Assets,* 3) *Working Capital/Total Assets*, 4) *Current Liabilities/Current Assets*, 5) *Leverage Dummy,* 6) *Net Income/Total Assets,* 7) *Funds from Operations/Total Liabilities,* 8) *Net Income Dummy,* 9) *Net Income Change Ratio*. Ohlson measure leverage through a dummy of 1 if total liabilities exceed total assets, and zero otherwise. He likewise introduces a dummy for net income, where the variable takes the value of 1 if net income has been negative for the two past years, and zero otherwise. On the variable of *size*, Ohlson introduces a domestic component, such that size equals the logarithm of total assets over GDP. It makes the variable universal to companies with different national backgrounds. Thus, this operation prevents that the large solid Latvian company become penalized in a portfolio with predominantly U.S. companies. Ohlson further creates the ratio of *net income change*, which takes positive value between 0 and 1 if the firm has experienced an increase in net income, and negative value between 0 and -1 if it experienced a decrease. This variable adds a dynamic factor to the framework, as it indicates in which direction the company is moving.

Generally, Ohlson was less focused on assets while he was more concerned with net income. As such, Altman denominated the majority of his ratios by total assets, where Ohlson frames a larger aspect of a business' financials. To take it a step further, Ohlson did not prohibit himself from capturing the same influences through utilizing features that indicated similar effects, such as net income in three different ratios. He also indicated the expected directions of his ratios, which can be summarized in the following table:

Table 4.2: James Ohlson (1980) – Direction of Financial Ratios on Probability of Default

Increasing Effect	*Decreasing Effect*	*Divergent Effect*
Total Liabilities / Total Assets	Size	Leverage Dummy
Current Liabilities / Current Assets	Working Capital / Total Assets	
Net Income Dummy	Net Income / Total Assets	
	Funds from Operation / Total Assets	
	Net Income Change Ratio	

Source: Personal collection

We re-estimate the coefficients of Ohlson's framework on our dataset with a similar methodology to the original one. As such, we build a logistic regression model including all the nine variables. Thus, the original Ohlson model will be evaluated based on the original coefficients multiplied onto original ratios, whereas the re-estimated will be considered with the original ratios but with new coefficients. In building our re-estimated Ohlson, we will build it with the maximum likelihood estimation (Agresti & Franklin, 2013).

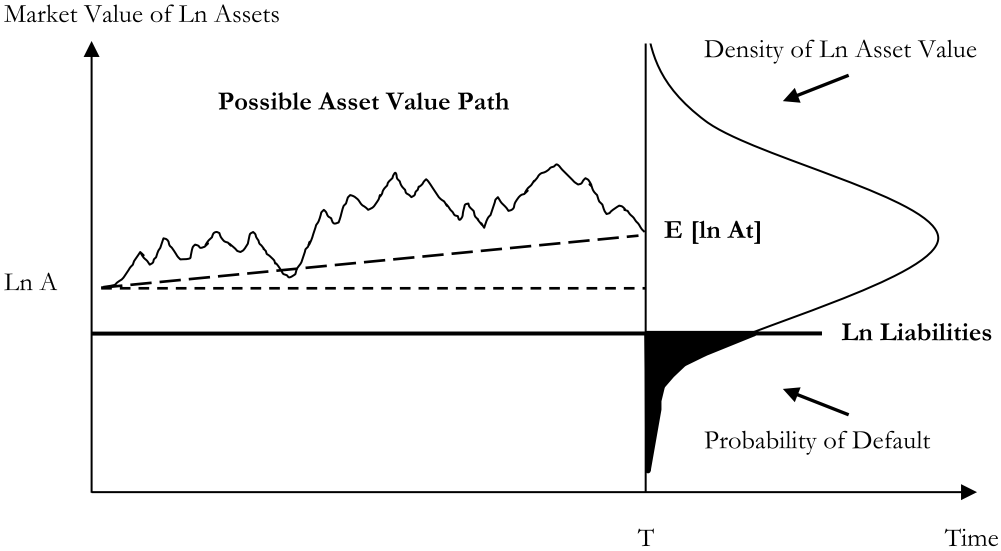
4.3 The Methodology of Merton: The Original and the Naïve

The structural approaches to modelling probability of default are not statistically estimated models like their ratio-based counterparts. Rather they are underpinned by the option pricing framework by Black and Scholes (1973) and build upon by Merton (1974) as presented in the theoretical framework of this thesis. The structural approaches are in essence using economic reasoning to identify conditions under which borrowers are expected to default and then estimate the probability that these conditions will take place. For any publicly traded firm, the probability of default can be calculated independently and requires no coefficient estimates. A variety of structural models and in turn different versions within each model exist. This thesis will employ two versions of the European call contingent claim approach, specifically the original model derived by Merton and a computationally less heavy and contemporary version presented by Bharath and Shumway (2008). The former will be referred to as the Original Merton Model and the latter as the Naïve Merton Model. As there are no coefficients to re-estimate per se, we refer to the Naïve version as the re-estimated model.

The Original Merton model assume a simple capital structure for the firm: debt plus equity. In this framework the liabilities consist of one zero-coupon bond with principal L maturing in T with no payments up until T (Löffler & Posch, 2011). The basic premise of the structural models is that default occurs if the value of the assets has plummeted to a level below the firm's liabilities at maturity. The insight of the Original Merton Model is that the payoffs to the shareholders of a firm are very similar to the payoffs they would have received had they purchased a call option on the value of the firm's assets with a strike price given by the amount of debt outstanding (Vassalou & Xing, 2004). The European call contingent claims models hence view equity as a European call option on a firm's assets with a strike price equal to the face value of its debt liabilities. As such, the option pricing techniques of Black and Scholes (1973) may be used to estimate the value of the option and the underlying probability of default (Tudela et al., 2003).

The option expires when the debt matures, at which point the equity holders either 1) exercise their option and pay off the debt if the value of the firm's assets is greater than that of its liabilities, or 2) let the option expire if the assets are not sufficient to cover the maturing debt, i.e. they exercise the walk-away option when the equity value is negative and leave the firm to the creditors (Jackson & Wood, 2013). If the option is left to expire, the firm is assumed to default and the residual claim to equity is assumed to be zero. To avoid foregoing a benefit from an increase in value, equity holders will wait until maturity before they decide whether to default or not (ibid.). This is the conditions under which the contingent claim models determine the probability of default.

Figure 4.1: Illustration of Default Probability in the Merton Framework



Source: Personal adaptation of Löffler & Posch, 2011

The probability of default as the probability that, at maturity, the value of the assets is below the value of the liabilities can be visualized as seen above. The distribution of the asset value at maturity is assumed to follow a lognormal distribution. The annualized variance of the logarithmic asset value changes is denoted by σ^2. The expected change in logarithmic asset values is denoted by $\mu - \frac{\sigma^2}{2}$ where μ is the firm's expected return. The logarithmic asset value in T thus follows a normal distribution with the following parameters (Löffler & Posch, 2011):

$$ln\, A_T \sim N\left(ln\, A_T + \left(\mu - \frac{\sigma^2}{2}\right)(T), \qquad \sigma^2(T)\right) \tag{4.6}$$

And consequently, the probability of default can be given by the cumulative standard normal distribution:

$$Probability\ of\ default = N\left(-\frac{ln\,\left[\frac{A}{L}\right] + \left(\mu - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right) \tag{4.7}$$

The number of standard deviations the expected asset value A_T is away from default, i.e. the term in brackets, is labeled the distance to default. The face value of long-term liabilities, L, are directly observable from the balance sheet and time to expiry, T, is in this thesis taken to be one year (Löffler & Posch, 2011). The market value of the firm's assets, the asset volatility and the firm's expected return, are not directly observable and must be estimated. As the observable book values may diverge considerably from the market values option pricing theory is used to determine the market value and volatility of a firm's assets (Vassalou & Xing, 2004).

The market value of equity for a publicly traded firm is observable given by the number of shares outstanding and the share price, and this is used to establish a relationship between the assets and equity (Löffler & Posch, 2011). At maturity, the value of equity will be zero as long as the asset value is below the value of liabilities. If the asset value is higher than the liabilities, then the residual value will flow to equity-holders and the pay-off increases linearly with the asset value. This pay-off can be viewed as the pay-off of a European call option:

$$E_T = max\ (0, A_T - L) \tag{4.8}$$

By assuming that the firm pays no dividend and thus using the standard Black-Scholes call option framework, equity is given with the following equations with r denoting the risk-free rate of return:

$$E_T = A_T * N(d_1) - Le^{-r}N(d_2) \tag{4.9}$$

$$d_1 = \frac{ln\ \frac{A_T}{L} + \left(r + \frac{\sigma^2}{2}\right)}{\sigma\sqrt{T}} \tag{4.10}$$

$$d_2 = d_1 - \sigma\sqrt{T} \tag{4.11}$$

The Black-Scholes framework consequently links the unobservable market value of assets and asset volatility to the observable market value of equity, E (Löffler & Posch, 2011). To solve the equation for equity with two unknowns we use the relationship between the volatility of equity and the volatility of asset value:

$$\sigma_E = \sigma\ N(d_1)\frac{A}{E} \tag{4.12}$$

As we can observe the market value of equity and can estimate the volatility of equity the result is two equations, (4.12) and (4.9), with two unknowns that can be solved with numerical or iterative routines. The firm's expected return on assets can then be estimated with CAPM and the probability of default according to the Original Merton Model can be determined (ibid.).

The Naïve Merton model simplifies the estimation difficulties of the original model. Instead of using the somewhat complex Black-Scholes equations (4.10) and (4.11), and the following iterative routines to solve for the unobservable variables, the model makes some simplified assumptions. The value of liabilities is now estimated as the sum of current liabilities and half of long-term debt. This value is added to the observed market value of equity to arrive at an estimate for the market value of assets. Asset volatility is estimated as a weighted average of volatility of debt and equity. Here, the volatility of debt is a simple linear transformation of the volatility of equity (Bharath & Shumway, 2008; Jackson & Wood, 2013). The same final equation for calculating probability of default is employed, but the parameters are substituted with the ones summarized below with subscript *N* for Naïve:

$$L_N = Current\ Liabilities + 0.5 * Long\ term\ Debt \tag{4.13}$$

$$A_N = E + L_N \tag{4.14}$$

$$\sigma_N = \frac{E}{A_N}\sigma_E + \frac{L_N}{A_N}\sigma_{DN} \tag{4.15}$$

$$\sigma_{DN} = 0.05 + 0.25\sigma_E \tag{4.16}$$

4.4 The Practical Modelling Methodology

The theory section presented how the practical approach to probability of default is predominantly carried out with the use of a logistic regression with accounting-based variables (Engelmann & Rauhmeier, 2011; De Laurentis et al., 2010; Neisen & Rosch, 2018). The logistic model has previously been described in the section regarding Ohlson, however this section will emphasize how the modelling procedures are carried out in practice. As such, the methodology presented here will serve as a template for the practical derived model presented later in this thesis. The steps included in this template comprise of the following: 1) The considerations to be made regarding the accounting-based inputs prior to conducting any modelling procedures, 2) The statistical treatment of these inputs, 3) The input selection techniques applied and 4) Deriving an optimal model. To further reflect a contrast between the practical model and the two other accounting-based models, we will use a number of new financial ratios from Plenborg et al. (2017) in addition to the ones provided by Altman and Ohlson. These are *Total Liabilities/EBITDA, Funds from Operations/Capital Expenditure, Funds from Operations/Current Liabilities* and *Working Capital/Revenue*. The motivation for including these additional ratios are two-fold. First the ratios are often employed when conducting a financial credit analysis and secondly, the total number of ratios will then to a greater extent reflect the range of ratios available to practitioners.

Financial ratios are structured by being different category labels. Like Altman and Ohlson, practitioners also use categories such as leverage and profitability. For each of the ratios an initial indication regarding the direction is formulated (Engelmann & Rauhmeier, 2011). The specific directional indication is to be grounded in economic reasoning such that a high degree of leverage is associated with an increased probability of default, par exemple. These considerations are all made prior to conducting any modelling steps. Practitioners employ few significant rather than many predictors, and this implies that the number of ratios to be included has to be reduced to secure the statistical appropriateness of the final model (ibid,). The categorization and formulation of directional indications will help with this matter.

To reduce the number of covariates the first selection stage involves assessing the statistical suitability. This implies that a test of whether the linearity assumption of the logistic regression applies to the ratios needs to be carried out. If a ratio is to be included it needs to satisfy a linear relationship between it and the log odds, i.e. that it is linearly associated with the log odds of probability of default (Engelmann & Rauhmeier, 2011). To check this, we aggregate observations for a given predictor into sorted groups. The first group contains the top 3% values the next values from *94%* to *97%* and so forth, resulting in a total of 33 groups per predictor. Hayden (2011) suggests that the data is aggregated into 50 groups, however, to ensure that each group includes defaulted observations we choose 33 groups. In each group we calculate the empirical default rate, i.e. the mean of the dependent default variable, the associated logit and the median of the predictor value. To check for linearity the logit and median predictor value is then plotted against each other (Hosmer & Lemeshow, 1980; Menard, 2010).

If there is evidence against the linearity assumptions, we apply different transformation techniques. These include either a quadratic, cubic or logarithmic transformation (Menard, 2010). The plots of each of the fits will also aid us in confirming the direction, and if the variable exhibits a wrong direction it is excluded. The second variable reduction technique involves analyzing a correlation matrix of the remaining variables. Two variables that are highly correlated with each other cannot both be included if the coefficients are to be statistically unbiased and significant (Engelmann & Rauhmeier, 2011). Variables that are highly correlated with each other are grouped and the ones with the lowest Somers' D is excluded. Somers' D is an evaluation metric that will be explained in a later section.

The initial modelling stage begins with the list of remaining variables. To arrive at the optimal model, only variables that are highly significant should be included. To determine the final variables to include in model a backward stepwise elimination is carried out. This implies estimating a preliminary model including all of the remaining variables, and then excluding the least significant variables until all variables exhibit the required significance level (Engelmann & Rauhmeier, 2011; Menard, 2010). The variables are excluded one at a time, each time estimating a new model prior to excluding the next. All of these steps ensure that the final model is including only highly significant and economically plausible variables.

4.5 Machine Learning Methodology

Having outlined the academic and practical methodologies employed in the paper, we turn to the machine learning algorithms considered in the research at hand. We first introduce the classification tree methodology and next the random forest algorithm.

4.5.1 Machine Learning Methodology: Classification Trees

The classification tree machine learning algorithm, which is the decision tree type for classification problems, is a rule-based technique that uses "if"-statements based on the values of the predictors to create separate and non-overlapping regions in the feature space (Lateef & Ruichek, 2019). In a classification setting, the optimal tree is one in which the two classes are assigned to their own regions. A tree diagram is a neat visualization of the feature space separation that allows the user to follow how the data is split. The tree diagram is drawn upside down such that the starting point of a tree is the root node containing all observations in the dataset. After the root node the dataset is split in branches defined by "if"-statements. Each branch from the root node leads to a decision node, representing a new threshold value of a predictor variable. Whenever the data is split by a new node, the feature space is further divided into additional regions. From a decision node, the tree can either continue branching down to the next decision node or to a terminal node. A terminal node resembles one of the regions and the associated class label for observations belonging to this region (Hastie et al., 2014).

The classification tree algorithm will split the data in the root node by using that predictor and threshold which separates the data the best. This is defined as the split creating the purest nodes where purity refers to optimal separation of classes (Hastie et al., 2008). As such, a node is pure if the observations in the corresponding region is all from the same class. In other words, the predictor used at the top split should enable us to divide the observations into groups that are as different as possible (ibid.). The subsequent splits of the data are made by maximizing the gain in purity. We use the Gini index to measure node purity:

$$G = \mathrm{p}_{\mathrm{rd}}(1 - \mathrm{p}_{\mathrm{rd}}) \tag{4.17}$$

Where p_{rd} refers to the proportion of observations in a given region belonging to class *default*. It will take a value of zero or one if all observations in the region belongs to the same class. The algorithm thus starts by considering all predictor variables and all possible threshold values before choosing the ones that maximizes gain in purity. This process is repeated at the next split, but instead of dividing the entire dataset we are dividing a region (Hastie et al., 2014). To mitigate the computational infeasibility of considering every possible combination of predictors and thresholds, the method employs a greedy top-down splitting approach. This implies that the model is deciding the best split at the particular step rather than looking ahead for future possibly more pure splits at later stages (Hastie et al., 2008). Thus, the classification algorithm is by design potentially foregoing a split that would have resulted in a better tree at a future step.

When employing a classification tree for predictions, each observation will receive a predicted score corresponding to the proportion of class *default* training data observations in the terminal node in which it belongs (Hastie et al., 2014). As such, if a firm ends in a terminal node where the proportion of training data observations from class *default* was *10%,* the predicted score will be *0.1*. The more decision nodes a tree has the purer the leaves can get, and the resulting class separation in the training data will be greater. However, the consequence of having more branches and terminal nodes is that the amount of observations within each terminal node decreases. We are therefore at risk of creating a model that is so specific to the data it is built with that it overfits on the data we intend to predict on (ibid.). This is represented by a trade-off between pure, deep trees and generalization error when employing the model for predictions (Hastie et al., 2008). To combat the potential over-complexity of the classification tree we first construct a large tree and then *prune* it back to obtain a subset of the tree. This implies that branches of decision nodes associated with small terminal nodes are removed. To decide the optimal subset of a tree and hence minimum number of observations in a terminal node, a repeated 10-fold leave-one-out cross validation procedure is carried out. That is, the data is first divided into ten folds and with nine folds we test how the subtree fared in prediction on the tenth. This procedure is repeated for varying number of minimum observations in a terminal node. The subtree with the lowest prediction error rate is chosen as the optimal pruned tree (ibid.).

4.5.2 Machine Learning Methodology: Random Forest

The discriminative power of the simple classification tree can be improved upon by employing the random forest algorithm. This machine learning approach use a collection of classification trees to boost the class separation performance. The simple classification tree is a high variance algorithm, i.e. it will yield differing results if applied to dissimilar datasets (Hastie et al., 2008). To reduce the variance and receive stable generalizable predictions, we use the concept of bagging. It is a frequently employed procedure in a classification tree setting. The background is that the average of a large set of observations returns a lesser variance. Bagging consists of generating a large number of bootstrapped datasets, i.e. repeated sampling from the same main dataset, growing a tree for each bootstrap. The number of trees needed in the forest relates to the weak law of large numbers as the resulting prediction will tend toward a mean value as the number of trees increases. This convergence can be seen at 200 trees and it is rarely necessary to construct more than 1000 trees (Hastie et al., 2014) why the random forest algorithm applied in this thesis will construct 1000 trees. The trees are constructed as deep unpruned trees, and the collection of trees is referred to as a forest. The resulting prediction is then the average of all the trees in the forest.

The random forest algorithm employs the concept of bagging but adds a refinement to the procedure. If there is one overarching predictor separating the classes the best, then the far majority of the trees in a bagged forest will use the same predictor at the top split. This implies that the predictions stemming from the individual trees will be correlated (Hastie et al., 2014). The reduction in variance and increase in ability to accurately generalize from using a bagged forest over a simple classification tree will be greater if the individual trees are uncorrelated. This is exactly what the random forest algorithm does. The trees are decorrelated by constraining which predictors are used at each split. Each time a split is considered a random sample among the predictors of size *m* is chosen and the algorithm can only choose among these *m* predictors. The predictor chosen for the split is still the one that maximizes the gain in purity (ibid.). The process is repeated across all trees in the forest and each tree will be much more distinct as the predictors at the top splits will differ.

The algorithm can be tuned with respect to the choice of *m* and minimum number of observations in each terminal node across the forest. The optimal number of these parameters is found by minimizing the out-of-bag error (Hastie et al., 2014). Each tree in the forest is constructed using two thirds of the total observations in the bootstrap. The remaining observations in the bootstrap, the out-of-bag observations, is used to calculate this error. For a given unique observation across all of the bootstrapped datasets we predict its class by using every tree where that observation was out-of-bag. The average of the predictions will then be that observations predicted class (ibid.). We repeat this procedure for every observation and the resulting out-of-bag error is used as an estimate of the test error for the random forest model.

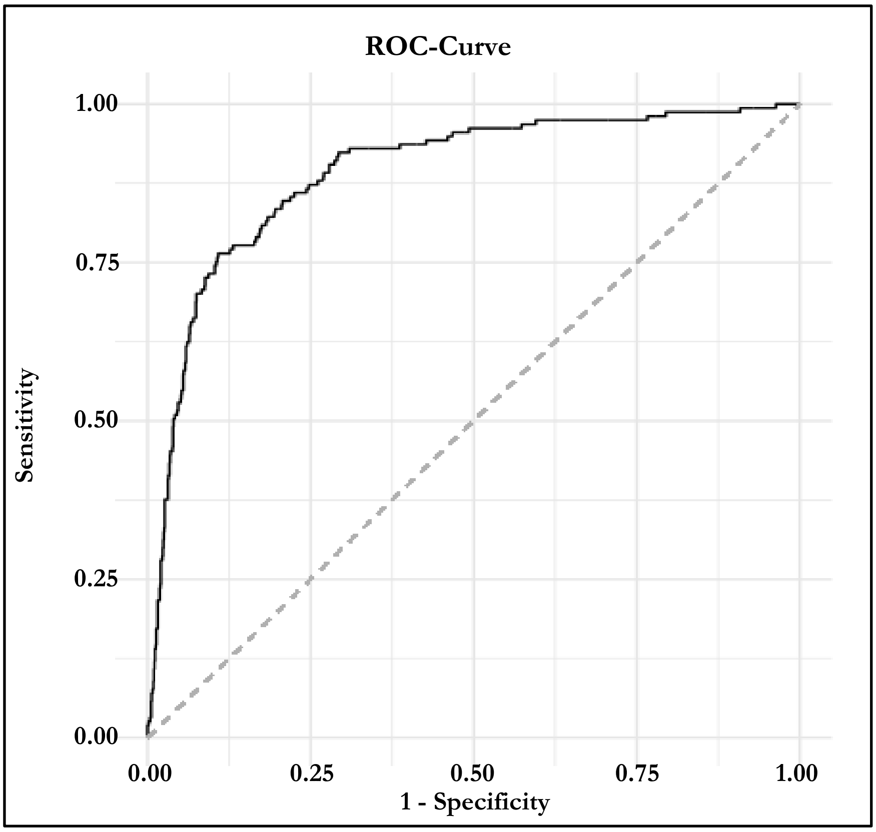
4.6 Metrics for Comparing Models

For comparing the models developed in our research, we utilize receiver operating characteristic (ROC) curves and Somers' D. To a large extent, the first is the primary metric for the academic field of default probability. The latter is somewhat more utilized in the practical approach. By these means, those two in combination fit the research of this paper particularly well, as it both has an academic and a practical scope. Below, we will uncover the two metrics' respective properties.

4.6.1 Receiver Operating Characteristic Curves

We will utilize the area under the ROC (AUROC), which indicates how efficient a classifier of default a given model is (Fawcett, 2006). Fundamentally, the AUROC indicates how well a model segregates the two groups in the classification problem in percentage terms. A value of 75% AUROC indicates that a randomly drawn default will be scored above a randomly drawn non-default, 75% of the times. Thus, a model with no discriminative power will have an area under the ROC of 50%, and go from (0,0) to (1,1) in a unit square. As the AUROC approaches 100% it will cover more of the unit square and move towards the (0,1) corner of the graph.

Figure 4.3: Theoretical Receiver Operating Characteristic-Curve



Source: Personal collection

Receiver operating characteristic curves are useful for comparing models of probability of default, as it overcomes the problem of arbitrary cut-off points. Arbitrary cut-off points are a topic of discussion within the PD-literature, which have attracted a lot of attention. Exemplified through Ohlson's original logistic regression, where he classified firms with a score above 0.038 as defaulting companies; a point which might as well have been anywhere between 0 and 1 (Ohlson 1980). The ROC-curve handles this problem by plotting a curve according to all cut-off points within the range, and the main focus is then left on the ability to segregate defaulting from solvent businesses across all cut-off points. Thus, it is a beneficial methodology for comparing different models that seek to explain the same problem. In reality, no PD-model is perfect, and how the AUROC handles misclassifications is a strong property of the methodology.

The two types of misclassifications are called false positives and false negatives, respectively. The first refers to solvent companies that gets incorrectly classified by the model as a default. This type of misclassification is called a type I error. The latter refers to a default firm, which gets incorrectly classified as a solvent company. This is termed a type II error. In extension hereof are true positives and true negatives, which indicate correct classifications by the model (Duan & Shresta, 2011). The type I and II errors are dependent on the cut-off point of a given model, which can be summarized in the following confusion matrix.

Table 4.1: Theoretical Confusion Matrix

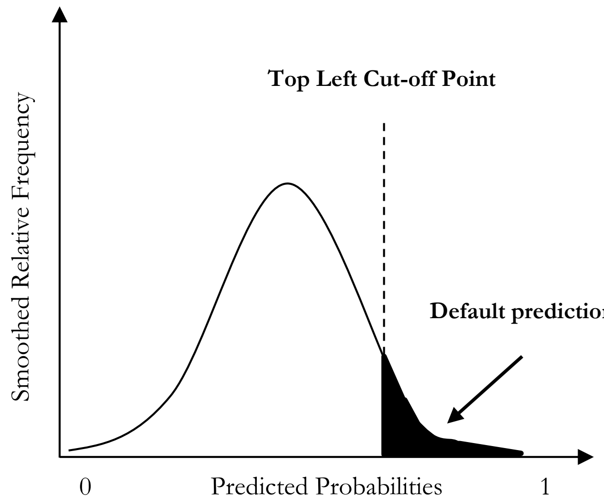
	Actual Observations	
Predicted by Model	Solvent	Default
Solvent	Score < Cut-off **(Correct Prediction)**	Score < Cut-off **(Type II-error)**
Default	Score ≥ Cut-off **(Type I-error)**	Score ≥ Cut-off **(Correct Prediction)**

Source: Personal collection

The proportion of true positives out of all defaulting companies in a given portfolio is termed true positive rate, or sensitivity. Likewise, the proportion of true negatives of all solvent companies in the portfolio is termed specificity. The concepts of sensitivity and specificity are denoting the x and y-axis of the ROC-curve, respectively. As such, the method is useful in comparing classification models, as it considers a tradeoff between both misclassified defaults and non-defaults. This is property that gives it strength compared to another analytical metric used in the discipline of statistical modelling, accuracy, which only considers correctly classified observations out of the total portfolio. To showcase the potential misleading the accuracy metric incurs, consider a portfolio of exactly 10% defaulting companies. A model that classifies all companies in the portfolio as solvent would receive an accuracy of 0.9 or 90%, in spite the fact that it classifies no firms as defaulting. We mitigate this problem by sticking to AUROC as our model performance metric, as its property of emphasizing both sensitivity and specificity make it attractive for skewed class distributions (Fawcett 2006).

When examining ROC curves throughout the thesis we will at times refer to the top-left point. We examine this point, not as an evaluation metric but to add an analytical layer and showcase how a specific model function. The top-left point in the ROC curve is the point that maximizes the sum of specificity and sensitivity. Choosing the optimal cut-off point is not within the scope of this thesis as it depends heavily on the situation at hand and the consequences an incorrect classification has. In a corporate loan setting one can imagine a trade-off between foregone profits from loans not given to actual non-defaulters and write-offs from loans given to actual defaulters. Consequently, we do not claim that this is the optimal point but rather give equal weight to specificity and sensitivity to remain unbiased. The cut-off value associated with the top-left point on the ROC curve is visualized in a generic example below. Here the shaded area on the density plot represents observations with a predicted default score above the cut-off value. Observations in the shaded area correspond to correctly classified defaulting firms and type II error and vice versa for observations in the non-shaded area.

Figure 4.4: Generic Density plot with cut-off value



Source: Personal collection

When using the area under the ROC as our metric for comparing model performance, we align ourselves with several of the pertinent fields. First, it is the primary metric within the academic PD-literature, as the importance of penalizing models for making too many misclassifications in both directions are prioritized when giving default probabilities to firms (Jackson & Wood, 2013; Duan & Shrestra, 2011). Second, the AUROC is one of the metrics amongst the preferred in practical credit risk models, due the above properties and its intuitive interpretation (Engelman & Rauhaimer, 2011). Third, the AUROC has also gained prevalence within the machine learning field, as the primary metric for evaluating classification models (Bradley 1997).

4.6.2 Somers' D

The other metric for evaluating models that we incorporate in our research is that of *Somers' D*. It is an evaluation tool provided by Robert Somers, as he identified a need for a pair of asymmetric coefficients in contrast to the symmetric measures of ordinality (Somers, 1962). As such, it captures the ordinal association between two possible response features X and Y. Here, the measure takes the value negative one when all pairs of variables disagree, and likewise the value of positive one when all pairs of the two features agree. Thus, it is a relevant measure in terms of evaluating a PD-model, as the performance of such can be dissected into a vector of the actual default or solvency of the observations, and a similar vector of the predicted values by the given model. If the observations ordered in terms of the latter vector completely segregates the defaults from the solvent firms, the Somers' D would equal one.

More specifically, Somers build his measure around what he denominates concordant and discordant pairs respectively. Concordant pairs would be the case of the larger of two X-values being associated to the larger of two Y-values. Similarly, the discordant pair occurs when the larger of two X-values would be associated with the smaller of two Y-values. The value of Somers' D is ultimately the difference between the two corresponding conditional probabilities, under the occurrence of the two X-values not being equal (Newson, 2002). By these means, it is a relevant metric to apply for different default probability models, as it evaluates one specific model's ability to discriminate between the defaulting and solvent companies. This property is shared with the AUROC, which it shares numerous assumptions with. More accurately, for models with a dichotomous outcome, the Somers' D has the relationship to AUROC of being $Somers'D = 2 * (AUROC - 0.5)$. However, we choose to present both measures in our thesis, in order to capture the duality of the research being both academically and practically oriented.

4.7 Training and Validation

When evaluating the models in our research, we choose to split our portfolio into training and testing data sets respectively. Doing this, we ensure the validity of our models being predicting mechanisms of defaulting companies. In its essence, a model is trained on the majority of the portfolio, and then a minority of the portfolio is *holdout* as a testing data set, in order to validate how well the given model discriminate between defaulting and solvent companies. There are several motivations for adopting the procedure, where most central is the prevention of an overfitted model. As such, a PD-model may be capable of correctly segregating all companies in the training portfolio, but if it is unable to achieve a higher than random discriminative power on new and unknown observations, the model is little useful.

By these means, we evaluate the models of our research on their performance on the testing data set, rather than their performance on the training portfolio when comparing our models. Furthermore, some models considered in the thesis at hand are already estimated in another framework, which we showcase the efficiency of by applying to our portfolio, including the original Altman, original Ohlson, and both Merton models. As such, these models do intrinsically not need a training and validating process, as they are already "trained" in another setting. However, in order to even the battlefield, we evaluate all models on the same testing portfolio. Specifically, the testing set is created from a randomly selected part of companies that constitute 25% of the original portfolio, which in total amounts to around 1500 firms. Furthermore, we check that the proportion of defaults is equal across both the training and testing portfolio. By adopting this approach, we not only comply to essential statistical properties, we also align with the academic literature of default probability (Jackson & Wood, 2013; Hillegeist et al., 2004; Giacosa et al., 2016).

5. Data

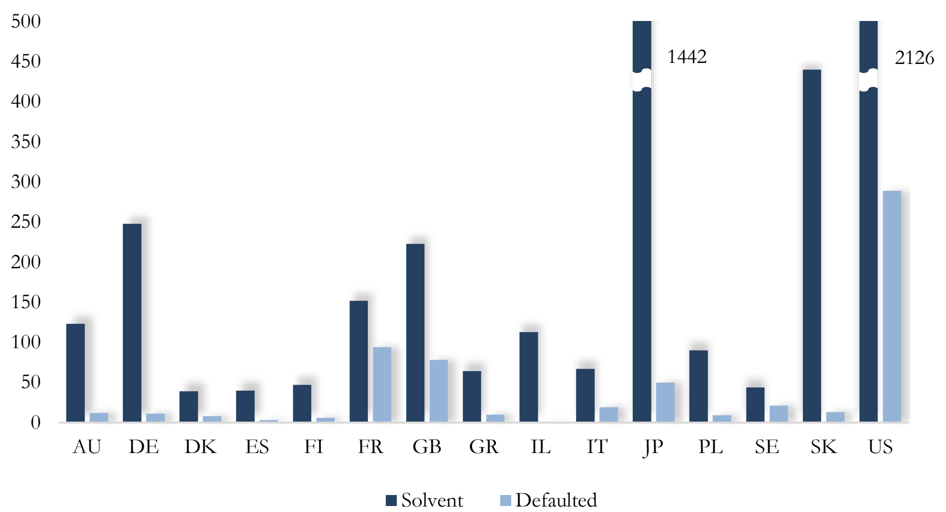
Multiple factors are driving the data collection design in this thesis. The multiple legs in the research question impose several strict requirements on the data behind the analyses. These requirements are two-fold. First the thesis seeks to model on a setting reflecting a true business environment. This includes considerations related to the number of defaulting relative to non-defaulting companies, the range of countries to include, whether certain industries should be excluded and a time period that echoes the economic cycle with both up and downturns. Secondly the thesis will build, compare and contrast a wide variety of both academically and practically motivated models. These differing models have either non or only minor overlapping input requirements. This consequently implies that the data behind need to be diverse as a prerequisite is that all data points are available for each unique observation. The following paragraphs will introduce and describe how the collection and preprocessing of the data used in this thesis has been conducted.

5.1 Data collection design

Corporate defaults are a somewhat rare event when dealing with firm level data. In order to obtain a meaningful amount of defaulting firm observations while still maintaining a natural relative composition, the data has been compiled as a combination of the "OSIRIS Global" and "COMPUSTAT North American" database. Both databases cover a wide range of publicly listed companies and corporate defaults. To ensure comparability across national accounting principles both databases applies a standardization procedure and have a high level of transparency. This facilitates data clarity and enables the researchers to conduct cross-border studies on a uniform dataset without sacrificing accuracy. A firm is classified as a defaulting firm when a deletion from one of the databases has occurred as a result of either a liquidation or bankruptcy corresponding to either a chapter VII or XI bankruptcy (US Courts, 2020). The observation of a defaulting firm includes the latest available balance sheet and income statement assuming that the observation has defaulted within 12 months.

To further ensure comparability across firms, explicit considerations has been made both with respect to nation and industry. This entails that the sample has excluded firm observations from countries not part of the OECD nations, which is parameter we choose to somewhat uniform the developmental level of companies' national origin. Additionally, the data extraction methodology involves only sampling companies adhering to the same overarching industrial definition. This implies that all companies operating within financial services has been excluded. This is done as this type of companies have vastly different balance sheets, earnings profiles and ways of reporting and are thus not comparable even post standardization efforts. Companies excluded due to this criterion includes banks, insurance companies, REITs, brokerages and other financial intermediaries. The exclusion of financial service companies is consistent with prior empery (Altman 1968; Ohlson 1980; Jackson & Wood 2013, Hillegeist et al. 2004) although we expand the scopes applied in these works, in order to evaluate PD models in a broader context than previously studied. The data collected is annual data spanning the years 2001 through 2019 representing balance sheet and income statement items at the fiscal year end. We have not made any discrimination between companies ending their fiscal year at differing months. Neither do we weigh companies differently based on the year they are observed. Data is delivered in US Dollars for accounting information across all countries. For retrieved market information, the data is provided in the native currency of each company. A unique ISO Currency Code represents each native currency and all native currencies have been converted to US Dollars using the exchange rate at year end 2019. The distribution of companies in the portfolio across countries is visualized on the below figure. The dataset is skewed towards countries which had a larger proportion of companies in the databases.

Figure 5.1: Distribution of Companies across Countries



Countries with above 40 firms in the portfolio. Y-axis capped for formatting purposes. Japan has 1442 solvent companies. U.S. has 2126 solvent companies.
Source: Personal collection

5.2 Data cleansing procedures

Prior to calculating financial ratios, we conduct several operationalizations on our portfolio of companies. This is done to reduce the initial 121.000 observations identified for possible analysis, as observations should only be included if sufficient data is available to allow for application of the probability of default models under consideration in this thesis.

We start by engaging in a comprehensive data cleansing. Here, we first remove observations that include missing values to ensure that all calculations are possible throughout the dataset. Second, we check for any duplicate observations, which fortunately was irrelevant due to good quality of data provided by the Compustat and Osiris databases. Third, we remove extremely small companies, as we consider those a different species of firms. We set thresholds of at least one million dollars in assets, and at least one hundred thousand in revenue. While we do this, we acknowledge that companies of this size do exist in the business landscape and do indeed have the probability of going default. It is nonetheless a measure we take to focus the models we build and ensure generality of the coefficients it find.

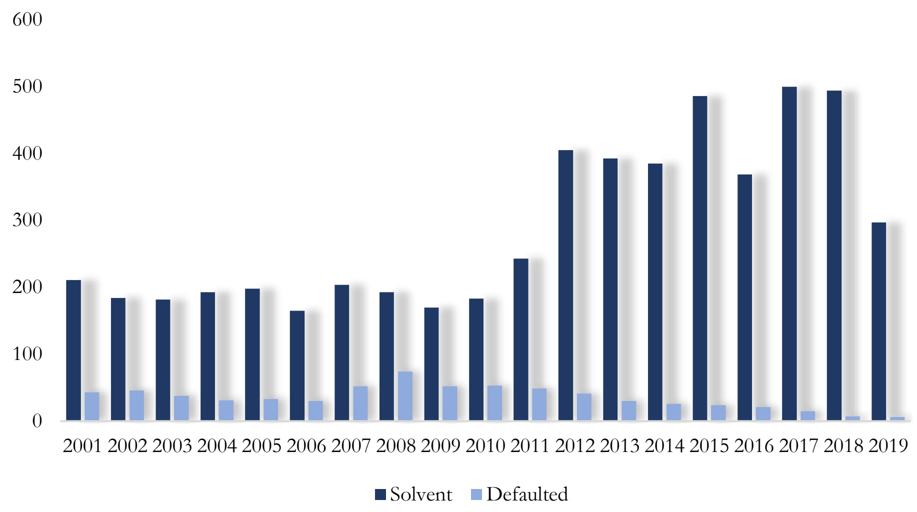
We then proceed to handle outliers through winsorizing. Winsorizing consists of replacing an outlier with "the nearest observation that is not seriously suspect" (Tukey 1962). According to several studies such as Dixon (1960; 1980), this approach shows to yield more stable results than other methods of dealing with outliers. The version of winsorizing that we are employing consists of re-coding every variable above and below the minimum and maximum 1% of the portfolio, i.e. having thresholds of 0.01 and 0.99. Specifically, we recode these observations to take the value of the threshold percentiles. The winsorizing is performed on all ordinary variables in the data set. Prior to the winsorizing process random checks of extreme values were looked up in their respective annual reports as a final data quality check. This was done to avoid including and giving weight to erroneous data points. Examples of companies with suspect values, which however add relevant information, are Exxon Mobil that have by far the largest total assets in the portfolio, and General Motors that was granted a bailout package of 13.4 billion by the U.S. Government in 2008.

Figure 5.2: Data Cleansing Process

Source: Personal collection

Finally, we balance the data set to some degree. The raw portfolio suffers from data becoming more comprehensive as it approximates 2019, mainly due to the quality of the databases. As such we remove observations, predominantly in the 2012-2019 period, through a randomization process. The event of default is treated as a dichotomous variable, representing a 1 for default and 0 for non-default, and is coded for the remaining observations in the cleansed dataset.

Figure 5.3: Distribution of Companies across Time



Source: Personal collection

5.3 Calculation of input parameters

There are several sources of financial information which constitute parameters our data set must comply with. These stem from the frameworks of Altman, Ohlson, Merton, and variables we introduce from Plenborg et al. (2017). The calculation hereof will be presented below.

5.3.1 Accounting-based models

The academic accounting-based models of Altman and Ohlson applied in this thesis require specific financial ratios derived from either the balance sheet or income statement of each observation included in the final sample. The majority of Altman's ratio are standard financial ratios whereas some of Ohlson's to a greater extent are unique to his model. A range of the ratios can be calculated directly from the corresponding financial items, however some require further operationalization.

The only ratio that is recurring in both models is *Working Capital/Total Assets*. The definition used for Working Capital in this thesis is current assets less current liabilities less cash at hand. We recognize that operational cash at hand ideally should be classified as working capital, however this separation has not been possible as it is rarely a unique balance sheet item. Altman's ratio of *Market Value of Equity/Book Value of Liabilities* require us to know the market value of equity. This is not an item directly available from the balance sheet of a firm. To calculate the market value of equity for each of the more than 6000 unique firms in the dataset we need to know both the number of shares outstanding and the share price at the end of each observation's fiscal year. The data required to calculate this is found from multiple data extractions from the merged COMPUSTAT/CRSP Global Stock database. This type of market data is delivered separately from the database of financials. We hence need to be able to pair the data with the correct observation to have one market record per company. Each of the observations in our dataset is provided a unique global index key that facilitates the pairing and merging of the extracted data.

A range of Ohlson's input parameters also requires attention in order to ensure the correct operationalization. This is due to some of them either not being standard financial ratios or because we need to clarify the definition applied in this thesis. These include *Size*, *Net Income Dummy*, *Net Income Change* ratio and *Funds from Operations/Total Liabilities*. In order to apply the Ohlson model in a true setting we follow his way of calculating *Size*. As our dataset contains observations from OECD countries rather than solely US firms, we choose to scale Size with OECD GDP. From the World Bank Database, we extract both current and constant 2010 OECD GDP in the year range of 2001 through 2019. These data extractions allow us to calculate a GDP Index comparable with Ohlson's. This index is added to the dataset as a separate column paired with the year of each unique observation. Common for both the *Net Income Dummy* and *Net Income Change* ratio is that we need Net Income for the observed year and one year prior, posing additional data requirements. Funds from Operations is not a financial item that was available in the databases for direct extraction. In order to ensure that the item is comparable across firms in the portfolio, we have defined Funds from Operations as EBITDA less change in Working Capital. This implies that an increase in Working Capital between the observed year and the year prior will affect Funds from Operations negatively. A consequence of this definition is that we need two years' worth of Working Capital data for each observation.

The practical probability of default model drives the inclusion of additional financial ratios worth considering when modeling probability of default. These include among other the two cash-flow based ratios of *Funds from Operations/Current Liabilities* and *Funds from Operations/Capital Expenditure* as well as the liquidity-based ratio of *Working Capital/Revenue*. To promote comparability across ratios we employ the same definition of Funds from Operation and Working Capital as in the above paragraphs. We follow the approximation from Plenborg et al. (2017) using Depreciation and Amortization as a proxy for Capital Expenditure, as the two over time should converge. In doing this we recognize that the *Funds from Operations/Capital Expenditure* ratio is at risk of reflecting Maintenance Capital Expenditure to a higher extent than actual Capital Expenditure for the period.

5.3.2 Structural Models

The application of the two structural approaches to probability of default is a demanding process with respect to data requirements. Part of the input parameters is already accessible from the dataset such as the book value of liabilities. Consequently, the main challenges lie within the estimation of the unobservable parameters of asset value and asset volatility for the Original Merton model. As put forward in the methodological part of this thesis, market value of equity and the associated volatility are both observable and can be used to solve for the unobservable values. Market value of equity is calculated in the same manner as described earlier as number of shares outstanding multiplied with the share price at the end of the observation's fiscal year. On the basis of historical daily stock returns for the entire fiscal year we calculate volatility of equity. However, both structural models applied require the annualized volatility. To facilitate this calculation, we convert the daily returns to logarithmic returns. The annualized volatility is then found by taking the standard deviation of the daily logarithmic returns and multiplying it with the square root of the number of trading days in the given year. The market value of equity and the associated annualized volatility is then paired with a risk-free rate proxied by the 12-month US treasury rate. For the Original Merton model, these inputs are then used to iteratively solve for asset value and asset volatility in equation (4.9) and (4.12), and the process is repeated for each of the 6121 observations in the sample. This computational complexity can be contrasted to the Naïve Merton Model as it has a closed form equation that does not require a numerical routine to solve for unknown values. The process for calculating market capitalization and volatility of equity is identical as above, however all of the remaining input parameters is directly available from the already established dataset.

5.4 Final Descriptive Statistics

Table 5.1: Descriptive Statistics of Ratios

Ratio	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
LevDummy	6,126	0.05	0.21	0.00	0.00	0.00	1.00
NIDummy	6,126	0.21	0.40	0.00	0.00	0.00	1.00
WC/TA	6,126	0.19	0.28	-7.08	0.07	0.30	0.97
RE/TA	6,126	-0.43	4.38	-249.00	-0.07	0.35	2.61
EBIT/TA	6,126	-0.01	0.37	-5.48	0.01	0.09	13.16
MV/BD	6,126	3.21	7.66	0.01	0.47	2.79	142.86
SA/TA	6,126	1.09	0.85	0.01	0.60	1.36	17.88
Size	6,126	8.07	1.90	3.26	6.80	9.36	12.84
TL/TA	6,126	0.61	3.20	0.02	0.36	0.68	244.07
CL/CA	6,126	1.02	10.81	0.01	0.40	0.91	809.804
NI/TA	6,126	-0.07	1.27	-94.26	-0.02	0.06	10.25
FPO/TL	6,126	0.11	0.91	-17.42	0.04	0.29	13.91
NIChange	6,126	-0.01	0.52	-1.00	-0.52	0.24	1.00
FFO/CAPEX	6,126	30.04	1,262.86	-9,641.98	0.63	4.34	84,947.54
TL/EBITDA	6,126	7.40	1,653.77	-84,783.47	1.71	8.07	97,543.15
FFO/CL	6,126	0.26	1.57	-33.25	0.06	0.59	36.83
WC/SA	6,126	0.44	11.05	-443.56	0.07	0.32	604.12
NäiveDD	6,126	4.08	9.15	-4.39	1.25	5.97	526.90

Source: Personal collection

The above table summarizes the descriptive statistics for the calculated financial ratios for the sample. It is evident that the minimum and maximum values are far from the mean, first and third quartile. In statistics, a distinction is made between *unusual* and *influential* data. Here, unusual refers to data points that lie outside the pattern set by other data and influential refers to data points that disproportionately influence the results of a model (Belsley & Kuh, 1980). In samples an observation may be both unusual and influential as an unusual value of a given predictor can have a disproportionate influence on the slope. However, for large samples, such as the one in this thesis, we are bound to obtain unusual observations simply due to the size of the distribution (Menard, 2010). That is, extreme values occur but the frequency is low. We constrained the disproportionally influential data points by first ensuring that they were free from measurement errors and then employing a winsorizing procedure that restricted rather than removing the values for the upper and lower percentile. In doing this we recognized that unusual observations 1) provide important information about the phenomenon under examination and 2) that they are valid for the population as the sample then reflects the true variability. We therefore deem the conclusions drawn from the sample correct.

6. Results

The results chapter of this research is split in three parts, which will serve the purpose of evaluating the three hypotheses that finalized the theoretical framework of the thesis. As such, the first section will be devoted to the academic approach to default probability. The second section will revolve around the practical discipline of credit risk. The third and concluding section of the results will be expanding upon the practical framework utilizing machine learning methodology. We evaluate the discriminative power of each model with respect to the AUROC and Somers' D metrics. These metrics present the overall discriminative ability of a model and we emphasize that comparison and evaluation between models can only truly be made with these. As such, the models should not be compared or evaluated in terms of accuracy at a single cut-off point, a specific threshold value or the predicted score assigned to individual firms. These values are model-specific and does not facilitate direct comparison. They do however add depth to the analysis of the individual model. To illustrate this analytical layer and gain further insights on how the models work we conduct two analyses for each. First, we illustrate how a model classifies observations at the cut-off point associated with the top-left point of the ROC graph. Secondly, we demonstrate throughout the results chapter each models' classification mechanism with a guiding example of the defaulted Thomas Cook Group (TCG).

6.1 The Academic Approach to Default Probability

The aim of the first section of the results chapter is to answer the thesis' first hypothesis: *The classic academic approaches to default probability have discriminative power on a modern portfolio.* To facilitate a verification of this hypothesis we estimate and subsequently evaluate the six different models that we have presented in the prior sections. These models include four accounting-based models and two structural models. The accounting-based models are the Altman and Ohlson models with their original coefficients and with coefficients re-estimated on our dataset respectively. The two structural models are the original Merton model and the more recent Naïve Merton model.

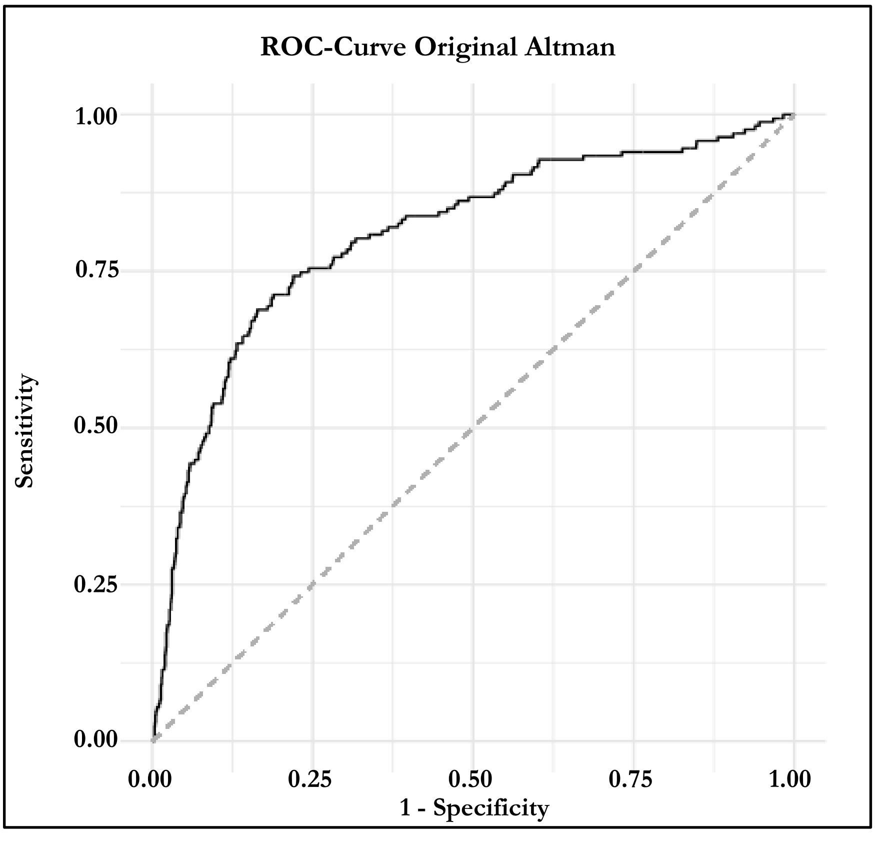
6.1.1 Original Framework of Altman

In Altman's seminal paper an increasing Z-score was associated with a lower likelihood of failure. This implies, that in order to facilitate comparisons between the models we need to flip the direction of the coefficients in the Altman model. This approach will not have any influence on the discriminative ability and the choice is purely made to simplify and avoid confusion when presenting the results. This alteration entails that a lower Z-score corresponds to a lesser likelihood of firm failure. The model will then be given by equation (6.1).

$$Z = -1.2 * WC/TA - 1.4 * RE/TA - 3.3 * EBIT/TA - 0.6 * MV/BD \\ -0.99 * SA/TA \tag{6.1}$$

Altman's Z-score framework does not provide us with a probability of default directly from the model. To calculate a probability of default from a given Z-score we follow the approach outlined in the methodology chapter and use the logistic transformation. By using this procedure, the ranking of observations provided by the modelled Z-score is transformed to a tangible probability of default.

Figure 6.1: ROC-Curve for Original Altman Model



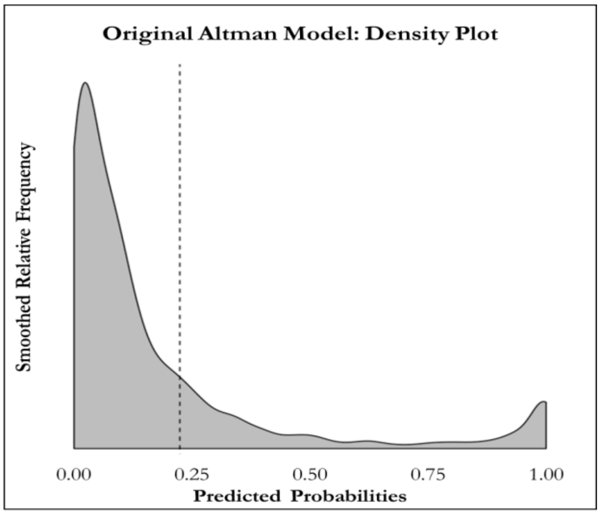
Source: Personal collection

The discriminative ability of the original Altman model is visualized on the ROC graph above. To recap, this graph expresses the quality of ranking associated with the classifier, i.e. the ability of the model to rank observations by increasing probability of default. In a perfect model all defaulting firms would be assigned a larger probability of default than the non-defaulting firms. The graph summarizes all the possible cut-off points of the model. An observation with a probability of default less than the cut-off point is classified as a non-defaulting firm and vice versa. At each cut-off point the associated sensitivity and specificity is mapped. To illustrate the discriminative mechanisms of the first Altman model we will zoom in on the top-left point of the ROC graph. The top-left point is displaying the cut-off point with the greatest proportion of correctly classified defaulting firms relative to the proportion of non-defaulting firms mistakenly classified.

Figure 6.2: Confusion Matrix and Density Plot in Original Altman Model

Original Altman Model – Top-left Cut-off Point

	Actual Observations	
Predicted by Model	Solvent	Default
Solvent	1140	52
Default	223	115



Source: Personal collection

The above density plot visualizes the frequency of the first Altman model's predicted probabilities. By examining the top-left point of the ROC graph we find that the associated cut-off point is at *22.4%*. Observations above this point will be classified as defaulting firms and this is visualized as the observations to the right of the vertical line. The confusion matrix beside the density plot tables the predicted classes relative to the actual classes of the firms in the testing set. The sum of the columns represents the actual number of defaults (167) and non-defaults (1363). Of the 1363 non-defaulting observations the model correctly classified 1140 and equivalently 115 of the 167 defaulting firms. We can see that the model predicted 223 false positives (Type I-error) and 52 false negatives (Type II-error). This corresponds to a false positive rate of *16.36%* and false negative rate of *31.13%*. If a false negative is costly, and we want to reduce this number we can choose a cut-off point associated with a higher sensitivity. However, this will come at the expense of increasing the number of false positives at a faster pace.

To further illustrate the mechanisms of the model, we can see on equation (6.2) and (6.3) how the probability of default is calculated for the guiding example of TCG. The corresponding Z-score is then transformed to a probability between zero and one using the logistic equation. The associated probability of default is *24.18%.* If the cut-off point is chosen at *22.4%* the model will correctly classify TCG as a defaulting firm.

$$\begin{aligned} Z &= -1.2 * -0.051 - 1.4 * -0.135 - 3.3 * 0.007 - 0.6 * 0.009 - 0.99 * 1.459 \\ Z &= -1.143 \end{aligned} \tag{6.2}$$

$$Predicted\ Probability\ of\ TCG\ Default = \frac{1}{1 + e^{-(-1.143)}} = 24.18\% \tag{6.3}$$

The Altman model has an AUROC of *80.74%* and the corresponding *95%* confidence interval is *76.87-84.61%*. The Somers' D associated with the Altman model is *0.6149* which further indicates a positive relationship between predicted probability of default and actual default.

6.1.2 Re-estimation of the Altman Model

As a consequence of the methodological choice of using a linear regression to re-estimate the coefficients of the Altman model, the ranking provided by the model will have only the majority of observations within the interval of 0 and 1. This implies that the size of the coefficients are not pairwise comparable between the two models. However, the direction and the relative within model size of the coefficients can be compared. The regression output for the re-estimated model can be seen in the table 6.1.

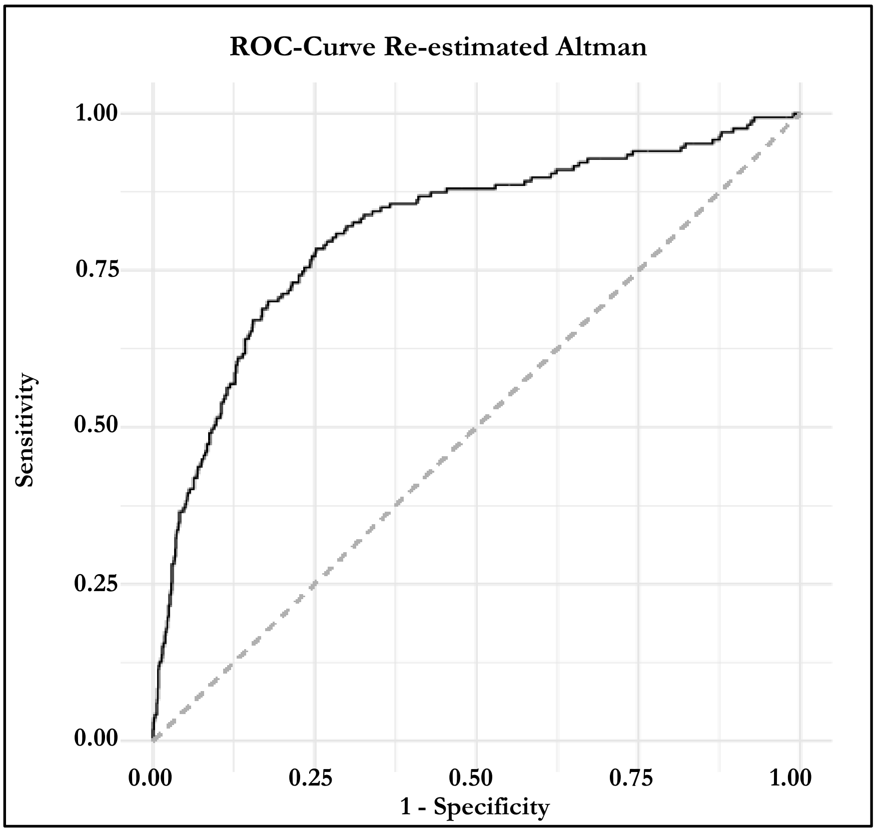
Table 6.1: Re-estimated Altman Model

	Default
	(1)
WC/TA	-0.065*** (0.016)
RE/TA	-0.006*** (0.001)
EBIT/TA	-0.177*** (0.013)
MV/BD	-0.003*** (0.001)
SA/TA	0.038*** (0.005)
Constant	1.087*** (0.008)

Note: 1) Model Coefficients. 2) Standard errors in parentheses.
3) *p<0.1; **p<0.05; ***p<0.01
Source: Personal collection

All of the variables including the intercept is significant at the *1%* level, signaling that the re-estimated model should possess discriminative power. Like in the original Altman model, the largest coefficient is the *EBIT/TA* ratio but the relative size of the remaining coefficients has changed. This is for instance reflected in *WC/TA* being ten times larger than *RE/TA* whereas they are close to identical in the original model. The largest deviation in the re-estimated model is related to the *SA/TA* ratio. This is the only ratio that has changed direction, i.e. an increase in the ratio is now associated with an increase in the predicted probability of default whereas the opposite was the case in the original model. Intuitively, the change in direction does not make sense and would present a sign of caution if the model was applied in practice. The Z-scores from the re-estimated Altman model is transformed to probabilities between zero and one by employing the logistic equation. The model's ranking ability is summarized in the below ROC graph. At first glance, the two Altman models' discriminative ability is very close to each other. The re-estimated Altman model does however seem to cover a greater area than the original as the top-left corner has been extended.

Figure 6.3: ROC-Curve in the Re-estimated Altman Model



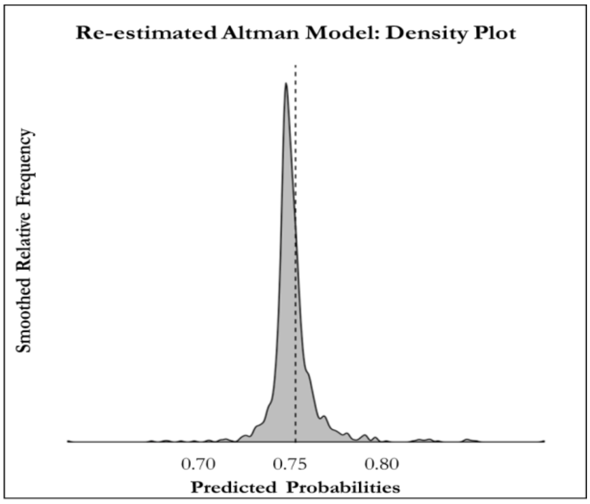
Source: Personal collection

As the Z-scores by design was already rather close to values between zero and one prior to the logistic transformation, the resulting frequency density plot of the predicted probabilities will not look like one shown earlier. This can be seen on the plot below. Our initial belief will consequently be that the cut-off point associated with the top-left point of the ROC graph will be high. If the cut-off point was low the false positive rate would be proportionally too high. We can examine this by looking at the confusion matrix corresponding to the aforementioned cut-off point below. The confusion matrix is the result of a cut-off point at *75.32%*, which, as expected, is higher than the original. The number of true default predictions is 131 and the number of wrong non-default prediction is 36. When looking at the number of true non-default predictions we observe 1020 and the number of mistakenly predicted defaults are 343. These predictions imply that at the top-left cut off point, the re-estimated model is more sensitive but less specific than the original.

Figure 6.4: Confusion Matrix and Density Plot in Re-estimated Altman Model

Re-estimated Altman Model – Top-left Cut-off Point

	Actual Observations	
Predicted by Model	Solvent	Default
Solvent	1020	36
Default	343	131



Source: Personal collection

We can illustrate how the re-estimated Altman model works by returning to the example of TCG. The critical Z-score value for the cut-off point of *75.32%* is 1.12. As TCG's Z-score of 1.145 and corresponding probability of default of *75.86%* is above this, the model correctly classifies the observation as a default. As with the original model, the predicted probability is just above the cut-off point.

$$\begin{aligned} Z &= 1.087 - 0.065 * -0.051 - 0.006 * -0.135 - 0.177 * 0.007 \\ &\quad -0.003 * 0.009 + 0.038 * 1.459 \\ Z &= 1.145 \end{aligned} \tag{6.4}$$

$$Predicted\ Probability\ of\ TCG\ Default = \frac{1}{1 + e^{-(1.145)}} = 75.86\% \tag{6.5}$$

The re-estimated Altman model's discriminative power, as summarized by the evaluation criteria, is slightly better than the original. The AUROC is *81.35%* with a confidence interval of *77.55-85.15%*. The corresponding Somers' D is 0.6270

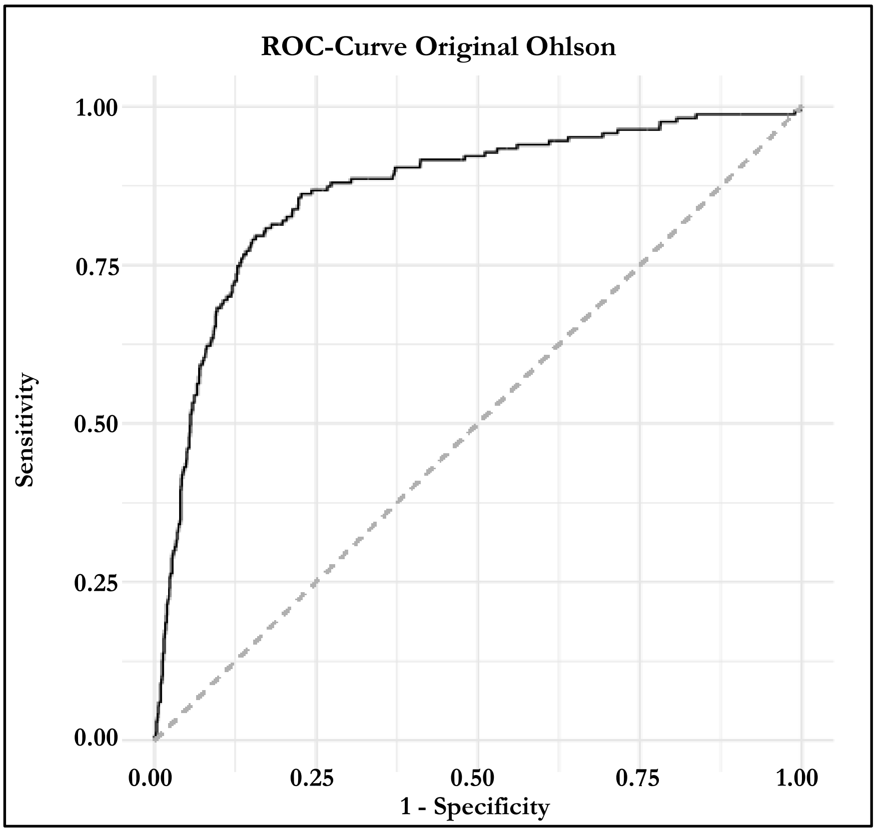
6.1.3 Original Framework of Ohlson

As presented in the methodology section of the thesis, the original Ohlson framework consists of nine features with coefficients estimated through a logistic regression. It yielded the following equation:

$$\begin{aligned} O = &-1.32 - 0.407 * Size + 6.03 * TL/TA - 1.43 * \mathrm{WC/TA} + 0.0757 * CL/CA \\ &-2.37 * NI/TA - 1.83 * \mathrm{FFO/TL} + 0.285 * NIDummy \\ &-1.72 * LevDummy - 0.521 * \mathrm{NIChange} \end{aligned} \tag{6.6}$$

As indicated by Ohlson in his paper of 1980, the ratios of *TL/TA, CL/CA,* and *NIDummy* should contribute to a larger probability of default. Likewise, should increases in S*ize, WC/TA, NI/TA, FFO/TL,* and *NIChange* ratio decrease the probability of default. At last, the *LevDummy* introduced by Ohlson could have a divergent effect on probability. As the score of the original framework of Ohlson come in the form of the log odds of default, we apply the logistic function to reach the probability of default. Having done so, we can create the ROC-curve for the original Ohlson framework seen below.

Figure 6.5: ROC-Curve in Original Ohlson Model



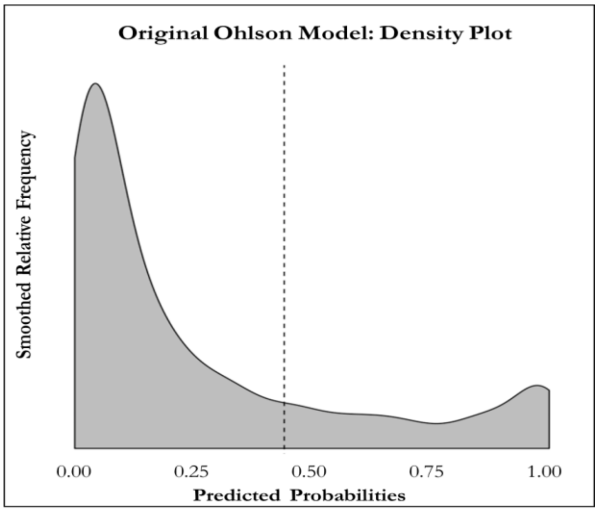
Source: Personal collection

From the curve we can find the top-left cut-off point, translated into the corresponding threshold value between zero and one. The cut-off equals *44.15%*, which indicates that companies with a default probability as calculated by Ohlson's original framework above this value is classified as default, and otherwise determined as solvent. At the cut-off point of *44.15%*, we can estimate the following confusion matrix.

Figure 6.6.: Confusion Matrix and Density Plot in Original Ohlson Model

Original Ohlson Model – Top-left Cut-off Point

	Actual Observations	
Predicted by Model	Solvent	Default
Solvent	1150	34
Default	213	133



Source: Personal collection

As such, the original Ohlson framework models 213 type I-errors, and 34 type II-errors. In total, it predicts 346 defaults, of which 133 are correct. In a similar manner, it predicts 1184 non-defaults, with 1150 correctly predicted non-defaults. Looking at the density plot of the predicted default probabilities, the 167 predicted defaults are the observations in the area to the right of the cut-off point. To illustrate the original model of Ohlson, we utilize TCG as an example by inserting those specific financial ratios into the framework. When applied to the framework, TCG receives a score 0.5242, and after being transformed by the logistic function, it is denoted by a probability of *62.81%*, well above the cut-off point of *44.15%*. By these means, the original Ohlson framework correctly predicts the default of TCG.

$$\begin{aligned} O &= -1.32 - 0.407 * 11.34 + 6.03 * 0.9557 - 1.43 * -0.051 \\ &\quad +0.0757 * 1.9981 - 2.37 * -0.0248 - 1.83 * 0.0586 \\ &\quad +0.285 * 0 - 1.72 * 0 - 0.521 * -1 \\ O &= 0.5242 \end{aligned} \tag{6.7}$$

$$Predicted\ Probability\ of\ TCG\ Default = \frac{1}{1 + e^{-(0.5242)}} = 62.81\% \tag{6.8}$$

Looking into the performance metrics of the original Ohlson framework, the area under the receiver operating characteristic curve equals *86.73%* with a confidence interval of *83.27 - 90.19%*. The Somers' D achieves a score of 0.7345.

6.1.4 Re-estimation of the Ohlson Model

When re-estimating the Ohlson framework on our portfolio of companies, we inherently estimate new coefficients for the original features. We do so by building a logistic regression model, with the default variable as the regressor, and the identical nine financial ratios as regressants. Doing so, we do not consider building a model where all coefficients are significant and comply with maximum likelihood estimation. This is reserved for the practical approach to default probability, which will be examined in the second part of the results. Thus, we stick to solely re-estimating the coefficients in the original framework.

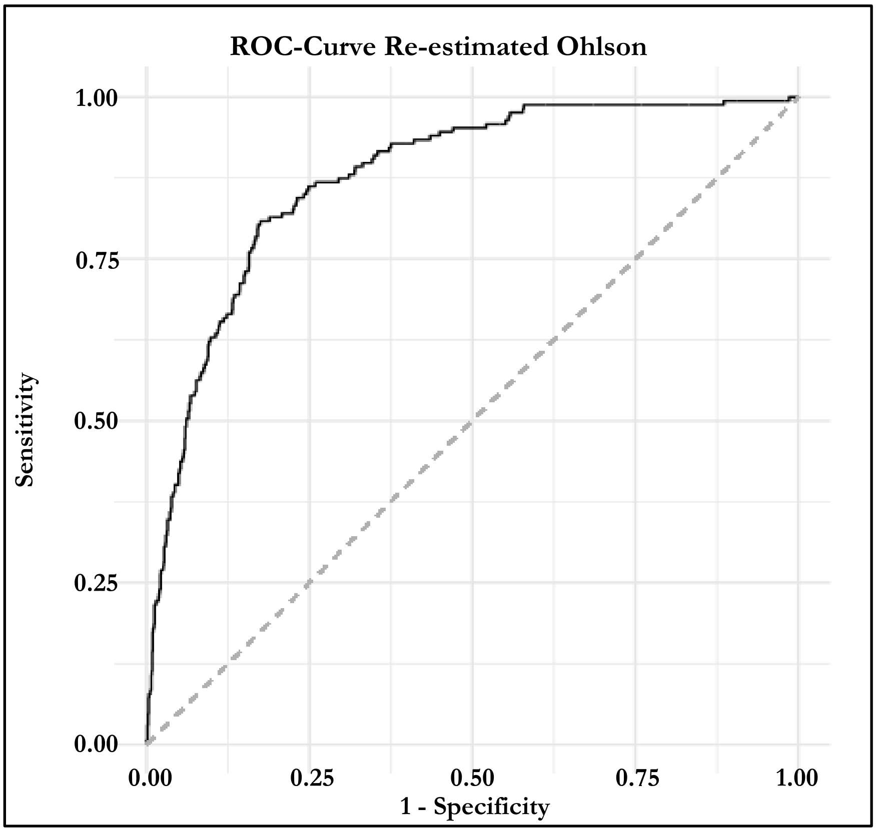
Table 6.2: Re-estimated Ohlson Model

	Default
	(1)
Size	-0.374*** (0.035)
TL/TA	2.176*** (0.244)
WC/TA	-0.021 (0.252)
CL/CA	-0.002 (0.013)
NI/TA	0.353*** (0.131)
FFO/TL	-0.197*** (0.054)
NIDummy	1.391*** (0.129)
LevDummy	-0.097 (0.253)
NIChange	-0.889*** (0.129)
Constant	-1.223*** (0.294)

Note: 1) Model Coefficients. 2) Standard errors in parentheses.
3) *p<0.1; **p<0.05; ***p<0.01
Source: Personal collection

Commenting exclusively on the regression of the re-estimation, all coefficients besides *WC/TA*, *CL/CA*, and *LevDummy* achieve significance at the *1%* level. In terms of indication of the coefficients, i.e. the direction of each coefficient, their effect shows relatively good alignment with Ohlson's framework. Of the significant regressants, *NI/TA* has a reverse relationship compared to the original model. It should be underlined that these changes make little economic sense, as a larger ratio of NI/TA should indicate a healthier financial state of the company, and thus a lower probability of default. Specifically, the interpretation of the variable at hand is that a one unit increase in the ratio equals an increase of 0.353 in the log odds of default. Ultimately, the ROC-curve for the re-estimated Ohlson model looks the following.

Figure 6.7: ROC-Curve in Re-estimated Ohlson Model

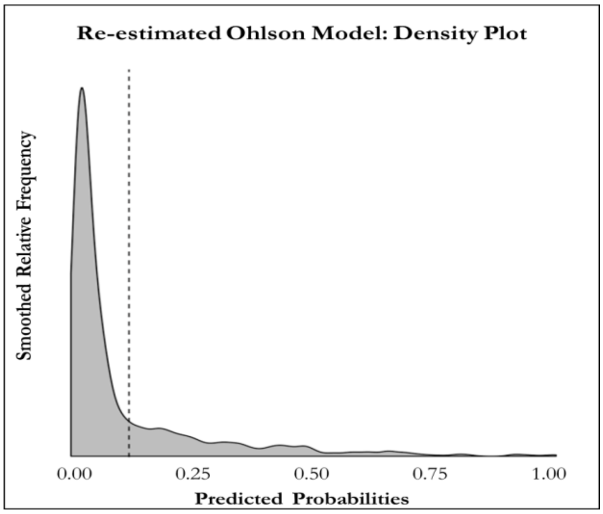


Source: Personal collection

From the top-left cut-off point at a threshold of *12.17%*, we can examine how the model segregates the two classes. This is a substantially different value than from the application of the original Ohlson model, which indicates that the re-estimation classifies the observations with a lower probability of default score. Below we present the resulting confusion matrix, which shows 238 type I-errors and 32 type II-errors. Compared to the original framework, the re-estimated Ohlson model correctly predicts two more defaulting companies, at the expense of labelling 25 more solvent companies as default. The default predictions are indicated on the density plot at the right.

Figure 6.8: Confusion Matrix and Density Plot in Re-estimated Ohlson Model

Re-estimated Ohlson Model – Top-left Cut-off Point		
	Actual Observations	
Predicted by Model	Solvent	Default
Solvent	1125	32
Default	238	135



Source: Personal collection

In terms of the TCG, the thesis' guiding example, it showcases how the discriminative ability of the Ohlson re-estimation model is different from the original one. This model gives the example a probability of default of 7*.45%*, which is below the cut-off value. This indicates that the re-estimated Ohlson model would incorrectly classify TCG as non-default, i.e. a false negative.

$$\begin{aligned} O = & -1.223 - 0.374 * 11.34 + 2.176 * 0.9557 - 0.021 * -0.051 \\ & -0.002 * 1.9981 + 0.353 * -0.0248 - 0.197 * 0.0586 \\ & +1.391 * 0 - 0.097 * 0 - 0.889 * -1 \\ O = & -2.52 \end{aligned} \tag{6.9}$$

$$Predicted\ Probability\ of\ TCG\ Default = \frac{1}{1 + e^{-(-2.52)}} = 7.45\% \tag{6.10}$$

The AUROC equals *87.54%* with a confidence interval of *85.2 - 89.88%*. As such, in spite of the re-estimated framework being unable to predict the default of Thomas Cook, the model segregates the two groups well across all possible cut-off values. The value of Somers' D is 0.7508.

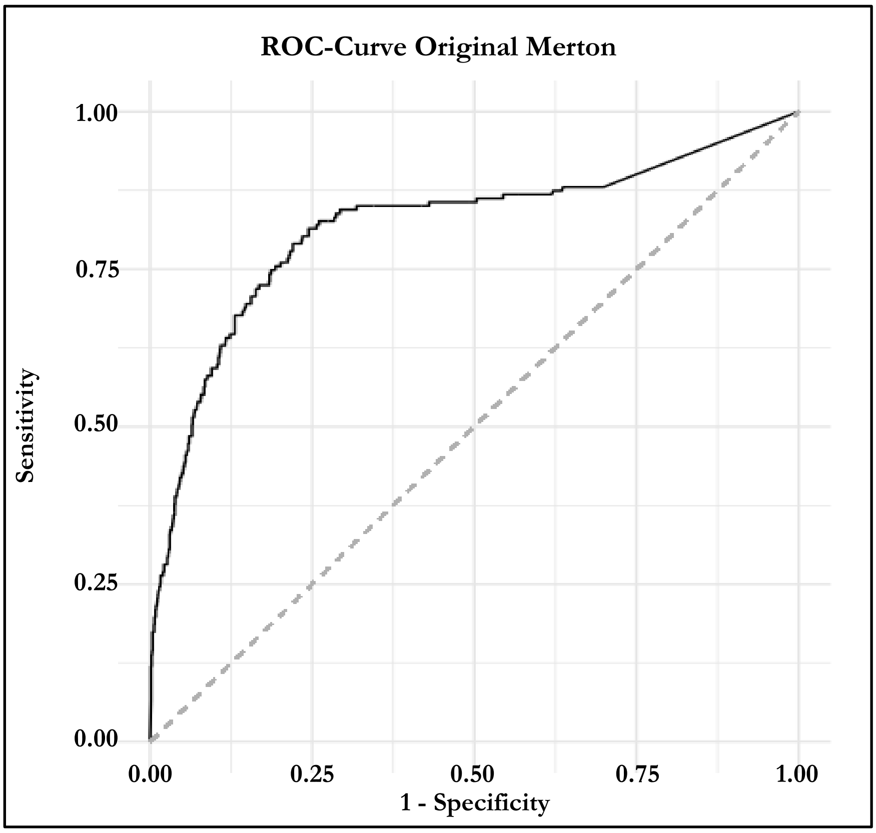
6.1.5 The Original Framework of Merton

We apply the European call contingent claim model through the original Merton framework. As stated in the methodology section of the thesis, this framework hypothesizes that default occurs if the value of the assets falls below the level of the firm's liabilities at maturity, which would happen through equity holders exercising their walk away option on the assets. We apply the framework through calculating market capitalization, stock returns and equity volatility to solve equation (4.9) and (4.12) before calculating probability of default in equation (6.11).

$$Probability\ of\ Default = N\left(-\frac{ln\left[\frac{A}{L}\right]+\left(\mu-\frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right) \tag{6.11}$$

The predicted values of the original Merton model measure the number of standard deviations the expected logarithm of asset value is from default, i.e. from the logarithm of liabilities. As presented in the methodological part of the thesis, this value is lognormally distributed. We can consequently calculate the predicted probability of default. Put differently, we can calculate how close it is to the colored region in figure 4.1. The ROC-curve for the model is presented below:

Figure 6.9: ROC-Curve in Original Merton Model



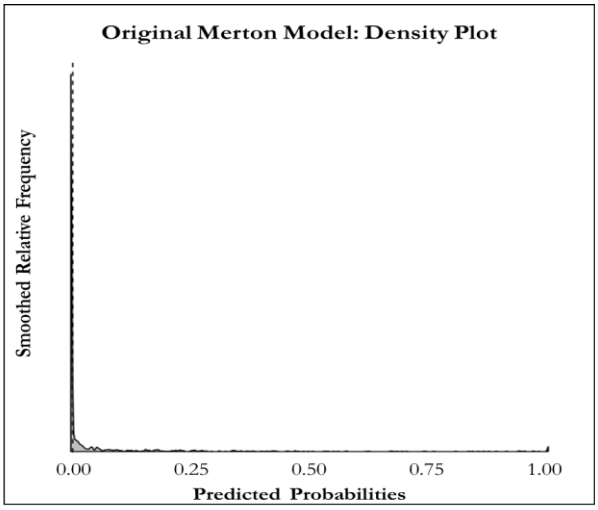
Source: Personal collection

The top-left cut-off point for the model is *0.05%*, which reflects a relatively low predicted probability of default mean. It implies, that non-defaulting firms are assigned a virtual to nothing probability of default. This is also visible from the ROC plot as the line is flat when the false positive rate is above *21%.* At this threshold, the Original Merton model has 300 type I-errors and 35 type II-errors. As can be seen on the density plot, the predicted probabilities are virtually zero for a large part of the observations.

Figure 6.10: Confusion Matrix and Density Plot in Original Merton Model

Original Merton Model – Top-left Cut-off Point

	Actual Observations	
Predicted by Model	Solvent	Default
Solvent	1063	35
Default	300	132



Source: Personal collection

In terms of our example, the distance to default as calculated by the model is 0.78. As such, Thomas Cook's expected asset value in the forecast horizon of one year is 0.78 standard deviations from default. This translates into a *21.63%* probability of default. At the chosen cut-off point, the model correctly classifies Thomas Cook as a default observation and with a large margin compared to the previous examples.

$$\begin{aligned} &Predicted\ Probability\ of\ TCG\ Default \\ &= N\left(-\frac{ln\left[\frac{7948054.26}{8187143.77}\right]+\left(5.91\%-\frac{1}{2}3.66\%^2\right)T}{3.66\%\sqrt{1}}\right) \\ &= 21.63\% \end{aligned} \tag{6.12}$$

The evaluation metrics for the original Merton framework reports that the AUROC is equal to *81.71%* with a confidence interval of *77.52 - 85.90%* and the Somers' D value is 0.5981.

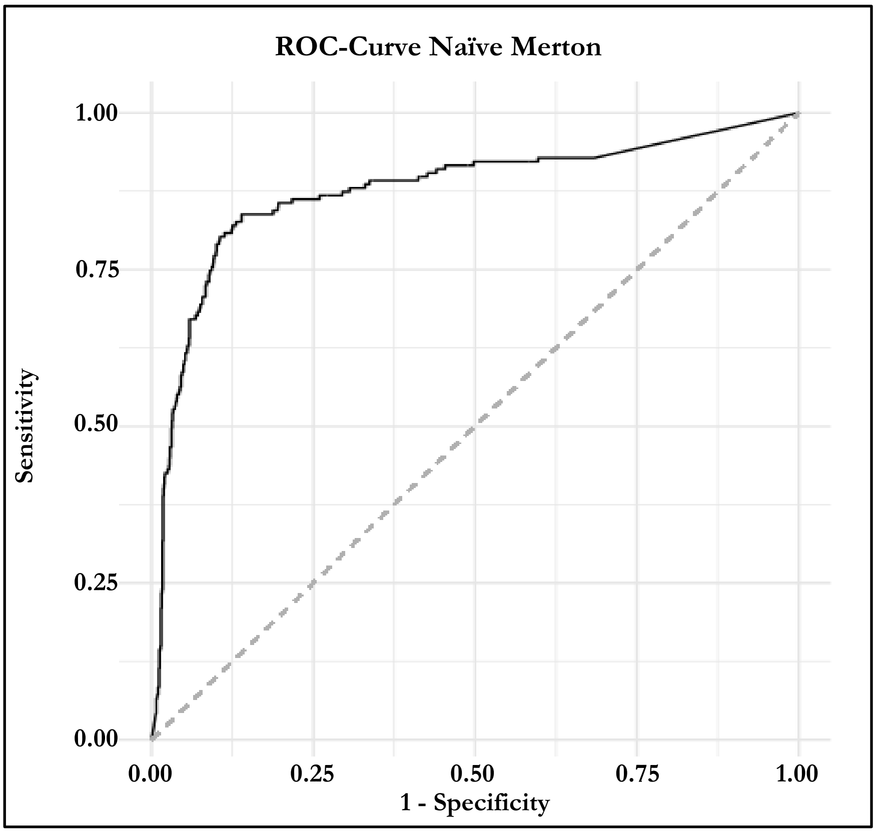
6.1.6 The Naïve Approach to the Merton Framework

In the re-estimated Merton model, or the Naïve approach to the contingent claim model, we utilize the computationally less heavy Bharath and Shumway approach (2008). We follow the simplifications of the framework put forward in the methodological chapter. This includes the value of liabilities now equaling current liabilities plus one half of the long-term debt, and the volatility of debt being a transformation of the volatility of equity. Doing this, we simplify the complex calculation of the unobservable variables in the original Merton framework in equation (6.11) and instead use equation (6.13).

$$Probability\ of\ Default = N\left(-\frac{ln\left[\frac{A_N}{L_N}\right] + \left(\mu_N - \frac{1}{2}\sigma_N^2\right)T}{\sigma_N\sqrt{T}}\right) \tag{6.13}$$

Having done the calculation described above, we are capable of calculating the distance to default in the Näive Merton framework. With equation (6.13) we transform the distance to default to a probability of default and estimate the ROC-curve below, which clearly covers a greater area than the original.

Figure 6.11: ROC-Curve in Naïve Merton Model



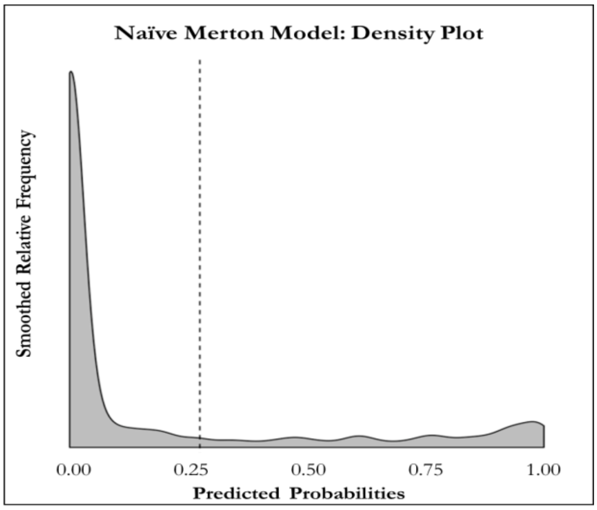
Source: Personal collection

The top-left cut-off point is *46.5%* from which we estimate the confusion matrix. At this threshold, the model has 190 type I-errors and 27 type II-errors. Thus, it correctly predicts 140 of the defaulted firms in the portfolio. On the right is the density graph of the predicted probabilities, with the indication of the cut-off point and the indication of the predicted defaulted companies.

Figure 6.12: Confusion Matrix and Density Plot in Naïve Merton Model

Naïve Merton Model – Top-left Cut-off Point

	Actual Observations	
Predicted by Model	Solvent	Default
Solvent	1173	27
Default	190	140



Source: Personal collection

The model's differing assumptions for the value of assets, asset volatility and debt give a very different picture of TCG than we have seen in the prior examples. Following the calculation presented below, the predicted probability of default for TCG is *98.74%*. This is well above the cut-off point and the resulting classification is a clear default. The evaluation metrics for the Naïve Merton model include an AUROC of *87.63%* with a confidence interval of *84.15 - 91.11%* and the Somers' D equals 0.7295.

$$
\begin{aligned}
&Predicted\ Probability\ of\ TCG\ Default \\
&= N\left(-\frac{ln\left[\frac{70175.81+6294893.75}{6294893.75}\right]+\left(-88.78\%-\frac{1}{2}43.35\%^2\right)T}{43.35\%\sqrt{1}}\right) \\
&= 98.74\%
\end{aligned}
\tag{6.14}
$$

6.1.7 Overview and Evaluation of Hypothesis 1

In the following overview section, we will emphasize the aspects of the six applied models, the three fundamental PD-frameworks and their re-estimations, which allow us to conclude upon the first hypothesis: *The classic academic approaches to default probability have discriminative power on a modern portfolio.* We will thus neglect some aspects of the specific models, and rather frame the focus of the overview on the evaluation parameters of the AUROC and Somers' D.

Table 6.3: Overview of Models

	Training AUROC	Testing AUROC	Somers' D	Solvent PD-Mean	Defaulted PD-Mean	Portfolio PD-Mean
Original Altman	80.65% [78.43-82.87]	80.74% [76.87-84.61]	0.6149	13.84%	49.12%	17.69%
Re-estimated Altman	80.01% [77.78-82.24]	81.35% [77.55-85.15]	0.6270	75.01%	76.64%	75.17%
Original Ohlson	87.08% [85.39-88.77]	86.73% [83.27-90.19]	0.7345	20.51%	70.57%	25.97%
Re-estimated Ohlson	87.26% [85.62-88.90]	87.54% [85.20-89.88]	0.7508	8.20%	33.57%	10.97%
Original Merton	80.72% [78.48-82.96]	81.71% [77.52-85.90]	0.5981	2.56%	26.08%	5.13%
Naïve Merton	88.74% [87.02-90.46]	87.63% [84.15-91.11]	0.7295	11.08%	69.89%	17.50%
Portfolio Total	-	-	-	-	-	10.95%

Source: Personal collection

As emphasized in the table above, all the three academic approaches to default probability reach an AUROC above 80% for both the training and testing set. This metric is specifically interesting as it allows us to compare the discriminative power between the models, which is most interesting on the testing data set, as it gives a real indication of the given model's discriminative power. The AUROC for the training portfolio has two properties. First, it is a measure of explanatory power similar to R^2 for regression models in classical statistics. Second, only minor deviations between training and testing AUROC implies a well-fitted model. However, we reinstate that the ultimate evaluation parameter is that of AUROC for the testing portfolio. Of the accounting-based models, the re-estimated Ohlson model outperforms it peers with an AUROC of 87.54%. This is however improved upon by the Naïve Merton model, which achieves the highest score across all models with 87.63% AUROC. However, in terms of the Somers' D, the Naïve Merton performs worse in relation to both of the logistic regressions. This indicates that the structural model is penalized relative more for having more ties within the predicted values, which prevents it from ordering the predictions accordingly.

Within the accounting-based paradigm of the academic approach to default probability, both the re-estimation of the Altman and the Ohlson model only slightly increase the discriminative power. This is aligned with the research of Hillegeist et al. (2004), which also found that the re-estimation of these two frameworks yielded very similar results to the original applications. With respect to the Merton model, the Naïve and less computationally heavy approach improve the discriminative power significantly, both in terms AUROC and Somers' D. While it may seem counterintuitive that a simpler model provides better segregation of defaults and solvent firms, it is in line with findings of both Bharath and Shumway (2008) and Jackson and Wood (2013).

The central aim of the analysis in this section has been to assess the discriminative power for all of the six models applied to our modern dataset. The above comparison and differentiation of the models with respect to the AUROC metric is consequently based on how well each model *rank* firms. Another evaluation metric independent of a model's discriminative power is *calibration*. Rather than a measure of ranking, calibration is related to what extent the models' predicted probabilities depart from the in-sample default frequency. This is an important point and is related to the interpretable part of *probability* in probability of default. As such, models can be said to have high discriminative power while at the same time being poorly calibrated. This has implications for the applicability of a model in a corporate loan setting. If a model is to be said to provide the user with a probability in its literal sense, it has to be well-calibrated. A poorly calibrated model is consequently providing the user with a score rather than a probability per se.

To recognize the importance of this aspect in model evaluation the above table includes the mean predicted probability of default. This is included for the testing sample in total as well as for the two classes. All the models give a higher mean predicted probability of default for the group of actually defaulted firms than they do for the non-defaulted firms. However, the actual proportion of defaults in the testing set is *10.95%* and it is evident that the majority of the models deviate from this proportion. The deviation is clearest for the re-estimated Altman model with a mean probability of *75.17%*. The model is capable of ranking the observations, but it is not well-calibrated and resembles a scoring-model rather than a probability-model. The re-estimated Ohlson model is a contrast with a mean probability of 10.97%. The difference between the two models reflects the two methodological branches. As the re-estimated Ohlson model is a result of a logistic regression it is by design calibrated to the data.

Ultimately, the findings of the first section of the results indicate that the three academic approaches to default probability segregate the classes of default and non-default companies well in the testing data set. It is emphasized with all models achieving an AUROC above

80%. Furthermore, it comes with the perspective that the re-estimations of the models only slightly improve their power, which only underline the strength of the original frameworks. As we in this research are interested in the models' ability to segregate classes rather than their ability to present an accurate probability of default, captured in calibration, we are able to verify the first hypothesis: *The classic academic approaches to default probability have discriminative power on a modern portfolio.*

Table 6.4: Verification of Hypothesis 1

#	Hypothesis	
1	*The classic academic approaches to default probability have discriminative power on a modern portfolio*	✔

6.2 The Practical Approach to Default Probability

The second part of the results chapter will revolve around the practical approach to default probability. As such, we will build two models in this part. The first utilizes accounting ratios as features exclusively, which is the most exact mimic of the methodology used in practice (Engelmann & Rauhmeier, 2011; De Laurentis et al., 2010; Neisen & Rosch, 2018). The second model will incorporate the structural approach to credit risk modelling, and thus establish a "Default Model of Synthesis" within the default probability literature. Afterwards, the section will provide an overview and conclude upon the second hypothesis of the research: *the practical model outperforms the academic but is improved through a synthesis of market and accounting theory.*

6.2.1 The Practitioners' Default Probability Model

As outlined above we intend to estimate a model mimicking one a practitioner will employ in a professional setting. As such, the variables used as inputs in the model should reflect those available to professionals. We acknowledge that those variables presented by Altman and Ohlson are not a complete list of financial ratios. To reflect this and to approximate the long-list format of financial ratios we have calculated four additional variables that is derived from Plenborg et al. (2017): *FFO/CL*, *WC/SA, FFO/CAPEX* and *TL/EBITDA*. We therefore have a long-list of 14 financial ratios and Ohlson's two dummy variables. The model can include a combination of variables from both the Altman and Ohlson framework in addition to the new range of variables we have included. This implies that we are now not constrained in terms of the mix of variables used as inputs in the model.

We are, however, constrained in terms of whether the individual variables satisfy the specific statistical requirements to be included in the final model. The section will consequently be driven by the methodological steps of the practical approach presented earlier in this thesis and we will in the following paragraphs walk through each to reach the final model.

The logistic model will, like the previous models, employ the binary default label variable as the dependent variable. In our preliminary long-list of variables, we assess the expected direction of each as the first step. That is, grounded in economic reasoning, the expected effect of a predictor will be negative if an increase will influence the probability of default downwards, *ceteris paribus*. This can be exemplified through an expected negative direction of the *FFO/CL* ratio, as a higher ratio indicates that the firm has liquidity to meet its short-term payables, which is associated with a lesser likelihood of bankruptcy. We couple the expected directions of the newly added variables with the directional relationships already highlighted by Altman and Ohlson. These can be seen on the table 6.5. The second pre-modelling step involves examining influential cases that will exert a disproportionate influence on the estimated parameters. Even though all the financial ratios have been calculated on the winsorized data, practitioners emphasize that extreme ratios still present should be removed. This is exemplified by the observation of nCoat Inc., which in their defaulting year of 2010 had total liabilities of 920 million, and -340 thousand in EBITDA, giving them a *TL/EBITDA* ratio of -2,703. By plotting the variables we identify ratios with such extreme influential data points. We impose round-numbered cut-off points to exclude these, and those together with the following number of removed observations can be found in table 6.5. In total, we remove 142 outliers, which essentially have no effect on the overall size of the portfolio.

Table 6.5: Predictor Cut-offs and Direction

Variable	Larger Than Cut-off	Less Than Cut-off	Outliers Removed*	Expected Direction	Observed Direction
WC/TA	-	-	-	↓	↓
RE/TA	-100.00	-	2	↓	↓
EBIT/TA	-	10.00	1	↓	↓
MV/DB	-	60.00	25	↓	↓
SA/TA	-	-	-	↓	↑
Size	-	15.00	2	↓	↓
TL/TA	-	15.00	2	↑	↑
CL/CA	-	33.00	3	↑	↑
NI/TA	-10.00	-	1	↓	↓
FFO/TL	-	-	-	↓	↓
NIChange	-	-	-	↓	↓
FFO/CAPEX	-200.00	200.00	55	↓	↓
TL/EBITDA	-200.00	200.00	60	↑	↑
FFO/CL	-	-	-	↓	↓
WC/SA	-200.00	200.00	2	↑	↑
LevDummy	-	-	-	↑↓	↑
NIDummy	-	-	-	↑	↑

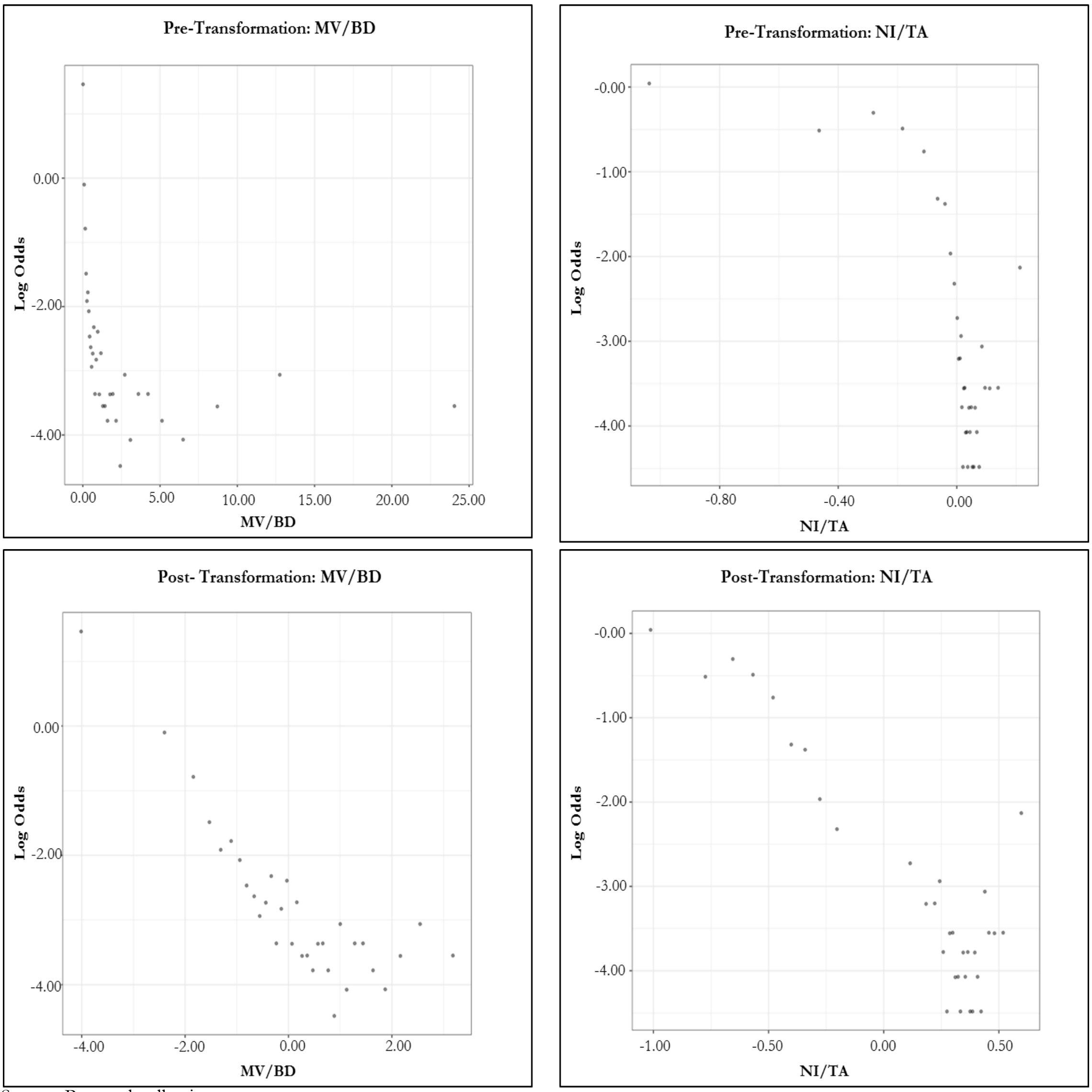
*The elimination of influential values is performed separately from the full data set. Thus, cut-offs in different variables may remove the same observations. In total, 142 observations are removed.

Source: Personal collection

The main statistical assumption of a logistic regression is that the logit of *y*, in our case the logit of the probability that the outcome is default, has a linear relationship with the predictors. This assumption makes intuitive sense by viewing it through an interpretation of the coefficients. If the relationship is linear in the logit, then the change in the logit of default for a one-unit change in a financial predictor is constant. If the relationship is non-linear in the logit, then the change is not constant and depends on the specific value of the predictor. To ensure that this assumption is satisfied we employ the check for linearity procedure suggested by Hayden (2011) and originally presented by Hosmer and Lemeshow (1980). For a given financial predictor we aggregate observations into groups defined by the value of the predictor. The first group contains the top *3%* values the next values from *94%* to *97%* and so forth, resulting in a total of 33 groups. Hayden suggests that the data is aggregated into 50 groups, however, to ensure that each group includes defaulted observations we choose 33 groups. In each group we calculate the empirical default rate, i.e. the mean of the dependent default variable, the associated logit and the median of the predictor value. To check for linearity the logit and median predictor value is then plotted against each other. The entire process is repeated for each individual financial predictor on our long-list and all the plots can be seen in Appendix 12.1.

From the plots we detect evidence of non-linearity among some of the predictors. This implies that we need to transform the variables in order to approach linearity with the logit before we can use them as predictors in the model. We apply a quadratic, cubic and logistic transformation before deciding which works best for the specific nonlinear variable. For each

Figure 6.13: Transformation Plots for *MV/BD* and *NI/TA*

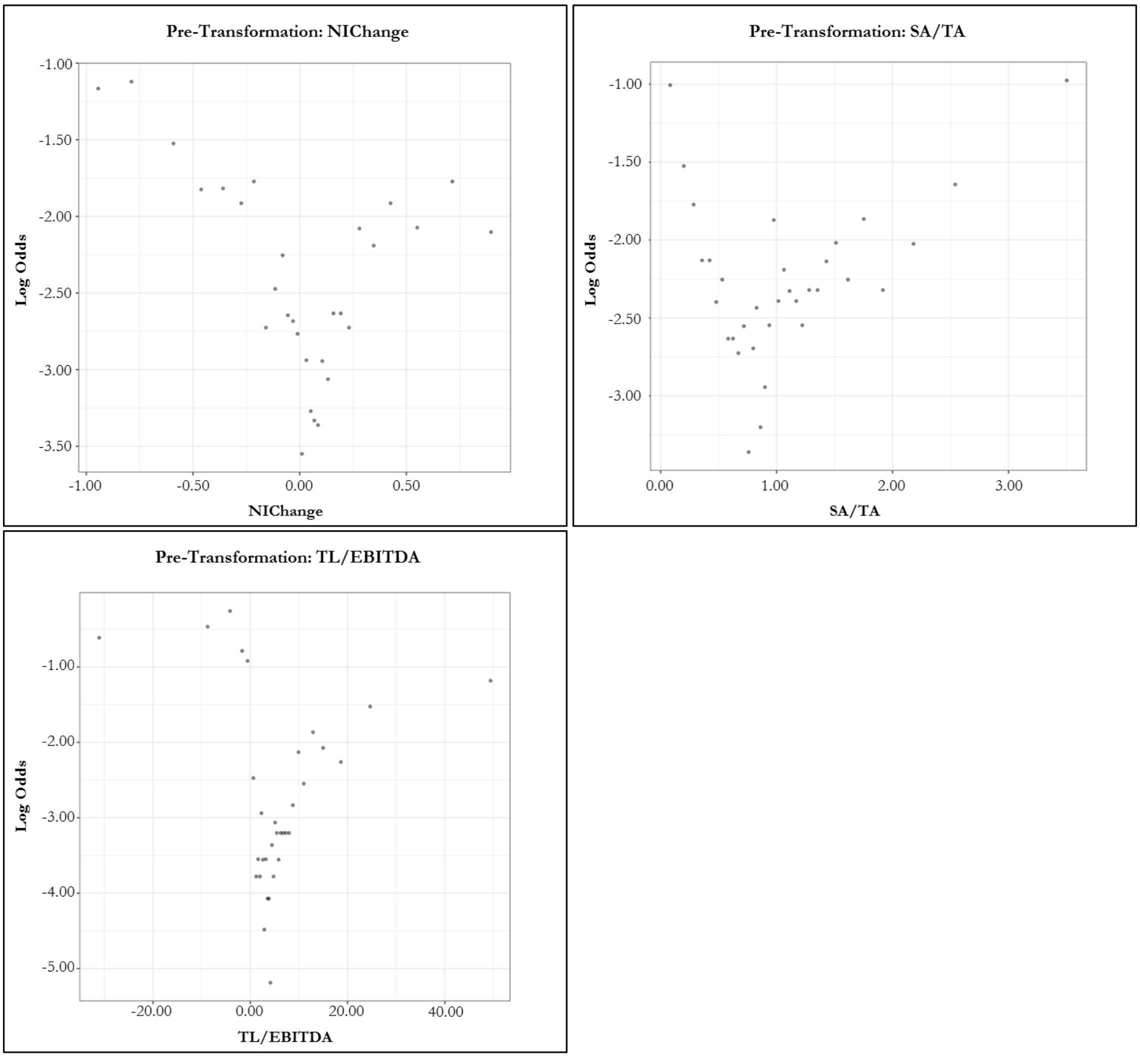


Source: Personal collection

transformation the above check-for-linearity procedure is repeated before re-plotting the relationship. Examples of such transformations can be seen in figure 6.13. Here we apply a logistic transformation to the MV/BD variable and a cubic transformation of the NI/TA variable, and the effect is clearly visible. From the plots we can also infer the directional relationship. As such, our initial belief that the direction of MV/BD and NI/TA was negative is confirmed.

Further following the recipe provided by Hayden, we engage in the variable selection process, as the optimal default model in practice is built with few rather than many predictors. We remove financial ratios due to four parameters. First, we remove variables with a directional relationship not consistent with economic reasoning. Secondly, we remove variables that do not satisfy the linearity assumption even after transformation. Thirdly, we remove highly correlated predictors. Lastly, we remove predictors with relatively low Somers'

Figure 6.14: Linearity Plots for *NIChange*, *SA/TA* and *TL/EBITDA*



Source: Personal collection

D metric (Engelmann & Rauhmeier, 2011). After transformations, the only variable with a conflicting directional relationship is the *SA/TA* ratio, and it is consequently excluded. The variables of *NIChange* and *TL/EBITDA* did not satisfy the linearity assumption and is also excluded. The plots for these three excluded predictors can be seen in figure 6.14.

For each of the remaining 13 variables we calculate the corresponding Somers' D metric which is used for the following two selection procedures. In the correlation matrix in table 6.6, we group highly correlated variables, that is variables with a correlation of more than 0.5 (Hayden, 2011). Three of such groups exists: a group related to profitability of *RE/TA, NI/TA* and *EBIT/TA*; a group related to leverage of *WC/TA, CL/CA* and *TL/TA*; and a group related to liquidity of *FFO/TL* and *FFO/CL*. In each of these three groups, only the variable with the largest Somers' D is included in the next step and all other variables are excluded. This implies that we keep *RE/TA, TL/TA* and *FFO/TL*. Of the remaining eight uncorrelated variables we lastly exclude variables with a Somers' D below 0.25. The *WC/SA* variable is excluded on this basis.

Table 6.6: Correlation Matrix - Practical Model of Default Probability

Ratio	Somers' D	WC/TA	RE/TA	EBIT/TA	MV/BD	Size	TL/TA	CL/CA	NI/TA	FFO/TL	FFO/CAPEX	FFO/CL	WC/SA
WC/TA	0.2090	1.00	0.19	0.32	0.22	-0.06	-0.57	-0.53	0.35	-0.04	0	-0.09	0.43
RE/TA	0.6508		1.00	0.57	-0.02	0.36	-0.31	-0.18	0.56	0.31	0.21	0.27	0.06
EBIT/TA	0.6454			1.00	0.02	0.37	-0.36	-0.27	0.88	0.39	0.22	0.35	0.01
MV/BD	0.6526				1.00	0.09	-0.28	-0.16	0.03	-0.03	-0.06	-0.07	0.11
Size	0.5122					1.00	-0.07	-0.07	0.35	0.21	0.17	0.23	0.11
TL/TA	0.4802						1.00	0.55	-0.46	-0.04	-0.08	-0.03	-0.24
CL/CA	0.3852							1.00	-0.31	-0.02	-0.03	-0.01	-0.25
NI/TA	0.6442								1.00	0.3	0.2	0.27	0.06
FFO/TL	0.5120									1.00	0.39	0.82	-0.17
FFO/CAPEX	0.5074										1.00	0.36	-0.07
FFO/CL	0.5088											1.00	-0.24
WC/SA	0.2438												1.00

Source: Personal collection

We are now left with one dummy variable and six financial ratios that all satisfy the linearity assumption, exhibit a Somers' D above 0.25 and capture distinctive effects with respect to predicting default. With the remaining predictors the initial modelling can begin where the goal is to build a parsimonious logistic regression, using the technique highlighted by Hayden. A parsimonious model is one which minimize the number of features while capturing the desired level of explanatory power, which in the case of this thesis means predicting default within the portfolio most efficiently (Wooldridge, 2016). As it is infeasible to model all possible combinations of financial predictors, we start by including all of the variables and gradually remove the least significant ratio until all coefficients show significance. Thus, it is the methodology of stepwise regression using a backwards elimination process (ibid).

Table 6.7: Practical Model of Default

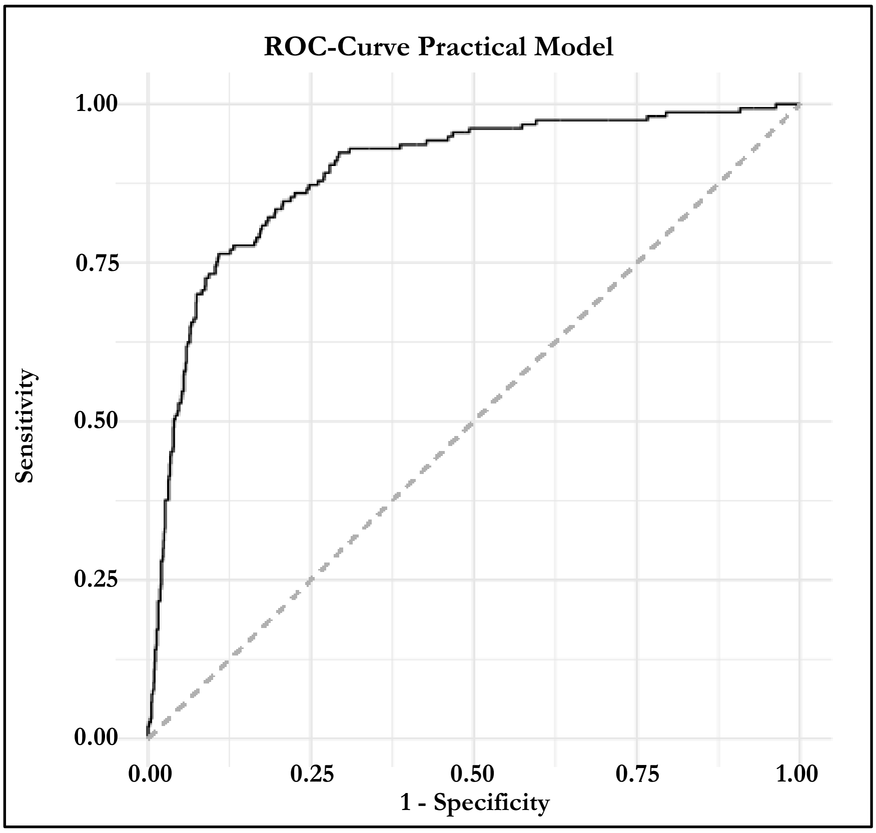
	Default		
	(1)	(2)	(3)
RE/TA	-0.040 (0.081)		
FFO/CAPEX	-0.003 (0.004)	-0.003 (0.004)	
MV/BD	-1.022*** (0.057)	-1.019*** (0.057)	-1.017*** (0.057)
Size	-0.472*** (0.046)	-0.481*** (0.043)	-0.481*** (0.043)
TL/TA	1.258** (0.532)	1.219** (0.516)	1.203** (0.516)
FFO/TL	-0.838*** (0.140)	-0.858*** (0.134)	-0.908*** (0.121)
NIDummy	0.679*** (0.155)	0.705*** (0.146)	0.704*** (0.145)
Constant	1.632*** (0.539)	1.642*** (0.539)	1.637*** (0.539)

Note: 1) Model Coefficients. 2) Standard errors in parentheses. 3) $*p<0.1$; $**p<0.05$; $***p<0.01$

Source: Personal collection

We start by including the remaining seven predictors, which yields a model where the *FFO/CAPEX* variable and the ratio of *RE/TA* are the only insignificant predictors at a *1%* significance level. The ratio of *RE/TA* has the lowest T-statistic, wherefore we remove it and re-run the logistic model according to our backward elimination methodology. We find that *FFO/CAPEX* is still insignificant. The variable is excluded, and we are left with a practical probability of default model including five features that are all significant at the *5%* level and all beside *TL/TA* are significant at the *1%* level. The elimination procedure can be seen on table 6.7 and the associated ROC curve is visualized in figure 6.15.

Figure 6.15: ROC-Curve in Practical Model



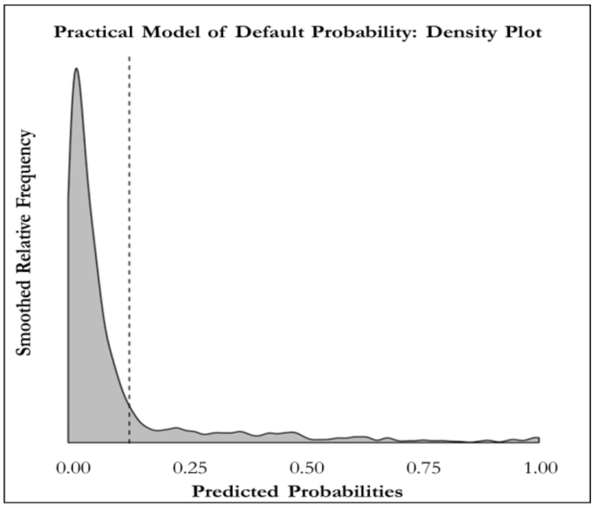
Source: Personal collection

The threshold associated with the top-left point in the ROC graph is *11.32%*. Here the practical model correctly classifies 111 defaulted observations and 1226 non-defaulted observations. By examining the mistakenly predicted classes, the model makes 113 Type I-errors and 46 Type II-errors.

Figure 6.16: Confusion Matrix and Density Plot in Practical Model

The Practical Model – Top-left Cut-off Point

	Actual Observations	
Predicted by Model	Solvent	Default
Solvent	1226	46
Default	113	111



Source: Personal collection

In terms of the example of TCG, the predicted probability of default according to the model is given by equation (6.15) and (6.16):

$$\begin{aligned} &\mathit{Predicted\ Log\ Odds\ TCG\ Default} \\ &= 1.637 - 1.017 * -4.759 - 0.481 * 11.34 + 1.203 * 0.985 \\ &\quad -0.908 * 0.389 + 0.704 * 0 \\ &= 1.86 \end{aligned} \tag{6.15}$$

$$\mathit{Predicted\ Probability\ of\ TCG\ Default} = \frac{1}{1 + e^{-(1.86)}} = 86.48\% \tag{6.16}$$

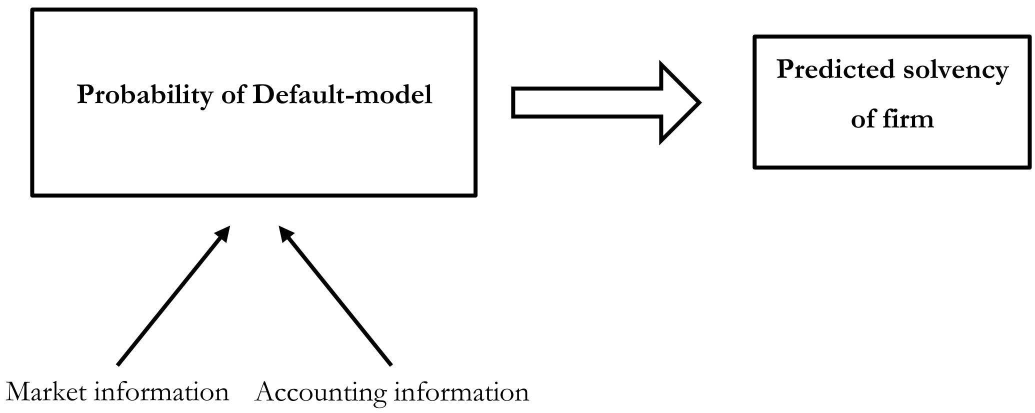
The observed values used in the calculation are the transformed ratios. We note that the corresponding probability of *86.48%* is above the cut-off value why the model correctly classifies TCG as a defaulting observation.

The final model, built using the practitioners' methodological steps, cuts the long-list of 14 financial ratios and two dummy variables to four ratios and one dummy. To evaluate the model's discriminative power, we employ it to classify on the out-of-sample testing set. The result is an AUROC of 89.5% with a confidence interval of 8*6.83 - 92.17%* and a Somers' D of 0.79.

6.2.2 The Default Model of Synthesis

The thesis' second hypothesis also concerns the clash between the two schools of probability of default modeling. The models within each academic school differ in terms of theoretical underpinning, ease of application, calculatory complexity and popularity. We recognize that both approaches have strengths and weaknesses and that each school may capture different information. In the assessment of the first hypothesis we uncovered that on our modern dataset the Naïve Merton model possessed the greatest discriminative power, however it was closely followed by the re-estimated Ohlson model. Rather than arguing in favor of one school, the thesis seeks to place itself in between this paradigm clash. To achieve a greater discriminative power, we will in the following evaluate a model built using the practical methodological framework that employs both accounting and market-based information. An illustration of this principle is shown below.

Figure 6.17: Probability of Default Model: The Default Model of Synthesis

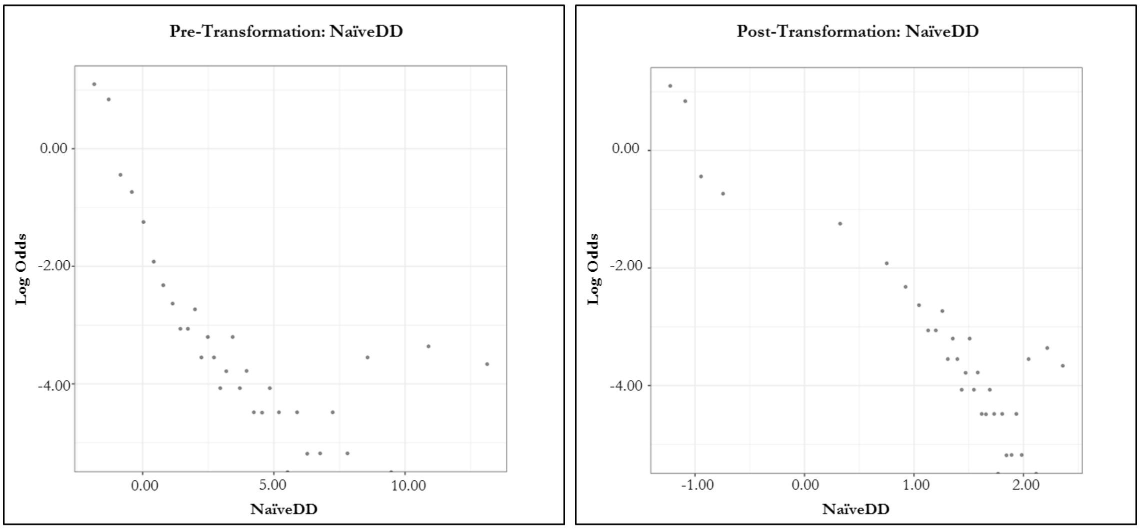


Source: Personal collection

By these means, we want to introduce the Naïve Merton to the practical PD-model. We choose the Naïve rather than the original Merton framework, as it showed to have a higher discriminative power in the first section of our research. Rather than inserting the Naïve Merton's probability of default into the logistic regression model directly, we comply with the point presented by Reisz and Perlich (2007) that an independent variable in the form of a probability is not consistent with the logistic model. As such, we utilize the *NaïveDD*, which is the value relied upon by the Naïve Merton model to calculate the probability of default. The distance to default from the Naïve Merton model is a measure of proximity to default. A low value is therefore an indication of default why the expected directional relationship is negative. We winsorize the *NaïveDD* variable prior to utilizing the parameter as a predictor, in order to keep it consistent with the remainder of the data set.

Both the linearity assumption and the correlation to other variables is investigated for the *NaïveDD*, in order to ensure the consistency with the properties of the practical approach to default modelling (Engelmann & Rauhmeier, 2011). Both investigations show that the variable is applicable for the desired model. The correlation matrix can be seen in Appendix 12.2 and it shows correlations below 0.5 between all variables. The predictor shows minor evidence against linearity, as the slope is curved towards the middle. We apply a cubic transformation to approach a more linear plot. The plots pre and post transformation can be seen below.

Figure 6.18: Transformation Plots for *NaïveDD*



Source: Personal collection

Like in the accounting-focused practical PD-model, we find the parsimonious regression model through steps of backward regression. Thus, we take the same starting point as in the previous section where we included all seven features. However, now we also include the *NaïveDD* resulting in eight initial predictors. It is visible from table 6.8 that three steps are needed to reach the final model. The first estimated model has three insignificant predictors and five significant at the 1% level. Among the insignificant predictors we again observe *RE/TA* and *FFO/CAPEX* as we did in the prior model. We also note that the direction of the *RE/TA* variable has changed such that an increase in the variable is associated with an increase in the probability of default, making it inconsistent with economic sense. However, most notably the *TL/TA* ratio is now also insignificant. As the T-statistic for the *TL/TA* ratio is the smallest the first step is to remove this variable. Subsequently, the *RE/TA* coefficient is still insignificant and is consequently removed. In the last step we remove the *FFO/CAPEX* ratio. The final model is reached at the fourth iteration where all variables are significant at the *5%* level and all beside *NIDummy* at the *1%* level. The fact that the both the market-based *NaïveDD* variable and the four accounting-based variables are significant implies that a model including both academic schools is aggregating more information than the frameworks individually. This underlines the relevance of the Default Model of Synthesis.

Table 6.8: Default Model of Synthesis

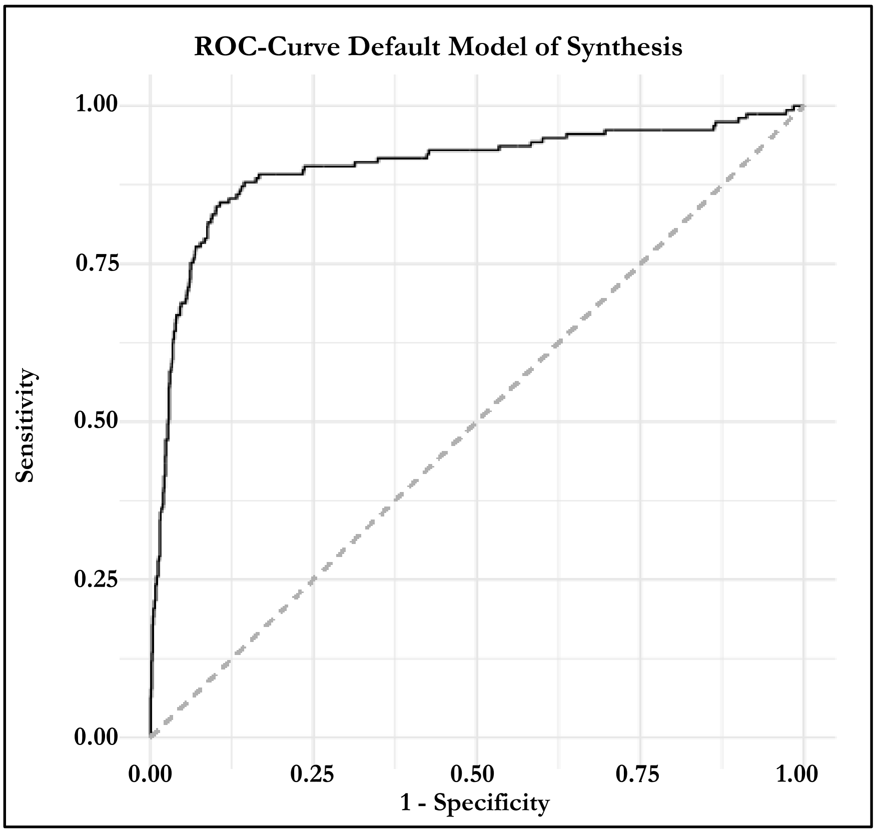
	Default			
	(1)	(2)	(3)	(4)
TL/TA	-0.289 (0.535)			
RE/TA	0.117 (0.085)	0.129 (0.082)		
FFO/CAPEX	-0.003 (0.004)	-0.003 (0.004)	-0.003 (0.004)	
MV/BD	-0.557*** (0.066)	-0.538*** (0.055)	-0.540*** (0.056)	-0.540*** (0.056)
Size	-0.431*** (0.047)	-0.430*** (0.047)	-0.401*** (0.054)	-0.402*** (0.043)
FFO/TL	-0.691*** (0.142)	-0.704*** (0.140)	-0.655*** (0.136)	-0.696*** (0.123)
NIDummy	0.484*** (0.162)	0.478*** (0.161)	0.383** (0.150)	0.385** (0.150)
NaïveDD	-0.832*** (0.075)	-0.839*** (0.074)	-0.824*** (0.074)	-0.823*** (0.073)
Constant	1.334** (0.540)	1.110*** (0.347)	0.887*** (0.316)	0.896*** (0.316)

Note: 1) Model Coefficients. 2) Standard errors in parentheses. 3) *p<0.1; **p<0.05; ***p<0.01

Source: Personal collection

The practical model employing both accounting and market information achieves an AUROC of *90.65%* on the testing set with a confidence interval of *87.47 - 93.83%* and a Somers' D of 0.813. The discriminative power of the model is an interesting finding in multiple ways. Firstly, it represents the influence the synthesis of the two paradigms has on the practical model's discriminative power. Secondly, the AUROC is greater than any of the prior estimated single-approach models.

Figure 6.19: ROC-Curve in Default Model of Synthesis



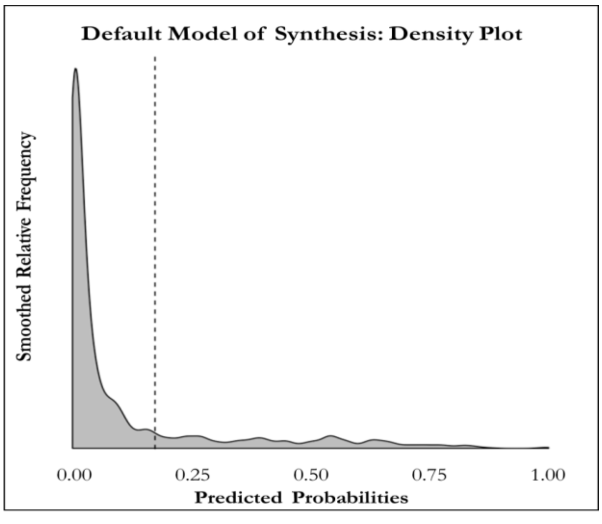
Source: Personal collection

The ROC curve is visualized above and the classification mechanism of the model at the top-left cutoff point of *13.61%* can be summarized as follows. The model correctly classifies a total of 1336 firms, here 133 of those are correct defaults and 1203 are correct non-defaults. Likewise, the number of Type I-errors is 136 and for Type II-errors the number is 24.

Figure 6.20: Confusion Matrix and Density Plot in Default Model of Synthesis

The Default Model of Synthesis – Top-left Cut-off Point

Predicted by Model	**Actual Observations** Solvent	Default
Solvent	1203	24
Default	136	133



Source: Personal collection

By using the coefficients from the final model and the logistic equation we can calculate the predicted probability of default for TCG in equation (6.17) and (6.18):

$$\begin{aligned}&\mathit{Predicted\ Log\ Odds\ TCG\ Default}\\&= 0.896 - 0.540 * -4.759 - 0.402 * 11.34 - 0.696 * 0.388\\&\quad +0.385 * 0 - 0.823 * -1.307\\&= -0.29\end{aligned} \tag{6.17}$$

$$\mathit{Predicted\ Probability\ of\ TCG\ Default} = \frac{1}{1 + e^{-(-0.29)}} = 42.87\% \tag{6.18}$$

The model assigns a probability of default to TCG of *42.87%* and as this is above the cut-off value the classification is correct.

6.2.3 Overview and Evaluation of Hypothesis 2

To answer the thesis' second hypothesis, we employed the practical probability of default methodology to our modern portfolio. The focus was centered on how a modelling procedure, which is unconstrained in terms of the mix of input variables used but instead constrained by specific variable requirements, leads to an improved model. Two models were built using this framework. The first model consisted only of accounting-based input variables and outperformed all prior models with an AUROC of 89.5%. The second model, named the "Default Model of Synthesis", bridged the gap between the competing accounting-based and structural schools. By combining the *NaïveDD* with several accounting-based variables the final practical model achieved an even greater AUROC of *90.65%*. The results of the two models confirms the thesis' second hypothesis.

Table 6.9: Verification of Hypothesis 2

#	Hypothesis	
2	*The practical model outperforms the academic but is improved through a synthesis of market and accounting theory*	✔

The evaluation metrics of the two practical approaches is shown in comparison to the six academic models in table 6.10. Here, it is visualized clearly how the practical model and the Model of Synthesis outperforms the frameworks from the prior section.

Table 6.10: Overview of Models

	Training AUROC	Testing AUROC	Somers' D	Solvent PD-Mean	Defaulted PD-Mean	Portfolio PD-Mean
Original Altman	80.65% [78.43-82.87]	80.74% [76.87-84.61]	0.6149	13.84%	49.12%	17.69%
Re-estimated Altman	80.01% [77.78-82.24]	81.35% [77.55-85.15]	0.6270	75.01%	76.64%	75.17%
Original Ohlson	87.08% [85.39-88.77]	86.73% [83.27-90.19]	0.7345	20.51%	70.57%	25.97%
Re-estimated Ohlson	87.26% [85.62-88.90]	87.54% [85.20-89.88]	0.7508	8.20%	33.57%	10.97%
Original Merton	80.72% [78.48-82.96]	81.71% [77.52-85.90]	0.5981	2.56%	26.08%	5.13%
Naïve Merton	88.74% [87.02-90.46]	87.63% [84.15-91.11]	0.7295	11.08%	69.89%	17.50%
Practical Model	88.30% [86.54-90.06]	89.50% [86.83-92.17]	0.7900	7.05%	35.42%	10.48%
Model of Synthesis	91.71% [90.15-93.27]	90.65% [87.47-93.83]	0.8130	6.45%	44.91%	10.93%
Portfolio Total	-	-	-	-	-	10.95%

Source: Personal collection

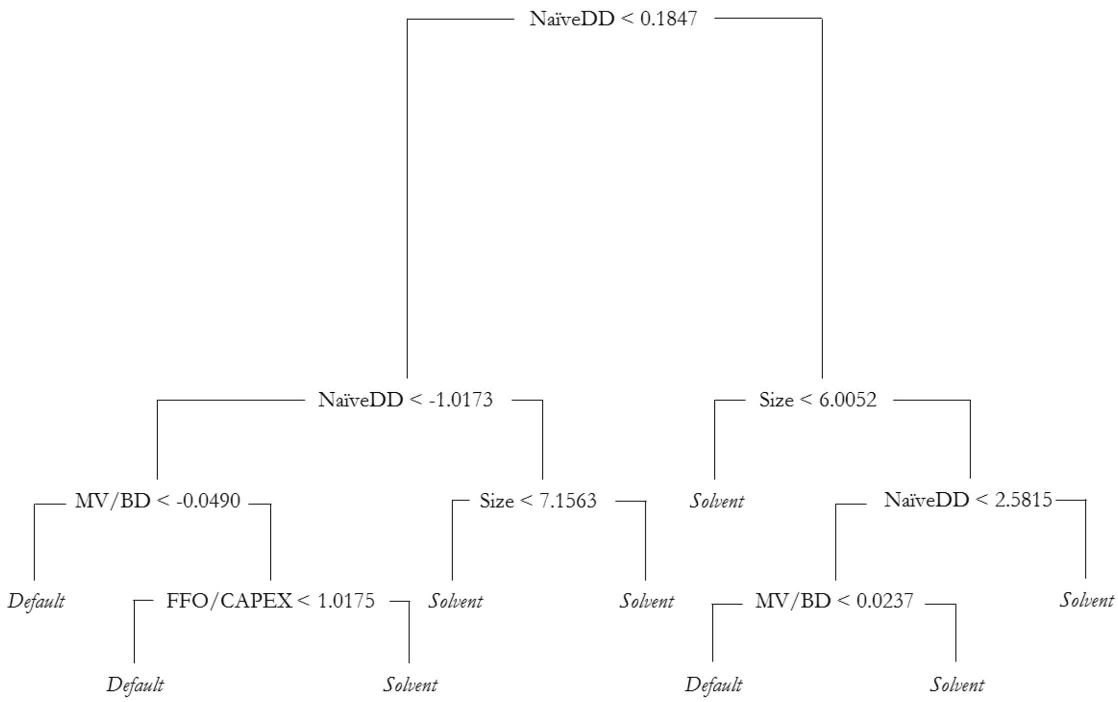
6.3 The Machine Learning Approach to Default Probability

Having presented the second part of our research which showcased that the practical approach to default outperformed the academically founded frameworks, we continue into the third section which is concerned with the introduction of machine learning methodology to credit risk modelling. The research manifested that the practical approach was indeed better than the academic frameworks, however, it was significantly improved upon by bridging the two paradigms within the academic sphere of default probability. We labelled this the "Default Model of Synthesis", as it incorporated both the accounting and the market approach to the discipline. This is the foundation we build upon in the third section of the thesis, which is exclusively methodologically driven, and serve the purpose of evaluating the third hypothesis: *The discriminative power of the practical model can be improved through the application of machine learning*. This means that the third section does not aim to build the best possible model from all the portfolio's predictors through a new framework. Instead, it underlines how the practical approach to PD can be improved even further through the application of machine learning methodology, i.e. only utilizing the exact same features as in the prior section. This part will be split in two, where the first introduces the algorithm of classification trees, and the second presents its ensemble extension of a random forest, which will constitute the final model of the research.

6.3.1 A Probability of Default Model using Classification Trees

As outlined above, we abstain from finding the most optimal predictors for a machine learning approach to default probability. Instead we build upon the finding of the second section of the thesis relating to the practical approach to default probability, which is the starting point of this section. Thus, we already know what predictors to build the initial classification tree with. Here we follow the approach outlined in the methodology section, where we start with an unconstrained classification tree and then test if we need to apply a pruning procedure before arriving at a final model. As such, we build a classification tree on the training data set, including the variables of *RE/TA, MV/BD, Size, TL/TA, FFO/TL, FFO/CAPEX, NIDummy,* and *NaïveDD*.

Figure 6.21: Classification Tree Diagram



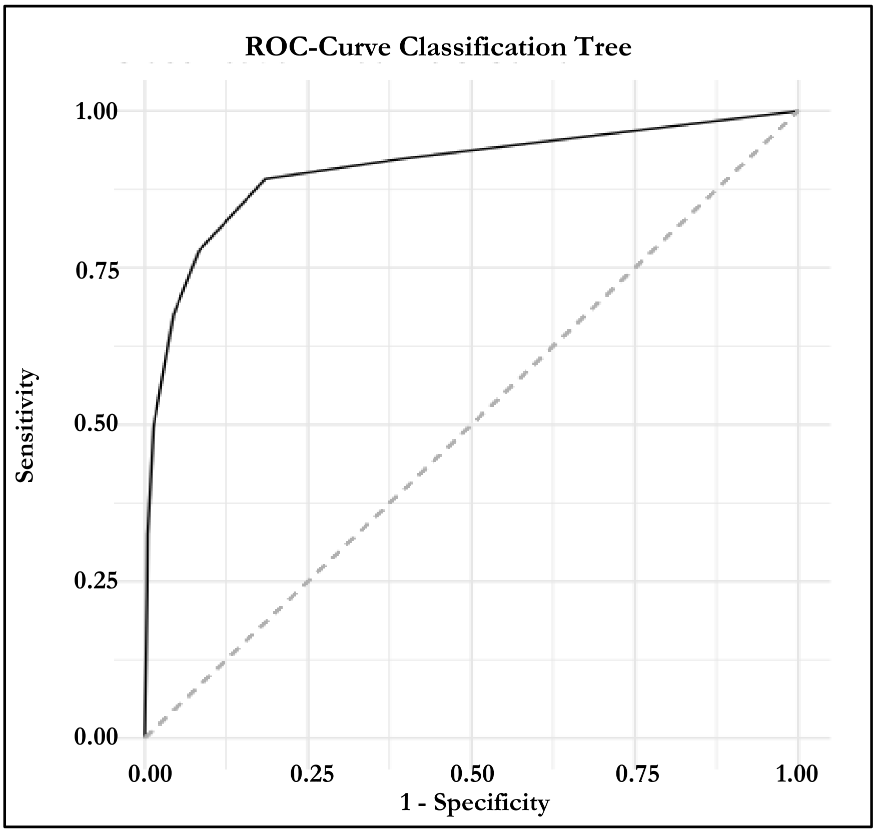
Source: Personal collection

The initial tree is visualized above and shows nine terminal nodes, where three is indicative of a default prediction and six give a solvent prediction. The single most decisive predictor according to the decision tree is that of the *distance to default* from the Naïve Merton framework. The *NaïveDD* is the root node containing all observations with an initial split happening at 0.1847. Firms with a value greater than this are then ordered by *Size*, where those with a *Size* less than 6 are considered solvent. Observations with a *Size* greater than 6 are once again ordered by *NaïveDD* at a cut-off of 2.5815. Above this value, firms are considered solvent, whereas below they are ordered by the *MV/BD-ratio* at 0.0237. Below this value, companies are predicted as default, whereas above they are deemed solvent. Going out the other branch of the tree, the *NaïveDD* again splits the tree at a value of -1.0173. Above this value, *Size* is a determining predictor, but only for node purity as both firms with a value above and below 7.1536 are considered solvent. Firms with a *NaïveDD* below -1.1398, are split by *MV/BD*, where companies with a ratio less than 0.049 are considered default. Those above are split by the ratio of *FFO/CAPEX* where firms with a ratio lower than 1.0175 are default and vice versa.

By inspecting the tree, we can observe that some of the terminal nodes has an interesting characteristic that may seem counterintuitive. At the mid splits made by *Size* the associated terminal nodes have the same predicted class. Regardless of the value of *Size* the response classification is the same. However, by including the extra split, the resulting nodes' purity increases corresponding to a better class segregation. Suppose an observation in the testing set has a *NaïveDD* below 0.1847 but above –1.1398. At this point, the node contains 415 observations with a split of *67.5%* non-default and *32.5%* default observations. We are more likely to make a non-default prediction for observations at this part of the tree, but the risk of a misclassification error is notably present. By including *Size* as a decision node, we can improve the class split considerably. As such, if the observation has a *Size* greater than 7.1535 then the resulting class split is *85.34%* and *14.66%*. By examining the terminal nodes further, we can highlight some rather pure nodes. Among observations with a *NaïveDD* below -1.1398 and a *MV/BD* below 0.05, *91.8%* are defaulting firms. The opposite is true for observations with a *NaïveDD* above 2.58 and a *Size* above 6 as 99.2% of these are non-defaulting firms.

Simple classification trees tend to suffer from high variance. As they function by splitting data to reach as pure nodes as possible, the data is split multiple times as we go down the tree. On the training set this works well as we are continuously narrowing down the data. As we go down the tree, and the terminal nodes are further away from the root node, the actual predictions will effectively be made by fewer and fewer data points. If the tree is too deep and specific to the data it is built with, we will make incorrect predictions for observations with slight deviations in financial ratios. A lower tree might generalize better, but at the cost of less pure nodes. As we by design have already pre-selected a subpart of the total long-list of financial ratios, the risk of overfitting has already been decreased. However, we do check if our tree needs to be pruned. We perform the pruning procedure described in the methodology with a 10-fold cross-validation on the training set, which shows that the optimal number of terminal nodes is nine. Therefore, the initial classification tree equals the optimal one.

Figure 6.22: ROC-Curve in Classification Tree Model



Source: Personal collection

Applying the model to the portfolio such that we predict on the testing data, we achieve the ROC-curve shown above. From here, we can extract the top-left cut-off point, which equals 0.09. At this point, we observe the confusion matrix seen below. At the top left cut-off point, the classification tree correctly predicts 140 defaulting companies, with 247 false positives, i.e. type I-errors. Simultaneously, it predicts correctly 1092 solvent companies, with 17 false negatives, i.e. type II-errors. If we apply the example of Thomas Cook Group to the classification tree, it showcases how it classifies such company as insolvent. TCG has a *NaïveDD* of -2.2304, which is below the models' initial cut-off point of 0.1849. It is also below the second threshold of the same parameter, which the classification tree sets at -1.0173. The following node considers the ratio of *MV/BD*, for which a threshold is set of 0.049. TCG has a *MV/BD*-ratio of 0.0086, wherefore it ends at its terminal node where it is correctly predicted as default.

Figure 6.23: Classification Tree Model – Top-left Cut-off Point

	Actual Observations	
Predicted by Model	Solvent	Default
Solvent	1092	17
Default	247	140

Source: Personal collection

The discriminative power of the classification tree is summarized with an AUROC of *90.26%,* with the confidence interval of *87.3 - 93.23%.* However, the Somers' D achieves a score of only 0.7286. With the given AUROC, the simple classification tree does not outperform the Default Model of Synthesis. Additionally, the nine terminal nodes make the classification tree inherently discriminate the portfolio into only nine different probabilities. The many ties related hereto make the Somers' D relatively low, compared to the high AUROC, which underline why the former is a poor metric to evaluate a classification tree. However, there are further challenges related to the classification tree.

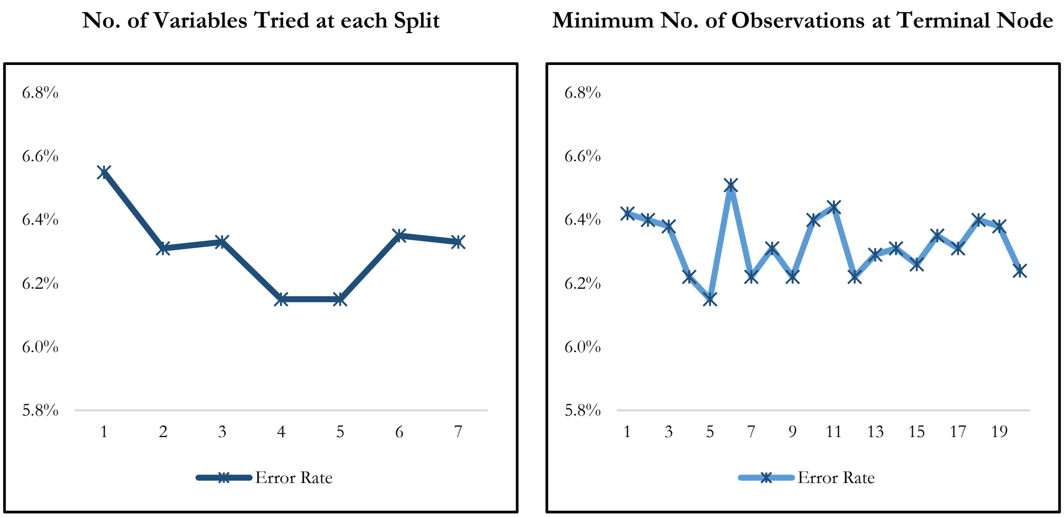
6.3.2 Random Forest: Extending the Classification Tree Framework

To improve the discriminative power, we use the simple tree as a building block for a more powerful model. As mentioned earlier, the simple classification trees suffer from high variance, i.e. the rule-based split of the training data into as pure nodes as possible generates a tree that is too detailed and specific to be effectively applied for classification of new observations. If the range of the financial ratios of the in-sample data are different from the range of ratios of the out-of-sample data, the discriminative power of the model will not be optimal. To mitigate the problem a pruning procedure can be applied, but the top structure of the tree remains the same independent of how far up we prune it. In other words, the financial ratios towards the top are the same. This implies that if we compare different pruned trees, they will tend to be correlated. These challenges are mitigated by methods of ensemble. As highlighted in the methodology chapter of the thesis, ensemble methods consist of estimating multiple analytical models instead of just one. By these means, we estimate a random forest model, which builds many hundreds of classification trees, prioritizing different combinations of features and compositions.

The random forest model contrasts the simple classification trees in terms of variance as each tree is grown on the basis of a different subpart of the dataset. To ensure that the individual bootstrapped sample contains the same number of defaulted observations we choose a stratified bootstrapping procedure. Additionally, for each tree built we decide a random sample of *m* size among the financial predictors as the basis for the splitting criterion. By considering only a subset of the total number of financial predictor variables at each split, multiple predictors will get a chance to be tested at the upper stages of the tree. We also vary the minimum number of observations at each terminal node. This implies that we mitigate the weakness of the simple tree. The average of the aggregated collection of trees will yield a lower variance, and each tree within the forest will be less correlated which will help us achieve a greater discriminative power.

Specifically, we build a random forest model, still solely with the variables of *RE/TA, MV/BD, Size, TL/TA, FFO/TL, FFO/CAPEX, NIDummy,* and *NaïveDD*. Then, we proceed to tune the model, where we estimate the random number of variables utilized for each split, and the minimum size of the terminal nodes in terms of observations. Doing so, we set the number of trees in each random forest to be 1000, as accounted for in the methodology chapter. We find the random number of variables at each split that minimizes the out-of-bag error at the value of 4 and 5, at the error rate of 6.15%. We choose to proceed with 4, however utilizing 5 should yield similar results. Then, we find that the minimal number of observations for each terminal node that minimizes the error rate at 5. This minimization process oscillates significantly through the tuning process, due to the granularity of changing the parameter with 1 per iteration. These optimizations are visualized on the graphs below.

Figure 6.24: Identification of Minimum Error Rate



Source: Personal collection

As such, we have our final random forest model. When we apply the random forest model to our test set, the predicted class is the average across the 1000 trees in the ensemble.

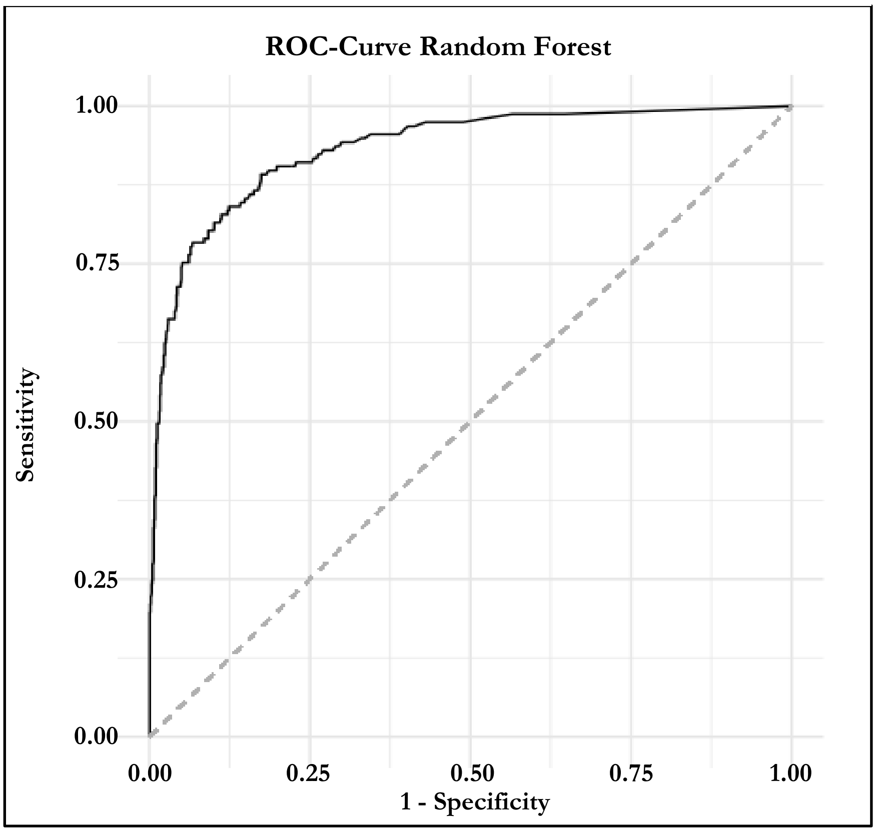
Figure 6.25: Random Forest Model – Top-left Cut-off Point

	Actual Observations	
Predicted by Model	Solvent	Default
Solvent	1106	17
Default	233	140

Source: Personal collection

The model correctly predicts 140 of the defaulted companies in the test portfolio. It is followed by 233 type I-errors, and 17 type II-errors. As such, the random forest model correctly predicts 1106 of the solvent companies. The confusion matrix and the ROC-curve can be seen in figure 6.25 and 6.26.

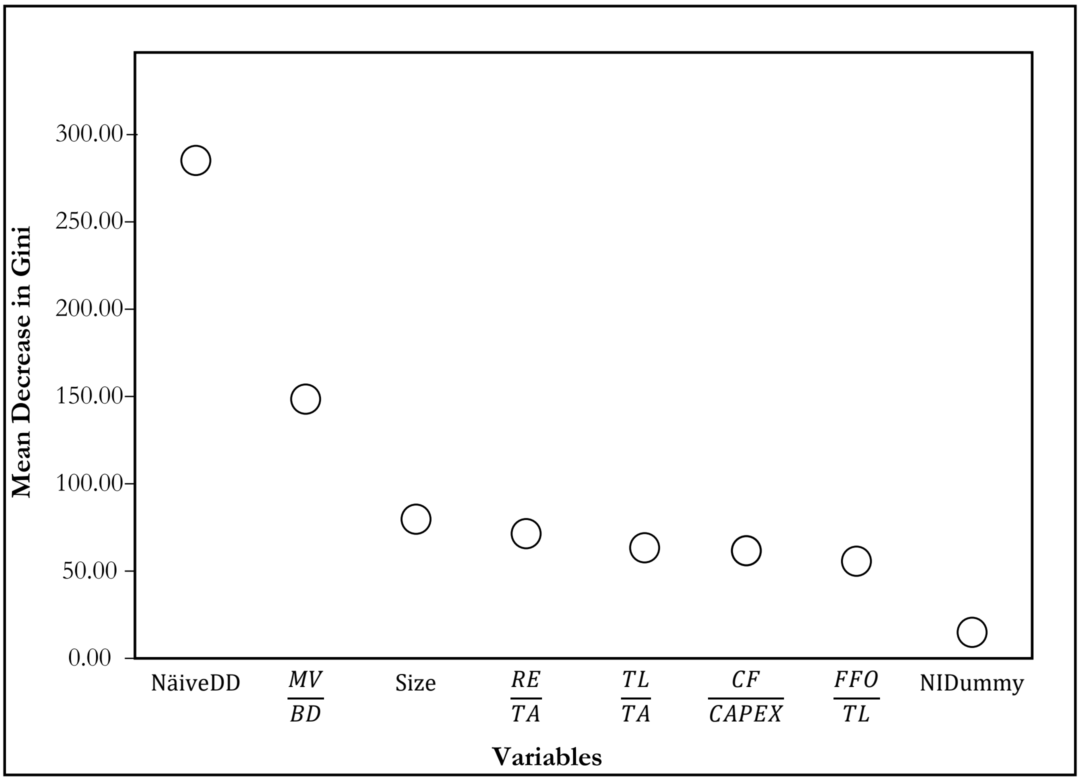
Figure 6.26: ROC-Curve in Random Forest Model



Source: Personal collection

The neatly visualized and easily interpretable tree diagram is no longer possible with a random forest. As the prediction is based on the average of 1000 trees, we are unable to make a single visualization. Furthermore, 1000 different visualization will not add any value to the analysis. As a consequence, the financial ratios corresponding to either terminal nodes or top decision nodes is not directly known. To examine the relative importance of the financial variables in the forest we need to make use of the Gini index. For an individual predictor, we add up and average the decrease in Gini index for all splits across the 1000 trees. A visual representation can be seen below, where the Naïve distance to default again is the most important variable.

Figure 6.27: Random Forest - Variable Importance



Source: Personal collection

In terms of the Thomas Cook Group, the properties of the random forest model make it unable to exemplify by an equation or diagram how the model arrives at a classification of either defaulting or solvent. This make the model somewhat a black box. However, we can create a dataset consisting of only the TCG observation, and have the model predict its solvency. By using the average across the 1000 trees in the ensemble, the model assigns the firm a predicted score of 0.4937, which is above the model's score threshold of 0.0901, wherefore the random forest model correctly classifies TCG as default. These two aspects underline the problems regarding interpretability with machine learning models, which we will dissect in relation to the discipline of default probability in the subsequent discussion chapter of the thesis.

However, when using the final random forest model to make predictions on the testing portfolio, we achieve an area under the receiver operating characteristic curve of *94.14%,* which is a significant improvement from the classification tree. This AUROC comes with a Somers' D of 0.8786, which constitute a massive improvement compared to the decision tree.

6.3.3 Overview and Evaluation of Hypothesis 3

While the machine learning model of a simple classification tree did not improve the area under the receiver operating characteristic curve in comparison to the Default Model of Synthesis, we utilized the framework as a building block in an ensemble model. Thus, through a random forest model, which optimizes and averages a sample of 1000 classification trees, we were able to achieve a discriminative power of *94.14%.* This is by far the largest AUROC found by the research at hand, and it clearly showcases the power of a machine learning methodology in the arena of default probability. Furthermore, this final model improved the discriminative power on the portfolio in comparison to the practical approach to the credit risk discipline. Through this extension of the practical framework, we address the gap identified in both the academic and practical fields of default probability. As such, the thesis confirms its third hypothesis.

Table 6.11: Verification of Hypothesis 3

#	Hypothesis	
3	*The discriminative power of the practical model can be improved through the application of machine learning*	✔

The evaluation metrics of the two machine learning approaches is shown in comparison to the academic and practical models in table 6.12. It is shown how the classification tree fails to improve on the discriminative power of the Default Model of Synthesis, however the random forest model clearly outperforms the other frameworks in the research.

Table 6.12: Overview of Models

	Training AUROC	Testing AUROC	Somers' D	Solvent PD-Mean	Defaulted PD-Mean	Portfolio PD-Mean
Original Altman	80.65% [78.43-82.87]	80.74% [76.87-84.61]	0.6149	13.84%	49.12%	17.69%
Re-estimated Altman	80.01% [77.78-82.24]	81.35% [77.55-85.15]	0.6270	75.01%	76.64%	75.17%
Original Ohlson	87.08% [85.39-88.77]	86.73% [83.27-90.19]	0.7345	20.51%	70.57%	25.97%
Re-estimated Ohlson	87.26% [85.62-88.90]	87.54% [85.20-89.88]	0.7508	8.20%	33.57%	10.97%
Original Merton	80.72% [78.48-82.96]	81.71% [77.52-85.90]	0.5981	2.56%	26.08%	5.13%
Naïve Merton	88.74% [87.02-90.46]	87.63% [84.15-91.11]	0.7295	11.08%	69.89%	17.50%
Practical Model	88.30% [86.54-90.06]	89.50% [86.83-92.17]	0.7900	7.05%	35.42%	10.48%
Model of Synthesis	91.71% [90.15-93.27]	90.65% [87.47-93.83]	0.8130	6.45%	44.91%	10.93%
Classification Tree	92.19% [90.97-93.58]	90.26% [87.30-93.23]	0.7286	n.a.*	n.a.*	n.a.*
Random Forest	93.22% [91.87-94.58]	94.14% [92.07-96.21]	0.8786	n.a.*	n.a.*	n.a.*
Portfolio Total	-	-	-	-	-	10.95%

*Probability of Default means are not relevant for Classification Trees as classification is based on class proportion in terminal nodes.
Source: Personal collection

7. Discussion

The discussion of this thesis will revolve around the practical approach to default probability, specifically the precedence of logistic regression and the deliberate refutation of machine learning methodologies. Thus, we discuss the dichotomy present in the field of default probability, which considers it being a field inherently concerned with building the most optimal discriminative model but simultaneously abstain from pursuing a practice that allow efficiency to reach new levels. By these means, the discussion chapter of the thesis is split in three parts. The first reflect upon the current precedence of the logistic regression model in the practical PD approach. The following two sections provide each an explanation for the current setting within the discipline, which are non-exclusive. The first explanation considers interpretability of machine learning models. The second draw upon a framework from political theory, which study the continuous adherence to the logistic regression through an institutional lens.

7.1 A Reflection on the Neglection of Machine Learning in Practice

As it was noted in the theoretical framework of the thesis, the research at hand presumes the methodology outlined by Hayden in Engelmann and Rauhmeier (2011) as the precedence within the practical approach to probability of default. However, confirmation of logistic regression as the overarching technique within the field is widespread (Medema et al., 2009; Bellini 2019; De Laurentis et al., 2010; Neisen & Rosch, 2018). It is described here that the somewhat uncontested precedence of the logistic PD model stems from its high interpretability, and foundation for evaluating the "economic sense" in the direction of explanatory variables. It is furthermore indicated that the current use of logistic models on accounting information consists of a high degree of reliance on historical tendencies. However, in a discipline occupied with achieving the clearest and most accurate discrimination between default and solvent companies, the devotion to logistic default probability models seem puzzling.

It is interesting that practitioners generally withstrain from pursuing other methodologies for modelling credit risk, it being both structural models and machine learning algorithms. Machine learning is generally acknowledged for being a superior predicting methodology, mainly due to its reliance on modelling predictive accuracy rather than statistically significant relationships (Breiman, 2001). While the practitioners of PD to a large extent utilize the methodology of using training and testing portfolios, the logistic regression technique comes inevitably short to its algorithm counterparts. Here, the explanation is that machine learning models are invented for uncovering deeply complex relationships in data and is inherently superior in this regard. This proposition is also showcased in the thesis at hand, where our research underlines that the practical approach to default probability comes short in comparison with a random forest model. Nonetheless, this approach enjoys none to very little weight in practical default probability modelling. This research intends to discuss some explanations for this environment within the field.

7.2 A Question of Interpretability

While machine learning excels at solving complex prediction problems it comes at the expense of interpretability. The sometimes inadequate interpretability is being expressed both theoretically and empirically. In a theoretical context, Doshi-Velez and Kim define interpretability as *"the ability to explain or to present in understandable terms to a human"* (2017). As such, the term can be translated into how simple it is to comprehend a model's motivation for a given output. Thus, in relation to this research, it denotes understanding why a company is predicted as default or solvent. Where it is preferable to increase discriminative power, it becomes less relevant if practitioners are unable to describe why a model predicts default rather than non-default for a given firm. For illustration, an economist is rarely capable of convincing a politician of a recession if he is unable to explain how the algorithm anticipates inverted yield curves.

Related to interpretability is the concept of computational burden, which considers the computational complexity in machine learning. It is defined by Goldreich as the "*intrinsic complexity of computational tasks*" (2008). It can be transcribed into how much time a model utilizes for finalizing its prediction. While this complexity is dependent on size of data sets and also iterations of calculation, it differs across algorithms as they have different costs. Like interpretability, computational burden is related to the subsequent discriminative power of a model, where a larger predictive accuracy comes hand in hand with more complexity. For illustration, a stock trader can be using a complex machine learning model to determine how he is rebalancing his portfolio, however, if the algorithm compute the result in multiple hours, the investment opportunities may be out the window. In relation to the research of this thesis, we present a table below consisting of the interpretability and computational burden of the practical and machine learning models developed.

Figure 7.1: Interpretability and Computational Burden

	Logistic Regression	Classification Tree	Random Forest
Interpretability	*High*	*Moderately High*	*Low*
Computational Burden	*Low*	*Moderate*	*High*

Source: Personal collection

Extending this discussion into an empirical context, the field of default probability is a discipline which is little constrained by computational burden as time and complexity are not at the center of concerns. On the other hand, interpretability is very central to the process. Here it is relevant that machine learning methodologies, such as the random forest, often can be considered somewhat a black box (Hastie et al., 2014). For practitioners of default probability, the inability to interpret credit risk models are problematic. It is pertinent to be capable of explaining why a firm is granted or denied a loan. As such, the interpretability of a model is forwarded to companies, authorities or other departments of lending institutions. This means that the credit risk department of a given bank is being held accountable for their models' prediction by both the supervisory authorities and the department facilitating client contact, wherefore it is crucial to have the ability of interpreting the given PD-model.

Our research further showcased these considerations, with the concepts of interpretability and computational burden being echoed by the thesis' illustrative example of the Thomas Cook Group. To improve the practical framework, we employed a random forest machine learning algorithm and the model greatly outperformed all prior approaches. Here, the example of the Thomas Cook Group underlined the drawback related to interpretability. The black box design of the random forest algorithm put us in a situation where we were unable to explicitly exemplify how the model reached its prediction. A byproduct of our research is that the pursuit for a greater discriminative power comes at the cost of an increased computational burden and a diminishing degree of interpretability.

7.3 Institutional Stickiness: A Possible Explanation for the Precedence of Logistic Regression

The precedence of logistic regression in the practical field of default probability can be explained through the framework of path dependency. Path dependency is a political theory framework developed by institutionalists, who claim that "history matters" (North, 1990). History matters such that an already established institutions continue to reign although it may be suboptimal on a broader scale, which is driven by the fact that changing an institution is related to certain costs. As such, the continuance of the historical institution is path dependent, as a result of the neglection of change driven by the cost implications of doing so. Within the framework of path dependency is the theorem of institutional stickiness. This term is developed within development economics, where it explains the reluctance to change an institution as a function of the institution's reality in the past time period (Boettke et al., 2008). As a continuation, an institution is going to continue its historical path, unless it is stimulated exogenously. While this framework originates in development economics, we can transcend it into the field of default probability modelling, where it offers relevant explanatory power.

Within the practical approach to credit risk, logistic regression has historically emerged as the single most dominant method for modelling due to its usefulness and relatively simple application (Engelmann & Rauhmeier, 2011; Neisen & Rosch, 2018). As such, the methodology is in fact an institution within the practical discipline. However, it has continued to be the main approach through years of emerging machine learning techniques, which are capable of outperforming the logistic model both theoretically and empirically, as showcased in the thesis at hand. This underlines the path dependent nature of the logistic regression within the field, where it has become unfeasible to change the modelling approach, due to its widespread recognition and application. In reality, the approach is both taught to new employees of the field, and maintained through their tenure in such lending institutions, making the logistic approach self-fulfilling, or in fact, path dependent. Furthermore, it has from a regulatory perspective been a requirement of further administrative work with the relevant supervisory authority, if a given lending institution would applicate another PD model than the logistic regression (ibid). These mechanisms make the logistic regression methodology embedded in the practical field of default probability. In other words, the status quo is inherently difficult to change, and logistic regression continues to exist as the overarching model. Thus, it has become a "sticky" institution. In order to change this, the institutional stickiness theorem suggests that an *"outside entity is required to provide a exogenous shock"* necessary for institutional change (Boettke et al., 2008). With the topic at hand, it would be the Basel Committee providing directions for a new approach to default probability modelling, which would have to be materialized into regulations by legislators. However, as the current situation witness, the logistic regression model constitutes an institution within the practical field of PD, which is dependent to its current path.

8. Reflections on Validity and Reliability

Our research concerns default probability modelling utilizing both an academic and a practical lens. We engage in an evaluation of the measurement validity, reliability and replicability, as well as both the internal and external validity. These reflections are important as they underpin the integrity of the findings of the thesis at hand.

8.1 Measurement Validity, Reliability and Replicability

First in line of these concepts relevant to the implications of the thesis' results is *reliability*, which refers to the consistency of the findings. In other words, would we be to find the same results, if we were to do the research again in another setting (Bryman, 2012). Here, the key focus is on the stability of our results, which is clarified through the area under the receiver operating characteristic curve that is stable from the training to the testing data set. Thus, this speaks to the stability of our findings throughout the research. In terms of reliability, if the research were to be performed again, in a different setting for instance, we would not necessarily find all the same directions and significance levels of all variables, but the hierarchy of discriminative power between the models would be similar.

As an extension of reliability is measurement validity, which encapsulates whether the measurements of the findings are true to what they are supposed to reflect. Here, we emphasize that the findings of the thesis to a large extent reflect both the theoretical and methodological frameworks that they build upon (Bryman, 2012; Thyer, 2001). Specifically, the classic academic frameworks do hold discriminative power in a modern setting, the practical models outperform the former, we ensure the direction of variables are grounded in theory, and finally the machine learning methodology do achieve a higher level of predictive accuracy. In this regard, we approached the research by deducting these properties from the underlying theory. Thus, the results consist of valid measures, which reflect the concepts they were hypothesized to capture.

A somewhat different reflection is that of replicability. It is important in academic research, as it further underpins the integrity of our findings and thus place our thesis within the pertinent literature. It naturally entails the ability for other researchers to replicate the study we have undertaken (Bryman, 2012; Heale 2017). Here, it is rather about making the mechanisms behind the research available to others than literally testing colleagues' work.

With this in mind, we highlight that we have clearly and explicitly outlined our methodology and data preparatory work, wherefore we believe the replicability of our thesis to be high. Additionally, we provide the code scripts from the R-software in a collected code book as appendix to the thesis, which further underlines the transparency provided.

8.2 Internal Validity

The concept of internal validity concerns the issue of causality. Embedded in this is the question of whether X causes Y, and whether there is a chance of the opposite. For the thesis at hand, we identify internal validity in two levels of the research. The first concerns the causality found in specific models, which can be simplified to whether specific ratios truly explain default (Thyer, 2001; Bryman, 2012). Here it should be emphasized that the ratios reported in the dataset inherently mirror the characteristics that drive companies towards default. As such, a low value of a significant ratio in the practical model, *funds from operations/total liabilities*, infers that a company is pushed towards default. Thus, it is not a default in the company that drives a low funds from operations. This is consistent in the models build in the thesis, where financial information, X, determines the defaults, Y, and not the reverse. At the second level, we have the overarching research of classic academic approaches, practical default probability and two extensions of the practical framework. Here the improvement of credit risk modelling that is showcased throughout the thesis is aligned with the theoretical and methodological underpinnings of the framework. As such, these improvements are caused by the application of complementary theoretical approaches and better methodological techniques, wherefore this relationship also exhibit convincing causality. As such, with both of these aspects highlighted above, the internal validity of our research is emphasized.

8.3 External Validity

Finally, we approach the concept of external validity, which entails whether the research's findings provide knowledge outside of the framework of the study itself. In other words, whether our results are generalizable to other settings (Heale, 2017; Bryman, 2012). Here, a motivation underpinning the overarching research was to contribute to the field of default probability and approach a model of credit risk which can be looked towards in other settings. We emphasize that the large data set consisting of more than 6000 companies, and the long time period of the research that span several economic fluctuations advocate for the external validity of the thesis. However, the generalizability is simultaneously subject to the limitations set by the data set, which include the overweight of U.S. companies, and the inherent focus on publicly traded companies, par exemple. In spite of this, the overall comprehensiveness of both our portfolio, time horizon and the applicated models speak to the external validity of our research for OECD countries. Yet, we wish to highlight the field of default probability as a dynamic discipline that intrinsically dictates constantly updated applications, wherefore the generalizability is characterized with caution. However, the thesis at hand have provided research which can be help guide these future applications both academically and practically.

9. Conclusion

The aim of this thesis was to answer the overarching research question:

"To what extent does the classic academic approaches to default probability have discriminative power, and through which measures can the practical approach to credit risk be improved?"

The research question was answered through three hypotheses, which were deducted on the back of a theoretical framework containing both the theory and the literature of default probability. Here, the thesis identified two major gaps in the literature, which created a catalyst for formulating the hypotheses.

The first hypothesis was academically motivated, and thus concerned with testing the discriminative power of the default probability models of Altman, Ohlson, Merton and their respective re-estimations. Therefore, it addressed the first part of the research question of *to what extent* the classic academic approaches to credit risk have discriminative power. Using the original methodologies, the research finds that these models all achieve a high discriminative power on a portfolio collected at least forty years after they were developed.

The second hypothesis was practically motivated and revolved around establishing the practical model of default probability, and further combine accounting theory with market theory. As such, it addressed the second part of the research question of *through which measures* the practical approach to credit risk can be improved. By employing the technique outlined by Hayden, the research found that the practical approach outperforms the academic frameworks. However, by combining its intrinsic accounting focus with the market theory approach to default probability, the thesis creates a superior framework for practical credit risk, which it coins the "Default Model of Synthesis".

The third hypothesis was methodologically motivated and investigated the possible use of machine learning in default probability modelling. Thus, it was also directed at the second part of the research question, introducing how the practical default model can be improved *through another measure*. By applying the machine learning algorithm of a random forest, the research showcased how the practical approach to default probability could be improved exclusively by transcending it into a more advanced methodological framework.

Through these findings, each of the three hypotheses was verified and can be summarized the following.

Table 9.1: Overview of Verified Hypotheses

#	Hypothesis	
1	*The classic academic approaches to default probability have discriminative power on a modern portfolio*	✔
2	*The practical model outperforms the academic but is improved through a synthesis of market and accounting theory*	✔
3	*The discriminative power of the practical model can be improved through the application of machine learning*	✔

The third hypothesis holds further implications for the practical discipline of default probability, as it showcases how a machine learning methodology could improve its discriminative power. Therefore, the research discussed the neglection of machine learning in the practical field. The thesis gives two non-exclusive explanations for the current prevalence of logistic regression amongst practitioners. First, the low interpretability of complex algorithms such as the random forest disables practitioners in explaining why certain companies are considered default and others solvent, which is an essential property for a discipline that is held accountable both by clients and supervisory institutions. Second, the logistic regression methodology is deeply embedded in both the practical field and the legal institutions that monitor it, wherefore it has become a "sticky" institution that is path dependent. In sum, the discussion finds that the status quo is unlikely to change.

Ultimately, the research of the thesis finds that the classic academic approaches to default risk do have discriminative power on a modern portfolio, which underlines their relevance in a present setting. Further, the goal of the thesis was not to induct a new universal methodology, but to conceptualize a framework where the practical approach to probability of default can be improved. The thesis find that it can be improved in two ways: 1) Through the application of the Default Model of Synthesis and 2) through the methodological application of a random forest algorithm, which however is constrained by its interpretability.

9.1 Contribution

The thesis contributes to both the academic literature of default probability and the practical field of credit risk. We identify two main contributions of the research at hand, and subsequently several supporting additions to the literature. First of the main contributions is the Default Model of Synthesis. The model showcases how a combination of accounting and market theory can allow for higher discriminative power, utilizing the same methodology as the practical approach to credit risk. This fills the gap in the literature facilitated by the clash between the two strands in the academic discipline. Second of the main contributions is the extension of the practical framework with the introduction of a random forest algorithm. This part of the research addresses a void in the literature, as scholars of default probability have neglected to grant machine learning a central position within the discipline, in spite of its indisputable predictive power.

Additionally, the research at hand has contributed to the field of default probability with several supporting elements. It has reviewed the frameworks of the main academic tenants in a modern portfolio setting, which underlined their relevance half a century after its origin. In comparison to the literature, the thesis has broadened the portfolio scope significantly, both in terms of geographical attention and industry focus. Finally, the research has provided a discussion of the reluctance showed towards machine learning by the practical side of default probability, where it points toward interpretability and "institutional stickiness" as explanations of the continued prevalence of logistic regression. The thesis is as such multi-layered, as it is first theory-testing, and subsequently framework-building both through theory and methodology.

Table 9.2: Overview of Thesis' Contributions

	Contribution	Type	Field Relevance
Main	Default Model of Synthesis	Theory testing & Framework-building	Practical
	Random forest extension	Theory testing & Framework-building	Practical
Supporting	Review of classic academic models	Theory testing	Academic
	Expanded portfolio scope	Theory testing	Academic
	Implications of machine learning in PD	Discussion	Practical

Source: Personal collection

10. Avenues of Further Research

In order to study probability of default in a more comprehensive setting, the data for the research in this thesis consisted of publicly listed firms across OECD countries. A heavyweight of the existing academic literature is carried out in settings characterized by either a single country or a single sector, or both simultaneously (Jackson & Wood, 2013; Altman, 1968; Giacosa et al, 2016; Duan & Shresta, 2011, Tudela et al., 2003). We therefore suggest that future research is conducted in such country or sector specific settings to facilitate further comparability to the existing academic literature. As such, it would be interesting to see whether our findings are replicated or altered if the Default Model of Synthesis is tested exclusively on American energy companies, or Indian manufacturing companies, par exemple. Ultimately, the application of the Default Model of Synthesis in new portfolio settings would provide a more complete picture of its properties.

Were the investigation of this thesis to be taken even further, several avenues of further research exist. Underpinning the default probability models are the company characteristics, which the frameworks utilize to predict future insolvency. By these lines, it would be relevant to consider an in-depth analysis of which companies the models categorically fail to predict as defaulted, i.e. a scrutiny of the models' misclassifications. While the focus in this thesis is on improving the discriminative power of the practical approach with *theoretical* and *methodological* extensions, we recognize that complementary analyses are possible. As such, an analytical focus of the firms misclassified by the models would to a large extend complement our research. This potential research is in line with the critical realism position in philosophy of science, where it is pronounced to look for quasi-regularities in the characteristics of the misclassified firms (Moses & Knutsen, 2012). This infiltration of the deep domain would contribute to a research that approach a more complete understanding of what drives company default. This however is up to future researchers to contribute.

11. Bibliography

Addo, Peter, Dominique Guegan, and Bertrand Hassani. "Credit Risk Analysis Using Machine And Deep Learning Models." *Risks* 6.2 (2018): 38. Web. 14 May 2020.

Agarwal, Vineet, and Richard Taffler. "Does Financial Distress Risk Drive The Momentum Anomaly?." *Financial Management* 37.3 (2008): 461-484. Web.

Agresti, Alan, and Christine A Franklin. *Statistics: The Art And Science Of Learning From Data*. 3rd ed. Pearson, 2013. Print.

Altman, Edward I. "The Prediction Of Corporate Bankruptcy: A Discriminant Analysis." *The Journal of Finance* 23.1 (1968): 193. Web.

Baesens, Bart. *Analytics In A Big Data World*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2014. Print.

Baldassarri, Giorgio, and Alma Chen. *Spglobal.com*. N.p., 2016. Web. 15 May 2020.

Basel Committe on Banking Supervision. "The Core Principles Methodology." *Bis.org*. N.p., 1999. Web. 15 May 2020.

Bastos, João Augusto S. L. de A. "Credit Scoring With Boosted Decision Trees." *MRA Paper* 1.1 (2008): n. pag. Web.

Beaver, William H. "Financial Ratios As Predictors Of Failure." *Journal of Accounting Research* 4 (1966): 71. Web.

Belfrage, Claes, and Felix Hauf. "The Gentle Art Of Retroduction: Critical Realism, Cultural Political Economy And Critical Grounded Theory." *Organization Studies* 38.2 (2016): 251-271. Web.

Bellini, Tiziano. "Introduction To Expected Credit Loss Modelling And Validation." *IFRS 9 and CECL Credit Risk Modelling and Validation* (2019): 1-30. Web. 14 May 2020.

Belsley, David A, and Edwin Kuh. *Regression Diagnostics*. New York: Wiley, 1980. Print.

Bharath, Sreedhar T., and Tyler Shumway. "Forecasting Default With The Merton Distance To Default Model." *Review of Financial Studies* 21.3 (2008): 1339-1369. Web.

Bhaskar, Roy. *A Realist Theory Of Science*. 1975. Print.

Black, Fischer, and Myron Scholes. "The Pricing Of Options And Corporate Liabilities." *Journal of Political Economy* 81.3 (1973): 637-654. Web.

Boettke, Peter J., Christopher J. Coyne, and Peter T. Leeson. "Institutional Stickiness And The New Development Economics." *American Journal of Economics and Sociology* 67.2 (2008): 331-358. Web.

Bradley, Andrew P. "The Use Of The Area Under The ROC Curve In The Evaluation Of Machine Learning Algorithms." *Pattern Recognition* 30.7 (1997): 1145-1159. Web.

Breiman, Leo. "Statistical Modeling: The Two Cultures (With Comments And A Rejoinder By The Author)." *Statistical Science* 16.3 (2001): 199-231. Web.

Brockman, Paul, and H.J. Turtle. "A Barrier Option Framework For Corporate Security Valuation." *Journal of Financial Economics* 67.3 (2003): 511-529. Web.

Bryman, Alan. *Social Research Methods*. 4th ed. Oxford: Oxford University Press, 2012. Print.

Buch-Hansen, Hubert, and Peter Nielsen. *Kritisk Realisme*. Kbh.: Nota, 2012. Print.

Canella, Cara, Alexandra Elias, and Alex Kremzier. (2018): n. pag. Web. 7 Nov. 2019.

Chang, Yung-Chia et al. "Establishing Decision Tree-Based Short-Term Default Credit Risk Assessment Models." *Communications in Statistics - Theory and Methods* 45.23 (2016): 6803-6815. Web. 14 May 2020.

Charitou, Andreas, Evi Neophytou, and Chris Charalambous. "Predicting Corporate Failure: Empirical Evidence For The UK." *European Accounting Review* 13.3 (2004): 465-497. Web.

Courts, U.S. "Chapter 7 - Bankruptcy Basics." *United States Courts*. N.p., 2020. Web. 15 May 2020.

Dixon, W J. "Efficient Analysis Of Experimental Observations." *Annual Review of Pharmacology and Toxicology* 20.1 (1980): 441-462. Web. 14 May 2020.

Dixon, W. J. "Simplified Estimation From Censored Normal Samples." *The Annals of Mathematical Statistics* 31.2 (1960): 385-391. Web. 14 May 2020.

Duan, Jin-Chuan, and Keshab Shrestha. "Statistical Credit Rating Methods." *Global Credit Review* (2011): n. pag. Print.

Duffie, Darrell, and David Lando. "Term Structures Of Credit Spreads With Incomplete Accounting Information." *Econometrica* 69.3 (2001): 633-664. Web.

Engelmann, Bernd, and Robert Rauhmeier. *The Basel II Risk Parameters*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2011. Print.

Fawcett, Tom. "An Introduction To ROC Analysis." *Pattern Recognition Letters* 27.8 (2006): 861-874. Web. 14 May 2020.

Geske, Robert. "The Valuation Of Corporate Liabilities As Compound Options." *The Journal of Financial and Quantitative Analysis* 12.4 (1977): 541. Web.

Giacosa, Elisa et al. "Re-Estimation Of Company Insolvency Prediction Models: Survey On Italian Manufacturing Companies." *Corporate Ownership and Control* 14.1 (2016): 159-174. Web.

Goldreich, Oded. "Computational Complexity." *ACM SIGACT News* 39.3 (2008): 35. Web.

Hamer, Michelle M. "Failure Prediction: Sensitivity Of Classification Accuracy To Alternative Statistical Methods And Variable Sets." *Journal of Accounting and Public Policy* 2.4 (1983): 289-307. Web. 14 May 2020.

Hastie, Trevor et al. *An Introduction To Statistical Learning*. Springer, 2014. Print.

Hastie, Trevor, Jerome Friedman, and Robert Tisbshirani. *The Elements Of Statistical Learning*. New York: Springer, 2008. Print.

Heale, Roberta. "Validity And Reliability In Quantitative Studies." *Evidence Based Nursing* 18.3 (2017): 66-67. Web.

Hillegeist, Stephen A. et al. "Assessing The Probability Of Bankruptcy." *Review of Accounting Studies* 9.1 (2004): 5-34. Web.

Hosmer, David W, and Stanley Lemeshow. *Applied Logistic Regression*. Enskede: TPB, 1980. Print.

Jackson, Richard H.G., and Anthony Wood. "The Performance Of Insolvency Prediction And Credit Risk Models In The UK: A Comparative Study." *The British Accounting Review* 45.3 (2013): 183-202. Web.

James, Gareth et al. *An Introduction To Statistical Learning*. 2013. Print.

Keasey, Kevin, and Robert Watson. "The Prediction Of Small Company Failure: Some Behavioural Evidence For The UK." *Accounting and Business Research* 17.65 (1986): 49-57. Web. 14 May 2020.

Keerthi, S. Sathiya, and Chih-Jen Lin. "Asymptotic Behaviors Of Support Vector Machines With Gaussian Kernel." *Neural Computation* 15.7 (2003): 1667-1689. Web.

Lamontagne, Luc, and Mario Marchand. *Advances In Artificial Intelligence*. Berlin: Springer, 2006. Print.

Lateef, Fahad, and Yassine Ruichek. "Survey On Semantic Segmentation Using Deep Learning Techniques." *Neurocomputing* 338 (2019): 321-348. Web.

Laurentis, Giacomo De, Renato Maino, and Luca Molteni. *Developing, Validating And Using Internal Ratings: Methodologies And Case S*. John Wiley & Sons, 2010. Print.

Löffler, Gunter Johannes, and Peter N Posch. *Credit Risk Modeling Using Excel And VBA, Second Edition*. Chichester, West Sussex, U.K.: John Wiley & Sons Ltd, 2011. Print.

Marsh, David, and Gerry Stoker. *Theory And Methods In Political Science*. Basingstoke: Palgrave Macmillan, 2010. Print.

Marsh, David, and Gerry Stoker. *Theory And Methods In Political Science*. 3rd ed. Palgrave Macmillan, 2010. Print.

Medema, Lydian, Ruud H. Koning, and Robert Lensink. "A Practical Approach To Validating A PD Model." *Journal of Banking & Finance* 33.4 (2009): 701-708. Web. 14 May 2020.

Menard, Scott W. *Logistic Regression: From Introductory To Advanced Concepts And Applications*. Los Angeles: SAGE, 2010. Print.

Merton, Robert C. "ON THE PRICING OF CORPORATE DEBT: THE RISK STRUCTURE OF INTEREST RATES*." *The Journal of Finance* 29.2 (1974): 449-470. Web.

Moses, Jonathon W, and Torbjørn Lindstrøm Knutsen. *Ways Of Knowing*. London: Macmillan International Higher Education, 2012. Print.

Neisen, Martin, and Stefan Röth. *Basel IV: The Next Generation Of Risk Weighted Assets*. 2nd ed. Wiley Finance, 2018. Print.

Newson, Roger. "Parameters Behind "Nonparametric" Statistics: Kendall's Tau, Somers' D And Median Differences." *The Stata Journal: Promoting communications on statistics and Stata* 2.1 (2002): 45-64. Web. 14 May 2020.

North, Douglass C., and Michael C. Munger. "Institutions, Institutional Change And Economic Performance." *Southern Economic Journal* 58.1 (1990): 296. Web.

Ohlson, James A. "Financial Ratios And The Probabilistic Prediction Of Bankruptcy." *Journal of Accounting Research* 18.1 (1980): 109. Web.

Petropoulos, Anastasios et al. "Predicting Bank Insolvencies Using Machine Learning Techniques." *International Journal of Forecasting* (2020): n. pag. Web.

Plenborg, Thomas, Finhn Kinserdal, and Christian V Petersen. *Financial Statement Analysis*. Bergen: Fagbokforlaget, 2017. Print.

Provost, Foster, and Tom Fawcett. *Data Science For Business* . 1st ed. O'Reilly Media, 2013. Print.

Reisz, Alexander S., and Claudia Perlich. "A Market-Based Framework For Bankruptcy Prediction." *Journal of Financial Stability* 3.2 (2007): 85-131. Web.

Scott, David. "Critical Realism And Empirical Research Methods In Education." *Journal of Philosophy of Education* 39.4 (2005): 633-646. Web.

Shalev-Shwartz, Shai, and Shai Ben-David. *Understanding Machine Learning*. 1st ed. New York: Cambridge University Press, 2014. Print.

Sobehart, Jorge, and Roger Stein. "Benchmarking Quantitative Default Risk Models: A Validation Methodology." *The Journal of Risk Model Validation* 1.1 (2000): 77-113. Web.

Somers, Robert H. "A New Asymmetric Measure Of Association For Ordinal Variables." *American Sociological Review* 27.6 (1962): 799. Web.

Thyer, Bruce A. *Handbook Of Social Work Research*. London: SAGE, 2001. Print.

Tudela, Merxe, and Garry Young. "A Merton-Model Approach To Assessing The Default Risk Of UK Public Companies." *SSRN Electronic Journal* (2003): n. pag. Web.

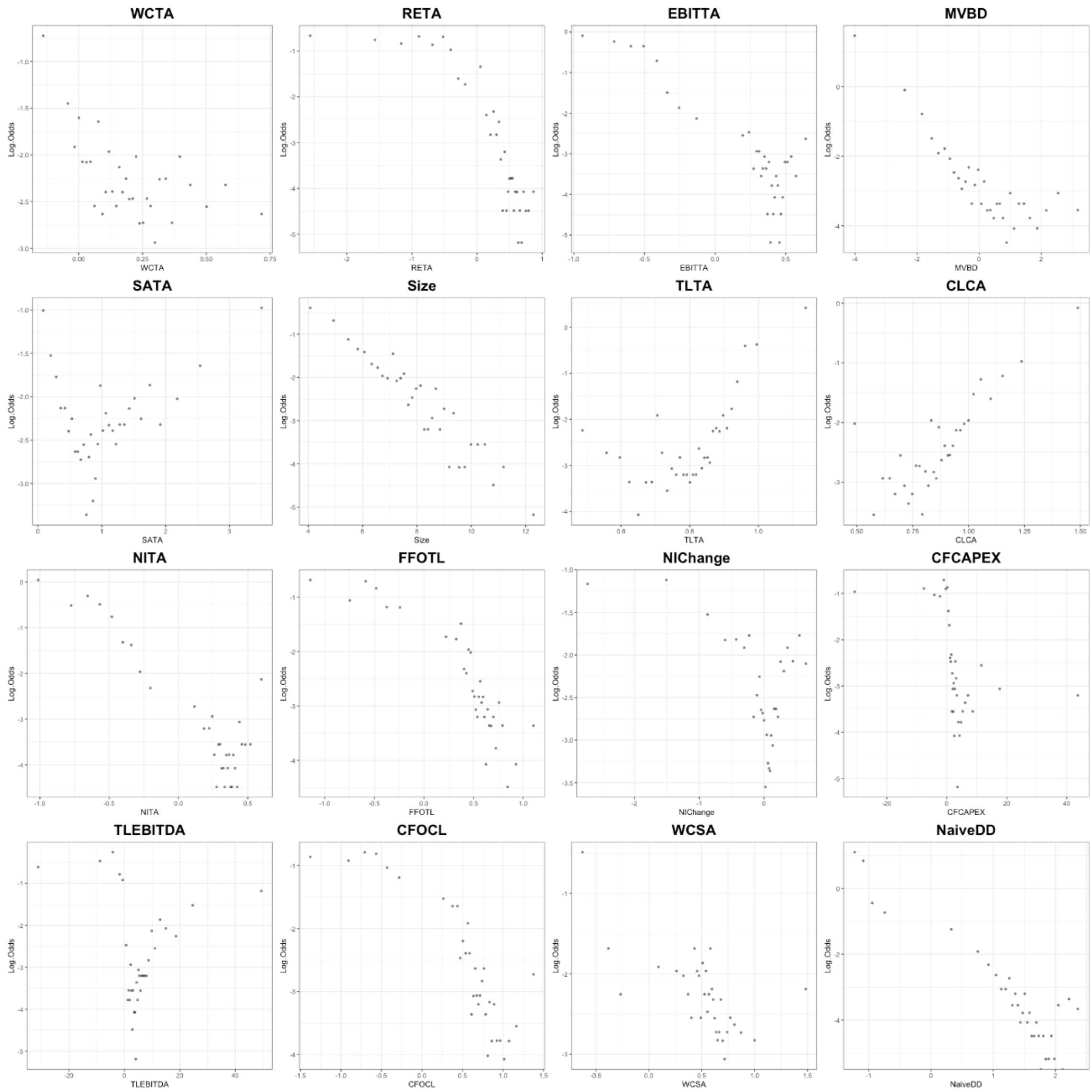
Tukey, John W. "The Future Of Data Analysis." *The Annals of Mathematical Statistics* 33.1 (1962): 1-67. Web. 14 May 2020.

Vassalou, Maria, and Yuhang Xing. "Default Risk In Equity Returns." *The Journal of Finance* 59.2 (2004): 831-868. Web.

Wooldridge, Jeffrey M. *Introductory Econometrics*. 6th ed. Cengage Learning, 2016. Print.

12. Appendices

12.1 Appendix: Linearity Plots



12.2 Appendix: Correlation Matrix from R-software

Correlation Matrix: The Default Model of Synthesis

Ratio	Somers' D	WC/TA	RE/TA	EBIT/TA	MV/BD	Size	TL/TA	CL/CA	NI/TA	FFO/TL	CF/CAPEX	CFO/CL	WC/SA	NaiveDD
WC/TA	0.2090	1.00	0.19	0.32	0.22	-0.06	-0.57	-0.53	0.35	-0.04	0	-0.09	0.43	0.12
RE/TA	0.6508		1.00	0.57	-0.02	0.36	-0.31	-0.18	0.56	0.31	0.21	0.27	0.06	0.42
EBIT/TA	0.6454			1.00	0.02	0.37	-0.36	-0.27	0.88	0.39	0.22	0.35	0.01	0.49
MV/BD	0.6526				1.00	0.09	-0.28	-0.16	0.03	-0.03	-0.06	-0.07	0.11	0.48
Size	0.5122					1.00	-0.07	-0.07	0.35	0.21	0.17	0.23	0.11	0.28
TL/TA	0.4802						1.00	0.55	-0.46	-0.04	-0.08	-0.03	-0.24	-0.45
CL/CA	0.3852							1.00	-0.31	-0.02	-0.03	-0.01	-0.25	-0.33
NI/TA	0.6442								1.00	0.3	0.2	0.27	0.06	0.48
FFO/TL	0.5120									1.00	0.39	0.82	-0.17	0.31
CF/CAPEX	0.5074										1.00	0.36	-0.07	0.13
CFO/CL	0.5088											1.00	-0.24	0.31
WC/SA	0.2438												1.00	0.15
NaiveDD	0.6864													1.00

12.3 Appendix: Code Book

Due to the comprehensiveness of the code in the R-software underlying the research at hand, we have created a separate code book, which will be submitted alongside with the thesis.