University of Bergamo

Department of Management, Economics and Quantitative Methods

XXVII ROUND OF THE PHD PROGRAM

"Economics, Applied Mathematics and Operational Research"

(supported by the Cotutelle agreement with Macquarie University, Sydney)

# DEFAULT PROBABILITIES IN CREDIT RISK MANAGEMENT: ESTIMATION, MODEL CALIBRATION, AND BACKTESTING

**Martin Gurny**

PhD Thesis

**Thesis Supervisors**

Prof. Sergio Ortobelli

Prof. Stefan Trück (Macquarie University)

Dr. Egon Kalotay (Macquarie University)

**Program Coordinator**

Prof. Marida Bertocchi

October 2015

# ACKNOWLEDGEMENTS

It is a pleasure to thank those who made this thesis possible.

My deepest gratitude goes to my supervisors, Professor Sergio Ortobelli from University of Bergamo and Professor Stefan Trück from Macquarie University. Their enthusiasm, patience, and expertise, positively added to my candidature experience. In particular, I would like to thank Professor Sergio Ortobelli for his guidance in early stages of my PhD candidature in Bergamo. He has always been available when I needed and his advice has been invaluable to me. Special thanks then goes to Professor Stefan Trück for his inspirational ideas, support, and all the help he provided, professionally and personally, during my stay at Macquarie University.

I am also thankful to my associate supervisor at Macquarie University, Dr. Egon Kalotay, for his insightful comments that enabled me to gain a better understanding of my work.

I would like to use this opportunity to thank fellow PhD candidates at both institutions for their friendship and emotional support through the difficult times.

Lastly, I would like to thank my loving family, mother Paulina, father Ivo, and brother Petr, for all the support they provided me throughout my candidature. I would not have finished this thesis without their enormous encouragement.

Martin Gurny

# ABSTRACT

This dissertation thesis is devoted to estimation and examination of default probabilities (PDs) within credit risk management. Assigning an appropriate PD is a widely employed strategy by many financial institutions as well as the supervisory authorities and providing accurate estimates can be considered as one of the key challenges in credit risk management. False estimation of PDs leads to, among other things, unreasonable ratings and incorrect pricing of financial instruments. As a matter of fact, these issues were among the key reasons for the global financial crisis (GFC) as undervaluation of the risk caused the collapse of the financial system.

In the first study, we discuss structural models based on the Merton's framework. First, we observe that the classical distributional assumption of the Merton model is generally rejected. Second, we implement a structural credit risk model based on stable non-Gaussian processes as a representative of subordinated models in order to overcome some drawbacks of the Merton one. Finally, following the Moody's KMV estimation methodology, we propose an empirical comparison between the results obtained from the classical Merton model and the stable Paretian one. In particular, we suggest alternative parameter estimation for subordinated processes, and we optimize the performance for the stable Paretian model. Our results indicate that PD is generally underestimated by the Merton model and that the stable Lévy model is substantially more sensitive to the periods of financial crises.

The second study is devoted to examination of the performance of static and multi-period credit-scoring models for determining PDs of financial institutions. We use an extensive database for the U.S. provided by the Federal Financial Institutions Examination Council (FFIEC). In fact, our extensive sample contains more than seven thousand U.S. commercial banks with over four hundred default events. Our analysis also focuses on evaluating the performance of the considered scoring techniques. We apply a substantial number of model evaluation methods, including techniques that have not yet been applied in the literature on credit scoring. We also provide an overall ranking of the models according to the different evaluation criteria and find that the considered scoring models provide a high predictive

accuracy in distinguishing between default and non-default financial institutions. Despite the difficulty of predicting defaults in the financial sector as it has been mentioned in the literature, the proposed models perform very well also in comparison to results on scoring techniques for the corporate sector.

Finally, in the third study, we investigate the question whether distressed renewable energy companies earn on average higher returns than low distress risk companies. Using the Expected Default Frequency (EDF) measure obtained from Moody's KMV, we demonstrate that there is a positive cross-sectional relationship between returns of both, equally-weighted (EW) and value-weighted (VW) portfolios, and evidence for a distress risk premium in the U.S. renewable energy sector. The positively priced distress premium is also confirmed by investigating returns corrected for common Fama and French and Carhart risk factors. We further show that raw and risk-adjusted returns of value-weighted portfolios that take a long position in the 20% most distressed stocks and a short position in the 20% safest stocks generally outperform the S&P 500 index throughout our sample period (2002–2014).

# Table of Contents

## List of Figures

## List of Tables

## List of Abbreviations

| | |
|---|---|
| AGINA | the Ardour Global Alternative Energy Index North America |
| AIRB | the advanced internal ratings based approach |
| AR | accuracy ratio |
| AVAR | Asymmetric Vector Autoregression model |
| AVR | asset value return |
| BCBS | the Basel Committee on Banking Supervision |
| BEKK | Baba-Engle-Kraft-Kroner model |
| BIS | Bank for International Settlements |
| BM | book-to-market (ratio) |
| C_IBD | cost of total interest bearing deposits |
| CAPM | Capital Asset Pricing Model |
| CDS | credit default spread |
| CELS | the NASDAQ Clean Edge Green Energy Index |
| CFPB | the Consumer Financial Protection Bureau |
| CRA | the Community Reinvestment Act |
| CRSP | the Center for Research in Security Prices |
| DA | discriminant analysis |
| DD | distance-to-default |
| DoV | distribution of values |
| EAD | exposure at default |
| ECAI | external credit assessment institution |
| ECO | the WilderHill Clean Energy Index |
| EDF | Expected Default Frequency |
| EL | expected loss |
| EQ_TA | total equity capital & minority interests / total assets |
| ER | efficiency ratio |
| EW | equally-weighted (portfolio, return) |

| | |
|---|---|
| EWS | early warning system |
| FDIC | the Federal Deposit Insurance Corporation |
| FFIEC | the Federal Financial Institutions Examination Council |
| FIRA | the Financial Institutions Regulatory and Interest Rate Control Act |
| FIRB | the foundation internal ratings based approach |
| FN | false negative |
| FP | false positive |
| FRB | the Board of Governors of the Federal Reserve System |
| GARCH | Generalized Autoregressive Conditional Heteroskedasticity model |
| GCorr | Global Correlation Model |
| GFC | global financial crisis |
| GWE | the ISE Global Wind Energy Index |
| HL | Hosmer-Lemeshow test / statistic |
| HMDA | the Home Mortgage Disclosure Act |
| HML | high-minus-low (value factor) |
| HSD | honesty significant difference |
| IE_II | total interest expense / total interest income |
| II_EA | total interest income / interest earning assets |
| IRB | internal ratings based approach |
| JB | Jarque-Bera test |
| KMV | KMV Corporation / Kealhofer-McQuown-Vasicek |
| KS | Kolmogorov-Smirnov test |
| LGD | loss given default |
| LLA_TLL | loans & leases allowance / total loans & leases |
| LP_AA | provision for loan & lease losses / average assets |
| M | maturity |
| MA | maturity adjustment |
| MC | market capitalization |
| MDA | multiple discriminant analysis |

| | |
|---|---|
| MLE | maximum likelihood estimation |
| MOM | winners-minus-losers (momentum factor) |
| NCRLL_GLL | noncurrent loans & leases / gross loans & leases |
| NCUA | the National Credit Union Administration |
| NEX | the WilderHill New Energy Global Innovation Index |
| NIG | Normal Inverse Gaussian process |
| NIM | net interest margin |
| NL_TLL | net loss / average total loans & leases |
| NLL_EQ | net loans & leases / total equity capital & minority interests |
| NLL_TA | net loans & leases / total assets |
| OCC | the Office of the Comptroller of the Currency |
| OTC | over-the-counter market |
| PD | probability of default |
| RCLL | right continuous with left-hand limits |
| RE_EQ | retained earnings / total equity capital & minority interests |
| RENIX | the Renewable Energy Industrial Index |
| RM | market portfolio (market factor) |
| ROA | return on assets |
| ROC | relative / receiver operating characteristic |
| ROE | return on equity |
| RR | recovery rate |
| RWA | risk weighted assets |
| S&P | Standard & Poor's |
| SMB | small-minus-big (size factor) |
| SPGCE | the S&P Global Clean Energy Index |
| STD | the standardized approach |
| T1LC | tier one leverage capital ratio |
| T1RBC_RWA | tier one risk-based capital / risk-weighted assets |
| TD_EQ | total deposits / total equity capital & minority interests |

| | |
|---|---|
| TN | true negative |
| TP | true positive |
| TRBC_RWA | total risk-based capital / risk-weighted assets |
| UBPR | the Uniform Bank Performance Report |
| UL | unexpected loss |
| VaR | Value-at-Risk |
| VAR | Conventional Vector Autoregression model |
| VG | Variance Gamma process |
| VK | Vasicek-Kealhofer model |
| VW | value-weighted (portfolio, return) |
| WCDR | worst-case default rate |
| WRDS | Wharton Research Data Services |

# Chapter 1

# Introduction

Credit risk and estimation of default probabilities (PDs), that represent a borrower's risk level, have become one of the most intensely studied topics in the financial literature and have undergone tremendous developments in the last decades. The PD indicates a probability that a given counterparty will not be able to meet its obligations and is one of the key input factors for the modeling and measurement of credit risk. Its estimation is nowadays a widely employed strategy by many financial institutions and supervisory authorities. The significance of this assessment substantially increased since 2008 when several countries had encountered a period of financial and economic turmoil often referred to as the global financial crisis (GFC). Providing accurate estimates of PDs can be considered as one of the key challenges in credit risk management. False estimation of PDs leads to unreasonable ratings and incorrect pricing of financial instruments. As a matter of fact, these issues were among the key reasons for the GFC as undervaluation of the risk caused the collapse of the financial system which had been extended through credit derivatives on global markets. Probabilities of default can also be considered as key parameters for the calculation of economic and regulatory capital of financial institutions under the Basel II and Basel III Accords that emphasise the risk sensitivity of the capital of commercial banks. These reasons highlight how important the estimation of PD is and why it has been a significant research topic for a long time.

This introductory chapter is mostly based on following sources: Crosbie and Bohn [42], Bluhm, Overbeck and Wagner [23], Duffie and Singleton [53], Hull [92], Sironi and Resti [159], and Trück and Rachev [169].

## 1.1 Credit risk, ratings, and probability of default (PD)

There are several definitions of credit risk in the literature. For instance, Duffie and Singleton [53] define credit risk as "*the risk of default or of reductions in market value caused by changes in the credit quality of issuers or counterparties*". According to Hull [92], "*credit risk arises from the possibility that borrowers, bond issuers, and counterparties in derivatives transactions may default*". Sironi and Resti [159] then define credit risk as "*the possibility that an unexpected change in a counterparty's creditworthiness may generate a corresponding unexpected change in the market value of the associated credit exposure*". These definitions combine the following three concepts.

1. **Default risk and migration risk** – Credit risk is not limited to the possibility of the counterparty's default: even a mere deterioration in its creditworthiness constitutes a manifestation of credit risk. Therefore, it comprises two different cases: the risk of default and the risk of migration. While the former represents the risk of loss resulting from the borrower's actual insolvency (whereby payments are interrupted), the latter expresses the risk of loss resulting from a mere deterioration in its credit rating.

2. **Risk as an unexpected event** – In order to be considered as a risk, the variation in the counterparty's credit rating must be unexpected. As a matter of fact, expected developments in the borrower's economic/financial status are always taken into account when the PD and associated interest rates are determined. The real risk is represented by the possibility that those evaluations could later prove incorrect. That is, that a deterioration in the counterparty unforeseen by the lender occurs. In this respect, proper risk only relates to events which, although foreseeable, are unexpected.

3. **Credit exposure** – With respect to the concept of credit exposure, credit risk is by no means limited to the "classic" forms of credit granted by a bank (on-balance-sheet loans and securities), but also includes off-balance-sheet operations such as guarantees, derivative contracts traded on OTC (over-the-counter) markets, and transactions in securities, foreign currencies or derivatives pending final settlement.

Apart from above mentioned default and migration risks, credit risk comprises several other types of risk:

- **spread risk**: the risk associated with a rise in the spreads required of borrowers (e.g. bond issuers) by the market; in the event of increased risk aversion by investors, the spread associated with a given PD (and therefore a given rating class) may increase; in such a case the market value of the securities declines, without any reduction in the issuer's credit rating;
- **recovery risk**: indicates the risk that the recovery rate actually recorded after the liquidation of the insolvent counterparty's assets will be less than the amount originally estimated (either because the liquidation value was lower than estimated or simply because the recovery process took longer than expected);
- **pre-settlement (substitution risk)**: indicates the risk that the bank's counterparty in an OTC derivative will become insolvent before the maturity of the contract, thus forcing the bank to "replace" it at new (and potentially less favourable) market conditions;
- **country risk**: indicates the risk that a non-resident counterparty will be unable to meet its obligations due to events of a political or legislative nature, such as the introduction of foreign exchange constraints, which prevent it from repaying its debt.

### 1.1.1 Expected (EL) and unexpected loss (UL)

There is a need of a loss protection in terms of an insurance for the bank. In terms of credit risk, we distinguish between expected and unexpected loss. The basic idea behind the expected loss is as follows. The bank assigns to every customer a default probability (**PD**), a loss fraction called the loss given default (**LGD**), describing the fraction of the loan's exposure expected to be lost in case of default, and the exposure at default (**EAD**) subject to be lost in the considered time period. The loss of any obligor is then defined by a loss variable

$$\tilde{L} = \mathrm{EAD} \times \mathrm{LGD} \times L \qquad \text{with} \qquad L = \mathbf{1}_D, \qquad \mathrm{P}(D) = \mathrm{PD}, \tag{1.1}$$

where $D$ denotes the event that the obligor defaults in a certain period of time (most often one year), and $\mathrm{P}(D)$ denotes the probability of $D$. There is a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ underlying this concept, consisting of a sample space $\Omega$, a *σ-Algebra* $\mathcal{F}$, and a probability measure $\mathrm{P}$. The elements of $\mathcal{F}$ are the measurable events of the model, and intuitively it makes sense to claim that the event of default should be measurable. Moreover, it is common to identify $\mathcal{F}$ with the information available, and the information if an obligor defaults or survives should be included in the set of measurable events.

We can now define the expected loss (**EL**) of any customer as the expectation of its corresponding loss variable $\tilde{L}$, namely

$$\mathrm{EL} = \mathrm{E}\left[\tilde{L}\right] = \mathrm{EAD} \times \mathrm{LGD} \times \mathrm{P}\left(D\right) = \mathrm{EAD} \times \mathrm{LGD} \times \mathrm{PD}, \tag{1.2}$$

because the expectation of any Bernoulli random variable, like $\mathbf{1}_D$, is its event probability.

The EL of a transaction is an insurance or loss reserve that covers losses the bank expects from historical default experience. But holding capital as a cushion against expected losses is not enough. In fact, the bank should in addition to the expected loss reserve also save money for covering unexpected losses exceeding the average experienced losses from past history. As a measure of the magnitude of the deviation of losses from the EL, the standard deviation of the loss variable $\tilde{L}$ as defined in (1.1) is a natural choice. For obvious reasons, this quantity is called the unexpected Loss (**UL**) and is defined by

$$\mathrm{UL} = \sqrt{\mathrm{var}\left[\tilde{L}\right]} = \sqrt{\mathrm{var}\left[\mathrm{EAD} \times \mathrm{LGD} \times L\right]}. \tag{1.3}$$

The distinction between EL and UL is important when dealing with a diversified portfolio of exposures. The EL on such a portfolio is simply equal to the sum of the ELs on the individual loans in it, whereas the volatility of the total portfolio loss is generally lower than the sum of the volatilities of the losses on individual loans (and much more so if the correlation between individual loans is low). In other words, while EL cannot be reduced by diversifying the portfolio, UL (i.e. the volatility of losses around the mean) can be reduced through a suitable portfolio strategy. This means that an effective loan portfolio diversification policy, while leaving total expected returns unchanged, can significantly reduce total credit risk.

### 1.1.2 Credit ratings

A rating is an indicator of creditworthiness of customers, where quantitative as well as qualitative information is used to evaluate a client. In practise, the rating procedure is often more based on the judgment and experience of the rating analyst than on pure mathematical procedures with strictly defined outcomes. Ratings are assigned to customers either by external rating agencies such as Moody's, Standard & Poor's (S&P), or Fitch, or bank-internal rating methodologies.

One of the objectives of rating agencies when they assign ratings is rating stability. Therefore, ratings change only when there is reason to believe that a long-term change in the company's creditworthiness has taken place. This goes hand in hand with the fact that rating agencies also try to "rate through the cycle". If the economy exhibits a downturn with subsequent effect of increasing the company's PD in the next six months, but makes very little difference to the company's PD over the next three to five year, then a rating agency would not usually change the company's credit rating in these circumstances. There are other companies (e.g. Moody's KMV), though, that provide PD estimates based on equity price and other variables. These estimates tend to respond more quickly to market information than credit ratings. The types of models that are used to produce the estimates will be discussed in Section 1.3.2.

The ratings published by rating agencies are available only for relatively large corporate clients. Many small and medium size businesses do not issue publicly traded bonds and therefore are not rated by rating agencies. That is the reason why most banks have procedures for rating the creditworthiness of their corporate and retail clients. The internal-ratings-based (IRB) approach in Basel II allows banks to use their internal ratings in determining the PD and will be discussed in Section 1.2.1.

### 1.1.3 Real world vs. risk-neutral PDs

Just as with interest-rate risk, differences between real world and risk-neutral PDs reflect risk premia required by market participants to take on the risks associated with default. In general, default-risk premia reflect aversion to both the risk of timing of default and to the risk of severity of loss in the event of default.

Risk-neutral PDs (PDs in a world where all investors are risk-neutral) are backed out of bond yields or credit default swap (CDS) spreads and are also sometimes called *implied default probabilities*. These probabilities are higher than real world ones and should be used for valuing credit derivatives and estimating the present value of the cost of default. By contrast, PDs implied from historical data are real world PDs, sometimes also called *physical default probabilities*. These probabilities should be used when carrying out scenario analyses to calculate potential future losses from defaults or for calculating credit VaR and regulatory capital.

There are several reasons why we can often see substantial differences between these two types. First, corporate bonds are relatively illiquid and the returns on bonds are higher than they would otherwise be to compensate for this. Second, the subjective PDs of bond traders may be much higher than the estimates from historical data. Bond traders might be allowing for depression scenarios much worse than anything seen in the period covered by their data. Third, and most importantly, bonds do not default independently of each other which leads to systematic risk that cannot be diversified away. Finally, bond returns are highly skewed with limited upside. Unlike stocks where idiosyncratic risk can be diversified away by choosing a portfolio of several dozen stocks, this is difficult for bonds and a proper diversification would require tens of thousands of different bonds to be held. As a result, bond traders may earn an extra return for bearing idiosyncratic as well as the systematic risk.

## 1.2 PD and Basel Accords

In 1988, the *Basel Committee on Banking Supervision*[1] introduced its BIS Accord (also known as **Basel I**). Although it improved the way capital requirements were determined, it had significant weaknesses. To name a few, all loans by a bank to a company had a risk weight of 100% and required the same amount of capital (independently on a credit rating of the company in question), no model for default correlation, etc. In June 1999, the Basel Committee proposed new rules with a more risk-sensitive framework, the *New Basel Capital Accord* (also known as **Basel II**). This Accord consists of three mutually reinforcing pillars, which together contribute to safety and soundness in the financial system.

1. **Minimum capital requirements** – The first pillar sets out the minimum capital requirements and defines the minimum ratio of capital to risk-weighted assets (RWA). The new framework maintains both the current definition of the total capital and the

[1] The Basel Committee on Banking Supervision (BCBS) is a committee of central banks and bank supervisors from the major industrialised countries that meet every three months at the *Bank for International Settlements* (BIS) in Basel.

minimum requirement of at least 8% of the bank's capital to its risk-weighted assets (RWA).

$$\text{Total Capital} = 0.08 \times (\text{credit risk RWA} + \text{market risk RWA} + \text{operational risk RWA}) \tag{1.4}$$

2. **Supervisory review** – The second pillar is concerned with the supervisory review process and requires supervisors to undertake a qualitative review of their bank's capital allocation techniques and compliance with relevant standards. It places more emphasis on early interventions when problems arise and supervisors are required, apart from ensuring that the minimum capital required is held, to encourage banks to develop and use better risk management techniques and to evaluate these techniques.

3. **Market discipline** – The third pillar aims to bolster market discipline through enhanced disclosure requirements by banks which facilitate market discipline. The idea here is that banks will be subjected to added pressure to make sound risk management decisions if shareholders and potential shareholders have more information about those decisions.

Following the 2007-2009 credit crisis, the Basel Committee realized that a major overhaul of Basel II was necessary. The final version of **Basel III** was published in December 2010[2] and the regulations are being implemented gradually between 2013 and 2019. There are six parts to the regulations: capital definition and requirements, capital conservation buffer, countercyclical buffer, leverage ratio, liquidity risk, and counterparty credit risk. The Tier 1 plus Tier 2 capital requirement is the same as under Basel II, however the definition of what qualifies as equity capital for regulatory purposes has been tightened (see, e.g., Hull [92]).

### 1.2.1 Credit risk capital under Basel II

There are three approaches for credit risk under Basel II which banks can choose from: the standardized approach (STD), the foundation internal ratings based approach (FIRB), and the advanced internal ratings based approach (AIRB).

[2] See Basel Committee on Banking Supervision [13-14].

**The Standardized Approach (STD)**

The STD approach is the simplest of the three broad approaches to credit risk and is used by banks that are not sufficiently sophisticated (from the regulator's point of view) to use the internal rating approaches. The bank allocates a risk weight to each of its assets and off-balance-sheet positions and produces a sum of RWA values. A risk weight of 100% means that an exposure is included in the calculation of RWA at its full value, which translates into a capital charge equal to 8% of that value. Similarly, a risk weight of 20% results in a capital charge of 1.6% (i.e. 20% of 8%). Individual risk weights depend on the broad category of the borrowers, which are sovereigns, banks and corporates. Under Basel II, the risk weights are refined by the reference to a rating provided by an *external credit assessment institution* (ECAI), such as rating agencies.

The risk weight for a country (sovereign) exposure ranges from 0% to 150% and the risk weight for an exposure to another bank or a corporation ranges from 20% to 150%. Supervisors are allowed to apply lower risk weights (20% rather than 50%, 50% rather than 100%, and 100% rather than 150%) when exposures are to the country in which the bank is incorporated or to that country's central bank. For claims on banks, national supervisors can choose to base capital requirements on the rating of the country in which the bank is incorporated. The risk weight assigned to the bank will be 20% if the country of incorporation has a rating between AAA and AA–, 50% if it is between A+ and A–, 100% if it is between BBB+ and B–, 150% if it is below B–, and 100% if it is unrated. The standard rule for retail lending is that a risk weight of 75% be applied. When claims are secured by a residential mortgage, the risk weight is 35%. Because of poor historical loss experience, the risk weight for claims secured by commercial real estate is 100%.

**The Internal Ratings Based Approach (IRB)**

Under the IRB approach, banks are allowed to use their internal estimates of borrower creditworthiness to assess credit risk in their portfolios, subject to strict methodological and disclosure standards, and translate the results into estimates of a potential future loss amount. Regulators base the capital requirement on the value at risk (VaR) calculated using a one-year horizon and a 99.9% confidence level. They recognize that expected losses are usually covered by the way banks price their products. The capital required is therefore the VaR minus the expected loss.

The VaR is calculated using the one-factor Gaussian copula model of time to default. Assume that a bank has a very large number of obligors and the *i*-th obligor has a one-year probability of default equal to $PD_i$. The formula for the so-called worst-case default rate (WCDR), defined so that the bank is 99.9% certain it will not be exceeded next year for the *i*-th counterparty, looks as follows (see, e.g., Trück and Rachev [169] for derivation) :

$$\mathrm{WCDR} = \Phi\left[\frac{\Phi^{-1}\left(\mathrm{PD}_i\right) + \sqrt{\rho}\,\Phi^{-1}\left(0.999\right)}{\sqrt{1-\rho}}\right] \tag{1.5}$$

where $\Phi$ denotes the cumulative standard normal distribution function and $\rho$ is the copula correlation between each pair of obligors. This can be considered as the core of the function for calculating the RWA in the IRB approach. Overall, the formula for risk weighted assets RWA is:

$$\mathrm{RWA} = 12.5 \times \mathrm{EAD} \times \mathrm{LGD} \times \left(\mathrm{WCDR} - \mathrm{PD}\right) \times \mathrm{MA}. \tag{1.6}$$

Note that the RWA equals 12.5 times the capital required, so that the required capital is 8% of RWA. Obviously, next to the probability of default PD and the worst-case default rate WCDR, also the factors exposure at default (EAD), loss given default (LGD), and a maturity adjustment (MA) enter the calculation of RWA. Furthermore, the calculated WCDR is dependent on the copula correlation parameter $\rho$.

Thus, the capital charge for the exposures depends on a set of following risk components (inputs) which are provided either through the application of standardized supervisory rules (FIRB approach) or internal assessments (AIRB approach), subject to supervisory minimum requirements.

**Probability of Default (PD)**: All banks, whether using the FIRB or the AIRB methodology, have to provide an internal estimate of the PD associated with the borrowers in each borrower grade. Each estimate of PD has to represent a conservative view of a long-run average PD for the grade in question and has to be grounded in historical experience and empirical evidence. The preparation of the estimates, the risk management processes, and the rating assignments that lay behind them have to reflect full compliance with supervisory minimum requirements to qualify for the IRB recognition.

**Loss Given Default (LGD)**: While the PD associated with a given borrower does not depend on the features of the specific transaction, LGD is facility-specific. Losses are generally understood to be influenced by key transaction characteristics such as the presence of collateral and the degree of subordination. It is equal to one minus the expected recovery rate (RR) on the exposure.

**Exposure at Default (EAD)**: As with LGD, EAD is also facility-specific. It is represented by the current exposure plus the possible variation in the size of the loan which may take from now to the date of possible default.

**Maturity (M)**: Where maturity is treated as an explicit risk component (AIRB approach), banks are expected to provide supervisors with the effective contractual maturity of their exposures. Where there is no explicit adjustment for maturity, a standard supervisory approach is presented for linking effective contractual maturity to capital requirements.

With regards to above described risk components, the main difference between the two IRB approaches is following. In the foundation methodology (FIRB), banks estimate the PD associated with each borrower, and the supervisors supply the other inputs (LGD, EAD, M). In the advanced methodology (AIRB), banks with sufficiently developed internal capital allocation processes are permitted to supply other necessary inputs as well. Under both IRB approaches, the range of risk weights are far more diverse than those in the STD approach, resulting in greater risk sensitivity.

Basel II assumes a relationship between the correlation parameter $\rho$ and the PD. Following Lopez [122], this relationship can be described by the following expression:

$$\rho(\mathrm{PD}) = 0.12\left(\frac{1-e^{-50\mathrm{PD}}}{1-e^{-50}}\right) + 0.24\left[1-\left(\frac{1-e^{-50\mathrm{PD}}}{1-e^{-50}}\right)\right]. \tag{1.7}$$

A very close approximation of this relationship is provided by the more simple expression:

$$\rho(\mathrm{PD}) = 0.12\left(1 + e^{-50\mathrm{PD}}\right). \tag{1.8}$$

Clearly, according to these expressions, the correlation declines with increasing PD. The reason usually given for this inverse relationship is as follows. As a company becomes less creditworthy, its PD increases, becomes more idiosyncratic and less affected by overall market conditions.

Finally, the maturity adjustment (MA) in equation (1.6) is defined as

$$\mathrm{MA} = \frac{1 + \left(\mathrm{M} - 2.5\right) \times b\left(\mathrm{PD}\right)}{1 - 1.5 \times b\left(\mathrm{PD}\right)} \tag{1.9}$$

with

$$b\left(\mathrm{PD}\right) = \left[0.11852 - 0.05478 \times \ln\left(\mathrm{PD}\right)\right]^2. \tag{1.10}$$

The maturity adjustment is designed to allow for the fact that, if an instrument lasts longer than one year, there is a one-year credit exposure arising from a possible decline in the creditworthiness of the counterparty as well as from a possible default by the counterparty.

As we have seen, for both the FIRB approach and the AIRB approach the probability of default (PD) is a key parameter for the modeling and measurement of credit risk. Therefore, next section will be devoted to various possibilities of its estimation and quantification.

## 1.3 PD and its quantification

In general, there are two types of approaches to estimating default probabilities.[3] First type are so-called credit-scoring models that are based on economic and financial indicators of a company. These models use various statistical methods such as discriminant analysis, regressions models, or inductive models. Second type is then the utilization of market valuation. In this case we are talking either about structural models (based on equity prices) or so-called reduced-form models (based on bond prices).

[3] As a matter of fact, there are also other possibilities how to estimate PDs. One can take advantage of credit ratings provided by rating agencies (if a company has issued publicly traded debt), hazard rates, credit default swap (CDS) spreads, or asset swap spreads. For more details, see, e.g., Hull [92].

### 1.3.1 Scoring systems

A class of statistical models, generally known as credit-scoring models, belongs to the most widely used models to predict a company's default. They can be found in virtually all types of credit analysis, from consumer credit to commercial loans. The idea is to pre-identify certain key factors that determine the PD and combine or weight them into a quantitative score. This score can be either directly interpreted as a probability of default or used as a classification system.

Two major seminal papers in the area on bankruptcy prediction have been published in the 1960's by Beaver [15] and Altman [4]. Since then an impressive body of theoretical and especially empirical research concerning this topic has evolved and we will address these studies in Chapter 3 in more detail. As the major methodologies for credit scoring should be mentioned logit models, probit models, discriminant analysis models and more recently, neural networks.

**Discriminant analysis**

Discriminant analysis (DA) or multiple discriminant analysis (MDA) tries to derive the linear combination of two or more independent variables that will discriminate best between a priori defined groups, which in the simplest case are failing and non-failing companies. A basic principal is to maximize the difference between two groups, while the differences among particular members of the same group are minimized.

DA can also be thought of as multiple regression. If we code the two groups in the analysis as 1 and 2 and use that variable as the dependent one in a multiple regression analysis, analogous results to using a discriminant analysis could be obtained (see Trück and Rachev [169]). This is due to the statistical decision rule of maximizing the between-group variance relative to the within group variance in the discriminant analysis technique. DA derives the linear combinations from an equation that takes the following form:

$$Z = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \ldots + \gamma_n X_n \qquad (1.11)$$

where $Z$ is the discriminant score ($Z-\text{score}$), $\gamma_0$ is a constant, $\gamma_i$ (*i* = 1, 2, … ,*n*) the discriminant coefficients, and $X_i$ (*i* = 1, 2, … ,*n*) the independent variables, i.e. the financial ratios.

**Logit and probit models**

Logit and probit models can be considered to be among the most popular approaches in the empirical default-prediction literature, see, e.g., Ohlson [141], Zmijewski [180], or Shumway [157]. Unlike the linear probabilistic model, where the outcome variable may be above 100% or below 0%, transformations used in logit a probit models guarantee that the dependent variable is always between 0 and 100%, and can therefore be correctly interpreted as a PD. These models can be easily applied to cases where the dependent variable is either nominal or ordinal, and has two or more levels. Further, the independent variables can be any mix of qualitative and quantitative predictors.

Logit and probit models allow for estimation of the probability for the occurrence of defined event. In credit scoring, the studied event is the default or credit failure of a company. Thus, the response variable $Y$ takes on the value $Y=1$ if company failed, and $Y=0$, otherwise. We are interested in modeling the probability $Y$ by specifying the following model:

$$Y = f\left(\alpha + \sum_{i=1}^{n} \beta_i X_i\right) \tag{1.12}$$

where $X_i$ (*i* = 1, 2, … ,*n*) are the explanatory variables, $\alpha$ is a constant, and $\beta_i$'s are the estimated weights of $X_i$.

The literature suggests various ways to specify the function $f$. In case of the logit model, we apply the so-called logistic transformation

$$Y = \frac{\exp\left(\alpha + \sum_{i=1}^{n} \beta_i X_i\right)}{1 + \exp\left(\alpha + \sum_{i=1}^{n} \beta_i X_i\right)} = \frac{1}{1 + \exp\left(-\alpha - \sum_{i=1}^{n} \beta_i X_i\right)}. \tag{1.13}$$

For the probit model, the cumulative distribution function of the normal distribution is used:

$$Y = \int_{-\infty}^{\alpha + \sum_{i=1}^{n} \beta_i X_i} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt. \tag{1.14}$$

We will discuss these two models in more detail in Chapter 3.

**Inductive models (neural networks and genetic algorithms)**

Discriminant analysis and logit and probit models have a common denominator – the attempt to identify the fundamental relationships which explain the economic/financial balance of a company, and can therefore be used to forecast default (Sironi and Resti [159]). These models follow a "structural" approach: they start with assumptions made by an analyst and seek confirmation for these assumptions in an empirical data sample. Inductive models, however, use a purely inductive process: if, starting from a data sample, a certain empirical regularity is found, it is used in a substantially uncritical way to forecast future defaults by other companies. Hence, instead of relying on deductively determined rules, a purely empirical approach is used.

These models are often referred to as "black boxes". They can be used to generate results rapidly, but their logic may not be fully understood. They have one significant drawback, though. As users, we do not really know what is happening in "hidden layers" between inputs and outputs. Despite this fact, inductive approach might be very useful, particularly in such cases where it is almost impossible to design the rules underlying a certain phenomenon (Sironi and Resti [159]).

A neural network consists of a large number of elements (neurons), which are connected to one another by elementary relations (synapses). The neurons are arranged in "layers"; each neuron in the outermost layer of the network receives an input of *n* variables and processes them with a linear or, more often, non-linear function, the result of which is passed on to the neurons in the next layer. These neurons also process the input received with a further function, and transmit a new output to the next layer in the network. After one or more hidden layers, the network generates a final result. In the case of default forecasting, the result may be, for example, a numerical score which must have a value as close as possible to 1 for abnormal companies and as close as possible to 0 for healthy ones.

The coefficients of the individual elementary functions that make up the network are estimated by means of an iterative mechanism. In practise, the values of the coefficients are gradually modified to obtain results as similar as possible to the desired ones. The learning process of a network is therefore a gradual attempt to identify the correct weights to be attributed to the input variables and the synapses of the hidden layers, so as to obtain a result similar to that of the (unknown) function to be approximated. For more detailed discussion on neural networks see, e.g., Bishop [20].

Genetic algorithms, like neural networks, are inspired by the behavior of biological organisms. Their operation is based on a transposition of Darwin's principles of natural selection and "survival of the fittest". A structure of genetic algorithms is based on the principles of natural evolution, where only the individuals with good characteristics to interact with the external environment have a high probability to survive. This evolution process therefore leads to a continuous improvement of the species. Genetic algorithms simulate this process with one difference. The "individuals" required to evolve are not living organisms, but possible solutions to a problem.

If, for the sake of simplicity, we consider a linier function:

$$z = \alpha_0 + \alpha_1 x_x + \alpha_2 x_2 + \ldots + \alpha_m x_m, \tag{1.15}$$

then each individual is represented by a vector $\alpha = [\alpha_0, \ldots, \alpha_m]'$ which indicates the algebraic sign and weight with which the various balance sheet indicators are included in the construction of *z*. Nil values of one or more $\alpha_j$ indicate that the corresponding economic/financial indicators are not used by this individual/solution.

Although genetic algorithms do not guarantee to identify the "ideal" solution to the problem for which they are used, they often allow good solutions to be obtained very quickly. This approach has proved to be particularly effective in areas where other research methods had been producing poor results due to a presence of a solutions space which is not only large, but also little-known or "noisy". For further discussion on genetic algorithms we refer to, e.g., Mitchell [136].

### 1.3.2 Structural models

The framework of structural models was introduced by Merton [135] in 1974. Unlike within estimation of PDs that rely on the companies' credit ratings, which are revised relatively infrequently, this approach is based on equity prices and, therefore, can provide more up-to-date information for default probabilities estimation.

The core concept of the Merton model is to treat company's equity and debt as a contingent claim written on company's asset value. Suppose, for simplicity, that a firm has one zero-coupon bond outstanding and that the bond matures at time $T$. Define value of company's assets today and at time $T$ as $V_0$ and $V_T$, respectively, value of company's equity today and at time $T$ as $E_0$ and $E_T$, respectively, amount of debt interest and principal due to be repaid at time $T$ as $D$, volatility of assets (assumed constant) as $\sigma_V$, and instantaneous volatility of equity as $\sigma_E$.

If $V_T < D$, it is rational for the company to default on the debt at time $T$. The value of the equity is then zero. If $V_T > D$, the company should make the debt repayment at time $T$ and the value of the equity at this time is $V_T - D$. The Merton model, therefore, gives the value of the firm's equity at time $T$ as

$$E_T = \max\left(V_T - D, 0\right). \tag{1.16}$$

In line with the Black-Scholes option pricing theory [22], the Merton model stipulates that the company's equity value satisfies the following equation for pricing the call option

$$E_0 = V_0\Phi\left(d_1\right) - De^{-rT}\Phi\left(d_2\right) \tag{1.17}$$

where

$$d_1 = \frac{\ln\left(V_0 / D\right) + \left(r + \sigma_V^2 / 2\right)T}{\sigma_V\sqrt{T}} \qquad d_2 = d_1 - \sigma_V\sqrt{T} \tag{1.18}$$

and $\Phi$ is the cumulative normal distribution function.

Under the Merton model, the company defaults when the option is not exercised. It can be shown that the probability of such event is $\Phi(-d_2)$. To calculate this, we require $V_0$ and $\sigma_V$, both not directly observable. However, if the company is publicly traded, we can observe $E_0$. This means that equation (1.17) provides one condition that must be satisfied by $V_0$ and $\sigma_V$. From a result in stochastic calculus known as Ito's lemma, we can also estimate $\sigma_E$

$$\sigma_E E_0 = \frac{\partial E}{\partial V} \sigma_V V_0 . \tag{1.19}$$

Here $\partial E / \partial V$ is the delta of the equity and is equal to $\Phi(d_1)$, so that

$$\sigma_E E_0 = \Phi(d_1) \sigma_V V_0 . \tag{1.20}$$

This provides another equation that must be satisfied by $V_0$ and $\sigma_V$. Equations (1.17) and (1.20) provide a pair of simultaneous equations that can be solved for $V_0$ and $\sigma_V$.

There are many extensions of the Merton's framework that try to overcome one or more of the unrealistic assumptions. For reviews covering structural models, see, e.g., Lando [111], Bielecki and Rutkowski [19], or Uhrig-Homburg [170]. We will discuss the Merton model and its distributional assumptions in more detail in Chapter 2.

### 1.3.3 Reduced-form models

Reduced-form models are another major class of models where, unlike structural models, defaults do not explicitly depend on the value of the firm. They are more general than structural models and assume that an exogenous random variable drives default and that the probability of default over any time interval is non-zero. An important input to determine the default probability and the price of a bond is the rating of the company. Thus, in order to determine the risk of a credit portfolio of rated issuers generally historical average defaults and transition probabilities for current rating classes have to be considered (Trück and Rachev [169]). Besides the fact that they allow for realistic short-term credit spreads, reduced-form models also give great flexibility in specifying the source of default.

Generally, reduced-form models allow for surprise defaults. They model the time of default as an exogenous process without making assumptions from some underlying asset. Therefore, the default event is modeled as more aggregated than in the structural models where the time of default depends on the firm value that may depend on or be correlated with other variables (Trück and Rachev [169]).

At the heart of these models lies the instantaneous rate of default, i.e. the default intensity $\lambda$. Let $\mathfrak{F}_t$ be the information up to time *t*, $\tau$ the default time, $\Delta t$ a marginally short time interval, and $\lambda(t)$ the default intensity as a function of time. Mathematically expressed is the default intensity (assuming no default up to time *t*)

$$\mathrm{PD}\left(\tau \in \left(t + \Delta t\right) | \mathfrak{F}_t\right) \approx \lambda\left(t\right)\Delta t \tag{1.21}$$

approximately the proportionality factor between the default probability within a given time interval $\Delta t$ and the length of this time interval. In other words, $\lambda$ is the intensity of the process that specifies the default time $\tau$. In the literature, often Poisson processes are used to model the default time as they suit to model rare and discretely countable events such as defaults. In this context, the time of default is interpreted as the first jump of the Poisson process. After default, the intensity is usually set equal to zero.

One important advantage of reduced-form models is that their framework is capable of reducing the technical difficulties of modeling defaultable claims to modeling the term-structure of non-defaultable bonds and related derivatives. Reduced-form models differ in their assumptions about the default time (indirectly the default intensity), the correlations between the risk-free interest rates and the default time and the way they model the recovery rate $\varphi$.

**The model of Fons (1994)**

Fons [72] was the first who developed a reduced-form model and derived credit spreads using historical default rates and recovery rate estimates. The approach is based on the results of Moody's corporate bond default studies, which at that time covered 473 defaults of issuers that ever held a Moody's corporate bond rating between January 1, 1970 and December 31,

1993. He found out that the term structure of credit risk, i.e. the behavior of credit spreads as maturity varies, seems to depend on the issuer's credit quality, i.e. its rating. For bonds rated investment grade, the term structures of credit risk have an upward sloping structure. On the other hand, speculative grade rated bonds behave in the opposite way: the term structures of the credit risk have a downward sloping structure. In every rating category, Fons compares term structures of credit spreads with weighted-average marginal default rates, using data from Moody's investigations. In his model, Fons assumes that investors are risk neutral. The risky bond price $B(0, T)$ with face value $B$ maturing at time $T$ supplied by Fons can be used to infer the credit spread on that bond by means of a formula which links the price of the bond to its yield to maturity. The price of a risky bond in $t = 0$ can be expressed in terms of its yield, with $r$ being the riskless yield and $s$ being the credit spread:

$$B(0, T) = B \cdot e^{-(r+s)T}, \tag{1.22}$$

whereas the price of a riskless security is

$$B'(0, T) = B \cdot e^{-rT}. \tag{1.23}$$

We denote $d_R(t)$ as the probability of default in year $t$ after the bond was assigned rating $R$, given that the bond has not defaulted before that date. Seen from date $t = 0$, $S_R(t)$ is the survival probability at date $t$. In the event of default the investor receives a fraction $\mu$ of par, the recovery rate. $S_R(t)$ is given by

$$S_R(t) = \prod_{j=1}^{t} \left(1 - d_R(j)\right), \tag{1.24}$$

whereas the probability that the bond rated $R$ will default in year $t$ is given by

$$D_R(t) = S_R(t-1) \cdot d_R(t) = \prod_{j=1}^{t-1} \left(1 - d_R(j)\right) \cdot d_R(t). \tag{1.25}$$

The expected value of the random flow $X_t$ received in $t$ is such that

$$E(X_t) = S_R(t-1) \cdot d_R(t) \cdot \mu \cdot B'(0, t). \tag{1.26}$$

The price of zero-coupon bond with initial rating $R$ maturing at $T$ is then the sum of the expected returns in each year:

$$\begin{aligned} B_R(0,T) &= \sum_{t=1}^{T} E(X_t) + S_R(T)\cdot B'(0,T) \\ &= \sum_{t=1}^{T} S_R(t-1)\cdot d_R(t)\cdot \mu\cdot B\cdot e^{-rt} + S_R(T)\cdot B\cdot e^{-rT}. \end{aligned} \tag{1.27}$$

Thus, with this formula we can compute the spread $s$ of the risky zero bond as follows:

$$s = -\frac{1}{T}\ln\left[\sum_{t=1}^{T} S_R(t-1)\cdot d_R(t)\cdot \mu\cdot e^{-r(t-T)} + S_R(T)\right]. \tag{1.28}$$

Fons determines the term structure of credit risk by calculating the spreads for zero bonds of every maturity $T$. Obviously, Fons' model also required an estimate of the recovery rate of a bond, which does usually not depend on the initial rating, but on its seniority and the bankruptcy laws of the issuer's home country. This model can be considered as one of the first reduced-form approaches to the modeling of credit spreads and default risk. Since then a variety of intensity models have been developed using ratings and corresponding default intensities as a starting point for the evaluation of credit risk.

**The model by Jarrow and Turnbull (1995)**

Jarrow and Turnbull [97] were the first ones to develop an intensity-based approach for valuation of risky debt. They propose three key assumptions for their model. First, there are no arbitrage opportunities and the market completeness. This is equivalent to the existence and uniqueness of an equivalent martingale measure $Q$ under which the discounted prices of the default-free and risky zero-coupon bonds are martingales. Second, there is a constant recovery-of-face value $\varphi$ that is given exogenously. And third, the authors assume the independence of the short-term spot interest rate $r(t)$ and the default process under the martingale measure $Q$.

Under these assumptions the price of a risky bond can be determined according to:

$$v(t,T) = E_t^Q\left(e^{-\int_t^T r(s)ds}\right)E_t^Q\left(\mathbf{1}_{(\tau > T)} + \varphi\mathbf{1}_{(\tau < T)}\right)$$
$$= p(t,T)E_t^Q\left(\mathbf{1}_{(\tau > T)} + \varphi\mathbf{1}_{(\tau < T)}\right). \tag{1.29}$$

Note that hereby it is implicitly assumed that the recovery payment is done at maturity. The equation would not change, however, if we assume a recovery payment at default. In that case, we would roll over the recovery payment $\varphi$ with the money market account until the maturity and then discount it again with the default-free zero bond. The price of the risky zero-coupon bond at time $t$ with maturity $T$ is equal to the expected payoff at maturity $T$ under the martingale measure discounted with the default-free zero-coupon bond with the same maturity. The equation can be further simplified to

$$v(t,T) = p(t,T)\left[\varphi + (1-\varphi)Q(\tau > T)\right]$$
$$= p(t,T)\varphi + p(t,T)(1-\varphi)Q(\tau > T) \tag{1.30}$$

with $Q(\tau > T)$ being the survival probability until maturity under the martingale measure. The first term on the right-hand side of the equation can be interpreted as the time $t$ value of the recovery rate that will be received surely at maturity. The second term is the time $t$ value of the promised payment if the zero-bond survives beyond the maturity.

**The model suggested by Madan and Unal (1998)**

Madan and Unal [123] decompose the risky debt into two securities: the *survival security* making the promised payments at maturity in case of survival and paying nothing otherwise; and the *default security* paying the recovery rate in default and nothing otherwise. Thus, different types of risk are addressed by different securities. While the survival security faces only the timing risk of default, the default security faces the recovery risk of default.

They are three assumptions underlying their model. First, the default payouts are independently and identically distributed across time and interest rate states. This implies the time-homogeneity of recovery rate $\varphi$. Second, default timing risks are functions of firm specific information that are independent of interest rate movements (further relaxation of Jarrow and Turnbull’s assumptions). Although the independence between short-term spot interest rate process and the default process remains, the default intensity is not constant anymore, but depends on the stock price of the firm. Thus, Madan and Unal build a bridge between the structural and the reduced-form models. And third, the recovery rate is referenced to an identical default-free zero-bond (recovery-of-treasury).

According to Madan and Unal [123], the firm‘s equity is a sign for the firm‘s financial strength and hence, changes in the equity levels will be reflected to the default probabilities. The authors use for their model the (by the money market account $B(t)=e^{-\int_0^t r(s)ds}$ discounted) equity value $s(t)$. The dynamics of the equity value is described by the following stochastic differential equation:

$$d\,s(t)=\sigma\, s(t)\,dW(t)$$

where $\sigma$ is the constant standard deviation of the equity value and $W$ is a standard Brownian motion.

Based on this assumption, the default intensity equals:

$$\lambda(s,t)=\frac{c}{\left(\ln\left(\frac{s}{s_{critic}}\right)\right)^2} \tag{1.31}$$

where $s_{critic}$ is the critical equity value and $c$ is constant parameter. The choice of such a function is, first, based on the requirement that equity value and default intensity should be inversely related. Second, if the exogenously given critical equity level is reached, the default probability goes to infinity, i.e. the firm defaults certainly.

After having set the foundation for the timing risk of default, which is only relevant for the survival security, Madan and Unal [123] model the recovery rate risk. The recovery rate $\varphi$ is random variable with a density function $q(\varphi)$. Thus, the expected payoff at default equals to:

$$E(\varphi)=\int_0^1 \varphi\, q(\varphi)\,d\varphi. \tag{1.32}$$

Based on the above models and under the assumption of the independence between the default intensity and the risk-free interest rate process, their fundamental equation for the value of the bond simplifies to:

$$\begin{aligned} v(t,T) &= E_t^Q\left(e^{-\int_t^T r(s)ds}\right)E_t^Q\left(\mathbf{1}_{(\tau > T)} + E(\varphi)\mathbf{1}_{(\tau < T)}\right) \\ &= p(t,T)E_t^Q\left(\mathbf{1}_{(\tau > T)} + E(\varphi)\mathbf{1}_{(\tau < T)}\right) \\ &= p(t,T)E(\varphi) + p(t,T)\left(1 - E(\varphi)\right)Q(\tau > T). \end{aligned} \tag{1.33}$$

**The model suggested by Lando (1998)**

The main feature of the approach suggested in Lando [112] is to model the default time using a Cox process. Hereby, it is assumed that the default intensity is a function of some state variable, the stochastic process $X(t)$ which may include riskless interest rates, stock prices, growth rate in the economy or other variables relevant to predict the likelihood of default. Thus, the state variable captures the correlation between the default time process and the interest rates, relaxing the key assumption made in the previous models. In Lando's model, the default time is the first jump time of the Cox process with intensity $\lambda\left(X(t)\right)$.

Assuming a recovery payment at maturity, Lando models it as

$$\varphi_T \mathbf{1}_{(\tau \le T)} = \varphi_T - \varphi_T \mathbf{1}_{(\tau > T)}. \tag{1.34}$$

Assuming a constant recovery rate, we obtain the following equation for the price of the risky bond:

$$\begin{aligned} v(t,T) &= E_t^Q\left(e^{-\int_t^T r(s)ds}\mathbf{1}_{(\tau > T)} + e^{-\int_t^T r(s)ds}\left(\varphi_T - \varphi_T \mathbf{1}_{(\tau > T)}\right)\right) \\ &= E_t^Q\left(e^{-\int_t^T r(s)ds}\mathbf{1}_{(\tau > T)}\right) + \varphi_T E_t^Q\left(e^{-\int_t^T r(s)ds} - e^{-\int_t^T r(s)ds}\mathbf{1}_{(\tau > T)}\right) \\ &= E_t^Q\left(e^{-\int_t^T r(s)ds}\mathbf{1}_{(\tau > T)}\right) + \varphi_T E_t^Q\left(e^{-\int_t^T r(s)ds}\right) - \varphi_T E_t^Q\left(e^{-\int_t^T r(s)ds}\mathbf{1}_{(\tau > T)}\right) \\ &= p(t,T)\varphi_T + (1 - \varphi_T)E_t^Q\left(e^{-\int_t^T r(s)ds}\mathbf{1}_{(\tau > T)}\right). \end{aligned} \tag{1.35}$$

Lando [112] further shows that the expectation on the right-hand side of the pricing equation can be expressed as:

$$E_t^Q\left(e^{-\int_t^T r(s)ds}\mathbf{1}_{(\tau > T)}\right) = E_t^Q\left(e^{-\int_t^T r(s)+\lambda(X(s))ds}\right). \tag{1.36}$$

That is the current value of the promised payment at maturity $T$, if there has been no default until $T$.

Overall, similar to the model suggested by Jarrow and Turnbull [97], the equation can be decomposed into two parts: a certain payment of the recovery rate and a promised payment in case of survival. While the certain payment is still the same, the promised payment additionally depends on the correlation between the interest rate and default processes. In the model of Jarrow and Turnbull [97], however, an interest rate change only changes the discounting factor of the promised payment, but not the default probabilities. Besides the value of a promised payment at maturity $T$, Lando [112] also derives equations for the value of a stream of payments (e.g. swaps), which terminates when default occurs, and for the resettlement payment at the time of default. For further details on the model, see Lando [112].

**The model of Duffie and Singleton (1999)**

Probably one of the most popular intensity based models goes back to Duffie and Singleton [54]. The special feature of their model is the recovery-of-market value assumption, i.e. the recovery rate is a fraction of the market value of the risky debt prior to default. Under this assumption, the authors construct an adjusted short-rate accounting for both the probability and the timing of default and the losses at default:

$$R(t) = r(t) + \lambda(t)(1-\varphi). \tag{1.37}$$

Given an exogenous default process and a recovery rate, the risky security can be valued as if it were default-free:

$$v(t,T) = E_t^Q\left(e^{-\int_t^T R(s)ds}\right). \quad (1.38)$$

As a special case of their model, Duffie and Singleton [54] also introduce some state variable $Y$, which both the short-term interest rate and the default processes are an exogenously given functions of. Hereby, the authors considers two cases for the state variable $Y$. The first one is that $Y$ is a continuous time Markov process under the martingale measure $Q$. The second approach considers a jump-diffusion process to allow sudden changes of $Y$. Also the case where the recovery rate and the default intensity depend in the current price of the risky security is discussed. Thus, the model is also able to incorporate the correlation between interest rates and default intensities. For further details on the framework, see Duffie and Singleton [54].

## 1.4 An overview of industry models

Several industry models for measuring credit portfolio risk have been developed in the 90's. Besides these commercial models there are various internal models employed in large international banks, which in most cases are more or less inspired by the well-known commercial products. For a comprehensive review of these models, see Crouhy, Galai and Mark [43]. For most of the models it is easy to find some technical documentation describing the mathematical framework of the model and giving some idea about the underlying data and the calibration of the model to the data. An exception is Moody's-KMV *PortfolioManager*$^{TM}$, where most of the documentation is proprietary or confidential.[4] However, even for this model the basic underlying idea can be explained without reference to non-public sources.

There are three types of credit portfolio models in use currently:

1. Structural (asset value) models – There are two vendor-supplied credit portfolio models of this type: Moody's-KMV **PortfolioManager**$^{\mathbf{TM}}$ released in 1993 and RiskMetrics Group's **CreditMetrics**$^{\mathbf{TM}}$ released in 1997.

[4] The model was originally developed by the KMV Corporation (Kealhofer, McQuown, Vasicek) founded in 1989. In 2002, the KMV Corporation was acquired by Moody's Analytics.

2. Macroeconomic (macro-factor) models – **CreditPortfolioView™** introduced by McKinsey and Company in 1998.

3. Actuarial (reduced-form) models – **CreditRisk+™** introduced by Credit Suisse First Boston in 1997.

### 1.4.1 CreditMetrics™

*CreditMetrics™* is one of the most well-known models for estimating credit risk on a portfolio of exposures (loans or bonds). It was originally introduced by the US bank J.P. Morgan. *CreditMetrics™* is a method for estimating the distribution of changes in the market value of a portfolio of credit exposures that may occur within a given risk horizon (generally one year). That distribution can be used to find the expected loss (EL) and various measurements of unexpected loss (UL) such as the standard deviation of losses, the percentiles and the associated Value-at-Risk.

It is a multinomial model, so it considers both the losses due to a default and those linked to migration of the obligor to a different rating class (in fact, this model is sometimes referred to as “migration approach”). Although it relies partly on the conceptual tools developed by Merton [135], it is a reduced-form model. Unlike the structural models, in fact, *CreditMetrics™* does not derive the probability of default (or migration) based on the characteristics of the company (market value and volatility of assets, value of debt), but simply uses as input historical data on default and migration rates by rating class.

It is assumed that all variables, except the current rating state of the issuer, behave deterministically over time. Thus, the value of the bond or loan at the risk time horizon $T$ is essentially dependent on the rating state $i$ of the issuer at this point of time. *CreditMetrics™* assumes that if the issuer is not in a state of default at the risk time horizon, the value of the bond or loan is determined by discounting the outstanding cash flows using credit spreads over the riskless interest rate $r$. The spreads correspond to the rating state $i$ of the issuer in $T$. The distribution of bond or loan values in $T$ is thus given by the probabilities $P(X = i)$ of the different rating states in $T$, together with the corresponding values of the bond $V_{i,T}$.

In the first stage of the model we determine the distribution of ratings of the exposure at the end of a given risk time horizon $t$. This is done with the help of a transition matrix $P$. Suppose that the initial rating of the exposure at time 0 is $i \in \{1, 2, \dots, K\}$. This initial setting can be represented by the vector $p_{i\cdot}(0) = \delta_i$. In the *CreditMetrics*[TM] framework in order to obtain the distribution of possible ratings at $t$ the initial rating vector is multiplied with a $t$-step transition matrix. If the risk horizon is more than one year it is suggested to compute the required vector of transition probabilities $p_{i\cdot}(t)$ either with a multiple of a one-year transition matrix $P$, thus, $p_{i\cdot}(t) = \delta_i \cdot P^t$ or, if available, with a directly estimated $t$-year transition matrix $p_{i\cdot}(t) = \delta_i \cdot P(t)$. Thus, we obtain all possible future ratings at time $t$ and the corresponding transition probabilities.

In a second step a risk-adjusted forward price is derived for each rating state. The case of default and non-default states are considered separately. The remaining cash flows from $t$ to $T$ in non-default categories are discounted with state specific forward rates. The forward zero curve for each rating category can be found by calibrating forward rates to observed credit spreads of different maturities.

In the case of non-default states agreed payments before $t$ will be fully received and can be added – including the earned interest until $t$ – to the risk-adjusted value of the bond at time $t$:

$$B_j(t, T) = \sum_{k=1}^{t} C_k \left(1 + f^*(k, t)\right)^{t-k} + \sum_{k=t+1}^{T} \frac{C_k}{\left(1 + f_j(t, k)\right)^{k-t}} + \frac{B}{\left(1 + f_j(t, T)\right)^{T-t}} \tag{1.39}$$

with $C_k$ denoting the nominal coupon in year $k$, $B$ the nominal principal, $f^*$ being the riskless forward rate, and $f_j$ the forward rate for $j$-rated bonds. In case that the bond defaults before $t$ a recovery payment is assigned:

$$B_K(t, T) = R \cdot \left( \sum_{k=1}^{T} C_k + B \right) \tag{1.40}$$

where $R$ is the expected fraction of the bond's nominal cash flows that is paid back. The parameter $R$ is estimated as the average return in prior default experience and depends on the seniority class of the bond. In *CreditMetrics™* the recovery rate is simulated by a beta distribution whose mean and standard deviation are calibrated in order to fit the parameters of the historically observed recovery rate that corresponds to the seniority of the item.

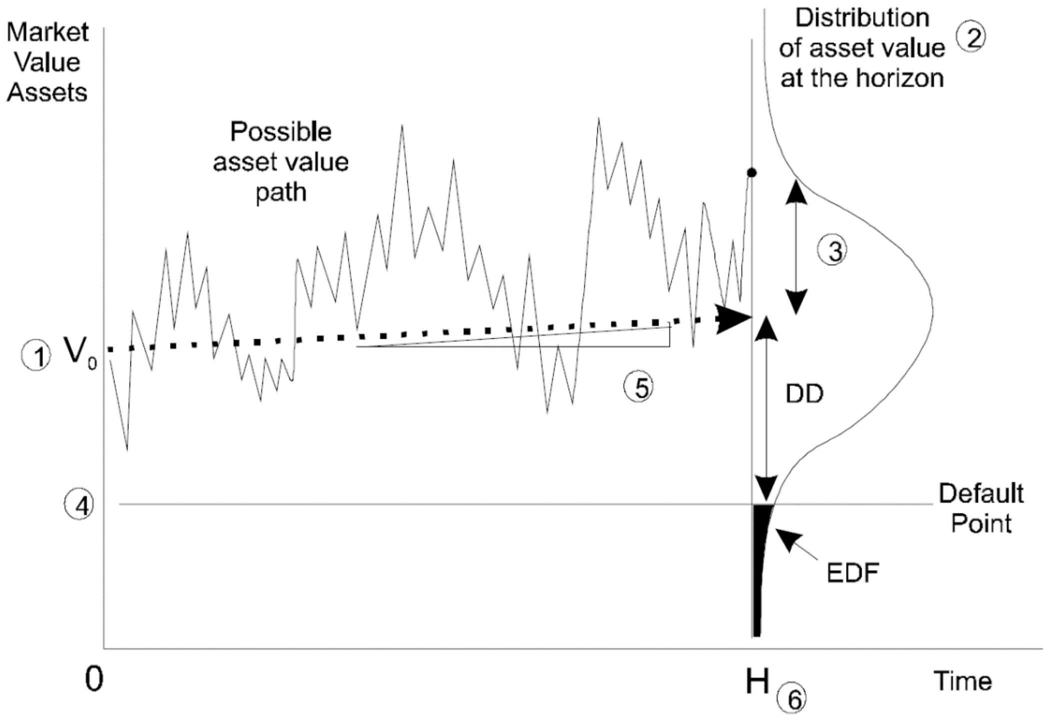
Regarding the bond price $B_j$ as a random variable, the mass distribution of this random variable is given by the vector $p_{i\cdot}$. Hence, the so-called "Distribution of Values" (DoV) for a given initial rating and the considered risk time horizon can be obtained by using adequate transition matrices and forward curves. Credit risk measures like the expected (EL) or unexpected (UL) loss can be derived from the DoV. The DoV gives for each predicted bond price the probability of being assigned to this rating.

Obviously, *CreditMetrics™* offers a quite different approach for measuring credit risk than the firm value models. The model provides a rather empirical Value-at-Risk approach for measuring credit risk that should be consistent with actual market prices. Besides, it is rather interested in potential losses during worst-case scenarios. In this framework historical transition matrices and forward prices are more important than the value of the firm. However, as it comes to deriving joint transition matrices for two or more individual companies, the company's asset value is considered as the key driver of rating changes. For measuring asset return correlations the issuers' equity returns from publicly available quotations are used.

### 1.4.2 PortfolioManager™

Moody's-KMV *PortfolioManager™* is based on Merton's insight that debt behaves like a short put option on the value of the firm's assets. With such a perspective, default will occur when the value of the firm's assets falls below the value of the firm's debt (or other fixed claims). There are six variables that determine the default probability of a firm over some horizon, from now until time $H$ (see Figure 1.1): 1. the current asset value, 2. the distribution of the asset value at time $H$, 3. the volatility of the future assets value at time $H$, 4. the level of the default point, the book value of the liabilities, 5. the expected rate of growth in the asset value over the horizon, 6. the length of the horizon $H$.

**Figure 1.1**
The Moody's-KMV model




Source: Crosbie and Bohn [42], page 13

Moody's-KMV *PortfolioManager*™ derives the Expected Default Frequency (EDF) for each firm based on firm's capital structure, the asset return value and its volatility using framework by Merton [135]. In this method, each value of the EDF can then be used to specify a credit rating. The default probabilities are derived in three steps. First, the value of the firm's assets is estimated based on a standard geometric Brownian motion as in the Merton's framework. Second, distance-to-default is computed. The distance-to-default is the number of standard deviations between the mean of the asset value and the default point where the default point is defined as the sum of the short-term debt liabilities and half of the long-term liabilities to be met over the risk horizon. The third and last step is to derive the default probabilities, EDFs, from the distance-to-default index. The probability of default is then the proportion of the firms of a given ranking of distance-to-default which actually defaulted over the risk horizon, usually one year. The EDFs can also be used as an indicator of the creditworthiness of the issuing firms.

Based on a sample of 100,000 companies, KMV showed that there would be a sharp increase in the slope of EDF prior to default of those firms that have defaulted or went bankrupt over a 20-year period. With this empirical evidence, each EDF index can be matched one-on-one to one of those conventional credit rating classes. While the lowest EDF corresponds to the highest credit rating, it increases as the credit rating goes down implying a negative relationship between the two (Onmus–Baykal [142]).

The basic idea behind default correlation in this model is that, for two obligors, the correlation between the values of their assets in combination with their individual default points will determine the probability that the two firms will default at the same time; and this joint probability of default can then be related to the default event correlation.

In the Moody's-KMV model, default correlation is computed in the *Global Correlation Model* (GCorr), which implements the asset-correlation approach via a factor model that generates correlated asset returns

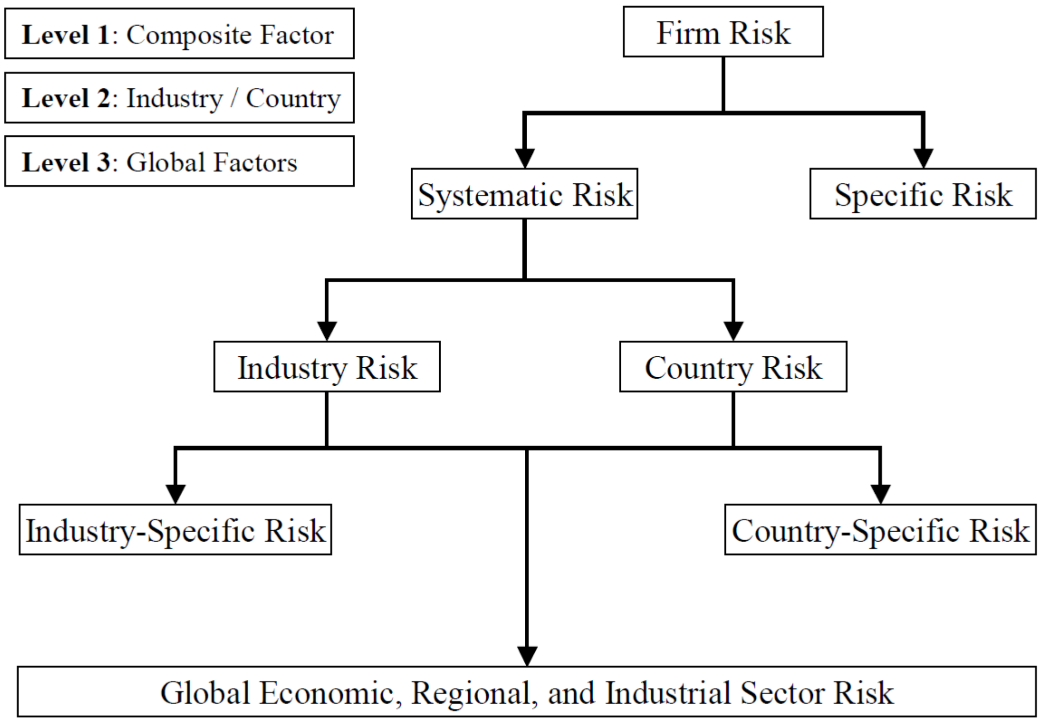
$$r_A(t) = \beta_A \, r_{CI,A}(t) \tag{1.41}$$

where $r_A(t)$ is the return of firm A's assets in period $t$, and $r_{CI,A}(t)$ is the return on a unique custom index (factor) for firm A in period $t$.

The custom index for each firm is constructed from industry and country factors (indices). The construction of the custom index for an individual firm proceeds as follows. First, the firm's assets and sales to the various industries in which it operates are allocated (from the 61 industries covered by the Moody's-KMV model). Second, the firm's assets and sales to the various countries in which it operates are allocated (from the 45 countries covered by the Moody's-KMV model). And third, the country and industry returns are combined. In the Moody's-KMV approach, the correlation of default events for two firms depends on the asset correlation for those firms and their individuals probabilities of default. In practise, this means that default correlations will be determined by the $R^2$ of the factor models and the EDFs of the individual companies.

To be more specific, the asset value return (AVR) of an individual company is mapped to a multi-factor model in three distinct phases (see Figure 1.2):

**Figure 1.2**
Three-level factor structure in the Moody's-KMV Global Correlation Model




Source: Bluhm, Overbeck and Wagner [23], page 43

(i) systematic and specific components are separated;

(ii) the systematic component is linked to several factors associated with various industries and countries;

(iii) the return of each factor associated with an industry/country is broken down, in turn, into a specific risk component (industry-specific risk and country-specific risk) and a systematic risk component (which depends, for example, on the exposure of the country/industry to global economic performance).

There are two key differences with respect to *CreditMetrics*™. First, the distribution of possible events in one year's time is not multinomial but binomial, it is therefore impossible to estimate the losses from downgrading, but we must focus on the losses related to default. Second, we need to know the spread the market requires of the obligor. In effect, if we wished, this spread (which we denote as $d$ ) could be readily determined by recalling that the market value of the credit must be the same for both risk-neutral investors and for risk-averse investors. The spread is then calculated as follows (see Sironi and Resti [159]):

$$d = \frac{(1+r)\,\mathrm{LGD}\cdot\mathrm{PD}^*}{1-\mathrm{PD}^*\cdot\mathrm{LGD}} \tag{1.42}$$

where $r$ is the risk-free rate, $\mathrm{PD}^*$ the risk neutral default probability, and $\mathrm{LGD}$ loss given default. This relationship is useful for the second variant offered by the Moody's-KMV – *CreditMonitor* which is designed for marking the portfolio to market. It involves constructing a certain number of discrete classes that group all the obligors with an EDF within a certain interval. A migration matrix must then be constructed for these EDF classes. Each class must also be associated with a credit spread based on its mean risk-neutral PD.

### 1.4.3 CreditPortfolioView$^{TM}$

*CreditPortfolioView$^{TM}$*, developed in 1997 by Tom Wilson, is based on the observation that credit cycles depend on the economic cycle. Therefore, during phases of economic growth the migrations toward higher rating classes (upgrades) tend to be more frequent, while migration rates toward lower classes (downgrades) and defaults decline. The opposite occurs during recessions. Thus, the transition matrices used in *CreditMetrics$^{TM}$* should be adjusted, depending on the current phase of the cycle. This approach therefore proposes to link the probabilities of migration and default to macroeconomic variables such as interest rate levels, the employment rate, real GDP growth and the savings rate, thus "conditioning them" to the state of the economic cycle.

Assume that the probability of default $\mathrm{PD}_{j,t}$ at time $t$ of a group or segment $j$ of companies reacts uniformly to changes in the economic cycle (generally companies in the same industry and same geographical area). *CreditPortfolioView$^{TM}$* assumes that this probability varies with the economic cycle; operationally, it is modelled according to a logit function:

$$\mathrm{PD}_{j,t} = \frac{1}{1+e^{-y_{j,t}}} \tag{1.43}$$

where $y_{j,t}$ represents the value at time $t$ of a "health index" of the segment $j$ based on macroeconomic factors. As index values rise, the default probability declines.

In turn, the index $y_{j,t}$ is a linier combination of several macroeconomic variables $x_{j,1}, x_{j,2}, \dots, x_{j,n}$ (the rate of real GDP growth, the employment rate, the level of long-term interest rates, the level of public spending, etc.):

$$y_{j,t} = \beta_{j,0} + \beta_{j,1}x_{j,1,t} + \beta_{j,2}x_{j,2,t} + \dots + \beta_{j,n}x_{j,n,t} + \upsilon_{j,t}\,. \tag{1.44}$$

The value of coefficients $\beta_{j,1}, \beta_{j,2}, \dots, \beta_{j,n}$ is estimated based on historical experience, analysing the data on past default frequencies. The last term $\upsilon_{j,t}$ represents random error (assumed to be independent from $x_{j,t}$ and characterized by a normal distribution with mean zero and volatility $\sigma_{\upsilon}$). While the terms linked to macroeconomic factors represent a systematic risk component (affecting several segments that can share the same macroeconomic factors), the random term identifies the specific risk component associated with segment $j$.

In order to use equations (1.43) and (1.44) as forecasting tools, we must produce an estimate of the future values of macroeconomic factors. To this end, for each factor *CreditPortfolioView™* uses a second-order auto-regressive model AR(2) like the following:

$$x_{j,i,t} = \gamma_{i,0} + \gamma_{i,1}x_{j,i,t-1} + \gamma_{i,2}x_{j,i,t-2} + \varepsilon_{j,i,t} \tag{1.45}$$

where coefficients $\gamma_{i,j}$ must be estimated empirically and $\varepsilon_{j,i,t}$ represents a normally distributed error term with mean zero.

The model is not limited to generating a projection of the conditional default probabilities of the various segments, but also uses them to condition the entire transition probability matrix. In fact, the mean long-term transition matrix (unconditional) is adjusted to reflect the expected default probabilities for the subsequent year. For closer discussion on estimating this conditional transition matrix, see, e.g., Sironi and Resti [159].

### 1.4.4 CreditRisk+™

*CreditRisk+™* was developed by Credit Suisse Financial Products in 1997. It applies to credit risk some instruments typical of the mathematics of insurance (actuarial mathematics). The losses of an insurance company derive from two fundamental variables: (i) the frequency with which a certain type of event occurs (event frequency) and (ii) the amount the company must pay out when the event occurs (loss severity). The idea is similar to credit risk, where the losses depend on the frequency of default events and the rate of loss given default. Based on this analogy, it is possible to use insurance-derived models for estimating credit losses.

Obviously, these models can focus only on default risk; migration risk is not considered. Moreover, exposures at default (EAD) and recovery rates are treated as deterministic variables. Therefore, neither exposure risk nor recovery risk can be estimated. Despite these limitations, *CreditRisk+TM* is highly effective in estimating the risk of portfolios with a large number of positions. It has therefore been applied extensively in the management of some traditional banking portfolios, such as loans to small and medium enterprises, consumer loans, and mortgages.

As mixture distribution this approach incorporates the gamma distribution which is defined by the probability density

$$\gamma_{\alpha,\beta}(x) = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} e^{-x/\beta} x^{\alpha-1} \qquad (x \geq 0),$$

where $\Gamma(\cdot)$ denotes the gamma function. The first and second moments of a gamma-distributed random variable $\Lambda$ are

$$\mathrm{E}[\Lambda] = \alpha\,\beta, \qquad \mathrm{V}[\Lambda] = \alpha\,\beta^2. \tag{1.46}$$

Instead of incorporating a factor model (as we have seen it in the case of *CreditMetricsTM* and *PortfolioManagerTM*), *CreditRisk+TM* implements a so-called sector model. However, somehow one can think of a sector as a “factor-inducing” entity, where every sector could be thought of as generated by a single underlying factor. In this way, sectors and factors are somehow comparable objects. From an interpretational point of view, sectors can be identified with industries, countries, or regions, or any other systematic influence on the economic performance of counterparties with a positive weight in this sector. Each sector

$s \in \{1, \dots, m_S\}$ has its own gamma-distributed random intensity $\Lambda^{(s)} \sim \Gamma(\alpha_s, \beta_s)$, where the variables $\Lambda^{(1)}, \dots, \Lambda^{(m_S)}$ are assumed to be independent.

Let us assume now that a credit portfolio of $m$ loans to $m$ different obligors is given. In the sector model of *CreditRisk+TM*, every obligor $i$ admits a breakdown into sector weights $w_{i,s} \geq 0$ with $\sum_{s=1}^{m_S} w_{i,s} = 1$, such that $w_{i,s}$ reflects the sensitivity of the default intensity of obligor $i$ to the systematic default risk arising from sector $s$. The risk of sector $s$ is captured by two parameters. The first driver is the mean default intensity of the sector

$$\lambda_{(s)} = \mathrm{E}\left[\Lambda^{(s)}\right] = \alpha_s \, \beta_s ;$$

the second driver is the default intensity's volatility

$$\sigma_{(s)} = \mathrm{V}\left[\Lambda^{(s)}\right] = \alpha_s \, \beta_s^2 .$$

Every obligor $i$ admits a random default intensity $\Lambda_i$ with mean value $\mathrm{E}\left[\Lambda_i\right] = \lambda_i$, which could be calibrated to the obligor's one-year default probability using following relationship:

$$\mathrm{PD}_i = \mathrm{P}\left[L_i' \geq 1\right] = 1 - e^{-\lambda_i} \approx \lambda_i \tag{1.47}$$

where $L_i'$ denotes a Poisson-distributed random variable. The sector parameterization of $\Lambda_i$ is as follows:

$$\Lambda_i = \sum_{s=1}^{m_S} w_{i,s} \, \lambda_i \, \frac{\Lambda^{(s)}}{\lambda_{(s)}} \quad \left(i = 1, \dots, m\right). \tag{1.48}$$

This shows that two obligors are correlated if and only if there is at least one sector such that both obligors have a positive sector weight with respect to this sector. Only in such cases two obligors admit a common source of systematic default risk. Note that equation (1.48) is consistent with the assumption $\lambda_i$ equals the expected default intensity of obligor $i$. The default risk of obligor $i$ is then modeled by a mixed Poisson random variable $L_i'$ with random intensity $\Lambda_i$.

Note that in accordance with equation (1.48) any conditional default intensity of obligor $i$ arising from realizations $\theta_1, \dots, \theta_{m_S}$ of the sector's default intensities $\Lambda^{(1)}, \dots, \Lambda^{(m_S)}$ generates a conditional one-year default probability $\mathrm{PD}_i\left(\theta_1, \dots, \theta_{m_S}\right)$ of obligor $i$ by setting

$$\begin{aligned} \mathrm{PD}_i\left(\theta_1, \dots, \theta_{m_S}\right) &= \mathrm{P}\left[L_i' \geq 1 \mid \Lambda_1 = \theta_1, \dots, \Lambda_{m_S} = \theta_{m_S}\right] \\ &= 1 - e^{-\lambda_i \sum_{s=1}^{m_S} w_{i,s}\, \theta_s / \lambda_{(s)}} . \end{aligned} \tag{1.49}$$

Let $L'$ denote the random variable representing the number of defaults in the portfolio. We already mentioned that *CreditRisk+™* is a Poisson mixture model. More explicitly, it is assumed that $L'$ is a Poisson variable with random intensity $\Lambda^{(1)}, \dots, \Lambda^{(m_S)}$. Additionally, it is naturally required to obtain the portfolio's defaults as the sum of single obligor defaults, and indeed equation (1.48) is obviously consistent with $L' = L_1' + \dots + L_m'$ when defining the sector's mean intensity by

$$\lambda_{(s)} = \sum_{i=1}^{m} w_{i,s}\, \lambda_i .$$

On the portfolio level, the "trick" *CreditRisk+™* uses in order to obtain a nice closed-form distribution of portfolio defaults is sector analysis. Given that we know distribution of defaults in every single sector, the portfolio's default distribution then just turns out to be the convolution of the sector distributions due to the independence of the sector variables $\Lambda^{(1)}, \dots, \Lambda^{(m_S)}$. So we only have to find the sector's default distributions.

When focusing on single sectors, it is a standard result from elementary statistics (see, e.g., Rice [149]) that any gamma-mixed Poisson distribution follows a negative binomial distribution. Therefore, every sector has its own individually parameterized negative binomial distribution of sector defaults, such that the portfolio's default distribution indeed can be obtained as a convolution of negative binomial distributions. As a consequence, the generating function of the portfolio loss can be explicitly written in a closed form. For formula and discussion on loss distribution, see Bluhm, Overbeck and Wagner [23].

## 1.5 Thesis chapters overview

This dissertation thesis covers several topics in credit risk management and comprises three various studies conducted in Chapters 2, 3, and 4, that are closely related to estimation and examination of default probabilities.

In our first study (Chapter 2)[5], we discuss structural models based on the Merton's framework. First, we observe that the classical distributional assumption of the Merton [135] model (company value follows the log-normal distribution) is generally rejected. Second, we implement a structural credit risk model based on stable non-Gaussian processes as a representative of subordinated models in order to overcome some drawbacks of the Merton one. In particular, we propose to use Hurst, Platen and Rachev [93] option pricing model based on the stable Paretian distributions which generalizes the standard Merton's methodology. Finally, following the Moody's KMV estimation methodology, we propose an empirical comparison between the results obtained from the classical Merton model and the stable Paretian one. In particular, we suggest alternative parameter estimation for subordinated processes, and we optimize the performance for the stable Paretian model. Our results suggest that PD is generally underestimated by the Merton model and that the stable Lévy model is substantially more sensitive to the periods of financial crises.

Structural models are not plausible for the estimation of PDs of banks, unless some adjustments are made, since financial institutions have significantly greater debt compared to corporates. Therefore, the second study (Chapter 3)[6] employs rating-based models applied to financial institutions. In fact, this chapter is devoted to examination of the performance of static and multi-period credit-scoring models for determining PDs of financial institutions. Academic research linked to the performance of rating models for financial institutions is rather limited as most studies mainly focus on corporates and, due to their different balance sheet structure, often exclude financial institutions from their sample. However, the importance of assessing the default risk of financial institutions has become even more obvious since the recent period of financial and economic turmoil during the financial crisis.

[5] Chapter 2 is based on the paper by Gurny, Ortobelli Lozza and Giacometti [80] which has been published in *Journal of Applied Mathematics*.

[6] Chapter 3 is based on a working paper by Gurny, Kalotay and Trück [79] which is intended to be submitted either to *Contemporary Accounting Research* or *Omega* at the beginning of 2016.

We use a unique database for the U.S. provided by the Federal Financial Institutions Examination Council (FFIEC). Our extensive sample contains more than seven thousand U.S. commercial banks with over four hundred default events during the period 2007-2013. Our analysis also focuses on evaluating the performance of the considered scoring techniques. We apply a substantial number of model evaluation methods, including techniques that have not yet been applied in the literature on credit scoring. We also provide an overall ranking of the models according to the different evaluation criteria and find that the considered scoring models provide a high predictive accuracy in distinguishing between default and non-default financial institutions. Despite the difficulty of predicting defaults in the financial sector as it has been mentioned in the literature, the proposed models perform very well also in comparison to results on scoring techniques for the corporate sector.

Finally, in our third study (Chapter 4)[7], we include credit risk topic in asset pricing framework. In particular, we investigate the question whether distressed renewable energy companies earn on average higher returns than low distress risk companies. Using the Expected Default Frequency (EDF) measure obtained from Moody's KMV, we demonstrate that there is a positive cross-sectional relationship between returns of both, equally-weighted (EW) and value-weighted (VW) portfolios, and evidence for a distress risk premium in the U.S. renewable energy sector. The positively priced distress premium is also confirmed by investigating returns corrected for common Fama and French [65] and Carhart [37] risk factors. We further show that raw and risk-adjusted returns of value-weighted portfolios that take a long position in the 20% most distressed stocks and a short position in the 20% safest stocks generally outperform the S&P 500 index throughout our sample period (2002–2014).

Chapter 5 then concludes and summarizes the results.

[7] Chapter 4 is based on a working paper by Gurny and Trück [81] which is intended to be submitted to *Energy Economics* by the end of 2015.

# Chapter 2

# Structural Credit Risk Models with Subordinated Processes

The structural approach to credit risk modeling was proposed in 1974 by Robert Merton in his seminal paper on the valuation of corporate debt [135]. Largely as a logical extension of the Black and Scholes [22] option pricing framework, he introduced a model for assessing the credit risk of a company by characterizing a company's equity as a derivative on its assets.

The Merton model requires a number of simplifying assumptions (the company can default only at debt's maturity time $T$ but not before, the model is not able to distinguish among the different types of debt, constant and flat term structure of interest rates, etc.). Notwithstanding, one of the most important drawbacks is an assumption that company value follows the log-normal distribution. It is well known that log-returns of equities are not Gaussian distributed, and several empirical investigations have shown that log-returns of equities present skew distributions with excess kurtosis which leads to a greater density in the tails, and that the normal distribution with a comparatively thinner tail simply cannot describe this phenomenon (see, e.g., Mandelbrot [126-128], Fama [62-64], or Rachev and Mittnik [147]).

The main contribution of this study is twofold. First, we introduce a structural credit risk model based on the stable Paretian distributions as a representative of subordinated models. Secondly, we show that it is possible to use this model in the Merton's framework, and we propose an empirical comparison of the Moody's KMV methodology applied to the Merton model and our subordinated one. In particular, we prove that the basic assumption of the

Merton model is generally rejected, and consequently the log-returns of the companies' asset values are not Gaussian distributed. For this reason, we discuss the possibility for using other subordinated processes to approximate the behaviour of the log-returns of the company value. Thus, we propose to use the Hurst, Platen and Rachev [93] option pricing model based on the stable Paretian distributions which generalizes the standard Merton methodology.

The practical and theoretical appeal of the stable non-Gaussian approach is given by its attractive properties that are almost the same as the normal ones. As a matter of fact, the Gaussian law is a particular stable Paretian one, and thus the stable Paretian model is a generalization of the Merton one. The first relevant desirable property of the stable distributional assumption is that stable distributions have domain of attraction. The generalized central limit theorem for the normalized sums of i.i.d. random variables determines the domain of attraction of each stable law. Therefore, any distribution in the domain of attraction of a specified stable distribution will have properties close to those of the stable distribution. Another attractive aspect of the stable Paretian assumption is the stability property; that is, stable distributions are stable with respect to summation of i.i.d. random stable variables. Hence, the stability governs the main properties of the underlying distribution. In addition, in the empirical financial literature, it is well documented that the asset returns have a distribution whose tail is heavier than that of the distributions with finite variance.

The idea of using subordinated stable Paretian processes goes back to the seminal work of Mandelbrot and Taylor [129]. Stable laws have been applied in several financial sectors (see Rachev [146] and Rachev and Mittnik [147]). For these reasons, the stable Paretian law is the first candidate as a subordinated model investigating for credit risk modeling, and in this study we discuss how to use the Hurst, Platen and Rachev [93] stable subordinated model in the framework of structural credit risk models. In particular, as for the Merton model, we propose two different methodologies for the parameter estimation: the first is to generalize the maximum likelihood parameter estimation proposed by Duan [49]; the second is a generalization of the Moody's KMV methodology.

This chapter is organized as follows. In Section 2.1, we firstly provide literature review on structural credit risk models. In Section 2.2, we review the theory and the distributional assumptions of the Merton model. Subsequently, we introduce the credit risk models with subordinated processes and describe the Mandelbrot-Taylor distributional assumptions. Section 2.3 is devoted to the parameters estimation for both the Merton and the subordinated models. We characterize empirical data and make a comparison between the obtained results in Section 2.4. Finally, we provide a brief summary in Section 2.5.

## 2.1 Literature review

The first generation structural credit risk models are based on Merton [135] model. In this approach, the company's default depends on the value of the company's assets. A firm will default when its market value is lower than the value of its liabilities. The payment to the debt holders at the maturity of debt is therefore the smaller of the face value of the debt or the market value of the firm's assets. Following this basic intuition, Merton derives a formula for risky bonds to estimate the probability of default of a firm and the yield gap between a risky bond and default-free bond. In addition to Merton [135], models by Black and Cox [21], Geske [75], and Vasicek [171] might be classified in the first generation structural credit risk models. These models try to improve the original Merton's framework by relaxing one or more of the unrealistic assumptions (Laajimi [109]).

Black and Cox [21] first describe some solution methods to be applied when the problem of valuation of contingent claims is discrete in time. They then examine the effects of safety covenants, subordination arrangements, and restrictions on the financing of interest and dividend payments on the value of the security. They find that in theory these provisions may have significant effects on the behavior of the firm's securities and may increase the value of the risky bonds.

Geske [75] modifies the original Merton's framework by allowing the risky bond to have discrete interest payments. Although, Black and Cox [21] looks at a similar problem, in their case, the interest payments are continuous in time and state that in general, there is no closed form solution when the interest payments are discrete in time. However, Geske [75] derives a general valuation equation for a risky coupon bond with an arbitrary number of discrete coupon payments and a principal payment using the compound option technique. He also discusses the effects of safety covenants, subordinated debt, and payout financing restrictions in the compound option case. In particular, the general valuation equation developed using the compound option technique is applied to the subordinated debt.

In addition to study by Geske [75], Vasicek [171] discusses the distinction between the long-term and short-term liabilities in valuing credit risk. However, the valuation of debt becomes more complicated when one considers a debt structure by priority and by term. When all debt matures at the same time, the senior bondholders need not to be concerned about any junior debt. Because, in this case, the senior bondholder faces a loss only if the firm's higher priority liabilities are greater than the firm's assets. However, if the maturity dates for the firm's debt differ, the lender should not only be concerned about his claim but also other claims on the firm's asset that mature earlier even if they are junior debt. He further points out that the size of the expected loss will depend on the market value of the firm's assets and that of its total maturing debt and higher priority debt. Moreover, Vasicek [171] states that the long-term debt is as good as the firm's capital. After describing the effects of debt structure by term on the probability of default and the expected loss, he gives a method to find the price of a short-term loan. He argues that the price of a short-term loan can be calculated by the difference between the loan face value and the expected loss discounted at the risk-free interest rate.

The second generation structural credit risk models then assume that a firm may default any time between the issuance and maturity of the debt, which relaxes another of the Merton's assumptions, and specify a stochastic process for the evolution of the short-term rates (Laajimi [109]). In this scenario, the default may occur whenever the market value of the firm goes below a lower limit determined by the lender and borrower in the debt contract. The second generation structural-form models include Kim, Ramaswamy and Sundaresan [102] and Longstaff and Schwartz [120].

Kim, Ramaswamy and Sundaresan [102] show that conventional contingent claims models are unsuccessful in generating the credit spreads observed empirically even when excessive debt ratios and high level business risk parameters are used in numerical simulations. Due to this finding, they modify the conventional contingent claims model in two directions. First, they allow the bankruptcy to occur anytime between the issuance and maturity of the bond. In particular, the issuing firm may default on its coupon payment obligations any time. Second, they relax the flat risk-free rate assumption by specifying a stochastic process for the evolution of the short rate. They also introduce in their study the call features to examine its effect in the yield spreads between corporate and Treasury bonds.

Longstaff and Schwartz [120] then modify the first generation models in three directions: (i) default can arise anytime between the issuance and the maturity of the bonds; (ii) interest rates are not flat, i.e. there exists interest rate risk; (iii) strict absolute priority is violated. In contrast to Kim, Ramaswamy and Sundaresan [102], this paper derives a closed form solution to the valuation equation of risky fixed-rate and floating-rate coupons in a model with complex capital structure. In an application of their model to value risky discount and coupon bonds, they show that credit spreads produced by the model are comparable in magnitude to actual spreads. Furthermore, the model implies that credit spreads may differ among the firms with same default risk. The main reason for this is that the value of these firms' assets may have a different degree of correlation with interest rates. This implication of the model is helpful in explaining the observed differences in credit spreads among the similar rated bonds across various industries.

There are many other extensions of the Merton's framework. Ju, Parrino, Poteshman and Weisbach [100] consider a dynamic model of optimal capital structure where the firm financing decision is determined by a balancing between corporate taxes advantage and bankruptcy costs (trade-off theory). Collin-Dufresne, Goldstein and Martin [41] also consider a dynamic capital structure by modeling a mean-reverting leverage ratio and stochastic interest rate. Acharya and Carpenter [1] develop a model with both stochastic interest rate and endogenous defaults. The interest rate is modeled as one-factor diffusion process and the issuer follows optimal call and default rules. Thus, they bridge the gap between endogenous default and stochastic interest rate literatures.

Hackbarth, Hennessy and Leland [82] distinguish between bank and public debt. They assume that renegotiation through private workout is only possible for bank's debt. This renegotiation possibility makes bank's debt more attractive, but limits bank's debt capacity for strong firms, e.g. firms with high bargaining power. Bourgeon and Dionne [26] extend the Hackbarth, Hennessy and Leland [82] model to allow banks to adopt a mixed strategy in which renegotiation is sometimes refused ex-post in order to raise debt capacity ex-ante. Carey and Gordy [36] suppose that holders of private debt, e.g. banks, with strong covenants control the choice of the bankruptcy threshold. Since the private debt is senior, the bank triggers bankruptcy only when the asset's value falls below the face value of the bank's debt. In accordance with their model, they find empirical evidence indicating that the recovery rate is sensitive to debt composition.

Other extensions include Mauer and Triantis [133], Childs, Mauer and Ott [40], and Sundaresan and Wang [165], who consider endogenous investment. The cash holding management policy is accounted for in Acharya, Huang, Subrahmanyam and Sundaram [2], Anderson and Carverhill [8], and Asvanunt, Broadie and Sundaresan [10]. Sarkar and Zapatero [155] consider mean reverting cash flows. Zhou [179], Duffie and Lando [51] and Giesecke and Goldberg [76] add a jump component to the value process of assets allowing for "surprise" default at the cost of closed-form solution. Alternatively, Hackbarth, Miao and Morellec [83] consider jumps in the cash flow process with regime change. Finally, Longstaff [119], Morellec [137], and Ericsson and Renault [58] include a liquidity premia to price corporate debt, while Duffie and Lando [51] consider accounting information uncertainty.

As mentioned in Section 1.4.2, the Merton's framework is the underlying idea behind a commercial model developed by Moody's KMV. Distance-to-default, which is the normalized distance measured in standard deviations, of a firm's asset value from its default threshold plays a central role in calculating the Expected Default Frequency (EDF) in this model (Laajimi [109]). Sobehart, Keenan and Stein [160] and Stein [163], among other studies, examine the accuracy of the Moody's KMV model. Both studies find the Moody's KMV model to be incomplete. Kealhofer and Kurbat [101] find opposite results, namely that the Moody's KMV model captures all the information contained in agency ratings migration and accounting ratios. Crosbie and Bohn [42] find that combining market prices and financial statements gives more effective default measurement. The authors empirically test the EDF, derived from the KMV methodology, versus the credit rating analysis, and show that the EDF obtains a better power curve.

The accuracy of default forecasting of the Moody's KMV model is studied in Bharath and Shumway [18]. The authors compare the accuracy of this model with simpler alternative. They find that implied default probabilities from credit default swaps and corporate bond yield spreads are only weakly correlated with Moody's KMV default probabilities. The authors conclude that this model does not provide a sufficient statistic for default, which can be obtained using relatively naïve hazard models. Hillegeist, Keating, Cram and Lundstedt [87] and Du and Suo [48] compare the Moody's KMV model to other models and conclude that it does not provide adequate predictive power. However, Duffie, Saita and Wang [52] discover a significant predictive strength over time within the Moody's KMV model.

## 2.2 Merton and subordinated credit risk models

The core concept of the Merton [135] model is to treat company's equity and debt as a contingent claim written on company's asset value. In this framework, the company is considered to have a very simple capital structure. It is assumed that the company is financed by one type of equity with a market value $E_t$ at time $t$ and a zero-coupon debt instrument at $t$ $(D_t)$ with a face value of $L$ maturing at time $T$ [8]. The exercise price of a call option is defined as the value $L$. Let $A_t$ be the company's asset value at time $t$. Naturally, the following accounting identity holds for every time point:

$$A_t = E_t + D_t. \tag{2.1}$$

In the Merton framework the value of company's equity at maturity time $T$ is given by

$$E_T = \max\left[A_T - L, 0\right]. \tag{2.2}$$

### 2.2.1 The Merton-Black-Scholes distributional assumptions

Under the Merton model, the assets value is assumed to follow a geometric Brownian motion (GBM) in the following form:

$$dA_t = \mu A_t dt + \sigma A_t dW_t, \tag{2.3}$$

where $\mu$ is the expected return (drift coefficient), $\sigma$ is the volatility (diffusion coefficient), both unobserved, and $W_t$ is the normal variable $N(0,1)$. Using Ito's lemma, we can obtain the solution of (2.3) as follows:

[8] Generally, in a credit risk models framework we assume one-year time horizon for debt maturity and subsequent estimation of PD. One year is perceived as being of sufficient length for a bank to raise additional capital on account of increase in portfolio credit risk (if any).

$$A_T = A_t \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)(T-t) + \sigma\sqrt{(T-t)}W_t\right], \tag{2.4}$$

where $(T-t)$ is a remaining maturity.

In accordance with the Black and Scholes [22] option pricing theory, the Merton model stipulates that the company's equity value satisfies the following equation for pricing the call option within a risk neutral framework:

$$E_t = A_t\Phi(d_1) - Le^{-r(T-t)}\Phi(d_2), \tag{2.5}$$

where

$$d_1 = \frac{\ln\left(\frac{A_t}{L}\right) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{(T-t)}}, \tag{2.6}$$

$$d_2 = d_1 - \sigma\sqrt{(T-t)}, \tag{2.7}$$

$r$ is the risk-free interest rate[9] and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal variable. Equation (2.7) is referred to as the *distance-to-default* (DD) by Moody's KMV. The larger the number in DD is, the less chance the company will default.

We can estimate PD by rearranging (2.4) as follows:

$$\begin{aligned} \mathrm{PD}_t &= \mathrm{P}\left[A_T \le L\right] \\ &= \mathrm{P}\left[\ln(A_t) + \left(\mu - \frac{1}{2}\sigma^2\right)(T-t) + \sigma\sqrt{(T-t)}W_t \le \ln(L)\right] \\ &= \int_{-\infty}^{-\frac{\ln\left(\frac{A_t}{L}\right) + \left(\mu - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{(T-t)}}} \phi(x)\,dx, \end{aligned} \tag{2.8}$$

[9] The Treasury bill yields are commonly used as the risk-free interest rate $r$. Their rates are considered an important benchmark because treasury securities are backed by the full faith and credit of the U.S. Treasury. Therefore, they represent the rate at which investment is considered risk-free.

where $\phi$ is the probability density function of a standard normal variable. Note that unlike (2.8), (2.5) is not a function of $\mu$, but it is a function of $r$ (we would get PD under the risk neutral probability measure). When we estimate PD, the risk-free interest rate $r$ has to be replaced with real company drift $\mu$ since this step has nothing to do with option pricing. Thereby, the default probability of the company under the objective probability measure is given by

$$\begin{aligned} \text{PD}_t &= \Phi\left(-\hat{d}_2\right) \\ &= \Phi\left(-\frac{\ln\left(\frac{A_t}{L}\right)+\left(\mu-\frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{(T-t)}}\right). \end{aligned} \tag{2.9}$$

Further discussion on this topic can be found in Delianedis and Geske [45] who showed that risk neutral PDs can serve as an upper bound to objective PDs.

### 2.2.2 Credit risk models with subordinated assumptions

Using subordinated processes, we are usually able to capture empirically observed anomalies which are presented in the evolution of return processes over time. That is, we substitute the physical (calendar) time with a so-called intrinsic (operational) time which provides distribution tail effects often observed in the market, see Hurst, Platen and Rachev [93] and Rachev and Mittnik [147]. Thus, if $W=\{W(t), t\geq 0\}$ is a stochastic process and $T=\{T(t), t\geq 0\}$ is a non-negative stochastic process defined on the same probability space and adapted to the same filtration, a new process $Z=\{Z(t)=W(T(t)), t\geq 0\}$ may be formed, and it is defined as subordinated to $W$ by the intrinsic time process $T$. Next, we will suppose that $W$ is a standard Brownian motion. In this case, if the intrinsic time process $T$ is the deterministic physical time, that is, $T(t)=t$, we obtain the classical lognormal model (see Osborne [144]). Typically, subordinated models with random intrinsic time are leptokurtic with heavier tails compared to the normal distribution. Feller [67] showed that if the intrinsic time process has non-negative stationary independent increments, then the subordinated process $Z$ also has stationary independent increments.

Generally, we assume frictionless markets, where the log-price process $Z$ is subordinated to a standard Brownian motion $W$ by the independent intrinsic time process $T$. Therefore, we model the assets price process $A_t$ (the company's asset value in our case) by using a stochastic equation as follows:

$$A(t) = A(t_0)\exp\left\{\int_{t_0}^{t}\mu(s)ds + \int_{t_0}^{t}\rho(s)dT(s) + \int_{t_0}^{t}\sigma(s)dW(T(s))\right\}, \tag{2.10}$$

where the drift in the physical time scale $\mu(s)$, the drift in the intrinsic time scale $\rho(s)$, and the volatility $\sigma(s)$ are generally assumed to be constant. The appeal of processes subordinated to a standard Brownian motion $W$ by an intrinsic time process $T$ with non-negative stationary independent increments is also due to the option pricing formula which follows from the classical Black-Scholes one in a frictionless complete market and a risk-minimizing strategy in incomplete markets.[10] Hurst, Platen and Rachev [93] stable subordinated model uses the unique continuous martingale that makes sense in a discrete setting, but a priori it is not derived from a risk-minimizing strategy even if the markets are incomplete (see Rachev and Mittnik [147]). Following the same notation as in the Merton's framework, the value of a European call option at time $t$ (the value of company's equity) with exercise price $L$ (face value of a zero coupon debt instrument) and time to maturity $t$[11] is given by

$$E_t = A(t_0)\,F_+\left(\ln\left(\frac{A(t_0)}{L_{r,t_0,t}}\right)\right) - L_{r,t_0,t}\,F_-\left(\ln\left(\frac{A(t_0)}{L_{r,t_0,t}}\right)\right), \tag{2.11}$$

where

$$F_\pm(x) = \int_0^{+\infty}\Phi\left(\frac{x \pm \frac{1}{2}y}{\sqrt{y}}\right)dF_Y(y), \tag{2.12}$$

[10] In incomplete markets, there exist non-redundant claims carrying an intrinsic risk. In order to evaluate a contingent claim, a risk-minimizing strategy is often applied (see Hofmann, Platen and Schweizer [89], Follmer and Sondermann [71], and Follmer and Schweizer [70]).

[11] Here, we change the notation of maturity time from $T$ (used in the Merton's framework) to $t$ since $T$ denotes the intrinsic time process in the subordinated option pricing models.

$\Phi(\cdot)$ is the cumulative distribution function of the standard normal variable, $F_Y$ is the cumulative distribution function of a random variable $Y = \int_{t_0}^{t} \sigma^2(s)dT(s)$, and $L_{r,t_0,t} = L\exp\left(-\int_{t_0}^{t} r(s)ds\right)$ is the discounted exercise price (the right continuous with left-hand limits (RCLL) time-dependent function $r(t)$ defines the short term interest rate). Considering a continuous distribution of the random variable $Y$ with density function $f_Y$, $F_{\pm}(x)$ can now be numerically integrated over the finite interval $[0,1]$ using the transformation $y = u(1-u)^{-3}$ (see Rachev and Mittnik [147]); that is,

$$\begin{aligned} F_{\pm}(x) &= \int_0^{+\infty} \Phi\left(\frac{x \pm \frac{1}{2}\lambda y}{\sqrt{\lambda y}}\right) f_Y(y)dy \\ &= \int_0^1 \Phi\left(\frac{x \pm \frac{1}{2}\lambda u(1-u)^{-3}}{\sqrt{\lambda u(1-u)^{-3}}}\right) f_Y\left(u(1-u)^{-3}\right)\frac{1+2u}{(1-u)^4}du. \end{aligned} \tag{2.13}$$

Moreover, as for the classical Black-Scholes model, in the case of subordinated models, we can also monitor the variation in the derivative price with respect to the parameters that enter into the option formula (the Greeks). For our purposes, it is sufficient to define delta, which is given by

$$delta = \Delta_E = \frac{\partial E_t}{\partial A} = F_+\left(\ln\left(\frac{A(t_0)}{L_{r,t_0,t}}\right)\right). \tag{2.14}$$

Analogously to the Merton model, the probability of default can be estimated under the risk neutral probability measure as follows:

$$\mathrm{PD}_t = F_+\left(\ln\left(\frac{L_{r,t_0,t}}{A(t_0)}\right)\right) = \int_0^{+\infty} \Phi\left(\frac{\ln\left(\frac{L_{r,t_0,t}}{A(t_0)}\right)+\frac{1}{2}y}{\sqrt{y}}\right) dF_Y(y). \tag{2.15}$$

Recall that under the risk neutral measure the stationary increment $Z(t+s)-Z(t)$ has mean $\mu_{Z,s}=0$ and variance $\sigma^2_{Z,s}=\mu_{T,s}\sigma^2$, where $\sigma$ and $\mu_{T,s}$ are, respectively, the volatility and the mean of the increment of the stationary process $T$ when they exist (see Hurst, Platen and Rachev [93]). The skewness coefficient of this increment is zero (models are symmetric around the zero mean). Kurtosis of the subordinated models is defined as $k_{Z,s}=3\frac{1+\sigma^2_{T,s}}{\mu_{T,s}}$ for all $s \ge 0$ (where $\sigma^2_{T,s}$ is the variance of the random variable $T(t+s)-T(t)$ when it exists); that is, subordinated models with intrinsic random time are leptokurtic. Thereby, the model we consider in the following presents heavier tails and higher peaks around the origin compared to normal distribution.

### 2.2.3 The Mandelbrot-Taylor distributional assumptions

Mandelbrot [126-128] and Mandelbrot and Taylor [129] have proposed the stable Paretian distribution to estimate the log-returns. An $\alpha$-stable distribution $S_\alpha=(\sigma,\beta,\mu)$ depends on four parameters: the index of stability $\alpha\in(0,2]$ ($\alpha=2$ in the Gaussian case), the skewness parameter $\beta\in[-1,1]$, the scale parameter $\sigma\in(0,+\infty)$, and the location parameter $\mu\in(-\infty,+\infty)$, see Samorodnitsky and Taqqu [153] for further details on stable distributions. Mandelbrot and Taylor [129] supposed that the intrinsic time process $T$ has stationary independent increments as follows:

$$T(t+s)-T(t) \overset{d}{=} S_{\alpha/2}\left(cs^{2/\alpha},1,0\right), \tag{2.16}$$

for all $s,t\ge 0$, $\alpha\in(0,2)$, and $c>0$. Here, the index of stability is $\alpha/2$; the scale parameter is $cs^{\alpha/2}$; the stable skewness is 1; and the location parameter is zero. Under the Mandelbrot-

Taylor assumptions, the subordinated process $Z(t)=\ln\left(A_{th}\right)$, is a symmetric $\alpha$-stable Lévy motion with stationary independent increments as follows:

$$Z\left(t+s\right)-Z\left(t\right)=\ln\left(A_{th}/A_{(t-s)h}\right)\overset{d}{=}S_{\alpha}\left(\nu s^{1/\alpha},0,0\right), \tag{2.17}$$

for all $s,t>0$, where

$$\nu=\frac{\sigma\sqrt{c}}{\sqrt{2}\left(\cos\left(\frac{\pi\alpha}{4}\right)\right)^{1/\alpha}}. \tag{2.18}$$

If we consider the constant scalar parameter $\sigma$, then the random variable $Y$ in (2.11) is as follows:

$$Y=\sigma^{2}\left(T\left(t\right)-T\left(t_{0}\right)\right)=\lambda V, \tag{2.19}$$

where $\lambda=c\sigma^{2}\left(t-t_{0}\right)^{2/\alpha}$ and $V=S_{\alpha/2}\left(1,1,0\right)$. Hence, with

$$c=2\left(\cos\left(\frac{\pi\alpha}{4}\right)\right)^{2/\alpha}, \tag{2.20}$$

it follows that $Z\left(t\right)\overset{d}{=}S_{\alpha}\left(\sigma t^{1/\alpha},0,0\right)$. Thus, we can estimate the index of stability $\alpha$ and the scalar parameter $\sigma$ using the maximum likelihood method (see Rachev and Mittnik [147] and the references therein). Moreover, considering the density function $f_V$ of the $\alpha/2$ stable random variable $V$, we obtain the following expression for $F_{\pm}\left(x\right)$:

$$F_{\pm}\left(x\right)=\int_{0}^{1}\Phi\left(\frac{x\pm\frac{1}{2}\lambda u\left(1-u\right)^{-3}}{\sqrt{\lambda u\left(1-u\right)^{-3}}}\right)f_{V}\left(u\left(1-u\right)^{-3}\right)\frac{1+2u}{\left(1-u\right)^{4}}du\,. \tag{2.21}$$

The probability of default under the risk neutral probability measure is then given by

$$\mathrm{PD}_t = \int_0^1 \Phi\left(\frac{\ln\left(\frac{L_{r,t_0,t}}{A(t_0)}\right)+\frac{1}{2}\lambda u(1-u)^{-3}}{\sqrt{\lambda u(1-u)^{-3}}}\right) f_V\left(u(1-u)^{-3}\right)\frac{1+2u}{(1-u)^4}\,du\,. \qquad (2.22).$$

## 2.3 Estimation methodology

While for the Merton model there are just three parameters necessary for the estimation of default probabilities — namely, the company's market value $A_t$ at time $t$, the asset drift $\mu$, and the asset volatility $\sigma$ — in the case of the subordinated models, we have to estimate the company's market value at time $t$ and the parameters of the subordinated process. Clearly, different distributional hypothesis of the subordinated model could require the estimation of several different parameters. For example, in the $\alpha$-stable Lévy process, once the index of stability $\alpha$ is estimated, the scalar parameter $\sigma$ is the unique parameter that should be estimated since the skewness parameter and the location parameter have been fixed equal to zero in the model.

### 2.3.1 Parameter estimates for the Merton model

The unknown parameters of the Merton model come from (2.5). Since the market value of assets is a random variable and cannot be observed directly, it is impossible to directly estimate the drift and the volatility in a movement of log-returns on $A_t$. Therefore, these three parameters have to be estimated in a different way. In fact, we use the observed market value of equity $E_t$ along with (2.5) to estimate them indirectly.

Generally, the starting point for the two iterative methodologies proposed in literature (the maximum likelihood estimation method and the Moody's KMV method) is based on the so-called *calibration method* (see Bluhm, Overbeck and Wagner [23], Crosbie and Bohn [42], Bruche [32], or Ericsson and Reneby [59]), which finds two unknown parameters ( $A_t$ and $\sigma$ ) by solving the system of two equations as follows:

$$\begin{aligned} E_t &= A_t\Phi(d_1) - Le^{-r(T-t)}\Phi(d_2) \\ \sigma_E &= \frac{A_t}{E_t}\Phi(d_1)\sigma \end{aligned}, \qquad (2.23)$$

where $\sigma_E$ is the standard deviation of the equity log returns $\ln\left(E_{th}/E_{(t-1)h}\right)$. Nevertheless, this method does not estimate asset drift $\mu$; it determines the risk neutral probability of default using the risk free asset $r$. As a consequence, Jovan [99] showed that this method provides different estimates of PDs for the same obligors compared to the two following iterative methodologies: the maximum likelihood estimation method and the Moody's KMV method.

**Maximum likelihood estimation (MLE) method**

This methodology was initially proposed by Duan [49] and enhanced later by Duan, Gauthier and Simonato [50]. The time series of daily market value of equity $E_t$ is equal to $n$ days, where $t=(0,\dots,n)$. In Duan, Gauthier and Simonato [50] the time step $h$ is introduced. Typically, the value of this coefficient for daily data would be $h$ = 1/250. The methodology is iterative and the following log-likelihood function for the estimation of $\mu$ and $\sigma$ of model (2.3), where $th=(0,\dots,nh)$, is defined on the basis of observed values of $E_t$ as follows:

$$\begin{aligned} L\left(\hat{\theta};\hat{A}_{th}\,\middle|\,E_{th}\right) = &-\frac{n}{2}\ln\left(2\pi\hat{\sigma}^2h\right) \\ &-\frac{1}{2}\sum_{t=1}^{n}\frac{\left(\hat{R}_t-\left(\hat{\mu}-(1/2)\hat{\sigma}^2\right)h\right)^2}{\hat{\sigma}^2h} \\ &-\sum_{t=1}^{n}\ln\left(\hat{A}_{th}\right)-\sum_{t=1}^{n}\ln\left(\Phi\left(d_1\right)\right), \end{aligned} \tag{2.24}$$

where

$$\hat{R}_t=\ln\left(\frac{\hat{A}_{th}}{\hat{A}_{(t-1)h}}\right), \tag{2.25}$$

and where $\hat{\theta}\equiv(\hat{\mu},\hat{\sigma})$ and $\hat{A}_{th}$ is estimated from (2.5). To launch the iteration process we could insert as initial values the values obtained by solving the system (2.23). Despite the fact that these estimates are not the best ones from a solution point of view, they can be good enough as the initial values for different kinds of iterative procedures. Each iteration produces a time series of daily values $\hat{A}_{th}^{(i)}$, where the debt maturity ranges over $1\le(T-th)\le T$. We maximize (2.24) to obtain estimates of the unobserved asset drift and volatility $\hat{\theta}^{(i)}$. Since this is an iterative procedure, we use the new estimates obtained from (2.24) and the new market value of assets obtained from (2.5) for maximizing (2.24) once again. The procedure is repeated until the differences in $\hat{\mu}^{(i)}$ and $\hat{\sigma}^{(i)}$ between the successive iterations are sufficiently small (i.e., until $\left|\hat{\mu}^{(i+1)} - \hat{\mu}^{(i)}\right| + \left|\hat{\sigma}^{(i+1)} - \hat{\sigma}^{(i)}\right| \leq \varepsilon$ for a given small $\varepsilon$ ).

Duan, Gauthier and Simonato [50] found that the Moody's KMV method provides the same estimates as the MLE method, even though they state that the latter method is preferable for inference statistics.

**Moody's KMV methodology**

This iterative procedure follows a disclosed part of Moody's KMV methodology for a calculation of Expected Default Frequency (see Duan, Gauthier and Simonato [50], Duffie, Saita and Wang [52], Crosbie and Bohn [42], or Vassalou and Xing [172]). This method is quite similar to the MLE method. The unique difference is that in order to obtain estimates of the asset drift and volatility, instead of maximizing the log-likelihood function, we have explicit formulas.

The first step is exactly the same, calculation of the daily value of $\hat{A}_{th}^{(i)}$, $th = \left(0, \dots, nh\right)$ from (2.5). As the initial values can be used again the estimates obtained by solving the system (2.23). Then, the arithmetic mean of the sample is given by

$$\bar{R}^{(i)} = \frac{1}{n} \sum_{t=1}^{n} \hat{R}_t^{(i)}, \tag{2.26}$$

where $\hat{R}_t$ is defined in (2.25). Another step is the calculation of estimates of the asset volatility $\hat{\sigma}$ and the drift $\hat{\mu}$ of model (2.3) which are defined as follows:

$$\begin{aligned} \hat{\sigma}^{(i+1)} &= \sqrt{\frac{1}{nh} \sum_{t=1}^{n} \left(\hat{R}_t^{(i)} - \bar{R}^{(i)}\right)^2}, \\ \hat{\mu}^{(i+1)} &= \bar{R}^{(i)} \frac{1}{h} + \frac{1}{2} \hat{\sigma}^{2(i+1)}. \end{aligned} \tag{2.27}$$

Since this is again an iterative procedure, we use the new estimates obtained from (2.27) to calculate $A_{th}^{(i+1)}$. The procedure is repeated until the differences in $\hat{\mu}$ and $\hat{\sigma}$ among successive iterations are sufficiently small.

It is worth to mention that the Merton model with parameters estimated according to the methodology described above differs from the one actually employed by Moody's KMV. How well the Merton model performs substantially relies on the simplifying assumptions facilitating its implementation. These simplifying assumptions are not really realistic in practice, though. That is why Moody's KMV does not rely solely on these assumptions. Indeed, the founders of KMV, Oldrich Vasicek and Stephen Kealhofer, developed a so-called Vasicek-Kealhofer (VK) model (see Arora, Bohn and Zhu [9]) to estimate the distance-to-default of an individual company. One of the most important differences is that while we use the cumulative normal distribution to convert distances-to-default into "real" default probabilities in classical Merton model, Moody's KMV uses its large historical database to estimate the real empirical distribution of distances-to-default, and it calculates default probabilities based on that distribution.

## 2.3.2 Parameter estimates for subordinated models

We can extend the estimation methodologies proposed for the Merton model in order to estimate the parameters of a subordinated model.

**Maximum likelihood estimation (MLE) method**

Obviously, in order to use this method, we have to revise (2.24). In fact, (2.24) can be derived from more general formula which can be used for the derivation of log-likelihood functions for any subordinated model. This formula is defined in the following way:

$$L\left(\hat{\theta};\, \hat{A}_{th} \mid E_{th}\right) = \sum_{t=1}^{n} \ln\left(f_Z\left(\hat{R}_t\right)\right) - \sum_{t=1}^{n} \ln\left(\hat{A}_{th}\right) - \sum_{t=1}^{n} \ln\left(\Delta_E\right), \tag{2.28}$$

where $\hat{\theta}$ represents the set of the parameters in the density function $f_Z\left(\hat{R}_t\right)$ of the stationary increment $\ln\left(A_{th} / A_{(t-1)h}\right) = Z(t+1) - Z(t)$, $\hat{A}_{th}$ is estimated from (2.11), $\hat{R}_t$ is defined in (2.25), and $\Delta_E$ is given by (2.14). The initial values $\hat{A}_{th}^{(1)}$ of the iteration process could be the ones obtained by solving the system (2.23). The procedure continues iteratively till the distance $\left\|\hat{\theta}^{(i+1)} - \hat{\theta}^{(i)}\right\|$ is sufficiently small. Typically, there are two problems regarding this maximum likelihood method. The first difficulty is related to computation time. This method generally presents more local optima, and it can be very time consuming to reach a global optimum. Secondly, it is often very problematic to implement this methodology since many subordinated models do not have close form equation for the density function $f_Z$ .

**An extended Moody's KMV methodology**

As for Moody's KMV iterative methodology, we have to first compute the daily value of $\hat{A}_{th}^{(i)}$, $th = (0, \dots, nh)$ solving (2.11), then the other parameters of the subordinated process $\hat{\theta}^{(i+1)}$ are estimated on the series $\hat{R}_t^{(i)} = \ln\left(\hat{A}_{th}^{(i)} / \hat{A}_{(t-1)h}^{(i)}\right)$ considering the distributional assumption of the subordinated model. The procedure continues iteratively till the distance $\left\|\hat{\theta}^{(i+1)} - \hat{\theta}^{(i)}\right\|$ is sufficiently small. In particular, for the $\alpha$ -stable Lévy model, we first suggest to determine the index of stability $\alpha$ . Secondly, the unique parameter that must be estimated is the scalar parameter $\sigma$ since the skewness parameter and the location parameter are fixed equal to zero. Clearly, even in this case, we need to insert some initial values $\hat{A}_{th}^{(1)}$ of the iteration process that could be the ones obtained by solving the system (2.23). Moreover, as for the Merton model (see Duan, Gauthier and Simonato [50]), the extended Moody's KMV methodology provides the same estimates as the MLE method when the parameter estimates $\hat{\theta}^{(i+1)}$ are the MLE on the series $\hat{R}_t^{(i)}$ .

## 2.4 Application and results

In this section, we first describe the data used in the computational analysis and apply the Merton model. Subsequently, we test the distributional assumption of this model. Finally, we apply the stable Lévy model and compare obtained results with the Merton's ones. We use Moody's KMV and the extended Moody's KMV methodology described in Section 2.3.1 and 2.3.2, respectively, whilst estimating parameters of the models.

To apply the above mentioned models to a particular company, we need the market value of equity $E_t$, the face value of the zero-coupon debt instrument $L$, and the risk-free interest rate $r$. We used 13-week Treasury bill for risk-free interest rate. Thomson Reuters Datastream dataset was used to obtain the market value of equity and the face value of the zero-coupon debt instrument. Our sample contains 24 U.S. companies with strong capitalization in the U.S. market.[12] Our data spans the period from January 3, 2000, to December 30, 2011. As the market value of equity, we used consolidated market value of a company which is defined as a share price multiplied by the number of ordinary shares in issue. Finally, for the face value of the zero-coupon debt instrument, we used the sum of the short-term debt, current portion of the long-term debt, and half of the long-term debt.[13] While the short-term debt and current portion of the long-term debt represent that portion of the debt payable within one year including current portion of the long-term debt and sinking fund requirements of preferred stock or debentures, the long-term debt represents all interest bearing financial obligations excluding amounts due within one year.

### 2.4.1 Analysis of the distributional assumptions of the company value log-returns

The Merton's model distributional assumption implies that the unobservable company value log-returns are Gaussian distributed. In order to test this assumption, we use the daily log-returns of the companies' asset values obtained from both the Merton model and the alpha stable Lévy model, from January 3, 2000, to December 30, 2011 (for a total of 3157 daily values).

First of all, we test the Gaussian and the stable non-Gaussian hypotheses on the company value log-returns obtained from the Merton model. Thus, we compute different statistics every day on the last 250 daily company values (1 year of daily values). Table 2.1 reports the

[12] The companies are (1) Boeing, (2) Cisco Systems, (3) Chevron, (4) E. I. du Pont de Nemours, (5) Walt Disney, (6)Home Depot, (7)Hewlett-Packard, (8) IBM, (9) Intel, (10) Johnson & Johnson, (11) Coca Cola, (12)McDonalds, (13) 3M, (14) Merck & Co., (15) Microsoft, (16) Pfizer, (17) Procter & Gamble, (18) AT & T, (19) UnitedHealth Group, (20) United Technologies, (21) Verizon Communications, (22) WalMart Stores, (23) Exxon Mobil, and (24) Travelers Companies.

[13] There needs to be chosen an amount of the debt that is relevant to a potential default during a one year period. Total debt is inadequate when not all of it is due in one year (it is assumed one-year time horizon for debt maturity and subsequent estimation of PD), as the firm may remain solvent even when the value of assets falls below its total liabilities. Using the short-term debt for the default barrier would be often wrong, for instance, when there are covenants that force the company to serve other debts when its financial situation deteriorates. Prior studies generally choose the short-term debt plus half of the long-term debt for the default barrier (see Bharath and Shumway [18], Vassalou and Xing [172], or Duffie, Saita and Wang [52]).

average among all the firms and for all the ex-post period of different statistics applied to company value log-returns to test the Gaussian hypothesis and the stable non-Gaussian hypothesis. In particular, we consider the average of the following statistics: the mean, the standard deviation, the skewness $E\left((X-E(X))^3\right)/E\left((X-E(X))^2\right)^{1.5}$, the kurtosis $E\left((X-E(X))^4\right)/E\left((X-E(X))^2\right)^2$, the percentage of rejection of the Gaussian hypothesis using the Jarque-Bera (JB) test at the 5% significance level (see Jarque and Bera [96]), the stable index of stability “alpha”, the stable index of skewness “beta”, the stable scalar parameter “sigma”, the stable location parameter “mu”, and the percentage of rejection of the stable Paretian hypothesis using the Kolmogorov-Smirnov (KS) test at the 5% significance level.

The results reported in Table 2.1 suggest that: (1) the returns exhibit heavy tails since the average of the stability parameters alpha is less than 2 and the average of kurtosis is much higher than 3; (2) the returns are slightly asymmetric since the average of the skewness parameter and the average of the stable parameter beta are different from zero; and (3) the Gaussian hypothesis is almost always rejected for all companies while the stable Paretian hypothesis is generally rejected only for four companies of the considered sample.

**Table 2.1**
Descriptive statistics for the log-returns of the companies’ asset values

| | | | |
|---|---|---|---|
| mean | 0.0000 | alpha | 1.7089 |
| st.dev. | 0.0196 | beta | 0.0062 |
| skewness | -0.6140 | sigma | 0.0106 |
| kurtosis | 33.4351 | mu | 0.0001 |
| JB test (95%) | 96.77% | KS test (95%) | 16.56% |

The table reports the average of chosen statistics among 24 companies in our sample, applied to the daily log-returns of the companies’ asset values obtained from the Merton model. We also test the Gaussian hypothesis using the Jarque-Bera (JB) test and the stable non-Gaussian hypothesis using the Kolmogorov-Smirnov (KS) test. Particular statistics are expressed in decimal numbers, whilst JB and KS tests denote the percentage of the hypotheses rejection.

Next, using a Kolmogorov-Smirnov (KS) test (at the 5% significance level) we test the different distributional hypothesis for the log-returns of the companies' asset values obtained from the stable Lévy model. We observe almost the same percentage of rejection (16.55%) as we get from the Merton model (16.56%). Similarly, applying the Jarque-Bera (JB) test we get 98.44% of rejection of the Gaussian hypothesis from the stable Lévy model (compared to 96.77% obtained from the Merton model). From this preliminary analysis, we deduce that the classical distributional hypothesis of the Merton model is almost never verified. Moreover, the stable non-Gaussian hypothesis appears more realistic than the Gaussian one. Therefore, it is appropriate to apply the stable Lévy model which is able to capture empirically observed anomalies that contradict the classical normality assumption. The results we get here are not a real surprise since the stable Paretian laws generalize the Gaussian one.

### 2.4.2 PD estimates from the Merton model

We applied Moody's KMV methodology[14] to estimation of the parameters for the Merton model and subsequently used these parameters for calculation of the probability of default for a given company. The results of the empirical analysis are reported in Figure 2.1 and Table 2.2. In Table 2.2, there are listed average values of the ratio between the debt and the companies' asset values and average values of PDs and distances-to-default obtained from the Merton model. In particular, we observe that when the average ratio between debt and company value is high, there is generally an analogous higher probability of default and a lower distance-to-default. This aspect could be a problem when using this model for calculating the risk neutral and real default probabilities of a bank since financial institutions have significantly greater debt compared to other companies. Therefore, the Merton model is not plausible for the estimation of PDs of financial institutions unless some adjustments are made.[15]

Figure 2.1 describes the evolution of the PDs on the monthly basis. These probabilities are almost null during all the decade. However, we can distinguish three periods of increased PDs for some companies from our sample. First, at the beginning of the century after the

[14] We perform our analysis using MATLAB.

[15] For example, Byström [33] shows that one of the main implications of his simplified "spread sheet" version of the Merton model is the fact that the default probability's insensitivity to the leverage ratio at high levels of debt makes it possible to apply his model to banks and other highly leveraged firms.

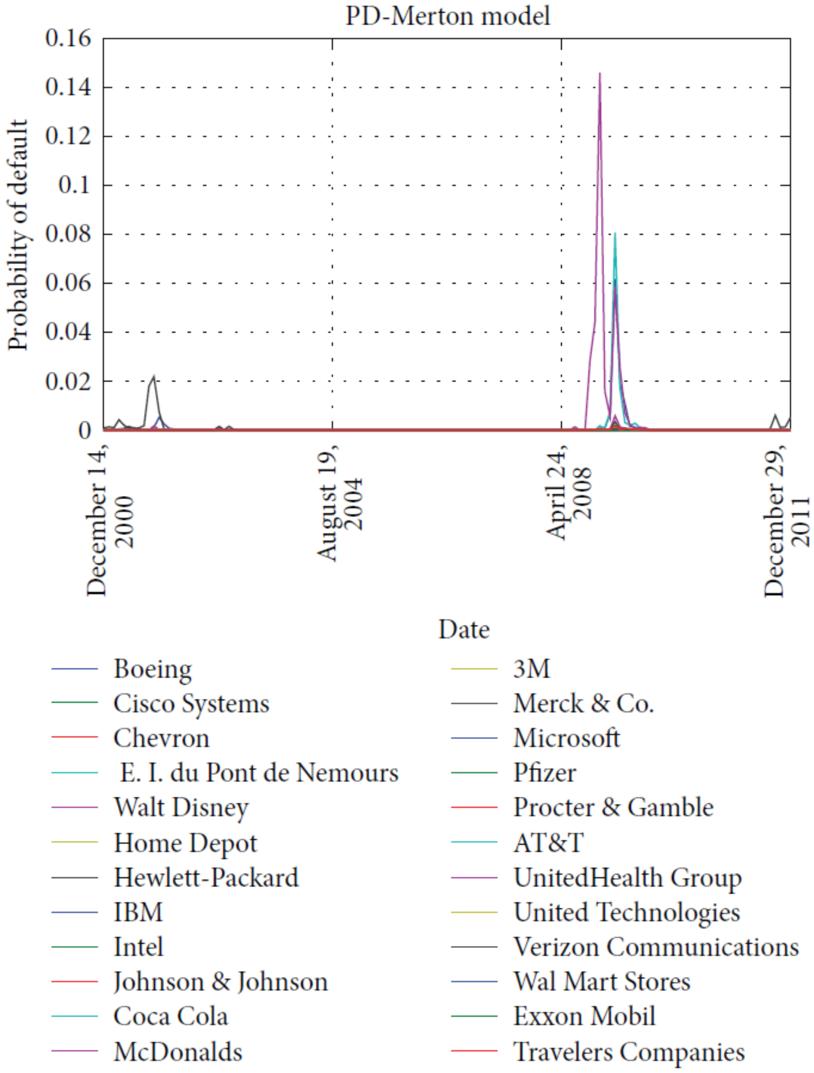
**Table 2.2**
Outcomes from the Merton model

| Company | Average ratio ($L/A$) | Average PD | Average DD |
|---|---|---|---|
| (1) Boeing | 0.1326 | 0.000830 | 8.9020 |
| (2) Cisco Systems | 0.0262 | 0.000000 | 20.6010 |
| (3) Chevron | 0.0613 | 0.000000 | 13.8524 |
| (4) E. I. du Pont de Nemours | 0.1169 | 0.000845 | 9.9706 |
| (5) Walt Disney | 0.1312 | 0.000083 | 8.5109 |
| (6) Home Depot | 0.0600 | 0.000002 | 11.8297 |
| (7) Hewlett-Packard | 0.0909 | 0.000511 | 8.3242 |
| (8) IBM | 0.1037 | 0.000000 | 11.4799 |
| (9) Intel | 0.0099 | 0.000000 | 14.2761 |
| (10) Johnson & Johnson | 0.0331 | 0.000000 | 22.8226 |
| (11) Coca Cola | 0.0615 | 0.000000 | 17.5142 |
| (12) McDonalds | 0.1031 | 0.000015 | 12.2037 |
| (13) 3M | 0.0493 | 0.000000 | 14.9342 |
| (14) Merck & Co. | 0.0611 | 0.000037 | 11.1672 |
| (15) Microsoft | 0.0068 | 0.000000 | 21.4008 |
| (16) Pfizer | 0.0815 | 0.000019 | 11.0915 |
| (17) Procter & Gamble | 0.1010 | 0.000000 | 13.9819 |
| (18) AT&T | 0.1619 | 0.000013 | 8.4346 |
| (19) UnitedHealth Group | 0.0924 | 0.002424 | 10.2912 |
| (20) United Technologies | 0.0800 | 0.000001 | 12.1045 |
| (21) Verizon Communications | 0.2117 | 0.000106 | 8.8750 |
| (22) Wal Mart Stores | 0.0957 | 0.000000 | 12.4895 |
| (23) Exxon Mobil | 0.0208 | 0.000000 | 18.0516 |
| (24) Travelers Companies | 0.1298 | 0.000035 | 8.9095 |

The table reports average monthly values of the ratio between the debt and the company value $(L/A)$, default probabilities (PD), and distances-to-default (DD) obtained from the Merton model for 24 companies in our sample. All values are expressed in decimal numbers.

high-tech crisis and September 11, 2001; second, during the subprime crisis in 2008 and 2009; and finally third, during the country credit risk crisis in 2011. During the first period and the country credit risk crisis, the most evident grown of PD is due to the Hewlett-Packard firm (its PD increased up to 2.1% in the first period and to 1% in the last one). The period with more significant growth in PDs is dated from September 2008. This might be easily explained by the subprime mortgage crisis that reached a critical stage during the first week of September 2008 and was characterized by severely contracted liquidity in the global credit markets and insolvency threats to investment banks and other institutions. Beginning with bankruptcy of Lehman Brothers on September 14, 2008, the financial crisis entered an acute phase marked by the failures of prominent American and European banks and efforts by the American and

**Figure 2.1**
The Merton model – monthly PDs




The figure plots the evolution of monthly PDs obtained from the Merton model for 24 companies in our sample.

European governments to rescue distressed financial institutions. Among the companies from our sample which were affected the most belong UnitedHealth Group, E. I. du Pont de Nemours, and Boeing. UnitedHealth Group is a care company which offers a spectrum of products and services. This company suffered a jump in PD from 0% in May 2008 up to 14.6% in November 2008. E. I. du Pont de Nemours is a chemical company and was the world's third largest chemical company based on market capitalization in 2009. This company's PD increased from 0% in October 2008 to 8.1% in February 2009. Finally, Boeing as a representative of aerospace industry suffered an increase in PD from 0% in October 2008 to 6.2% in February 2009. This phase of financial crisis lasted approximately one year and the values of PD of observed companies went back to zero in October 2009.

### 2.4.3 PD estimates from the stable Lévy model

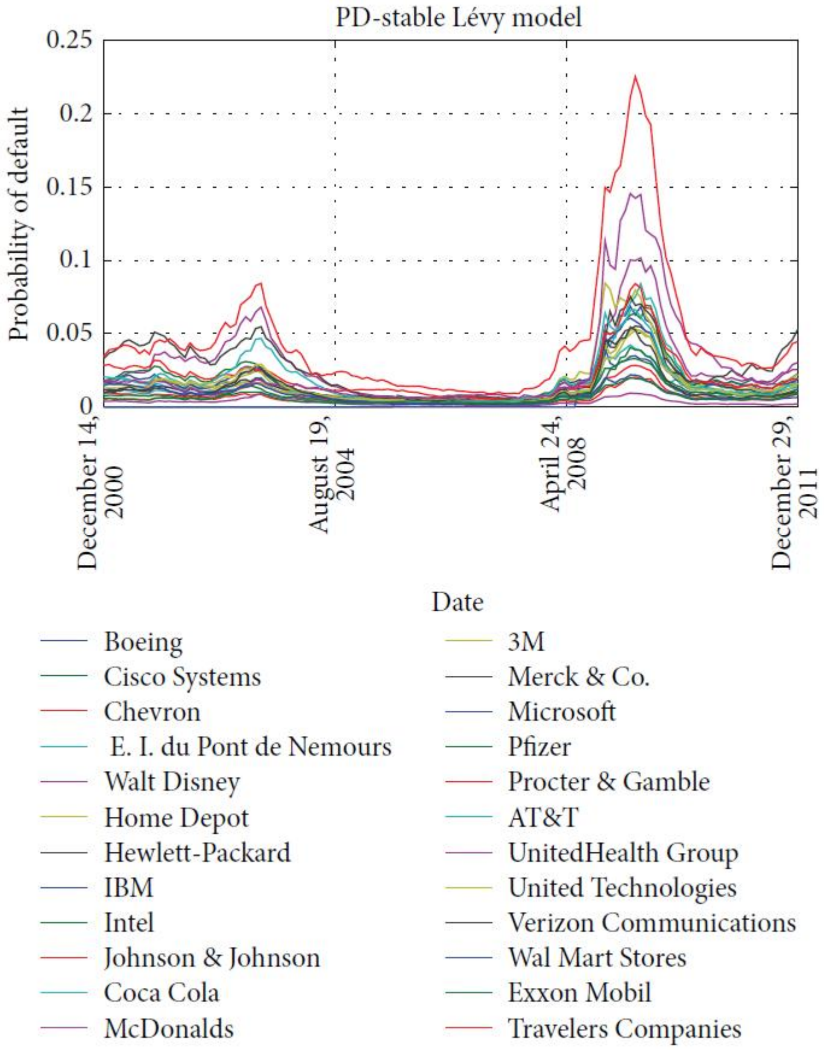
In order to evaluate the stable Lévy model, we estimate the parameters using the extended Moody's KMV methodology. First of all, we calculate the indices of stability (alphas) on the daily log-returns of the companies' asset values, obtained by the stable Lévy model, which are reported in Table 2.3. To evaluate the stable parameters and the distributions of subordinator $f_V$ in (2.21), we perform a maximum likelihood estimator that uses the fast Fourier transform (see Rachev and Mittnik [147] or Nolan [140]). The estimated index of stability is maintained constant for each firm and for all the period of analysis. Clearly, we could have adapted the model more dynamically requiring that the index of stability changes periodically with the scalar and location stable parameters. However, this would require the

**Table 2.3**
Outcomes from the stable Lévy model

| Company | Alpha | Average ratio ($L/A$) | Average PD | Average DD |
|---|---|---|---|---|
| (1) Boeing | 1.6619 | 0.1308 | 0.0149 | 8.9153 |
| (2) Cisco Systems | 1.5756 | 0.0262 | 0.0116 | 20.4104 |
| (3) Chevron | 1.6671 | 0.0606 | 0.0067 | 13.7868 |
| (4) E. I. du Pont de Nemours | 1.6575 | 0.1169 | 0.0137 | 10.0480 |
| (5) Walt Disney | 1.5680 | 0.1305 | 0.0265 | 8.5155 |
| (6) Home Depot | 1.6101 | 0.0599 | 0.0173 | 11.9741 |
| (7) Hewlett-Packard | 1.5850 | 0.0914 | 0.0253 | 8.3069 |
| (8) IBM | 1.6110 | 0.1032 | 0.0120 | 11.5404 |
| (9) Intel | 1.6411 | 0.0098 | 0.0131 | 14.3321 |
| (10) Johnson & Johnson | 1.5803 | 0.0330 | 0.0068 | 22.9854 |
| (11) Coca Cola | 1.5505 | 0.0614 | 0.0120 | 17.6094 |
| (12) McDonalds | 1.7570 | 0.1012 | 0.0032 | 12.3247 |
| (13) 3M | 1.5590 | 0.0494 | 0.0136 | 14.9028 |
| (14) Merck & Co. | 1.5909 | 0.0610 | 0.0150 | 11.1738 |
| (15) Microsoft | 1.5459 | 0.0068 | 0.0082 | 21.1204 |
| (16) Pfizer | 1.6691 | 0.0813 | 0.0085 | 11.2040 |
| (17) Procter & Gamble | 1.4745 | 0.1010 | 0.0204 | 13.9846 |
| (18) AT&T | 1.5985 | 0.1607 | 0.0176 | 8.5163 |
| (19) UnitedHealth Group | 1.5839 | 0.0925 | 0.0256 | 10.3436 |
| (20) United Technologies | 1.6064 | 0.0798 | 0.0138 | 12.0951 |
| (21) Verizon Communications | 1.6645 | 0.2106 | 0.0114 | 8.9470 |
| (22) Wal Mart Stores | 1.6398 | 0.0955 | 0.0080 | 12.5641 |
| (23) Exxon Mobil | 1.6494 | 0.0207 | 0.0060 | 18.1822 |
| (24) Travelers Companies | 1.4659 | 0.1291 | 0.0464 | 8.9419 |

The table reports the indices of stability (alphas) and average monthly values of the ratio between the debt and the company value $(L/A)$, default probabilities (PD), and distances-to-default (DD) obtained from the stable Lévy model for 24 companies in our sample. All values are expressed in decimal numbers.

**Figure 2.2**
The stable Lévy model – monthly PDs




The figure plots the evolution of monthly PDs obtained from the stable Lévy model for 24 companies in our sample.

knowledge of the subordinator density distribution $f_V$ that changes with the index of stability. Since this distribution is obtained by inverting the Fourier transform, the iterating procedure of the Moody's KMV methodology would require too much computational time in that case. In Table 2.3, there are also listed the average values of the ratio between the debt and the companies' asset values and average values of PDs and distances-to-default obtained from the stable Lévy model. Figure 2.2 then describes the evolution of PDs on the monthly basis.

### 2.4.4 Comparison of the Merton and stable Lévy model

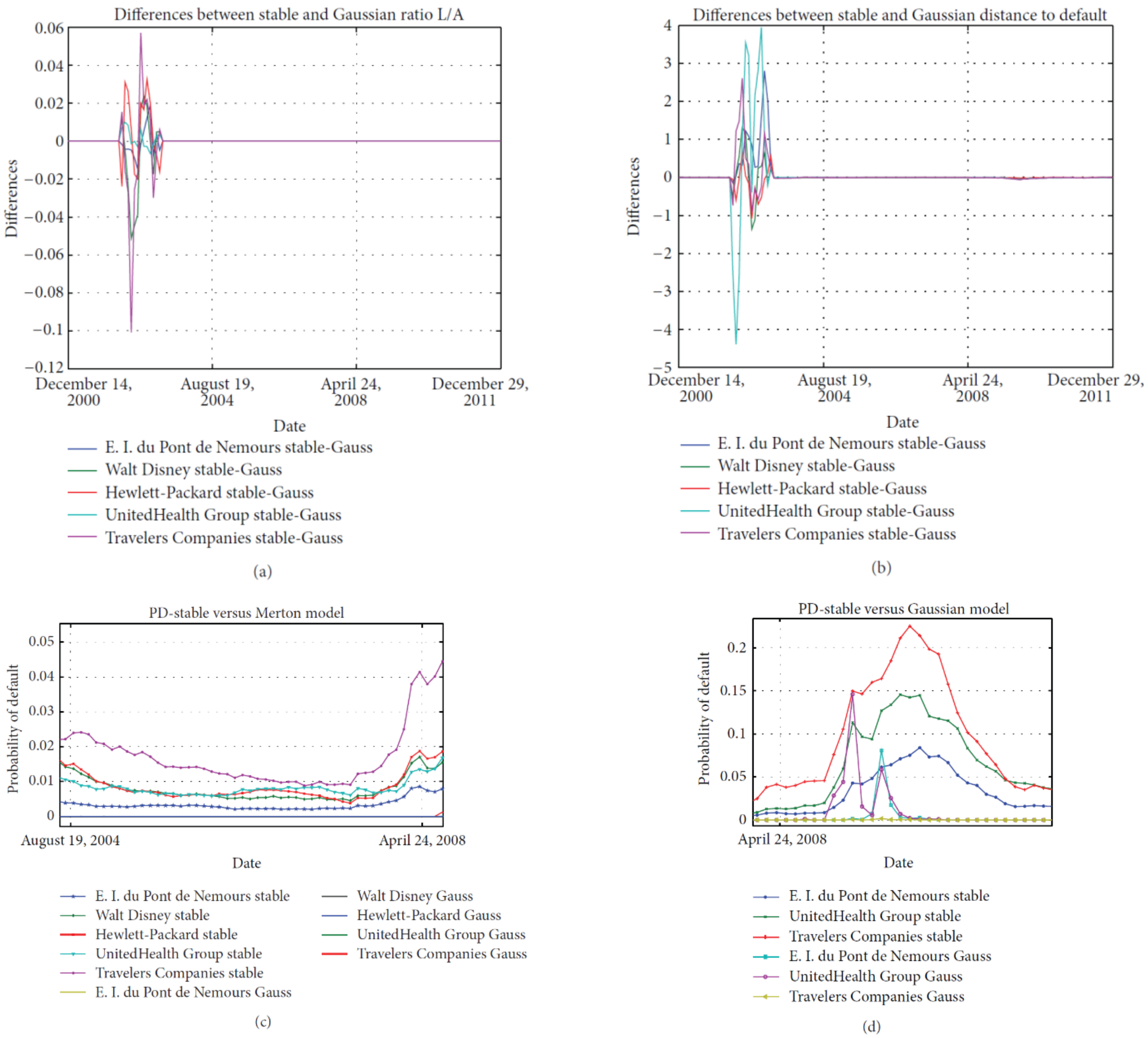
Comparing the outcomes of the two models, we observe that there are not very large differences between the companies' values obtained by the stable Lévy model and the companies' values obtained by the Merton model. This finding is not very surprising as we could not have expected strong differences in these values that represent an unobservable objective variable whose big differences could not be easily justifiable. This observation implies that there are not large differences between the two models with respect to: (1) the average ratio between the debt and the company value; (2) the average distance-to-default.

Figure 2.3 reports the main differences between the two models for those companies that present the highest peaks in default probabilities (E. I. du Pont de Nemours, Walt Disney, Hewlett-Packard, UnitedHealth Group, and Travelers Companies). In particular, Figures 2.3(a) and 2.3(b) show that the main differences in the ratio between the debt and the company value and between the distances-to-default, respectively, are concentrated during the high volatility period after September 11, 2001. However, this difference (as remarked previously) is almost null during the big crisis following the Lehman Brothers bankruptcy. Figures 2.3(c) and 2.3(d) show default probabilities of chosen companies during "calm" periods and during periods of the crisis, respectively. In this case, we observe very big differences between estimated PDs. On one hand, the probabilities of default computed by the Merton model are almost null during the "calm" periods and increase during one or two months of the crisis. On the other hand, the default probabilities computed by the stable Lévy model are never null during the "calm" periods and become very high during the months of the crisis and in the close subsequent periods.

Particularly, we observe the biggest difference for the Travelers Companies for which the Merton model does not register any significant difference in the default probabilities while the stable Lévy model shows the highest values. This difference is essentially caused by the combination of two aspects. First, the index of stability of the Travelers Companies is very small, which means very fat tails with high probability of losses. Second, the ratio between the debt and the Travelers Companies assets value is high. This analysis confirms the previous finding that the average default probabilities obtained by the stable Lévy model are much

**Figure 2.3**

Differences between the models for chosen companies




The figure plots the main differences between the Merton and stable Lévy model for companies that present the highest peaks in default probabilities (E. I. du Pont de Nemours, Walt Disney, Hewlett-Packard, UnitedHealth Group, and Travelers Companies). In particular, (a) plots the differences between stable and Gaussian ratio $(L/A)$ over the whole sample period; (b) plots the differences between stable and Gaussian distances-to-default over the whole sample period; (c) plots probabilities of default during "calm" periods; (d) plots probabilities of defaults during the crisis.

higher than those obtained by the Merton model. This is not a real surprise since while the probability tails of the Gaussian distribution tend to zero exponentially, the probability tails of stable non-Gaussian distribution tend to zero in polynomial order. Therefore, the probability of losses calculated by the stable Lévy model is much higher than the probability of losses obtained from the Merton one. This effect is also emphasized in Figure 2.2 that reports the evolution of default probabilities during the decade 2001–2011. It shows much higher sensitivity of these probabilities for all companies with respect to the periods of crises. Moreover, since all the tests have shown that the stable non-Gaussian hypothesis appears more realistic than the Gaussian one, we conclude that the Merton model generally underestimates the probability of default.

## 2.5 Conclusions

The structural approach to credit risk modeling, initially developed by Merton [135], has been widely used over the last decades. The basic idea behind this framework is to treat company's equity and debt as a contingent claim written on company's asset value. However, the classical version of this model requires a number of simplifying and unrealistic assumptions. In this study, we focus on overcoming the assumption that company value follows the log-normal distribution. In fact, we prove that this assumption is generally rejected, and consequently the log-returns of the companies' asset values are not Gaussian distributed. For this reason, we propose alternative structural credit risk model and discuss how to evaluate the probability of default of a given firm under different distributional hypotheses.

In particular, we implement a structural credit risk model based on the stable Paretian distributions as a representative of subordinated models. The practical and theoretical appeal of the stable non-Gaussian approach is given by its attractive properties that are almost the same as the normal ones. We argue that it is possible to use this model in the Merton's framework. In fact, we propose an empirical comparison of the Moody's KMV methodology applied to the Merton model and our subordinated one. Particularly, we suggest alternative parameter estimation for subordinated processes and optimize the performance for the stable Lévy model.

The empirical analysis suggests that the probability of default is generally underestimated by the Merton model. Clearly, these results should be further discussed and compared to other distributional models in a future research. As a matter of fact, two alternative structural credit risk models based on well-known symmetric Lévy processes (the Variance Gamma (VG) process and the Normal Inverse Gaussian (NIG) process) were proposed by Brambilla, Gurny and Ortobelli Lozza [30]. Once the framework of these models has been established, the authors focus on empirical comparison of estimated default probabilities. On the same data set used in Gurny, Ortobelli Lozza and Giacometti [80], the authors demonstrate that both models are able to capture the situation of instability that affects each company in considered period and, in fact, are very sensitive to the periods of the crises. Specifically, default probabilities from the NIG model exhibit a greater level of variability compared to the VG model.[16] Furthermore, they observe that increased PDs are also present in the aftermath of the crises. Overall, the authors find that PD estimates obtained from the NIG model are significantly higher than those from the VG model.

[16] The authors state that within application of the NIG model it was often necessary to approximate PD estimates, since the extended Moody's KMV methodology required many more iterations to converge. This fact implies that the probabilities are not always well approximated. Nevertheless, the same general trend of PDs obtained from the VG model is also found in the NIG one.

# Chapter 3

# Prediction of U.S. Commercial Bank Failures via Scoring Models: The FFIEC Database Case

In previous chapter, we have dealt with structural credit risk models and applied this framework to 24 U.S. companies with strong capitalization in the U.S. market. In fact, all these companies were components of the Dow Jones Industrial Average index at the time our analysis was conducted. However, we dropped financial institutions from our analysis since one of our findings is that the companies with a higher value of the ratio between the debt and the companies' asset values tend to have a higher average value of default probability. One of the consequences of this fact is that structural credit risk models based on the Merton's framework are not plausible for the estimation of PDs of financial institutions (due to their different balance sheet structure), unless some adjustments are made. This is the reason why we devote our second study to estimation of PDs of financial institutions.

In this chapter, we examine the performance of static and multi-period credit-scoring models for determining default probabilities of financial institutions. Due to their simplicity, credit-scoring models are among the most popular and widely used approaches for the estimation of PDs. These multivariate models use financial indicators of a company as input and attribute a weight to each of these indicators that reflects its relative importance in predicting the risk of default.

The main contribution of this study is threefold. First, unlike many other studies that focus on estimating rating models for the corporate sector, we provide a study where rating models for financial institutions are derived and investigated. Literature on this topic is still rather limited, mainly due to insufficient number of historical defaults in the financial sector, which is essential for estimating such models. However, the importance of such assessment for financial institutions has become even more obvious since the recent period of financial and economic turmoil during the financial crisis. We take advantage of the fact that there were 492 defaults of commercial banks in the U.S. from February 2, 2007 to December 31, 2013.[17] This fact encouraged us to compile and examine a significant database of historical financial ratios for defaulted banks. To the best of our knowledge, we provide the first empirical study to use this extensive sample of financial institutions for the estimation and evaluation of default prediction models. While, for example, Canbas, Cabuk and Kilic [35] work with 40 privately owned Turkish commercial banks and 21 defaults, Kolari, Glennon, Shin and Caputo [104] use over 1,000 large U.S. commercial banks (they define large banks to be greater than $250 million in total assets) in each year with 55 defaults in total. They split their sample of failed banks into an original sample used to build a model (containing 18 large failed banks) and a holdout sample (containing remaining 37 large failed banks). In comparison, our sample contains more than seven thousand U.S. commercial banks with up to 405 banks defaulted during the period 2007-2013. We use a framework called the walk-forward approach, see, e.g., Stein [162], with out-of-time validation. The approach allows us in each period to use the maximum number of available data to fit and test the models, such that we are not restricted to dividing our sample into an estimation and holdout sample.

Second, we provide one of the first studies to look at the Federal Financial Institutions Examination Council (FFIEC) database and to provide scoring models for these banks. This publicly accessible database includes complete and detailed financial reports on most FDIC-insured institutions. Even though there were authors such as Vitale and Laux [173], who used information from this database in order to examine the hypothesis that mergers and acquisitions did not produce better performing institutions during the 2006 to 2008 period, the full sample of banks contained in this database has not been used so far to build a credit-scoring model.

Third, we focus and provide a variety of methods for evaluating the performance of the considered models. Unfortunately, the literature does not provide a large number of studies

[17] http://www.fdic.gov/bank/individual/failed/banklist.html

that distinguish between the performance of the models and if they do, they often satisfy with a comparison to the $Z$-score proposed by Altman [4] or the $O$-score suggested by Ohlson [141]. Clearly, more sophisticated techniques for model comparison are required as those scores were derived several decades ago. A key study in this regard might be considered Stein [162] where an overview of some evaluation and comparison techniques is provided with a focus on potential challenges of model validation under real-world conditions. We apply some of the techniques suggested in Stein [162], including the walk-forward approach with out-of-time validation, ROC curve analysis, calibration accuracy tests and bootstrapping of ROC curve areas. Building on existing work, we also suggest a number of additional performance evaluation techniques that have not yet been applied in the literature on scoring models. We suggest to use nonparametric tests such as the Kruskal-Wallis and Tukey's multiple comparison test to investigate significant differences between the particular models in terms of bootstrapped ROC areas. Although DeLong, DeLong and Clarke-Pearson [46] provide a test for the difference between the areas under the ROC curves of two rating models, this test relies on assumptions of asymptotic normality what is often violated as pointed out by Engelmann, Hayden and Tasche [57]. Unlike this test, the proposed nonparametric Kruskal-Wallis and Tukey's multiple comparison tests for our study do not require the assumption of normality. Further, as an extension of log-likelihoods calculated within calibration accuracy test suggested in Stein [162], we apply the Vuong's closeness test for non-nested models (see Vuong [174]) to determine whether calculated log-likelihoods for various models are statistically different. Finally, we also apply the Hosmer-Lemeshow's chi-squared goodness-of-fit test (see Hosmer Jr and Lemeshow [91]) to examine the overall fit of the estimated models. Due to the number of estimated models and the fact that different models perform best according to different criteria, we also create a simple ranking system to provide an overall summary on the performance of estimated models.

This chapter is organized as follows. In Section 3.1, we firstly provide literature review on credit-scoring models. Section 3.2 is devoted to description of the FFIEC council and its database. The theoretical aspects of particular models used in this paper, including static single-period and multi-period discrete hazard models based on logistic and probit regression techniques, along with the proposed evaluation techniques, are described in more detail in Section 3.3. Section 3.4 provides empirical results on model estimation and validation. Finally, Section 3.5 concludes and summarizes the results.

## 3.1 Literature review

Although the techniques underlying credit-scoring models were devised in 1930s by authors such as Fisher [69] and Durand [55], the decisive boost to a development and spread of these models came in 1960s with the studies by Beaver [15] and Altman [4]. The latter one has been considered by many as the most significant study in this field. The resulting $Z$ -score (derived from multiple discriminant analysis) has often been considered as a benchmark model and has often been compared to the performance of models presented in the literature at a later stage.

Other seminal contributions in the field are attributed to McFadden [134] who, from the statistical point of view, contrasted discriminant analysis with logit models. Altman, Haldeman and Narayanan [6] investigated the predictive performance of a seven variable discriminant analysis model ("Zeta model") which improved upon Altman's [4] earlier five variable model. The study by Santomero and Vinso [154] systematically developed probabilistic estimates of failure. Ohlson [141] showed that the predictive power of any model depends upon when the information (financial report) is assumed to be available and the predictive power of linear transforms of a vector of ratios seems to be robust across estimation procedures. The natural Hausman specification test of distributional assumptions for discriminant and logit analysis by comparing the two estimators is proposed by Lo [117]. Queen and Roll [145] used market indicators in order to predict survival of firms in their study. More recently, a simple hazard model for forecasting bankruptcy in the corporate sector has been developed by Shumway [157] who demonstrates that this model corrects for the period at risk and allows for time-varying covariates. Altman, Rijken, Balan, Mina, Forero and Watt [7] developed a new " $Z$ -Metrics" model for the *RiskMetrics Group* which is in fact an updated and improved version of the $Z$ -score methodology.

It is commonly thought that just as banks and other lending institutions examine the financial statements of prospective borrowers, the financial statements of banks themselves need to be analysed by regulators to assess the risk of bank failure. However, the majority of previously proposed credit-scoring models have been derived from samples of non-financial institutions, mainly due to their different balance sheet structure and insufficient number of financial institutions' defaults occurring in the past. Nevertheless, there were several attempts to identify the key factors for healthy financial institutions originating from financial statements.

Among the first authors to apply these models to commercial banks and develop so-called early warning system (EWS) of bank failure are Stuhr and Van Wicklen [164], Korobow and Stuhr [106], Sinkey [158], and Korobow, Stuhr and Martin [107]. These authors used multiple discriminant function or arctangent regression in order to distinguish between banks that were accorded high summary ratings by bank supervisory authorities and banks that were given low summary ratings. Martin [132] as first used a logistic regression approach for early warning of bank failure. An excellent overview and critique of the literature for scoring models up to year 1981 can be found in Altman, Avery, Eisenbeis and Sinkey [5]. West [176] implemented a factor-analysis approach along with logit regression to measure the condition of individual institutions and to assign each of them a probability of being a problem bank. Other contributions to research on failed banks can be found in Bovenzi, Marino and McFadden [27], Korobow and Stuhr [105], Lane, Looney and Wansley [113], Maddala [124], Whalen and Thomson [178], Espahbodi [60], Thomson [168], Kolari, Caputo and Wagner [103]. This research has confirmed that scoring models perform well as EWSs. More recent work on this topic has been conducted by Logan [118] who implemented a logit model to identify leading indicators of failure for U.K. small banks. His analysis focuses on a small banks' crisis of the early 1990s.

Other recent studies tend to combine parametric and nonparametric approaches for the prediction of bank failures. Tam and Kiang [167] implement a neural network approach to perform discriminant analysis. An integrated model approach for bankruptcy prediction has been introduced by Jo and Han [98]. The authors use discriminant analysis with two artificial intelligence models (neural network and case-based forecasting) and conclude that the integrated models produce higher prediction accuracy than individual models. Alam, Booth, Lee and Thordarson [3] identifies potentially failing banks using fuzzy clustering algorithm and self-organizing neural networks. Kolari, Glennon, Shin and Caputo [104] apply both logit analysis and the nonparametric approach of trait recognition to the problem of predicting large U.S. commercial bank failures. They conclude that both models performed well in terms of classification results, however with regards to the prediction results using holdout samples, trait recognition outperforms logit in most tests in terms of minimizing Type I and II errors. A very similar approach is employed by Lanine and Vennet [114] to predict failures among Russian commercial banks. The study tests if bank-specific characteristics can be used to predict vulnerability to failures and shows that liquidity, asset quality and capital adequacy are important determinants of bankruptcy. Lam and Moy [110] combine several discriminant methods and perform simulation analysis to enhance the accuracy of classification results. Canbas, Cabuk and Kilic [35] conduct research on bank failure prediction in Turkey and use principal component analysis to explore the basic financial characteristics of the banks. The authors also subsequently estimate discriminant, logit and probit models based on these characteristics. The most recent methods often use neural networks as representatives of the latest developments in intelligence techniques. A key advantage of this approach is that the models do not require assumptions about the statistical distribution or properties of the data and can capture nonlinear relationships between the explanatory variables and default risk. Authors who have been recently dealing with this approach are, for example, Boyacioglu, Kara and Baykan [28] and Ioannidis, Pasiouras and Zopounidis [95].

In general, there is no overall agreement on what is the best statistical technique or method for building credit-scoring models. Approaches have been designed with regards to the details of the problem, the data structure, the characteristics used, the extent to which it is possible to segregate the classes by using those characteristics, and the objective of the classification (Hand and Henley [84]). However, more simple classification techniques, such as linear discriminant analysis and logistic regression, are generally considered to provide good results also in comparison to advanced statistical techniques, such as neural networks and fuzzy algorithms, and for the majority of the cases the results are not statistically different (Baesens, Van Gestel, Viaene, Stepanova, Suykens and Vanthienen [12]).

## 3.2 The FFIEC council & database

The Federal Financial Institutions Examination Council (FFIEC)[18] is a formal interagency body to prescribe uniform principles, standards, and report forms for the federal examination of financial institutions. It comprises the following five United States' federal banking regulators:

[18] http://www.ffiec.gov/

- the Board of Governors of the Federal Reserve System (FRB),
- the Federal Deposit Insurance Corporation (FDIC),
- the National Credit Union Administration (NCUA),
- the Office of the Comptroller of the Currency (OCC),
- the Consumer Financial Protection Bureau (CFPB).

The Council was established on March 10, 1979, pursuant to title X of the Financial Institutions Regulatory and Interest Rate Control Act of 1978 (FIRA), Public Law 95-630. It is responsible for developing uniform reporting systems for federally supervised financial institutions, their holding companies, and the nonfinancial institution subsidiaries of those institutions and holding companies. It also patronages the Home Mortgage Disclosure Act (HMDA), which provides public loan data, and the Community Reinvestment Act (CRA), which is intended to encourage depository institutions to help meet the credit needs of the communities in which they operate. The act also requires a periodical evaluation of each insured depository institution's record in helping meet the credit needs of the community, yielding a so-called CRA rating. This record is taken into account in considering an institution's application for deposit facilities, including mergers and acquisitions.

The data used in this study is collected from the FFIEC database. This publicly accessible database includes complete and detailed financial reports on financial institutions. Through the FFIEC Central Data Repository's Public Data Distribution web page[19], financial and structural information for most FDIC-insured institutions is available from March 31, 2001 onwards.

Tables 3.1 - 3.3 provide some descriptive statistics on the banks included in the FFIEC database. In particular, we report statistics as of December 31, 2013 on the institution type (Table 3.1) and the location of the banks (Table 3.2). We also report some descriptive statistics on the size of the banks (Table 3.3) as of December 31, 2006 until December 31, 2012.

[19] https://cdr.ffiec.gov/public/

**Table 3.1**
U.S. banks according to institution type

| Institution Type | Number | Percentage |
|---|---|---|
| Non-member Bank | 3,911 | 57% |
| National Bank | 1,150 | 17% |
| State Member Bank | 883 | 13% |
| Federal Savings Bank | 545 | 8% |
| State Savings Bank | 385 | 6% |
| Total number of banks | 6,877 | 100% |

The table shows the numbers and percentages of particular institution types contained in the FFIEC database as of December 31, 2013. Non-member Banks are represented by 57%, National Banks by 17%, State Member Banks by 13%, Federal Savings Banks by 8%, and State Savings Banks by 6%.

In Table 3.1, non-member Banks (defined as commercial banks, state charters and Fed non-members, supervised by the FDIC) are represented by 57% in the FFIEC database, followed by National Banks (defined as commercial banks, national/federal charters and Fed members, supervised by the OCC) by 17%, State Member Banks (defined as commercial banks, state charters and Fed members, supervised by the FRB) by 13%, Federal Savings Banks (defined as savings associations, state/federal charters, supervised by the OTS[20]) by 8%, and State Savings Banks (defined as savings banks, state charters, supervised by the FDIC) by 6%.

With regards to location[21] of the banks in the FFIEC database, Table 3.2 shows that the highest number is represented in Illinois (553) and Texas (536), the lowest number in Guam (3), Virgin Islands (2) and Federated States of Micronesia (1).

Table 3.3 illustrates the wide range of bank size included in the FFIEC database. The largest bank in 2012 in terms of total assets was JPMorgan Chase Bank, Columbus (OH) with approximately $1,897 billion in total assets. The mean value of total assets among banks included in the FFIEC database has increased from $1.3 billion in 2006 to $2 billion in 2012.

[20] As of June 30, 2011, the Office of Thrift Supervision (OTS) is no longer an active regulatory agency. It was merged with the OCC, FDIC, and CFPB as of July 21, 2011.
[21] The state in which the institution is physically located. The FDIC Act defines state as any State of the United States, the District of Columbia, and any territory of the United States, Puerto Rico, Guam, American Samoa, the Trust Territory of the Pacific Islands, the Virgin Island, and the Northern Mariana Islands.

**Table 3.2**
Number of U.S. banks in particular states

| State / Country | Number | State / Country | Number | State / Country | Number |
|---|---|---|---|---|---|
| Alabama | 139 | Kentucky | 189 | Ohio | 228 |
| Alaska | 5 | Louisiana | 144 | Oklahoma | 230 |
| Arizona | 24 | Maine | 30 | Oregon | 29 |
| Arkansas | 118 | Maryland | 73 | Pennsylvania | 200 |
| California | 222 | Massachusetts | 154 | Puerto Rico | 6 |
| Colorado | 100 | Michigan | 128 | Rhode Island | 10 |
| Connecticut | 46 | Minnesota | 361 | South Carolina | 68 |
| Delaware | 24 | Mississippi | 87 | South Dakota | 76 |
| District of Columbia | 4 | Missouri | 315 | Tennessee | 182 |
| Fed. St. of Micronesia | 1 | Montana | 65 | Texas | 536 |
| Florida | 194 | Nebraska | 208 | Utah | 55 |
| Georgia | 225 | Nevada | 19 | Vermont | 13 |
| Guam | 3 | New Hampshire | 21 | Virgin Islands | 2 |
| Hawaii | 9 | New Jersey | 103 | Virginia | 103 |
| Idaho | 14 | New Mexico | 46 | Washington | 62 |
| Illinois | 553 | New York | 167 | West Virginia | 62 |
| Indiana | 131 | North Carolina | 74 | Wisconsin | 263 |
| Iowa | 336 | North Dakota | 89 | Wyoming | 34 |
| Kansas | 296 | | | | |

The table reports the numbers of banks contained in the FFIEC database as of December 31, 2013 sorted by location, i.e. the state in which the institution is physically located. The highest number of banks is located in Illinois (553) and Texas (536), while the lowest number is located in Guam (3), Virgin Islands (2) and Federated States of Micronesia (1).

Through the FFIEC CDR web page Reports of Condition and Income (Call Report) data can be obtained for individual institutions. The Uniform Bank Performance Reports (UBPR) are also available online. The UBPR is an analytical tool created for bank supervisory, examination, and management purposes. In a concise format, it shows the impact of management decisions and economic conditions on a bank's performance and balance-sheet composition. The performance and composition data contained in the report can be used as an aid in evaluating the adequacy of earnings, liquidity, capital, asset and liability management, and growth management.

The UBPR is produced for every commercial and savings bank insured by the FDIC. The report is computer-generated from a database derived from public and non-public sources. It contains several years' worth of data, which is updated quarterly. This data is presented in the form of ratios, percentages, and dollar amounts computed mainly from Call Reports submitted by the bank. Each UBPR also contains corresponding average data for the bank's peer group and percentile rankings for most ratios. The UBPR therefore permits evaluation of

**Table 3.3**
Mean value and percentiles of total assets (in $000) for U.S. banks

| | Total Assets (in $000) | | | | | |
|---|---|---|---|---|---|---|
| Date | Mean | q (0.05) | q (0.25) | q (0.50) | q (0.75) | q (0.95) |
| 31/12/2006 | $1,313,590 | $19,451 | $56,060 | $120,735 | $287,388 | $1,447,302 |
| 31/12/2007 | $1,475,439 | $20,659 | $58,930 | $127,647 | $299,173 | $1,509,570 |
| 31/12/2008 | $1,667,495 | $23,132 | $64,585 | $136,231 | $315,951 | $1,583,903 |
| 31/12/2009 | $1,662,707 | $24,446 | $70,023 | $146,715 | $331,758 | $1,602,073 |
| 31/12/2010 | $1,774,114 | $25,848 | $73,627 | $148,954 | $328,321 | $1,641,129 |
| 31/12/2011 | $1,919,396 | $27,234 | $76,460 | $155,546 | $346,476 | $1,745,165 |
| 31/12/2012 | $2,022,651 | $28,840 | $81,599 | $165,704 | $368,499 | $2,017,260 |

The table reports descriptive statistics on the size of the banks (mean value and chosen percentiles of total assets in thousands of dollars) contained in the FFIEC database from year 2006 to 2012. The mean of total assets among considered FFIEC banks has increased from $1,314 million in 2006 to $2,023 million in 2012.

a bank's current condition, trends in its financial performance, and comparisons with the performance of its peer group.

## 3.3 Credit-scoring and model evaluation techniques

In this section, we review the techniques of logistic and probit regression as representatives of credit-scoring models. Subsequently, we describe static and dynamic discrete hazard models that will be applied in the empirical analysis. The section also reviews a number of model evaluation techniques such as ROC analysis, bootstrapping, calibration accuracy tests and the use of nonparametric techniques such as the Kruskal-Wallis test and Tukey's multiple comparison procedure for comparison of model performance.

### 3.3.1 Logistic and probit regressions

Logistic and probit regressions are multivariate techniques that belong to the class of probabilistic statistical classification models and have been heavily used for credit scoring, see, e.g., Martin [132], West [176], Logan [118], Shumway [157]. They are typically used to predict a binary response based on one or more predictor variables and allow for estimation of the probability for the occurrence of an event using a set of independent variables. In credit scoring, the studied event is the default or credit failure of a corporation or, in our case, of a financial institution. Thus, the response variable $y_i$ takes on the value $y_i = 1$ if bank $i$ failed (with probability $\mathrm{PD}_i$), and $y_i = 0$ otherwise (with probability $1 - \mathrm{PD}_i$). We are interested in modeling the probability $\mathrm{PD}_i$ for the occurrence of a default event by specifying the following model:

$$\mathrm{PD}_i = f\left(\alpha + \sum_{j=1}^{n} \beta_j x_{i,j}\right), \tag{3.1}$$

where $x_{i,j}$ denotes particular explanatory variable of $i$-th bank used to forecast the probability of default and $\alpha, \beta_j$ are the estimated parameters of the model. The right-hand side of the equation (3.1) then enters into a distribution function, depending on a given model.

The literature suggests various ways to specify the probability $\mathrm{PD}_i$. In our study, we will concentrate on the application of logistic and probit regressions, also referred to as logit and probit models. For the logit model, the so-called logistic transformation

$$\mathrm{PD}_i = \frac{\exp\left(\alpha + \sum_{j=1}^{n} \beta_j x_{i,j}\right)}{1 + \exp\left(\alpha + \sum_{j=1}^{n} \beta_j x_{i,j}\right)} = \frac{1}{1 + \exp\left(-\alpha - \sum_{j=1}^{n} \beta_j x_{i,j}\right)} \tag{3.2}$$

is applied. For the probit model, the cumulative distribution function of the normal distribution is used:

$$\mathrm{PD}_i = \int_{-\infty}^{\alpha + \sum_{j=1}^{n} \beta_j x_{i,j}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} t^2\right) dt. \tag{3.3}$$

Due to nonlinear nature of these models it is necessary to use maximum likelihood estimation in order to obtain the model parameters. Given $\mathrm{PD}_i$ and assuming that defaults are independent, we can express the maximum likelihood function as follows:

$$L = \prod_{i=1}^{n} \mathrm{PD}_i^{y_i} \left(1 - \mathrm{PD}_i\right)^{1 - y_i}. \tag{3.4}$$

Since it is easier to maximize a summation rather than a product, it is a common practise to work with the logarithm of the maximum likelihood function:

$$\ln L = \sum_{i=1}^{n} y_i \ln \mathrm{PD}_i + \sum_{i=1}^{n} \left(1 - y_i\right) \ln \left(1 - \mathrm{PD}_i\right). \tag{3.5}$$

Therefore, by combining (3.5) and (3.2) we get the logarithm of the maximum likelihood function for the logit model as follows:

$$\begin{aligned}\ln L = &\sum_{i=1}^{n} y_i \ \ln\left[\frac{1}{1+\exp\left(-\alpha-\sum_{j=1}^{n}\beta_j x_{i,j}\right)}\right] \\ &+\sum_{i=1}^{n}(1-y_i)\ \ln\left[1-\frac{1}{1+\exp\left(-\alpha-\sum_{j=1}^{n}\beta_j x_{i,j}\right)}\right],\end{aligned} \tag{3.6}$$

and by combining (3.5) and (3.3) we get the logarithm of the maximum likelihood function for the probit model as follows:

$$\begin{aligned}\ln L = &\sum_{i=1}^{n} y_i \ \ln\left[\int_{-\infty}^{\alpha+\sum_{j=1}^{n}\beta_j x_{i,j}} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right)dt\right] \\ &+\sum_{i=1}^{n}(1-y_i)\ \ln\left[\int_{-\infty}^{\alpha+\sum_{j=1}^{n}\beta_j x_{i,j}} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right)dt\right].\end{aligned} \tag{3.7}$$

For model evaluation or testing the significance of a model, log-likelihood ratio test or Wald test can be used, see, e.g., Tabachnick and Fidell [166] or Hosmer Jr and Lemeshow [91]. Logit and probit models typically provide rather similar results. However, one of the main differences between the techniques is that the logistic function exhibits heavier tails (see, e.g., Trück and Rachev [169]).

## 3.3.2 Static and discrete hazard models

Static (single-period) models, even though widely used in the past, may not fully appropriate for estimation of default probabilities, see Shumway [157] or Hillegeist, Keating, Cram and Lundstedt [87]. Firstly, there are often multiple-period data sets available. Since static models only consider one set of explanatory variables for each bank, they neglect the fact that the characteristics of most banks change from year to year. Also, through time, various observations for the explanatory variables such as, e.g., financial ratios of a corporation or bank, become available and the question rises which of these observations should be included into the model to predict default events in an optimal way. A common practise is to use information on financial ratios one year prior to the default or non-default event what may actually introduce an unnecessary selection bias into the estimation process. Secondly, by ignoring the fact that banks and their performances change through time, static models produce inconsistent estimates of default probabilities and test statistics may subsequently be biased and provide incorrect inference.

On the other hand, dynamic discrete-time hazard (multi-period) models take advantage of multiple-period data sets and consider several observations on each bank that existed for some time throughout the sample period. Each bank either defaults during the sample period, survives, or may leave the sample for a reason other than default (for example a merger, takeover, or if the bank failed to provide financial ratios, etc.). Unlike static models, hazard models are also able to incorporate explanatory variables that change through time (time-varying covariates). Therefore, the approach also allows for the inclusion of additional macroeconomic or market indicators into the model. Clearly, these variables typically take on the same value for all banks at a given point of time, but may provide some additional explanatory power through time.

The discrete hazard model estimates the PDs as:

$$\mathrm{PD}_{i,t} = \mathrm{P}\left(y_{i,t} = 1\right) = \mathrm{E}\left(y_{i,t} \mid x_{i,t}\right), \tag{3.8}$$

where $\mathrm{PD}_{i,t}$ denotes the probability that bank $i$ will default in period $t$, conditional on surviving until the end of period $t-1$ and on the observed covariates $x_{i,t}$. These covariates represent bank-specific independent variables that are observable at the beginning of period $t$. The response variable $y_{i,t}$ equals one if bank $i$ defaults in period $t$, and equals zero otherwise. The discrete hazard model has then the following form:

$$\mathrm{PD}_{i,t} = f\left[\alpha(t) + \sum_{j=1}^{n} \beta_j x_{i,t,j}\right], \tag{3.9}$$

where $\alpha(t)$ is a time-varying, system-wide variable that captures the baseline hazard rate. Again, the link function $f$ may be specified in various ways, for example using a logit or probit model. Note the two key differences between equations (3.1) and (3.9): first, in model (3.9), the constant $\alpha$ is replaced by the baseline hazard rate $\alpha(t)$, while, secondly, the subscript $t$ reflects the use of multiple bank-year observations of data for the same bank $i$.

Furthermore, hazard models treat all observations of a particular bank as dependent observations.[22] Therefore, an adjustment of the sample size to account for the lack of independence between bank-year observations is necessary for calculating correct test statistics of coefficients. In fact, hazard models often produce different statistical inferences.[23]

## 3.3.3 Evaluation techniques

Once an appropriate model has been identified, the performance of the model can be validated across a variety of criteria. This section outlines approaches to model validation as they have been suggested in the literature for credit-scoring models such as ROC curve analysis or likelihood based measures. We also suggest a number of possible directions for new validation techniques. In particular, we suggest the use of econometric techniques that provide statistical power to distinguish between models that provide relatively similar results. Particularly we propose nonparametric techniques such as the Kruskal-Wallis test and Tukey's multiple comparison procedure (see Hochberg and Tamhane [88]). We further propose the use of Vuong's closeness test (see Vuong [174]) that is based on comparing the log-likelihood of non-nested models. Finally, we propose the use of the Hosmer-Lemeshow chi-squared goodness-of-fit test (see Hosmer Jr and Lemeshow [91]) that allows for a comparison of the predicted and actually observed default frequencies for sub-groups of the entire sample.

### 3.3.3.1 ROC analysis

ROC (*relative* or *receiver operating characteristic*) curves, see, e.g., Green and Swets [77], Hanley [85], Hosmer Jr and Lemeshow (2004), Stein [162], are among the most powerful tools to quantify the predictive power and are widely used for evaluation of credit default models. A ROC curve plots the Type II error against one minus the Type I error. Unlike contingency

---

[22] This is a unique difference between hazard and "pooled" models that are estimated with data on each bank in each year of its existence as if each bank-year observation was an independent observation ("pooled" models treat each bank-year as a separate observation).

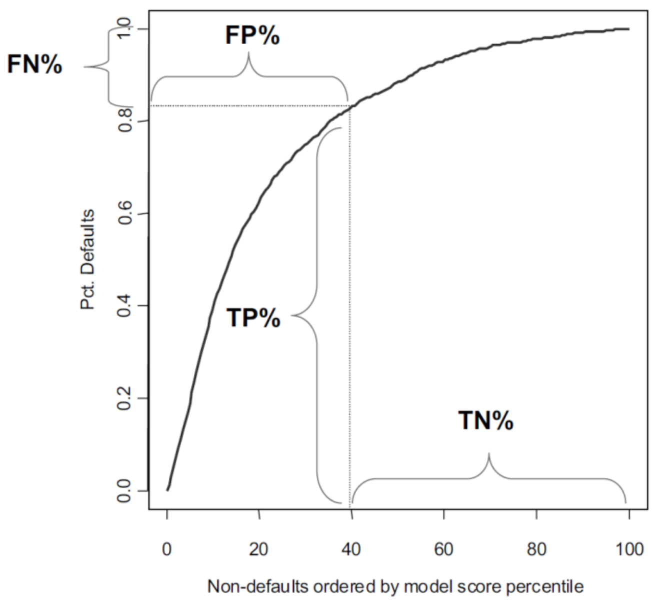
[23] For more detailed discussion on discrete-time hazard models and their econometric properties we refer to Shumway [157].

table analysis, where a specific model cut-off point needs to be chosen, ROC curves provide information on the performance of a model at any cut-off point that might be chosen. The ROC curve is also known as the trade-off curve, because it shows the trade-off between 'goods' and 'bads' – the percentage of total bads that must be accepted in order to accept a given percentage of total goods.

An example of the ROC curve is given in Figure 3.1, where a TP (true positive) is a predicted default that actually occurs; a TN (true negative) is a predicted non-default that actually occurs (the company does not default); a FP (false positive) is a predicted default that does not occur, and a FN (false negative) is a predicted non-default where the company actually defaults. The errors of the model are FN and FP shown on the off diagonal, where FN represents a Type I error and FP represents a Type II error.

A convenient measure for summarizing the ROC curve is the area under the curve (the ROC area), which is calculated as the integral of the ROC curve: the proportion of the area below the ROC curve relative to the total area of the unit square. A value of 0.5 indicates a random model, while a value of 1 indicates perfect discrimination. A similar measure, the accuracy ratio (AR), can also be calculated and Engelmann, Hayden and Tasche [57] provide the following identity relationship between the ROC area and the AR: $\text{AR} = 2(\text{ROC area} - 0.5)$.

**Figure 3.1**
An example of the ROC curve




Source: Stein [162], page 82

#### 3.3.3.2 Bootstrapping, Kruskal-Wallis and Tukey's multiple comparison test

Since the results of model testing are subject to sample variability, one may also be interested in conducting a variety of resampling techniques such as, e.g., bootstrapping (Efron and Tibshirani [56]) which allows to leverage the available data and reduce the dependence on the particular sample. As described in Stein [162], a typical resampling technique proceeds as follows. From the result set, a sub-sample is selected at random. The performance measure of interest (e.g. ROC area) is calculated for this sub-sample and recorded. Another sub-sample is then drawn and the process is repeated. This continues for many repetitions until a distribution of the performance measure is established. The sampling distribution is used to calculate statistics of interest (standard error, percentiles of the distribution, etc.).

For testing whether the performance measures calculated from bootstrapping are significantly different among the various models, we suggest to use a nonparametric Kruskal-Wallis test or Tukey's mutiple comparison test. The Kruskal-Wallis test is a nonparametric version of the classical one-way analysis of variance (ANOVA), and tests the null hypothesis that all samples are drawn from the same population, or equivalently, from a different population with the same distribution (Hollander and Wolfe [90]). Rejecting the null hypothesis means that at least one of the samples stochastically dominates at least one other sample. Unlike a standard one-way ANOVA, the test does not require the assumption that all samples come from a population with a normal distribution.

To perform the test, we have to put the data in ascending order and write down the ranking of each observation in the sample. Specifically, let group $j$, where $j=1,\dots,k$, have $n_j$ observations and $n=n_1+n_2+\dots+n_k$ be the total number of observations. We put all of the observations into one big group, and rank them, with the rank of 1 for the smallest observation and the rank of $n$ for the largest one. We need to keep track of which observation and rank goes with which of the $k$ groups. In the case of tied observations we average the ranks. Finally, we add up the ranks for each separate group and denote the rank sum for group $j$ by $T_j$. The Kruskal-Wallis statistic $K$ is then given by the following expression:

$$K=\left[\frac{12}{n\left(n+1\right)}\sum_{j=1}^{k}\frac{T_j^2}{n_j}\right]-3\left(n+1\right). \tag{3.10}$$

This test is always one-sided and its statistic is chi-squared distributed (under the assumption that $n_j \geq 5$) with $k-1$ degrees of freedom. Note that the Kruskal-Wallis statistic $K$ is an omnibus test statistic and cannot tell you which specific groups of independent variable are statistically different from each other. It only tells you that at least two groups are different.

This drawback is overcome by the Tukey's test (Hochberg and Tamhane [88]). It is a multiple comparison procedure which allows to further investigate which of the samples are significantly different. The test uses Tukey's honestly significant difference (Tukey's HSD) criterion, that is optimal for the comparison of groups with equal sample sizes, to test for significant differences with respect to the performance of the various models. It basically compares the means of every treatment to the means of every other treatment. Therefore, the test is simultaneously applied to the set of all pairwise comparisons $\mu_i - \mu_j$ and identifies any difference between two means that is greater than the expected standard error.

Tukey's test is based on a formula very similar to that of the $t$-test. In fact, Tukey's test is essentially a $t$-test, except that it corrects for experiment-wise error rate. When there are multiple comparisons being made, the probability of making a Type I error increases – Tukey's test corrects for that, and is thus more suitable for multiple comparisons than doing a number of $t$-tests would be (Linton and Harder [116]). The formula for Tukey's statistic is:

$$q_s = \frac{Y_A - Y_B}{SE}, \tag{3.11}$$

where $Y_A$ is the larger of the two means being compared, $Y_B$ is the smaller of the two means being compared, and $SE$ is the standard error of the data in question. The value of the test statistic can then be compared to a cut-off value from the studentized range distribution.

#### 3.3.3.3 Calibration accuracy test and Vuong's closeness test

The second dimension within validating credit models (after examination of a models' power) is model calibration. Calibration examines how well the estimated model PDs match with actual outcomes. Using a calibration accuracy test and its likelihood estimates, we are able to determine which model's PDs (from a set of candidate models) are closest to actual PDs given a set of empirical data. We refer to Stein [162] for a closer discussion about calibration and likelihood-based measures of calibration. The higher the likelihood the more accurate is the model in predicting default probabilities.

If a model predicts a binary event (default/no default), its estimate of the probability for the occurrence of a single event $y$ given data $x$ is

$$prob(y|x) = p(x)^y \left[1 - p(x)\right]^{(1-y)}, \tag{3.12}$$

where $p(x)$ is the PD predicted by the model, conditional on the input variables $x$, while the event $y$ is defined as one if the bank defaults and zero otherwise. Using these two inputs (a vector of estimated PDs and a vector of default outcomes) the likelihood measure $L$ for the estimated model can then be calculated as follows:

$$L = \prod_{i=1}^{n} prob(y_i | x_i) = \prod_{i=1}^{n} p(x_i)^{y_i} \left(1 - p(x_i)\right)^{(1-y_i)}. \tag{3.13}$$

Since, in general, it is more convenient to work with summations than products, by convention we work with the log of the likelihood $\ell$ defined as:

$$\ln L = \ell = \sum_{i=1}^{n} y_i \ln\left[p(x_i)\right] + (1 - y_i)\ln\left[1 - p(x_i)\right]. \tag{3.14}$$

In order to determine whether calculated log-likelihoods for various models are significantly different, we can use the Vuong's closeness test for non-nested models, see Vuong [174]. It is a likelihood-ratio based test for model selection using the Kullback-Leibler information criterion that makes probabilistic statements about two models that can be nested, non-nested or overlapping. The test examines the null hypothesis that the two models are equally close to the actual one, against the alternative hypothesis that one model is closer.

With non-nested models and i.i.d. exogenous variables, model A is preferred with significance level $\alpha$, if the $Z$ statistic exceeds the positive (falls below negative) $(1-\alpha)$-quantile of the standard normal distribution. The $Z$ statistic is defined as:

$$Z = \frac{LR_N\left(\beta_{ML,A}, \beta_{ML,B}\right)}{\sqrt{N}\omega_N}, \tag{3.15}$$

where

$$LR_N\left(\beta_{ML,A},\beta_{ML,B}\right) = L_N^A - L_N^B - \frac{K_A - K_B}{2}\log N. \tag{3.16}$$

Hereby, $L_N^j$ denotes the log-likelihood of model $j$, $K_j$ is the number of parameters in model $j$, and $N$ is the number of observations. The denominator in the expression for $Z$, $\omega_N$, is defined by setting $\omega_N^2$ equal to either the mean of the squares of the pointwise log-likelihood ratios $\ell_i$, or to the sample variance of these values, where

$$\ell_i = \log\frac{f_A\left(y_i \mid x_i,\beta_{ML,A}\right)}{f_B\left(y_i \mid x_i,\beta_{ML,B}\right)}. \tag{3.17}$$

#### 3.3.3.4 Hosmer-Lemeshow test

Hosmer-Lemeshow's chi-squared goodness-of-fit test (Hosmer Jr and Lemeshow [91]) is a test based on grouping the values of the estimated probabilities. It consists of dividing the ranked predicted probabilities into $k$ groups (probabilities are often divided based on deciles, such that $k=10$) and computing the Pearson chi-squared statistic that compares the predicted and actually observed frequencies in a $2\times k$ contingency table. The $HL$ test statistic follows a chi-squared distribution with $k$ degrees of freedom[24],

$$HL = \sum_{i=1}^{k}\left[\frac{\left(O_i^{ND} - E_i^{ND}\right)^2}{E_i^{ND}} + \frac{\left(O_i^{D} - E_i^{D}\right)^2}{E_i^{D}}\right], \tag{3.18}$$

where $O_i^{ND}$ is the observed number of non-defaults in group $i$ and $E_i^{ND}$ is the expected (predicted) number of non-defaults based on the model. Similarly, $O_i^D$ is the observed number of defaults in group $i$ and $E_i^D$ is the expected number according to the estimated

[24] In general, for a $j\times k$ contingency table there are $\left(j-1\right)\left(k-1\right)$ degrees of freedom in the Pearson chi-squared statistic, which implies $k-1$ degrees of freedom in our case. However, in case of the out-of-sample validation the distribution, if we use $k$ groups, is $\chi^2\left(k\right)$.

model. [25] The closer the agreement between the observed and expected values, the smaller will be the value of the $HL$ test statistic, which indicates a good fit to the data and, therefore, good overall model fit.

The appropriateness of the $p$-value calculated using the $HL$ statistic depends on the validity of the assumption that the estimated expected frequencies are large. In general, all expected frequencies should be greater than 5, what might pose a problem for sub-groups with very low probabilities of default. The advantage of a summary goodness-of-fit statistic like $HL$ is that it provides a single, easily interpretable value that can be used to assess the model fit. The disadvantage is that in the process of grouping the data, we may miss important information on the deviation of model probabilities and actual occurrences of defaults, due to a small number of individual data points (Hosmer Jr and Lemeshow [91]). Tables listing the observed and estimated expected frequencies in each decile contain valuable descriptive information for assessing the adequacy of the fitted model over the deciles. Comparison of the observed and expected frequencies within each cell then may indicate regions where the model does not perform satisfactory.

## 3.4 Application and results

The following section provides empirical results of the study. We first describe the data used in our analysis, in particular, the number of observations in our sample for each year. Then we provide results on the estimated credit-scoring models, i.e. static and discrete hazard models based on logistic and probit regressions. Finally, all of the estimated models are validated on control samples. Following Stein [162], we apply a rolling window methodology with out-of-time validation within estimating and validating the models.

### 3.4.1 Data description

As mentioned in the introduction, there were in total 492 defaults of commercial banks in the U.S. from February 2, 2007 to December 31, 2013. A defaulted (failed) bank can be defined in

---

[25] Alternatively, the $HL$ test statistic might be defined as $HL = \sum_{i=1}^{k} \frac{n_i \left(\hat{\pi}_i - \pi_i\right)^2}{\pi_i \left(1-\pi_i\right)}$, where $\hat{\pi}_i$ are observed default rates, $\pi_i$ are corresponding expected rates, $n_i$ are the number of observations in group $i$ and $k$ is the number of groups for which frequencies are being analysed.

a variety of ways. In our study a defaulted bank is defined as a financial institution which has been closed by a federal or state regulator.

For collection of financial ratios we use the FFIEC database, in particular, we use ratios acquired from UBPR reports in the database. Table 3.4 provides a comparison of the number of banks used for estimation of the models in this study[26] and the total number of banks in the U.S.[27], along with the number of defaulted banks in the FFIEC database and the U.S. in total for particular years within the sample period. There are two reasons why the numbers of banks in the FFIEC database are lower: (a) Savings & Loan Associations are not included in the FFIEC database, (b) for some banks the data in the FFIEC database is available rather later.

For the applied hazard models, we use a rolling window methodology (the walk-forward approach) with out-of-time validation as it is closest to the actual application of default prediction models in practise and gives a realistic view of how a particular model would perform over time. We refer to Stein [162] for a more thorough discussion of this approach.

An important question is what time lag should be taken into account between the observation of balance sheet data and the default event, when compiling a database of financial indicators for defaulted and non-defaulted banks. A common practise is to use at least a one year lag. To ensure that financial ratio values are collected at least one year prior

**Table 3.4**
Comparison of the number of banks (FFIEC vs. U.S. in total)

| # of banks | | | # of defaulted banks | | |
|---|---|---|---|---|---|
| Date | FFIEC database | U.S. in total | Model / Year | FFIEC database | U.S. in total[28] |
| 31/12/2006 | 7,768 | 8,691 | 2008 | 19 | 25 |
| 31/12/2007 | 7,579 | 8,544 | 2009 | 120 | 140 |
| 31/12/2008 | 7,261 | 8,314 | 2010 | 138 | 158 |
| 31/12/2009 | 6,996 | 8,021 | 2011 | 86 | 92 |
| 31/12/2010 | 6,799 | 7,666 | 2012 | 42 | 51 |

This table shows a comparison of the total number of banks and the number of defaulted banks between the FFIEC database and the actual number of banks in the U.S. for the time period 2006-2010. Note that, for example, for 2008 model (based on 19 defaulted banks in 2008) we use balance sheet data from 31/12/2006.[29]

[26] The exact number of banks used within estimation of models depends on a particular type of a model (static vs. hazard) and a particular year, and is specified for each estimated model in Tables 3.7 and 3.8.
[27] http://www.usbanklocations.com/bank-rank/total-assets.html
[28] See Footnote 17.
[29] For detailed explanation, see Section 3.4.2.

**Table 3.5**
Data collection dates for the models

| MODEL | Defaulted banks | Data collection dates (financial ratios) |
|---|---|---|
| 2008 | defaulted in year 2008 | 31/12/2006 |
| 2009 | defaulted in years 2008 + 2009 | 31/12/2006<br>31/12/2007 |
| 2010 | defaulted in years 2008 + 2009 + 2010 | 31/12/2006<br>31/12/2007<br>31/12/2008 |
| 2011 | defaulted in years 2008 + 2009 + 2010 + 2011 | 31/12/2006<br>31/12/2007<br>31/12/2008<br>31/12/2009 |
| 2012 | defaulted in years 2008 + 2009 + 2010 + 2011 + 2012 | 31/12/2006<br>31/12/2007<br>31/12/2008<br>31/12/2009<br>31/12/2010 |

The table reports data collection dates for individual models. For example, for 2012 discrete hazard models (based on defaults from 2008 to 2012) we use balance sheet data up to 31/12/2010.[30]

to the default event, we use a 12-24 months horizon before the actual default. For example, for banks defaulted in 2008 balance sheet data and financial ratios are collected on December 31, 2006. Data collection dates for the individual models are summarized in Table 3.5. The financial ratios used as explanatory variables for the estimation of the scoring models are provided in Table 3.6.

We decided to examine a total of nineteen financial ratios that are expected to describe the financial health of a bank. In particular, we use indicators describing the profitability, efficiency, liquidity, assets quality and capital adequacy of a bank.[31] Chosen variables come from the FFIEC database and are often used by regulators for comparison purposes. To

[30] Again, see Section 3.4.2 for detailed explanation.

[31] In addition to these financial indicators we also included four macroeconomic (GDP growth, unemployment and inflation rate, difference between 10-year and 3-month Treasury Bill rates) and five market indicators (the VIX index, the TED spread, excess returns on NASDAQ, KBW and Dow Jones U.S. bank indices) to better reflect the economic situation on the market. Since within inclusion of these variables the collinearity issues arise, we used these variables only within hazard model 2012 as we already had 5 different observations. However, inclusion of these variables did not significantly improve the performance of the model. This is not a real surprise as we worked only with 5-year time period and would need a few credit cycles covered to benefit from macroeconomic variables (market indicators proved to be statistically insignificant in our model).

**Table 3.6**
List of explanatory variables

| Indicator | Description | Indicator's Group |
|---|---|---|
| $x_1$: ROA | Return on Assets (%) | Profitability |
| $x_2$: ROE | Return on Equity (%) | Profitability |
| $x_3$: NIM | Net Interest Margin (%) | Profitability |
| $x_4$: IE_II | Total Interest Expense / Total Interest Income (%) | Profitability |
| $x_5$: II_EA | Total Interest Income / Interest Earning Assets (%) | Profitability |
| $x_6$: C_IBD | Cost of Total Interest Bearing Deposits (%) | Profitability |
| $x_7$: ER | Efficiency Ratio (%) | Efficiency |
| $x_8$: NLL_TA | Net Loans & Leases / Total Assets (%) | Liquidity |
| $x_9$: LP_AA | Provision for Loan & Lease Losses / Average Assets (%) | Assets Quality |
| $x_{10}$: NL_TLL | Net Loss / Average Total Loans & Leases (%) | Assets Quality |
| $x_{11}$: NCRLL_GLL | Noncurrent Loans & Leases / Gross Loans & Leases (%) | Assets Quality |
| $x_{12}$: LLA_TLL | Loans & Leases Allowance / Total Loans & Leases (%) | Assets Quality |
| $x_{13}$: EQ_TA | Total Equity Capital & Minority Interests / Total Assets (%) | Capital Adequacy |
| $x_{14}$: TD_EQ | Total Deposits / Total Equity Capital & Minority Interests (times) | Capital Adequacy |
| $x_{15}$: RE_EQ | Retained Earnings / Average Total Equity Capital (%) | Capital Adequacy |
| $x_{16}$: NLL_EQ | Net Loans & Leases / Total Equity Capital (times) | Capital Adequacy |
| $x_{17}$: T1RBC_RWA | Tier One Risk-Based Capital / Risk-Weighted Assets (%) | Capital Adequacy |
| $x_{18}$: TRBC_RWA | Total Risk-Based Capital / Risk-Weighted Assets (%) | Capital Adequacy |
| $x_{19}$: T1LC | Tier One Leverage Capital Ratio (%) | Capital Adequacy |

The table lists financial ratios used as explanatory variables within estimation of particular models. There are nineteen financial ratios in total that describe a financial health of banks. These ratios are divided into five indicator groups (profitability, efficiency, liquidity, assets quality, and capital adequacy).

examine whether particular variables affect PDs in a way they are supposed to, we initially conduct a univariate regression. Based on this analysis, the following variables yielded an opposite sign than expected under the economic hypothesis, and therefore were removed: $x_5$**:** Total Interest Income / Interest Earning Assets (%), $x_9$**:** Provision for Loan & Lease Losses / Average Assets (%), and $x_{12}$**:** Loans & Leases Allowance / Total Loans & Leases (%).

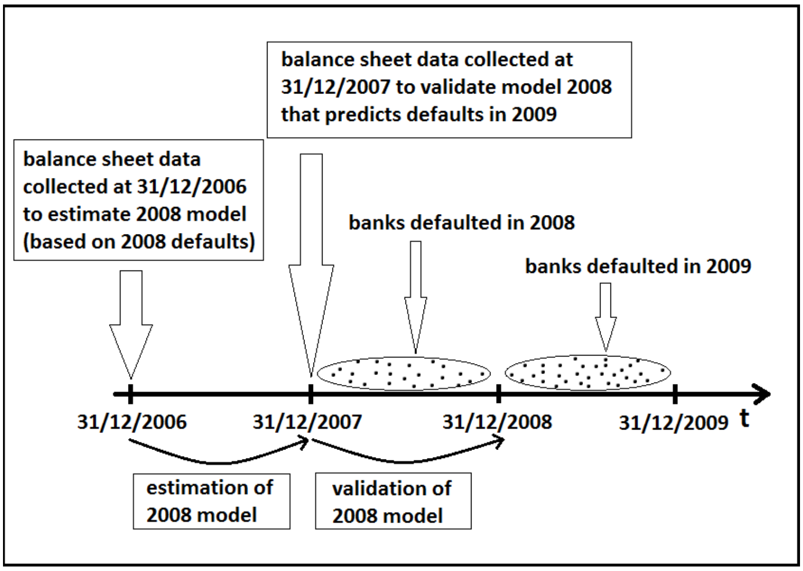
We also closely investigated outliers among the observations of financial ratios. To ensure that statistical results are not heavily influenced by outliers or errors in the collected data, we decided to use winsorized data and set all observations for the considered financial ratios that exceeded the 99$^{th}$ percentile or were below the 1$^{st}$ percentile equal to these values.[32]

[32] The same approach was used, for instance, by Shumway [157].

### 3.4.2 Model estimation

As mentioned earlier, we estimate both static and dynamic discrete hazard models using equations (3.1) and (3.9), respectively. For each of these approaches, we apply logistic and probit regressions in order to calibrate the models. Note that while estimating, for example, the 2010 static model, we used banks defaulted in 2010 and financial ratios collected on December 31, 2008. While validating this model (with the parameters we have estimated), we are predicting defaults in 2011 using financial ratios collected on December 31, 2009. This process is described in more detail for estimation of the 2008 static model in Figure 3.2. On the other hand, for the 2010 discrete hazard model, we use data on defaulted and non-defaulted banks in 2008, 2009, and 2010 and financial ratios collected on December 31, 2006, December 31, 2007 and December 31, 2008 to estimate the models (see Table 3.5). Clearly, data used for model validation is the same as for the static models, i.e., we validate the 2010 hazard models by predicting defaults in 2011 using financial ratios collected on December 31, 2009.

**Figure 3.2**
Illustration of data used for estimation and validation of the 2008 static model




This figure describes the data used for estimation and subsequent validation of 2008 static model. For estimation of this model (based on banks defaulted in 2008) we used financial ratios collected at December 31, 2006. Within subsequent validation of this model (with the parameters we have estimated), we are predicting defaults in 2009 using financial ratios collected at December 31, 2007.

Typically models are estimated based on a list of variables that are thought to be relevant in explaining default events, while the actual choice of the variables is often based on individual judgment of an analyst in an iterative procedure. The selection decision is usually based on the statistical significance and relative contribution of each independent variable, the evaluation of inter-correlations between the relevant variables, observations on the predictive accuracy of the various profiles, and individual judgment of the analysts, see, e.g., Altman [4]. The iterative procedure is finalized when adding another variable could not significantly improve the results (Altman, Haldeman and Narayanan [6]). Alternatively, for example, a stepwise regression technique could be applied, see, e.g., Kolari, Glennon, Shin and Caputo [104], that provides an algorithm for inclusion of relevant variables that is purely based on statistical significance of the variables and improved estimation results for the model.

We applied stepwise regression initially, but found that several of the statistically significant variables did not provide the expected signs for the estimated coefficients. While models based on stepwise regression may provide a good in-sample fit with high explanatory or discriminative power, they may suffer from poor interpretation of the estimated coefficients and often fail to provide good results in an out-of-sample environment. Therefore, we decided to apply the following algorithm instead of using a method that is purely guided by statistical significance. For the estimation of the models, we start with one variable only from each of the indicator groups, i.e. we include the variable with the highest explanatory power in the univariate regression for the categories profitability, efficiency, liquidity, asset quality, and capital adequacy. Subsequently, we examine whether all these variables have the correct sign and are statistically significant also in the multivariate model. In case that a particular variable had the incorrect sign or was statistically insignificant, we replaced this variable by another variable from the same group of indicators, namely by the one with the second highest explanatory power in the univariate model. Once all the chosen variables had the correct sign and were statistically significant, we tried to add additional variables to the model. Note that the inclusion of new variables may cause a change in statistical significance and possibly also a change in the sign of the coefficients for variables previously included into the model. Therefore, the algorithm stops when no additional variable with a meaningful contribution to the model's explanatory power could be added to the model.

All explanatory variables eventually used in the models are not highly correlated with each other.[33] Overall, the majority of included variables is rather similar for the different years and typically provides a good mixture of financial ratios relating to profitability, liquidity, asset quality and capital adequacy. Also, among the models that employ the same variables for different years, the coefficients often change only marginally over time which implies that our models might be considered robust and stable.

Results for the estimated static logit and probit models for 2008 to 2012 are summarized in Table 3.7. The table contains information on the included variables, estimated coefficients, statistical significance and standard errors for the coefficients, log-likelihood of the model, pseudo-$R^2$ value[34], and statistical significance of the entire model (likelihood-ratio test[35]), as well as the total number of defaulted and non-defaulted banks used for the estimation. All estimated models contain between three and five explanatory variables, usually based on a subset of the following variables: $x_1$: Return on Assets (%), $x_2$: Return on Equity (%), $x_6$: Cost of Total Interest Bearing Deposits (%), $x_8$: Net Loans & Leases / Total Assets (%), $x_{11}$: Noncurrent Loans & Leases / Gross Loans & Leases (%), $x_{13}$: Total Equity Capital & Minority Interests / Total Assets (%), $x_{17}$: Tier One Risk-Based Capital / Risk-Weighted Assets (%), $x_{18}$: Total Risk-Based Capital / Risk-Weighted Assets (%), and $x_{19}$: Tier One Leverage Capital Ratio (%). Majority of these variables is statistically significant at the 1% level of significance. It is obvious that models' power is increasing over time (with pseudo-$R^2$ values starting at 0.2293 for the 2008 logit and 0.1983 for the 2008 probit models and ending up at 0.5263 for the 2012 logit and 0.5126 for the 2012 probit models) which will be confirmed by the ROC curve analysis during the out-of-sample validation.

[33] In all cases, correlation coefficient does not exceed 0.6.

[34] Pseudo-$R^2$ (specifically, the *McFadden's Pseudo R-Squared* is reported here) cannot be interpreted as an OLS coefficient of determination ( $R^2$ ) since calculation of maximum likelihood estimates is rather done through an iterative process and is not based on minimization of variance. Nevertheless, higher values still indicate a better model fit.

[35] We also conducted Wald tests, however, test statistics and $p$-values for the estimated models are not reported here as they yielded the same results as the conducted likelihood-ratio tests (all models are statistically significant at the 1% significance level).

**Table 3.7**

Estimated static models (years 2008 – 2012)

| | models 2008 | | models 2009 | | models 2010 | | models 2011 | | models 2012 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Logit | Probit | Logit | Probit | Logit | Probit | Logit | Probit | Logit | Probit |
| Intercept | -16.46*** | -6.85*** | -11.16*** | -4.83*** | -8.98*** | -5.17*** | -9.11*** | -4.62*** | -2.72* | -1.79*** |
| | (2.63) | (0.80) | (0.97) | (0.39) | (1.25) | (0.50) | (1.57) | (0.71) | (1.54) | (0.56) |
| $x_1$: ROA | | | | -5.70* | | | | | | |
| | | | | (3.46) | | | | | | |
| $x_2$: ROE | -2.27* | | | | -2.55*** | -1.35*** | -2.73*** | -1.34*** | -2.06*** | -1.03*** |
| | (1.36) | | | | (0.43) | (0.22) | (0.56) | (0.26) | (0.73) | (0.30) |
| $x_6$: C_IBD | 206.71*** | 60.47*** | 161.47*** | 62.84*** | 116.88*** | 53.70*** | 83.65*** | 41.04*** | | |
| | (45.00) | (13.63) | (19.89) | (8.34) | (18.38) | (8.59) | (23.93) | (11.10) | | |
| $x_8$: NLL_TA | 5.34* | 2.26*** | 2.17** | 0.93** | 3.86*** | 2.57*** | 5.33*** | 2.46*** | | |
| | (2.83) | (0.82) | (0.94) | (0.39) | (1.26) | (0.54) | (1.61) | (0.74) | | |
| $x_{11}$: NCRLL_GLL | 21.84* | 13.29*** | 28.32*** | 13.54*** | 25.97*** | 13.17*** | 21.01*** | 9.72*** | 21.69*** | 8.78*** |
| | (11.96) | (3.66) | (3.05) | (1.60) | (2.82) | (1.41) | (3.46) | (1.63) | (5.07) | (2.06) |
| $x_{13}$: EQ_TA | | | | | | | | | -57.82*** | -19.23*** |
| | | | | | | | | | (17.90) | (6.30) |
| $x_{17}$: T1RBC_RWA | | | | | -24.91*** | | | | | |
| | | | | | (4.82) | | | | | |
| $x_{18}$: TRBC_RWA | -9.08* | | | | | | | | | |
| | (2.63) | | | | | | | | | |
| $x_{19}$: T1LC | | | -17.10*** | -4.83*** | | -12.43*** | -34.58*** | -4.62*** | | |
| | | | (4.09) | (0.39) | | (2.59) | (9.78) | (0.71) | | |
| Log-likelihood | -102.68 | -106.82 | -467.69 | -465.40 | -389.70 | -386.99 | -228.24 | -228.89 | -121.04 | -124.54 |
| LR test (chi2 value) | 61.10 | 52.83 | 272.73 | 277.30 | 579.83 | 585.23 | 471.03 | 469.75 | 268.96 | 261.95 |
| LR test (Prob > chi2) | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** |
| Pseudo R2 | 0.2293 | 0.1983 | 0.2257 | 0.2295 | 0.4266 | 0.4306 | 0.5078 | 0.5065 | 0.5263 | 0.5126 |
| # of non-defaulted banks | 7,749 | | 7,459 | | 7,123 | | 6,910 | | 6,757 | |
| # of defaulted banks | 19 | | 120 | | 138 | | 86 | | 42 | |

The table reports results for the estimated logit and probit static models for the years 2008 to 2012. It provides information on estimated coefficients, their statistical significance and standard errors (numbers in parentheses), log-likelihood, pseudo-$R^2$ values, and statistical significance of the entire model (likelihood-ratio test), as well as the total number of defaulted and non-defaulted banks used within the estimation for each year. Variables are defined in Table 3.6. *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

**Table 3.8**
Estimated discrete hazard models (years 2009 – 2012)

| | models 2009 | | models 2010 | | models 2011 | | models 2012 | |
|---|---|---|---|---|---|---|---|---|
| | Logit | Probit | Logit | Probit | Logit | Probit | Logit | Probit |
| Intercept | -8.22*** | -3.30*** | -7.82*** | -2.93*** | -7.88*** | -3.03*** | -6.17*** | -3.09*** |
| | (0.72) | (0.26) | (0.69) | (0.17) | (0.63) | (0.15) | (0.35) | (0.13) |
| $x_1$: ROA | | -6.53* | | | | | | |
| | | (2.82) | | | | | | |
| $x_2$: ROE | | | -2.10*** | -1.02*** | -2.28*** | -1.08*** | -2.23*** | -1.11*** |
| | | | (0.33) | (0.15) | (0.27) | (0.12) | (0.24) | (0.11) |
| $x_6$: C_IBD | 132.55*** | 41.01*** | 91.32*** | 35.18*** | 96.48*** | 40.02*** | 104.36*** | 41.50*** |
| | (14.20) | (5.05) | (9.46) | (3.53) | (7.36) | (2.90) | (6.26) | (2.56) |
| $x_8$: NLL_TA | | | 2.11* | | 2.24* | | | |
| | | | (0.70) | | (0.65) | | | |
| $x_{11}$: NCRLL_GLL | 26.53*** | 11.91*** | 24.32*** | 11.40*** | 22.07*** | 10.15*** | 21.49*** | 9.76*** |
| | (2.68) | (1.38) | (1.99) | (0.93) | (1.64) | (0.77) | (1.54) | (0.72) |
| $x_{17}$: T1RBC_RWA | -15.50*** | -5.72*** | -15.10*** | -6.20*** | -16.94*** | -6.78*** | -19.91*** | -6.67*** |
| | (2.88) | (1.02) | (2.42) | (0.83) | (2.35) | (-3.03) | (2.20) | (0.77) |
| Log-likelihood | -631.71 | -637.10 | -1,096.04 | -1,106.57 | -1,334.68 | -1,349.87 | -1,470.94 | -1,485.4 |
| LR test (chi2 value) | 153.11 | 147.61 | 259.95 | 252.60 | 316.08 | 307.94 | 309.83 | 303.49 |
| LR test (Prob > chi2) | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** |
| Pseudo R2 | 0.1918 | 0.1849 | 0.2535 | 0.2463 | 0.3065 | 0.2986 | 0.3247 | 0.3181 |
| # of non-defaulted banks | 7,631 | | 7,496 | | 7,411 | | 7,370 | |
| # of defaulted banks | 139 | | 277 | | 363 | | 405 | |
| # of observations | 15,208 | | 22,331 | | 29,241 | | 35,998 | |

The table reports results for the estimated logit and probit discrete hazard model for the years 2009 to 2012. It provides information on estimated coefficients, their statistical significance and standard errors (numbers in parentheses), log-likelihood, pseudo-$R^2$ value, and statistical significance of the entire model (likelihood-ratio test), as well as the total number of defaulted and non-defaulted banks and observations used within the estimation for each year. Variables are defined in Table 3.6. *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Results for the estimated discrete hazard models for the years 2009[36] to 2012 are reported in Table 3.8. Also the dynamic models typically contain between three and five explanatory variables that represent a subset of the following variables: $x_1$: Return on Assets (%), $x_2$: Return on Equity (%), $x_6$: Cost of Total Interest Bearing Deposits (%), $x_8$: Net Loans & Leases / Total Assets (%), $x_{11}$: Noncurrent Loans & Leases / Gross Loans & Leases (%), and $x_{17}$: Tier One Risk-Based Capital / Risk-Weighted Assets (%). Note that based on the applied algorithm, for most of the years, a very similar subset of variables were included into the final model for the estimated static and discrete hazard models.[37] As in case of the static models, majority of the variables in the hazard models is statistically significant at the 1% level of significance and also models' power is increasing over time. However, increase in pseudo-$R^2$ values is not that significant here and the values themselves are lower compared to static models (e.g. 0.5263 for the 2012 static logit model and 0.3247 for the 2012 hazard logit model).

In order to supplement measures such as the pseudo-$R^2$ and the statistical significance of the entire model, we also examine the calibration accuracy of the models by applying Vuong's closeness test (Vuong [174]). The test examines how well estimated models' PDs match with actual outcomes of defaulted and non-defaulted banks. Using a calibration accuracy test and its likelihood estimates we are able to determine which model's PDs (from a set of candidate models) are closest to the actual PDs given a set of empirical data.[38] The higher the likelihood, the more accurately a model predicts actual defaults.

Using a vector of model outputs (estimated PDs) and a vector of default outcomes (one for defaulted banks and zero for non-defaulted banks) we calculate the log-likelihood[39]

[36] We did not estimate hazard models for 2008, since no pooling of the data can be done for the first year of our sample. As a result, the estimated discrete hazard models for 2008 would be identical to the static ones.

[37] As mentioned in Section 3.3.2, for the discrete hazard models it is necessary to adjust the sample size to account for the lack of independence between bank-year observations. This is what differentiates hazard models from simple "pooled" models and guarantees appropriate test statistics for the estimated coefficients. Based on the adjustment of statistical significance of particular coefficients within each estimated model, the variable $x_8$ (ratio of net loans & leases to total assets) had to be excluded for the estimated probit models in 2010, 2011, and 2012, and for the logit model in 2012. Even though this variable was statistically significant at the 5% (sometimes even at the 1%) level of significance for the "pooled" models, after the adjustment of the coefficients the variable was not significant anymore, even at the 10% level of significance.

[38] For a closer discussion on calibration and likelihood-based measures of calibration, see Stein [162].

[39] Logarithm of the likelihood is a monotonic transformation of the likelihood and thus the fact the model with higher log-likelihood is better calibrated is still valid.

**Table 3.9**
Calibration accuracy test (in-sample calibration)

| | model 2008 (log-likelihood) | model 2009 (log-likelihood) | model 2010 (log-likelihood) | model 2011 (log-likelihood) | model 2012 (log-likelihood) |
|---|---|---|---|---|---|
| Static logit | **-102.2307** | **-467.1505** | -389.7131 | **-228.2497** | **-120.6940** |
| Static probit | -106.5046 | -488.3219 | **-386.9624** | -321.3570 | -124.3151 |
| Hazard logit | -------- | **-631.3825** | **-1,095.7732** | **-1,334.3447** | **-1,470.3562** |
| Hazard probit | -------- | -636.8508 | -1,106.2798 | -1,349.5912 | -1,484.9912 |

The table reports the log-likelihood measures obtained from the calibration accuracy test for in-sample. Note that the higher value of log-likelihood, the better calibrated a given model is (model predicts PDs more accurately). These models are highlighted in bold.

measure for each of the estimated models. Since different datasets were used for estimation of the models, we could not use these tests for making a cross-comparison (static against hazard models) in this case, but only for a comparison of the results within the group of either static or dynamic probit and logit models for each year. Results for conducted calibration accuracy test are reported in Table 3.9. In order to determine whether the calculated log-likelihoods for various models are significantly different, we used Vuong's test for non-nested models. Results are reported in Table 3.10.

Our findings for in-sample calibration indicate that the logit models typically provide a better fit to the data than the probit models. The only exception is the 2010 static model, however, as indicated by Table 3.10, the difference between the logit and probit model is not statistically significant. We are also able to statistically distinguish between most of the models at the 5% level of significance pointing towards a significantly better fit of the applied logit models in 2009, 2011 and 2012. For 2008 and 2010, the performance of the estimated logit and probit models cannot be statistically distinguished. The $Z$ statistic for the 2008 static

**Table 3.10**
Vuong's closeness test for non-nested models (in-sample calibration)

| | model 2008 | model 2009 | model 2010 | model 2011 | model 2012 |
|---|---|---|---|---|---|
| static logit / | -1.8285 | **3.8247** | 0.7837 | **8.5623** | **2.5873** |
| static probit | (0.9663) | **(0.0001)** | (0.2166) | **(0.0000)** | **(0.0048)** |
| hazard logit / | ------ | **2.7908** | 1.2374 | **1.8877** | **2.8121** |
| hazard probit | ------ | **(0.0026)** | (0.1080) | **(0.0295)** | **(0.0025)** |

The table reports Vuong's closeness test for non-nested models. Top number represents $Z$ statistics while the number in parenthesis is a $p$-value. Statistically significant differences are highlighted in bold.

models is very low due to a relatively small number of defaulted banks (19) in the sample. Overall, based on the conducted tests we state that logit models are better calibrated and, therefore, produce more accurate default probability estimates compared to probit models.

### 3.4.3 Model validation

As mentioned above, we apply a framework called the walk-forward approach with out-of-time validation that allows testing models while controlling for time dependence, see Stein [162] for a more thorough discussion of this approach. This technique suggests to use a different set of data in validating the out-of-sample performance of the estimated models. At the same time, the approach allows to use as much of the data as possible to fit and to test the models.[40] Numbers of non-defaulted and defaulted banks used for the validation along with data collection dates are reported in Table 3.11.

As explained in Section 3.4.2 and Figure 3.2, for validation of the estimated models we use financial ratios collected one year after estimation of the models (see Table 3.5). For

**Table 3.11**
Numbers of banks and data collection dates for control samples

| VALIDATION OF | # of defaulted banks | # of non-defaulted banks | Date (financial ratios) |
|---|---|---|---|
| model 2008 | 120 *(defaulted in 2009)* | 7,515 | 31/12/2007 |
| model 2009 | 138 *(defaulted in 2010)* | 7,185 | 31/12/2008 |
| model 2010 | 86 *(defaulted in 2011)* | 6,978 | 31/12/2009 |
| model 2011 | 42 *(defaulted in 2012)* | 6,834 | 31/12/2010 |
| model 2012 | 23 *(defaulted in 2013)* | 6,618 | 31/12/2011 |

The table reports the number of defaulted and non-defaulted banks along with the data collection dates used for validation of the estimated models. For out-of-sample validation of the models, we use financial ratios collected one year after the estimation period (see Table 3.5) and then investigate the performance of the models in predicting defaults of the next year. For example, while estimating the 2008 static model (based on banks defaulted in 2008), we used financial ratios collected on December 31, 2006. For validation of the estimated model, we are predicting defaults in 2009 using financial ratios collected on December 31, 2007.

[40] While Stein [162] describes this approach with out-of-sample and out-of-time sampling, we used only out-of-time sampling in order to avoid the reduction in the number of defaulted banks within estimation and testing.

example, for the estimation of the 2008 static model, we use data on defaulted and non-defaulted banks in 2008 and financial ratios collected on December 31, 2006. The performance of the model is then validated predicting defaults in 2009 using financial ratios collected on December 31, 2007.

#### 3.4.3.1 Distributions of estimated PDs

Let us first have a look at the estimated out-of-sample default probabilities obtained from the models. In Table 3.12, we provide descriptive statistics for the distribution of PDs (mean value, standard deviation, skewness, and kurtosis), calculated separately for non-defaulted and defaulted banks during the validation period. As expected, we find that for all models the mean of the estimated PDs is significantly lower for non-defaulted banks in comparison to defaulted banks. For example, estimated average PDs for the static logit model in the non-default group are between 0.69% (for 2012) and 1.97% (for 2011), while in the default group they range from 8.24% (for 2008) up to 44.87% (for 2012). Quite similar results are obtained for the static probit models as well as for the dynamic discrete hazard models. Typically estimated PDs for each group are highly skewed and exhibit excess kurtosis.

#### 3.4.3.2 ROC curve analysis

After examining the distributions of estimated default probabilities for the individual models, we now compare our estimated models in terms of areas under the ROC curves (the ROC area) and accuracy ratios (AR). The information for each model and year is summarized in Table 3.13, while Figure 3.3 provides a plot of the ROC curve for the estimated static and dynamic probit model for 2010 and the static and dynamic logit model for 2012.

The results in Table 3.13 illustrate that the areas under the ROC curve increase over time and reach almost 99% for the static models in 2012. This is mainly a result of the large sample size, i.e. the high number of non-defaulted banks, in comparison to the very low number of defaulted banks for the years 2011 and 2012. For 2011, the sample contained 6,834 non-defaulted banks and only 42 defaults, while for 2012 the sample contains 6,618 non-defaults and only 23 defaults. From a first glance, we observe that for a specific year, ROC areas and accuracy ratios are typically very similar for all models, what makes it hard to decide whether any of the models is able to outperform the others. The only exception is 2012, where the

**Table 3.12**
PDs' statistics of the distributions (non-defaulted and defaulted banks)

| | | | model 2008 | model 2009 | model 2010 | model 2011 | model 2012 |
|---|---|---|---|---|---|---|---|
| Static logit | non-defaulted | mean | 1.03% | 0.75% | 1.72% | 1.97% | 0.69% |
| | | st.dev. | 0.05 | 0.03 | 0.09 | 0.12 | 0.06 |
| | | skew | 14.71 | 15.34 | 8.13 | 7.26 | 12.69 |
| | | kurt | 275.26 | 304.16 | 71.20 | 53.76 | 178.35 |
| | defaulted | mean | 8.24% | 14.72% | 40.54% | 38.74% | 44.87% |
| | | st.dev. | 0.14 | 0.24 | 0.35 | 0.36 | 0.39 |
| | | skew | 3.26 | 2.20 | 0.49 | 0.59 | 0.43 |
| | | kurt | 11.64 | 3.71 | -1.31 | -1.28 | -1.65 |
| Hazard logit | non-defaulted | mean | ------- | 0.72% | 1.61% | 1.83% | 0.58% |
| | | st.dev. | ------- | 0.03 | 0.08 | 0.11 | 0.05 |
| | | skew | ------- | 16.19 | 8.84 | 7.70 | 14.68 |
| | | kurt | ------- | 358.56 | 86.56 | 61.27 | 241.54 |
| | defaulted | mean | ------- | 13.51% | 34.15% | 35.91% | 39.27% |
| | | st.dev. | ------- | 0.23 | 0.33 | 0.35 | 0.40 |
| | | skew | ------- | 2.36 | 0.81 | 0.69 | 0.60 |
| | | kurt | ------- | 4.61 | -0.76 | -1.18 | -1.50 |
| Static probit | non-defaulted | mean | 0.90% | 1.07% | 1.84% | 3.04% | 0.62% |
| | | st.dev. | 0.04 | 0.05 | 0.09 | 0.12 | 0.05 |
| | | skew | 17.32 | 9.91 | 7.66 | 6.22 | 13.79 |
| | | kurt | 351.32 | 130.73 | 64.81 | 40.83 | 221.44 |
| | defaulted | mean | 6.02% | 15.60% | 40.37% | 45.88% | 40.27% |
| | | st.dev. | 0.11 | 0.25 | 0.34 | 0.33 | 0.38 |
| | | skew | 3.85 | 1.80 | 0.52 | 0.33 | 0.68 |
| | | kurt | 16.07 | 2.37 | -1.13 | -1.29 | -1.35 |
| Hazard probit | non-defaulted | mean | ------- | 1.07% | 1.80% | 1.88% | 0.61% |
| | | st.dev. | ------- | 0.04 | 0.08 | 0.11 | 0.05 |
| | | skew | ------- | 13.12 | 8.23 | 7.61 | 14.21 |
| | | kurt | ------- | 238.76 | 78.04 | 60.91 | 237.25 |
| | defaulted | mean | ------- | 15.60% | 34.04% | 35.50% | 37.97% |
| | | st.dev. | ------- | 0.21 | 0.30 | 0.32 | 0.38 |
| | | skew | ------- | 2.28 | 0.81 | 0.73 | 0.71 |
| | | kurt | ------- | 4.68 | -0.57 | -0.91 | -1.26 |

The table reports statistics of the distributions (mean value, standard deviation, skewness, and kurtosis) of estimated PDs for individual models, calculated separately for non-defaulted and defaulted banks. Mean values of PDs are expressed in percentage units, while other statistics are expressed in decimal numbers.

static logit and probit models seem to clearly outperform their dynamic counterparts. While for the static logit model, the ROC area is 0.9881, we obtain a value of 0.9490 for the hazard logit model. For the static probit model we obtain a ROC area of 0.9882, while the dynamic probit model yields a value of 0.9511. This is quite an interesting finding and is most likely

**Table 3.13**
ROC areas and accuracy ratios (AR)

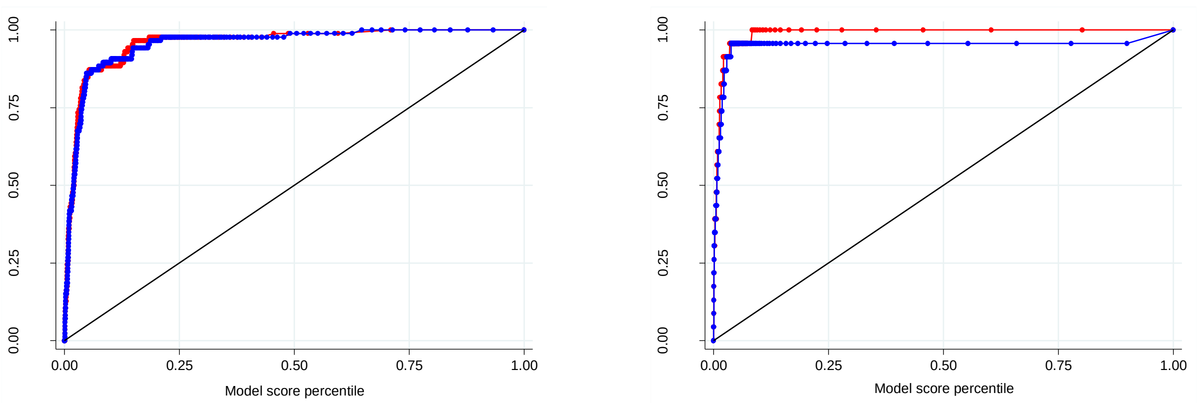
| | | model 2008 | model 2009 | model 2010 | model 2011 | model 2012 |
|---|---|---|---|---|---|---|
| Static logit | ROC area: | 0.8536 | 0.9333 | 0.9605 | 0.9624 | 0.9881 |
| | AR: | 0.7072 | 0.8666 | 0.9210 | 0.9248 | 0.9762 |
| Hazard logit | ROC area: | ------ | 0.9333 | 0.9592 | 0.9619 | 0.9490 |
| | AR: | ------ | 0.8666 | 0.9184 | 0.9238 | 0.8980 |
| Static probit | ROC area: | 0.8359 | 0.9383 | 0.9578 | 0.9595 | 0.9882 |
| | AR: | 0.6718 | 0.8766 | 0.9156 | 0.9190 | 0.9764 |
| Hazard probit | ROC area: | ------ | 0.9389 | 0.9556 | 0.9595 | 0.9511 |
| | AR: | ------ | 0.8778 | 0.9112 | 0.9190 | 0.9022 |

The table shows calculated areas under the ROC curves (ROC area) and accuracy ratios (AR) for each of the estimated models. Reported values are very high due to a high number of non-defaulted banks compared to defaulted banks.

caused by the exclusion of the financial ratio *net loans & leases to total assets* ($x_8$). As stated in Footnote 37, we did estimate the simple “pooled” models to find a possible cause of the significant difference between the static and discrete hazard models for 2012. Including the variable $x_8$ into the discrete hazard models for predicting defaults in 2012, we obtain areas under the ROC curve of 0.9891 for the dynamic logit and 0.9894 for the dynamic probit models, i.e. results almost identical to those of the static models. However, even though this variable was statistically significant at the 5% (sometimes even at 1%) level in the “pooled” models, it was insignificant even at the 10% level, after the necessary adjustment of statistical inference.

**Figure 3.3**
ROC curves

A) Probit 2010 models B) Logit 2012 models




The figure plots ROC curves for 2010 probit models (panel A) and 2012 logit models (panel B). Static models are represented by the red curves, while hazard models by the blue curves.

#### 3.4.3.3 Sizing the variability of ROC areas – bootstrapping

As pointed out by Stein [162], results of model testing are subject to sample variability. The author also illustrates that it is typically rather the number of defaults than the total number of total observations in the sample that tends to drive the stability of performance measures such as the accuracy ratio or the ROC area. Small numbers of defaults lead to a very high variability in the results. Stein [162] concludes that the best one can do is to size and understand this variability as under normal circumstances it is not possible to reduce it.

A common approach to sizing the variability of a particular statistic given an empirical sample is to use resampling techniques to leverage the available data and reduce the dependence of the results on a particular sample. Therefore, we bootstrapped our control sample data sets (sampled with replacement) 1,000 times to examine the stability of the results.[41] Figure 3.4 provides an exemplary plot of the distribution of ROC areas for the hazard

**Table 3.14**
ROC area statistics of the distributions (bootstrapping)

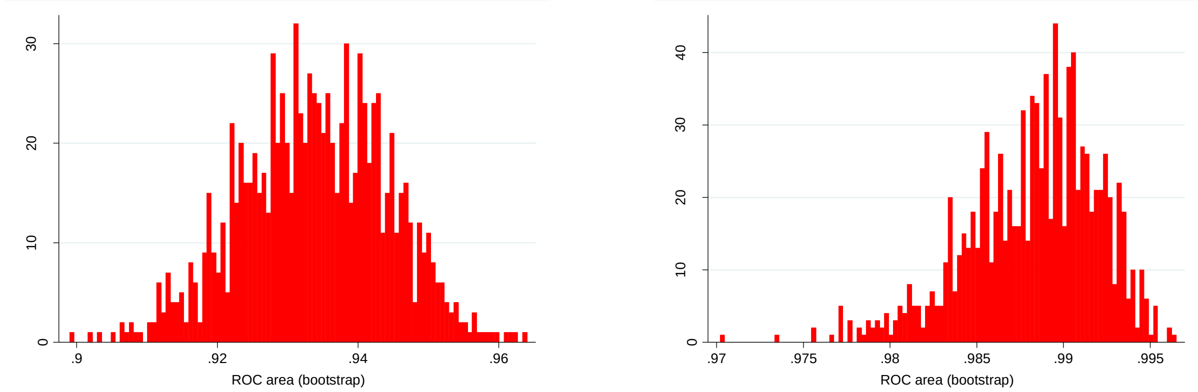
| | | model 2008 | model 2009 | model 2010 | model 2011 | model 2012 |
|---|---|---|---|---|---|---|
| Static logit | mean | 0.8545 | 0.9335 | 0.9605 | 0.9621 | 0.9882 |
| | std.dev. | 0.0170 | 0.0100 | 0.0100 | 0.0130 | 0.0042 |
| | skew | -0.0720 | -0.1522 | -0.5641 | -0.9772 | -0.8208 |
| | kurt | 3.0798 | 2.8460 | 3.3229 | 3.9371 | 3.8913 |
| Hazard logit | mean | -------- | 0.9338 | 0.9592 | 0.9616 | 0.9504 |
| | std.dev. | -------- | 0.0104 | 0.0102 | 0.0127 | 0.0429 |
| | skew | -------- | -0.1364 | -0.6403 | -0.9586 | -1.2150 |
| | kurt | -------- | 2.8264 | 3.5545 | 3.9441 | 6.1221 |
| Static probit | mean | 0.8370 | 0.9385 | 0.9579 | 0.9593 | 0.9883 |
| | std.dev. | 0.0196 | 0.0097 | 0.0104 | 0.0128 | 0.0038 |
| | skew | -0.0417 | -0.1745 | -0.5992 | -0.9191 | -0.7230 |
| | kurt | 3.0058 | 2.9205 | 3.5269 | 3.8446 | 3.7050 |
| Hazard probit | mean | -------- | 0.9392 | 0.9557 | 0.9592 | 0.9524 |
| | std.dev. | -------- | 0.0104 | 0.0108 | 0.0142 | 0.0397 |
| | skew | -------- | -0.1843 | -0.6639 | -0.9536 | -1.2093 |
| | kurt | -------- | 2.8707 | 3.7224 | 3.9271 | 6.0884 |

The table reports descriptive statistics of the distributions (mean value, standard deviation, skewness, and kurtosis) of areas under the ROC curves (ROC area) for each model and year based on 1,000 bootstrapped resamples for each year.

[41] Creating random samples for bootstrapping we combine defaulted and non-defaulted banks into one pool and calculate the ROC areas for static and hazard models based on 1,000 resamples from this pool. We keep the bootstrap size same as sample size.

**Figure 3.4**
Distribution of ROC areas based on bootstrapping

A) Hazard logit 2009 model B) Static probit 2012 model




The figure plots the distribution of areas under the ROC curves (ROC area) for 2009 hazard logit model (panel A) and 2012 static probit model (panel B) based on bootstrapping.

logit model for 2009 and the static probit model for 2012. Statistics of the distributions for all models are then listed in Table 3.14.

Results for ROC areas based on the bootstrap resamples suggest that the ROC area statistics are not affected by dependence on the particular sample as the mean value of the bootstrapped ROC areas (Table 3.14) typically differs only slightly from its original value (Table 3.13).[42] We also performed a Kruskal-Wallis test to examine significant differences between the particular models in terms of the bootstrapped ROC areas. Note that this test does not require the assumption that all samples come from a population with a normal distribution, which would not be justified in our case. Results of this test are reported in Table 3.15.

Since for each of the years we reject the null hypothesis that all samples of ROC areas are drawn from the same population, we can say that, in terms of bootstrapped ROC areas, at least one model stochastically dominates at least one other model. In order to further investigate this issue, we also conducted a multiple comparison procedure. This procedure uses Tukey's honestly significant difference (Tukey's HSD) criterion that is optimal for the comparison of groups with equal sample sizes. The procedure allows to test for significant differences with regards to the performance (ROC area) of the particular models. The test is conducted with a significance level of $\alpha$ = 0.05. For each year, Table 3.16 indicates for all four

[42] For example, mean value of ROC area statistic calculated from bootstrapping for 2009 hazard probit model is 0.9392 while without resampling this statistic is 0.9389 for this model.

**Table 3.15**
Kruskal-Wallis test

| | model 2008 (rank sum) | model 2009 (rank sum) | model 2010 (rank sum) | model 2011 (rank sum) | model 2012 (rank sum) |
|---|---|---|---|---|---|
| Static logit | 1.25E+06 | 1.70E+06 | 2.24E+06 | 2.15E+06 | 2.52E+06 |
| Hazard logit | ------ | 1.73E+06 | 2.10E+06 | 2.09E+06 | 1.52E+06 |
| Static probit | 7.49E+05 | 2.25E+06 | 1.95E+06 | 1.86E+06 | 2.51E+06 |
| Hazard probit | ------ | 2.32E+06 | 1.72E+06 | 1.90E+06 | 1.45E+06 |
| K-statistic | 380.216 | 245.537 | 112.939 | 46.743 | 791.703 |
| p-value | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

The table reports results of the Kruskal-Wallis test. $K$ statistic has a chi-squared distribution with 1 degree of freedom for 2008 models and 3 degrees of freedom for the rest of the models. In all cases, we reject null hypothesis implying that there are statistically significant differences between bootstrapped ROC areas among the models in particular years.

estimated models (in 2008 only two static models have been estimated) which of the other models perform significantly worse or significantly better.[43]

Let us consider the 2008 models to illustrate the results of the conducted tests. Overall, we have four estimated models: (1) the static logit model, (2) the discrete hazard logit model, (3) the static probit model, (4) the discrete hazard probit model. For 2008, the population of the bootstrap sample of the static logit model (1) is significantly to the right, which implies significantly higher bootstrapped ROC area values compared to the probit model (3). For 2012, there is no statistical difference between the static models (1) and (3) and between the hazard models (2) and (4). However, both static models have population of the bootstrap samples significantly to the right, which means significantly higher bootstrapped ROC area values compared to the hazard models. This is in accordance to our results reported in Table 3.14, where mean values of bootstrapped ROC area for static models are 0.9882 and 0.9883, respectively, compared to hazard models with values 0.9504 and 0.9524. Thus, using the Kruskal-Wallis and Tukey's tests we managed to demonstrate significant statistical difference between the models' power.

Overall, Tukey's test provides a very powerful tool for distinguishing between individual models and should be considered as an additional testing procedure for comparing the out-of-sample performance of credit-scoring models. To the best of our knowledge, so far the test

[43] Test statistics, along with the number of degrees of freedom and $p$-values, are identical to the values in the Kruskal-Wallis test.

**Table 3.16**
Tukey's test

| | model 2008 (worse / better) | model 2009 (worse / better) | model 2010 (worse / better) | model 2011 (worse / better) | model 2012 (worse / better) |
|---|---|---|---|---|---|
| (1) static logit | All / - | - / {3,4} | All / - | {3,4} / - | {2,4} / - |
| (2) hazard logit | ------ | - / {3,4} | {3,4} / {1} | {3,4} / - | - / {1,3} |
| (3) static probit | - / All | {1,2} / - | {4} / {1,2} | - / {1,2} | {2,4} / - |
| (4) hazard probit | ------ | {1,2} / - | - / All | - / {1,2} | - / {1,3} |

The table provides results for the multiple comparison procedure of the mean ranks of particular models using Tukey's HSD criterion. In particular, it illustrates for each of the four estimated models - (1) static logit model, (2) discrete hazard logit model, (3) static probit model, (4) discrete hazard probit model - which of the other models performs significantly worse or significantly better. 'All' means that all other models were significantly worse/better, while '-' indicates that none of the other models were significantly worse/better.

has not been applied to examining the discriminatory power of credit rating models. Note that this test was able to statistically distinguish between all 2010 models, where, for example, the difference between mean values of bootstrapped ROC areas for static and dynamic probit models is relatively small (0.0022). Examining the performance of different scoring models is one of the key tasks to develop appropriate models, while often it is quite difficult to distinguish between the models with regards to their discriminatory power. DeLong, DeLong and Clarke-Pearson [46] provide a test for the difference between the areas under the ROC curves of two rating models, which relies on the assumption of asymptotic normality. Engelmann, Hayden and Tasche [57] then discuss this approach and test the validity of this assumption. Their analysis indicates that reliability of this method is not guaranteed in the case of a validation sample containing only a small number of defaults. On the other hand, despite the fact that Tukey's test does not require the assumption of normality, it proved to be a very powerful test in distinguishing between individual scoring models.

### 3.4.3.4 Calibration accuracy test

Several of the previous sections suggest that the performance of individual models is very similar for a particular year. So far we have focused on examining the discriminative power of the models which clearly is one of the key criteria to be applied when validating credit models. Another key task is to examine the performance of the models with respect to their likelihood. This section is devoted to examining model calibration along with the application of Vuong's closeness test (Vuong [174]). Results for log-likelihoods for the out-of-sample validation of the models are reported in Table 3.17. Note that in comparison to Section 3.4.2 where results

**Table 3.17**
Calibration accuracy test (out-of-sample validation period)

| | model 2008 (log-likelihood) | model 2009 (log-likelihood) | model 2010 (log-likelihood) | model 2011 (log-likelihood) | model 2012 (log-likelihood) |
|---|---|---|---|---|---|
| Static logit | **-605.3952** | -498.1433 | -126.1653 | -0.2820 | -153.0582 |
| Hazard logit | -------- | **-498.0313** | **-99.7176** | **-0.2515** | -150.3667 |
| Static probit | -651.0178 | -506.4024 | -155.3029 | -1.6970 | **-147.9030** |
| Hazard probit | -------- | -506.2495 | -119.6402 | -0.2755 | -167.6681 |

The table reports the log-likelihood - equation (3.14) – for each model and year, based on the out-of-sample calibration period. Note that higher values of the log-likelihood indicate a better calibration of the model to default and non-default data, i.e. the model predicts PDs more accurately. For each year, results for the best model are indicated in bold.

on model estimation were examined using the log-likelihood, results in Table 3.17 are based only on out-of-sample results. Thus, the log-likelihood in equation (3.14) is calculated by comparing the vector of predicted PDs to actually observed defaults and non-defaults during the out-of-sample validation period.

The results in Table 3.17 indicate that static and dynamic logit models typically outperform their probit counterparts with regards to accuracy. We also observe that for 2009, 2010 and 2011 the discrete hazard logit models yield the highest log-likelihood of all models. Recall that for 2008 the static and dynamic hazard models are identical such that only results for the static models are reported. Interestingly, for 2012, we find that the static probit model provides the best result in terms of the log-likelihood measure.

In order to determine whether calculated log-likelihoods for various models are significantly different, we use the Vuong's closeness test for non-nested models. In a first step, we test the best model, i.e. the model with the highest log-likelihood in a particular year, against all other models. Then we test the second best model against the remaining models and so on. In this way, we conducted this test for six different pairwise combinations for each year 2009, 2010, 2011 and 2012, while we only have one pairwise combination in 2008. The higher the value of the $Z$ statistic, the greater is the difference between the model with the higher log-likelihood and the other model. Results of conducted tests are reported in Table 3.18.

Unfortunately, in terms of log-likelihoods calculated from the calibration accuracy test, we cannot distinguish between the majority of the models (unlike for the in-sample calibration results). There are only a few pairwise combinations (highlighted in bold) that

**Table 3.18**
Vuong's closeness test for non-nested models (out-of-sample validation period)

| | static logit / hazard logit | static logit / static probit | static logit / hazard probit | hazard logit / static probit | hazard logit / hazard probit | static probit / hazard probit |
|---|---|---|---|---|---|---|
| model 2008 | ------ | 1.1787 | ------ | ------ | ------ | ------ |
| | ------ | (0.1193) | ------ | ------ | ------ | ------ |
| model 2009 | **1.3433** | 0.3821 | 0.2633 | 0.5134 | 0.4134 | 0.5899 |
| | **(0.0896)** | (0.3512) | (0.3961) | (0.3038) | (0.3397) | (0.2776) |
| model 2010 | **2.9600** | **1.6880** | 0.4842 | **2.7187** | 0.8103 | **2.5622** |
| | **(0.0015)** | **(0.0457)** | (0.3141) | **(0.0033)** | (0.2089) | **(0.0052)** |
| model 2011 | 0.0008 | 0.0253 | 0.0981 | 0.0212 | -0.0802 | 0.1040 |
| | (0.4997) | (0.4899) | (0.4609) | (0.4915) | (0.5320) | (0.4586) |
| model 2012 | -0.0995 | 0.5577 | 0.7600 | 0.3507 | **1.5453** | 1.0171 |
| | (0.5396) | (0.2885) | (0.2236) | (0.3629) | **(0.0611)** | (0.1546) |

The table reports Vuong's closeness test for non-nested models. We report $Z$ statistics and $p$-values (in parenthesis) for each conducted test. There are six pairwise combinations for the years 2009-2012 (as we have four estimated models in each year) and only one pairwise combination in 2008 (only two estimated models). Tests that yield a significance outperformance of the model with the higher log-likelihood are indicated in bold letters.

indicate a statistically significant difference between the log-likelihoods of the models at the 10% level. However, we got similar results to our findings obtained from ROC analysis and bootstrapping ROC areas, where we showed that the 2012 static models have a higher discriminatory power than the hazard models. The 2012 static probit model with a log-likelihood value of -147.90 seems to perform better than the hazard probit model with a log-likelihood value of -167.67 (although the value of the $Z$ statistic of 1.02 from Vuong's closeness test suggests that the difference is not significant at the 10% level). The 2012 hazard logit model with a log-likelihood value of -150.37 might be slightly better calibrated compared to static logit model with the log-likelihood value of -153.09, nonetheless this difference is not significant whatsoever ( $p$ -value of 0.5396).

#### 3.4.3.5 PD analysis for the entire score sample

Likelihood measures make relative comparisons between competing models. Unfortunately, it is not possible to use them for evaluating whether a specific model is correctly calibrated or not. Therefore, it is often useful to conduct an additional analysis by comparing the expected and actually observed number of defaults. Using our out-of-sample validation periods, we started with calculating the mean values of estimated PDs (defaulted and non-defaulted banks together) and the expected number of defaults, along with the ratios of

**Table 3.19**
E(PD), actual and expected # of defaults, and ratios for out-of-time validation

| | | model 2008 | model 2009 | model 2010 | model 2011 | model 2012 |
|---|---|---|---|---|---|---|
| Static logit | E(PD): | 1.15% | 1.01% | 2.20% | 2.19% | 0.84% |
| | actual # of D: | 120 | 138 | 86 | 42 | 23 |
| | expected # of D: | 88 | 74 | 155 | 151 | 56 |
| | ratio: | 136% | 186% | 55% | 28% | 41% |
| Hazard logit | E(PD): | ------ | 0.96% | 2.01% | 2.04% | 0.72% |
| | actual # of D: | ------ | 138 | 86 | 42 | 23 |
| | expected # of D: | ------ | 70 | 142 | 140 | 48 |
| | ratio: | ------ | 197% | 61% | 30% | 48% |
| Static probit | E(PD): | 0.98% | 1.96% | 2.31% | 3.30% | 0.75% |
| | actual # of D: | 120 | 138 | 86 | 42 | 23 |
| | expected # of D: | 75 | 144 | 163 | 227 | 50 |
| | ratio: | 160% | 96% | 53% | 19% | 46% |
| Hazard probit | E(PD): | ------ | 1.35% | 2.20% | 2.09% | 0.74% |
| | actual # of D: | ------ | 138 | 86 | 42 | 23 |
| | expected # of D: | ------ | 99 | 155 | 144 | 49 |
| | ratio: | ------ | 139% | 55% | 29% | 47% |

The table reports mean values of estimated PDs (E(PD)), calculated for defaulted and non-defaulted banks together, actual and expected (under a given model) number of defaults (D), along with the ratios of actual number of defaults over expected number of defaults for each of the models.

actual over expected number of defaults for every particular model. Results are summarized in Table 3.19.

Our findings suggest that overall the calibrated models underestimate the actual number of defaults for the years 2008 and 2009, while they clearly overestimate the number of defaults for the years 2010, 2011 and 2012. A possible explanation for this behaviour is that the models for 2008 and 2009 were estimated during periods of lower default rates using financial ratios from December 31, 2006 and December 31, 2007, respectively. On the other hand, models for later years were calibrated during the financial crisis using data from periods of relatively high number of defaults and may, therefore, overestimate the actual number of defaults in later periods.

A common approach for researchers to determine the accuracy of estimated probabilities is to run experiments in which they attempt to estimate the goodness-of-fit between expected (under a given model) and actual default rates, see, e.g., Stein [161]. Such a comparison for each model, along with the Hosmer-Lemeshow's chi-squared goodness-of-fit

**Table 3.20**

Expected vs. actual default rates (Hosmer-Lemeshow test)

| | | model 2008 | | model 2009 | | model 2010 | | model 2011 | | model 2012 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | exp. DR | act. DR | exp. DR | act. DR | exp. DR | act. DR | exp. DR | act. DR | exp. DR | act. DR |
| | (1 ; 0.90) | 8.71% | 8.78% | 7.72% | 14.75% | 20.33% | 10.76% | 21.23% | 5.97% | 8.20% | 3.46% |
| | (0.90 ; 0.80) | 1.45% | 2.49% | 0.97% | 2.46% | 0.84% | 0.99% | 0.32% | 0.00% | 0.09% | 0.00% |
| | (0.80 ; 0.70) | 0.64% | 1.57% | 0.54% | 0.68% | 0.35% | 0.14% | 0.15% | 0.00% | 0.05% | 0.00% |
| | (0.70 ; 0.60) | 0.32% | 1.31% | 0.34% | 0.41% | 0.20% | 0.14% | 0.09% | 0.00% | 0.04% | 0.00% |
| | (0.60 ; 0.50) | 0.18% | 0.52% | 0.22% | 0.00% | 0.12% | 0.00% | 0.06% | 0.00% | 0.03% | 0.00% |
| Static logit | (0.50 ; 0.40) | 0.10% | 0.52% | 0.15% | 0.41% | 0.07% | 0.00% | 0.04% | 0.15% | 0.02% | 0.00% |
| | (0.40 ; 0.30) | 0.05% | 0.26% | 0.10% | 0.14% | 0.04% | 0.14% | 0.03% | 0.00% | 0.01% | 0.00% |
| | (0.30 ; 0.20) | 0.03% | 0.13% | 0.06% | 0.00% | 0.02% | 0.00% | 0.02% | 0.00% | 0.01% | 0.00% |
| | (0.20 ; 0.10) | 0.01% | 0.13% | 0.04% | 0.00% | 0.01% | 0.00% | 0.01% | 0.00% | 0.00% | 0.00% |
| | (0.10 ; 0) | 0.00% | 0.00% | 0.01% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| | | HL stat. = | 79.52 | HL stat. = | 73.92 | HL stat. = | 44.54 | HL stat. = | 102.07 | HL stat. = | 21.41 |
| | | *p*-value = | 0.0000 | *p*-value = | 0.0000 | *p*-value = | 0.0000 | *p*-value = | 0.0000 | *p*-value = | 0.0184 |
| | (1 ; 0.90) | | | 6.92% | 14.62% | 17.56% | 10.76% | 19.41% | 5.97% | 6.75% | 3.31% |
| | (0.90 ; 0.80) | | | 1.00% | 2.46% | 1.05% | 1.13% | 0.41% | 0.00% | 0.15% | 0.00% |
| | (0.80 ; 0.70) | | | 0.58% | 0.96% | 0.51% | 0.00% | 0.20% | 0.00% | 0.09% | 0.00% |
| | (0.70 ; 0.60) | | | 0.39% | 0.14% | 0.33% | 0.00% | 0.13% | 0.00% | 0.06% | 0.00% |
| | (0.60 ; 0.50) | | | 0.27% | 0.27% | 0.23% | 0.14% | 0.09% | 0.00% | 0.05% | 0.00% |
| Hazard logit | (0.50 ; 0.40) | ---------- | | 0.19% | 0.14% | 0.16% | 0.00% | 0.07% | 0.15% | 0.04% | 0.00% |
| | (0.40 ; 0.30) | | | 0.13% | 0.14% | 0.11% | 0.14% | 0.05% | 0.00% | 0.03% | 0.00% |
| | (0.30 ; 0.20) | | | 0.08% | 0.00% | 0.07% | 0.00% | 0.03% | 0.00% | 0.02% | 0.00% |
| | (0.20 ; 0.10) | | | 0.04% | 0.14% | 0.03% | 0.00% | 0.01% | 0.00% | 0.01% | 0.00% |
| | (0.10 ; 0) | | | 0.01% | 0.00% | 0.01% | 0.00% | 0.00% | 0.00% | 0.00% | 0.15% |
| | | HL stat. = | ----- | HL stat. = | 88.47 | HL stat. = | 30.81 | HL stat. = | 86.43 | HL stat. = | 90.96 |
| | | *p*-value = | ----- | *p*-value = | 0.0000 | *p*-value = | 0.0006 | *p*-value = | 0.0000 | *p*-value = | 0.0000 |
| | (1 ; 0.90) | 6.92% | 8.78% | 14.22% | 15.44% | 21.74% | 10.76% | 28.60% | 5.82% | 7.38% | 3.46% |
| | (0.90 ; 0.80) | 1.36% | 2.36% | 2.43% | 1.91% | 0.88% | 1.13% | 2.15% | 0.15% | 0.08% | 0.00% |
| | (0.80 ; 0.70) | 0.68% | 1.70% | 1.24% | 0.82% | 0.26% | 0.00% | 0.94% | 0.00% | 0.04% | 0.00% |
| | (0.70 ; 0.60) | 0.37% | 0.79% | 0.74% | 0.14% | 0.11% | 0.00% | 0.56% | 0.00% | 0.02% | 0.00% |
| | (0.60 ; 0.50) | 0.21% | 0.66% | 0.44% | 0.14% | 0.05% | 0.14% | 0.34% | 0.15% | 0.01% | 0.00% |
| Static probit | (0.50 ; 0.40) | 0.12% | 0.52% | 0.28% | 0.14% | 0.02% | 0.00% | 0.22% | 0.00% | 0.01% | 0.00% |
| | (0.40 ; 0.30) | 0.06% | 0.13% | 0.17% | 0.27% | 0.01% | 0.14% | 0.13% | 0.00% | 0.01% | 0.00% |
| | (0.30 ; 0.20) | 0.03% | 0.52% | 0.10% | 0.00% | 0.00% | 0.00% | 0.07% | 0.00% | 0.00% | 0.00% |
| | (0.20 ; 0.10) | 0.01% | 0.00% | 0.04% | 0.00% | 0.00% | 0.00% | 0.03% | 0.00% | 0.00% | 0.00% |
| | (0.10 ; 0) | 0.00% | 0.26% | 0.01% | 0.00% | 0.00% | 0.00% | 0.01% | 0.00% | 0.00% | 0.00% |
| | | HL stat. = | 384.77 | HL stat. = | 9.96 | HL stat. = | 66.98 | HL stat. = | 202.04 | HL stat. = | 16.01 |
| | | *p*-value = | 0.0000 | *p*-value = | 0.4439 | *p*-value = | 0.0000 | *p*-value = | 0.0000 | *p*-value = | 0.0995 |
| | (1 ; 0.90) | | | 9.69% | 15.85% | 19.03% | 10.91% | 20.03% | 5.82% | 7.12% | 3.31% |
| | (0.90 ; 0.80) | | | 1.53% | 1.50% | 1.43% | 0.85% | 0.47% | 0.15% | 0.13% | 0.00% |
| | (0.80 ; 0.70) | | | 0.83% | 0.68% | 0.60% | 0.14% | 0.17% | 0.00% | 0.06% | 0.00% |
| | (0.70 ; 0.60) | | | 0.54% | 0.00% | 0.35% | 0.00% | 0.10% | 0.00% | 0.03% | 0.00% |
| | (0.60 ; 0.50) | | | 0.36% | 0.27% | 0.22% | 0.14% | 0.06% | 0.00% | 0.02% | 0.00% |
| Hazard probit | (0.50 ; 0.40) | ---------- | | 0.24% | 0.41% | 0.15% | 0.00% | 0.04% | 0.15% | 0.02% | 0.00% |
| | (0.40 ; 0.30) | | | 0.16% | 0.00% | 0.10% | 0.14% | 0.03% | 0.00% | 0.01% | 0.00% |
| | (0.30 ; 0.20) | | | 0.09% | 0.00% | 0.05% | 0.00% | 0.01% | 0.00% | 0.01% | 0.00% |
| | (0.20 ; 0.10) | | | 0.04% | 0.14% | 0.02% | 0.00% | 0.01% | 0.00% | 0.00% | 0.00% |
| | (0.10 ; 0) | | | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.15% |
| | | HL stat. = | ----- | HL stat. = | 40.65 | HL stat. = | 38.93 | HL stat. = | 92.58 | HL stat. = | 545.11 |
| | | *p*-value = | ----- | *p*-value = | 0.0000 | *p*-value = | 0.0000 | *p*-value = | 0.0000 | *p*-value = | 0.0000 |

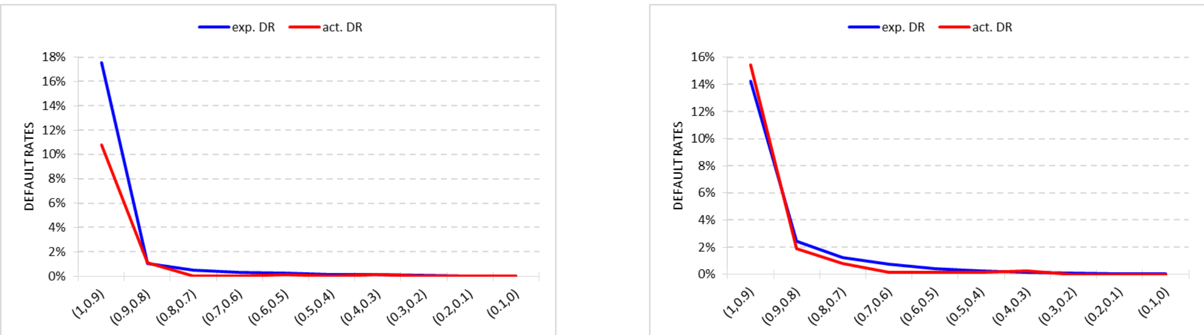
The table reports a comparison between expected (exp. DR) and actual (act. DR) default rates along with the Hosmer-Lemeshow's goodness-of-fit tests. For each of the models the ranked PD estimates were divided into ten intervals of the same size. The $HL$ test statistic (HL stat.) follows a chi-squared distribution with 10 d.f.

tests, is carried out in Table 3.20, where we divided the ranked PD estimates into ten intervals of the same size (each interval contains the same number of banks).[44] Moreover, we also illustrate the comparison procedure in more detail for the 2010 hazard logit and the 2009 static probit model in Figure 3.5.

Our findings imply that the expected and actual default rates are typically very similar for banks with a credit score that falls into one of the bands [0,0.10], (0.10,0.20], …., (0.80,0.90], i.e. for all those that do not fall into the decile of banks worst credit scores. As expected, the difference between expected and actual default rates is most substantial for banks with the worst score, i.e. banks that fall into the decile (0.90,1]. While for 2008 and 2009, actual default rates exceed expected rates as they have been estimated by the model, for the remaining years actual default rates are below expected default rates according to the estimated models. Note that these results are in line with our findings in Table 3.19 for the entire sample. The Hosmer-Lemeshow test results in a rejection of the hypothesis of an accurate prediction of the number of defaults for most of the models even at the 10% level of significance. Exceptions include the 2009 and 2012 static probit models. Overall, the results suggest that expected and actual default rates are not statistically equal for the (0.90,1]. In order to further investigate this issue, we also conducted the same type of analysis for particular deciles themselves.

**Figure 3.5**
Expected vs. actual default rates

A) Hazard logit 2010 model B) Static probit 2009 model




This figure shows a comparison between expected (under a given model) and actual default rates for 2010 hazard logit model (panel A) and 2009 static probit model (panel B).

[44] The first interval (0.90,1] contains the 10% banks with the lowest credit score (banks with the highest estimated PDs), the seconds interval (0.80,0.90] contains the next 10% of the banks, etc.

**Table 3.21**

Expected vs. actual default rates (Hosmer-Lemeshow test for deciles)

| | | model 2008 | | model 2009 | | model 2010 | | model 2011 | | model 2012 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | HL stat. | *p*-value | HL stat. | *p*-value | HL stat. | *p*-value | HL stat. | *p*-value | HL stat. | *p*-value |
| | (1 ; 0.90) | 0.00 | **0.945** | 50.76 | 0.000 | 39.88 | 0.000 | 95.73 | 0.000 | 19.78 | 0.000 |
| | (0.90 ; 0.80) | 5.75 | 0.017 | 16.95 | 0.000 | 0.19 | **0.660** | 2.21 | **0.137** | 0.60 | **0.438** |
| | (0.80 ; 0.70) | 10.59 | 0.001 | 0.30 | **0.584** | 0.86 | **0.354** | 1.01 | **0.314** | 0.34 | **0.560** |
| | (0.70 ; 0.60) | 23.20 | 0.000 | 0.11 | **0.740** | 0.11 | **0.742** | 0.64 | **0.424** | 0.24 | **0.625** |
| Static logit | (0.60 ; 0.50) | 5.31 | 0.021 | 1.62 | **0.203** | 0.85 | **0.357** | 0.44 | **0.510** | 0.17 | **0.678** |
| | (0.50 ; 0.40) | 13.95 | 0.000 | 3.28 | **0.070** | 0.49 | **0.483** | 1.66 | **0.120** | 0.12 | **0.727** |
| | (0.40 ; 0.30) | 6.25 | 0.012 | 0.09 | **0.760** | 1.98 | **0.159** | 0.20 | **0.656** | 0.08 | **0.773** |
| | (0.30 ; 0.20) | 3.15 | **0.076** | 0.47 | **0.494** | 0.13 | **0.722** | 0.12 | **0.737** | 0.05 | **0.826** |
| | (0.20 ; 0.10) | 11.30 | 0.000 | 0.26 | **0.611** | 0.04 | **0.837** | 0.06 | **0.809** | 0.02 | **0.882** |
| | (0.10 ; 0) | 0.01 | **0.914** | 0.07 | **0.785** | 0.00 | **0.945** | 0.01 | **0.908** | 0.00 | **0.950** |
| | (1 ; 0.90) | | | 67.33 | 0.000 | 22.54 | 0.000 | 79.39 | 0.000 | 12.43 | 0.000 |
| | (0.90 ; 0.80) | | | 15.68 | 0.000 | 0.04 | **0.833** | 2.85 | **0.092** | 1.00 | **0.317** |
| | (0.80 ; 0.70) | | | 1.81 | **0.179** | 3.65 | **0.056** | 1.39 | **0.239** | 0.57 | **0.448** |
| | (0.70 ; 0.60) | | | 1.20 | **0.272** | 2.37 | **0.124** | 0.91 | **0.340** | 0.42 | **0.519** |
| Hazard logit | (0.60 ; 0.50) | --------- | | 0.00 | **0.974** | 0.23 | **0.630** | 0.65 | **0.421** | 0.31 | **0.575** |
| | (0.50 ; 0.40) | | | 0.11 | **0.744** | 1.14 | **0.286** | 0.61 | **0.433** | 0.24 | **0.626** |
| | (0.40 ; 0.30) | | | 0.00 | **0.952** | 0.07 | **0.786** | 0.32 | **0.574** | 0.17 | **0.678** |
| | (0.30 ; 0.20) | | | 0.59 | **0.442** | 0.47 | **0.492** | 0.20 | **0.658** | 0.12 | **0.734** |
| | (0.20 ; 0.10) | | | 1.69 | **0.193** | 0.24 | **0.627** | 0.10 | **0.755** | 0.06 | **0.808** |
| | (0.10 ; 0) | | | 0.06 | **0.808** | 0.05 | **0.816** | 0.02 | **0.883** | 75.64 | 0.000 |
| | (1 ; 0.90) | 4.08 | 0.043 | 0.89 | **0.344** | 49.98 | 0.000 | 174.56 | 0.000 | 14.90 | 0.000 |
| | (0.90 ; 0.80) | 5.61 | 0.018 | 0.82 | **0.366** | 0.52 | **0.473** | 13.15 | 0.000 | 0.52 | **0.472** |
| | (0.80 ; 0.70) | 11.97 | 0.001 | 1.07 | **0.301** | 1.83 | **0.176** | 6.50 | 0.011 | 0.23 | **0.629** |
| | (0.70 ; 0.60) | 3.64 | **0.057** | 3.59 | **0.058** | 0.80 | **0.372** | 3.84 | 0.050 | 0.14 | **0.705** |
| Static probit | (0.60 ; 0.50) | 7.18 | 0.007 | 1.51 | **0.219** | 1.15 | **0.284** | 0.79 | **0.375** | 0.09 | **0.761** |
| | (0.50 ; 0.40) | 10.33 | 0.001 | 0.51 | **0.474** | 0.16 | **0.686** | 1.52 | **0.218** | 0.06 | **0.808** |
| | (0.40 ; 0.30) | 0.52 | **0.471** | 0.48 | **0.490** | 12.51 | 0.000 | 0.92 | **0.337** | 0.03 | **0.852** |
| | (0.30 ; 0.20) | 58.37 | 0.000 | 0.70 | **0.404** | 0.02 | **0.877** | 0.50 | **0.479** | 0.02 | **0.897** |
| | (0.20 ; 0.10) | 0.09 | **0.764** | 0.32 | **0.573** | 0.00 | **0.944** | 0.21 | **0.647** | 0.01 | **0.942** |
| | (0.10 ; 0) | 282.97 | 0.000 | 0.07 | **0.791** | 0.00 | **0.986** | 0.04 | **0.835** | 0.00 | **0.983** |
| | (1 ; 0.90) | | | 31.77 | 0.000 | 30.24 | 0.000 | 86.60 | 0.000 | 14.55 | 0.000 |
| | (0.90 ; 0.80) | | | 0.00 | **0.959** | 1.68 | **0.195** | 1.53 | **0.216** | 0.86 | **0.354** |
| | (0.80 ; 0.70) | | | 0.19 | **0.659** | 2.51 | **0.113** | 1.18 | **0.278** | 0.37 | **0.542** |
| | (0.70 ; 0.60) | | | 3.96 | **0.047** | 2.49 | **0.114** | 0.67 | **0.414** | 0.23 | **0.632** |
| Hazard probit | (0.60 ; 0.50) | --------- | | 0.14 | **0.706** | 0.22 | **0.641** | 0.43 | **0.514** | 0.15 | **0.695** |
| | (0.50 ; 0.40) | | | 0.83 | **0.361** | 1.06 | **0.302** | 1.85 | **0.173** | 0.10 | **0.749** |
| | (0.40 ; 0.30) | | | 1.14 | **0.286** | 0.15 | **0.695** | 0.18 | **0.674** | 0.07 | **0.780** |
| | (0.30 ; 0.20) | | | 0.66 | **0.418** | 0.38 | **0.537** | 0.10 | **0.755** | 0.04 | **0.849** |
| | (0.20 ; 0.10) | | | 1.91 | **0.167** | 0.16 | **0.686** | 0.04 | **0.843** | 0.01 | **0.903** |
| | (0.10 ; 0) | | | 0.04 | **0.848** | 0.02 | **0.875** | 0.00 | **0.944** | 528.72 | 0.000 |

The table reports $HL$ test statistics (HL stat.) for particular deciles calculated from a comparison between expected and actual default rates, where for each of the models the ranked PD estimates were divided into ten intervals of the same size. The $HL$ test statistic follows a chi-squared distribution with 1 d.f. in this case. Particular cases of acceptance of null hypothesis at the 10% confidence level are highlighted in bold.

Results of the Hosmer-Lemeshow test applied for particular deciles are reported in Table 3.21. Again, we divided the ranked PD estimates into ten intervals (deciles) of the same size (each interval contains the same number of banks). There is a visible pattern (with a few exceptions such as the 2012 hazard models, where these models incorrectly ranked one of the 23 defaulted banks into the last decile causing a high value of the *HL* statistic) of statistical equality between expected and actual default rates for the models 2009-2012 for all the deciles except of the first one containing 10% banks with the worst rating. Particular cases where we accept null hypothesis of the Hosmer-Lemeshow test at the 10% confidence level are highlighted in bold. Overall, we can see the highest reported values of the *HL* statistic are in the first decile implying that the expected and actual default rates differ the most for the banks with the worst rating. This is what causes the rejection of the Hosmer-Lemeshow test for most of the models in Table 3.20.

There is a reason why these results might be biased. As stated in Section 3.3.3.4, the appropriateness of the $p$-value calculated using $HL$ statistic depends on the validity of the assumption that the estimated expected frequencies are large. These should be greater than 5. Unfortunately, for many of our models this assumption is violated for most of the deciles.

#### 3.4.3.6 Focus on the tails

There are three reasons why we decided to devote this section to the tails of the distribution of estimated PDs and to incorporate the tail-based measures within validation of the models: (1) our control samples contain a very high number of non-defaulted banks compared to defaulted banks and we wanted to focus on predicted defaults rather than majority of non-defaults; (2) ROC areas in particular years were not very helpful in distinguishing between the models; (3) the expected and actual default rates significantly differ for the banks with the highest 10% PDs.

First of all, we calculated a ratio of number of defaulted banks to the number of non-defaulted banks within the highest 10% PDs (divided into 20 intervals of size 0.5%) along with the information of number of captured defaulted banks within this interval. Results are reported in Table 3.22 (to save space, we do not report values in intervals (0.92,0.925] – (0.975,0.98]).

We can see that the power of the models increases over time (which is in accordance with the ROC analysis conclusions). In fact, for the 2011 and 2012 models more than 95% of defaulted banks are captured within the group of banks with the 10% lowest credit scores. For example, twenty-two out of twenty-three defaulted banks were captured within the 4.5% of banks with the highest PDs for the 2012 logit and probit hazard models. In other words, these twenty-two defaulted banks (as mentioned above, we are predicting defaults in 2013 here) are among the 299 banks (out of 6641 banks) with the highest estimated PDs. These

**Table 3.22**

Ratios of # of defaulted over # of non-defaulted banks (banks with the highest 10% PDs)

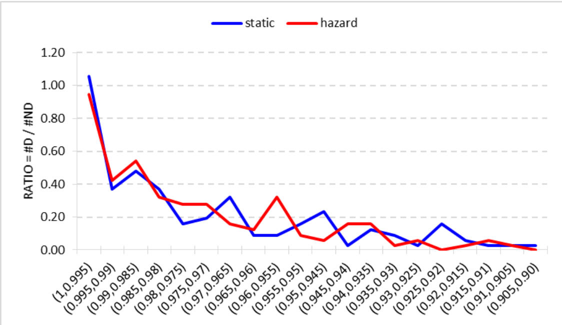
| | | model 2008 | model 2009 | model 2010 | model 2011 | model 2012 |
|---|---|---|---|---|---|---|
| | | Ratios | Ratios | Ratios | Ratios | Ratios |
| Static logit | (1 ; 0.995) | 0.36 | 1.06 | 0.52 | 0.03 | 0.38 |
| | (0.995 ; 0.99) | 0.19 | 0.37 | 0.40 | 0.21 | 0.14 |
| | (0.99 ; 0.985) | 0.19 | 0.48 | 0.40 | 0.13 | 0.10 |
| | (0.985 ; 0.98) | 0.15 | 0.37 | 0.25 | 0.21 | 0.10 |
| | … | … | … | … | … | … |
| | (0.92 ; 0.915) | 0.09 | 0.06 | 0.00 | 0.00 | 0.00 |
| | (0.915 ; 0.91) | 0.06 | 0.03 | 0.00 | 0.00 | 0.03 |
| | (0.91 ; 0.905) | 0.09 | 0.03 | 0.00 | 0.00 | 0.00 |
| | (0.905 ; 0.90) | 0.03 | 0.03 | 0.00 | 0.03 | 0.00 |
| | D banks captured: | 67/120 (55.83%) | 108/138 (78.26%) | 76/86 (88.37%) | 41/42 (97.62%) | 23/23 (100%) |
| Hazard logit | (1 ; 0.995) | ---------- | 0.95 | 0.52 | 0.03 | 0.38 |
| | (0.995 ; 0.99) | | 0.42 | 0.30 | 0.17 | 0.10 |
| | (0.99 ; 0.985) | | 0.54 | 0.52 | 0.26 | 0.10 |
| | (0.985 ; 0.98) | | 0.32 | 0.17 | 0.10 | 0.06 |
| | … | | … | … | … | … |
| | (0.92 ; 0.915) | | 0.03 | 0.00 | 0.00 | 0.00 |
| | (0.915 ; 0.91) | | 0.06 | 0.03 | 0.03 | 0.00 |
| | (0.91 ; 0.905) | | 0.03 | 0.00 | 0.03 | 0.00 |
| | (0.905 ; 0.90) | | 0.00 | 0.00 | 0.00 | 0.00 |
| | D banks captured: | ---------- | 107/138 (77.54%) | 76/86 (88.37%) | 41/42 (97.62%) | 22/23 (95.65%) |
| Static probit | (1 ; 0.995) | 0.36 | 0.85 | 0.46 | 0.03 | 0.38 |
| | (0.995 ; 0.99) | 0.21 | 0.48 | 0.40 | 0.21 | 0.10 |
| | (0.99 ; 0.985) | 0.09 | 0.37 | 0.40 | 0.10 | 0.14 |
| | (0.985 ; 0.98) | 0.12 | 0.37 | 0.21 | 0.21 | 0.10 |
| | … | … | … | … | … | … |
| | (0.92 ; 0.915) | 0.03 | 0.09 | 0.00 | 0.03 | 0.00 |
| | (0.915 ; 0.91) | 0.06 | 0.03 | 0.00 | 0.03 | 0.00 |
| | (0.91 ; 0.905) | 0.09 | 0.00 | 0.03 | 0.00 | 0.00 |
| | (0.905 ; 0.90) | 0.03 | 0.06 | 0.00 | 0.00 | 0.00 |
| | D banks captured: | 67/120 (55.83%) | 112/138 (81.16%) | 76/86 (88.37%) | 40/42 (95.24%) | 23/23 (100%) |
| Hazard probit | (1 ; 0.995) | ---------- | 0.85 | 0.59 | 0.03 | 0.32 |
| | (0.995 ; 0.99) | | 0.32 | 0.21 | 0.21 | 0.14 |
| | (0.99 ; 0.985) | | 0.61 | 0.67 | 0.17 | 0.06 |
| | (0.985 ; 0.98) | | 0.42 | 0.09 | 0.13 | 0.06 |
| | … | | … | … | … | … |
| | (0.92 ; 0.915) | | 0.12 | 0.00 | 0.00 | 0.00 |
| | (0.915 ; 0.91) | | 0.06 | 0.03 | 0.03 | 0.00 |
| | (0.91 ; 0.905) | | 0.03 | 0.03 | 0.00 | 0.00 |
| | (0.905 ; 0.90) | | 0.00 | 0.00 | 0.03 | 0.00 |
| | D banks captured: | ---------- | 116/138 (84.06%) | 76/86 (88.37%) | 40/42 (95.24%) | 22/23 (95.65%) |

The table reports ratios of number of defaulted banks over number of non-defaulted banks (ratios) for the banks with the highest 10% PDs (divided into 20 intervals of size 0.5%) along with the information of number of captured defaulted banks within this interval.

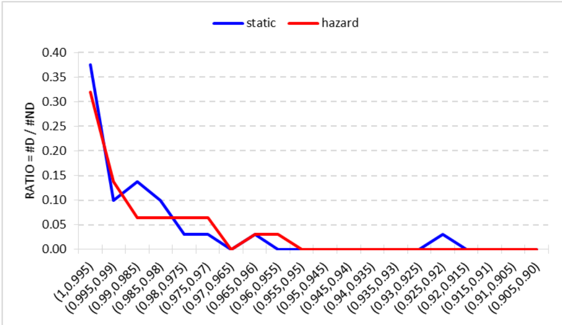
findings are quite promising and achieve very good results compared to recent studies conducted in the corporate sector. Beaver, McNichols and Rhie [16] builds a model based on accounting ratios which captures 80.3% of the year-ahead defaulting corporations in the lowest two deciles (period 1994-2002). Once the authors include additional variables

**Figure 3.6**
Ratios of # of defaulted over # of non-defaulted banks

A) Logit 2009 models




B) Probit 2012 models



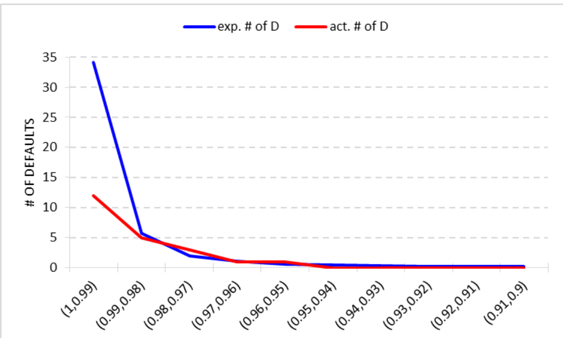

The figure shows a comparison between ratios (number of defaulted banks over number of non-defaulted banks) calculated for the static and the hazard models for the banks with the highest 10% PDs. The 2009 logit models are considered in panel A while the 2012 probit models in panel B.

derived from equity markets, this measure rose to 88.1%. After allowing their model coefficients to adjust over time, this measure even increases up to 92%. The model of Duffie, Saita and Wang [52] places 94% of the one-year ahead defaults in the lowest two deciles (period 1993-2004). We typically obtain similar or even slightly better results considering only the lowest decile for our 2011 and 2012 models.

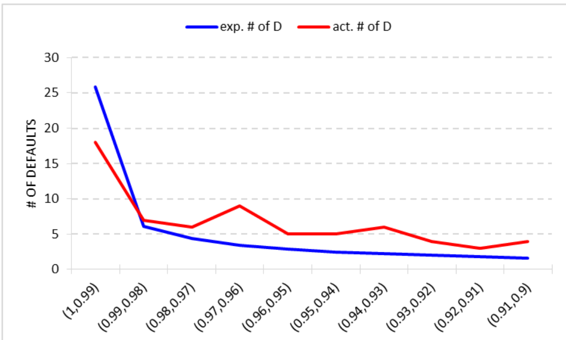
Moreover, we also illustrate a comparison of these ratios in more detail for the 2009 logit and the 2012 probit models in Figure 3.6*.* As expected, their values have a downward trend with decreasing values of estimated PDs.

**Figure 3.7**
Expected vs. actual # of defaults

A) Hazard logit 2012 model




B) Static probit 2008 model




The figure shows a comparison between expected (under a given model) and actual number of defaults for the banks with the highest 10% PDs. The 2012 hazard logit model is considered in panel A while the 2008 static probit model in panel B.

**Table 3.23**

Expected vs. actual number of defaults (Hosmer-Lemeshow test)

| | | model 2008 | | model 2009 | | model 2010 | | model 2011 | | model 2012 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | interval | exp. # of D | act. # of D | exp. # of D | act. # of D | exp. # of D | act. # of D | exp. # of D | act. # of D | exp. # of D | act. # of D |
| | (1 ; 0.99) | 28.91 | 16 | 31.20 | 29 | 64.47 | 22 | 67.25 | 7 | 41.05 | 13 |
| | (0.99 ; 0.98) | 9.28 | 11 | 8.18 | 22 | 37.28 | 17 | 46.48 | 10 | 8.33 | 6 |
| | (0.98 ; 0.97) | 6.53 | 10 | 4.60 | 11 | 18.18 | 11 | 18.00 | 6 | 2.59 | 2 |
| | (0.97 ; 0.96) | 5.02 | 8 | 3.14 | 12 | 8.80 | 15 | 6.72 | 11 | 1.07 | 1 |
| | (0.96 ; 0.95) | 3.99 | 3 | 2.33 | 8 | 5.00 | 8 | 2.87 | 0 | 0.51 | 0 |
| | (0.95 ; 0.94) | 3.33 | 4 | 1.88 | 8 | 3.26 | 1 | 1.62 | 3 | 0.30 | 0 |
| Static logit | (0.94 ; 0.93) | 2.78 | 1 | 1.60 | 7 | 2.28 | 1 | 1.08 | 2 | 0.21 | 0 |
| | (0.93 ; 0.92) | 2.43 | 5 | 1.40 | 6 | 1.69 | 1 | 0.76 | 1 | 0.15 | 0 |
| | (0.92 ; 0.91) | 2.17 | 5 | 1.23 | 3 | 1.35 | 0 | 0.58 | 0 | 0.12 | 1 |
| | (0.91;0.90) | 1.93 | 4 | 1.10 | 2 | 1.12 | 0 | 0.47 | 1 | 0.10 | 0 |
| | | 66 | 67 | 57 | 108 | 143 | 76 | 146 | 41 | 54 | 23 |
| | | HL stat. = | 24.00 | HL stat. = | 134.25 | HL stat. = | 393.97 | HL stat. = | 5017.44 | HL stat. = | 59.23 |
| | | *p* -value = | 0.0043 | *p* -value = | 0.0000 | *p* -value = | 0.0000 | *p* -value = | 0.0000 | *p* -value = | 0.0000 |
| | (1 ; 0.99) | | | 27.57 | 29 | 58.81 | 20 | 66.12 | 6 | 34.13 | 12 |
| | (0.99 ; 0.98) | | | 7.19 | 22 | 28.57 | 17 | 39.23 | 10 | 5.67 | 5 |
| | (0.98 ; 0.97) | | | 4.04 | 16 | 14.15 | 14 | 14.36 | 9 | 1.99 | 3 |
| | (0.97 ; 0.96) | | | 2.80 | 9 | 7.37 | 10 | 5.89 | 9 | 1.04 | 1 |
| | (0.96 ; 0.95) | ---------- | | 2.16 | 12 | 4.56 | 11 | 2.80 | 0 | 0.61 | 1 |
| | (0.95 ; 0.94) | | | 1.78 | 7 | 3.24 | 2 | 1.63 | 2 | 0.42 | 0 |
| Hazard logit | (0.94 ; 0.93) | | | 1.54 | 6 | 2.42 | 0 | 1.16 | 1 | 0.31 | 0 |
| | (0.93 ; 0.92) | | | 1.37 | 2 | 1.90 | 1 | 0.88 | 2 | 0.25 | 0 |
| | (0.92 ; 0.91) | | | 1.22 | 3 | 1.56 | 1 | 0.70 | 1 | 0.20 | 0 |
| | (0.91;0.90) | | | 1.10 | 1 | 1.31 | 0 | 0.56 | 1 | 0.17 | 0 |
| | | ------ | ------ | 51 | 107 | 124 | 76 | 133 | 41 | 45 | 22 |
| | | HL stat. = | ------ | HL stat. = | 163.63 | HL stat. = | 183.88 | HL stat. = | 2036.69 | HL stat. = | 31.92 |
| | | *p* -value = | ------ | *p* -value = | 0.0000 | *p* -value = | 0.0000 | *p* -value = | 0.0000 | *p* -value = | 0.0002 |
| | (1 ; 0.99) | 25.82 | 18 | 42.51 | 29 | 62.79 | 21 | 67.05 | 7 | 34.37 | 12 |
| | (0.99 ; 0.98) | 6.10 | 7 | 17.42 | 20 | 36.48 | 16 | 49.11 | 9 | 8.12 | 7 |
| | (0.98 ; 0.97) | 4.42 | 6 | 11.08 | 9 | 21.35 | 14 | 28.41 | 7 | 3.24 | 2 |
| | (0.97 ; 0.96) | 3.44 | 9 | 8.03 | 17 | 11.62 | 12 | 17.26 | 7 | 1.47 | 1 |
| | (0.96 ; 0.95) | 2.88 | 5 | 6.17 | 9 | 7.10 | 7 | 10.60 | 4 | 0.70 | 0 |
| | (0.95 ; 0.94) | 2.50 | 5 | 5.07 | 8 | 4.82 | 3 | 7.52 | 1 | 0.41 | 0 |
| Static probit | (0.94 ; 0.93) | 2.22 | 6 | 4.34 | 11 | 3.40 | 2 | 5.67 | 1 | 0.26 | 0 |
| | (0.93 ; 0.92) | 1.99 | 4 | 3.72 | 3 | 2.48 | 0 | 4.32 | 2 | 0.18 | 1 |
| | (0.92 ; 0.91) | 1.79 | 3 | 3.16 | 4 | 1.88 | 0 | 3.47 | 2 | 0.13 | 0 |
| | (0.91;0.90) | 1.61 | 4 | 2.85 | 2 | 1.46 | 1 | 2.81 | 0 | 0.11 | 0 |
| | | 53 | 67 | 104 | 112 | 153 | 76 | 196 | 40 | 49 | 23 |
| | | HL stat. = | 31.13 | HL stat. = | 37.00 | HL stat. = | 303.50 | HL stat. = | 4019.50 | HL stat. = | 36.61 |
| | | *p* -value = | 0.0003 | *p* -value = | 0.0000 | *p* -value = | 0.0000 | *p* -value = | 0.0000 | *p* -value = | 0.0000 |
| | (1 ; 0.99) | | | 32.38 | 26 | 55.61 | 19 | 64.65 | 7 | 32.66 | 12 |
| | (0.99 ; 0.98) | | | 11.13 | 25 | 28.62 | 17 | 36.58 | 9 | 7.49 | 4 |
| | (0.98 ; 0.97) | | | 7.13 | 14 | 16.62 | 12 | 16.50 | 9 | 2.95 | 4 |
| | (0.97 ; 0.96) | | | 4.99 | 18 | 10.12 | 10 | 8.00 | 6 | 1.56 | 1 |
| | (0.96 ; 0.95) | ---------- | | 3.74 | 9 | 6.89 | 9 | 4.28 | 4 | 0.91 | 1 |
| | (0.95 ; 0.94) | | | 3.08 | 6 | 5.10 | 7 | 2.62 | 1 | 0.60 | 0 |
| Hazard probit | (0.94 ; 0.93) | | | 2.64 | 7 | 3.82 | 0 | 1.85 | 0 | 0.40 | 0 |
| | (0.93 ; 0.92) | | | 2.28 | 4 | 3.01 | 0 | 1.31 | 2 | 0.29 | 0 |
| | (0.92 ; 0.91) | | | 1.96 | 6 | 2.42 | 1 | 0.98 | 1 | 0.22 | 0 |
| | (0.91;0.90) | | | 1.75 | 1 | 1.99 | 1 | 0.79 | 1 | 0.18 | 0 |
| | | ------ | ------ | 71 | 116 | 134 | 76 | 138 | 40 | 47 | 22 |
| | | HL stat. = | ------ | HL stat. = | 94.67 | HL stat. = | 136.92 | HL stat. = | 1096.67 | HL stat. = | 30.02 |
| | | *p* -value = | ------ | *p* -value = | 0.0000 | *p* -value = | 0.0000 | *p* -value = | 0.0000 | *p* -value = | 0.0004 |

The table reports a comparison between expected (exp. # of D) and actual (act. # of D) number of defaults for the banks with the highest 10% PDs (divided into 10 intervals of size 1%) along with the Hosmer-Lemeshow chi-squared tests. For all of the models we reject null hypothesis suggesting that there is a significant difference between expected and actual number of defaults.

We also conducted Hosmer-Lemeshow's chi-squared goodness-of-fit tests on the banks with the highest 10% PDs (10 intervals of size 1%). $P$-values of these tests for particular models along with the expected and actual numbers of defaulted banks within a particular interval are reported in Table 3.23.

We use these tests to examine whether the expected number of defaults is equal to the actual number of defaults for the set of banks with the highest 10% PDs. Results reported in Table 3.23 indicate that for all models we reject the null hypothesis at the 5% significance level. Also, the same trend as in the entire distribution (Table 3.19) regarding expected number of defaults compared to actual number of defaults is present in the interval of the highest 10% PDs. That is, the calibrated models underestimate the actual number of defaults for the years 2008 and 2009, while they clearly overestimate the number of defaults for the years 2010, 2011 and 2012. Note that the same issue as in Section 3.4.3.5 is present here. That is, for many intervals (intervals with lower PD clusters) the expected number of defaults is lower than 5.

Furthermore, we also illustrate a comparison of expected and actual number of defaults for the banks with the highest 10% PDs in more detail for the 2012 hazard logit and the 2008 static probit model in Figure 3.7.

#### 3.4.3.7 Estimated PDs and "real" PDs

Overall, we state that our estimated default probabilities are slightly biased (see Table 3.23, where we got a rejection on the banks with the highest 10% PDs for all of the models, or Table 3.19, where we got underestimated PDs for the years 2008 and 2009, and overestimated PDs for the years 2010, 2011 and 2012). On the other hand, they still might be considered reasonable estimates since, except for the first decile that contains the 10% banks with the worst rating, the analysis in Tables 3.20, 3.21 and Figure 3.5 illustrates that the expected and actual default rates are the same or do not differ much from each other. These results were also confirmed by the conducted Hosmer-Lemeshow's chi-squared test for most of the deciles. However, there is one more reason why our default probability estimates can be considered as being close to "real" PDs.

Generally, there is a difference between a discriminatory model and a model for determination of PDs. A ratio of defaulted and non-defaulted banks is not that crucial within a discriminatory model as the key idea is to find a cut-off point that best discriminates between these two groups. Estimated PDs then cannot be considered as real PDs. There are some approaches how to calibrate these estimated PDs to real ones, though. For example, estimated model might be adjusted by a constant which will correct a bias caused by not using an empirical ratio. One might also use adjusted maximum likelihood functions within estimation that takes into account an empirical ratio (e.g. Zmijewski [180]) or translate estimated PDs into real ones using various transforms (e.g. Neagu, Keenan and Chalermkraivuth [139]). As mentioned earlier, we worked with all of the available information on U.S. commercial banks (using the FFIEC database) in our case and so avoided choice-based samples within the estimation procedure. This means that we got ratios of defaulted and non-defaulted banks very close to the actual empirical ones. Therefore, our estimated PDs can be considered as "real" PDs and may be used for activities such as calculation of economic capital, credit Value-at-Risk, for scenario analysis purposes etc.

#### 3.4.3.8 A summary of the results on model comparison

Since we have estimated a number of different models and have examined the performance of these models across various criteria, we now provide a summary of the results for all models. Results for the comparison across different criteria are provided in Table 3.24. Note that we report the results separately for each year, i.e. for 2008, 2009, 2010, 2011, and 2012. The selected criteria are divided into two groups, reflecting results for the estimation and the validation stage. The former one is represented by number of variables, Pseudo $R^2$, the log-likelihood, and results for the calibration accuracy test (in-sample), while the latter one by mean values of estimated default probabilities, calculated separately for non-defaulted and defaulted banks, areas under the ROC curves, mean values of bootstrapped ROC areas, Tukey's test, log-likelihood of the calibration accuracy test (out-of-sample), ratio of actual and expected number of defaults, Hosmer-Lemeshow test, and percentage of defaulted banks captured in the 10% highest PDs. Note that for the in-sample calibration measures only a comparison between static-static and hazard-hazard models is possible due to the different datasets that have been used for the estimation.

Let us first recall that each of estimated models is statistically significant at the 1% level of significance (tested using the log-likelihood ratio test and the Wald test). In terms of Pseudo $R^2$ measure, the static models outperform the hazard models, with the static logit models providing better results for the years 2008, 2011, and 2012. In terms of log-likelihoods, the logit models outperform probit models in all years, with the exception of 2010, where no significant difference between the models can be detected.

For the validation measures, we find that the areas under the ROC curves typically differ only slightly for all models in a particular year, with the only exception in 2012, where the static models are clearly superior to the hazard models. Despite this fact, we managed to distinguish between the models using the Kruskal-Wallis test (see Table 3.15) and Tukey's test applied to the bootstrapped ROC areas. With regards to the conducted out-of-sample calibration accuracy tests, we find that the static logit and hazard logit models outperform the static probit and hazard probit models. The results also indicate that overall hazard models produce more accurate PD estimates compared to the static models. However, unlike for the in-sample calibration, we cannot distinguish significantly between the majority of the models based on the conducted Vuong's closeness test for non-nested models. While comparing actual and expected number of defaults for particular models, we concluded that for the 2008 and 2009 models estimated PDs are too low (the only exception is the 2009 static probit model), while for 2010, 2011 and 2012 estimated model PDs are too high. For the later years, the static probit models seem to perform best as the ratio of expected over actual number of defaults is closest to 1. In terms of defaulted banks captured within the 10% banks with the lowest credit score (i.e. highest PDs), we do not find a clear pattern with regards to one model outperforming all the others.

In order to put all of this information together, we have created a simple ranking system in Table 3.25 for the criteria considered in this study. We have excluded the number of explanatory variables, the log-likelihood as well as results for the calibration accuracy test (in-sample) as we cannot use these measures to compare the static against the dynamic hazard models. With regards to the measures that are used to examine the performance of the models for the validation period, we exclude the mean value of the bootstrapped ROC areas

**Table 3.24**

A summary of the results on model comparison

| | | Estimation | | | Validation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | # of variables | Pseudo R2 | Log-likelihood of CA test (in-sample) | E (PD) ND | E (PD) D | ROC area | bootstrapped E (ROC area) | Tukey's test worse / better | Log-likelihood of CA test (out-of-sample) | Ratio of act. and exp. # of D | HL test | D banks captured in 10% highest PDs |
| model 2008 | static logit | 5 | **0.2293** | **-102.2307*** | 1.03% | **8.24%** | **0.8536** | **0.8545** | **All / -** | **-605.3952** | **136%** | reject | **55.83%** |
| | hazard logit | ----- | ----- | ----- | ------ | ------ | ------ | ------ | ------ | ------ | ------ | ------ | ------ |
| | static probit | 3 | 0.1983 | -106.5046* | **0.90%** | 6.02% | 0.8359 | 0.8370 | - / All | -651.0178 | 160% | reject | **55.83%** |
| | hazard probit | ----- | ----- | ----- | ------ | ------ | ------ | ------ | ------ | ------ | ------ | ------ | ------ |
| model 2009 | static logit | 4 | 0.2257 | **-467.1505*** | 0.75% | 14.72% | 0.9333 | 0.9335 | - / {3,4} | -498.1433 | 186% | reject | 78.26% |
| | hazard logit | 3 | 0.1918 | **-631.3825**** | **0.72%** | 13.51% | 0.9333 | 0.9338 | - / {3,4} | **-498.0313** | 197% | reject | 81.16% |
| | static probit | 5 | **0.2295** | -488.3219* | 1.07% | **15.60%** | 0.9383 | 0.9385 | **{1,2} / -** | -506.4024 | **96%** | **accept** | 77.54% |
| | hazard probit | 4 | 0.1849 | -636.8508** | 1.07% | **15.60%** | **0.9389** | **0.9392** | **{1,2} / -** | -506.2495 | 139% | reject | **84.06%** |
| model 2010 | static logit | 5 | 0.4266 | -389.7131* | 1.72% | **40.54%** | **0.9605** | **0.9605** | **All / -** | -126.1653 | 55% | reject | **88.37%** |
| | hazard logit | 5 | 0.2535 | **-1,095.7732**** | **1.61%** | 34.15% | 0.9592 | 0.9592 | {3,4} / {1} | **-99.71760** | **61%** | reject | **88.37%** |
| | static probit | 5 | **0.4306** | **-386.9624*** | 1.84% | 40.37% | 0.9578 | 0.9579 | {4} / {1,2} | -155.3029 | 53% | reject | **88.37%** |
| | hazard probit | 4 | 0.2463 | -1,106.2798** | 1.80% | 34.04% | 0.9556 | 0.9557 | - / All | -119.6402 | 55% | reject | **88.37%** |
| model 2011 | static logit | 5 | **0.5078** | **-228.2497*** | 1.97% | 38.74% | **0.9624** | **0.9621** | **{3,4} / -** | -0.2820 | 28% | reject | **97.62%** |
| | hazard logit | 5 | 0.3065 | **-1,334.3447**** | **1.83%** | 35.91% | 0.9619 | 0.9616 | **{3,4} / -** | **-0.2515** | **30%** | reject | 95.24% |
| | static probit | 5 | 0.5065 | -321.3570* | 3.04% | **45.88%** | 0.9595 | 0.9593 | - / {1,2} | -1.6970 | 19% | reject | **97.62%** |
| | hazard probit | 4 | 0.2986 | -1,349.5912** | 1.88% | 35.50% | 0.9595 | 0.9592 | - / {1,2} | -0.2755 | 29% | reject | 95.24% |
| model 2012 | static logit | 3 | **0.5263** | **-120.6940*** | 0.69% | **44.87%** | 0.9881 | 0.9882 | **{2,4} / -** | -153.0582 | 41% | reject | **100.00%** |
| | hazard logit | 4 | 0.3247 | **-1,470.3562**** | **0.58%** | 39.27% | 0.9490 | 0.9504 | - / {1,3} | -150.3667 | **48%** | reject | **100.00%** |
| | static probit | 3 | 0.5126 | -124.3151* | 0.62% | 40.27% | **0.9882** | **0.9883** | **{2,4} / -** | **-147.9030** | 46% | **accept** | 95.65% |
| | hazard probit | 4 | 0.3181 | -1,484.9912** | 0.61% | 37.97% | 0.9511 | 0.9524 | - / {1,3} | -167.6681 | 47% | reject | 95.65% |

The table provides a comparison of estimated models in terms of various criteria. Estimation stage represent number of variables, Pseudo $R^2$, and log-likelihood of calibration accuracy (CA) test (in-sample), while for validation stage these are mean values of estimated default probabilities (E(PD)), calculated separately for non-defaulted (ND) and defaulted (D) banks, areas under the ROC curves (ROC area), mean values of bootstrapped ROC areas, Tukey's test, log-likelihood of CA test (out-of-sample), ratio of actual (act.) and expected (exp.) number of defaults (D), Hosmer-Lemeshow (HL) test, and percentage of defaulted (D) banks captured in 10% highest PDs. Comparison is done separately for particular year and the best value of individual criterion is highlighted in bold. For in-sample CA test only a comparison between static-static (*) and hazard-hazard (**) models is possible due to different datasets used within estimation. Regarding a ratio of actual and expected number of defaults value closest to 100% is the best.

**Table 3.25**
Rankings of the models

| | | Estimation | Validation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Pseudo R2 | E (PD) ND | E (PD) D | ROC area | Tukey's test | Log-likelihood of CA test (out-of-sample) | Ratio of act. and exp. # of D | HL test | D banks captured in 10% highest PDs | Total points | Final ranking |
| model 2008 | static logit | 1 (4) | 2 (3) | 1 (4) | 1 (4) | 1 (4) | 1 (4) | 1 (4) | reject (0) | 1-2 (4) | 31 | 1 |
| | hazard logit | ----- | ------ | ------ | ------ | ------ | ------ | ------ | ------ | ------ | ------ | ------ |
| | static probit | 2 (3) | 1 (4) | 2 (3) | 2 (3) | 2 (3) | 2 (3) | 2 (3) | reject (0) | 1-2 (4) | 26 | 2 |
| | hazard probit | ----- | ------ | ------ | ------ | ------ | ------ | ------ | ------ | ------ | ------ | ------ |
| model 2009 | static logit | 2 (3) | 2 (3) | 3 (2) | 3-4 (2) | 3-4 (2) | 2 (3) | 3 (2) | reject (0) | 3 (2) | 19 | 3-4 |
| | hazard logit | 3 (2) | 1 (4) | 4 (1) | 3-4 (2) | 3-4 (2) | 1 (4) | 4 (1) | reject (0) | 2 (3) | 19 | 3-4 |
| | static probit | 1 (4) | 3-4 (2) | 1-2 (4) | 2 (3) | 1-2 (4) | 4 (1) | 1 (4) | accept (1) | 4 (1) | 24 | 1-2 |
| | hazard probit | 4 (1) | 3-4 (2) | 1-2 (4) | 1 (4) | 1-2 (4) | 3 (2) | 2 (3) | reject (0) | 1 (4) | 24 | 1-2 |
| model 2010 | static logit | 2 (3) | 2 (3) | 1 (4) | 1 (4) | 1 (4) | 3 (2) | 2-3 (3) | reject (0) | 1-4 (4) | 27 | 1 |
| | hazard logit | 3 (2) | 1 (4) | 3 (2) | 2 (3) | 2 (3) | 1 (4) | 1 (4) | reject (0) | 1-4 (4) | 26 | 2 |
| | static probit | 1 (4) | 4 (1) | 2 (3) | 3 (2) | 3 (2) | 4 (1) | 4 (1) | reject (0) | 1-4 (4) | 18 | 3 |
| | hazard probit | 4 (1) | 3 (2) | 4 (1) | 4 (1) | 4 (1) | 2 (3) | 2-3 (3) | reject (0) | 1-4 (4) | 16 | 4 |
| model 2011 | static logit | 1 (4) | 3 (2) | 2 (3) | 1 (4) | 1-2 (4) | 3 (2) | 3 (2) | reject (0) | 1-2 (4) | 25 | 1-2 |
| | hazard logit | 3 (2) | 1 (4) | 3 (2) | 2 (3) | 1-2 (4) | 1 (4) | 1 (4) | reject (0) | 3-4 (2) | 25 | 1-2 |
| | static probit | 2 (3) | 4 (1) | 1 (4) | 3-4 (2) | 3-4 (2) | 4 (1) | 4 (1) | reject (0) | 1-2 (4) | 18 | 3 |
| | hazard probit | 4 (1) | 2 (3) | 4 (1) | 3-4 (2) | 3-4 (2) | 2 (3) | 2 (3) | reject (0) | 3-4 (2) | 17 | 4 |
| model 2012 | static logit | 1 (4) | 4 (1) | 1 (4) | 2 (3) | 1-2 (4) | 3 (2) | 4 (1) | reject (0) | 1-2 (4) | 23 | 2 |
| | hazard logit | 3 (2) | 1 (4) | 3 (2) | 4 (1) | 3-4 (2) | 2 (3) | 1 (4) | reject (0) | 1-2 (4) | 22 | 3 |
| | static probit | 2 (3) | 3 (2) | 2 (3) | 1 (4) | 1-2 (4) | 1 (4) | 3 (2) | accept (1) | 3-4 (2) | 25 | 1 |
| | hazard probit | 4 (1) | 2 (3) | 4 (1) | 3 (2) | 3-4 (2) | 4 (1) | 2 (3) | reject (0) | 3-4 (2) | 15 | 4 |

The table provides overall rankings of the models for the criteria considered in our study. We have excluded number of variables, log-likelihood of calibration accuracy (CA) test (in-sample), as we cannot use cross comparison (static versus hazard model) here, and mean value of bootstrapped ROC areas in order to avoid double consideration of this measure. Models in particular years are ranked according to individual criterion and assign with points as follows: the model ranked highest gets 4 points, the second best 3 points, the third one 2 points, and the last one 1 point. For Hosmer-Lemeshow (HL) test, acceptation is rewarded by 1 point while rejection by none. First number denotes ranking while second number in parenthesis denotes assigned points. Final rankings of the models are then given in the last column.

in order to avoid double consideration of this measure. In a first step, all models for a particular year are ranked with respect to each of the considered criteria. We then assign points for each model in the following way: the model ranked highest gets 4 points, the second best 3 points, the third one 2 points, and the last one 1 point. For the Hosmer-Lemeshow test, if the null hypothesis cannot be rejected, the model receives 1 point, while a rejection of the model yields zero points. The table also provides the final rankings of the models for each particular year. For 2008, we find that the static logit model outperforms the static probit model. For 2009, the static probit and hazard probit model obtain the same score and outperform the static and hazard logit models which have also the same score. For 2010, the static logit model is ranked first, followed by the hazard logit model, the static probit model, and the hazard probit model. For 2011, the static and hazard logit models have the same score and seem to outperform the static probit model by 7 points and the hazard probit model ranked last by 8 points. Finally, for 2012, the static probit model is ranked first, followed by the static logit model, the hazard logit model, and the hazard probit model.

## 3.5 Conclusions

In this study, we estimate and investigate credit-scoring models for determining default probabilities of financial institutions. We contribute to the existing literature on rating models for financial institutions by taking advantage of the fact that many U.S. commercial banks defaulted during the GFC and subsequent periods, which enabled us to compile and examine a significant database of historical financial ratios for defaulted banks. We provide the first empirical study to use the Federal Financial Institutions Examination Council (FFIEC) database and to provide scoring models for these banks. This database contains an extensive sample of more than seven thousand U.S. commercial banks with over 400 defaults during our sample period 2007-2013. We compare two types of models in this study: static models and dynamic discrete hazard models. We apply logistic and probit regression techniques in order to calibrate our models and a rolling window methodology (the walk-forward approach) allowing for out-of-time validation of the estimated models.

Substantial part of this study is devoted to the application of model evaluation techniques. Apart from well-known techniques, such as ROC analysis with bootstrapping of areas under the ROC curves or calibration accuracy tests, we also apply the Kruskal-Wallis and the Tukey's test to investigate significant differences between the particular models in terms of bootstrapped ROC areas. Furthermore, we apply Vuong's closeness test for non-nested models to determine whether calculated log-likelihoods for various models are statistically different for the estimated models. Finally, we use the Hosmer-Lemeshow's chi-squared goodness-of-fit test to examine the overall fit of the estimated models.

The majority of the estimated models builds on variables that form a reasonable mixture of profitability, liquidity, assets quality, and capital adequacy indicators. We find that our models have a high default/non-default classification and predictive accuracy. Specifically, for the models that were calibrated using defaults in 2011 and 2012, more than 95% of defaulted banks being captured within the banks with the highest 10% PDs. These are very good results compared to recent studies conducted on the corporate sector. Since all the models perform very well and their performances are similar in terms of power (areas under the ROC curves) we use the Kruskal-Wallis and the Tukey's multiple comparison test to examine significant differences between the particular models in terms of bootstrapped ROC areas. Specifically, the Tukey's test proves to be a very powerful tool as it is able to distinguish between the models where the differences between mean values of bootstrapped ROC areas are very small. Using a calibration accuracy test and its likelihood estimates we show that logit models typically outperform probit models in accuracy of estimated PDs in particular years. We also find that multi-period hazard models generally produce more accurate default probability estimates compared to static models.

We state that our estimated default probabilities might be considered as reasonable estimates since we show and prove by accepting the null hypothesis in Hosmer-Lemeshow's chi-squared tests (except of the first deciles containing 10% banks with the worst rating) that the expected and actual default rates are statistically equal for most of the deciles. Also, due to the fact that we work with all of the available information on U.S. commercial banks and thus avoid choice-based samples within estimation, we obtain ratios of defaulted and non-defaulted banks very close to empirical ones. This is necessary in order to produce estimates that are close to "real" PDs.

Finally, due to the number of estimated models and the fact that different models perform best according to different criteria, we provide a summary of comparison for all the models in terms of the chosen criteria and create a simple ranking system in order to determine which model works the best for a particular year.

# Chapter 4

# Distress Risk and Stock Returns of U.S. Renewable Energy Companies

While in the previous chapters we have focused on estimation of default probabilities using various methods (structural credit risk models in Chapter 2 and credit-scoring models in Chapter 3), in our last study, we take advantage of the "outputs" of Moody's KMV model – Expected Default Frequencies (EDFs) – and use these structural-based default probability indicators in asset pricing framework applied to U.S. renewable energy companies.

During the last decade, the renewable energy sector has undergone significant overall growth in the global economy and several renewable energy, clean energy or so-called alternative energy stock indices have been created. Prominent indices for the sector include, for example, the WilderHill Clean Energy Index (ECO), the WilderHill New Energy Global Innovation Index (NEX), or the S&P Global Clean Energy Index (SPGCE) (Inchauspe, Ripple and Trück [94]). At the same time, similar to technology stocks or venture capital, investments into renewable energy stocks can be considered as being relatively risky, see, e.g., Henriques and Sadorsky [86], Kumar, Managi and Matsuda [108], Sadorsky [150], or Managi and Okimoto [125]. In general, only a fraction of renewable energy companies become really successful, while many others go bankrupt or are acquired after some time. This goes hand in hand with the nature of their business – there is often a significant gap between innovation, adoption, and a phase where the company really becomes established on the market. This gap is often referred to as the "Valley of Death", see, e.g., Weyant [177]. As a result, one may argue that in particular investors who buy shares in small and/or highly risky renewable energy companies, i.e. stocks with typically higher volatility and probability of default, will also expect higher average returns for bearing this risk. Our study aims to shed light into this important question and thoroughly examines the relationship between distress risk and returns in the U.S. renewable energy sector.

For other sectors, there has been some controversy with regards to the relationship between distress risk and (expected) returns in equity markets, arising from several prominent studies. Two major studies report a positive cross-sectional relationship between default risk and equity returns. Vassalou and Xing [172] argue that firms with high default risk on average earn higher returns than low default risk firms, however, this holds only to the extent that they are small in size and have high book-to-market (BM) ratios. Chava and Purnanandam [39], using estimated ex-ante expected returns based on the implied cost of capital, also find strong support for this positive relationship. On the other hand, several other key studies suggest that distress risk is priced negatively, i.e. stocks of companies with higher default risk usually yield lower average returns. In the literature, this controversial relationship is often referred to as the "distress risk puzzle", see, e.g., Dichev [47], Griffin and Lemmon [78], Campbell, Hilscher and Szilagyi [34], Garlappi, Shu and Yan [73], or Avramov, Chordia, Jostova and Philipov [11].

Recently, there has also been a rising interest in examining returns of renewable energy companies, as well as identifying potential drivers of these returns, see, e.g., Henriques and Sadorsky [86], Kumar, Managi and Matsuda [108], Sadorsky [150], Bohl, Kaufmann and Stephan [25], or Managi and Okimoto [125]. These studies typically focus on the relationship between renewable energy stocks, changes in the oil price, equity indices and carbon prices. The authors report evidence for the impact of several of these variables on renewable energy stock prices or returns and suggest that in particular returns of high technology and renewable energy stocks seem to be significantly correlated.[45] However, none of these studies has examined how distress risk is priced in the renewable energy sector.

In this study, we contribute to the literature by combining work on the relationship between distress risk and equity returns with studies that focus on the driving factors of

[45] As Inchauspe et al. (2015) argue, a possible explanation for this phenomenon is that high technology and renewable energy companies often compete for the same inputs. These resources might include highly-qualified engineers and researchers, research facilities, semi-conductors, integrated circuits and thermoelectric materials, among others.

returns of renewable energy companies. In particular, we provide the first empirical study that investigates the question whether distressed renewable energy companies earn on average higher returns than renewable stocks of companies with low default risk. Thus, we examine whether, on top of the widely used Fama and French [65] and Carhart [37] risk factors, distress risk is priced in the renewable energy sector.

We use the *Expected Default Frequency* (EDF) measure obtained from one of the major rating agencies (Moody's KMV) as a proxy for distress risk. The EDF measures the probability that a company will default over a specified period of time (typically one year). It is based on the so-called structural approach to modeling default risk for a borrower, initially introduced by Merton [135]. One key advantage of this measure is its availability at a daily frequency, what clearly distinguishes it from other measures of default risk that are based on balance sheet data and updated only very infrequently. Thus, using EDFs allows us to construct portfolios of renewable energy stocks sorted by distress risk on a relatively high frequency, such as e.g. a monthly basis. This also allows us to investigate the performance of the constructed portfolios on a monthly basis as it is typically done in the literature that motivates our study, see, e.g., Fama and French [65], Carhart [37], Vassalou and Xing [172], Boyer and Filion [29], Campbell, Hilscher and Szilagyi [34], Garlappi, Shu and Yan [73]. Note that in comparison to most earlier studies focusing in particular on the renewable energy sector, e.g., Henriques and Sadorsky [86], Kumar, Managi and Matsuda [108], Sadorsky [150], Bohl, Kaufmann and Stephan [25], Managi and Okimoto [125], we also significantly extend the time period considered by using a data set of monthly returns from 2002 up to 2014. Thus, our sample period includes observations for the period of the global financial crisis as well as a significant sample period after the crisis. Furthermore, unlike many above mentioned studies that typically look at one of the renewable energy stock indices, we examine returns of individual renewable energy companies in the U.S. market. Thus, next to examining the pricing of distress risk in the renewable sector, our analysis is expected to provide additional insights on how market risk (measured by beta), size, and book-to-market (BM) effects are priced for renewable energy companies.

This chapter is organized as follows. We provide a review of the existing literature on the pricing of distress risk in equity markets as well as on investigating returns of renewable energy companies in Section 2. Section 3 is devoted to a brief description of three well-known asset pricing models and the construction of particular pricing factors. The data used in this study and statistics of returns and the EDF measure are described in Section 4. In Section 5, we investigate the relationship between distress risk and equity returns in the renewable energy sector, along with examining a possible link between pricing factors such as the size effect, the book-to-market effect and distress risk. Results for the pricing of distress risk are reported in Section 6. Finally, we conclude in Section 7 with a summary of our results.

## 4.1 Literature review

The trade-off between distress risk and stock returns has important implications for the risk-reward relationship in financial markets. In line with the fundamental principle of financial theory, investors will require higher average returns for bearing additional risk. Thus, investors should also expect a compensation for holding more distressed stocks. This risk-reward trade-off is the main idea behind the conceptual framework of asset pricing and investment decision making in efficient markets. However, the existing empirical literature has not produced consistent evidence to confirm the above conjecture for distress risk. In fact, several studies have shown the opposite – more distress stocks usually earn lower average returns, see, e.g., Dichev [47], Griffin and Lemmon [78], or Campbell, Hilscher and Szilagyi [34].

Dichev [47] was among the first to demonstrate the negative cross-sectional relationship between default risk and future stock returns, measuring default risk by the Altman [4] *Z*-score and Ohlson [141] *O*-score. These results suggest that default risk is not rewarded by higher returns, hence it casts doubt on the notion of a market premium for distress risk. Moreover, the results suggest that the relation between default risk and book-to-market is not monotonic: distressed firms generally have high book-to-market values but the most distressed firms have lower book-to-market values. Griffin and Lemmon [78] confirm Dichev [47] findings and by using Ohlson [141] *O*-score also find a negative relationship between distress risk and realized stock returns. The authors also report that the difference in returns between high and low book-to-market stocks is more than twice as large as that in other firms, suggesting that they may be mispriced.

On the other hand, using default likelihood indicators based on the Merton [135] model, Vassalou and Xing [172] find evidence for distressed stocks earning higher returns, in particular in the small value segment. Therefore, these results suggest the presence of an equity return premium for distress risk. The authors also argue that default risk is closely related to size and book-to-market effects and that these two characteristics can be viewed as default effects. The findings also indicate that book-to-market and size effects are concentrated in high default risk companies – the size effect exists only within the quintile with the highest default risk and the book-to-market effect only in the two quintiles with the highest default risk. Moreover, they demonstrate that default risk is systematic and therefore priced in the cross-section of equity returns.

Based on a hazard model that incorporates accounting and market variables as covariates in the spirit of Shumway [157], Campbell, Hilscher and Szilagyi [34] show that firms with high default probabilities have abnormally low expected returns. Thus, they argue that distress risk cannot explain the size and value premiums. In fact, distressed portfolios have low average returns, but high standard deviations, market betas, and loadings on Fama and French [65] size and value factors. They also tend to do poorly when market-wide implied volatility increases. Interestingly, Campbell, Hilscher and Szilagyi [34] also find evidence for the ʼdistress effect' being most pronounced among small and illiquid stocks, however, this means that these stocks yield particularly low returns for these stocks. Garlappi, Shu and Yan [73] use Moody's KMV default measure and confirm the negative relationship between default risk and stock returns. Their proposed mechanism relies on the effects of strategic interactions between equity holders and debt holders on equity returns. They argue that potential violations of the absolute priority rule for claimants at bankruptcy can help explain this negative correlation, because distressed stocks have lower betas and, therefore, earn lower returns. Avramov, Chordia, Jostova and Philipov [11] demonstrate that most of the negative returns for high default risk stocks are concentrated around rating downgrades. Consistent with Campbell, Hilscher and Szilagyi [34], they find that this effect is even more limited in the cross-section and is driven by a small segment of the worst-rated stocks. Their study indicates that profits of momentum strategies that buy 'winners' and sell 'losers' are remarkably concentrated among a small subset of firms with low credit ratings, which adds a new dimension to the complex relationship between financial distress and cross-sectional properties of equity returns. Their work also sheds new light on the debate about a priced distress risk factor in equity returns. While Chan and Chen [38] and Fama and French [66] argue that the size and book-to-market effects proxy for a priced distress risk factor, and Vassalou and Xing [172] find evidence that the size and book-to-market factors contain some default-related information, their results are in line with Dichev [47] and Campbell, Hilscher and Szilagyi [34], who conclude that distress risk is unlikely to be systematic.

Apart from Vassalou and Xing [172], the study by Chava and Purnanandam [39] is another one where strong support for the positive cross-sectional relationship between default risk and stock returns is found. These authors construct indices based on accounting numbers, options, and hazard models, and unlike prior studies that use noisy ex post realized returns to estimate expected returns, they use ex ante estimates based on the implied cost of capital. Their results suggest that investors expected higher returns for bearing default risk, but they were negatively surprised by lower-than-expected returns on high default risk stocks in the 1980s. Finally, Garlappi and Yan [74] explicitly consider financial leverage and study the cross-sectional implications of potential shareholder recovery upon resolution of financial distress. Contrary to Griffin and Lemmon [78] and Vassalou and Xing [172], they document that the value premium is hump-shaped instead of monotonically increasing in default probability. It increases when levels of EDF are low and declines sharply at very high levels of EDF.

Increased interest in the effects of energy and stock market prices (oil prices in particular) on the financial performance of the renewable sector has been well documented by a number of empirical studies. Faff and Brailsford [61] examine the relationship between oil prices and stock market indices of various industries in Australia and find significant effects of the oil price on equity returns, in particular for the oil, gas, resource and building industry stocks. Sadorsky [152] finds positive effects of an increasing oil price on Canadian oil and gas stocks. His results are confirmed by Boyer and Filion [29] who find evidence of a significant relationship between oil and natural gas prices, respectively, and stock returns of Canadian oil and gas companies.

Henriques and Sadorsky [86] use a four-variable vector-autoregressive model to account for the relationship between returns on renewable energy stocks, technology stocks, crude oil price and interest rates. They report evidence of Granger causality from crude oil prices to stock prices for renewable energy companies listed on major U.S. stock exchanges, and of the behaviour of renewable energy stock prices closely mirrored those of technology stock prices. Sadorsky [150] applies multivariate GARCH and dynamic conditional correlation models to examine volatility spillover effects between oil prices, technology stocks and clean energy companies. The results of this study suggest that renewable energy stock prices correlate more intensively with technology stock prices than with oil prices. Consequently, they argue that technology stocks cannot be considered a good hedge, while due to significantly lower correlations oil provides a more useful hedge for clean energy stocks. Using a variable beta model, Sadorsky [151] studies the determinants of systematic risk for U.S. listed renewable energy stocks between 2001 and 2007 and documents that renewable energy stocks exhibit substantial market risk. In fact, the study shows that a rise in oil prices has a positive impact on the beta of renewable energy stocks.

Kumar, Managi and Matsuda [108] also examine the relationship between alternate energy prices, oil prices, technology stocks and interest rates, but extend the analysis by including carbon prices. Similar to Henriques and Sadorsky [86], they apply a vector-autoregressive model and suggest that both the oil price and technology stock prices separately affect stock prices of clean energy firms. However, carbon allowance prices had no significant effects on renewable energy stocks. Managi and Okimoto [125] extend previous work by analysing data up to 2010 and apply Markov-switching vector autoregressive models to detect possible structural changes in the oil-renewable energy stock price relationship. They find evidence of a structural change occurring in late 2007, a period where a significant increase in the price of oil coincides with the U.S. economy entering into a recession. In contrast to Henriques and Sadorsky [86], the authors find a positive relationship between oil and the prices of clean energy stocks after 2007, suggesting a movement from conventional energy to clean energy. A copula approach is applied in Reboredo [148], where the author examines systemic risk and dependence between oil and renewable energy markets. By computing the conditional value-at-risk as a measure of systemic risk, the author finds significant time-varying dependence as well as symmetric tail dependence between oil returns and several global and sectoral renewable energy indices.

Broadstock, Cao and Zhang [31] and Wen, Guo, Wei and Huang [175] are then studies predominantly focusing on renewable energy markets in China. While Broadstock, Cao and

Zhang [31] show that oil price dynamics impacted on energy stocks in China, especially after the onset of the recent global financial crisis, when correlation increased significantly, Wen, Guo, Wei and Huang [175] use an asymmetric Baba–Engle–Kraft–Kroner (BEKK) model and document mean and volatility spillover effects between Chinese renewable energy and fossil fuel companies. Finally, Cummins, Garry and Kearney [44] perform a price discovery analysis to determine Granger causality relationships for a range of prominent green equity indices with broader equity and commodity markets. Contrary to Henriques and Sadorsky [86], who use one global index, or Kumar, Managi and Matsuda [108], who use three specific global indices, their study use an expanded database of green energy indices by including two prominent global indices, one sectoral index, and one regional index. Also, in order to overcome drawbacks of the conventional vector autoregression (VAR) model, they apply an asymmetric vector autoregressive (AVAR) model as a first layer of robustness to examine Granger causality between the variables of interest in their study.

A number of studies have also examined the factors that drive the performance of renewable energy stocks. There is also some literature investigating the impact of the Fukushima Daiichi nuclear disaster in Japan in March 2011 on nuclear and renewable energy stocks. Ferstl, Utz and Wimmer [68] examine this impact on alternative energy stocks in France, Germany, and Japan, and find positive abnormal returns for these stocks. Next, a study by Betzer, Doumet and Rinne [17] examines the severe reaction of the German Federal Government that included the temporary shutdown of almost half of the nation’s nuclear power plants. Lopatta and Kaspereit [121] argue that the more an energy company had relied on nuclear power, the more its share price declined after the Fukushima accident. Further investigating the issue, they suggest that energy companies could prevent increases in market beta due to such events by shifting some of their energy production from nuclear to renewable or other sources.

Bohl, Kaufmann and Stephan [25] apply a four-factor asset pricing model and study the behaviour of German renewable energy stocks. Their results suggest that while renewable energy stocks earned considerable risk-adjusted returns between 2004 and 2007, the performance has deteriorated significantly, delivering negative returns since 2008. Ortas and Moneva [143] study the time-varying beta behaviour of 21 clean-technology equity indices, finding that these indices yield higher returns and risk than conventional stock indices.

Moreover, they also find a structural change in the dynamics of clean technology indices' return/risk performance that coincides with the beginning of the financial crisis. The dynamics of excess returns for the NEX index are examined in Inchauspe, Ripple and Trück [94]. The authors propose a multi-factor asset pricing model with time-varying coefficients to study the role of energy prices and stock market indices as explanatory factors and find a strong influence of the MSCI World index and technology stocks throughout the sample period. Finally, Bohl, Kaufmann and Siklos [24] analyse whether the explosive price behavior of renewable energy stocks during the mid-2000s was driven by rising crude oil prices and overall bullish market sentiment. They suggest strong evidence of explosive price behavior for European and global sector indices, even after controlling for a set of explanatory variables.

Overall, during the last decade, due to substantial growth in the sector there has been an increased interest in examining the performance of renewable energy companies, as well as in identifying potential drivers of this performance. While some of the recent studies, see, e.g., Bohl, Kaufmann and Stephan [25], Inchauspe, Ripple and Trück [94] , have also included standard pricing factors such as market risk, Fama and French [66] size and value factors or a Carhart [37] momentum factor, none of these studies has examined how distress risk is priced in the renewable energy sector. We believe, however, that given the structure of the renewable energy sector with a significant gap between innovation, adoption, and a phase where the company really becomes established, i.e. a high number of small, innovative but also highly risky companies, distress risk may play a significant role as it comes to determining investors’ return expectations for individual companies.

## 4.2 Asset pricing models

Typically, the literature investigates the existence or pricing of a distress risk premium in a factor model set-up. This section briefly summarises three well-known asset pricing models that have been heavily used in the past.

Building on the Markowitz framework [130, 131], Sharpe [156], Lintner [115], and Mossin [138] independently developed the so-called Capital Asset Pricing Model (CAPM) to explain the behavior of common stock returns. In this model, all investors combine the market portfolio and the risk-free asset such that the only risk investors are compensated for is the systematic risk associated with the market portfolio. Therefore, the CAPM is often denoted as a so-called one-factor model. Several empirical studies however have shown that this model actually does not perform that well. In 1992, an influential paper by Fama and French [66] was published summarizing much of the earlier empirical work in the area. As a result, Fama and French [65] introduced a new three-factor model where, in addition to a market risk factor, a size (market capitalization) and value (book-to-market ratio) factor were added. It has been shown that this model tends to produce significant coefficients on all three factors and that the three factors are capturing much of the common variation in portfolio returns. Following the success of the model, other factors based on individual stock characteristics have been proposed in the literature, most notably the momentum factor introduced by Carhart [37], which is based on the observation that stocks with a high past performance (winners) outperform stocks with a low past performance (losers) in the next 3-12 months.

### 4.2.1 The Capital Asset Pricing Model (CAPM)

The CAPM is defined as

$$R_{i,t} - R_{F,t} = \alpha_i^{CAPM} + \beta_i \left( R_{M,t} - R_{F,t} \right) + \varepsilon_{i,t} , \tag{4.1}$$

where $R_{i,t}$ denotes the return of a company or portfolio $i$ at time $t$, $R_{F,t}$ is the risk-free interest rate at time $t$, and $R_{M,t}$ is the market return at time $t$. $\alpha_i^{CAPM}$ and $\beta_i$ are estimated coefficients, where $\alpha_i^{CAPM}$ represents the average return in excess of the reward for the exposure to the market factor (it is often referred to as the *abnormal* or *active return* of an asset), and $\varepsilon_{i,t}$ is the independently and normally distributed error term from this regression, $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$. The idea behind this model is that excess return of a particular company or portfolio $\left(R_{i,t} - R_{F,t}\right)$ can be explained by their relationship with a market risk factor $\left(R_{M,t} - R_{F,t}\right)$.

### 4.2.2 The Fama-French three-factor model

The Fama-French three-factor model is denoted by the following equation

$$R_{i,t} - R_{F,t} = \alpha_i^{3F} + \beta_{1,i}\left(R_{M,t} - R_{F,t}\right) + \beta_{2,i}\text{SMB}_t + \beta_{3,i}\text{HML}_t + \varepsilon_{i,t}\,. \tag{4.2}$$

Hereby $\text{SMB}_t$ is the realization on a capitalization-based factor portfolio that buys small cap stocks and sells large cap stocks. Similarly, $\text{HML}_t$ is the realization on a factor portfolio that buys high BM (book-to-market) stocks and sells low BM stocks. The $\beta_{2,i}$ and $\beta_{3,i}$ coefficients measure the sensitivity of the portfolio's return to the small-minus-big (SMB) and high-minus-low (HML) factors, respectively.

### 4.2.3 The Carhart four-factor model

The Carhart four-factor model then introduces an additional momentum factor and can be denoted by

$$R_{i,t} - R_{F,t} = \alpha_i^{4F} + \beta_{1,i}\left(R_{M,t} - R_{F,t}\right) + \beta_{2,i}\text{SMB}_t + \beta_{3,i}\text{HML}_t + \beta_{4,i}\text{MOM}_t + \varepsilon_{i,t}\,, \tag{4.3}$$

where $\text{MOM}_t$ is the prior one-year price momentum factor that captures the return spread between portfolios of past winner and past loser stocks.

The size, value, and momentum factors are constructed in the following way. First, monthly stock returns are calculated and sorted according to the value of firm characteristics (the explanatory factors). Second, the stocks are divided into relevant groups (portfolios), according to their factor rank, and the difference in portfolio returns between high rated and low rated stocks according to these characteristics is calculated. In particular, the SMB (small-minus-big) factor is based on the difference in portfolio returns between stocks with a small market capitalization and stocks with a big market capitalization, the HML (high-minus-low) factor is based on the difference between stocks with a high book-to-market equity ratio and a low book-to-market equity ratio, and the MOM factor is based on the difference between winner and loser portfolios. For closer discussion on construction of these factors we refer to Fama and French [65], Carhart [37], or Professor Kenneth French's website[46]. For the risk-free interest rate $R_F$, we use one-month Treasury Bill rates and the market return $R_M$ is calculated as the value-weighted return of all CRSP firms incorporated in the U.S. and listed on the NYSE, AMEX, or NASDAQ.

[46] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

## 4.3 Data description

Our sample contains U.S. renewable energy companies listed on the NYSE, AMEX, or NASDAQ stock exchanges that are or were components of the following renewable, clean or alternative energy indices: the WilderHill Clean Energy Index (ECO), the WilderHill New Energy Global Innovation Index (NEX), the Ardour Global Alternative Energy Index North America (AGINA), the Renewable Energy Industrial Index (RENIXX World), the ALTEXGlobal Index (ALTEXGlobal), the NASDAQ Clean Edge Green Energy Index (CELS), and the ISE Global Wind Energy Index (GWE). As a matter of fact, many companies are or were components of two or more of these indices.

The WilderHill Clean Energy Index (**ECO**) tracked 48 Clean Energy companies as of July 2015. Specifically, businesses that stand to benefit substantially from a societal transition towards the use of cleaner energy and conservation. Stocks and sector weightings within the ECO Index are based on their significance for clean energy, technological influence and relevance to preventing pollution in the first place.[47] The index has six sub-sectors: renewable energy harvesting (25% sector weight, 11 stocks), power delivery and conservation (21%, 9 stocks), energy conversion (19%, 10 stocks), greener utilities (17%, 7 stocks), energy storage (9%, 5 stocks), and cleaner fuels (9%, 6 stocks). The largest company accounts for 3.30% and the top 5 holdings account for 15.52% of total investments into the ECO. There is a strong focus in favour of pure-play companies in wind power, solar power, hydrogen and fuel cells, biofuels, and related fields. Market capitalization for a majority of Clean Energy Index stocks is typically $200 million and above. The index focuses on North American companies and is listed in the U.S. only.

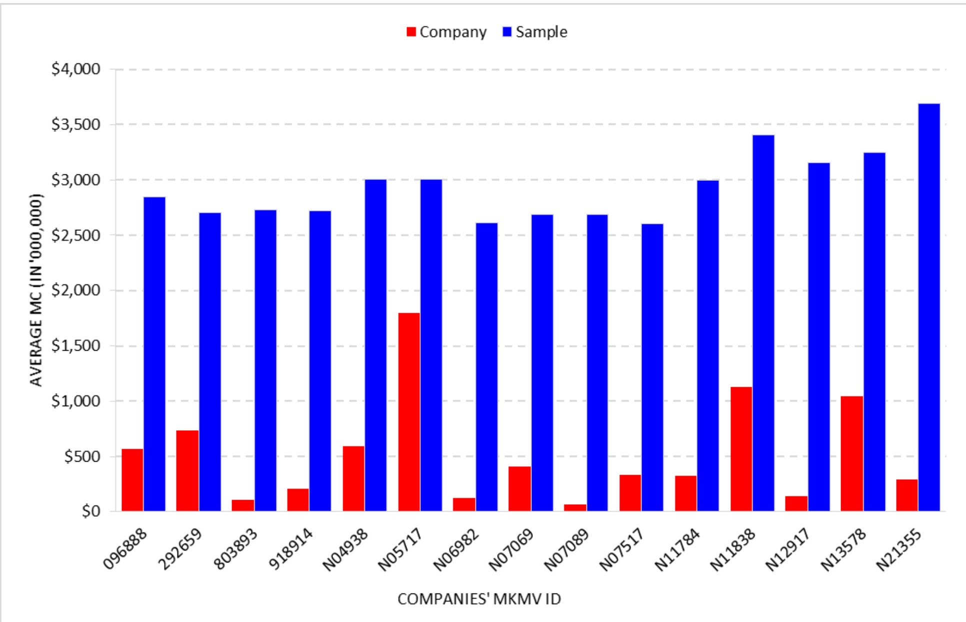
The WilderHill New Energy Global Innovation Index (**NEX**) focuses on the generation and use of renewable energy, and the efficiency, conservation and advancement in renewable energy in general.[48] The index was composed of 107 companies in 27 countries as of July 2015. The largest company accounts for 1.87% and the top 10 holdings account for 18.7% of total NEX investment. As of September 2014, the index was composed of seven sub-sectors: storage (2.3%), energy conversion (1.4%), and projects related to renewable energy other

[47] Source: http://www.wildershares.com/. Accessed: July 2015.
[48] Source: http://www.nexindex.com/. Accessed: July 2015.

than the above (12.2%). The investments are distributed by regions with weights of 41.2% for the Americas, 29.6% for Asia and Oceania, and 29.2% for Europe, the Middle East and Africa. For a stock to be included in this index, the company must be identified as one that has a meaningful exposure to clean energy, either as a technology, equipment, service or finance provider, such that profitable growth of the industry can be expected to have a positive impact on that company's performance. Market capitalization for a majority of NEX index stocks is typically $250 million and above.

The **AGINA** index, as a part of the Ardour Global Alternative Energy Indices, merely focuses on North American renewable companies and tracked 55 companies as of June 2015. The largest company accounts for 2.02% and the top 5 holdings account for 9.75% of total AGINA investment. Companies included in this index are involved in alternative energy resources (solar, wind, hydro, tidal, wave, geothermal and bio-energy), energy efficiency, and others. The **RENIXX World** index is run by the International Economic Platform for Renewable Energies and was established in May 2006. It is the first global stock index, which tracks the performance of the world's 30 largest companies in the renewable energy sector. Companies

**Figure 4.1**
Defaulted companies (average size)




The figure compares the average size, expressed by market capitalization (MC) in '000,000 of $, of 15 defaulted companies in our sample with the average size of the sample over the period when a given company was active on the market in our sample period.

must achieve at least 50 percent of their revenue in the renewable energy industry coming from wind energy, solar power, biomass, geothermal energy, hydropower or fuel cells to be included in the index. The **ALTEXGlobal** index is run by Bakers Investment Group and serves as a benchmark index for Alternate Energy internationally. Tracking 138 companies it is the world's largest Alternative Energy Index with an aggregated market capitalization of $1.16 trillion USD. The **CELS** index is a modified market capitalization-weighted index designed to track the performance of U.S.-traded clean energy companies. As of March 2015, the index was composed of 46 companies. Finally, the **GWE** index provides a benchmark for investors interested in tracking public companies that are identified as providing goods and services exclusively to the wind energy industry. This global index was composed of 44 companies (the largest company accounts for 8.49% and the top 5 holdings account for 37.62% of total GWE investment) as of July 2015.

We match the MKMV (*Moody's KMV*) database with the CRSP (*The Center for Research in Security Prices*) and COMPUSTAT databases, both available through *Wharton Research Data Services* (WRDS). In order to be included in our sample, all chosen companies need to be present simultaneously in all three databases. Specifically, for a given month, the following

**Table 4.1**
**Defaulted companies**

| MKMVID | Company's name | Date of Chapter 11 filling | Date of delisted return | Date of first reported EDF = 35 | Date of last reported EDF = 35 | Mean EDF | Trend EDF | Trend price |
|---|---|---|---|---|---|---|---|---|
| 096888 | ENER1 INC | 26/01/2012 | Oct - 11 | Nov - 11 | Mar - 12 | 9.51 | ↗ | ↘ |
| 292659 | ENERGY CONVERSION DEV | 14/02/2012 | Feb - 12 | May - 11 | Aug - 12 | 7.46 | ↗ | ↘ |
| 803893 | SATCON TECHNOLOGY CORP | 17/10/2012 | Oct - 12 | Jul - 12 | Mar - 14 | 10.30 | ↗ | ↘ |
| 918914 | VALENCE TECHNOLOGY INC | 12/07/2012 | Jul - 12 | Jun - 12 | Mar - 14 | 10.52 | ↗ | ↘ |
| N04938* | USEC INC / CENTRUS ENERGY CORP | 3/05/2014 | ----- | Apr - 12 | Dec - 14 | 12.22 | ↗ | ↘ |
| N05717** | QUICKSILVER RESOURCES INC | 17/03/2015 | ----- | Oct - 14 | Dec - 14 | 7.47 | ↗ | ↘ |
| N06982 | DISTRIBUTED ENERGY SYS CORP | 4/06/2008 | Jun - 08 | Jun - 08 | Jun - 10 | 11.00 | ↗ | ↘ |
| N07069 | EVERGREEN SOLAR INC | 15/08/2011 | Aug - 11 | Jan - 11 | Jun - 12 | 9.19 | ↗ | ↘ |
| N07089 | BEACON POWER CORP | 30/10/2011 | Nov - 11 | Sep - 11 | Apr - 13 | 8.16 | ↗ | ↘ |
| N07517 | MEDIS TECHNOLOGIES LTD | xx/09/2011 | Aug - 09 | Nov - 09 | Mar - 11 | 7.47 | ↗ | ↘ |
| N11784 | RASER TECHNOLOGIES INC | 29/04/2011 | Nov - 10 | Jun - 10 | Sep - 11 | 14.40 | ↗ | ↘ |
| N11838 | VERASUN ENERGY CORP | 31/10/2008 | Nov - 08 | Oct - 08 | Jun - 10 | 16.39 | ↗ | ↘ |
| N12917 | NOVA BIOSOURCE FUELS INC | 31/03/2009 | Apr - 09 | Sep - 08 | Jan - 11 | 26.81 | ↗ | ↘ |
| N13578 | GT ADVANCED TECHNOLOGIES INC | 6/10/2014 | Oct - 14 | Oct - 14 | Dec - 14 | 6.85 | ↗ | ↘ |
| N21355 | KIOR INC | 9/11/2014 | Oct - 14 | Mar - 14 | Dec - 14 | 11.91 | ↗ | ↘ |

* emerged from bankruptcy (restructuring) as Centrus Energy Corp. on 30/09/2014
** defaulted in 2015

The table reports information about 15 defaulted companies in our sample. Namely, we report company's name, date when company filed for bankruptcy protection under Chapter 11, date of delisted return in CRSP database, dates of first and last maximum value (35) of EDF measure reported in MKMV database, along with average EDF value, and EDF and price trends towards default.

information on a company is required: share price, shares outstanding, return data from CRSP; accounting data from COMPUSTAT; and the EDF as a measure of default risk available from MKMV. Our sample spans the period from January 2002 to December 2014.

In order to avoid a survivorship bias by taking into account only companies currently being the components of the above mentioned indices, we also include companies that left the ECO index, i.e. the index whose components represent the biggest part of our sample, in earlier years. Overall, we work with a total of 141 companies, where 15 companies have defaulted (filed for bankruptcy protection under Chapter 11), 12 companies have been acquired, and another 16 companies have left the ECO index (but are still active). In total our sample contains 10.6% companies that have defaulted and 8.5% companies that have been acquired.

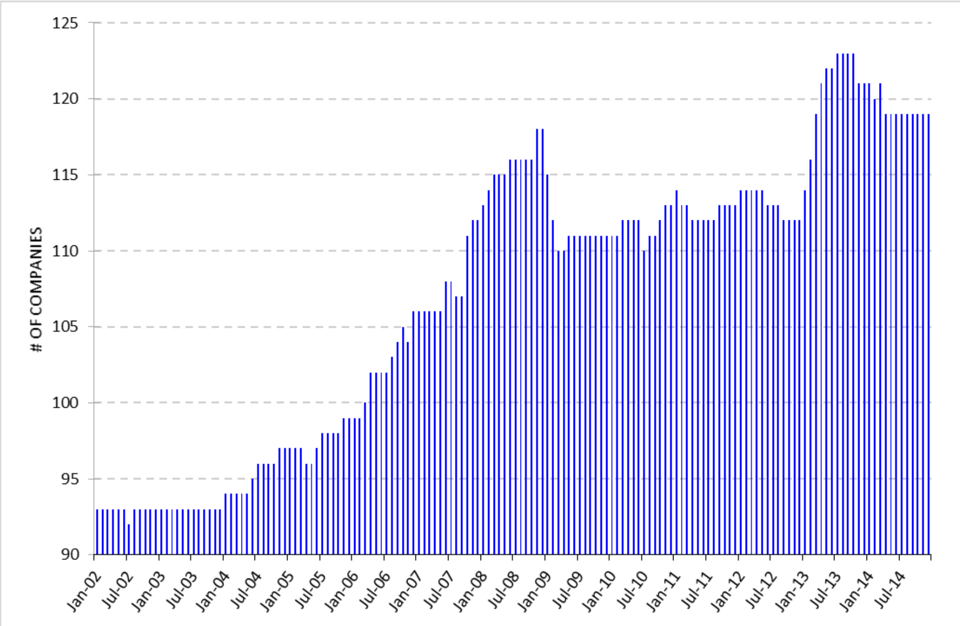
We investigate defaulted companies, in terms of size and corresponding EDF values, in Figure 4.1 and Table 4.1. We can see that all 15 defaulted companies are significantly smaller compared to average size of the sample (see Figure 4.1). Reported information from Table 4.1 show that average EDF values for these companies are considerably high with expected increasing trend and decreasing trend in stock prices towards default. In fact, most of the times the first maximum EDF value (EDF = 35) reported in MKMV database predate the date of filing for bankruptcy protection under Chapter 11 and the date of delisted return in CRSP database. Regarding the acquired companies, there are in general two possible reasons why

**Table 4.2**
**Acquired companies**

| MKMVID | Company's name | Acquired by | Date of acquisition | Last reported EDF | Trend EDF | Trend price |
|---|---|---|---|---|---|---|
| 029066 | AMERICAN POWER CONVERSION CP | SCHEIDER ELECTRIC | 14/02/2007 | 0.04 | ↘ | ↗ |
| 155771 | CENTRAL VERMONT PUB SERV | QUEBEC'S GAZ METRO | 27/06/2012 | 0.11 | ↘ | ↗ |
| 283695 | EL PASO CORP | KINDER MORGAN | 24/05/2012 | 0.10 | ↘ | ↗ |
| 458771 | INTERMAGNETICS GENERAL CORP | ROYAL PHILIPS ELECTRONICS | 9/11/2006 | 0.04 | ↘ | ↗ |
| 460254 | INTL RECTIFIER CORP | INFINEON TECHNOLOGIES | 13/01/2015 | 0.02 | ↘ | ↗ |
| 486587 | KAYDON CORP | SKF GROUP | 16/10/2013 | 0.05 | ↘ | ↗ |
| 834090 | SOLA INTERNATIONAL INC | CARL ZEISS VISION HOLDING | 22/03/2005 | 0.12 | ↘ | ↗ |
| 98975W | ZOLTEK COS INC | TORAY INDUSTRIES | 3/03/2014 | 0.07 | ↘ | ↗ |
| N03918 | POWER-ONE INC | ABB LTD | 25/07/2013 | 0.46 | ↘ | ↗ |
| N06112 | VERENIUM CORP | BASF | 31/10/2013 | 2.09 | ↘ | ↗ |
| N10271 | COLOR KINETICS INC | ROYAL PHILIPS ELECTRONICS | 27/08/2007 | 0.07 | ↘ | ↗ |
| N12496 | COMVERGE INC | H.I.G. CAPITAL | 22/05/2012 | 10.95 | ↗ | ↘ |

The table reports information about 12 acquired companies in our sample. Specifically, Moody's KMV ID (MKMVID), company's name, name of the company it was acquired by, date of acquisition, last reported EDF in MKMV database, and trend of EDF and trend of the price towards the date of acquisition.

**Figure 4.2**
Number of companies in the sample through time




The figure plots number of companies in our sample over the period from January 2002 to December 2014. We start with 93 companies in January 2002 and end with 119 companies in December 2014. Minimum number of companies is 92 (July 2002), maximum is then 123 (July-October 2013).

a given company may be taken over. First, the company does not perform well and the acquisition is the only way how to prevent likely default. Second, on the contrary, the company is performing very well and becomes desirable for acquisition by another more established firm, from which both companies could benefit. We examine acquired companies in our sample and probable reason for the acquisition in Table 4.2. Upon close investigation of the last reported EDFs, together with the EDF and price trends towards the date of acquisition, we conclude that 11 out of 12 companies were acquired due to a very good performance on the market. For these companies, EDF remained relatively low and was generally decreasing, while the stock price was generally increasing. The only exception is the Comverge Inc. (N12496) whose last reported EDF was relatively high (10.95). Also, the EDF was increasing towards the date of acquisition, while the stock price was decreasing. These findings suggest that the company was saved from potential bankruptcy by the acquisition with H.I.G. Capital.

Note that as it is typically done in asset pricing studies, we use monthly returns to measure the performance of the individual companies. Figure 4.2 provides a plot of the number of companies at each point in time throughout our sample period from January 2002 to December 2014.

We are particularly interested in the relationship between risk and return for renewable energy companies with a focus on the performance of distressed stocks (represented by the EDF measure) in this sector. Thus, in the following we devote two sections to descriptive statistics of these two crucial variables and make a comparison between our sample and the U.S. market as a whole.

### 4.3.1 Returns

As mentioned above, returns are collected from the CRSP database. Because we focus on studying the returns on distressed stocks, we follow Campbell, Hilscher and Szilagyi [34] to deal with the problem of delisted firms and use the delisted return for the final month of the company's life reported in the CRSP database for our defaulted and acquired companies.

Summary statistics for returns and the volatility of returns (volatility of equity) are reported in Table 4.3. We make a comparison between returns in our sample and in the overall U.S. market in Panel A. All returns are pooled together before summary statistics are calculated. That is, for our sample period we have 16,927 monthly observations of returns for 141 companies, while there are 1,052,610 monthly observations of returns for 13,239 companies contained in CRSP. We can see that on average returns in our sample are slightly

**Table 4.3**
**Summary statistics of returns and volatilities of returns**

| Panel A: Returns | | | Panel B: Volatilities of Returns | | |
|---|---|---|---|---|---|
| | Our Sample | US Universe | | Our Sample | US Universe |
| # of obs. | 16,927 | 1,052,610 | # of obs. | 141 | 13,239 |
| Mean | 0.012 | 0.010 | Mean | 0.193 | 0.142 |
| Median | 0.004 | 0.006 | Median | 0.163 | 0.119 |
| Std. | 0.219 | 0.155 | Std. | 0.282 | 0.109 |
| Skewness | 24.104 | 5.416 | Skewness | 10.575 | 4.332 |
| Kurtosis | 1,623.438 | 234.311 | Kurtosis | 120.855 | 75.414 |

The table reports summary statistics of returns (panel A) and of volatilities of returns (panel B) for our sample and the whole U.S. market. Specifically, apart from the number of observations (# of obs.), we report the mean, median, standard deviation (Std.), skewness, and kurtosis. In panel A, all returns are pooled together before summary statistics are calculated, while in panel B, for each stock we calculate the return volatility and summary statistics are subsequently calculated from the distribution of volatilities. All values are expressed in decimal units.

higher (1.2% compared to 1.0% for the U.S. market), although the median is lower (0.4% compared to 0.6%). Together with the fact that returns for our sample are more skewed and leptokurtic, these statistics suggest that we have more extreme return observations in our sample, i.e. observations with relatively high positive returns. Moreover, comparing the standard deviation of 0.219 in our sample to the significantly lower standard deviation of 0.155 in the entire U.S. market, we conclude that our companies, in terms of returns, are typically far more volatile.

This finding is confirmed by information reported in Panel B, where we look at individual companies and calculate the standard deviation of returns for each of them. As expected, on average standard deviations are higher in our sample (0.193 compared to 0.142 for the U.S. market). The distribution of standard deviations is also more skewed and has a higher kurtosis, and also the variation of the estimated volatilities is significantly higher in our sample (standard deviation of 0.282 compared to 0.109 in the U.S. market).

Overall, by examining these returns we confirm that renewable energy stocks are typically more volatile (or risky) in comparison to the entire universe of U.S. equities.

### 4.3.2 Expected Default Frequencies (EDFs)

Another key variable in our analysis is a distress risk factor represented by the *Expected Default Frequency* (EDF) obtained from the MKMV database.[49] EDF is a measure of the probability that a company will default over a specified period of time (typically one year). It is based on the structural approach to modeling default risk for a borrower described originally by Merton [135].[50] This approach assumes that there are three major drivers of a company's default probability: market value of assets, asset volatility, and default point. When the market value of assets falls to a level insufficient to repay the liabilities (default point), the company is considered to be in default. MKMV combines this framework with its own default database to derive an empirical probability of default for a company, the EDF. Thus, in this approach the *Distance-to-Default* (based on Merton's model) is mapped into an EDF credit measure that takes on values from 0-35%.

---

[49] This measure have been used in the study by Garlappi et al. (2008), while Vassalou and Xing (2004) used their own EDF-mimicking measure "DLI" for default likelihood.
[50] See Sections 1.3.2 and 2.2.1 for closer discussion of this model.

**Table 4.4**
Summary statistics of the EDF measure

| Month | # Company | Mean | Std. | Min | Max | Median | Quart 1 | Quart 3 |
|---|---|---|---|---|---|---|---|---|
| Dec-02 | 93 | 5.92 | 9.38 | 0.06 | 35.00 | 1.58 | 0.41 | 6.53 |
| Dec-03 | 93 | 1.92 | 4.51 | 0.03 | 35.00 | 0.45 | 0.15 | 1.60 |
| Dec-04 | 97 | 0.94 | 2.42 | 0.02 | 21.85 | 0.25 | 0.13 | 0.74 |
| Dec-05 | 99 | 1.02 | 4.26 | 0.01 | 35.00 | 0.15 | 0.08 | 0.45 |
| Dec-06 | 106 | 0.86 | 3.56 | 0.01 | 35.00 | 0.14 | 0.06 | 0.33 |
| Dec-07 | 112 | 1.02 | 4.24 | 0.01 | 35.00 | 0.10 | 0.06 | 0.32 |
| Dec-08 | 118 | 4.85 | 8.33 | 0.04 | 35.00 | 1.05 | 0.23 | 4.47 |
| Dec-09 | 111 | 4.17 | 8.42 | 0.05 | 35.00 | 0.73 | 0.29 | 3.28 |
| Dec-10 | 113 | 3.59 | 7.58 | 0.04 | 35.00 | 0.69 | 0.30 | 2.10 |
| Dec-11 | 113 | 4.85 | 8.88 | 0.04 | 35.00 | 0.63 | 0.19 | 4.13 |
| Dec-12 | 112 | 4.48 | 8.78 | 0.02 | 35.00 | 0.41 | 0.11 | 3.67 |
| Dec-13 | 121 | 2.24 | 6.15 | 0.01 | 35.00 | 0.28 | 0.07 | 1.31 |
| Dec-14 | 119 | 3.59 | 8.43 | 0.01 | 35.00 | 0.20 | 0.06 | 1.15 |
| Full sample | 16,703 | 2.91 | 6.90 | 0.01 | 35.00 | 0.37 | 0.11 | 1.68 |

Our sample spans a period from January 2002 to December 2014. The table reports the number of renewable energy companies in our sample, as well as the mean, standard deviation, minimum and maximum, median, and first and third quartile of the EDF distribution at the end of each year (2002-2014). EDF quantities are expressed in percent units. The number of companies in a "Full sample" denotes the total number of observations.

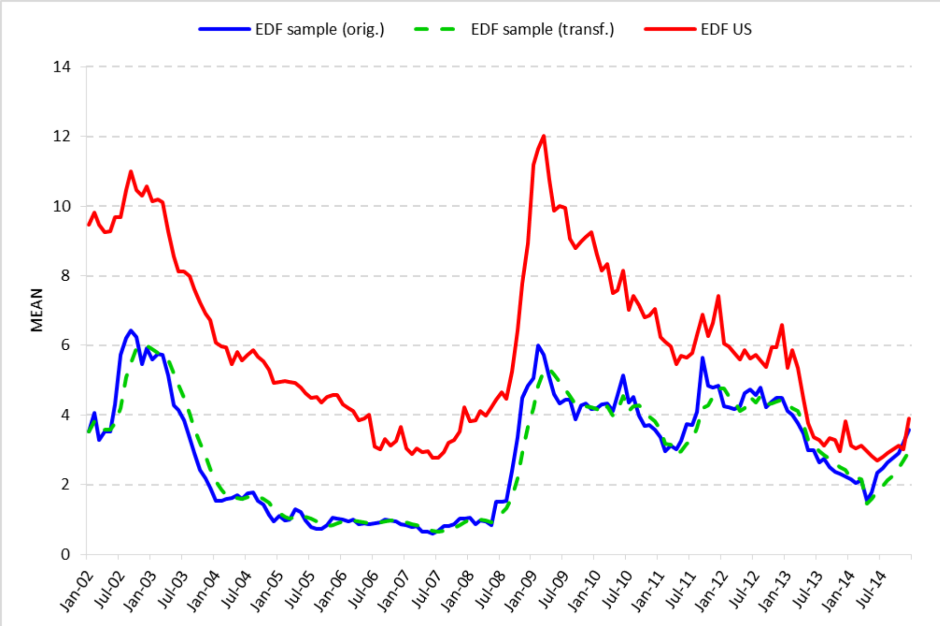
Summary statistics for the EDF measure are reported in Table 4.4. The average EDF measure in our sample is 2.91% with a median of 0.37%. The reported results show that there are substantial variations in the average as well as in the distribution of this measure over time. We can also see that the majority of companies in our sample during the sample period typically have an EDF score below 1.7%.

Because the EDF measure is based on market prices, we follow Garlappi, Shu and Yan [73] and use an exponentially smoothed version of this measure, based on a time-weighted average, in order to mitigate the effect of noisy stock prices on default scores. Specifically, for default probability in month $t$, we use

$$\overline{\mathrm{EDF}_t} = \frac{\sum_{s=0}^{5} e^{-sv}\,\mathrm{EDF}_{t-s}}{\sum_{s=0}^{5} e^{-sv}} \tag{4.4}$$

where $v$ is chosen to satisfy $e^{-5v} = 1/2$, such that the 5-month lagged EDF measure receives half the weight of the current EDF measure. Our empirical results are reported based on this transformed $\overline{\mathrm{EDF}_t}$ measure, which we will still refer to as EDF for notational convenience. A comparison of the monthly averages of the original and transformed EDF measure for our sample along with the average EDF for the entire U.S. market is provided in Figure 4.3. We

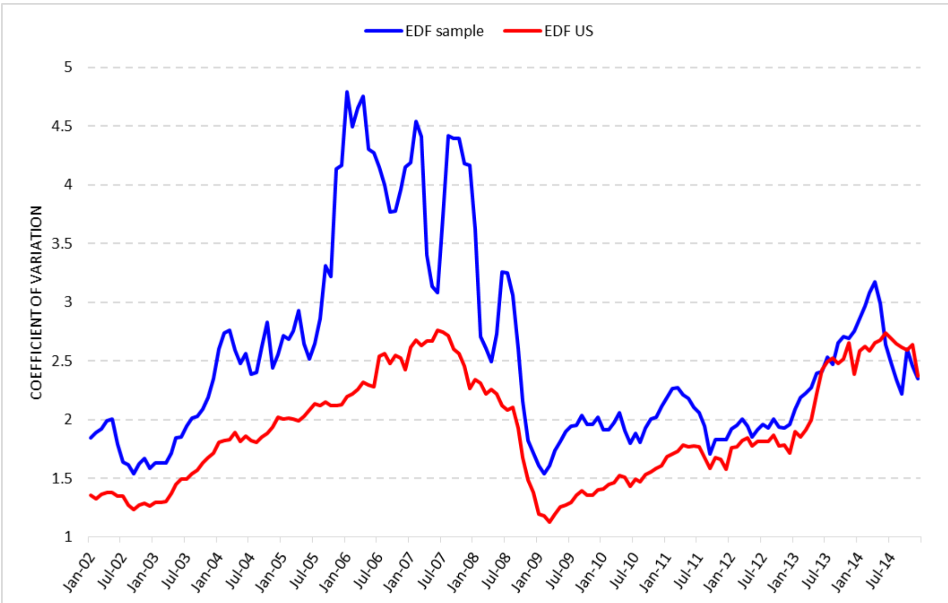
**Figure 4.3**
Mean of EDF measure




The figure plots the mean of original monthly EDF measure against the mean of transformed one based on Equation (4) for our sample along with the average EDF for U.S. market over the period from January 2002 to December 2014.

provide a comparison of the coefficients of variation (defined as a ratio of standard deviation over the mean) between EDFs of our sample and the entire U.S. market in Figure 4.4.

**Figure 4.4**
EDF – coefficient of variation




The figure plots the coefficient of variation (defined as a ratio of standard deviation and mean) between EDF of our sample and U.S. market over the period from January 2002 to December 2014.

Figure 4.3 illustrates that there are only marginal differences between the average monthly original and transformed EDF measure. Interestingly, we also observe that in terms of EDF, on average, renewable energy stocks are less risky than stocks in the entire U.S. market for the considered sample period. However, Figure 4.4 indicates that the coefficient of variation is typically higher for our sub-sample of renewable energy companies throughout the time period considered. This implies that per unit of default risk there is a higher variation in our sample. In other words, the discrepancy between the low-risk and the high-risk companies in our sample is larger than for the overall U.S. market.

## 4.4 Distress risk and equity returns

We start our analysis by investigating the relationship between distress risk (measured by EDF) and equity returns. Specifically, we examine whether portfolios with different default risk characteristics provide significantly different returns. The results are reported in Table 4.5. At the end of each month $t$ from January 2002 to November 2014, we form five portfolios of stocks according to each company's transformed EDF score. Therefore, based on the number of companies in our sample for a particular month (between 89 and 119), we form portfolios that contain between 18 and 24 companies. We then analyse the equally-weighted (EW) and value-weighted (VW) returns of these portfolios in month $t+1$. Portfolio 1 represents the portfolio of the 20% companies with the lowest distress risk, while portfolio 5 is the portfolio of the 20% companies with the highest distress risk. Furthermore, we compute returns for the portfolio that is formed by taking a long position in stocks with the highest EDF and a short position in stocks with the lowest EDF. For each of these portfolios we also report the average EDF score, the average size (market capitalization, expressed in '000,000 of $), and the average book-to-market (BM) ratio.

As illustrated in Table 4.5, we find a positive relationship between returns of both, equally-weighted (EW) and value-weighted (VW) portfolios, and distress risk – the higher the EDF, the higher the corresponding return. This positive relationship is consistent with findings of Vassalou and Xing [172], who use their own "DLI" measure based on the Merton [135] model as a measure of distress risk. The return difference between equally-weighted (EW) high default risk portfolios and low default risk portfolios is 1.19% per month (14.28% p.a.). The difference in returns for value-weighted (VW) portfolios is 1.73% per month (20.76% p.a.)

**Table 4.5**
Raw returns on portfolios sorted on the basis of the EDF measure

| Portfolios | Low<br>1 | 2 | EDF<br>3 | 4 | High<br>5 | High-Low | *t*-stat |
|---|---|---|---|---|---|---|---|
| Raw returns (EW) | 0.92 | 1.15 | 1.00 | 0.94 | 2.11 | 1.19 | (1.3161) |
| Raw returns (VW) | 1.17 | 1.56 | 1.46 | 1.93 | 2.90 | 1.73 | (1.9682)* |
| Average EDF | 0.08 | 0.22 | 0.54 | 1.56 | 9.13 | | |
| Average size | 7.92 | 3.27 | 1.94 | 1.42 | 0.59 | | |
| Average BM | 0.41 | 0.50 | 0.55 | 0.61 | 0.78 | | |

At the end of each month $t$ from January 2002 to November 2014, we sort our companies into quintiles based on their weighted EDF measures, as defined in (4.4). We then compute equally (EW) and value-weighted (VW) returns of these portfolios over the next month (month $t + 1$). In this table, we report the time-series averages of returns of these portfolios. Returns are expressed in percent units. Portfolio 1 is the portfolio with the lowest default risk and portfolio 5 with the highest one. The "High-Low" column is the difference between a quantity of the high EDF quantile and that of the low EDF quantile. The $t$-stat are the $t$-statistics of these differences and are calculated from Newey-West standard errors. The value of the truncation parameter $q$ was selected in each case to be equal to the number of autocorrelations in returns that are significant at the 5 percent level. *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level. "Average EDF", "Average size" and "Average BM" denote the average EDF, size and book-to-market ratio for particular portfolios, respectively. Size (market capitalization) is expressed in billions of $.

and is statistically significant at the 10% level. Thus, similar to Vassalou and Xing [172] we argue that the observed pattern is indicative of positively priced default risk.

Also note that small-capitalization stocks have on average higher EDF scores, and as a result, they provide higher returns than big-capitalization stocks. In addition, the average size of a portfolio and its BM ratio vary monotonically with the average EDF score of the portfolio. In particular, the average size increases as default risk of the portfolio decreases, whereas the opposite is true for the BM ratio. These results imply that the size and BM effect may be linked to default risk of stocks. Therefore, we follow Vassalou and Xing [172] and further investigate this possible link between the size and BM effects and default risk. We will focus on EW portfolios, since this is the weighting scheme typically employed in studies that deal with the size and BM effects, see, e.g., Fama and French [66] and Vassalou and Xing [172].

### 4.4.1 Size, BM, and distress risk

To further examine the extent to which the size and BM effects are related to default effects, in the following we perform two-way sorts and then examine each of the two effects for different default risk portfolios.

**Table 4.6**
Size effect controlled by default risk

| Size | Small | Medium | Big | Small-Big | *t*-stat |
|---|---|---|---|---|---|
| | | Panel A: Average Returns | | | |
| Low EDF | 1.12 | 0.99 | 0.95 | 0.17 | (0.3946) |
| Medium EDF | 1.50 | 0.95 | 0.67 | 0.83 | (1.7725)* |
| High EDF | 3.18 | 0.89 | 0.78 | 2.40 | (1.9671)* |
| | | Panel B: Average Size | | | |
| Low EDF | 0.629 | 2.364 | 15.714 | | |
| Medium EDF | 0.187 | 0.722 | 5.012 | | |
| High EDF | 0.055 | 0.221 | 2.376 | | |
| | | Panel C: Average EDF | | | |
| Low EDF | 0.143 | 0.128 | 0.097 | | |
| Medium EDF | 0.635 | 0.579 | 0.521 | | |
| High EDF | 7.974 | 5.449 | 5.302 | | |
| | | Panel D: Average BM | | | |
| Low EDF | 0.470 | 0.468 | 0.385 | | |
| Medium EDF | 0.562 | 0.549 | 0.547 | | |
| High EDF | 0.763 | 0.647 | 0.738 | | |

At the end of each month from February 2002 to December 2014, stocks are first sorted into three portfolios based on their weighted EDF measures (low, medium, high) in the previous month. Within each portfolio, we subsequently sort stocks into three size portfolios (small, medium, big), based on their market capitalization in the previous month. The equally-weighted average returns of the portfolios in Panel A are reported in percent units. “Small-Big” is the return difference between the smallest and biggest size portfolios within each default group. *t*-stat are the corresponding *t*-statistics of these differences and are calculated from Newey-West standard errors. The value of the truncation parameter *q* was selected in each case to be equal to the number of autocorrelations in returns that are significant at the 5 percent level. *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level. Average size (market capitalization) in Panel B is expressed in billions of $, while average EDF in Panel C is expressed in percent units, and average BM (book-to-market) in Panel D in decimal units.

Table 4.6 provides results from sequential sorts. Stocks are first sorted into three groups according to their default risk (low, medium, high). Subsequently, the stocks within each EDF group are sorted into three size portfolios (small, medium, big). Using these nine created portfolios, we investigate whether there is a size effect in all default risk groups.

Reported results in Panel A suggest that the size effect is present in particular for portfolios that contain more distressed stocks. This effect is more pronounced and statistically significant for the high EDF portfolio where the average return difference between small and big firms is 2.40% per month (28.80% p.a.). This is about fourteen times more than for a portfolio containing low-distress stocks (0.17% per month). We investigate to what extent we are truly capturing the size effect in Panel B. We can see that there really is substantial variation in the market capitalization of stocks within the high EDF portfolio. However, it is not necessary always a case that renewable firms with a high EDF are also small in size. In fact, the biggest firms in the created 'high distress risk' portfolios are rather medium-sized renewable energy companies. Their average size is $2.376 billion and, therefore, still bigger than the average small and medium sized firms in the 'low distress risk' and 'medium distress risk' categories (ranging from 0.187 to 2.364). This basically means that when we are sorting our stocks according to their EDF, it is clearly not a sorting by size only. On the other hand, high EDF/small size portfolio do typically contain the smallest of the small firms. These results show that the size effect is concentrated in the smallest firms, which also happen to be among those renewable companies with the highest distress risk.

In Panel C, we examine how much riskier stocks in high EDF portfolios are in comparison to other default risk groups. The results show that they are indeed much riskier. On average, small firms in high default risk portfolios are about thirteen times riskier in terms of the applied EDF measure than small firms in medium EDF portfolios, and about 56 times riskier than small firms in low distress risk portfolios. Thus, the large average returns earned by small high-default risk companies (see Panel A) compared to the rest of the portfolios can be explained by a possible compensation for the large distress risk they have. Moreover, we can see that in all default risk groups the average EDF monotonically decreases as size increases and that the difference between small and big firms is significantly higher for high default risk groups (2.672) compared to medium and low default risk groups (0.114 and 0.046). This also explains the large difference in returns between small and big stocks in the high EDF portfolio. Finally, the average BM ratios of the default- and size-sorted portfolios are reported in Panel D. The results show that the average BM ratios are the highest for the high EDF group.

Overall, the results in Table 4.6 imply that the size effect might be partially interpreted as a default effect, however, sorting the stocks according to their EDF is not the same thing as just sorting by size. The size effect is significant only in the segment of our sample with the highest distress risk, where the difference in returns between small and big firms can be explained by the difference in their default risk. For the remaining stocks in our sample, where no significant size effect has been detected, also the difference in default risk between small and big stocks is only minimal.

Table 4.7 presents results from sequential portfolio sorting, where stocks are first sorted into three groups according to their EDF (low, medium, high), and subsequently each of these

**Table 4.7**
BM effect controlled by default risk

| BM | High | Medium | Low | High-Low | *t*-stat |
|---|---|---|---|---|---|
| | | Panel A: Average Returns | | | |
| Low EDF | 1.07 | 1.08 | 0.91 | 0.16 | (0.4163) |
| Medium EDF | 1.41 | 1.34 | 0.45 | 0.96 | (2.1374)** |
| High EDF | 1.97 | 0.72 | 1.36 | 0.61 | (0.7281) |
| | | Panel B: Average BM | | | |
| Low EDF | 0.703 | 0.408 | 0.212 | | |
| Medium EDF | 0.931 | 0.495 | 0.232 | | |
| High EDF | 1.434 | 0.565 | 0.146 | | |
| | | Panel C: Average EDF | | | |
| Low EDF | 0.136 | 0.120 | 0.111 | | |
| Medium EDF | 0.615 | 0.559 | 0.561 | | |
| High EDF | 7.300 | 4.166 | 7.102 | | |
| | | Panel D: Average Size | | | |
| Low EDF | 5.208 | 5.929 | 7.493 | | |
| Medium EDF | 2.188 | 2.469 | 1.304 | | |
| High EDF | 0.778 | 1.222 | 0.674 | | |

At the end of each month from February 2002 to December 2014, stocks are first sorted into three portfolios based on their weighted EDF measures (low, medium, high) in the previous month. Within each portfolio, we subsequently sort stocks into three BM (book-to-market) portfolios (high, medium, low), based on their past month's BM ratio. The equally-weighted average returns of the portfolios in Panel A are reported in percent units. High-Low" is the return difference between the highest BM and lowest BM portfolios within each default group. *t*-stat are the corresponding *t*-statistics of these differences and are calculated from Newey-West standard errors. The value of the truncation parameter *q* was selected in each case to be equal to the number of autocorrelations in returns that are significant at the 5 percent level. *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level. Average BM in Panel B is expressed in decimal units, average EDF in Panel C is expressed in percent units, and average size (market capitalization) in Panel D in billions of $.

groups is sorted into three BM portfolios (high, medium, low). In the following we will examine the BM effect within each of the EDF groups.

Panel A shows that the BM effect is only statistically significant for the constructed medium EDF portfolios, with a return differential of 0.96% per month (11.52% p.a.). This is about one and a half times more than the difference for a portfolio containing high-distress stocks (0.61% per month) and six times more than for low EDF group (0.16%). However, note that the differences between average EDF for value stocks (high BM) and growth stocks (low BM) for all three default portfolios in Panel C are rather marginal. These results rather suggest that for our sample of U.S. renewable energy companies, unlike the size effect, the BM effect is not a default risk effect.

The differences in average BM ratios within particular EDF portfolios reported in Panel B are relatively low. For instance, the difference between the value and growth firms within medium EDF portfolio is only 0.7, which suggests that the return differential of these portfolios observed in Panel A are not truly caused by the BM effect. We can also see that the average BM ratios are higher in portfolios with highly distressed stocks, however, this is not true for the low BM group where it is the medium EDF portfolio that has the highest average BM value. Furthermore, the average EDF in Panel C exhibits a monotonic relation with BM only in the low EDF category, that is, the portfolio with the lowest default risk. For the other two groups, i.e. medium and high distress risk, the relation is not monotonic. This is in contrast with our results from Panel C in Table 4.6, where we clearly find a monotonic relationship between default risk and size of portfolios.

**Table 4.8**
Default effect controlled by size

| EDF | Low | Medium | High | High-Low | *t*-stat |
|---|---|---|---|---|---|
| | | Panel A: Average Returns | | | |
| Small | 1.68 | 0.95 | 2.99 | 1.31 | (0.9836) |
| Medium | 0.63 | 0.68 | 1.05 | 0.42 | (0.6623) |
| Big | 0.98 | 1.11 | 1.04 | 0.06 | (0.1089) |
| | | Panel B: Average Size | | | |
| Small | 0.168 | 0.129 | 0.091 | | |
| Medium | 0.823 | 0.733 | 0.640 | | |
| Big | 11.532 | 7.361 | 5.245 | | |
| | | Panel C: Average EDF | | | |
| Small | 0.475 | 2.146 | 10.636 | | |
| Medium | 0.130 | 0.460 | 3.927 | | |
| Big | 0.073 | 0.224 | 3.287 | | |
| | | Panel D: Average BM | | | |
| Small | 0.572 | 0.628 | 0.790 | | |
| Medium | 0.460 | 0.514 | 0.671 | | |
| Big | 0.394 | 0.492 | 0.633 | | |

At the end of each month from February 2002 to December 2014, stocks are first sorted into three portfolios based on their past month's market capitalization (small, medium, big). Within each portfolio, we subsequently sort stocks into three EDF portfolios (low, medium, high), based on their weighted EDF measures in the previous month. The equally-weighted average returns of the portfolios in Panel A are reported in percent units. "High-Low" is the return difference between the highest and lowest default risk portfolios within each size group. *t*-stat are the corresponding *t*-statistics of these differences and are calculated from Newey-West standard errors. The value of the truncation parameter *q* was selected in each case to be equal to the number of autocorrelations in returns that are significant at the 5 percent level. *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level. Average size (market capitalization) in Panel B is expressed in billions of $, while average EDF in Panel C is expressed in percent units, and average BM in Panel D in decimal units.

Finally, Panel D shows that a portfolio with high distressed stocks contains mainly small firms. This time even the highest value of 1.22 is lower than any of the values from medium and high EDF groups. And again, contrary to our findings from Table 4.6, size varies monotonically only within the low EDF portfolio.

Our findings from EDF-BM sorting imply that, unlike the size effect that can be to some extent interpreted as a default effect, the BM effect is not truly related to default risk. Moreover, the monthly return premium of small firms over big firms for the high EDF portfolio is 2.4%, and therefore about 1.8% larger than that of value stocks over growth stocks (0.6%).

### 4.4.2 The default effect

Tables 4.6 and 4.7 illustrate that while the size effect is somewhat related to default risk, we did not find much evidence that the same is true for the BM effect. In what follows, we investigate whether default risk is rewarded differently depending on the size and BM characteristics of a company. We follow Vassalou and Xing [172] and define the default effect as a positive average return differential between high and low default risk firms.

In Table 4.8, we reverse the sorting procedure applied in Table 4.6 and examine whether there is a default effect in size-sorted portfolios. Thus, we first sort stocks into three groups according to their size (small, medium, big), and subsequently within each of this size group we sort stocks into three distress risk portfolios (low, medium, high).

Reported results in Panel A show that there is no statistically significant default effect in any of the size-sorted portfolios, although the differences in returns are positive. The highest difference between average monthly returns for high risk and low risk companies is 1.31% per month (15.72% p.a.) for portfolios containing small firms. Thus, this implies that in particular in the small size segment distressed firms earn on average higher returns than low distress risk firms. Panel C also emphasizes the substantially higher default risk for the high EDF categories, independent of the market capitalization of the stocks. Note that within the small size portfolio, the average EDF varies between 10.64% (for the high distress risk category) and 0.48% (for the low distress risk category), which suggests that small firms likely significantly differ with respect to their default risk characteristics. The same is also true with respect to their returns, as illustrated by Panel A. Note, however, that we do not find a monotonic relationship: for the small size category, for example, medium distress risk firms on average yield a higher return than low distress risk firms. However, the highest returns are provided by small firms with the highest distress risk, confirming earlier results.

We can also see in Panel C that the average EDF monotonically decreases as firm's size increases. This confirms the close relation between size and default risk observed in Table 4.6. Finally, Panels B and D show that small size/high EDF portfolios contain the smallest stocks with the highest BM ratios, while big size/low EDF portfolios contain the largest companies with the lowest BM ratios.

In the last sequential sort, we investigate the presence of a default effect in BM-sorted portfolios. The results are reported in Table 4.9. Stocks are first sorted into three groups

**Table 4.9**

Default effect controlled by BM

| EDF | Low | Medium | High | High-Low | *t*-stat |
|---|---|---|---|---|---|
| | | Panel A: Average Returns | | | |
| High BM | 1.32 | 1.60 | 1.88 | 0.56 | (0.7499) |
| Medium BM | 1.07 | 1.45 | 0.33 | -0.74 | (-1.3477) |
| Low BM | 0.73 | 0.65 | 1.38 | 0.65 | (0.8528) |
| | | Panel B: Average Size | | | |
| High BM | 4.212 | 1.929 | 0.843 | | |
| Medium BM | 6.346 | 3.082 | 1.240 | | |
| Low BM | 7.436 | 1.502 | 0.649 | | |
| | | Panel C: Average EDF | | | |
| High BM | 0.259 | 1.244 | 8.465 | | |
| Medium BM | 0.112 | 0.397 | 2.716 | | |
| Low BM | 0.096 | 0.505 | 6.925 | | |
| | | Panel D: Average BM | | | |
| High BM | 0.834 | 0.986 | 1.352 | | |
| Medium BM | 0.466 | 0.478 | 0.483 | | |
| Low BM | 0.222 | 0.213 | 0.135 | | |

At the end of each month from February 2002 to December 2014, stocks are first sorted into three portfolios based on their past month's BM (book-to-market) ratio (high, medium, low). Within each portfolio, we subsequently sort stocks into three EDF portfolios (low, medium, high), based on their weighted EDF measures in the previous month. The equally-weighted average returns of the portfolios in Panel A are reported in percent units. "High-Low" is the return difference between the highest and lowest default risk portfolios within each BM group. *t*-stat are the corresponding *t*-statistics of these differences and are calculated from Newey-West standard errors. The value of the truncation parameter *q* was selected in each case to be equal to the number of autocorrelations in returns that are significant at the 5 percent level. *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level. Average size (market capitalization) in Panel B is expressed in billions of $, while average EDF in Panel C is expressed in percent units, and average BM in Panel D in decimal units.

according to their BM ratio (high, medium, low), and subsequently within each of this BM groups we sort stock into three default portfolios (low, medium, high).

Panel A shows that no statistically significant default effect is present in any of the BM-sorted portfolios, which is consistent with the findings from Table 4.8. Moreover, the difference between high and low EDF portfolios for medium BM groups actually yields negative return. The highest average monthly return is only 0.65% per month (7.8% p.a.) for portfolios containing low BM firms. Next, we can see that the return difference between high and low BM portfolios is relatively small.

Once again, Panel C shows that value stocks can differ a lot with respect to their default risk characteristics. However, the same thing can be said about growth stocks too. The smallest firms are typically firms with the lowest BM ratios and are contained in high EDF/low BM portfolios.

## 4.5 Pricing of distress risk

In this section, using the asset pricing models described in Section 4.2, we investigate whether distress risk is systematic, and, therefore, whether it is priced in the cross-section of equity returns. In other words, we measure the premium that investors receive for holding distressed stocks.

Following the same approach as in Table 4.5, at the end of each month from January 2002 to November 2014 we sort the companies in our sample into quintiles based on their EDF measures and form five equally-weighted (EW) and value-weighted (VW) portfolios. For each month, portfolio 1 contains the 20% companies with the lowest distress risk, while portfolio 5 contains the 20% companies with the highest distress risk. We also construct the long-short portfolio that takes a long position in the 20% of stocks with the highest distress risk (these stocks will provide higher returns), and a short position in the 20% of stocks with the lowest distress risk (stocks providing lower returns). A key questions in our analysis is also whether returns of the created portfolios can be explained by the factors included into standard asset pricing models. Further, we want to examine whether portfolios of distressed companies in the renewable sector as well as the created long-short strategy based on distress risk yields abnormal or active returns beyond what would be suggested by standard asset pricing models.

In Table 4.10 and Table 4.11 we report the results from regressions using the excess returns of equally-weighted (EW) and value-weighted (VW) portfolios, respectively. Panel A in these tables reports monthly alphas expressed in annualized percent units with respect to the CAPM (4.1), the Fama-French three-factor model (4.2), and the Carhart four-factor model (4.3) with corresponding *t*-statistics below in parenthesis. We estimate these models using the standard factor-mimicking portfolios available on Professor Kenneth French's website (see Footnote 46). Panels B, C, and D then report estimated factor loadings for excess returns on the CAPM market factor, on the Fama-French market, size, and value factors, and on the four Carhart factors (including momentum), respectively, again with corresponding *t*-statistics. Finally, there are reported R-squared measures from these regressions in Panel E. Figure 4.5 then graphically summarizes the behavior of alphas across particular portfolios, while Figure 4.6 shows the evolution of factor loadings from the four-factor model across the created distress risk portfolios. We also provide correlation coefficients between raw returns and the applied factors in Table 4.12.

The risk-adjusted returns (alphas) corrected for given risk factors are reported in Panel A of Table 4.10 and Table 4.11, respectively. They are generally increasing across our portfolios, although this pattern is not monotonic. In case of EW portfolios, alphas are decreasing for portfolios 3 and 4, but for all three models they significantly increase for portfolio 5 containing the 20% of renewable stocks with the highest EDF (see also panel A in Figure 4.5). In fact, also the "High-Low" strategy where we hold the riskiest quintile of stocks and sell the quintile of stocks with the lowest failure risk provides positive returns from 4.06% to 6.68% p.a., depending on the applied asset pricing model. In case of VW portfolios, alphas are decreasing only for portfolio 3. However, in comparison to EW portfolios, all alphas are positive (see also panel B in Figure 4.5) and their values for the long-short strategy are about twice as high, ranging from 10.79% p.a. to 13.41% p.a. Note that generally results on alpha are also relatively stable with respect to the applied pricing models. This is true in particular for the VW portfolios, where the calculated annualized active returns are hardly affected by the choice of model.

**Table 4.10**
Risk-adjusted returns on EW portfolios sorted on the basis of the EDF measure

| | Low | | EDF | | High | |
|---|---|---|---|---|---|---|
| EW portfolios | 1 | 2 | 3 | 4 | 5 | High-Low |
| | | | Panel A: Portfolio Alphas (EW) | | | |
| CAPM alpha | 1.91 | 2.97 | 0.39 | -1.67 | 8.81 | 5.50 |
| | (0.66) | (0.85) | (0.09) | (0.30) | (0.92) | (0.60) |
| 3-factor alpha | 0.96 | 1.89 | -1.22 | -3.08 | 6.42 | 4.06 |
| | (0.37) | (0.64) | (0.32) | (0.60) | (0.70) | (0.45) |
| 4-factor alpha | 0.48 | 1.88 | -1.44 | -1.94 | 8.54 | 6.68 |
| | (0.18) | (0.62) | (0.38) | (0.38) | (0.89) | (0.71) |
| | | | Panel B: CAPM Regression Coefficients (EW) | | | |
| RM | 1.091 | 1.341 | 1.441 | 1.627 | 2.133 | 1.045 |
| | (20.73)*** | (18.71)*** | (18.21)*** | (13.72)*** | (7.99)*** | (4.02)*** |
| | | | Panel C: Three-factor Regression Coefficients (EW) | | | |
| RM | 0.987 | 1.212 | 1.263 | 1.463 | 1.872 | 0.888 |
| | (16.73)*** | (17.41)*** | (16.09)*** | (11.41)*** | (6.99)*** | (3.37)*** |
| SMB | 0.539 | 0.788 | 0.911 | 0.950 | 1.320 | 0.784 |
| | (5.42)*** | (6.72)*** | (6.52)*** | (4.53)*** | (4.10)*** | (2.42)** |
| HML | -0.053 | -0.300 | -0.079 | -0.283 | -0.079 | -0.031 |
| | (0.44) | (2.49)** | (0.56) | (1.34) | (0.18) | (0.07) |
| | | | Panel D: Four-factor Regression Coefficients (EW) | | | |
| RM | 1.045 | 1.213 | 1.289 | 1.327 | 1.619 | 0.575 |
| | (15.98)*** | (14.29)*** | (14.29)*** | (11.14)*** | (7.07)*** | (2.60)*** |
| SMB | 0.514 | 0.787 | 0.900 | 1.009 | 1.431 | 0.921 |
| | (5.06)*** | (6.34)*** | (6.29)*** | (4.93)*** | (4.10)*** | (2.58)** |
| HML | -0.029 | -0.299 | -0.068 | -0.340 | -0.186 | -0.163 |
| | (0.24) | (2.60)** | (0.48) | (1.74) | (0.45) | (0.42) |
| MOM | 0.116 | 0.003 | 0.052 | -0.272 | -0.507 | -0.625 |
| | (1.94) | (0.03) | (0.61) | (2.46)** | (1.43) | (1.63) |
| | | | Panel E: R-squared (EW) | | | |
| CAPM | 0.7215 | 0.7200 | 0.6692 | 0.6049 | 0.4319 | 0.1635 |
| 3-factor | 0.7679 | 0.7913 | 0.7398 | 0.6616 | 0.4755 | 0.1879 |
| 4-factor | 0.7763 | 0.7913 | 0.7407 | 0.6789 | 0.5004 | 0.2476 |

At the end of each month *t* from January 2002 to November 2014, we sort our companies into quintiles based on their weighted EDF measures, as defined in (4.4). Portfolio 1 is the portfolio with the lowest default risk and portfolio 5 with the highest one. The "High-Low" column denotes a portfolio that takes a long position in the 20% stocks with high EDF and a short position in the 20% stocks with low EDF. In this table, we show results from regressions of equally-weighted (EW) excess returns of month $t + 1$ (period from February 2002 to December 2014) on a constant (alpha), market returns (RM), as well as three factor Fama-French (RM, SMB, HML) and four factor Carhart (RM, SMB, HML, MOM) regressions. Panel A shows monthly alphas or active returns (in annualized percent units) from these regressions and the corresponding absolute values of *t*-statistics (in parenthesis). Panel B shows loadings on the market factor and the corresponding absolute values of *t*-statistics (in parentheses) from the CAPM model. Panels C and D show loadings on three factors and four factors, respectively, and the corresponding absolute values of the *t*-statistics (in parentheses) from the applied three-factor and four-factor regressions. R-squared are then reported in Panel E. *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

**Table 4.11**
Risk-adjusted returns on VW portfolios sorted on the basis of the EDF measure

| VW portfolios | Low<br>1 | 2 | EDF<br>3 | 4 | High<br>5 | High-Low |
|---|---|---|---|---|---|---|
| Panel A: Portfolio Alphas (VW) | | | | | | |
| CAPM alpha | 5.53 | 8.62 | 6.76 | 10.13 | 18.91 | 11.98 |
| | (1.75)* | (2.26)** | (1.30) | (1.56) | (1.86)* | (1.25) |
| 3-factor alpha | 5.60 | 8.46 | 5.84 | 9.27 | 17.79 | 10.79 |
| | (1.89)* | (2.31)** | (0.62) | (1.47) | (1.74)* | (1.12) |
| 4-factor alpha | 5.37 | 8.12 | 5.66 | 10.32 | 20.17 | 13.41 |
| | (1.79)* | (2.21)** | (1.08) | (1.63) | (1.98)** | (1.41) |
| Panel B: CAPM Regression Coefficients (VW) | | | | | | |
| RM | 1.012 | 1.227 | 1.320 | 1.649 | 2.046 | 1.037 |
| | (14.30)*** | (13.82)*** | (12.26)*** | (11.26)*** | (8.78)*** | (5.04)*** |
| Panel C: Three-factor Regression Coefficients (VW) | | | | | | |
| RM | 1.004 | 1.200 | 1.224 | 1.553 | 1.926 | 0.925 |
| | (14.63)*** | (11.42)*** | (11.53)*** | (9.18)*** | (7.99)*** | (4.18)*** |
| SMB | 0.247 | 0.250 | 0.420 | 0.517 | 0.569 | 0.325 |
| | (2.08)** | (1.78)* | (2.05)** | (1.96)* | (1.19) | (0.72) |
| HML | -0.396 | -0.227 | 0.093 | -0.086 | 0.033 | 0.424 |
| | (3.11)*** | (1.68)* | (0.44) | (0.38) | (0.08) | (1.11) |
| Panel D: Four-factor Regression Coefficients (VW) | | | | | | |
| RM | 1.032 | 1.241 | 1.246 | 1.427 | 1.643 | 0.612 |
| | (13.94)*** | (11.82)*** | (10.76)*** | (9.29)*** | (7.25)*** | (2.93)*** |
| SMB | 0.234 | 0.232 | 0.411 | 0.572 | 0.693 | 0.462 |
| | (1.95)* | (1.64) | (1.86)* | (2.20)** | (1.47) | (1.04) |
| HML | -0.384 | -0.209 | 0.102 | -0.138 | -0.087 | 0.292 |
| | (2.90)*** | (1.50) | (0.49) | (0.62) | (0.22) | (0.83) |
| MOM | 0.057 | 0.081 | 0.043 | -0.251 | -0.568 | -0.627 |
| | (0.85) | (1.01) | (0.25) | (1.69)* | (2.21)** | (2.21)** |
| Panel E: R-squared (VW) | | | | | | |
| CAPM | 0.6578 | 0.6777 | 0.5423 | 0.5419 | 0.4207 | 0.1712 |
| 3-factor | 0.6930 | 0.6903 | 0.5583 | 0.5560 | 0.4295 | 0.1848 |
| 4-factor | 0.6951 | 0.6933 | 0.5589 | 0.5688 | 0.4625 | 0.2487 |

At the end of each month $t$ from January 2002 to November 2014, we sort our companies into quintiles based on their weighted EDF measures, as defined in (4.4). Portfolio 1 is the portfolio with the lowest default risk and portfolio 5 with the highest one. The "High-Low" column denotes a portfolio that takes a long position in the 20% stocks with high EDF and a short position in the 20% stocks with low EDF. In this table, we show results from regressions of value-weighted (VW) excess returns of month $t + 1$ (period from February 2002 to December 2014) on a constant (alpha), market return (RM), as well as three factor Fama-French (RM, SMB, HML) and four factor Carhart (RM, SMB, HML, MOM) regressions. Panel A shows monthly alphas or active returns (in annualized percent units) from these regressions and the corresponding absolute values of $t$-statistics (in parenthesis). Panel B shows loadings on the market factor and the corresponding absolute values of $t$-statistics (in parentheses) from the CAPM model. Panels C and D show loadings on the three factors and four factors, respectively, and the corresponding absolute values of the $t$-statistics (in parentheses) from three-factor and four-factor regressions. R-squared are then reported in Panel E. *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

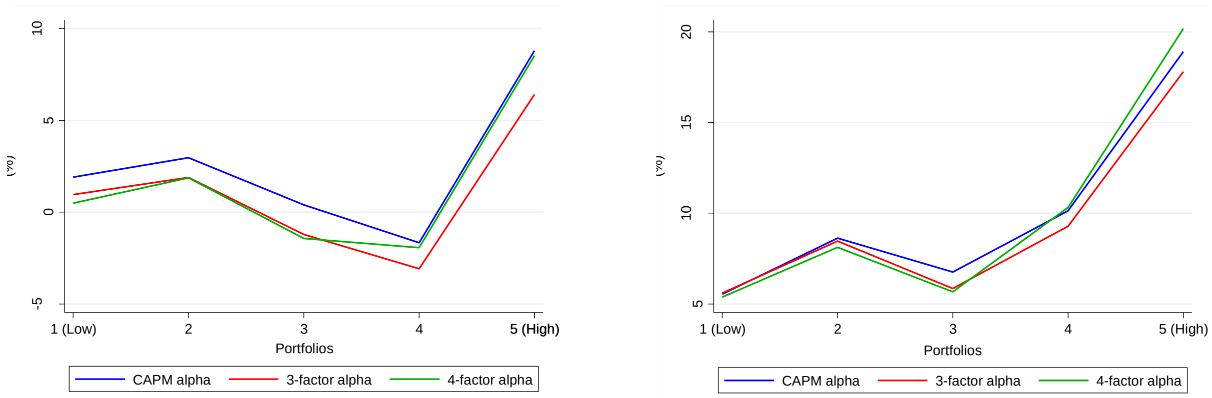
Overall, our results indicate again that the distress risk premium seems to be an effect that is mainly concentrated in the companies with substantial default risk, i.e. the highest quintile. Only for these portfolios we get active annualized returns of substantial magnitude, i.e. greater than 5% for EW portfolios and even greater than 15% for VW portfolios. Note, however that due to the significant standard deviation in returns for the portfolios containing high distress risk stocks, only for VW portfolios the active returns are also statistically significant.

Regarding the factor loadings that are reported in panels B, C, and D in Tables 4.10 and 4.11, the market factor RM is increasing and statistically significant for all models. We can see that stocks in portfolio 5 (stocks with high probability of default) have beta-factors about twice the size of those in portfolio 1 (stocks with low probability of default). The size factor SMB is also almost monotonically increasing in both three-factor and four-factor models implying that the small companies prevail among distressed stocks. Finally, the value factor HML is rather humped-shaped for VW portfolios and irregular for EW portfolios, while the momentum factor MOM has a decreasing pattern, with positive loadings on the first three portfolios with lower risk and negative loadings for the remaining two portfolios with higher distress risk (see also Figure 4.6).

**Figure 4.5**
Portfolio alphas from the regressions of excess returns

A) Equally-weighted (EW) portfolios

B) Value-weighted (VW) portfolios

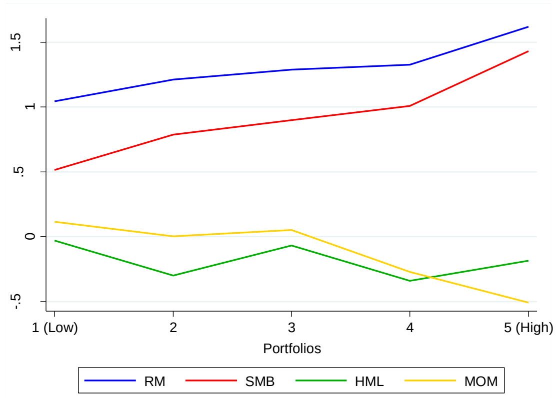



The figure plots monthly alphas (in annualized percent units) from the applied CAPM model, the Fama-French three-factor model, and the Carhart four-factor model for 5 distress risk-sorted equally-weighted (EW) portfolios (Panel A) and value-weighted (VW) portfolios (Panel B) from February 2002 to December 2014 (see panels A of Tables 4.10 and 4.11). Portfolios are formed at the end of each month from January 2002 to November 2014, when we sort our companies into quintiles based on their weighted EDF measures, as defined in (4.4).

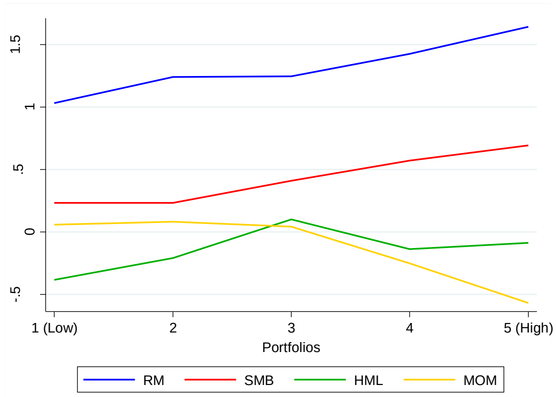
**Figure 4.6**

Factor loadings from the four-factor regression of excess returns

A) Equally-weighted (EW) portfolios




B) Value-weighted (VW) portfolios




The figure plots loadings on excess market return (RM), size factor (SMB), value factor (HML), and momentum factor (MOM) from four-factor regression (see panels D of Tables 4.10 and 4.11) for 5 distress risk-sorted equally-weighted (EW) portfolios (Panel A) and value-weighted (VW) portfolios (Panel B) from February 2002 to December 2014. Portfolios are formed at the end of each month from January 2002 to November 2014, when we sort our companies into quintiles based on their weighted EDF measures, as defined in (4.4).

Thus, contrary to the findings of Campbell, Hilscher and Szilagyi [34], we consistently find that stocks with high risk of failure also have high average returns, both raw and risk-adjusted, implying that distress risk is positively priced in the U.S. stock market for renewable energy companies. However, as pointed out previously, our results also indicate that distress risk seems to be predominantly priced in the highest quintile, i.e. for companies with a relatively high probability of default. In particular for the created VW portfolios, we get high and statistically significant active annualized returns with magnitudes between 17.8% and 20.2%.

Reported R-squared measures from these regressions are relatively high, particularly for low-risk portfolios, and generally decreasing with portfolios that hold more distressed stocks. The values start at 0.72 – 0.78 for EW portfolio 1 and end at 0.43 – 0.50 for EW portfolio 5, while we have 0.66 – 0.70 for VW portfolio 1 and 0.42 – 0.46 for VW portfolio 5. In general, R-squared measures for EW portfolios are slightly higher than for VW ones. Note that for our "High-Low" strategy where we take a long position in the 20% of high-distress stocks and a short position in the 20% of safest stocks the values drop significantly to 0.16 – 0.25 for EW portfolio and 0.17 – 0.25 for VW portfolio, respectively. This indicates that returns created through setting up a long-short strategy based on distress risk in the renewable energy sector cannot be explained by standard factors in asset pricing models. We interpret this as an additional confirmation for a distress risk factor that is systematically priced in the renewable energy sector.

From reported results in Table 4.12 we can see that the raw returns of our five portfolios are relatively highly correlated and exhibit correlations ranging from 0.65 to 0.86 for EW portfolios and from 0.66 to 0.81 for VW portfolios. Regarding the correlation between raw returns of our five portfolios and the pricing factors, we find that the market factor RM is relatively highly correlated with returns of the created portfolios. Note, however, that both for EW and VW portfolios, the correlation is generally decreasing for portfolios that hold more distressed stocks. For EW portfolio returns the correlation coefficients have values between 0.85 (portfolio 1) and 0.66 (portfolio 5), for VW portfolio returns correlations range from 0.81 (portfolio 1) to 0.65 (portfolio 5). The size factor SMB, the value factor HML, and the momentum factor MOM show much lower degrees of correlation with raw returns of our portfolios. However, the correlation between returns from the generated distress risk portfolios and the Fama-French SMB factor is typically still above 0.5 (for EW portfolios) and

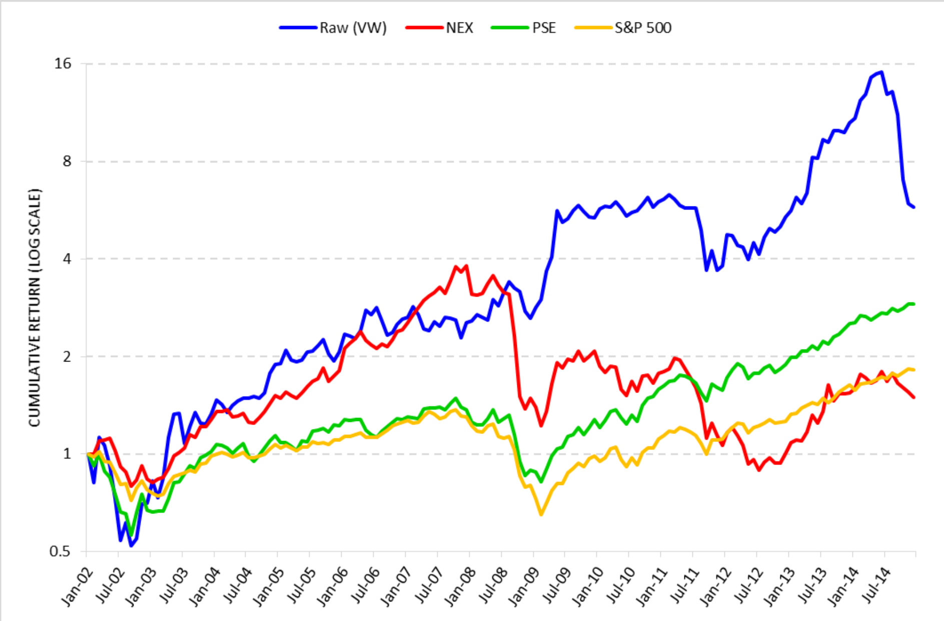
**Table 4.12**
Correlation coefficients between raw returns and given factors

| | 1 (Low) | 2 | 3 | 4 | 5 (High) | High-Low | RM | SMB | HML | MOM |
|---|---|---|---|---|---|---|---|---|---|---|
| Panel A: Equally-weighted returns (EW) | | | | | | | | | | |
| 1 (Low) | 1 | | | | | | | | | |
| 2 | 0.85 | 1 | | | | | | | | |
| 3 | 0.85 | 0.86 | 1 | | | | | | | |
| 4 | 0.78 | 0.82 | 0.82 | 1 | | | | | | |
| 5 (High) | 0.66 | 0.70 | 0.65 | 0.73 | 1 | | | | | |
| High-Low | 0.33 | 0.46 | 0.40 | 0.53 | 0.93 | 1 | | | | |
| RM | 0.85 | 0.85 | 0.82 | 0.78 | 0.66 | 0.40 | 1 | | | |
| SMB | 0.51 | 0.54 | 0.55 | 0.50 | 0.44 | 0.29 | 0.37 | 1 | | |
| HML | 0.17 | 0.09 | 0.16 | 0.11 | 0.14 | 0.09 | 0.20 | 0.16 | 1 | |
| MOM | -0.25 | -0.32 | -0.29 | -0.42 | -0.40 | -0.37 | -0.42 | -0.07 | -0.17 | 1 |
| Panel B: Value-weighted returns (VW) | | | | | | | | | | |
| | 1 (Low) | 2 | 3 | 4 | 5 (High) | High-Low | RM | SMB | HML | MOM |
| 1 (Low) | 1 | | | | | | | | | |
| 2 | 0.81 | 1 | | | | | | | | |
| 3 | 0.72 | 0.69 | 1 | | | | | | | |
| 4 | 0.74 | 0.79 | 0.66 | 1 | | | | | | |
| 5 (High) | 0.66 | 0.66 | 0.68 | 0.75 | 1 | | | | | |
| High-Low | 0.34 | 0.43 | 0.49 | 0.57 | 0.93 | 1 | | | | |
| RM | 0.81 | 0.82 | 0.74 | 0.73 | 0.65 | 0.41 | 1 | | | |
| SMB | 0.38 | 0.37 | 0.38 | 0.38 | 0.33 | 0.22 | 0.37 | 1 | | |
| HML | 0.01 | 0.09 | 0.18 | 0.14 | 0.14 | 0.18 | 0.20 | 0.16 | 1 | |
| MOM | -0.28 | -0.28 | -0.28 | -0.40 | -0.43 | -0.40 | -0.42 | -0.07 | -0.17 | 1 |

The table reports correlation coefficients between raw returns of equally-weighted (EW) portfolios in Panel A and value-weighted portfolios (VW) in Panel B, and the CAPM, Fama-French, and Carhart factors.

**Figure 4.7**

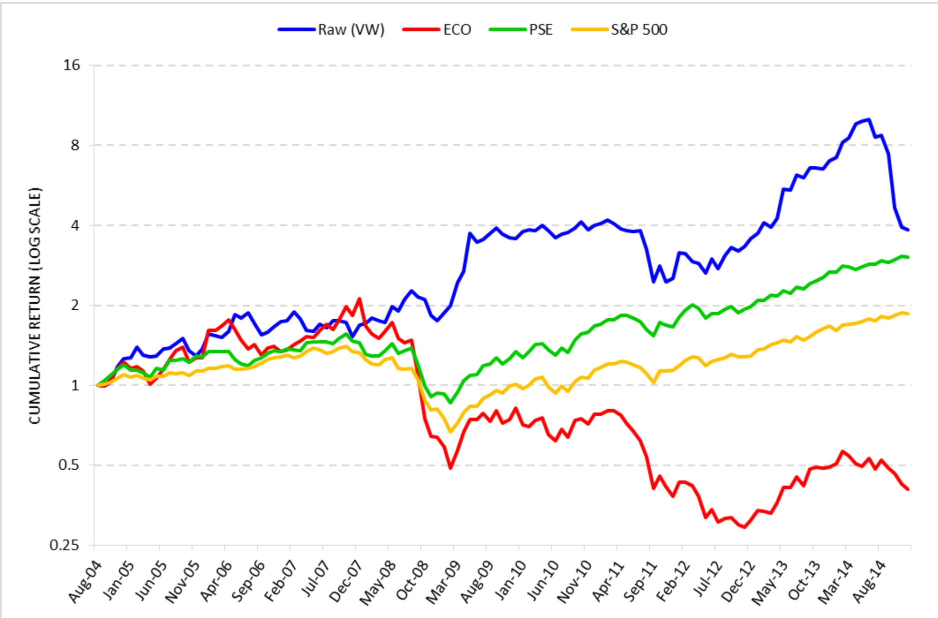
Cumulative raw returns on "High-Low" VW portfolio and chosen indices




The figure plots cumulative raw returns on "High-Low" value-weighted (VW) portfolio that takes a long position in the 20% most distressed stocks and a short position in the 20% safest stocks, along with cumulative returns on the NEX, PSE (Arca Tech 100), and S&P 500 indices over the period from February 2002 to December 2014.

**Figure 4.8**

Cumulative raw returns on "High-Low" VW portfolio and chosen indices (Aug 04 – Dec 14)




The figure plots cumulative raw returns on "High-Low" value-weighted (VW) portfolio that takes a long position in the 20% most distressed stocks and a short position in the 20% safest stocks, along with cumulative returns on the ECO, PSE (Arca Tech 100), and S&P 500 indices. Since the ECO index is only available from September 2004, considered period is September 2004 - December 2014.

greater than 0.3 (for VW portfolios). Again, this confirms the relationship between distress risk and the size effect that we pointed out earlier. Also, the momentum factor has a relatively higher correlation (in absolute values) in comparison to the SMB factor. Finally, there is a very high correlation of 0.93 for both EW and VW raw returns of portfolio 5 (high risk) and our strategy “High-Low”, confirming that it is predominantly the high distressed stocks that play a crucial role in our long-short strategy.

Lastly, we focus on a comparison of the calculated raw and risk-adjusted returns for our long-short strategy to various benchmark models, including two renewable energy indices (ECO and NEX), a technology index PSE (Arca Tech 100), and the market index S&P 500. We start in Figure 4.7 with a graphical comparison between cumulative raw returns on the “High–Low” value-weighted (VW) portfolio and cumulative returns on the NEX, PSE, and S&P 500 indices over the period from February 2002 to December 2014. We can see that raw returns from the created distress risk portfolios clearly outperform these three indices throughout the sample period. Note, however that in particular during the beginning of our sample period

**Table 4.13**

Correlation coefficients between excess returns

| Panel A: Full sample period | | | | | | |
|---|---|---|---|---|---|---|
| | EW | VW | S&P 500 | PSE | NEX | WTI |
| EW | 1 | | | | | |
| VW | 0.65 | 1 | | | | |
| S&P 500 | 0.38 | 0.40 | 1 | | | |
| PSE | 0.47 | 0.44 | 0.90 | 1 | | |
| NEX | 0.43 | 0.46 | 0.74 | 0.73 | 1 | |
| WTI | 0.20 | 0.27 | 0.25 | 0.23 | 0.44 | 1 |
| **Panel B: Period Aug 2004 - Dec 2014** | | | | | | |
| | EW | VW | S&P 500 | PSE | ECO | WTI |
| EW | 1 | | | | | |
| VW | 0.64 | 1 | | | | |
| S&P 500 | 0.32 | 0.32 | 1 | | | |
| PSE | 0.39 | 0.33 | 0.91 | 1 | | |
| ECO | 0.50 | 0.53 | 0.76 | 0.80 | 1 | |
| WTI | 0.28 | 0.40 | 0.40 | 0.42 | 0.53 | 1 |

In Panel A, the table reports correlation coefficients between excess returns on equally-weighted (EW) and value-weighted (VW) "High-Low" portfolio, that takes a long position in the 20% most distressed stocks and a short position in the 20% safest stocks, excess returns on the S&P 500, PSE (Arca Tech 100), and NEX indices, and excess returns on the U.S. WTI crude oil. Considered period is February 2002 - December 2014. In panel B, instead of the NEX the index ECO index is reported. Since the ECO index is only available from September 2004, considered period is September 2004 - December 2014.

in 2002 as well as for the second half of 2014 the "High–Low" portfolios yield relatively high negative returns. This is particularly surprising for 2014, since during this time period the benchmark indices perform significantly better than the created long-short strategy. We provide the same comparison with the ECO index over the period from August 2004 to December 2014 in Figure 4.8.[51]

We also report correlations between excess returns of EW and VW "High–Low" portfolios, chosen indices (S&P 500, PSE, NEX), and the WTI crude oil price over the period from February 2002 to December 2014 in Panel A of Table 4.13. As expected, there are relatively high correlations between excess returns in the S&P 500 and the PSE index (0.90), between the S&P 500 and the NEX (0.74), and the NEX and PSE index (0.73). However, excess returns on our long-short strategy are not highly correlated with these indices (the highest correlation coefficient is 0.47 between EW and PSE). We also find that correlations between excess returns from our long-short strategy and excess returns from WTI crude oil prices are quite low, ranging from 0.20 for EW portfolios to 0.27 for VW. While correlations between NEX or ECO index returns and returns from the WTI are typically quite pronounced (between 0.44 and 0.53), the identified distress risk premium for the renewable sector does not seem

**Table 4.14**
Descriptive statistics of monthly risk-adjusted returns (alphas) on "High-Low" portfolios

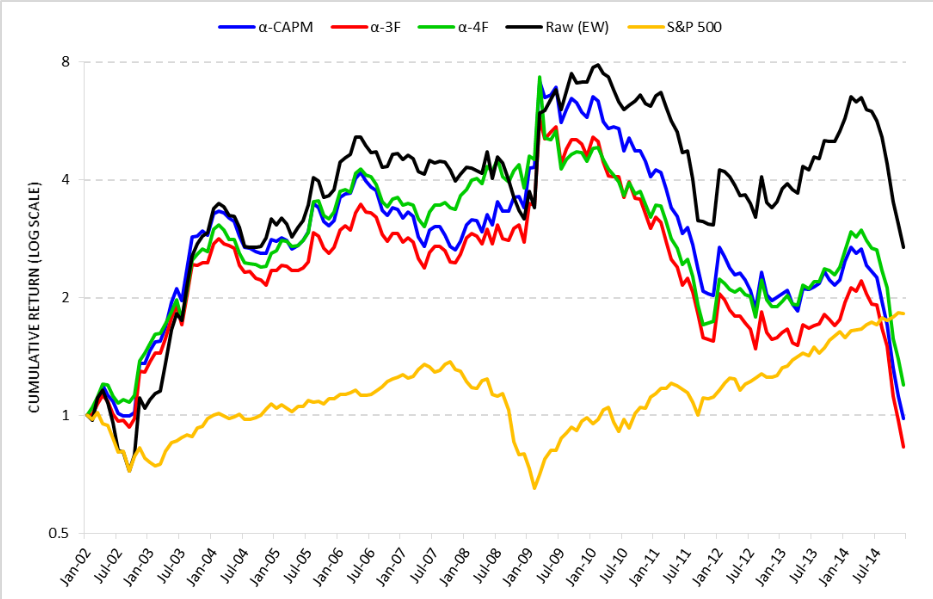
| | Mean | Std. | Min | Max | Median | Quart 1 | Quart 3 |
|---|---|---|---|---|---|---|---|
| Panel A: Equally-weighted (EW) | | | | | | | |
| α-CAPM | 0.46 | 10.26 | -22.65 | 65.73 | -0.86 | -5.44 | 3.22 |
| α-3F | 0.34 | 10.11 | -25.58 | 66.62 | -0.59 | -4.50 | 4.21 |
| α-4F | 0.56 | 9.73 | -30.20 | 62.57 | -0.15 | -4.24 | 4.40 |
| Panel B: Value-weighted (VW) | | | | | | | |
| | Mean | Std. | Min | Max | Median | Quart 1 | Quart 3 |
| α-CAPM | 1.00 | 9.91 | -39.90 | 34.29 | 0.07 | -4.39 | 6.45 |
| α-3F | 0.90 | 9.83 | -40.17 | 34.77 | 0.86 | -4.56 | 6.50 |
| α-4F | 1.12 | 9.44 | -40.14 | 32.68 | 0.78 | -3.69 | 6.15 |

The table reports descriptive statistics of monthly risk-adjusted returns (alphas) on "High-Low" equally-weighted (EW) portfolios in Panel A and value-weighted (VW) portfolios in Panel B. These portfolios takes a long position in the 20% most distressed stocks and a short position in the 20% safest stocks over the period from February 2002 to December 2014. Returns are calculated using estimated coefficients from the CAPM, Fama-French model (3F), and Carhart model (4F). Specifically, we take the difference between raw and expected returns based on the models. All values are expressed in percent units.

[51] The ECO index is only available from August 2004.

**Figure 4.9**

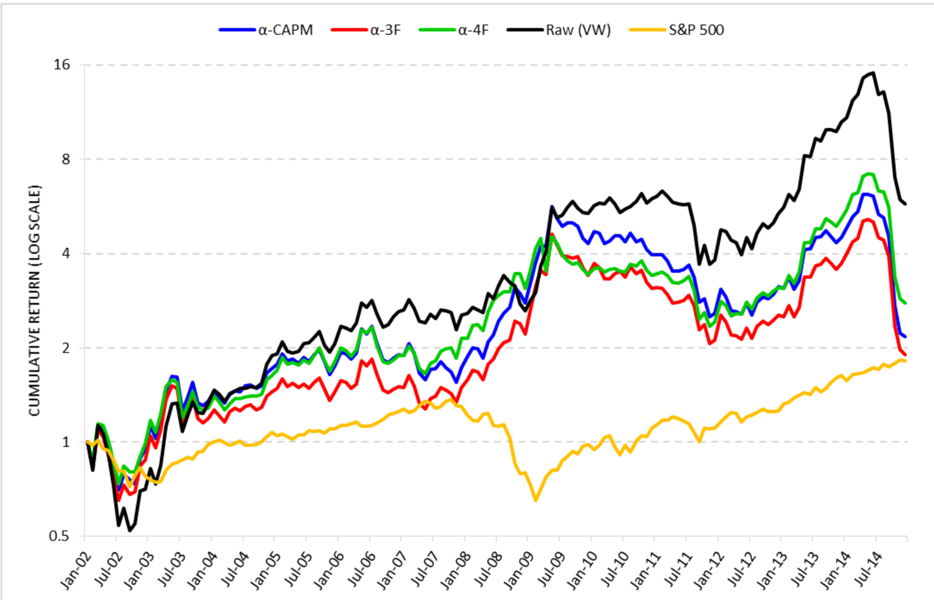
Cumulative raw and risk-adjusted returns on “High-Low” EW portfolio




The figure plots cumulative raw returns on "High-Low" equally-weighted (EW) portfolio that takes a long position in the 20% most distressed stocks and a short position in the 20% safest stocks, along with risk-adjusted returns (alphas) from CAPM model, Fama-French three-factor model, and Carhart four-factor model over the period from February 2002 to December 2014. For comparison purposes, the figure also plots cumulative return on the S&P 500 index.

**Figure 4.10**

Cumulative raw and risk-adjusted returns on “High-Low” VW portfolio




The figure plots cumulative raw returns on "High-Low" value-weighted (VW) portfolio that takes a long position in the 20% most distressed stocks and a short position in the 20% safest stocks, along with risk-adjusted returns (alphas) from CAPM model, Fama-French three-factor model, and Carhart four-factor model over the period from February 2002 to December 2014. For comparison purposes, the figure also plots cumulative return on the S&P 500 index.

to be influenced too much by returns in the oil market. Therefore, while movements in the oil price clearly are one of the driving factors of returns in the renewable sector, we do not find clear evidence for a different impact of oil returns on high and low distress risk renewable energy companies.

Note that as a robustness check we provide the same comparison of correlations using the ECO instead of the NEX for the period from August 2004 to December 2014 in Panel B of Table 4.13. Overall, results are qualitatively the same as for the NEX.

Furthermore, we report descriptive statistics of monthly risk-adjusted active returns (alphas) on our “High–Low” EW and VW portfolios over the period February 2002 to December 2014 in Table 4.14. In order to calculate these alphas we proceed as follows: first, we calculate the expected returns for each month based on the estimated coefficients from the CAPM, Fama-French three-factor model and Carhart four-factor model. Subsequently, each month we take the difference between raw returns and these expected returns. The average alphas for our constructed portfolios vary from 0.34% to 0.56% for EW long-short portfolios and from 0.90% to 1.12% for VW long-short portfolios for the different asset pricing models. Note that these are the monthly alphas. After multiplying these average values by 12, we get the same alphas as reported in Tables 4.10 and 4.11 (reported in annualized percent units). We observe that the active returns for the created portfolios are relatively volatile, with monthly standard deviations ranging from 9.73% to 10.26% for EW portfolios and from 9.44% to 9.91% for VW portfolios.

Moreover, we illustrate the cumulative performance of these risk-adjusted returns for the EW long-short portfolio in Figure 4.9 and for the VW long-short portfolio in Figure 4.10. For comparison purposes, we also plot cumulative returns of the S&P 500 index. We find that cumulative raw returns of our “High–Low” EW and VW portfolios outperform returns of the S&P 500 index over the considered sample period. We also plot the cumulative alphas from the CAPM model, the Fama-French three-factor model, and the Carhart four-factor model for the created distress risk investment strategy. We find that the alphas are highly correlated and range from 0.90 to 0.99 for the EW and from 0.94 to 0.99 for the VW long-short portfolios. As illustrated by these figures, also the performance of the cumulative risk-adjusted active returns is typically above the cumulative performance of the S&P 500 index throughout our sample period (2002–2014). However, we observe a significant drop in the performance for the second half of 2014 that was already indicated in Figures 4.7 and 4.8. It seems like in particular during this period the created distress risk portfolio strategy provided substantial negative returns. However, overall the excellent performance of the "High–Low" EW and VW portfolios also from a risk-adjusted perspective is confirmed in this section.

## 4.6 Conclusions

The trade-off between distress risk and stock returns has important implications for the risk-reward relationship in financial markets and contributes to the conceptual framework of asset pricing and investment decision making. During the last decade, investments in renewable energy stocks have accomplished tremendous growth rates in the global economy, mostly due to the conjunction of rising oil prices, increasing market liquidity for investments in renewable energy sector, and government policies. Consequently, several renewable, clean and alternative energy stock indices have been created, including the WilderHill Clean Energy Index (ECO), the WilderHill New Energy Global Innovation Index (NEX), or the S&P Global Clean Energy Index (SPGCE). At the same time, companies involved in renewable energy business are relatively highly risky firms with high profitability potential.

In this study, we contribute to the literature by combining work on the relationship between distress risk and equity returns with studies that focus on the driving factors of returns of renewable energy companies. Specifically, we investigate the relationship between distress risk and realized returns of U.S. renewable energy companies and examine risk-adjusted returns corrected for common Fama and French [65] and Carhart [37] risk factors to show whether distress risk is positively priced in the renewable sector.

Using the *Expected Default Frequency* (EDF) from *Moody's KMV* as a proxy for distress risk, we find a positive relationship between realized equity returns of both, equally-weighted (EW) and value-weighted (VW) portfolios and distress risk in the renewable energy sector. Thus, we confirm findings of Vassalou and Xing [172] and Chava and Purnanandam [39] on positive distress risk premiums. Investors expect higher average returns for bearing the additional risk of holding more distressed stocks in the renewable sector. We find a significant difference between returns of VW portfolio consisting of the riskiest quintile of stocks and those consisting of the quintile with the lowest failure risk. This positively priced distress premium in the U.S. renewable energy sector is also confirmed by applying three major asset pricing models – the CAPM, the Fama and French [65] three-factor model, and the Carhart [37] four-factor model, that correct returns for given risk factors such as market risk, size premiums, value premiums, and momentum.

We further investigate a possible link between the size and value (book-to-market) effects and default risk, and find that the size effect is concentrated in the smallest firms, which also happen to be among those with the highest distress risk. Thus, as suggested by Vassalou and Xing [172] the size effect may partially be interpreted as a default effect, however, sorting renewable stocks according to their EDF does not yield the same results as sorting them by size. The size effect is significant only in the segment of our sample with the highest distress risk, where the difference in returns between small and big firms can be explained by the difference in their default risk. We show that distressed firms earn on average higher returns than low distress risk firms, and that significantly higher returns are earned by firms that are also small in size. Unlike for the size effect, our results suggest that the book-to-market effect is not truly related to distress risk.

Our findings complement other conducted studies that mostly focus on examining returns of renewable energy companies and on identifying potential drivers of these returns. Our study is particularly closely related to Bohl, Kaufmann and Stephan [25] who investigate stocks of German renewable energy companies and show that the outperformance of German renewable energy stocks was completely reversed between 2008 and 2011, where significantly negative active returns were delivered. We find similar pattern for this time period in the U.S. market. However, we also demonstrate that raw and risk-adjusted (active) returns of VW portfolios that take a long position in the 20% most distressed stocks and a short position in the 20% safest stocks generally outperform S&P 500 index throughout our sample period (2002–2014). Returns for portfolios that implement such a “High-Low” distress risk trading strategy typically exhibit rather low correlations with standard factors in asset pricing models. Interestingly, we also find that returns for these portfolios are also not highly correlated to pricing factors for renewable energy stocks such as returns from technology stocks and oil prices. Overall, these results indicate that distress risk is systematically priced in the renewable energy sector and should be considered as an additional pricing factor for these companies.

# Chapter 5

# Summary and Conclusions

In this dissertation thesis, we have investigated several dimensions for estimation and examination of default probabilities in credit risk management. This topic has undergone substantial development in the last decades and become one of the most intensely studied topics in the financial literature. Assigning an appropriate PD, which is the key input factor for modeling and measurement of credit risk, is a widely employed strategy by financial institutions as well as the supervisory authorities around the world. Providing accurate estimates can be considered as one of the key challenges in credit risk management since false estimation of PDs might lead to unreasonable ratings and incorrect pricing of financial instruments. In fact, these were the reasons that stood behind the emergence of recent global financial crisis as undervaluation of the risk caused the collapse of the financial system which had been extended through credit derivatives on global markets.

This thesis consists of three various studies. One of the most significant approaches for estimation of default probabilities are structural credit risk models. This approach was introduced in 1974 by Merton [135] and is based on an idea of treating company's equity and debt as a contingent claim written on company's asset value. A significant attention has been given to this framework in past and the Merton model has become very popular, despite the fact that the classical version of this model is based on a number of simplifying and unrealistic assumptions. In our first study (Chapter 2), we have firstly confirmed several empirical investigations that have shown that log-returns of equities present skew distributions with excess kurtosis, which leads to a greater density in the tails, by demonstrating that the distributional assumption of the Merton model (company value follows the log-normal distribution) is generally rejected. Therefore, we have discussed the possibility for using other subordinated processes to approximate the behaviour of the log-returns of the company value. In fact, we have introduced a structural credit risk model based on stable non-Gaussian processes as a representative of subordinated models and shown that it is possible to use this model in the Merton's framework. In particular, we have proposed to use Hurst, Platen and Rachev [93] option pricing model based on the stable Paretian distributions which generalizes the standard Merton's methodology.

The practical and theoretical appeal of the stable non-Gaussian approach is given by its attractive properties that are almost the same as the normal ones. As a matter of fact, the Gaussian law is a particular stable Paretian one, and thus the stable Paretian model is a generalization of the Merton one. The first relevant desirable property of the stable distributional assumption is that stable distributions have domain of attraction. The generalized central limit theorem for the normalized sums of i.i.d. random variables determines the domain of attraction of each stable law. Therefore, any distribution in the domain of attraction of a specified stable distribution will have properties close to those of the stable distribution. Another attractive aspect of the stable Paretian assumption is then the stability property; that is, stable distributions are stable with respect to summation of i.i.d. random stable variables. Hence, the stability governs the main properties of the underlying distribution. In addition, in the empirical financial literature, it is well documented that the asset returns have a distribution whose tail is heavier than that of the distributions with finite variance. The idea of using subordinated stable Paretian processes goes back to the seminal work of Mandelbrot and Taylor [129] and stable laws then have been applied in several financial sectors. For these reasons, the stable Paretian law is the first candidate as a subordinated model.

We have proposed two different methodologies for the parameter estimation: the first is to generalize the maximum likelihood parameter estimation proposed by Duan [49]; the second is a generalization of the Moody's KMV methodology. Moreover, we have optimized the performance for the stable Lévy model and conducted an empirical comparison between the results obtained from the classical Merton model and the stable Lévy one. Besides confirming a hypothesis that the companies with a higher average value of the ratio between the debt and the companies' asset values tend to have a higher average value of default probability, our findings also suggest that PD is generally underestimated by the Merton model and that the stable Lévy model is substantially more sensitive to the periods of financial crises. We have also referred to a study conducted by Brambilla, Gurny and Ortobelli Lozza [30] who extended our work and applied two alternative structural credit risk models based on well-known symmetric Lévy processes (the Variance Gamma (VG) process and the Normal Inverse Gaussian (NIG) process). These authors concluded that both models are able to capture the situation of instability that affects each company in considered period and, in fact, are very sensitive to the periods of the crises, similar to our stable Lévy model.

One of the implications of our findings that the more leveraged companies tend to have a higher average value of PD is that the structural credit risk models based on the Merton's framework are not appropriate for estimation of PDs for financial institutions, unless some adjustments are made. This is the reason why we have devoted our second study (Chapter 3) to estimation of PDs of banks. In particular, we have derived and investigated the performance of static and multi-period credit-scoring models, which is another significant approach for determining default probabilities. Due to their simplicity, credit-scoring models are among the most popular and widely used approaches for the estimation of PDs. These multivariate models use financial indicators of a company as input and attribute a weight to each of these indicators that reflects its relative importance in predicting the risk of default.

The main contribution of this study was threefold. First, we have taken the advantage of the fact that many U.S. commercial banks defaulted during the GFC and subsequent periods, which enabled us to compile and examine a significant database of historical financial ratios for defaulted banks. Sufficient number of historical defaults is essential for estimating such models. In fact, our sample contained more than seven thousand U.S. commercial banks with over four hundred default events during our sample period 2007-2013. To the best of our knowledge, we have provided the first empirical study to use such an extensive sample of financial institutions for the estimation and evaluation of default prediction models. For instance, Canbas, Cabuk and Kilic [35] worked with 40 privately owned Turkish commercial banks and 21 defaults; or Kolari, Glennon, Shin and Caputo [104] used over 1,000 large U.S. commercial banks in each year with 55 defaults in total. Following general approach for estimation and subsequent validation of a scoring model, they split their sample of failed banks into an original sample used to build a model (containing 18 large failed banks) and a holdout sample (containing remaining 37 large failed banks). In comparison, we have used the walk-forward approach with out-of-time validation. This approach is closest to the actual application of default prediction models in practise and gives a realistic view of how a particular model would perform over time. At the same time, it allowed us to use the maximum number of available data in each period to fit and test the models while controlling for time dependence, as we were not restricted to dividing our sample into an estimation and holdout sample.

Second, we have provided the first empirical study to use the Federal Financial Institutions Examination Council (FFIEC) database and to estimate scoring models for these banks. The full sample of banks contained in this database has not been used so far to build a credit-scoring model. Specifically, we have compared static and dynamic discrete hazard models and applied logistic and probit regression techniques in order to calibrate our models.

Finally, substantial part of this study was devoted to the application of various model evaluation techniques, including techniques that have not yet been applied in the literature on credit scoring before. We have used some of the well-known techniques, such as the walk-forward approach with out-of-time validation, ROC curve analysis, calibration accuracy tests, or bootstrapping of ROC curve areas. Furthermore, building on existing work, we have applied the Kruskal-Wallis and the Tukey's multiple comparison procedure to investigate significant differences between the particular models in terms of bootstrapped ROC areas. The main advantage of these two nonparametric approaches is that they do not require the assumption of normality which would not be justified in our case. As an extension of log-likelihoods calculated within calibration accuracy test suggested in Stein [162], we have applied the Vuong's closeness test for non-nested models to determine whether calculated log-likelihoods for various models are statistically different. Moreover, we have also applied the Hosmer-Lemeshow's chi-squared goodness-of-fit test to examine the overall fit of the estimated models.

The majority of our estimated models builds on variables that form a reasonable mixture of profitability, liquidity, assets quality, and capital adequacy indicators. We have found that our models have a high default/non-default classification and predictive accuracy. Specifically, for the models that were calibrated using defaults in 2011 and 2012, more than

95% of defaulted banks were captured within the banks with the highest 10% PDs. These are very good results compared to recent studies conducted on the corporate sector. Since all the models performed very well and their performances were similar in terms of power (areas under the ROC curves) we have applied the Kruskal-Wallis and the Tukey's multiple comparison test to examine significant differences between the particular models in terms of bootstrapped ROC areas. Especially the Tukey's test has proved to be a very powerful tool as it was able to distinguish between the models where the differences between mean values of bootstrapped ROC areas were very small. Using a calibration accuracy test and its likelihood estimates we have shown that logit models typically outperform probit models in accuracy of estimated PDs in particular years. We have also found that multi-period hazard models generally produce more accurate default probability estimates compared to static models.

Moreover, since we have shown by applying the Hosmer-Lemeshow's chi-squared test that the expected and actual default rates are statistically equal for most of the deciles, we have stated that our estimated default probabilities might be considered as reasonable estimates. Also, due to the fact that we have worked in this study with all of the available information on U.S. commercial banks and thus avoided choice-based samples within estimation, we have obtained ratios of defaulted and non-defaulted banks very close to empirical ones. This was necessary in order to produce estimates that are close to "real" PDs and might be subsequently used for purposes of calculation of economic capital, credit Value-at-Risk, scenario analysis purposes etc.

Due to the number of estimated models and the fact that different models performed best according to different criteria, we have provided a summary of comparison for all the models in terms of the chosen criteria and created a simple ranking system in order to determine which model works the best for a particular year.

Unlike first two studies, where we have focused on estimation of default probabilities, in our last study (Chapter 4), we have taken advantage of Moody's KMV database and used their structural-based default probability indicators (Expected Default Frequencies – EDFs) in asset pricing framework. In particular, we have investigated whether U.S. distressed renewable energy companies earn on average higher returns than low distress risk companies.

Renewable energy sector is considered to be a relatively risky sector with high profitability potential, similar to high-tech sector or venture capital. Therefore, based on the fundamental principle of financial theory where individuals expect higher average returns for bearing risk, investors who buy stocks of renewable energy companies should expect higher average returns. Nevertheless, there has been controversy with regards to this hypothesis as the existing empirical literature has not produced consistent evidence to confirm this conjecture. In fact, only two major studies, conducted by Vassalou and Xing [172] and Chava and Purnanandam [39], found a positive cross-sectional relationship between distress risk and returns. Several other key studies (e.g. Dichev [47], Campbell, Hilscher and Szilagyi [34], or Garlappi, Shu and Yan [73]) suggest that distress risk is priced negatively - more distress stocks usually earn lower average returns (often referred to as a “distress risk puzzle”). None of these studies was applied directly to the renewable energy sector, though.

We have combined two streams of the literature in this study. Apart from studies that describe the relationship between distress risk and equity returns, we have also contributed to the literature that investigates returns on renewable energy sector. Increased interest in the effects of energy and stock market prices on the financial performance of renewable sector has been well documented by a number of empirical studies. However, these studies either focus on the relationship between renewable energy stocks and other variables, or on the effects of energy and stock market prices on the renewable sector. On the other hand, our study provides first empirical research that examines the relationship between returns of renewable energy companies and distress risk premium. We have used the Expected Default Frequency (EDF) obtained from Moody’s KMV database as a distress risk measure. Moreover, we have significantly extended the time period considered in previous studies by using a data set from 2002 up to 2014 that includes observations for the period of the global financial crisis and beyond. Also, unlike many other studies that typically look at one of the renewable energy stock indices, we have worked and examined individual companies.

After sorting the companies according to their EDF measures and subsequent evaluation of the performance of portfolios that are based on this sorting procedure, we have demonstrated that there is a positive relationship between equity returns of both, equally-weighted (EW) and value-weighted (VW) portfolios, and default risk. Thus, distressed renewable energy companies earn on average higher expected returns than renewables with low default risk. Therefore, our results confirm a pattern also suggested by Vassalou and Xing [172] and Chava and Purnanandam [39]. We have found a significant difference between returns of value-weighted (VW) portfolios consisting of the riskiest quintile of stocks and one consisting of the quintile with the lowest failure risk.

We have further examined a possible link between pricing factors such as the size effect, the BM effect and distress risk, and found that the size effect is concentrated in the smallest firms that are typically also the firms with the highest distress risk. In other words, we have shown that default risk is particularly priced for small renewable energy companies. At the same time, the size effect is most pronounced for companies with high default risk such that the highest average returns are typically observed for companies that are small in size and at the same time exhibit a relatively high risk of financial distress. Note that unlike for the size effect, our results indicate that the BM effect is not truly related to default effect.

Positively priced distress risk in the U.S. stock market for the renewable energy sector has been also confirmed by applying three major asset pricing models, namely the Capital Asset Pricing Model (CAPM), the Fama and French [65] three-factor model, and the Carhart [37] four-factor model. These models correct observed returns of the constructed portfolios for given risk factors (market return, size premium, value premium, momentum). Finally, we have shown that raw and risk-adjusted returns of VW portfolios that take a long position in the 20% most distressed renewable stocks and a short position in the 20% renewable companies with the lowest default risk, generally outperform the S&P 500 index throughout our sample period (2002–2014).

# Bibliography


[1] **Acharya VV, and Carpenter JN (2002)**. Corporate bond valuation and hedging with stochastic interest rates and endogenous bankruptcy. *Review of Financial Studies*, 15 (5), 1355-1383.

[2] **Acharya VV, Huang JZ, Subrahmanyam M, and Sundaram RK (2006)**. When does strategic debt-service matter? *Economic Theory*, 29 (2), 363-378.

[3] **Alam P, Booth D, Lee K, and Thordarson T (2000)**. The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study. *Expert Systems with Applications*, 18 (3), 185-199.

[4] **Altman EI (1968)**. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23 (4), 589-609.

[5] **Altman EI, Avery RB, Eisenbeis RA, and Sinkey JF (1981)**. *Application of Classification Techniques in Business, Banking and Finance*. Greenwich,USA: JAI press.

[6] **Altman EI, Haldeman RG, and Narayanan P (1977)**. ZETA$^{TM}$ analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1 (1), 29-54.

[7] **Altman EI, Rijken H, Balan D, Mina J, Forero J, and Watt M (2010)**. The Z-Metrics™ Methodology for Estimating Company Credit Ratings and Default Risk Probabilities. *RiskMetrics Group*.

[8] **Anderson RW, and Carverhill AP (2007)**. *Liquidity and capital structure*. Working paper No. 6044. Center for Economic Policy Research.

[9] **Arora N, Bohn JR, and Zhu F (2005)**. Reduced form vs. structural models of credit risk: A case study of three models. *Journal of Investment Management*, 3 (4), 43.

[10] **Asvanunt A, Broadie M, and Sundaresan S (2007)**. *Managing corporate liquidity: Welfare and pricing implications*. Working paper. Columbia University.

[11] **Avramov D, Chordia T, Jostova G, and Philipov A (2009)**. Credit ratings and the cross-section of stock returns. *Journal of Financial Markets*, 12 (3), 469-499.

[12] **Baesens B, Van Gestel T, Viaene S, Stepanova M, Suykens J, and Vanthienen J (2003)**. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54 (6), 627-635.

[13] **Basel_Committee_on_Banking_Supervision (2010)**. *Basel III: A global regulatory framework for more resilient banks and banking systems*.

[14] **Basel_Committee_on_Banking_Supervision (2010)**. *Basel III: International framework for liquidity risk measurement, standards and monitoring*.

[15] **Beaver WH (1966)**. Financial ratios as predictors of failure. *Journal of Accounting Research*, 4 (-), 71-111.

[16] **Beaver WH, McNichols MF, and Rhie J-W (2005)**. Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies*, 10 (1), 93-122.

[17] **Betzer A, Doumet M, and Rinne U (2013)**. How policy changes affect shareholder wealth: the case of the Fukushima Dai-ichi nuclear disaster. *Applied Economics Letters*, 20 (8), 799-803.

[18] **Bharath ST, and Shumway T (2008)**. Forecasting default with the Merton distance to default model. *Review of Financial Studies*, 21 (3), 1339-1369.

[19] **Bielecki TR, and Rutkowski M (2002)**. *Credit Risk: Modeling, Valuation and Hedging*. Berlin, Germany: Springer Science & Business Media.

[20] **Bishop CM (1995)**. *Neural Networks for Pattern Recognition*. New York, USA: Oxford University Press.

[21] **Black F, and Cox JC (1976)**. Valuing corporate securities: Some effects of bond indenture provisions. *The Journal of Finance*, 31 (2), 351-367.

[22] **Black F, and Scholes M (1973)**. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81 (3), 637-654.

[23] **Bluhm C, Overbeck L, and Wagner C (2003)**. *An Introduction to Credit Risk Modeling*. Boca Raton, USA: CRC Press.

[24] **Bohl MT, Kaufmann P, and Siklos PL (2015)**. What Drove the Mid-2000s Explosiveness in Alternative Energy Stock Prices? Evidence from US, European and Global Indices. *International Review of Financial Analysis*, 40 (-), 194-206.

[25] **Bohl MT, Kaufmann P, and Stephan PM (2013)**. From hero to zero: Evidence of performance reversal and speculative bubbles in german renewable energy stocks. *Energy Economics*, 37 (-), 40-51.

[26] **Bourgeon J-M, and Dionne G (2007)**. *On debt service and renegotiation when debt-holders are more strategic*. Working paper. CIRPEE.

[27] **Bovenzi JF, Marino JA, and McFadden FE (1983)**. Commercial bank failure prediction models. *Economic Review*, 68 (1), 14-26.

[28] **Boyacioglu MA, Kara Y, and Baykan ÖK (2009)**. Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications*, 36 (2), 3355-3366.

[29] **Boyer MM, and Filion D (2007)**. Common and fundamental factors in stock returns of Canadian oil and gas companies. *Energy Economics*, 29 (3), 428-453.

[30] **Brambilla C, Gurny M, and Ortobelli Lozza S (2015)**. Structural Credit Risk Models with Lévy Processes: The VG and NIG Cases. *Far East Journal of Mathematical Sciences*, 97 (1), 101-119.

[31] **Broadstock DC, Cao H, and Zhang D (2012)**. Oil shocks and their impact on energy related stocks in China. *Energy Economics*, 34 (6), 1888-1895.

[32] **Bruche M (2005)**. *Estimating Structural Bond Pricing Models via Simulated Maximum Likelihood*. Working paper. London School of Economics.

[33] **Byström HN (2006)**. Merton unraveled: A flexible way of modeling default risk. *The Journal of Alternative Investments*, 8 (4), 39-47.

[34] **Campbell JY, Hilscher J, and Szilagyi J (2008)**. In search of distress risk. *The Journal of Finance*, 63 (6), 2899-2939.

[35] **Canbas S, Cabuk A, and Kilic SB (2005)**. Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The Turkish case. *European Journal of Operational Research*, 166 (2), 528-546.

[36] **Carey M, and Gordy M (2007)**. *The bank as grim reaper: Debt composition and recoveries on defaulted debt*. Working Paper. Board of Governors of the Federal Reserve System

[37] **Carhart MM (1997)**. On persistence in mutual fund performance. *The Journal of Finance*, 52 (1), 57-82.

[38] **Chan K, and Chen N-f (1991)**. Structural and return characteristics of small and large firms. *The Journal of Finance*, 46 (4), 1467-1484.

[39] **Chava S, and Purnanandam A (2010)**. Is default risk negatively related to stock returns? *Review of Financial Studies*, 23 (6), 2523-2559.

[40] **Childs PD, Mauer DC, and Ott SH (2005)**. Interactions of corporate financing and investment decisions: The effects of agency conflicts. *Journal of Financial Economics*, 76 (3), 667-690.

[41] **Collin-Dufresne P, Goldstein RS, and Martin JS (2001)**. The determinants of credit spread changes. *The Journal of Finance*, 56 (6), 2177-2207.

[42] **Crosbie P, and Bohn J (2003)**. *Modeling default risk*. Moody's KMV.

[43] **Crouhy M, Galai D, and Mark R (2000)**. A comparative analysis of current credit risk models. *Journal of Banking & Finance*, 24 (1), 59-117.

[44] **Cummins M, Garry O, and Kearney C (2014)**. Price discovery analysis of green equity indices using robust asymmetric vector autoregression. *International Review of Financial Analysis*, 35 (-), 261-267.

[45] **Delianedis G, and Geske RL (2003)**. *Credit risk and risk neutral default probabilities: Information about rating migrations and defaults*. Working paper. EFA 2003.

[46] **DeLong ER, DeLong DM, and Clarke-Pearson DL (1988)**. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44 (3), 837-845.

[47] **Dichev ID (1998)**. Is the risk of bankruptcy a systematic risk? *The Journal of Finance*, 53 (3), 1131-1147.

[48] **Du Y, and Suo W (2007)**. Assessing credit quality from equity markets: Is a structural approach a better approach? *Canadian Journal of Administrative Sciences*, 24 (3), 212-218.

[49] **Duan JC (1994)**. Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, 4 (2), 155-167.

[50] **Duan JC, Gauthier G, and Simonato JG (2005)**. *On the equivalence of the KMV and maximum likelihood methods for structural credit risk models*. Working paper.

[51] **Duffie D, and Lando D (2001)**. Term structures of credit spreads with incomplete accounting information. *Econometrica*, 69 (3), 633-664.

[52] **Duffie D, Saita L, and Wang K (2007)**. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83 (3), 635-665.

[53] **Duffie D, and Singleton KJ (2003)**. *Credit Risk: Pricing, Measurement, and Management*. Princeton, USA: Princeton University Press.

[54] **Duffie D, and Singleton KJ (1999)**. Modeling term structures of defaultable bonds. *Review of Financial studies*, 12 (4), 687-720.

[55] **Durand D (1941)**. *Risk Elements in Consumer Instalment Financing*. New York, USA: National Bureau of Economic Research.

[56] **Efron B, and Tibshirani RJ (1994)**. *An Introduction to the Bootstrap*. Boca Raton, USA: Chapman & Hall/CRC.

[57] **Engelmann B, Hayden E, and Tasche D (2003)**. Testing rating accuracy. *Risk*, 16 (1), 82-86.

[58] **Ericsson J, and Renault O (2006)**. Liquidity and credit risk. *The Journal of Finance*, 61 (5), 2219-2250.

[59] **Ericsson J, and Reneby J (2005)**. Estimating Structural Bond Pricing Models. *Journal of Business*, 78 (2), 707-736.

[60] **Espahbodi P (1991)**. Identification of problem banks and binary choice models. *Journal of Banking & Finance*, 15 (1), 53-71.

[61] **Faff RW, and Brailsford TJ (1999)**. Oil price risk and the Australian stock market. *Journal of Energy Finance & Development*, 4 (1), 69-87.

[62] **Fama EF (1965)**. The behavior of stock-market prices. *The Journal of Business*, 38 (1), 34-105.

[63] **Fama EF (1963)**. Mandelbrot and the stable Paretian hypothesis. *The Journal of Business*, 36 (4), 420-429.

[64] **Fama EF (1965)**. Portfolio analysis in a stable Paretian market. *Management Science*, 11 (3), 404-419.

[65] **Fama EF, and French KR (1993)**. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33 (1), 3-56.

[66] **Fama EF, and French KR (1992)**. The cross-section of expected stock returns. *The Journal of Finance*, 47 (2), 427-465.

[67] **Feller W (1966)**. *An Introduction to Probability Theory and Its Applications II*. New York, USA: John Wiley & Sons.

[68] **Ferstl R, Utz S, and Wimmer M (2012)**. The effect of the Japan 2011 disaster on nuclear and alternative energy stocks worldwide: an event study. *BuR-Business Research*, 5 (1), 25-41.

[69] **Fisher RA (1936)**. The use of multiple measurements in taxonomic problems. *Annals of Human Genetics*, 7 (2), 179-188.

[70] **Follmer H, and Schweizer M (1993)**. A microeconomic approach to diffusion models for stock prices. *Mathematical Finance*, 3 (1), 1-23.

[71] **Follmer H, and Sondermann D (1986)**. *Contributions to Mathematical Economics*. Amsterdam, The Netherlands: North Holland.

[72] **Fons JS (1994)**. Using default rates to model the term structure of credit risk. *Financial Analysts Journal*, 50 (5), 25-32.

[73] **Garlappi L, Shu T, and Yan H (2008)**. Default risk, shareholder advantage, and stock returns. *Review of Financial Studies*, 21 (6), 2743-2778.

[74] **Garlappi L, and Yan H (2011)**. Financial Distress and the Cross-section of Equity Returns. *The Journal of Finance*, 66 (3), 789-822.

[75] **Geske R (1977)**. The valuation of corporate liabilities as compound options. *Journal of Financial and Quantitative Analysis*, 12 (4), 541-552.

[76] **Giesecke K, and Goldberg LR (2004)**. Sequential defaults and incomplete information. *Journal of Risk*, 7 (-), 1-26.

[77] **Green DM, and Swets JA (1966)**. *Signal Detection Theory and Psychophysics*. New York, USA: John Wiley & Sons.

[78] **Griffin JM, and Lemmon ML (2002)**. Book-to-market equity, distress risk, and stock returns. *The Journal of Finance*, 57 (5), 2317-2336.

[79] **Gurny M, Kalotay E, and Trück S (2015)**. *Prediction of U.S. Commercial Bank Failures via Scoring Models: The FFIEC Database Case*. Working paper. Macquarie University.

[80] **Gurny M, Ortobelli Lozza S, and Giacometti R (2013)**. Structural Credit Risk Models with Subordinated Processes. *Journal of Applied Mathematics*, 2013 (-), 12.

[81] **Gurny M, and Trück S (2015)**. *Distress Risk and Stock Returns of U.S. Renewable Energy Companies*. Working paper. Macquarie University.

[82] **Hackbarth D, Hennessy CA, and Leland HE (2007)**. Can the trade-off theory explain debt structure? *Review of Financial Studies*, 20 (5), 1389-1428.

[83] **Hackbarth D, Miao J, and Morellec E (2006)**. Capital structure, credit risk, and macroeconomic conditions. *Journal of Financial Economics*, 82 (3), 519-550.

[84] **Hand DJ, and Henley WE (1997)**. Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160 (3), 523-541.

[85] **Hanley JA (1989)**. Receiver operating characteristic (ROC) methodology: the state of the art. *Critical Reviews in Diagnostic Imaging*, 29 (3), 307-335.

[86] **Henriques I, and Sadorsky P (2008)**. Oil prices and the stock prices of alternative energy companies. *Energy Economics*, 30 (3), 998-1010.

[87] **Hillegeist SA, Keating EK, Cram DP, and Lundstedt KG (2004)**. Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9 (1), 5-34.

[88] **Hochberg Y, and Tamhane AC (1987)**. *Multiple Comparison Procedures*. New York, USA: John Wiley & Sons.

[89] **Hofmann N, Platen E, and Schweizer M (1992)**. Option pricing under incompleteness and stochastic volatility. *Mathematical Finance*, 2 (3), 153-187.

[90] **Hollander M, and Wolfe D (1999)**. *Nonparametric Statistical Methods*. New York, USA: John Wiley & Sons.

[91] **Hosmer Jr DW, and Lemeshow S (2004)**. *Applied Logistic Regression*. John Wiley & Sons.

[92] **Hull J (2012)**. *Risk Management and Financial Institutions*. Hoboken, USA: John Wiley & Sons.

[93] **Hurst SR, Platen E, and Rachev ST (1999)**. Option pricing for a logstable asset price model. *Mathematical and Computer Modelling*, 29 (10), 105-119.

[94] **Inchauspe J, Ripple RD, and Trück S (2015)**. The dynamics of returns on renewable energy companies: A state-space approach. *Energy Economics*, 48 (-), 325-335.

[95] **Ioannidis C, Pasiouras F, and Zopounidis C (2010)**. Assessing bank soundness with classification techniques. *Omega*, 38 (5), 345-357.

[96] **Jarque CM, and Bera AK (1987)**. A test for normality of observations and regression residuals. *International Statistical Review*, 55 (2), 163-172.

[97] **Jarrow RA, and Turnbull SM (1995)**. Pricing derivatives on financial securities subject to credit risk. *The Journal of Finance*, 50 (1), 53-53.

[98] **Jo H, and Han I (1996)**. Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction. *Expert Systems with Applications*, 11 (4), 415-422.

[99] **Jovan M (2010)**. The Merton structural model and IRB compliance. *Metodološki Zvezki*, 7 (1), 39-57.

[100] **Ju N, Parrino R, Poteshman AM, and Weisbach MS (2005)**. Horses and rabbits? Trade-off theory and optimal capital structure. *Journal of Financial and Quantitative Analysis*, 40 (2), 259-281.

[101] **Kealhofer S, and Kurbat M (2002)**. *The Default Prediction Power of the Merton Approach, relative to Debt Ratings and Accounting Variables*. KMV LLC.

[102] **Kim IJ, Ramaswamy K, and Sundaresan S (1993)**. Does default risk in coupons affect the valuation of corporate bonds?: A contingent claims model. *Financial Management*, 22 (3), 117-131.

[103] **Kolari J, Caputo M, and Wagner D (1996)**. Trait recognition: An alternative approach to early warning systems in commercial banking. *Journal of Business Finance & Accounting*, 23 (9-10), 1415-1434.

[104] **Kolari J, Glennon D, Shin H, and Caputo M (2002)**. Predicting large US commercial bank failures. *Journal of Economics and Business*, 54 (4), 361-387.

[105] **Korobow L, and Stuhr DP (1985)**. Performance measurement of early warning models: Comments on West and other weakness/failure prediction models. *Journal of Banking & Finance*, 9 (2), 267-273.

[106] **Korobow L, and Stuhr DP (1975)**. Toward early warning of changes in banks' financial condition: A progress report. *Monthly Review*.

[107] **Korobow L, Stuhr DP, and Martin D (1976)**. A probabilistic approach to early warning of changes in bank financial condition. *Monthly Review*.

[108] **Kumar S, Managi S, and Matsuda A (2012)**. Stock prices of clean energy firms, oil and carbon markets: A vector autoregressive analysis. *Energy Economics*, 34 (1), 215-226.

[109] **Laajimi S (2012)**. Structural Credit Risk Models: A Review. *Insurance and Risk Management*, 80 (1), 53-93.

[110] **Lam KF, and Moy JW (2002)**. Combining discriminant methods in solving classification problems in two-group discriminant analysis. *European Journal of Operational Research*, 138 (2), 294-301.

[111] **Lando D (2009)**. *Credit Risk Modeling: Theory and Applications*. Princeton, UK: Princeton University Press.

[112] **Lando D (1998)**. On Cox processes and credit risky securities. *Review of Derivatives research*, 2 (2-3), 99-120.

[113] **Lane WR, Looney SW, and Wansley JW (1986)**. An application of the Cox proportional hazards model to bank failure. *Journal of Banking & Finance*, 10 (4), 511-531.

[114] **Lanine G, and Vennet RV (2006)**. Failure prediction in the Russian bank sector with logit and trait recognition models. *Expert Systems with Applications*, 30 (3), 463-478.

[115] **Lintner J (1965)**. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47 (1), 13-37.

[116] **Linton L, and Harder L (2007)**. Biology 315–Quantitative Biology Lecture Notes. *Calgary, AB: University of Calgary*.

[117] **Lo AW (1986)**. Logit versus discriminant analysis: A specification test and application to corporate bankruptcies. *Journal of Econometrics*, 31 (2), 151-178.

[118] **Logan A (2001)**. *The United Kingdom's small banks' crisis of the early 1990s: what were the leading indicators of failure?* Bank of England.

[119] **Longstaff FA (1995)**. How much can marketability affect security values? *The Journal of Finance*, 50 (5), 1767-1774.

[120] **Longstaff FA, and Schwartz ES (1995)**. A simple approach to valuing risky fixed and floating rate debt. *The Journal of Finance*, 50 (3), 789-819.

[121] **Lopatta K, and Kaspereit T (2014)**. The cross-section of returns, benchmark model parameters, and idiosyncratic volatility of nuclear energy firms after Fukushima Daiichi. *Energy Economics*, 41 (-), 125-136.

[122] **Lopez JA (2004)**. The empirical relationship between average asset correlation, firm probability of default, and asset size. *Journal of Financial Intermediation*, 13 (2), 265-283.

[123] **Madan DB, and Unal H (1998)**. Pricing the risks of default. *Review of Derivatives Research*, 2 (2-3), 121-160.

[124] **Maddala G (1986)**. *Econometric issues in the empirical analysis of thrift institutions' insolvency and failure*. Office of Policy and Economic Research, Federal Home Loan Bank Board.

[125] **Managi S, and Okimoto T (2013)**. Does the price of oil interact with clean energy prices in the stock market? *Japan and the World Economy*, 27 (-), 1-9.

[126] **Mandelbrot B (1963)**. New Methods in Statistical Economics. *Journal of Political Economy*, 71 (-), 421-440.

[127] **Mandelbrot B (1963)**. The variation of Certain Speculative Prices. *Journal of Business*, 36 (-), 394-419.

[128] **Mandelbrot B (1967)**. The variation of some other speculative prices. *The Journal of Business*, 40 (4), 393-413.

[129] **Mandelbrot B, and Taylor HM (1967)**. On the distribution of stock price differences. *Operations Research*, 15 (6), 1057-1062.

[130] **Markowitz H (1952)**. Portfolio selection. *The Journal of Finance*, 7 (1), 77-91.

[131] **Markowitz H (1959)**. *Portfolio Selection: Efficient Diversification of Investments*. New York, USA: John Wiley & Sons.

[132] **Martin D (1977)**. Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance*, 1 (3), 249-276.

[133] **Mauer DC, and Triantis AJ (1994)**. Interactions of corporate financing and investment decisions: A dynamic framework. *The Journal of Finance*, 49 (4), 1253-1277.

[134] **McFadden D (1976)**. A comment on discriminant analysis" versus" logit analysis. In: *Annals of Economic and Social Measurement*. New York, USA: National Bureau of Economic Research, p. 511-523.

[135] **Merton RC (1974)**. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29 (2), 449-470.

[136] **Mitchell M (1998)**. *An Introduction to Genetic Algorithms*. London, UK: MIT Press.

[137] **Morellec E (2001)**. Asset liquidity, capital structure, and secured debt. *Journal of Financial Economics*, 61 (2), 173-206.

[138] **Mossin J (1966)**. Equilibrium in a capital asset market. *Econometrica: Journal of the Econometric Society*, 34 (4), 768-783.

[139] **Neagu R, Keenan S, and Chalermkraivuth K (2009)**. Internal credit rating systems: Methodology and economic value. *The Journal of Risk Model Validation*, 3 (2), 11-34.

[140] **Nolan JP (1997)**. Numerical calculation of stable densities and distribution functions. *Communications in Statistics - Stochastic models*, 13 (4), 759-774.

[141] **Ohlson JA (1980)**. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18 (1), 109-131.

[142] **Onmus–Baykal E (2010)**. *A Literature Review of Credit Risk Modeling*. Working paper. Georgetown University.

[143] **Ortas E, and Moneva JM (2013)**. The Clean Techs equity indexes at stake: Risk and return dynamics analysis. *Energy*, 57 (-), 259-269.

[144] **Osborne MM (1959)**. Brownian motion in the stock market. *Operations Research*, 7 (2), 145-173.

[145] **Queen M, and Roll R (1987)**. Firm mortality: using market indicators to predict survival. *Financial Analysts Journal*, 43 (3), 9-26.

[146] **Rachev ST (2003)**. *Handbook of Heavy Tailed Distributions in Finance*. Amsterdam, The Netherlands: Elsevier.

[147] **Rachev ST, and Mittnik S (2000)**. *Stable Paretian models in finance*. Chichester, UK: John Wiley & Sons.

[148] **Reboredo JC (2015)**. Is there dependence and systemic risk between oil and renewable energy stock prices? *Energy Economics*, 48 (-), 32-45.

[149] **Rice JA (1995)**. *Mathematical Statistics and Data Analysis*. Belmont, USA: Duxbury Press.

[150] **Sadorsky P (2012)**. Correlations and volatility spillovers between oil prices and the stock prices of clean energy and technology companies. *Energy Economics*, 34 (1), 248-255.

[151] **Sadorsky P (2012)**. Modeling renewable energy company risk. *Energy Policy*, 40 (-), 39-48.

[152] **Sadorsky P (2001)**. Risk factors in stock returns of Canadian oil and gas companies. *Energy Economics*, 23 (1), 17-28.

[153] **Samorodnitsky G, and Taqqu M (1994)**. *Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*. New York, USA: Chapman & Hall.

[154] **Santomero AM, and Vinso JD (1977)**. Estimating the probability of failure for commercial banks and the banking system. *Journal of Banking & Finance*, 1 (2), 185-205.

[155] **Sarkar S, and Zapatero F (2003)**. The Trade-off Model with Mean Reverting Earnings: Theory and Empirical Tests. *The Economic Journal*, 113 (490), 834-860.

[156] **Sharpe WF (1964)**. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19 (3), 425-442.

[157] **Shumway T (2001)**. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74 (1), 101-124.

[158] **Sinkey JF (1975)**. A multivariate statistical analysis of the characteristics of problem banks. *The Journal of Finance*, 30 (1), 21-36.

[159] **Sironi A, and Resti A (2007)**. *Risk Management and Shareholders' Value in Banking: From Risk Measurement Models to Capital Allocation Policies*. Chichester, UK: John Wiley & Sons.

[160] **Sobehart JR, Keenan SC, and Stein RM (2000)**. *Benchmarking Quantitative Default Risk Models: A Validation Methodology*. Moody’s Investors Service, Global Credit Research.

[161] **Stein RM (2006)**. Are the probabilities right? Dependent defaults and the number of observations required to test for default rate accuracy. *Journal of Investment Management*, 4 (2), 61.

[162] **Stein RM (2007)**. Benchmarking default prediction models: Pitfalls and remedies in model validation. *Journal of Risk Model Validation*, 1 (1), 77-113.

[163] **Stein RM (2002)**. *Benchmarking default prediction models: Pitfalls and remedies in model validation*. Moody’s KMV, Technical Report.

[164] **Stuhr DP, and Van Wicklen R (1974)**. Rating the financial condition of banks: A statistical approach to aid bank supervision. *Federal Reserve Bank of New York Monthly Review*, 56 (9).

[165] **Sundaresan S, and Wang N (2007)**. Investment under uncertainty with strategic debt service. *The American Economic Review*, 97 (2), 256-261.

[166] **Tabachnick BG, and Fidell LS (2007)**. *Using Multivariate Statistics* Boston, USA: Pearson Education Inc.

[167] **Tam KY, and Kiang MY (1992)**. Managerial applications of neural networks: the case of bank failure predictions. *Management Science*, 38 (7), 926-947.

[168] **Thomson JB (1991)**. Predicting bank failures in the 1980s. *Economic Review*, 27 (1), 9-20.

[169] **Trück S, and Rachev ST (2009)**. *Rating Based Modeling of Credit Risk: Theory and Application of Migration Matrices*. Norwell, USA: Academic Press.

[170] **Uhrig-Homburg M (2002)**. Valuation of Defaultable Claims - A Survey. *Schmalenbach Business Review*, 54 (-), 24-57.

[171] **Vasicek OA (1984)**. *Credit valuation*. KMV Corporation.

[172] **Vassalou M, and Xing Y (2004)**. Default risk in equity returns. *The Journal of Finance*, 59 (2), 831-868.

[173] **Vitale R, and Laux JA (2012)**. The Economic Efficacy Of Banking Mergers: 2006-2008. *Journal of Applied Business Research (JABR)*, 28 (6), 1211-1216.

[174] **Vuong QH (1989)**. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57 (2), 307-333.

[175] **Wen X, Guo Y, Wei Y, and Huang D (2014)**. How do the stock prices of new energy and fossil fuel companies correlate? Evidence from China. *Energy Economics*, 41 (-), 63-75.

[176] **West RC (1985)**. A factor-analytic approach to bank condition. *Journal of Banking & Finance*, 9 (2), 253-266.

[177] **Weyant JP (2011)**. Accelerating the development and diffusion of new energy technologies: Beyond the "valley of death". *Energy Economics*, 33 (4), 674-682.

[178] **Whalen G, and Thomson JB (1988)**. Using financial data to identify changes in bank condition. *Economic Review*, 24 (-), 17-26.

[179] **Zhou C (2001)**. The term structure of credit spreads with jump risk. *Journal of Banking & Finance*, 25 (11), 2015-2040.

[180] **Zmijewski ME (1984)**. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22 (-), 59-82.