

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

*Department of Applied Finance and Actuarial Studies, Faculty of Business and Economics,
Macquarie University Sydney, Australia*

and

*Department of Management, Economics and Quantitative Methods, University of Bergamo,
Italy*

Martin Gurny, M.Sc.

*Thesis submitted to the Faculty of Business and Economics, Macquarie University,
in fulfilment of requirement for the degree of Doctor in Philosophy*

Sydney, December 2015

DECLARATION OF THE CANDIDATE

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material that to a substantial extent has been accepted for the award of any other degree or diploma at any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by colleagues with whom I have worked elsewhere during my candidature is fully acknowledged.

The work presented in this thesis has been submitted to Macquarie University Sydney, Australia and University of Bergamo, Italy and supported by the joint PhD supervision (Cotutelle) agreement.

I agree that this thesis be accessible for the purpose of study and research in accordance with the normal conditions established by the Executive Director, Library Services or nominee, for the care, loan and reproduction of theses.



Sydney, 8th December 2015

Martin Gurny

ACKNOWLEDGEMENTS

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

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

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.

The thesis was copyedited by Jera Editing Services.

Martin Gurny

ABSTRACT

This PhD thesis is devoted to the 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 of PDs is considered as one of the key challenges in credit risk management. False estimation of these probabilities 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 default risks in the mortgage market almost caused the collapse of the financial system.

The first research paper, titled *Structural Credit Risk Models with Subordinated Processes*, analyses structural models based on the Merton framework. First, we observe that the classical distributional assumption of the Merton model is 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 original Merton approach. 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 techniques for subordinated processes, and we optimise the performance for the stable Paretian model. Our results indicate that PDs are generally underestimated by the Merton model and that the stable Lévy model is substantially more sensitive to the period of the financial crises.

The second study, *Prediction of U.S. Commercial Bank Failures via Scoring Models: The FFIEC Database Case*, is devoted to examining 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). Our sample contains more than 7,000 U.S. commercial banks with 405 default events. Our analysis also focuses on evaluating the performance of the considered scoring techniques. We apply a substantial number of model evaluation techniques, including

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

Finally, the last research paper, titled *Distress Risk and Stock Returns of U.S. Renewable Energy Companies*, investigates the question of whether distressed risk is priced in the renewable energy sector. Using the Expected Default Frequency (EDF) measure obtained from Moody's KMV, we demonstrate that there is a positive cross-sectional relationship between distress risk and returns of equally-weighted (EW) portfolios, providing 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 EW portfolio that takes a long position in the 10% most distressed stocks, and a short position in the 10% safest stocks, generally outperform the S&P 500 index throughout our sample period (2002–2013).

Table of Contents

1	Introduction	1
1.1	Structural models.....	3
1.2	Scoring systems (credit-scoring models)	5
1.3	Distress risk within an asset-pricing framework.....	9
1.4	Structure of the thesis	11
2	Structural Credit Risk Models with Subordinated Processes	15
2.1	Introduction	16
2.2	Literature review	18
2.3	Merton and subordinated credit risk models.....	23
2.3.1	The Merton-Black-Scholes distributional assumptions	23
2.3.2	Credit risk models with subordinated assumptions	25
2.3.3	The Mandelbrot-Taylor distributional assumptions	28
2.4	Estimation methodology	30
2.4.1	Parameter estimates for the Merton model	30
2.4.2	Parameter estimates for subordinated models.....	33
2.5	Application and results	34
2.5.1	Analysis of the distributional assumptions of the company value log-returns	35
2.5.2	PD estimates from the Merton model	37
2.5.3	PD estimates from the stable Lévy model	40
2.5.4	Comparison of the Merton and stable Lévy models.....	42
2.6	Conclusions	44
3	Prediction of U.S. Commercial Bank Failures via Scoring Models: The FFIEC Database Case	50
3.1	Introduction	51
3.2	Literature review	55
3.3	The FFIEC council & database.....	58
3.4	Credit-scoring and model evaluation techniques.....	62
3.4.1	Logistic and probit regressions	62
3.4.2	Static and discrete hazard models	64
3.4.3	Evaluation techniques.....	66
3.4.3.1	ROC analysis	67
3.4.3.2	Bootstrapping, Kruskal-Wallis and Tukey's multiple comparison test.....	68

3.4.3.3	Calibration accuracy test and Vuong's closeness test	70
3.4.3.4	Hosmer-Lemeshow test	71
3.5	Application and results	73
3.5.1	Data description	73
3.5.2	Model estimation	76
3.5.3	Model validation	84
3.5.3.1	Distributions of estimated PDs	85
3.5.3.2	ROC curve analysis	85
3.5.3.3	Sizing the variability of ROC areas – bootstrapping	88
3.5.3.4	Calibration accuracy test	91
3.5.3.5	PD analysis for the entire score sample	93
3.5.3.6	Focus on the tails	98
3.5.3.7	Estimated PDs and “real” PDs	102
3.5.3.8	A summary of the results on model comparison	103
3.6	Conclusions	107
4	Distress Risk and Stock Returns of U.S. Renewable Energy Companies	113
4.1	Introduction	114
4.2	Literature review	117
4.3	Asset pricing models	123
4.3.1	The Capital Asset Pricing Model (CAPM)	123
4.3.2	The Fama-French three-factor model	124
4.3.3	The Carhart four-factor model	124
4.3.4	Fama-MacBeth regression	125
4.4	Data description	126
4.4.1	Returns	130
4.4.2	Expected Default Frequencies (EDFs)	131
4.5	Distress risk and equity returns	135
4.5.1	Size, BM, and distress risk	136
4.5.2	The default effect	140
4.6	Pricing of distress risk	143
4.7	Conclusions	153
5	Summary and Conclusions	165
	Bibliography	173

List of Figures

Structural Credit Risk Models with Subordinated Processes

Fig. 1	<i>The Merton model – monthly PDs</i>	39
Fig. 2	<i>The stable Lévy model – monthly PDs</i>	41
Fig. 3	<i>Difference between the models for chosen companies</i>	43

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

Fig. 1	<i>An example of the ROC curve</i>	68
Fig. 2	<i>Illustration of data used for estimation and validation of the 2008 static model</i>	77
Fig. 3	<i>ROC curves</i>	87
Fig. 4	<i>Distribution of ROC areas based on bootstrapping</i>	89
Fig. 5	<i>Expected vs. actual default rates</i>	96
Fig. 6	<i>Ratios of # of defaulted over # of non-defaulted banks</i>	100
Fig. 7	<i>Expected vs. actual # of defaults</i>	100

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

Fig. 1	<i>Number of companies in the sample through time</i>	129
Fig. 2	<i>Mean of EDF measure</i>	134
Fig. 3	<i>EDF – coefficient of variation</i>	134
Fig. 4	<i>Portfolio alphas from the regressions of excess returns</i>	145
Fig. 5	<i>Factor loadings from the four-factor regression of excess returns</i>	146
Fig. 6	<i>Cumulative raw returns on “High-Low” EW portfolio and chosen indices</i>	150
Fig. 7	<i>Cumulative raw and risk-adjusted returns on “High-Low” EW portfolio</i>	152
Fig. A1	<i>Defaulted companies (average size)</i>	158
Fig. A2	<i>Scatterplot matrix of equally-weighted (EW) raw returns and given factors</i>	163
Fig. A3	<i>Cumulative raw returns on “High-Low” EW portfolio and chosen indices (Aug 04 – Dec 13)</i>	164

List of Tables

Structural Credit Risk Models with Subordinated Processes

Tab. 1	<i>Descriptive statistics for the log-returns of the companies' asset values</i>	36
Tab. 2	<i>Outcomes from the Merton model</i>	38
Tab. 3	<i>Outcomes from the stable Lévy model</i>	40

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

Tab. 1	<i>U.S. banks according to institution type</i>	59
Tab. 2	<i>Number of U.S. banks in particular states</i>	61
Tab. 3	<i>Mean value and percentiles of total assets (in \$US000) for U.S. banks</i>	62
Tab. 4	<i>Comparison of the number of banks (FFIEC vs. U.S. in total)</i>	73
Tab. 5	<i>Data collection dates for the models</i>	74
Tab. 6	<i>List of explanatory variables</i>	75
Tab. 7	<i>Estimated static models (years 2008 – 2012)</i>	80
Tab. 8	<i>Estimated discrete hazard models (years 2009 – 2012)</i>	81
Tab. 9	<i>Calibration accuracy test (in-sample calibration)</i>	83
Tab. 10	<i>Vuong's closeness test for non-nested models (in-sample calibration)</i>	83
Tab. 11	<i>Number of banks and data collection dates for control samples</i>	84
Tab. 12	<i>PDs' statistics of the distributions (non-defaulted and defaulted banks)</i>	86
Tab. 13	<i>ROC areas and accuracy ratios (AR)</i>	87
Tab. 14	<i>ROC area statistics of the distribution (bootstrapping)</i>	88
Tab. 15	<i>Kruskal-Wallis test</i>	90
Tab. 16	<i>Tukey's test</i>	91
Tab. 17	<i>Calibration accuracy test (out-of-sample validation period)</i>	92
Tab. 18	<i>Vuong's closeness test for non-nested models (out-of-sample validation period)</i>	93
Tab. 19	<i>E(PD), actual and expected # of defaults, and ratios for out-of-time validation</i> .	94
Tab. 20	<i>Expected vs. actual default rates (Hosmer-Lemeshow test)</i>	95

Tab. 21	<i>Expected vs. actual default rates (Hosmer-Lemeshow test for deciles)</i>	97
Tab. 22	<i>Ratios of # of defaulted over # of non-defaulted banks (banks with the highest 10% PDs)</i>	99
Tab. 23	<i>Expected vs. actual number of defaults (Hosmer-Lemeshow test)</i>	101
Tab. 24	<i>A summary of the results on model comparison</i>	105
Tab. 25	<i>Rankings of the models</i>	106

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

Tab. 1	<i>Summary statistics of returns and volatilities of returns</i>	130
Tab. 2	<i>Summary statistics of the EDF measure</i>	131
Tab. 3	<i>Raw returns on portfolios sorted on the basis of the EDF measure</i>	135
Tab. 4	<i>Size effect controlled by default risk</i>	137
Tab. 5	<i>BM effect controlled by default risk</i>	139
Tab. 6	<i>Default effect controlled by size</i>	141
Tab. 7	<i>Default effect controlled by BM</i>	142
Tab. 8	<i>Risk-adjusted returns on EW portfolios sorted on the basis of the EDF measure</i>	144
Tab. 9	<i>Correlation coefficients between raw returns and given factors</i>	148
Tab. 10	<i>The Fama-MacBeth regression with the distress risk factor</i>	149
Tab. 11	<i>Correlation coefficients between excess returns</i>	151
Tab. 12	<i>Descriptive statistics of monthly risk-adjusted returns (alphas) on “High-Low” portfolios</i>	151
Tab. A1	<i>Defaulted companies</i>	159
Tab. A2	<i>Acquired companies</i>	160
Tab. A3	<i>Correlation coefficients between excess returns (Sep 04 – Dec 13)</i>	164

List of Abbreviations

AGINA	the Ardour Global Alternative Energy Index North America
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
ECO	the WilderHill Clean Energy Index
EDF	Expected Default Frequency
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
FN	false negative

FP	false positive
FRB	the Board of Governors of the Federal Reserve System
GARCH	Generalized Autoregressive Conditional Heteroskedasticity 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
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
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
RWA	risk weighted assets
S&P	Standard & Poor's
SMB	small-minus-big (size factor)
SPGCE	the S&P Global Clean Energy Index
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
VAR	Conventional Vector Autoregression model
VG	Variance Gamma process
VK	Vasicek-Kealhofer model
WRDS	Wharton Research Data Services

1 Introduction

Credit risk modeling and the estimation of default probabilities (PDs), that represent a borrower's risk level, have become some of the most intensely studied topics in the financial literature and have undergone extensive development in recent 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 has increased substantially 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, when the undervaluation of default risk in the mortgage market and for structured credit products such as collateralized debt obligations almost caused the collapse of the financial system which had been over-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 PDs is and why it has been a significant research topic for a long time.

Probability of default itself, and its estimation and subsequent validation in particular, is the core of this thesis. The thesis provides significant contribution to the state-of-the-art of respective sub-topics of credit risk. In each chapter relevant open questions are discussed, new methods and approaches are suggested, and extensive study utilising real market data is provided.

There are several ways to proceed with estimating PDs, see, for example, Hull (2012) or Trück and Rachev (2009) for more details. In general, there are two model-based approaches, both with their own advantages and drawbacks, and their application largely depends on situation in which we want to make our estimation. The first approach focuses on so-called

credit-scoring models that are based on economic and financial indicators of a company (we will discuss these models in Section 1.2 and apply some of them in Chapter 3). These models are among the ones most widely used to predict a company's default. They can be found in virtually all types of credit analysis, from consumer credit to commercial loans, and are mostly applied when a given counterparty has not issued bonds and, therefore, market valuation is not possible. Since we work with the entire database of US banks in Chapter 3, where many of them are not publicly traded, we apply credit-scoring models in our analysis.

The second approach is then the utilisation of market valuation. In this case we are either talking about structural models (based on equity prices) or so-called reduced-form models (based on bond prices). The framework of structural models was introduced by Merton (1974). Unlike instruments 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 probability estimation. We apply this approach (along with a structural model which differs from the Merton one in its distributional assumptions) in Chapter 2.

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 (2009)). 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. However, we will not be discussing this approach in further detail in this thesis. Instead, we refer to Duffie and Singleton (2003) or Trück and Rachev (2009), for example, for an overview.

This PhD thesis deals with topics related to the estimation and examination of default probabilities within credit risk management. The contribution of the thesis concentrates on the following three research areas:

- structural models

- scoring systems (credit-scoring models)
- distress risk within an asset-pricing framework.

This introduction discusses the motivation and objectives for each of these research topics and provides an overview of the thesis.

1.1 Structural models

The framework of structural models was introduced by Merton (1974). Unlike instruments 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 a company's equity and debt as a contingent claim written on a company's asset value. Suppose, for simplicity, that a firm has one zero-coupon bond outstanding and that the bond matures at time T . In the following we denote the value of the company's assets today, and at time T , as V_0 and V_T , respectively, and the value of the company's equity today and at time T as E_0 and E_T , respectively, the amount of debt interest and principal due to be repaid at time T as D , the volatility of assets (assumed constant) as σ_V , and the instantaneous volatility of equity as σ_E .

If $V_T < D$, it is rational for the company to default on the debt at time T .¹ 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

$$E_T = \max(V_T - D, 0). \quad (1)$$

In line with the Black-Scholes option pricing theory² (Black and Scholes, 1973), the Merton model stipulates that the company's equity can be considered as a European call option on

¹ It is important to make a distinction between short and long term debts. It is often supposed that the default should occur only if the asset value drops below the former; i.e. it can be below the total debts for some time. We discuss the relevant debt usually used in this framework in more detail in Footnote 12.

² In fact, a so called Black and Scholes model and relevant PDE (partial differential equation) is often credited to Merton as well (for example as BMS PDE), since Merton developed the same result independently at more or

the value of the total assets of the firm and satisfies the following equation (derived under a risk-neutral framework)

$$E_0 = V_0 \Phi(d_1) - D e^{-rT} \Phi(d_2) \quad (2)$$

where

$$d_1 = \frac{\ln(V_0 / D) + (r + \sigma_V^2 / 2)T}{\sigma_V \sqrt{T}} \quad d_2 = d_1 - \sigma_V \sqrt{T} \quad (3)$$

and Φ is the cumulative standard 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 probability, we require V_0 and σ_V , both not directly observable. However, if the company is publicly traded, we can observe E_0 . This means that equation (2) provides one condition that must be satisfied by V_0 and σ_V . From a result in stochastic calculus known as Ito's lemma, we can also estimate σ_E , using the following relationship

$$\sigma_E E_0 = \frac{\partial E}{\partial V} \sigma_V V_0. \quad (4)$$

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

$$\sigma_E E_0 = \Phi(d_1) \sigma_V V_0. \quad (5)$$

This provides another equation that must be satisfied by V_0 and σ_V . Equations (2) and (5) provide a pair of simultaneous equations that can be solved for V_0 and σ_V .

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 between the different types of debt; constant and flat term structure of interest rates, etc.)

less the same time, see Merton (1973). The extension for credit risk was developed later. However, even Black and Scholes (1973, p. 649) suggested to regard company's stock as an option on its assets.

and many extensions of this framework have been proposed in order to overcome one or more of these unrealistic assumptions. For reviews covering structural models, see, for example, Lando (2009), Bielecki and Rutkowski (2002), Uhrig-Homburg (2002), or Section 2.1. One of the most important drawbacks is the assumption that the value of the company can be described by a log-normal distribution. It is well known that log-returns of equities are not normally 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, for example, Mandelbrot (1963a, 1963b, 1967), Fama (1963, 1965a, 1965b), or Rachev and Mitnik (2000).

Chapter 2 of the thesis focuses on overcoming this assumption and contributes to the literature by introducing a structural credit risk model based on the stable Paretian distributions as a representative of subordinated models. The idea of using subordinated stable Paretian processes goes back to the seminal work of Mandelbrot and Taylor (1967) and stable laws have been applied in several financial sectors (see Rachev (2003) and Rachev and Mitnik (2000)). For these reasons, the stable Paretian law is the first candidate used as a subordinated model investigating credit risk modeling. In our study, we discuss how to use the Hurst et al. (1999) option pricing model, which is based on the stable Paretian distributions, in the framework of structural credit risk models. In fact, we show that it is possible to use this subordinated model in the Merton framework and, as for the Merton model, we propose two different methodologies for the parameter estimation: the first is to generalise the maximum likelihood parameter estimation proposed by Duan (1994); the second is a generalisation of the Moody's KMV methodology. Finally, we provide an empirical comparison of the Moody's KMV methodology applied to the Merton model and our subordinated one.

1.2 Scoring systems (credit-scoring models)

In the first paper, we deal with structural credit risk models and apply this framework to 24 U.S. companies with strong capitalisation 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 is that structural credit risk models are typically not feasible for the estimation of PDs of financial institutions due to their different balance sheet structure. This is why we devote our second study to estimation of PDs of financial institutions using credit-scoring models.

This class of statistical models is among the models most widely used 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 1960s by Beaver (1966) and Altman (1968). Since then an impressive body of research concerning this topic has evolved and we will address these studies in Section 3.2 in more detail. The major methodologies used for credit scoring include logit and probit models, or discriminant analysis models.

Discriminant analysis (DA) 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 principle is to maximise the difference between two groups, while the differences among particular members of the same group are minimised.

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 (2009). This is due to the statistical decision rule of maximising 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:

$$Z = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_n X_n, \quad (6)$$

where Z is the discriminant score (Z -score), γ_0 is a constant, $\gamma_i (i = 1, 2, \dots, n)$ the discriminant coefficients, and $X_i (i = 1, 2, \dots, n)$ the independent variables, that is, the financial ratios.

Logit and probit models can be considered to be among the most popular approaches in the empirical default-prediction literature, see, for example, Ohlson (1980), Zmijewski (1984), or Shumway (2001). Unlike the linear probabilistic model, where the outcome variable may be above 100% or below 0%, transformations used in logit and 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 a 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 the company failed, and $Y = 0$, otherwise. We are interested in modeling the probability Y by specifying the following model:

$$Y = f\left(\alpha + \sum_{i=1}^n \beta_i X_i\right) \quad (7)$$

where $X_i (i = 1, 2, \dots, n)$ are the explanatory variables, α is a constant, and the β_i denote the estimated coefficients for the explanatory variables 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

$$Y = \frac{\exp\left(\alpha + \sum_{i=1}^n \beta_i X_i\right)}{1 + \exp\left(\alpha + \sum_{i=1}^n \beta_i X_i\right)} = \frac{1}{1 + \exp\left(-\alpha - \sum_{i=1}^n \beta_i X_i\right)}. \quad (8)$$

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

$$Y = \int_{-\infty}^{\alpha + \sum_{i=1}^n \beta_i X_i} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt. \quad (9)$$

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

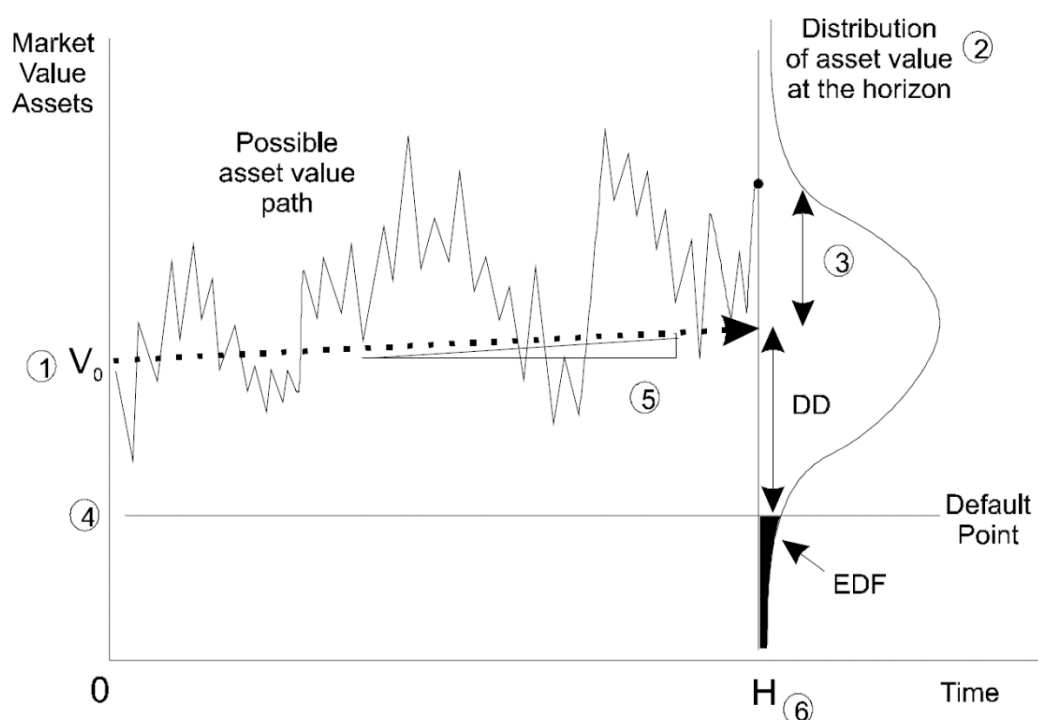
Chapter 3 of the thesis focuses on examination of the performance of static and multi-period logit and probit models for determining PDs of financial institutions. We contribute to the literature on credit-scoring models in the following way. First, unlike many other studies that focus on estimating rating models for the corporate sector, we provide a study where rating models for financial institutions are derived and investigated. Our sample contains more than 7,000 U.S. commercial banks with 4 default events between the years 2007 to 2013, which is substantially more than previous studies that estimated similar models for banks, see, for example, Canbas et al. (2005) or Kolari et al. (2002). 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. Finally, we provide a variety of methods for evaluating the performance of the considered models. Following Stein (2007), arguably a key study in this regard, we apply the techniques such as the walk-forward approach with out-of-time validation, ROC (relative or receiver operating characteristic) curve analysis, calibration accuracy tests and bootstrapping of ROC curve areas. We also suggest a number of additional performance evaluation techniques that have not yet been applied in the literature on scoring models. In particular, we suggest the use of non-parametric 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. These tests do not require the assumption of normality, which would not be satisfied in our case. Further, as an extension of log-likelihoods calculated within calibration accuracy test suggested in Stein (2007), we apply the Vuong's closeness test for non-nested models (Vuong (1989) 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 (Hosmer Jr and Lemeshow (2004)) to examine the overall fit of the estimated models.

1.3 Distress risk within an asset-pricing framework

While the first two studies are devoted to the estimation of PDs using two different approaches (structural models in Chapter 2 and credit-scoring models in Chapter 3), in the last research paper, we take advantage of the Moody's KMV³ database and use its structural-based default probability indicators (Expected Default Frequencies – EDFs) in an asset-pricing framework applied to the renewable energy sector in the U.S.

The Moody's KMV model (officially called *PortfolioManager*TM) is based on Merton's insight that debt behaves like a European short put option on the value of the firm's assets. With such a perspective, default occurs when the value of a firm's assets falls below the value of the firm's debt (or other fixed claims). There are six variables that determine the

Figure 1
The Moody's-KMV model



Source: Crosbie and Bohn (2003), page 13

³ KMV was founded by Kealhofer, McQuown, and Vasicek and it was an independent company until 2002 when it was acquired by Moody's Analytics and became Moody's KMV. This company provides various products for monitoring credit risk. For example, *Credit Monitor* integrates with *RiskAnalyst*TM, Moody's KMV *Financial Analyst*[®], and Moody's KMV *RiskCalc*[®] to support the analysis of private and public companies EDF credit measures. In addition, it also provides data for the analysis of credit portfolio risk within Moody's KMV *Portfolio Manager*TM.

default probability of a firm over some horizon, from now until time H (see Figure 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 .

Moody's KMV derive the so-called Expected Default Frequency (EDF) for a firm based on its capital structure, the asset return value and its volatility using the framework of Merton (1974). 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 framework. Second, the 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 PD is then the proportion of the firms of a given ranking of distance-to-default which have historically defaulted over a risk horizon, usually chosen to be one year. The EDFs can then also be used as an indicator of the creditworthiness of the issuing firms.

As mentioned above, in Chapter 4 of the thesis we use the EDF measure as a proxy for distress risk in an asset-pricing framework applied to U.S. renewable energy companies. The renewable energy sector has exhibited significant growth rates in the global economy during the last decade. At the same time, similar to the high-tech sector or venture capital, investments in renewable energy companies can be considered as being relatively risky; due to the nature of their business, there is often a significant gap between innovation, adoption, and a phase where the company really becomes established in the market (referred to as the "Valley of Death"). Therefore, investors who buy shares of renewable energy companies, that is, stocks of firms with potentially higher volatility and probabilities of distress or default, may expect higher average returns for bearing this risk. Our study aims to shed light on this important question, and thoroughly examines the relationship between distress risk and returns in the renewable sector.

For other sectors, there has been controversy with regards to the relationship between distress risk and (expected) returns in equity markets arising from several prominent studies. While Vassalou and Xing (2004) found a positive cross-sectional relationship, other key studies, for example, Campbell et al. (2008) or Garlappi et al. (2008), suggest that distress risk is priced negatively; that is, stocks with higher distress risk typically earn lower average returns. In the literature on financial economics, this controversial relationship is often referred to as the “distress risk puzzle”. Recently, there has also been a growing interest in examining returns of renewable energy companies, as well as identifying potential drivers of these returns. See, for example, Henriques and Sadorsky (2008), Kumar et al. (2012), Sadorsky (2012a), Bohl et al. (2013) and Inchauspe et al. (2015). These studies typically focus on the relationship between renewable energy stocks, changes in the oil price, other equity indices and carbon prices. The authors report evidence for the impact of several of these variables on renewable energy stock prices and suggest that, in particular, returns of high technology and renewable energy stocks seem to be highly correlated. However, none of these studies has examined how distress risk is priced in the renewable energy sector.

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. In particular, we provide the first empirical study that investigates 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 (1993) and Carhart (1997) risk factors, distress risk is priced in the renewable energy sector.

1.4 Structure of the thesis

This PhD thesis consists of three research papers which can be assigned to the research areas discussed above in the following way:

- Structural models
 - *Structural Credit Risk Models with Subordinated Processes*, co-authored with Sergio Ortobelli and Rosella Giacometti
- Scoring systems (credit-scoring models)
 - *Prediction of U.S. Commercial Bank Failures via Scoring Models: The FFIEC Database Case*, co-authored with Egon Kalotay and Stefan Trück

- Distress risk within an asset-pricing framework
 - *Distress Risk and Stock Returns of U.S. Renewable Energy Companies*, co-authored with Stefan Trück.

Bibliography

- ALTMAN, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23, 589-609.
- BEAVER, W. H. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71-111.
- BIELECKI, T. R. & RUTKOWSKI, M. 2002. *Credit Risk: Modeling, Valuation and Hedging*, Berlin, Germany, Springer Science & Business Media.
- BISHOP, C. M. 1995. *Neural Networks for Pattern Recognition*, New York, USA, Oxford University Press.
- BLACK, F. & SCHOLES, M. 1973. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81, 637-654.
- BOHL, M. T., KAUFMANN, P. & STEPHAN, P. M. 2013. From hero to zero: Evidence of performance reversal and speculative bubbles in german renewable energy stocks. *Energy Economics*, 37, 40-51.
- CAMPBELL, J. Y., HILSCHER, J. & SZILAGYI, J. 2008. In search of distress risk. *The Journal of Finance*, 63, 2899-2939.
- CANBAS, S., CABUK, A. & KILIC, S. B. 2005. Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The Turkish case. *European Journal of Operational Research*, 166, 528-546.
- CARHART, M. M. 1997. On persistence in mutual fund performance. *The Journal of Finance*, 52, 57-82.
- CROSBIE, P. & BOHN, J. 2003. Modeling default risk. Moody's KMV.
http://www.defaultrisk.com/pp_model_35.htm [April 11, 2016].
- DUAN, J. C. 1994. Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, 4, 155-167.
- DUFFIE, D. & SINGLETON, K. J. 1999. Modeling term structures of defaultable bonds. *Review of Financial studies*, 12, 687-720.
- FAMA, E. F. 1963. Mandelbrot and the stable Paretian hypothesis. *The Journal of Business*, 36, 420-429.
- FAMA, E. F. 1965a. The behavior of stock-market prices. *The Journal of Business*, 38, 34-105.
- FAMA, E. F. 1965b. Portfolio analysis in a stable Paretian market. *Management Science*, 11, 404-419.
- FAMA, E. F. & FRENCH, K. R. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.
- GARLAPPI, L., SHU, T. & YAN, H. 2008. Default risk, shareholder advantage, and stock returns. *Review of Financial Studies*, 21, 2743-2778.
- HENRIQUES, I. & SADORSKY, P. 2008. Oil prices and the stock prices of alternative energy companies. *Energy Economics*, 30, 998-1010.

- HOSMER JR, D. W. & LEMESHOW, S. 2004. *Applied Logistic Regression*, John Wiley & Sons.
- HULL, J. 2012. *Risk Management and Financial Institutions*, Hoboken, USA, John Wiley & Sons.
- HURST, S. R., PLATEN, E. & RACHEV, S. T. 1999. Option pricing for a logstable asset price model. *Mathematical and Computer Modelling*, 29, 105-119.
- INCHAUSPE, J., RIPPLE, R. D. & TRÜCK, S. 2015. The dynamics of returns on renewable energy companies: A state-space approach. *Energy Economics*, 48, 325-335.
- KOLARI, J., GLENNON, D., SHIN, H. & CAPUTO, M. 2002. Predicting large US commercial bank failures. *Journal of Economics and Business*, 54, 361-387.
- KUMAR, S., MANAGI, S. & MATSUDA, A. 2012. Stock prices of clean energy firms, oil and carbon markets: A vector autoregressive analysis. *Energy Economics*, 34, 215-226.
- LANDO, D. 2009. *Credit Risk Modeling: Theory and Applications*, Princeton, UK, Princeton University Press.
- MANDELBROT, B. 1963a. New Methods in Statistical Economics. *Journal of Political Economy*, 71, 421-440.
- MANDELBROT, B. 1963b. The variation of Certain Speculative Prices. *Journal of Business*, 36, 394-419.
- MANDELBROT, B. 1967. The variation of some other speculative prices. *The Journal of Business*, 40, 393-413.
- MANDELBROT, B. & TAYLOR, H. M. 1967. On the distribution of stock price differences. *Operations Research*, 15, 1057-1062.
- MERTON, R. C. 1973. Theory of rational option pricing. *The Bell Journal of economics and management science*, 4, 141-183.
- MERTON, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29, 449-470.
- MITCHELL, M. 1998. *An Introduction to Genetic Algorithms*, London, UK, MIT Press.
- OHLSON, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18, 109-131.
- RACHEV, S. T. 2003. *Handbook of Heavy Tailed Distributions in Finance*, Amsterdam, The Netherlands, Elsevier.
- RACHEV, S. T. & MITTNIK, S. 2000. *Stable Paretian models in finance*, Chichester, UK, John Wiley & Sons.
- SADORSKY, P. 2012. Correlations and volatility spillovers between oil prices and the stock prices of clean energy and technology companies. *Energy Economics*, 34, 248-255.
- SHUMWAY, T. 2001. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74, 101-124.
- SIRONI, A. & RESTI, A. 2007. *Risk Management and Shareholders' Value in Banking: From Risk Measurement Models to Capital Allocation Policies*, Chichester, UK, John Wiley & Sons.
- STEIN, R. M. 2007. Benchmarking default prediction models: Pitfalls and remedies in model validation. *Journal of Risk Model Validation*, 1, 77-113.
- TRÜCK, S. & RACHEV, S. T. 2009. *Rating Based Modeling of Credit Risk: Theory and Application of Migration Matrices*, Norwell, USA, Academic Press.
- UHRIG-HOMBURG, M. 2002. Valuation of Defaultable Claims - A Survey. *Schmalenbach Business Review*, 54, 24-57.

- VASSALOU, M. & XING, Y. 2004. Default risk in equity returns. *The Journal of Finance*, 59, 831-868.
- VUONG, Q. H. 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57, 307-333.
- ZMIJEWSKI, M. E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59-82.

2 Structural Credit Risk Models with Subordinated Processes

*Martin Gurny (contribution: 65%), Sergio Ortobelli (contribution: 30%),
Rosella Giacometti (contribution: 5%)*

This paper was presented at:

- The 9th International Scientific Conference – Financial Management of Firms and Financial Institutions, Ostrava, Czech Republic, September 2013;
- The AFAS/ECON PhD Workshop 2013, Sydney, Australia, September 2013.

This paper has been published in the *Journal of Applied Mathematics 2013* (2013), p.12.

Abstract

In this paper, structural models based on the Merton framework are discussed. First, we observe that the classical distributional assumption of the Merton model is 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 option pricing model based on the stable Paretian distributions which generalises the standard Merton 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 optimise the performance for the stable Paretian model. Our results suggest that probability of default (PD) is generally underestimated by the Merton model, and that the stable Lévy model is substantially more sensitive to the periods of financial crises.

Keywords: Credit risk, Probability of default (PD), Merton model, Stable Paretian distributions, Subordinated processes

2.1 Introduction

Estimating a borrower's risk level, namely the probability of default (PD), by assigning an appropriate PD, is a widely employed strategy by many financial institutions as well as the supervisory authorities. PD indicates a probability that a given counterparty will not be able to meet its obligations. The incorrect estimation of PD leads to, among other things, unreasonable ratings and incorrect pricing of financial instruments, and thereby it had been one of the causes of the recent global financial crisis (GFC). Undervaluation of the risk caused the collapse of the financial system which had been over-extended through credit derivatives on the global markets. PD is also a crucial parameter used in the calculation of economic and regulatory capital, under the Basel II and Basel III Accords for banking institutions. These reasons highlight how important the estimation of PD is and why it has been a significant research topic for a long time.

The probability of default, as one of the key risk parameters in the IRB approach⁴, has many methodologies for its estimation. In general, we can classify the existing methodologies into three groups: structural models, reduced-form models, and credit-scoring (statistical) models. We discuss the first type of models in this paper. The structural approach to credit risk modeling was proposed by Robert Merton (1974) in his seminal paper on the valuation of corporate debt. Largely as a logical extension of the Black and Scholes (1973) option pricing framework, he introduced a model for assessing the credit risk of a company by characterising 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 between the different types of debt; constant and flat term structure of interest rates, etc.). Notwithstanding this fact, 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

⁴ The Internal Rating Based (IRB) Approach is linked to capital requirements under Basel II Accord. Under this 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.

density in the tails, and that the normal distribution with a comparatively thinner tail simply cannot describe this phenomenon (see, for example, Mandelbrot (1963a, 1963b, 1967), Fama (1963, 1965a, 1965b), or Rachev and Mitnik (2000)).

The main contribution of this paper is twofold. First, we introduce a structural credit risk model based on the stable Paretian⁵ distributions as a representative of subordinated models. Second, we show that it is possible to use this model in the Merton 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 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 et al. (1999) option pricing model based on the stable Paretian distributions which generalises 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 generalisation of the Merton one. The first relevant desirable property of the stable distributional assumption is that stable distributions have a domain of attraction. The generalised central limit theorem for the normalised 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.

⁵ Throughout this paper we use the 'stable Paretian' term interchangeably with 'stable non-Gaussian', 'stable Lévy', and ' α -stable' terms.

The idea of using subordinated processes goes back to the studies by Bochner (1955) and Clark (1973) who used it for investigation of evolution of stock prices. Subordinated stable Paretian processes then have been introduced in the seminal work of Mandelbrot and Taylor (1967). Stable laws have been applied in several financial sectors (see Rachev (2003) and Rachev and Mittnik (2000)). For these reasons, the stable Paretian law is the first candidate as a subordinated model for investigating credit risk modeling, and in this paper we discuss how to use the Hurst et al. (1999) 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 generalise the maximum likelihood parameter estimation proposed by Duan (1994); the second is a generalisation of the Moody's KMV methodology.

This paper is organised as follows. In Section 2.2, we first provide literature review on structural credit risk models. In Section 2.3, 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 2.4 is devoted to the parameters estimation for both the Merton and the subordinated models. We characterise empirical data and make a comparison between the obtained results in Section 2.5. Finally, we provide a brief summary in Section 2.6.

2.2 Literature review

The first generation structural credit risk models are based on the Merton (1974) model.⁶ The basic idea behind this model is to treat a company's equity and debt as a contingent claim written on the company's asset value. If the asset value is lower than the amount of debt interest and principal, it is rational for the company to default on this debt. 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

⁶ It is worth to mention here an important issue: the separation of disciplines. While in finance it seems that Merton (1974) is treated as a revelation, the accounting identity and interpretation of the debt and equity holders having claims on the company's assets is an accounting truism known for over a century, and used in bankruptcy research reported in accounting journals since the early 1970s. For example, work by Balcaen and Ooghe (2006) provides an excellent overview of the classic statistical methodologies in bankruptcy prediction and their related problems.

between a risky bond and default-free bond. In addition to Merton (1974), models by Black and Cox (1976), Geske (1977), and Vasicek (1984) might be classified in the first generation structural credit risk models. These models try to improve the original Merton framework by relaxing one or more of the unrealistic assumptions (Laajimi (2012)).

Black and Cox (1976) 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 value of the security from the financing of interest and dividend payments. 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 (1977) modifies the original Merton framework by allowing the risky bond to have discrete interest payments. Although Black and Cox (1976) look at a similar problem, in their case, the interest payments are continuous in time and they state that in general, there is no closed form solution when the interest payments are discrete in time. However, Geske (1977) 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 the study by Geske (1977), Vasicek (1984) 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 its 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 also on that of its total maturing debt and higher priority debt. Moreover, Vasicek (1984) 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 assumptions, and specify a stochastic process for the evolution of the short-term rates (Laajimi (2012)). 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 et al. (1993) and Longstaff and Schwartz (1995).

Kim et al. (1993) 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 any time 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 feature to examine its effect on the yield spreads between corporate and Treasury bonds.

Longstaff and Schwartz (1995) then modify the first generation models in three directions: (i) default can arise any time between the issuance and the maturity of the bonds; (ii) interest rates are not flat. That is, there exists interest rate risk; (iii) strict absolute priority is violated. In contrast to Kim et al. (1993), 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 structures. 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 firms with the 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 framework. Ju et al. (2005) 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 et al. (2001) also consider a dynamic capital structure by modeling a mean-reverting leverage ratio and stochastic interest rate. Acharya and Carpenter (2002) develop a model with both stochastic interest rate and endogenous defaults. The interest rate is modeled as a 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 et al. (2007) distinguish between bank and public debt. They assume that renegotiation through private workout is only possible for bank debt. This renegotiation possibility makes bank debt more attractive, but limits banks' debt capacity to strong firms, that is, firms with high bargaining power. Bourgeon and Dionne (2007) extend the Hackbarth et al. (2007) 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 (2007) suppose that holders of private debt, for example, 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 (1994), Childs et al. (2005), and Sundaresan and Wang (2007), who consider endogenous investment. The cash holding management policy is accounted for in Acharya et al. (2006), Anderson and Carverhill (2007), and Asvanunt et al. (2011). Sarkar and Zapatero (2003) consider mean reverting cash flows. Zhou (2001), Duffie and Lando (2001) and Giesecke and Goldberg (2004) add a jump component to the value process of assets allowing for "surprise" default at the cost of closed-form solution. Alternatively, Hackbarth et al. (2006) consider jumps in the cash flow process with regime change. Finally, Longstaff (1995), Morellec (2001), and Ericsson and Renault (2006) include a liquidity premium to price corporate debt, while Duffie and Lando (2001) consider accounting information uncertainty.

It is also important to mention that the Merton framework is the underlying idea behind a commercial model originally developed by KMV which is nowadays referred to as the Moody's KMV model (see Footnote 3 for a more thorough discussion of the founders of KMV and its acquisition by Moody's Analytics). Distance-to-default, which is the normalised 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 (2012)). Sobehart et al. (2000) and Stein (2002), 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 (2002) find opposite results, namely that the Moody's KMV model captures all the information contained in agency ratings migration and accounting ratios. This is not really surprising as Kealhofer is one of the founders of this model and logically supports its usage. Crosbie and Bohn (2003) find that combining market prices and financial statements gives more effective default measurement. The authors empirically test the predictive power of EDFs derived from the KMV methodology versus credit ratings, and show that the model based on EDFs yields a better power curve, i.e. better forecasting results.

The accuracy of default forecasting of the Moody's KMV model is studied in Bharath and Shumway (2008). 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 et al. (2004) and Du and Suo (2007) compare the Moody's KMV model to other models and conclude that it does not provide adequate predictive power. However, Duffie et al. (2007) discover a significant predictive strength over time within the Moody's KMV model.

Despite the fact that there have been many various studies which tried to improve the original Merton work by relaxing one or more of the unrealistic assumptions (valuation of contingent claims in continuous time; the risky bond having continuous interest payments; the company can default only at debt's maturity time T but not before; the model is not able to distinguish between the different types of debt; constant and flat term structure of interest rates, etc.), none of these studies really looked at the assumption of the underlying distribution for the company value – i.e. the log-normal distribution. However, it is well known

that log-returns of equities are not normally distributed, and several empirical investigations have shown that log-returns of equities exhibit skewness and excess kurtosis which leads to a greater density in the tails.

2.3 Merton and subordinated credit risk models

The core concept of the Merton (1974) model is to treat a company's equity and debt as a contingent claim written on the 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 time t (D_t) with a face value of L maturing at time T ⁷. The exercise price of a European 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:

$$A_t = E_t + D_t. \quad (1)$$

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

$$E_T = \max[A_T - L, 0]. \quad (2)$$

2.3.1 The Merton-Black-Scholes distributional assumptions

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

$$dA_t = \mu A_t dt + \sigma A_t dW_t, \quad (3)$$

where μ is the expected return (drift coefficient), σ 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 (3) as follows:

⁷ Generally, in a credit risk model 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).

$$A_T = A_t \exp \left[\left(\mu - \frac{1}{2} \sigma^2 \right) (T-t) + \sigma \sqrt{(T-t)} W_t \right], \quad (4)$$

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

In accordance with the Black and Scholes (1973) option pricing theory (see Footnote 2 for closer discussion on relation between Black and Scholes and Merton work), the Merton model stipulates that the company's equity value satisfies the following equation for pricing the European call option within a risk-neutral framework:

$$E_t = A_t \Phi(d_1) - L e^{-r(T-t)} \Phi(d_2), \quad (5)$$

where

$$d_1 = \frac{\ln\left(\frac{A_t}{L}\right) + \left(r + \frac{1}{2} \sigma^2\right)(T-t)}{\sigma \sqrt{(T-t)}}, \quad (6)$$

$$d_2 = d_1 - \sigma \sqrt{(T-t)}, \quad (7)$$

r is the risk-free interest rate⁸ and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal variable. Equation (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 (4) as follows:

$$\begin{aligned} \text{PD}_t &= \text{P} [A_T \leq L] \\ &= \text{P} \left[\ln(A_t) + \left(\mu - \frac{1}{2} \sigma^2 \right) (T-t) + \sigma \sqrt{(T-t)} W_t \leq \ln(L) \right] \\ &= \int_{-\infty}^{\frac{\ln\left(\frac{A_t}{L}\right) + \left(\mu - \frac{1}{2} \sigma^2\right)(T-t)}{\sigma \sqrt{(T-t)}}} \phi(x) dx, \end{aligned} \quad (8)$$

⁸ 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 ϕ is the probability density function of a standard normal variable. Note that unlike (8), (5) is not a function of μ , 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 μ , 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

$$\begin{aligned} \text{PD}_t &= \Phi(-\hat{d}_2) \\ &= \Phi\left(-\frac{\ln\left(\frac{A_t}{L}\right) + \left(\mu - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{(T-t)}}\right). \end{aligned} \quad (9)$$

Further discussion on this topic can be found in Delianedis and Geske (2003) who showed that risk-neutral PDs can serve as an upper bound to objective PDs.

2.3.2 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 et al. (1999) and Rachev and Mitnik (2000). 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 (1959)). Typically, subordinated models with random intrinsic time are leptokurtic with heavier tails compared to the normal distribution. Feller (1966) 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:

$$A(t) = A(t_0) \exp \left\{ \int_{t_0}^t \mu(s) ds + \int_{t_0}^t \rho(s) dT(s) + \int_{t_0}^t \sigma(s) dW(T(s)) \right\}, \quad (10)$$

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-minimising strategy in incomplete markets.⁹ The stable subordinated model in Hurst et al. (1999) uses the unique continuous martingale that makes sense in a discrete setting, but a priori it is not derived from a risk-minimising strategy even if the markets are incomplete (see Rachev and Mitnik (2000)). Following the same notation used in the Merton framework, the value of a European call option at time t (the value of a company's equity) with exercise price L (face value of a zero coupon debt instrument) and time to maturity t ¹⁰, is given by

$$E_t = A(t_0) F_+ \left(\ln \left(\frac{A(t_0)}{L_{r,t_0,t}} \right) \right) - L_{r,t_0,t} F_- \left(\ln \left(\frac{A(t_0)}{L_{r,t_0,t}} \right) \right), \quad (11)$$

where

$$F_{\pm}(x) = \int_0^{+\infty} \Phi \left(\frac{x \pm \frac{1}{2}y}{\sqrt{y}} \right) dF_Y(y), \quad (12)$$

⁹ In incomplete markets, there exist non-redundant claims carrying an intrinsic risk. In order to evaluate a contingent claim, a risk-minimising strategy is often applied (see Hofmann et al. (1992), Follmer and Sondermann (1986), and Follmer and Schweizer (1993)).

¹⁰ Here, we change the notation of maturity time from T (used in the Merton 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 Mitnik (2000)); that is,

$$\begin{aligned} F_{\pm}(x) &= \int_0^{+\infty} \Phi\left(\frac{x \pm \frac{1}{2}\lambda y}{\sqrt{\lambda y}}\right) f_Y(y) dy \\ &= \int_0^1 \Phi\left(\frac{x \pm \frac{1}{2}\lambda u(1-u)^{-3}}{\sqrt{\lambda u(1-u)^{-3}}}\right) f_Y\left(u(1-u)^{-3}\right) \frac{1+2u}{(1-u)^4} du. \end{aligned} \quad (13)$$

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

$$\text{delta} = \Delta_E = \frac{\partial E_t}{\partial A} = F_+ \left(\ln \left(\frac{A(t_0)}{L_{r,t_0,t}} \right) \right). \quad (14)$$

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

$$\begin{aligned}
\text{PD}_t &= F_+ \left(\ln \left(\frac{L_{r,t_0,t}}{A(t_0)} \right) \right) \\
&= \int_0^{+\infty} \Phi \left(\frac{\ln \left(\frac{L_{r,t_0,t}}{A(t_0)} \right) + \frac{1}{2} y}{\sqrt{y}} \right) dF_Y(y). \tag{15}
\end{aligned}$$

Recall that under the risk-neutral measure the stationary increment $Z(t+s) - Z(t)$ has mean $\mu_{Z,s} = 0$ and variance $\sigma_{Z,s}^2 = \mu_{T,s} \sigma^2$, where σ and $\mu_{T,s}$ are, respectively, the volatility and the mean of the increment of the stationary process T when they exist (see Hurst et al. (1999)). The skewness coefficient of this increment is zero (models are symmetric around the zero mean). Kurtosis of the subordinated models is defined as $k_{Z,s} = 3 \frac{1 + \sigma_{T,s}^2}{\mu_{T,s}}$ for all $s \geq 0$ (where $\sigma_{T,s}^2$ 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. Therefore, the model we consider in the following presents heavier tails and higher peaks around the origin compared to a normal distribution.

2.3.3 The Mandelbrot-Taylor distributional assumptions

Mandelbrot (1963a, 1963b, 1967) and Mandelbrot and Taylor (1967) have proposed the stable Paretian distribution to estimate the log-returns. An α -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 (1994) for further details on stable distributions. Mandelbrot and Taylor (1967) supposed that the intrinsic time process T has stationary independent increments as follows:

$$T(t+s) - T(t) \stackrel{d}{=} S_{\alpha/2}(cs^{2/\alpha}, 1, 0), \tag{16}$$

for all $s, t \geq 0$, $\alpha \in (0, 2)$, 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 $Z(t) = \ln(A_{th})$, is a symmetric α -stable Lévy motion with stationary independent increments as follows:

$$Z(t+s) - Z(t) = \ln(A_{th} / A_{(t-s)h}) \stackrel{d}{=} S_\alpha(\nu s^{1/\alpha}, 0, 0), \quad (17)$$

for all $s, t > 0$, where

$$\nu = \frac{\sigma \sqrt{c}}{\sqrt{2} \left(\cos\left(\frac{\pi\alpha}{4}\right) \right)^{1/\alpha}}. \quad (18)$$

If we consider the constant scale parameter σ , then the random variable Y in (11) is as follows:

$$Y = \sigma^2 (T(t) - T(t_0)) = \lambda V, \quad (19)$$

where $\lambda = c\sigma^2(t - t_0)^{2/\alpha}$ and $V = S_{\alpha/2}(1, 1, 0)$. Hence, with

$$c = 2 \left(\cos\left(\frac{\pi\alpha}{4}\right) \right)^{2/\alpha}, \quad (20)$$

it follows that $Z(t) \stackrel{d}{=} S_\alpha(\sigma t^{1/\alpha}, 0, 0)$. Thus, we can estimate the index of stability α and the scale parameter σ using the maximum likelihood method (see Rachev and Mitnik (2000) 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 $F_\pm(x)$:

$$F_\pm(x) = \int_0^1 \Phi \left(\frac{x \pm \frac{1}{2} \lambda u (1-u)^{-3}}{\sqrt{\lambda u (1-u)^{-3}}} \right) f_V \left(u (1-u)^{-3} \right) \frac{1+2u}{(1-u)^4} du. \quad (21)$$

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

$$\text{PD}_t = \int_0^1 \Phi \left(\frac{\ln \left(\frac{L_{r,t_0,t}}{A(t_0)} \right) + \frac{1}{2} \lambda u (1-u)^{-3}}{\sqrt{\lambda u (1-u)^{-3}}} \right) f_V \left(u (1-u)^{-3} \right) \frac{1+2u}{(1-u)^4} du . \quad (22).$$

2.4 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 μ , and the asset volatility σ — 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 hypotheses of the subordinated model could require the estimation of several different parameters. For example, in the α -stable Lévy process, once the index of stability α is estimated, the scale parameter σ 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.

2.4.1 Parameter estimates for the Merton model

The unknown parameters of the Merton model come from (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 (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 et al. (2003), Crosbie and Bohn (2003), Bruche (2005), or Ericsson and Reneby (2005)), which finds two unknown parameters (A_t and σ) by solving the system of two equations as follows:

$$\begin{aligned} E_t &= A_t \Phi(d_1) - L e^{-r(T-t)} \Phi(d_2) \\ \sigma_E &= \frac{A_t}{E_t} \Phi(d_1) \sigma \end{aligned} \quad , \quad (23)$$

where σ_E is the standard deviation of the equity log returns $\ln(E_{th} / E_{(t-1)h})$. Nevertheless, this method does not estimate asset drift μ ; it determines the risk-neutral probability of default using the risk-free asset r . As a consequence, Jovan (2010) 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 (1994) and enhanced later by Duan et al. (2004). The time series of daily market value of equity E_t is equal to n days, where $t = (0, \dots, n)$. In Duan et al. (2004) 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 μ and σ of model (3), where $th = (0, \dots, nh)$, is defined on the basis of observed values of E_t as follows:

$$\begin{aligned} L(\hat{\theta}; \hat{A}_{th} | E_{th}) = & -\frac{n}{2} \ln(2\pi\hat{\sigma}^2 h) \\ & - \frac{1}{2} \sum_{t=1}^n \frac{(\hat{R}_t - (\hat{\mu} - (1/2)\hat{\sigma}^2)h)^2}{\hat{\sigma}^2 h} \\ & - \sum_{t=1}^n \ln(\hat{A}_{th}) - \sum_{t=1}^n \ln(\Phi(d_1)), \end{aligned} \quad (24)$$

where

$$\hat{R}_t = \ln \left(\frac{\hat{A}_{th}}{\hat{A}_{(t-1)h}} \right), \quad (25)$$

and where $\hat{\theta} \equiv (\hat{\mu}, \hat{\sigma})$ and \hat{A}_{th} is estimated from (5). To launch the iteration process we could insert as initial values the values obtained by solving the system (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 $1 \leq (T - th) \leq T$. We

maximise (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 (24) and the new market value of assets obtained from (5) for maximising (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 $|\hat{\mu}^{(i+1)} - \hat{\mu}^{(i)}| + |\hat{\sigma}^{(i+1)} - \hat{\sigma}^{(i)}| \leq \varepsilon$ for a given small ε).

Duan et al. (2004) 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 et al. (2004), Duffie et al. (2007), Crosbie and Bohn (2003), or Vassalou and Xing (2004)). 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 maximising the log-likelihood function, we have explicit formulas.

The first step is exactly the same, calculation of the daily value of $\hat{A}_{th}^{(i)}$, $th = (0, \dots, nh)$ from (5). As the initial values can be used again the estimates obtained by solving the system (23). Then, the arithmetic mean of the sample is given by

$$\bar{R}^{(i)} = \frac{1}{n} \sum_{t=1}^n \hat{R}_t^{(i)}, \quad (26)$$

where \hat{R}_t is defined in (25). Another step is the calculation of estimates of the asset volatility $\hat{\sigma}$ and the drift $\hat{\mu}$ of model (3), which are defined as follows:

$$\begin{aligned} \hat{\sigma}^{(i+1)} &= \sqrt{\frac{1}{nh} \sum_{t=1}^n \left(\hat{R}_t^{(i)} - \bar{R}^{(i)} \right)^2}, \\ \hat{\mu}^{(i+1)} &= \bar{R}^{(i)} \frac{1}{h} + \frac{1}{2} \hat{\sigma}^{2(i+1)}. \end{aligned} \quad (27)$$

Since this is again an iterative procedure, we use the new estimates obtained from (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 mentioning that the Merton model with parameters estimated according to the methodology described above differs from the one actually employed by Moody's KMV. The Merton model performance relies substantially 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 them. Indeed, the founders of KMV, Oldrich Vasicek and Stephen Kealhofer, developed a so-called Vasicek-Kealhofer (VK) model (see Arora et al. (2005)) 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 the 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.

2.4.2 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 (24). In fact, (24) can be derived from a more general formula that can be used for the derivation of log-likelihood functions for any subordinated model. This formula is defined in the following way:

$$L(\hat{\theta}; \hat{A}_{th} | E_{th}) = \sum_{t=1}^n \ln \left(f_Z(\hat{R}_t) \right) - \sum_{t=1}^n \ln \left(\hat{A}_{th} \right) - \sum_{t=1}^n \ln \left(\Delta_E \right), \quad (28)$$

where $\hat{\theta}$ represents the set of the parameters in the density function $f_Z(\hat{R}_t)$ of the stationary increment $\ln \left(A_{th} / A_{(t-1)h} \right) = Z(t+1) - Z(t)$, \hat{A}_{th} is estimated from (11), \hat{R}_t is defined in (25), and Δ_E is given by (14). The initial values $\hat{A}_{th}^{(1)}$ of the iteration process could be the ones obtained by solving the system (23). The procedure continues iteratively till the

distance $\|\hat{\theta}^{(i+1)} - \hat{\theta}^{(i)}\|$ 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. Second, it is often very problematic to implement this methodology since many subordinated models do not have close form equations 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 (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 $\|\hat{\theta}^{(i+1)} - \hat{\theta}^{(i)}\|$ is sufficiently small. In particular, for the α -stable Lévy model, we first suggest determination of the index of stability α . Second, the unique parameter that must be estimated is the scale parameter σ , 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 (23). Moreover, as for the Merton model (see Duan et al. (2004)), 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 $\hat{R}_t^{(i)}$.

2.5 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 ones. We use Moody's KMV and the extended Moody's KMV methodology described in Section 2.4.1 and 2.4.2, respectively, while 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 the 13-week Treasury bill for the 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 capitalisation in the U.S. market.¹¹ Our data spans the period from January 3, 2000, to December 30, 2011. For the market value of equity, we used the consolidated market value of a company which is defined as a share price multiplied by the number of ordinary shares on 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.¹² While the short-term debt and current portion of the long-term debt represent that portion of the debt payable within one year, including the 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.

2.5.1 Analysis of the distributional assumptions of the company value log-returns

The Merton model's 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 3,157 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

¹¹ 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. While the set of liquidly traded U.S. firms may be much larger, in our analysis we decided to analyse 24 companies from the Dow Jones Industrial Average index as these companies presumably should have very low PD. Our analysis aims to investigate whether there will be substantial differences in estimated PDs obtained from the two models even for these companies. The applied sample of 24 companies from various industries is considered to be sufficient for this task.

¹² 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 (a one-year time horizon is assumed 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 service 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 (2008), Vassalou and Xing (2004), or Duffie et al. (2007)).

every day on the last 250 daily company values (1 year of daily values). Table 1 reports the 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 $E((X - E(X))^3)/E((X - E(X))^2)^{1.5}$, the kurtosis $E((X - E(X))^4)/E((X - E(X))^2)^2$, the percentage of rejection of the Gaussian hypothesis using the Jarque-Bera (JB) test at the 5% significance level (see Jarque and Bera (1987)), the stable index of stability “alpha”, the stable index of skewness “beta”, the stable scale parameter “sigma”, the stable location parameter “mu”, and the percentage of rejection of the stable Paretian hypothesis using the Kolmogorov-Smirnov (KS) test¹³ at the 5% significance level.

The results reported in Table 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 rejected only for four companies of the considered sample.

Table 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.

¹³Note that as an alternative, one could also apply an Anderson-Darling test which places more weight on observations in the tail of the distribution.

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 generalise the Gaussian one.

2.5.2 PD estimates from the Merton model

We applied Moody's KMV methodology¹⁴ 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 1 and Table 2. In Table 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 1 describes the evolution of the PDs on a monthly basis. These probabilities are almost null during the whole decade. However, we can distinguish three periods of increased

¹⁴ We perform our analysis using MATLAB.

¹⁵ For example, Byström (2006) 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.

PDs for some companies from our sample. First, at the beginning of the century after the 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 characterised by severely contracted liquidity in the global credit markets and insolvency threats to investment banks and other institutions. Beginning with the bankruptcy

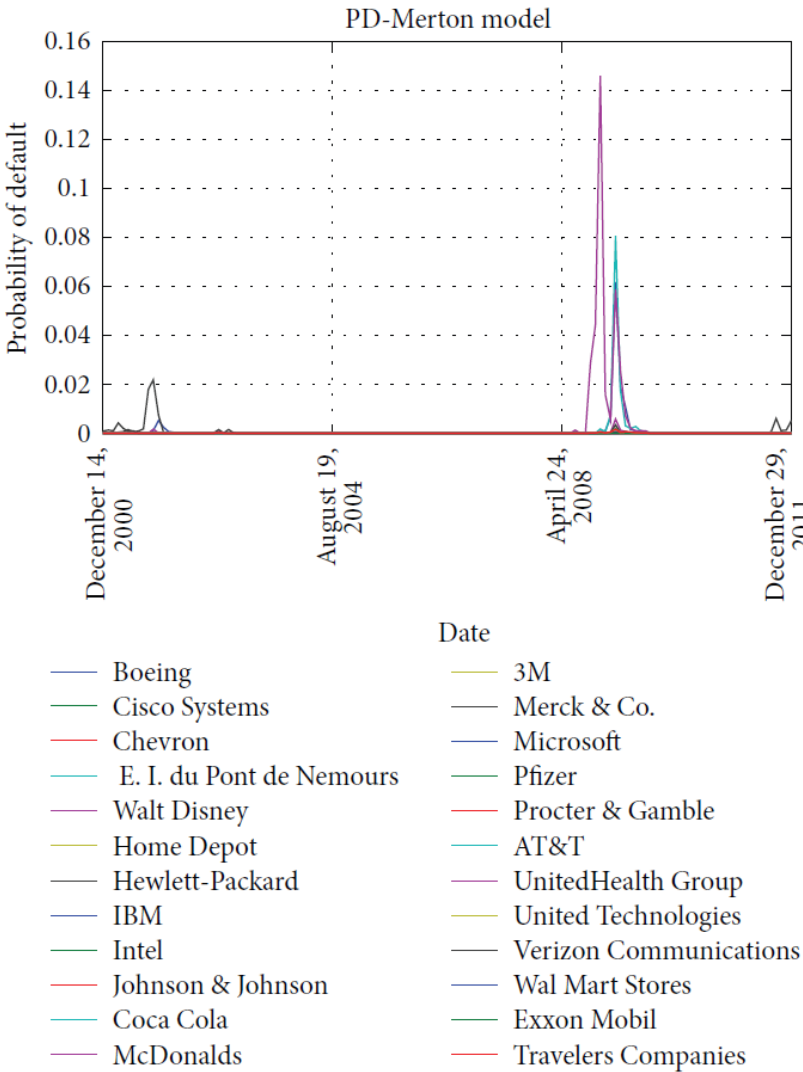
Table 2
Outcomes from the Merton model

Company	Average ratio (L/A)	Average PD	Average DD
(1) Boeing	0.1308	0.000830	8.9020
(2) Cisco Systems	0.0262	0.000000	20.6010
(3) Chevron	0.0606	0.000000	13.8524
(4) E. I. du Pont de Nemours	0.1169	0.000845	9.9706
(5) Walt Disney	0.1305	0.000083	8.5109
(6) Home Depot	0.0599	0.000002	11.8297
(7) Hewlett-Packard	0.0914	0.000511	8.3242
(8) IBM	0.1032	0.000000	11.4799
(9) Intel	0.0098	0.000000	14.2761
(10) Johnson & Johnson	0.0330	0.000000	22.8226
(11) Coca Cola	0.0614	0.000000	17.5142
(12) McDonalds	0.1012	0.000015	12.2037
(13) 3M	0.0494	0.000000	14.9342
(14) Merck & Co.	0.0610	0.000037	11.1672
(15) Microsoft	0.0068	0.000000	21.4008
(16) Pfizer	0.0813	0.000019	11.0915
(17) Procter & Gamble	0.1010	0.000000	13.9819
(18) AT&T	0.1607	0.000013	8.4346
(19) UnitedHealth Group	0.0925	0.002424	10.2912
(20) United Technologies	0.0798	0.000001	12.1045
(21) Verizon Communications	0.2106	0.000106	8.8750
(22) Wal Mart Stores	0.0955	0.000000	12.4895
(23) Exxon Mobil	0.0207	0.000000	18.0516
(24) Travelers Companies	0.1291	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. Considered time period is from December 2000 to December 2011.

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 European governments to rescue distressed financial institutions. Among the companies from our sample that were most affected were 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 was the world's third largest chemical company based on market capitalisation 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

Figure 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.

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 by October 2009.

2.5.3 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, 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 3. To evaluate the stable parameters and the distributions of subordinator f_V in (21), we perform a maximum likelihood estimator that uses the fast Fourier transform

Table 3

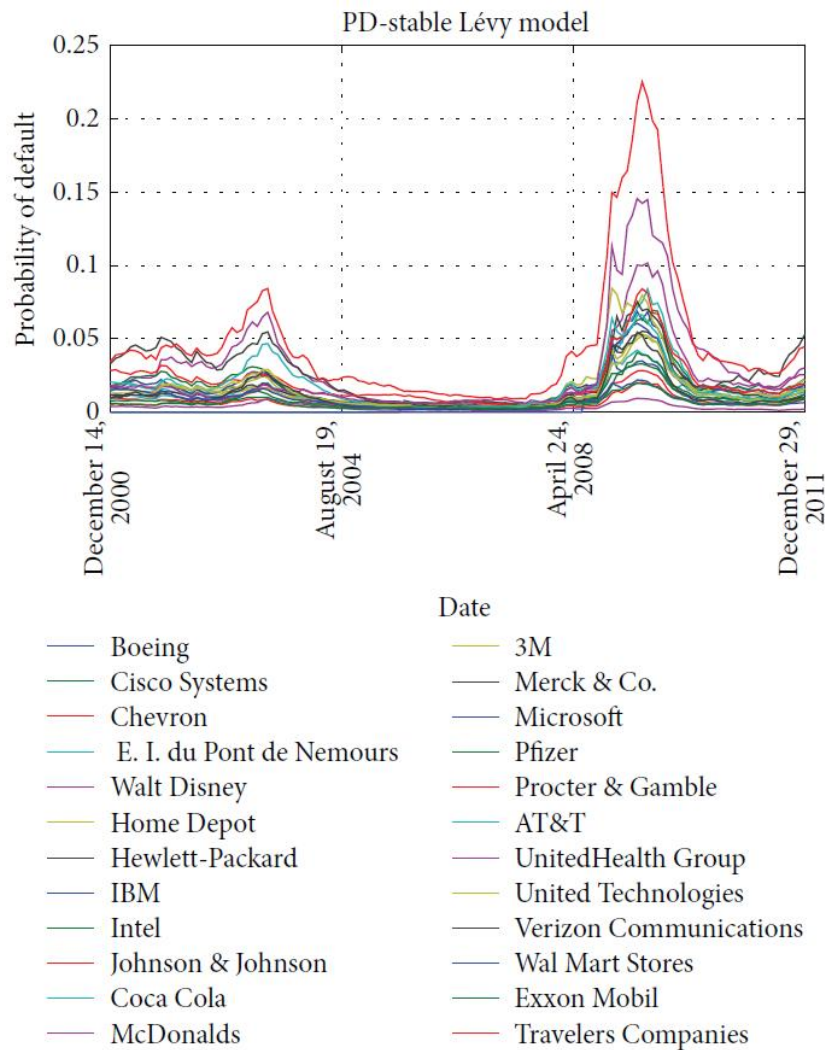
Outcomes from the stable Lévy model

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. Considered time period is from December 2000 to December 2011.

Figure 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.

(see Rachev and Mitnik (2000) or Nolan (1997)). 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 scale and location stable parameters. However, this would require 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 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 then describes the evolution of PDs on a monthly basis.

2.5.4 Comparison of the Merton and stable Lévy models

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 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.

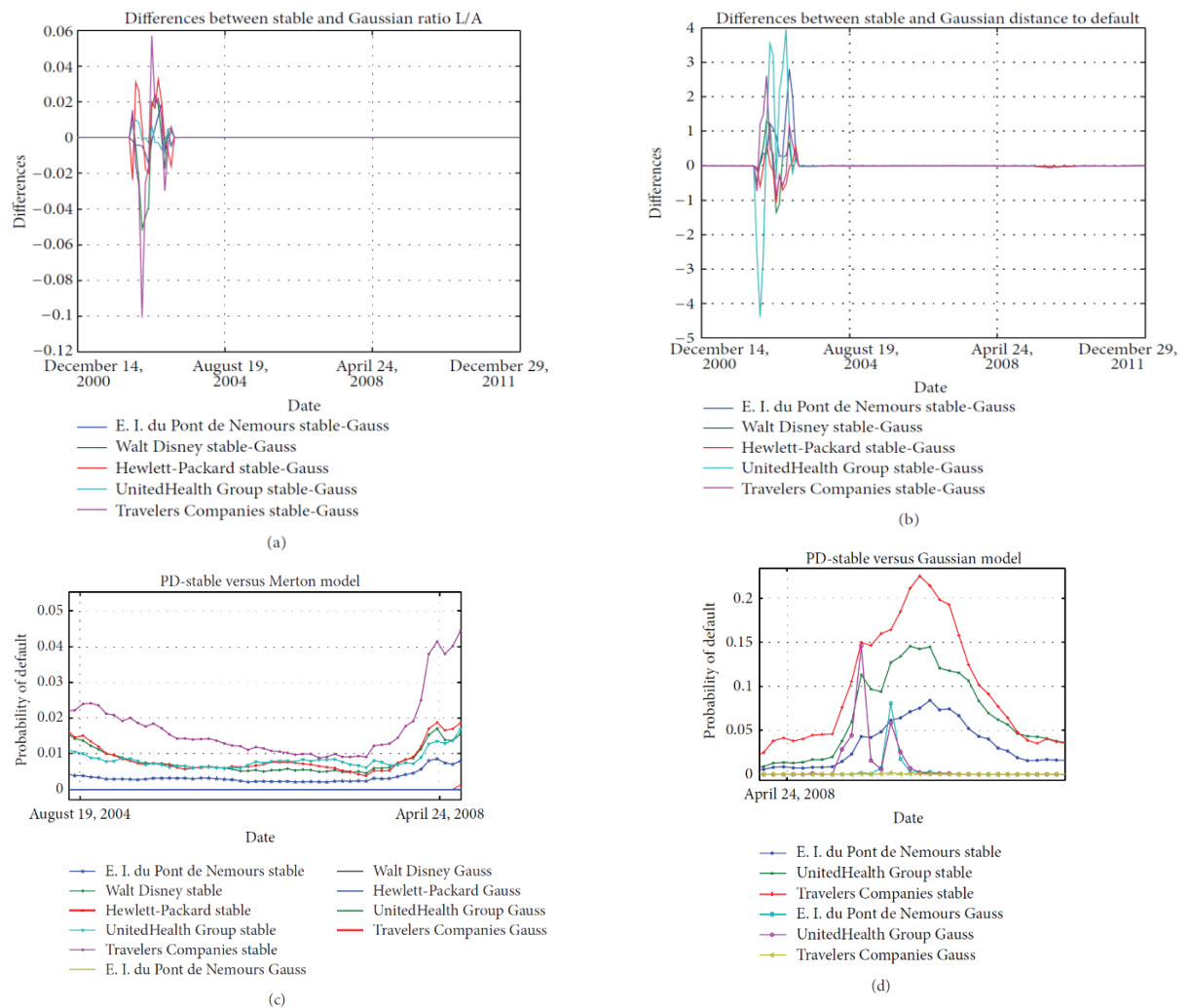
Figure 3 reports the main differences between the two models for those companies that present the highest peaks in default probabilities (E. I. du Pont de Nemours, Walt Disney, Hewlett-Packard, UnitedHealth Group, and Travelers Companies). In particular, Figures 3(a) and 3(b) show that the main differences in the ratio between the debt and the company value, and the debt and 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 3(c) and 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 the 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.

In particular, 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 higher than those obtained by the Merton model. This is not a real surprise, because 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 emphasised in Figure 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.

Figure 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.

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.

2.6 Conclusions

The structural approach to credit risk modeling, initially developed by Merton (1974), has been widely used over recent decades. The basic idea behind this framework is to treat a company's equity and debt as a contingent claim written on that company's asset value. However, the classical version of this model requires a number of simplifying and unrealistic assumptions. In this paper, we focus on overcoming the assumption that company value follows the log-normal distribution. In fact, we prove that this assumption is rejected, and consequently the log-returns of the companies' asset values are not Gaussian distributed. For this reason, we propose an 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 framework. In fact, we propose an empirical comparison of the Moody's KMV methodology applied to the Merton model and our subordinated one. In particular, we suggest alternative parameter estimation for subordinated processes, and optimise the performance for the stable Lévy model.

The results of our conducted empirical analysis suggest that the probability of default is generally underestimated by the Merton model. Nevertheless, it is worth to note that there are some limitations in this study. Some studies have found that structural models may not be very accurate for measuring default probabilities and fail to provide appropriate calibration and backtesting results. This critique relates to both models presented and we will discuss other approaches in subsequent chapters of this thesis. The resulting PD estimates for both models are significantly lower than the ones reported in the literature, see, e.g., Bharath and Shumway (2008). This is caused by the sample selection. While these authors examined

all firms in the intersection of the Compustat and CRSP databases (excluding financial firms), we analyse 24 firms from the Dow Jones Industrial Average index as these companies presumably should have very low PDs. Our analysis aims to examine whether there are substantial differences in the estimated PDs obtained from the two models even for these companies. Also, the computation of the debt ratios may be further scrutinised in a robustness check as the resulting debt ratios are low and central to the analyses. All these issues could be the content of possible extensions in future research.

Clearly, our results should be further discussed and compared to other distributional models in a future research program. As a matter of fact, two alternative structural credit risk models, based on well-known Lévy processes (the Variance Gamma (VG) process and the Normal Inverse Gaussian (NIG) process), were proposed by Brambilla et al. (2015). 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 et al. (2013)), the authors demonstrate that both models are able to capture the situation of instability that affects each company in the 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.

Bibliography

- ACHARYA, V. V. & CARPENTER, J. N. 2002. Corporate bond valuation and hedging with stochastic interest rates and endogenous bankruptcy. *Review of Financial Studies*, 15, 1355-1383.
- ACHARYA, V. V., HUANG, J. Z., SUBRAHMANYAM, M. & SUNDARAM, R. K. 2006. When does strategic debt-service matter? *Economic Theory*, 29, 363-378.
- ANDERSON, R. W. & CARVERHILL, A. P. 2007. Liquidity and capital structure. Working paper No. 6044. Center for Economic Policy Research.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=997024 [April 11, 2016].

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.

- ARORA, N., BOHN, J. R. & ZHU, F. 2005. Reduced form vs. structural models of credit risk: A case study of three models. *Journal of Investment Management*, 3, 43.
- ASVANUNT, A., BROADIE, M. & SUNDARESAN, S. 2011. Managing Corporate Liquidity: Strategies and Pricing Implications. *International Journal of Theoretical and Applied Finance*, 14, 369-406.
- BALCAEN, S. & OOGHE, H. 2006. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38, 63-93.
- BHARATH, S. T. & SHUMWAY, T. 2008. Forecasting default with the Merton distance to default model. *Review of Financial Studies*, 21, 1339-1369.
- BLACK, F. & COX, J. C. 1976. Valuing corporate securities: Some effects of bond indenture provisions. *The Journal of Finance*, 31, 351-367.
- BLACK, F. & SCHOLES, M. 1973. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81, 637-654.
- BLUHM, C., OVERBECK, L. & WAGNER, C. 2003. *An Introduction to Credit Risk Modeling*, Boca Raton, USA, CRC Press.
- BOCHNER, S. 1955. *Harmonic Analysis and the Theory of Probability*, University of California Press, Berkeley. CA.
- BOURGEON, J.-M. & DIONNE, G. 2007. On debt service and renegotiation when debt-holders are more strategic. Working paper. CIRPEE.
http://www.cirpee.org/fileadmin/documents/Cahiers_2007/CIRPEE07-29.pdf [April 11, 2016].
- BRAMBILLA, C., GURNY, M. & ORTOBELLI LOZZA, S. 2015. Structural Credit Risk Models with Lévy Processes: The VG and NIG Cases. *Far East Journal of Mathematical Sciences*, 97, 101-119.
- BRUCHE, M. 2005. Estimating Structural Bond Pricing Models via Simulated Maximum Likelihood. Working paper. London School of Economics.
<http://eprints.lse.ac.uk/24647/1/dp534.pdf> [April 11, 2016].
- BYSTRÖM, H. N. 2006. Merton unraveled: A flexible way of modeling default risk. *The Journal of Alternative Investments*, 8, 39-47.
- CAREY, M. & 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. <https://www.mnb.hu/letoltes/gordy.pdf> [April 11, 2016].
- CLARK, P. K. 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: Journal of the Econometric Society*, 135-155.
- CHILDS, P. D., MAUER, D. C. & OTT, S. H. 2005. Interactions of corporate financing and investment decisions: The effects of agency conflicts. *Journal of Financial Economics*, 76, 667-690.
- COLLIN-DUFRESNE, P., GOLDSTEIN, R. S. & MARTIN, J. S. 2001. The determinants of credit spread changes. *The Journal of Finance*, 56, 2177-2207.
- CROSBIE, P. & BOHN, J. 2003. Modeling default risk. Moody's KMV.
http://www.defaultrisk.com/pp_model_35.htm [April 11, 2016].

- DELIANEDIS, G. & GESKE, R. L. 2003. Credit risk and risk neutral default probabilities: Information about rating migrations and defaults. Working paper. EFA 2003. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=424301 [April 11, 2016].
- DU, Y. & SUO, W. 2007. Assessing credit quality from equity markets: Is a structural approach a better approach? *Canadian Journal of Administrative Sciences*, 24, 212-218.
- DUAN, J. C. 1994. Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, 4, 155-167.
- DUAN, J. C., GAUTHIER, G. & SIMONATO, J. G. 2004. On the equivalence of the KMV and maximum likelihood methods for structural credit risk models. Working paper. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.121.2998> [April 11, 2016].
- DUFFIE, D. & LANDO, D. 2001. Term structures of credit spreads with incomplete accounting information. *Econometrica*, 69, 633-664.
- DUFFIE, D., SAITA, L. & WANG, K. 2007. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83, 635-665.
- ERICSSON, J. & RENAULT, O. 2006. Liquidity and credit risk. *The Journal of Finance*, 61, 2219-2250.
- ERICSSON, J. & RENEBY, J. 2005. Estimating Structural Bond Pricing Models. *Journal of Business*, 78, 707-736.
- FAMA, E. F. 1963. Mandelbrot and the stable Paretian hypothesis. *The Journal of Business*, 36, 420-429.
- FAMA, E. F. 1965a. The behavior of stock-market prices. *The Journal of Business*, 38, 34-105.
- FAMA, E. F. 1965b. Portfolio analysis in a stable Paretian market. *Management Science*, 11, 404-419.
- FELLER, W. 1966. *An Introduction to Probability Theory and Its Applications II*, New York, USA, John Wiley & Sons.
- FOLLMER, H. & SCHWEIZER, M. 1993. A microeconomic approach to diffusion models for stock prices. *Mathematical Finance*, 3, 1-23.
- FOLLMER, H. & SONDERMANN, D. 1986. *Contributions to Mathematical Economics*, Amsterdam, The Netherlands, North Holland.
- GESKE, R. 1977. The valuation of corporate liabilities as compound options. *Journal of Financial and Quantitative Analysis*, 12, 541-552.
- GIESECKE, K. & GOLDBERG, L. R. 2004. Sequential defaults and incomplete information. *Journal of Risk*, 7, 1-26.
- GURNY, M., ORTOBELLI LOZZA, S. & GIACOMETTI, R. 2013. Structural Credit Risk Models with Subordinated Processes. *Journal of Applied Mathematics*, 2013, 1-12.
- HACKBARTH, D., HENNESSY, C. A. & LELAND, H. E. 2007. Can the trade-off theory explain debt structure? *Review of Financial Studies*, 20, 1389-1428.
- HACKBARTH, D., MIAO, J. & MORELLEC, E. 2006. Capital structure, credit risk, and macroeconomic conditions. *Journal of Financial Economics*, 82, 519-550.
- HILLEGEIST, S. A., KEATING, E. K., CRAM, D. P. & LUNDSTEDT, K. G. 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9, 5-34.
- HOFMANN, N., PLATEN, E. & SCHWEIZER, M. 1992. Option pricing under incompleteness and stochastic volatility. *Mathematical Finance*, 2, 153-187.

- HURST, S. R., PLATEN, E. & RACHEV, S. T. 1999. Option pricing for a logstable asset price model. *Mathematical and Computer Modelling*, 29, 105-119.
- JARQUE, C. M. & BERA, A. K. 1987. A test for normality of observations and regression residuals. *International Statistical Review*, 55, 163-172.
- JOVAN, M. 2010. The Merton structural model and IRB compliance. *Metodološki Zvezki*, 7, 39-57.
- JU, N., PARRINO, R., POTESHMAN, A. M. & WEISBACH, M. S. 2005. Horses and rabbits? Trade-off theory and optimal capital structure. *Journal of Financial and Quantitative Analysis*, 40, 259-281.
- KEALHOFER, S. & KURBAT, M. 2002. The Default Prediction Power of the Merton Approach, relative to Debt Ratings and Accounting Variables. KMV LLC.
<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=6936BCA6636D5133B152D3A5E1BB14CC?doi=10.1.1.101.7279&rep=rep1&type=pdf> [April 11, 2016].
- KIM, I. J., RAMASWAMY, K. & SUNDARESAN, S. 1993. Does default risk in coupons affect the valuation of corporate bonds?: A contingent claims model. *Financial Management*, 22, 117-131.
- LAAJIMI, S. 2012. Structural Credit Risk Models: A Review. *Insurance and Risk Management*, 80, 53-93.
- LONGSTAFF, F. A. 1995. How much can marketability affect security values? *The Journal of Finance*, 50, 1767-1774.
- LONGSTAFF, F. A. & SCHWARTZ, E. S. 1995. A simple approach to valuing risky fixed and floating rate debt. *The Journal of Finance*, 50, 789-819.
- MANDELBROT, B. 1963a. New Methods in Statistical Economics. *Journal of Political Economy*, 71, 421-440.
- MANDELBROT, B. 1963b. The variation of Certain Speculative Prices. *Journal of Business*, 36, 394-419.
- MANDELBROT, B. 1967. The variation of some other speculative prices. *The Journal of Business*, 40, 393-413.
- MANDELBROT, B. & TAYLOR, H. M. 1967. On the distribution of stock price differences. *Operations Research*, 15, 1057-1062.
- MAUER, D. C. & TRIANTIS, A. J. 1994. Interactions of corporate financing and investment decisions: A dynamic framework. *The Journal of Finance*, 49, 1253-1277.
- MERTON, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29, 449-470.
- MORELLEC, E. 2001. Asset liquidity, capital structure, and secured debt. *Journal of Financial Economics*, 61, 173-206.
- NOLAN, J. P. 1997. Numerical calculation of stable densities and distribution functions. *Communications in Statistics - Stochastic models*, 13, 759-774.
- OSBORNE, M. M. 1959. Brownian motion in the stock market. *Operations Research*, 7, 145-173.
- RACHEV, S. T. 2003. *Handbook of Heavy Tailed Distributions in Finance*, Amsterdam, The Netherlands, Elsevier.
- RACHEV, S. T. & MITTNIK, S. 2000. *Stable Paretian models in finance*, Chichester, UK, John Wiley & Sons.

- SAMORODNITSKY, G. & TAQQU, M. 1994. *Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*, New York, USA, Chapman & Hall.
- SARKAR, S. & ZAPATERO, F. 2003. The Trade-off Model with Mean Reverting Earnings: Theory and Empirical Tests. *The Economic Journal*, 113, 834-860.
- SOBEHART, J. R., KEENAN, S. C. & STEIN, R. M. 2000. Benchmarking Quantitative Default Risk Models: A Validation Methodology. Moody's Investors Service, Global Credit Research. <http://www.rogermstein.com/wp-content/uploads/53621.pdf> [April 11, 2016].
- STEIN, R. M. 2002. Benchmarking default prediction models: Pitfalls and remedies in model validation. Moody's KMV, Technical Report. https://riskcalc.moodyrms.com/us/research/crm/validation_tech_report_020305.pdf [April 11, 2016].
- SUNDARESAN, S. & WANG, N. 2007. Investment under uncertainty with strategic debt service. *The American Economic Review*, 97, 256-261.
- VASICEK, O. A. 1984. Credit valuation. KMV Corporation. http://www.defaultrisk.com/pp_model241.htm [April 11, 2016].
- VASSALOU, M. & XING, Y. 2004. Default risk in equity returns. *The Journal of Finance*, 59, 831-868.
- ZHOU, C. 2001. The term structure of credit spreads with jump risk. *Journal of Banking & Finance*, 25, 2015-2040.

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

*Martin Gurny (contribution: 70%), Egon Kalotay (contribution: 15%),
Stefan Trück (contribution: 15%)*

This paper was presented at:

- The 5th International Conference of the Financial Engineering and Banking Society (FEBS), Nantes, France, June 2015;
- The Faculty Research Seminar – VSB Technical University of Ostrava, Czech Republic, June 2015;
- The 22nd Annual Conference of the Multinational Finance Society (MFS), Halkidiki, Greece, July 2015.

Abstract

We examine the performance of static and multi-period credit-scoring models for determining probabilities of default (PDs) of financial institutions. We use an extensive data base for the U.S. provided by the Federal Financial Institutions Examination Council (FFIEC). 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. We use an extensive sample of more than 7,000 U.S. commercial banks that also contains 405 default events, using the unique Federal Financial Institutions Examination Council (FFIEC) database. 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. We 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 has been mentioned in the literature, the proposed models also perform very well in comparison to results on scoring techniques for the corporate sector.

Keywords: Credit-scoring models, Bank failure, Probability of default (PD), The FFIEC database, Discrete hazard model, Model evaluation techniques

3.1 Introduction

Credit risk and estimation of default probabilities (PDs) have become some of the most intensely studied topics in the financial literature and have undergone extensive development in recent 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 has increased substantially 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. 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.

In this paper, 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 paper 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.¹⁷ 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, 2013.¹⁸ 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 et al. (2005) work with 40 privately owned Turkish commercial banks and 21 defaults, Kolari et al. (2002) use over 1,000 large U.S. commercial banks (they define large banks to be greater than \$US250 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 the remaining 37 large failed banks). In comparison, our sample contains more than 7,000 U.S. commercial banks, with up to 405 banks defaulted during the period 2007-2013.¹⁹

We use a framework called the walk-forward approach, see, for example, Stein (2007), with out-of-time validation. The approach allows us, in each period, to use the maximum

¹⁷ One of the reasons why banking is usually excluded from bankruptcy analysis with scoring models and financial ratios, are the various regulatory changes and constraints in the banking sector and their impacts on the values of balance sheet items. See, e.g., KPMG (2014), for a general overview on imposed regulatory changes in the banking industry. Further, studies such as Cohen (2013), Lee and Chih (2013) illustrate the impact regulatory changes and constraints may have on reported balance sheet items and financial ratios of financial institutions. Note, however, that given the relatively high explanatory and predictive power of the applied scoring models in our study, we suggest that financial ratios can still successfully be applied to default prediction also in the banking sector.

¹⁸ <http://www.fdic.gov/bank/individual/failed/banklist.html>

¹⁹ While there were 492 defaults in considered period, our sample contains only 405 defaulted banks. As it will be explained later, Savings & Loan Associations are not included in the FFIEC database.

number of available data points to fit and test the models, such that we are not restricted to dividing our sample into estimation and holdout samples. This procedure works as follows: First, we select a year t . Second, we fit the model using all the data available on or before the selected year. Third, once the model's form and parameters are established for the selected time period, we generate the model outputs for all of the banks available during the following year $t + 1$. Fourth, we save the prediction as part of a result set. Fifth, we move the window one step forward (to $t + 1$) so that all of the data through that year can be used for fitting, and the data for the next year can be used for testing. Sixth, we repeat these steps adding the new predictions to the result set for every year.

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 (2012), 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 that distinguish between the performance of the models, and if they do, they often seem satisfied with a comparison to the Z -score proposed by Altman (1968) or the O -score suggested by Ohlson (1980). 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 to be Stein (2007), 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 (2007), including the walk-forward approach with out-of-time validation, ROC (relative or receiver operating characteristic) 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 of 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 et al. (1988) 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 that is often violated, as pointed out by Engelmann et al. (2003). 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 (2007), we apply the Vuong closeness test for non-nested models (see Vuong (1989)) 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 (2004)) 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.

The majority of the estimated models build on variables that form a reasonable mixture of profitability, liquidity, assets quality and capital adequacy indicators.²⁰ We find that our models have a high default/non-default classification and predictive accuracy, with more than 95% of defaulted banks being captured within the banks with the highest 10% PDs. These figures can be compared to results from recent studies conducted for the corporate sector, for instance, Beaver et al. (2005) with 80.3% - 92%, or Duffie et al. (2007) with 94% of the year-ahead defaulters being captured in the lowest two deciles. 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 multiple comparison test to examine significant differences between the particular models in terms of bootstrapped ROC areas. Specifically, the Tukey 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 outperform

²⁰ Unfortunately, there is no theoretical justification why or why not a certain financial accounting variable (or ratio) should be included into a default model. 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, and observations on the predictive accuracy of the various profiles.

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. Finally, 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 so 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.

This paper is organised as follows. In Section 3.2, we provide a literature review on credit-scoring models. Section 3.3 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 3.4. Section 3.5 provides empirical results on model estimation and validation. Finally, Section 3.6 concludes and summarises the results.

3.2 Literature review

Although the techniques underlying credit-scoring models were devised in 1930s by authors such as Fisher (1936) and Durand (1941), the decisive boost to a development and spread of these models came in the 1960s with the studies by Beaver (1966) and Altman (1968). The latter study has been considered by many as the most significant 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 (1976) who, from the statistical point of view, contrasted discriminant analysis with logit models. Altman et al. (1977) investigated the predictive performance of a seven-variable discriminant analysis model ("Zeta model") which improved on Altman's (1968) earlier five-variable model. The

study by Santomero and Vinso (1977) systematically developed probabilistic estimates of failure. Ohlson (1980) showed that the predictive power of any model depends on 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 (1986). Queen and Roll (1987) 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 (2001) who demonstrates that this model corrects for the period at risk and allows for time-varying covariates. Altman et al. (2010) developed a new “*Z* -Metrics” model for the *RiskMetrics Group* which is, in fact, an updated and improved version of the *Z* -score methodology. An excellent overview of corporate failure prediction models, the classic cross-sectional statistical methods and their related shortcomings is then provided by Balcaen and Ooghe (2006).

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 (1974), Korobow and Stuhr (1975), Sinkey (1975), and Korobow et al. (1976). 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 (1977) 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 1981 can be found in Altman et al. (1981). West (1985) 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 et al. (1983), Korobow and Stuhr (1985), Lane et al. (1986), Maddala (1986), Whalen and Thomson (1988), Espahbodi (1991), Thomson (1991), Kolari et al. (1996). This research has confirmed that scoring models perform well as EWSs. More recent work on this topic has been conducted by Logan (2001) who implemented a logit model to identify leading indicators of failure for U.K. small banks. His analysis focuses on the small banks' crisis of the early 1990s.

Another stream of studies tend to combine parametric and nonparametric approaches for the prediction of bank failures. Tam and Kiang (1992) implement a neural network approach to perform discriminant analysis. An integrated model approach for bankruptcy prediction has been introduced by Jo and Han (1996). 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 et al. (2000) identifies potentially failing banks using fuzzy clustering algorithms and self-organising neural networks. Kolari et al. (2002) 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 minimising Type I and II errors. A very similar approach is employed by Lanine and Vennet (2006) 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 (2002) combine several discriminant methods and perform simulation analysis to enhance the accuracy of classification results. Canbas et al. (2005) 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 artificial 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 et al. (2009) and Ioannidis et al. (2010).

Despite the fact that business failure prediction has become a major research domain within corporate finance over the last decades, regrettably, within the past 10 years, there are not that many studies which focus on financial sector. In fact, to the authors knowledge, there are only a couple of them - Canbas et al. (2005), Lanine and Vennet (2006), Boyacioglu et al. (2009), and Ioannidis et al. (2010). We could also mention here a study by Berger and Bouwman (2013) that does not deal directly with PD estimation, however, it analyses the default and survival risk of U.S. banks by examining how capital affects a bank's performance and how this effect varies across banking crises, market crises, and normal times that occurred in the U.S. over the past 25 years. This obvious lack of relevant studies, along with the recent GFC, encouraged us to apply our analysis on the FFIEC database.

In general, there is no overall agreement on 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 (1997)). However, more simple classification techniques, such as linear discriminant analysis and logistic regression, are generally considered to provide good results 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 et al. (2003)).

3.3 The FFIEC council & database

The Federal Financial Institutions Examination Council (FFIEC)²¹ 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:

- the Board of Governors of the Federal Reserve System (FRB),
- the Federal Deposit Insurance Corporation (FDIC),
- the National Credit Union Administration (NCUA),

²¹ <http://www.ffiec.gov/>

- 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 provides patronage for 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 page²², financial and structural information for most FDIC-insured institutions is available from March 31, 2001.

Table 1-3 provide some descriptive statistics on the banks included in the FFIEC

Table 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%.

²² <https://cdr.ffiec.gov/public/>

database. In particular, we report statistics as of December 31, 2013 on the institution type (Table 1) and the location of the banks (Table 2). We also report some descriptive statistics on the size of the banks (Table 3) as of December 31, 2006 until December 31, 2012.

In Table 1, non-member Banks (defined as commercial banks, state charters and Federal non-members, supervised by the FDIC) comprise 57% of the FFIEC database, followed by National Banks (defined as commercial banks, national/federal charters and Federal members, supervised by the OCC) 17%, State Member Banks (defined as commercial banks, state charters and Fed members, supervised by the FRB) 13%, Federal Savings Banks (defined as savings associations, state/federal charters, supervised by the OTS²³) 8%, and State Savings Banks (defined as savings banks, state charters, supervised by the FDIC) 6%.

With regards to location²⁴ of the banks in the FFIEC database, Table 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 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 \$US1,897 billion in total assets. The mean value of total assets among banks included in the FFIEC database has increased from \$US1.3 billion in 2006 to \$US2 billion in 2012. It is obvious that there is a domination of larger banks in this database. In other words, our sample is skewed which justifies the analysis that will be applied later on – segregation into subsamples according to the size.

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.

²³ 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.

²⁴ 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 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).

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

Table 3

Mean value and percentiles of total assets (in \$US000) for U.S. banks

Total Assets (in \$US000)						
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 \$US1,314 million in 2006 to \$US2,023 million in 2012.

3.4 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.

3.4.1 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, for example, Martin (1977), West (1985), Logan (2001), Shumway (2001). 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 $y_i = 1$ (with probability PD_i), if bank i fails, and $y_i = 0$ (with probability $1 - PD_i$) otherwise. We are interested in modeling the probability PD_i for the occurrence of a default event by specifying the following model:

$$PD_i = f\left(\alpha + \sum_{j=1}^n \beta_j x_{i,j}\right), \quad (1)$$

where $x_{i,j}$ denotes particular explanatory variable of i^{th} bank used to forecast the probability of default, and α, β_j are the estimated parameters of the model. The functional form of f then may assume specific distribution functions, depending on a given model.

The literature suggests various ways to specify the probability 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

$$PD_i = \frac{\exp\left(\alpha + \sum_{j=1}^n \beta_j x_{i,j}\right)}{1 + \exp\left(\alpha + \sum_{j=1}^n \beta_j x_{i,j}\right)} = \frac{1}{1 + \exp\left(-\alpha - \sum_{j=1}^n \beta_j x_{i,j}\right)} \quad (2)$$

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

$$PD_i = \int_{-\infty}^{\alpha + \sum_{j=1}^n \beta_j x_{i,j}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt. \quad (3)$$

Due to nonlinear nature of these models it is necessary to use maximum likelihood estimation in order to obtain the model parameters. Given PD_i and assuming that defaults are independent²⁵, we can express the maximum likelihood function as follows:

$$L = \prod_{i=1}^n PD_i^{y_i} (1 - PD_i)^{1 - y_i}. \quad (4)$$

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

²⁵ It is obvious that in our analysis we make the assumption of default independence, as this is widely used in similar default studies. Nevertheless, a possible question remains to what degree this assumption holds considering the GFC, the frequent mergers observed in the banking industry and their dependence on monetary and economic US policies. This question would be an interesting direction for future research, however, is beyond the scope of this study.

$$\ln L = \sum_{i=1}^n y_i \ln PD_i + \sum_{i=1}^n (1 - y_i) \ln (1 - PD_i). \quad (5)$$

Therefore, by combining (5) and (2) we get the logarithm of the maximum likelihood function for the logit model as follows:

$$\begin{aligned} \ln L = \sum_{i=1}^n y_i \ln \left[\frac{1}{1 + \exp \left(-\alpha - \sum_{j=1}^n \beta_j x_{i,j} \right)} \right] \\ + \sum_{i=1}^n (1 - y_i) \ln \left[1 - \frac{1}{1 + \exp \left(-\alpha - \sum_{j=1}^n \beta_j x_{i,j} \right)} \right], \end{aligned} \quad (6)$$

and by combining (5) and (3) we get the logarithm of the maximum likelihood function for the probit model as follows:

$$\begin{aligned} \ln L = \sum_{i=1}^n y_i \ln \left[\int_{-\infty}^{\alpha + \sum_{j=1}^n \beta_j x_{i,j}} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} t^2 \right) dt \right] \\ + \sum_{i=1}^n (1 - y_i) \ln \left[1 - \int_{-\infty}^{\alpha + \sum_{j=1}^n \beta_j x_{i,j}} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} t^2 \right) dt \right]. \end{aligned} \quad (7)$$

For model evaluation or testing the significance of a model, the log-likelihood ratio test or Wald test can be used; see, for example, Tabachnick and Fidell (2007) or Hosmer Jr and Lemeshow (2004). 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, for example, Trück and Rachev (2009)).

3.4.2 Static and discrete hazard models

Static (single-period) models, even though widely used in the past, may not be fully appropriate for estimation of default probabilities, see Shumway (2001) or Hillegeist et al. (2004). First, there are often multiple-period data sets available. Since static models consider only 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, for example, 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 practice is to use information on financial ratios one year prior to the default or non-default event that may actually introduce an unnecessary selection bias into the estimation process. Second, by ignoring the fact that banks and their performance change over time, static models produce inconsistent estimates of default probabilities, and test statistics may subsequently be biased and provide incorrect inferences.

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 over 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 over time.

The discrete hazard model estimates the PDs as:

$$PD_{i,t} = P(y_{i,t} = 1) = E(y_{i,t} | x_{i,t}), \quad (8)$$

where $PD_{i,t}$ 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:

$$PD_{i,t} = f \left[\alpha(t) + \sum_{j=1}^n \beta_j x_{i,t,j} \right], \quad (9)$$

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 (1) and (9): first, in model (9), the constant α is replaced by the baseline hazard rate $\alpha(t)$, while, second, 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.²⁶ 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.²⁷

3.4.3 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. In particular, we propose nonparametric techniques such as the Kruskal-Wallis test and Tukey's multiple comparison procedure (see Hochberg and Tamhane (1987)). We further propose the use of Vuong's closeness test (see Vuong (1989)) 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 (2004)) that allows for a comparison of the predicted and actually observed default frequencies for sub-groups of the entire sample.

²⁶ 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).

²⁷ For more detailed discussion on discrete-time hazard models and their econometric properties we refer to Shumway (2001).

3.4.3.1 ROC analysis

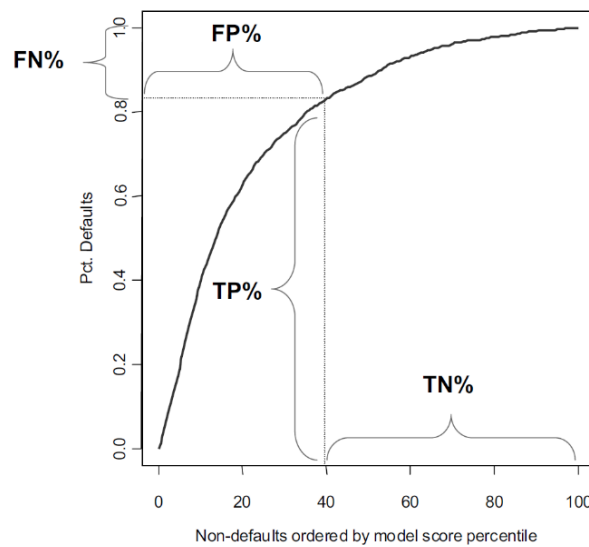
ROC (relative or receiver operating characteristic) curves, see, for example, Green and Swets (1966), Hanley (1989), Hosmer Jr and Lemeshow (2004), Stein (2007), are among the most powerful tools to quantify the predictive power of models and are widely used for evaluation of credit default models. A ROC curve plots the Type II error against one minus the Type I error. Unlike contingency 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 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 summarising 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 et al. (2003) provide the following identity relationship between the ROC area and the AR: $AR = 2(\text{ROC area} - 0.5)$.

Figure 1

An example of the ROC curve



Source: Stein (2007), page 82.

3.4.3.2 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, for example, bootstrapping (Efron and Tibshirani (1994)) which allows leveraging of the available data to reduce the dependence on the particular sample. As described in Stein (2007), 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 the use of a nonparametric Kruskal-Wallis test or Tukey's multiple 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 (1999)). 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, it is necessary 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. 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, and keep track of which observation and rank goes with which of the k groups. In the case of tied observations, average the ranks. Finally, 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:

$$K = \left[\frac{12}{n(n+1)} \sum_{j=1}^k \frac{T_j^2}{n_j} \right] - 3(n+1). \quad (10)$$

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 tells you only that at least two groups are different.

This drawback is overcome by the Tukey test (Hochberg and Tamhane (1987)). It is a multiple comparison procedure which allows further investigation of 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 an experiment-wide 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 (2007)). The formula for Tukey's statistic is:

$$q_s = \frac{Y_A - Y_B}{SE}, \quad (11)$$

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 studentised range distribution.

3.4.3.3 Calibration accuracy test and Vuong's closeness test

The second dimension within validating credit models (after examination of a model's 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, it is possible to determine which model's PDs (from a set of candidate models) are closest to actual PDs given a set of empirical data. Refer to Stein (2007) 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

$$prob(y|x) = p(x)^y [1 - p(x)]^{(1-y)}, \quad (12)$$

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:

$$L = \prod_{i=1}^n prob(y_i | x_i) = \prod_{i=1}^n p(x_i)^{y_i} (1 - p(x_i))^{(1-y_i)}. \quad (13)$$

Since, in general, it is more convenient to work with summations than products, by convention it is normal to work with the log of the likelihood ℓ , defined as:

$$\ln L = \ell = \sum_{i=1}^n y_i \ln [p(x_i)] + (1 - y_i) \ln [1 - p(x_i)]. \quad (14)$$

In order to determine whether calculated log-likelihoods for various models are significantly different, Vuong's closeness test for non-nested models may be used, see Vuong (1989). 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 α , if the Z statistic exceeds the positive (falls below negative) $(1 - \alpha)$ -quantile of the standard normal distribution. The Z statistic is defined as:

$$Z = \frac{LR_N(\beta_{ML,A}, \beta_{ML,B})}{\sqrt{N\omega_N}}, \quad (15)$$

where

$$LR_N(\beta_{ML,A}, \beta_{ML,B}) = L_N^A - L_N^B - \frac{K_A - K_B}{2} \log N. \quad (16)$$

Therefore, 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 , ω_N , is defined by setting ω_N^2 equal to either the mean of the squares of the pointwise log-likelihood ratios ℓ_i , or to the sample variance of these values, where

$$\ell_i = \log \frac{f_A(y_i | x_i, \beta_{ML,A})}{f_B(y_i | x_i, \beta_{ML,B})}. \quad (17)$$

3.4.3.4 Hosmer-Lemeshow test

Hosmer-Lemeshow's chi-squared goodness-of-fit test (Hosmer Jr and Lemeshow (2004)) 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 freedom²⁸, where O_i^{ND} is the

$$HL = \sum_{i=1}^k \left[\frac{(O_i^{ND} - E_i^{ND})^2}{E_i^{ND}} + \frac{(O_i^D - E_i^D)^2}{E_i^D} \right], \quad (18)$$

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 model.²⁹ 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 (2004)). 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 satisfactorily.

²⁸ In general, for a $j \times k$ contingency table there are $(j-1)(k-1)$ 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 $\chi^2(k)$.

²⁹ Alternatively, the HL test statistic might be defined as $HL = \sum_{i=1}^k \frac{n_i (\hat{\pi}_i - \pi_i)^2}{\pi_i (1 - \pi_i)}$, where $\hat{\pi}_i$ are observed default

rates, π_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.

3.5 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, that is, static and discrete hazard models based on logistic and probit regressions. Finally, all of the estimated models are validated on control samples. Following Stein (2007), we apply a rolling window methodology with out-of-time validation within estimating and validating the models.

3.5.1 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 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.

The FFIEC database is used for collection of financial ratios, in particular, the ratios acquired from UBPR reports in the database. Table 4 provides a comparison of the number of banks used for estimation of the models in this study³⁰ and the total number of banks in the U.S.³¹, 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

Table 4
Comparison of the number of banks (FFIEC vs. U.S. in total)

Date	# of banks		Model / Year	# of defaulted banks	
	FFIEC database	U.S. in total		FFIEC database	U.S. in total ³²
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 31/12/2006.³³

³⁰ 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 7 and 8.

³¹ <http://www.usbanklocations.com/bank-rank/total-assets.html>

³² See Footnote 18.

³³ For detailed explanation, see Section 3.5.2.

Table 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 31/12/2007
2010	defaulted in years 2008 + 2009 + 2010	31/12/2006 31/12/2007 31/12/2008
2011	defaulted in years 2008 + 2009 + 2010 + 2011	31/12/2006 31/12/2007 31/12/2008 31/12/2009
2012	defaulted in years 2008 + 2009 + 2010 + 2011 + 2012	31/12/2006 31/12/2007 31/12/2008 31/12/2009 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 31/12/2010.³⁴

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, a rolling window methodology (the walk-forward approach) with out-of-time validation is used, as it is closest to the actual application of default prediction models in practice and gives a realistic view of how a particular model would perform over time. Refer to Stein (2007) 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 to the default event, we use a 12 to 24-month horizon before the actual default. For example, for banks defaulted in 2008 balance sheet data and financial ratios are collected on December

³⁴ Again, see Section 3.5.2 for detailed explanation.

31, 2006. Data collection dates for the individual models are summarised in Table 5. The financial ratios used as explanatory variables for the estimation of the scoring models are provided in Table 6.

We decided to examine a total of 19 financial ratios that are expected by bank regulators 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.³⁵ Chosen

Table 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 ³⁶
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 the financial health of banks. These ratios are divided into five indicator groups (profitability, efficiency, liquidity, assets quality, and capital adequacy).

³⁵ 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).

³⁶ Since the literature (see, e.g., Berger and Bouwman (2009)) has found liquidity risk as an important contributor to default (next to the profitability and capital structure of firms), it could be interesting to analyse alternative liquidity proxies such as the liquidity coverage ratio, the net stable funding ratio under Basel II, or the liquidity creation measure (see Berger and Bouwman (2009)). However, these measures are not available in the FFIEC database.

variables come from the FFIEC database and are often used by regulators for comparison purposes. To 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 to that expected under the economic hypothesis³⁷, and therefore were removed (refer Table 6 above): 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 winsorised data and set all observations for the considered financial ratios that exceeded the 99th percentile or were below the 1st percentile, equal to these values.³⁸

3.5.2 Model estimation

As mentioned earlier, we estimate both static and dynamic discrete hazard models using equations (1) and (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 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 5). Clearly, data used for model validation is the same as for the static models, that is, we validate the 2010 hazard models by predicting defaults in 2011 using financial ratios collected on December 31, 2009.

³⁷ For example, we would assume that the higher total interest income as a proportion of interest earning assets (x_5), the lower estimated default probability for a given bank should be. However, estimated coefficient of this variable yielded a positive sign (the higher value of ratio, the higher estimated PD).

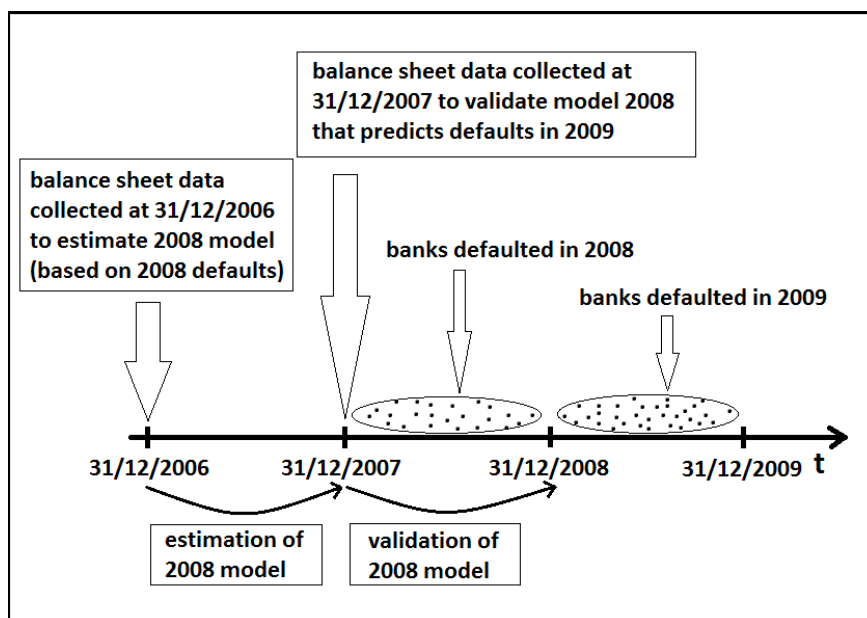
³⁸ The same approach was used, for instance, by Shumway (2001). One might argue that since we operate a heavily skewed sample, the symmetrical Winsorising approach might introduce a bias. However, in many cases the excluded observations appear to be a result of possible errors in the balance sheet data, such that we believe that excluding these observations from our sample is necessary.

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, for example, Altman (1968). The iterative procedure is finalised when adding another variable could not significantly improve the results (Altman et al. (1977)). Alternatively, for example, a stepwise regression technique could be applied. See, for example, Kolari et al. (2002), that provides an algorithm for inclusion of relevant variables that is based purely 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

Figure 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 the 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.

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 guided purely by statistical significance. For the estimation of the models, we start with one variable only from each of the indicator groups, that is, we include the variable with the highest explanatory power in the univariate regression for the category's 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.

Note that all explanatory variables eventually used in the models are not statistically dependent on each other.³⁹ Overall, the scale of 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.

Results for the estimated static logit and probit models for 2008 to 2012 are summarised in Table 7. The table contains information on the included variables, estimated coefficients, statistical significance and standard errors for the coefficients, log-likelihood of the model,

³⁹ In all cases, the coefficient of correlation does not exceed a value of 0.6. Further note that the applied "stepwise" algorithm will exclude highly correlated variables since they do not contribute to the explanatory power of the model. As illustrated in the subsequent analysis, our models do not use financial ratios based on equal denominators. We therefore strongly believe that the applied regression models do not suffer from multicollinearity and the provided results are not spurious.

pseudo- R^2 value⁴⁰, and statistical significance of the entire model (likelihood-ratio test⁴¹), 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 (%). The 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.

Regarding the economic impact of the estimated parameters of the considered ratios, we can see that all coefficients yield the correct sign that is consistent with the economic interpretation of the balance sheet ratios. For example, we obtain a negative sign for the variables x_1 and x_2 which implies that the higher the net income, the higher will be these ratios. Consequently, lower income corresponds to a higher PD for a given bank. Similarly, we have estimated negative coefficients for the variables x_{17} , x_{18} , and x_{19} . All these three ratios regard the required capital and, therefore, the higher the amount of capital held by a bank hold, the lower will typically be the PD for this bank. On the contrary, we obtain positive signs for the variables x_6 , x_8 , and x_{11} . This means that the higher the cost of total interest bearing deposits (x_6), the higher will be the estimated PD; the higher the ratio of net loans and leases to total assets (x_8), the higher will be the estimated PD; and finally the higher the ratio of noncurrent loans and leases to gross loans and leases (x_{11}), the higher will be the estimated PD.

⁴⁰ 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 minimisation of variance. Nevertheless, higher values still indicate a better model fit.

⁴¹ 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 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*** (2.63)	-6.85*** (0.80)	-11.16*** (0.97)	-4.83*** (0.39)	-8.98*** (1.25)	-5.17*** (0.50)	-9.11*** (1.57)	-4.62*** (0.71)	-2.72* (1.54)	-1.79*** (0.56)
x ₁ : ROA				-5.70* (3.46)						
x ₂ : ROE	-2.27* (1.36)				-2.55*** (0.43)	-1.35*** (0.22)	-2.73*** (0.56)	-1.34*** (0.26)	-2.06*** (0.73)	-1.03*** (0.30)
x ₆ : C_IBD	206.71*** (45.00)	60.47*** (13.63)	161.47*** (19.89)	62.84*** (8.34)	116.88*** (18.38)	53.70*** (8.59)	83.65*** (23.93)	41.04*** (11.10)		
x ₈ : NULL_TA	5.34* (2.83)	2.26*** (0.82)	2.17** (0.94)	0.93** (0.39)	3.86*** (1.26)	2.57*** (0.54)	5.33*** (1.61)	2.46*** (0.74)		
x ₁₁ : NCRLL_GLL	21.84* (11.96)	13.29*** (3.66)	28.32*** (3.05)	13.54*** (1.60)	25.97*** (2.82)	13.17*** (1.41)	21.01*** (3.46)	9.72*** (1.63)	21.69*** (5.07)	8.78*** (2.06)
x ₁₃ : EQ_TA					-24.91*** (4.82)				-57.82*** (17.90)	-19.23*** (6.30)
x ₁₇ : T1RBC_RWA										
x ₁₈ : TRBC_RWA	-9.08* (2.63)									
x ₁₉ : T1LC			-17.10*** (4.09)	-4.83*** (0.39)		-12.43*** (2.59)	-34.58*** (9.78)	-4.62*** (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-R2 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 6. *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Table 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*** (0.72)	-3.30*** (0.26)	-7.82*** (0.69)	-2.93*** (0.17)	-7.88*** (0.63)	-3.03*** (0.15)	-6.17*** (0.35)	-3.09*** (0.13)
x ₁ : ROA		-6.53* (2.82)						
x ₂ : ROE			-2.10*** (0.33)	-1.02*** (0.15)	-2.28*** (0.27)	-1.08*** (0.12)	-2.23*** (0.24)	-1.11*** (0.11)
x ₆ : C_IBD	132.55*** (14.20)	41.01*** (5.05)	91.32*** (9.46)	35.18*** (3.53)	96.48*** (7.36)	40.02*** (2.90)	104.36*** (6.26)	41.50*** (2.56)
x ₈ : NULL_TA			2.11* (0.70)		2.24* (0.65)			
x ₁₁ : NCRLL_GLL	26.53*** (2.68)	11.91*** (1.38)	24.32*** (1.99)	11.40*** (0.93)	22.07*** (1.64)	10.15*** (0.77)	21.49*** (1.54)	9.76*** (0.72)
x ₁₇ : TIRBC_RWA	-15.50*** (2.88)	-5.72*** (1.02)	-15.10*** (2.42)	-6.20*** (0.83)	-16.94*** (2.35)	-6.78*** (-3.03)	-19.91*** (2.20)	-6.67*** (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-R2 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 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 2009⁴² to 2012 are reported in Table 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 (%). Also here we got the expected signs for all estimated coefficients. Interpretation of the economic impact is then the same as in the case of the static models. Note that based on the applied algorithm, for most of the years, a very similar subset of variables was included into the final model for the estimated static and discrete hazard models.⁴³ As in the case of the static models, the 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 (1989)). 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.⁴⁴ The higher the likelihood, the more accurately a model predicts actual defaults.

⁴² 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.

⁴³ As mentioned in Section 3.4.2, 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.

⁴⁴ For a closer discussion on calibration and likelihood-based measures of calibration, see Stein (2007).

Table 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.

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-likelihood⁴⁵ 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 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 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 10, the difference between the logit and probit model is not statistically significant. We are also able to statistically distinguish between most of the

Table 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. The top number represents Z statistics while the number in parenthesis is a p -value. Statistically significant differences are highlighted in bold.

⁴⁵ 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.

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 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.

3.5.3 Model validation

As mentioned above, we apply a framework called the walk-forward approach with out-of-time validation that allows for testing of models while controlling for time dependence, see Stein (2007) for a more thorough discussion of this approach. This technique suggests use of a different set of data in validating the out-of-sample performance of the estimated models. At the same time, the approach allows for use of as much of the data as possible to fit and to test the models.⁴⁶ Numbers of non-defaulted and defaulted banks used for the validation along with data collection dates are reported in Table 11.

Table 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 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.

⁴⁶ While Stein (2007) 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.

As explained in Section 3.5.2 and Figure 2, for validation of the estimated models we use financial ratios collected one year after estimation of the models (see Table 5). For 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.

3.5.3.1 Distributions of estimated PDs

Let us first have a look at the estimated out-of-sample default probabilities obtained from the models. In Table 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.

3.5.3.2 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 summarised in Table 13, while Figure 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 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, that is, 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, which makes it hard to decide whether

Table 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%
		std.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%
		std.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%
		std.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%
		std.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%
		std.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%
		std.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%
		std.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%
		std.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.

any of the models is able to outperform the others. The only exception is 2012, where the 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 caused by the exclusion of the financial ratio *net loans & leases to total assets* (x_8). As stated

Table 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

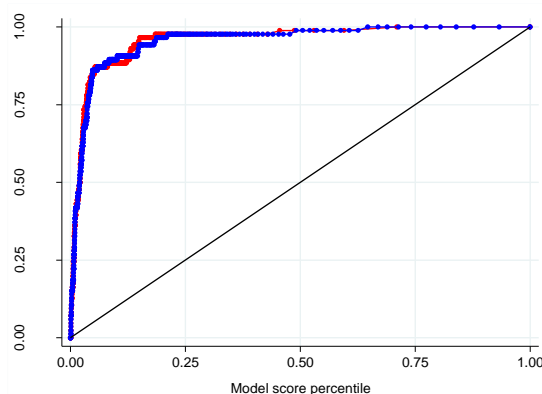
The table shows calculated areas under the ROC curves (ROC area) and accuracy ratios (AR) for each of the estimated models. Reported values are very high due to a high number of non-defaulted banks compared to defaulted banks.

in Footnote 43, 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, that is, 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.

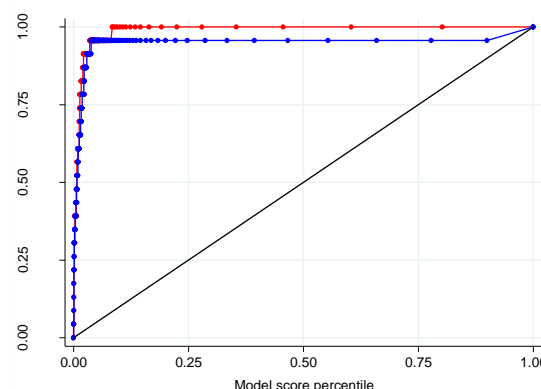
Figure 3

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.

3.5.3.3 Sizing the variability of ROC areas – bootstrapping

As pointed out by Stein (2007), results of model testing are subject to sample variability. The author also illustrates that it is typically the number of defaults rather 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 (2007) 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.⁴⁷ Figure 4 provides an exemplary plot of the distribution of ROC areas for the hazard

Table 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

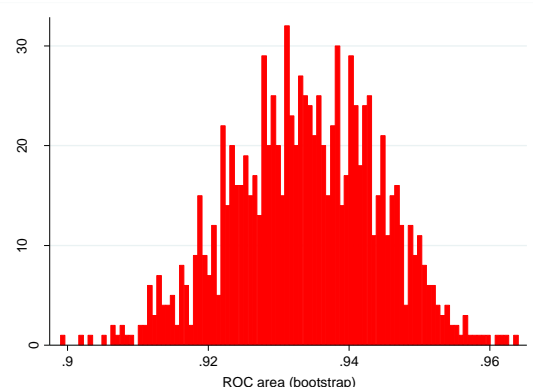
The table reports descriptive statistics of the distributions (mean value, standard deviation, skewness, and kurtosis) of areas under the ROC curves (ROC area) for each model and year based on 1,000 bootstrapped re-samples for each year.

⁴⁷ 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 re-samples from this pool. We keep the bootstrap size same as sample size.

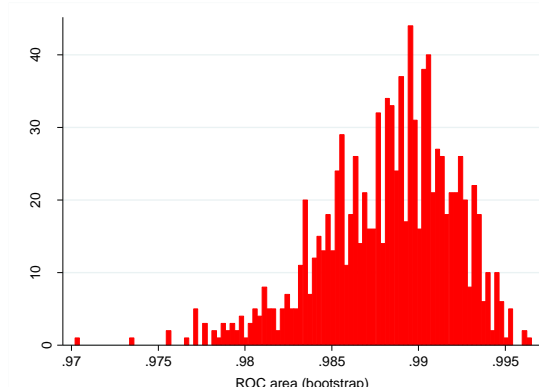
Figure 4

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 14.

Results for ROC areas based on the bootstrap re-samples 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 14) typically differs only slightly from its original value (Table 13).⁴⁸ 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 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 testing 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 16 indicates for all four

⁴⁸ For example, mean value of ROC area statistic calculated from bootstrapping for 2009 hazard probit model is 0.9392 while without re-sampling this statistic is 0.9389 for this model.

Table 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.⁴⁹

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 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

⁴⁹ Test statistics, along with the number of degrees of freedom and p -values, are identical to the values in the Kruskal-Wallis test.

Table 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 was 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 et al. (1988) 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 et al. (2003) 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.

3.5.3.4 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 (1989)). Results for log-likelihoods for the out-of-sample validation of the models are reported in Table 17. Note that in comparison to Section 3.5.2

Table 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 (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, that is, the model predicts PDs more accurately. For each year, results for the best model are indicated in bold.

where results on model estimation were examined using the log-likelihood, results in Table 17 are based only on out-of-sample results. Thus, the log-likelihood in equation (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 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 closeness test for non-nested models. In a first step, we test the best model, that is, 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 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 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 (0.0896)	0.3821 (0.3512)	0.2633 (0.3961)	0.5134 (0.3038)	0.4134 (0.3397)	0.5899 (0.2776)
model 2010	2.9600 (0.0015)	1.6880 (0.0457)	0.4842 (0.3141)	2.7187 (0.0033)	0.8103 (0.2089)	2.5622 (0.0052)
model 2011	0.0008 (0.4997)	0.0253 (0.4899)	0.0981 (0.4609)	0.0212 (0.4915)	-0.0802 (0.5320)	0.1040 (0.4586)
model 2012	-0.0995 (0.5396)	0.5577 (0.2885)	0.7600 (0.2236)	0.3507 (0.3629)	1.5453 (0.0611)	1.0171 (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 the static logit model with the log-likelihood value of -153.09, nonetheless this difference is not significant whatsoever (p -value of 0.5396).

3.5.3.5 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 or not a specific model is correctly calibrated. 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 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 summarised in Table 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 2012.⁵⁰ 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

⁵⁰ The reversal in under- to overestimation between 2009 and 2010 between observation and prediction might imply a data-driven result which would indicate deficiencies in the models. Generally, it is acknowledge that one-equation models might be insufficient in representing the entire default process, see, for example, Laitinen (1991) or Mecaj (2013). Note, however, that the relatively convincing results for default prediction – also in comparison to other recent major studies such as Beaver et al. (2005) or Duffie et al. (2007) – would suggest that the applied approach and the estimated models are quite useful.

expected (under a given model) and actual default rates; see, for example, Stein (2006). Such a comparison for each model, along with the Hosmer-Lemeshow's chi-squared goodness-of-

Table 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
Static logit	(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%
	(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	
Hazard logit	(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%
	(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	
Static probit	(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%
	(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	
Hazard probit	(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%
	(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.

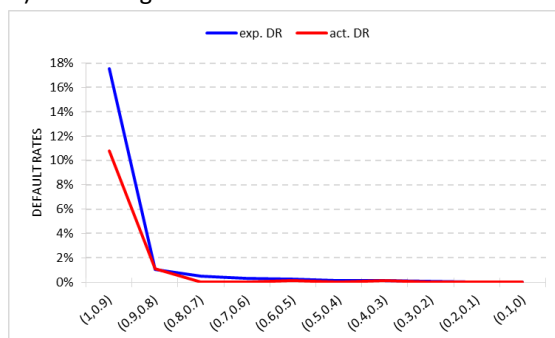
fit tests, is carried out in Table 20, where we divided the ranked PD estimates into ten intervals of the same size (each interval contains the same number of banks).⁵¹ Moreover, we also illustrate the comparison procedure in more detail for the 2010 hazard logit and the 2009 static probit model in Figure 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 $[0,0.10]$, $(0.10,0.20]$, ..., $(0.80,0.90]$, that is, 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, that is, 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 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.

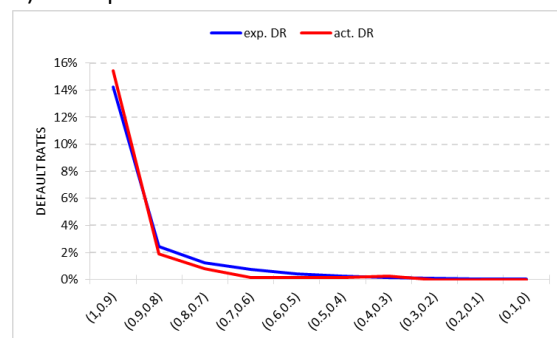
Figure 5

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).

⁵¹ The first interval $(0.90,1]$ contains the 10% banks with the lowest credit score (banks with the highest estimated PDs), the second interval $(0.80,0.90]$ contains the next 10% of the banks, etc.

Table 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
Static logit	(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
	(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
Hazard logit	(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
	(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
Static probit	(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
	(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
Hazard probit	(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
	(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 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 20.

There is a reason why these results might be biased. As stated in Section 3.4.3.4, 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.

3.5.3.6 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 differ significantly 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 22 (to save space, we do not report values in intervals (0.92,0.925] – (0.975,0.98]).⁵²

We can see that the power of the models increases over time (which is in accordance with the ROC analysis conclusions).⁵³ 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.

⁵² Generally, there is a downward trend with gradual decline of the ratios over the small intervals.

⁵³ In terms of the number of defaulted banks captured by the models, the 2008 models yield a relatively poor accuracy. Interestingly, the reverse is the case for Table 20 which produces the smallest differences between expected and actual default rates for the 2008 models. This might imply data-driven results (see Footnote 50).

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

Table 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%)

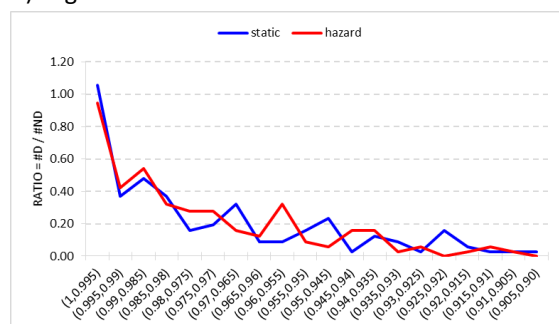
The table reports ratios of number of defaulted banks over number of non-defaulted banks (ratios) for the banks with the highest 10% PDs (divided into 20 intervals of size 0.5%) along with the information of number of captured defaulted banks within this interval.

here) are among the 299 banks (out of 6,641 banks)⁵⁴ with the highest estimated PDs. These findings are quite promising and achieve very good results compared to recent studies conducted in the corporate sector. Beaver et al. (2005) build a model based on accounting ratios which captures 80.3% of the year-ahead defaulting corporations in the lowest two

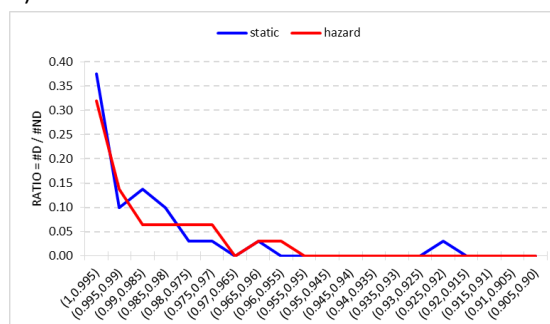
Figure 6

Ratios of # of defaulted over # of non-defaulted banks

A) Logit 2009 models



B) Probit 2012 models



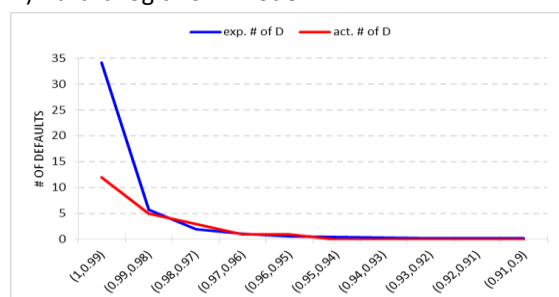
The figure shows a comparison between ratios (number of defaulted banks over number of non-defaulted banks) calculated for the static and the hazard models for the banks with the highest 10% PDs. The 2009 logit models are considered in panel A while the 2012 probit models in panel B.

deciles (period 1994-2002). Once the authors include additional variables 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 et al. (2007) places 94% of the one-year ahead defaults in the lowest two deciles (period 1993-2004). We typically obtain

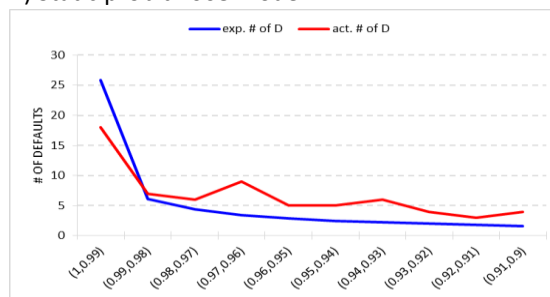
Figure 7

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.

⁵⁴ For validation of 2012 models, there are 23 defaulted banks and 6,618 non-defaulted banks in our sample (see Table 11). 4.5% out of 6,641 (23 + 6,618) is then 299 banks.

Table 23

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

	interval	model 2008		model 2009		model 2010		model 2011		model 2012	
		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
Static logit	(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
	(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
Hazard logit	(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
	(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
Static probit	(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
	(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
Hazard probit	(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
	(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.

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 6. As expected, their values have a downward trend with decreasing values of estimated PDs.

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%). The 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 23.⁵⁵

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 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 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 3.5.3.5 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 7.

3.5.3.7 Estimated PDs and "real" PDs

Overall, we state that our estimated default probabilities are slightly biased (see Table 23, where we got a rejection on the banks with the highest 10% PDs for all of the models, or Table 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

⁵⁵ Results in this table also imply data-driven results. Again, see Footnote 50.

reasonable estimates since, except for the first decile that contains the 10% banks with the worst rating, the analysis in Tables 20, 21 and Figure 5 illustrates that the expected and actual default rates 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 showing 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 (1984)) or translate estimated PDs into real ones using various transforms (e.g. Neagu et al. (2009)). 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.

3.5.3.8 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 24. Note that we report the results separately for each year, that is, 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 first group is represented by number of variables, Pseudo R^2 , the log-likelihood, and results for the calibration accuracy test (in-sample); the other 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 being 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 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 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 the estimated PDs are too low (the only exception is the 2009 static probit model), while for 2010, 2011 and 2012 the 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.

Table 24

A summary of the results on model comparison

Estimation				Validation								
# of variables	Pseudo R ²	Log-likelihood of CA test (in-sample)	E (PD) ND	ROC area D	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		
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												
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%
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%
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%
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 R², 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 25
Rankings of the models

Estimation		Validation								Total points	Final ranking
Pseudo R2	ND	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			
model 2008	static logit	1 (4)	2 (3)	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)	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)	reject (0)	3 (2)	19	3-4
	hazard logit	3 (2)	1 (4)	4 (1)	3-4 (2)	3-4 (2)	1 (4)	reject (0)	2 (3)	19	3-4
	static probit	1 (4)	3-4 (2)	1-2 (4)	2 (3)	1-2 (4)	1 (4)	accept (1)	4 (1)	24	1-2
	hazard probit	4 (1)	3-4 (2)	1-2 (4)	1 (4)	1-2 (4)	2 (3)	reject (0)	1 (4)	24	1-2
model 2010	static logit	2 (3)	2 (3)	1 (4)	1 (4)	1 (4)	2-3 (3)	reject (0)	1-4 (4)	27	1
	hazard logit	3 (2)	1 (4)	3 (2)	2 (3)	2 (3)	1 (4)	reject (0)	1-4 (4)	26	2
	static probit	1 (4)	4 (1)	2 (3)	3 (2)	3 (2)	4 (1)	reject (0)	1-4 (4)	18	3
	hazard probit	4 (1)	3 (2)	4 (1)	4 (1)	4 (1)	2 (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)	reject (0)	1-2 (4)	25	1-2
	hazard logit	3 (2)	1 (4)	3 (2)	2 (3)	1-2 (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)	reject (0)	1-2 (4)	18	3
	hazard probit	4 (1)	2 (3)	4 (1)	3-4 (2)	3-4 (2)	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)	reject (0)	1-2 (4)	23	2
	hazard logit	3 (2)	1 (4)	3 (2)	4 (1)	3-4 (2)	2 (3)	reject (0)	1-2 (4)	22	3
	static probit	2 (3)	3 (2)	2 (3)	1 (4)	1-2 (4)	1 (4)	accept (1)	3-4 (2)	25	1
	hazard probit	4 (1)	2 (3)	4 (1)	3 (2)	3-4 (2)	4 (1)	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 assigned 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, acceptance 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 put all of this information together, we have created a simple ranking system in Table 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 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 one 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.

3.6 Conclusions

In this paper, 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 7,000 U.S. commercial banks with 405 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.

A substantial part of this paper 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 test to investigate significant differences between the particular models in terms of bootstrapped ROC areas. Furthermore, we apply the Vuong 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 build 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 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 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 multiple comparison test to examine significant differences between the particular models in terms of bootstrapped ROC areas. Specifically, the Tukey 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.

Bibliography

- ALAM, P., BOOTH, D., LEE, K. & 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, 185-199.
- ALTMAN, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23, 589-609.
- ALTMAN, E. I., AVERY, R. B., EISENBEIS, R. A. & SINKEY, J. F. 1981. *Application of Classification Techniques in Business, Banking and Finance*, Greenwich, USA, JAI press.
- ALTMAN, E. I., HALDEMAN, R. G. & NARAYANAN, P. 1977. ZETATM analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1, 29-54.
- ALTMAN, E. I., RIJKEN, H., BALAN, D., MINA, J., FORERO, J. & WATT, M. 2010. The Z-Metrics™ Methodology for Estimating Company Credit Ratings and Default Risk Probabilities. RiskMetrics Group.
<https://www.fdic.gov/regulations/reform/altman2.pdf> [April 11, 2016].
- ALTMAN, E. I. & SAUNDERS, A. 1997. Credit risk measurement: Developments over the last 20 years. *Journal of Banking & Finance*, 21, 1721-1742.
- BAESENS, B., VAN GESTEL, T., VIAENE, S., STEPANOVA, M., SUYKENS, J. & VANTHIENEN, J. 2003. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54, 627-635.
- BALCAEN, S. & OOGHE, H. 2006. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38, 63-93.
- BEAVER, W. H. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71-111.
- BEAVER, W. H., MCNICHOLS, M. F. & 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, 93-122.
- BERGER, A. N. & BOUWMAN, C. H. 2009. Bank liquidity creation. *Review of Financial Studies*, 22, 3779-3837.
- BERGER, A. N. & BOUWMAN, C. H. 2013. How does capital affect bank performance during financial crises? *Journal of Financial Economics*, 109, 146-176.

- BOVENZI, J. F., MARINO, J. A. & MCFADDEN, F. E. 1983. Commercial bank failure prediction models. *Economic Review*, 68, 14-26.
- BOYACIOGLU, M. A., KARA, Y. & BAYKAN, Ö. K. 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, 3355-3366.
- CANBAS, S., CABUK, A. & KILIC, S. B. 2005. Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The Turkish case. *European Journal of Operational Research*, 166, 528-546.
- COHEN, B. H. 2013. How have banks adjusted to higher capital requirements? BIS Quarterly Review. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2401569 [April 19, 2016].
- DELONG, E. R., DELONG, D. M. & CLARKE-PEARSON, D. L. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44, 837-845.
- DUFFIE, D., SAITA, L. & WANG, K. 2007. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83, 635-665.
- DURAND, D. 1941. *Risk Elements in Consumer Instalment Financing*, New York, USA, National Bureau of Economic Research.
- EFRON, B. & TIBSHIRANI, R. J. 1994. *An Introduction to the Bootstrap*, Boca Raton, USA, Chapman & Hall/CRC.
- ENGELMANN, B., HAYDEN, E. & TASCHE, D. 2003. Testing rating accuracy. *Risk*, 16, 82-86.
- ESPAHBODI, P. 1991. Identification of problem banks and binary choice models. *Journal of Banking & Finance*, 15, 53-71.
- FISHER, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Human Genetics*, 7, 179-188.
- GREEN, D. M. & SWETS, J. A. 1966. *Signal Detection Theory and Psychophysics*, New York, USA, John Wiley & Sons.
- HAND, D. J. & HENLEY, W. E. 1997. Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160, 523-541.
- HANLEY, J. A. 1989. Receiver operating characteristic (ROC) methodology: the state of the art. *Critical Reviews in Diagnostic Imaging*, 29, 307-35.
- HILLEGEIST, S. A., KEATING, E. K., CRAM, D. P. & LUNDSTEDT, K. G. 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9, 5-34.
- HOCHBERG, Y. & TAMHANE, A. C. 1987. *Multiple Comparison Procedures*, New York, USA, John Wiley & Sons.
- HOLLANDER, M. & WOLFE, D. 1999. *Nonparametric Statistical Methods*, New York, USA, John Wiley & Sons.
- HOSMER JR, D. W. & LEMESHOW, S. 2004. *Applied Logistic Regression*, John Wiley & Sons.
- IOANNIDIS, C., PASIOURAS, F. & ZOPOUNIDIS, C. 2010. Assessing bank soundness with classification techniques. *Omega*, 38, 345-357.
- JO, H. & HAN, I. 1996. Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction. *Expert Systems with Applications*, 11, 415-422.

- KOLARI, J., CAPUTO, M. & WAGNER, D. 1996. Trait recognition: An alternative approach to early warning systems in commercial banking. *Journal of Business Finance & Accounting*, 23, 1415-1434.
- KOLARI, J., GLENNON, D., SHIN, H. & CAPUTO, M. 2002. Predicting large US commercial bank failures. *Journal of Economics and Business*, 54, 361-387.
- KOROBOW, L. & STUHR, D. P. 1975. Toward early warning of changes in banks' financial condition: A progress report. Federal Reserve Bank of New York.
https://www.newyorkfed.org/medialibrary/media/research/monthly_review/1975_pdf/07_3_75.pdf [April 11, 2016].
- KOROBOW, L. & STUHR, D. P. 1985. Performance measurement of early warning models: Comments on West and other weakness/failure prediction models. *Journal of Banking & Finance*, 9, 267-273.
- KOROBOW, L., STUHR, D. P. & MARTIN, D. 1976. A probabilistic approach to early warning of changes in bank financial condition. Federal Reserve Bank of New York.
https://www.newyorkfed.org/medialibrary/media/research/monthly_review/1976_pdf/07_3_76.pdf [April 11, 2016].
- KPMG 2014. Evolving Banking Regulation: Is the End in Sight? KPMG.
<http://www.kpmg.com/KY/en/IssuesAndInsights/ArticlesPublications/PublishingImages/evolving-banking-regulation-ema-2014-v3-fs.pdf> [April 19, 2016].
- LAITINEN, E. K. 1991. Financial ratios and different failure processes. *Journal of Business Finance & Accounting*, 18, 649-673.
- LAM, K. F. & MOY, J. W. 2002. Combining discriminant methods in solving classification problems in two-group discriminant analysis. *European Journal of Operational Research*, 138, 294-301.
- LANE, W. R., LOONEY, S. W. & WANSLEY, J. W. 1986. An application of the Cox proportional hazards model to bank failure. *Journal of Banking & Finance*, 10, 511-531.
- LANINE, G. & VENNET, R. V. 2006. Failure prediction in the Russian bank sector with logit and trait recognition models. *Expert Systems with Applications*, 30, 463-478.
- LEE, T.-H., & CHIH, S.-H. 2013. Does financial regulation affect the profit efficiency and risk of banks? Evidence from China's commercial banks. *The North American Journal of Economics and Finance*, 26, 705-724.
- LINTON, L. & HARDER, L. 2007. Biology 315—Quantitative Biology Lecture Notes. Calgary, AB: University of Calgary.
- LO, A. W. 1986. Logit versus discriminant analysis: A specification test and application to corporate bankruptcies. *Journal of Econometrics*, 31, 151-178.
- LOGAN, A. 2001. The United Kingdom's small banks' crisis of the early 1990s: what were the leading indicators of failure?, Bank of England.
- MECAJ, A. 2013. Structural and evolutionary patterns of financially distressed firms: Strategies to overcome decline. Universidad de Salamanca.
http://gredos.usal.es/jspui/bitstream/10366/124180/1/DAEE_ArjolaMecaj_Tesis.pdf [April 18, 2016].
- 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.

- MARTIN, D. 1977. Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance*, 1, 249-276.
- MCFADDEN, D. 1976. A comment on discriminant analysis" versus" logit analysis. *Annals of Economic and Social Measurement*, 5, 511-523.
- NEAGU, R., KEENAN, S. & CHALERMKRAIVUTH, K. 2009. Internal credit rating systems: Methodology and economic value. *The Journal of Risk Model Validation*, 3, 11-34.
- OHLSON, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18, 109-131.
- QUEEN, M. & ROLL, R. 1987. Firm mortality: using market indicators to predict survival. *Financial Analysts Journal*, 43, 9-26.
- SANTOMERO, A. M. & VINSO, J. D. 1977. Estimating the probability of failure for commercial banks and the banking system. *Journal of Banking & Finance*, 1, 185-205.
- SHUMWAY, T. 2001. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74, 101-124.
- SINKEY, J. F. 1975. A multivariate statistical analysis of the characteristics of problem banks. *The Journal of Finance*, 30, 21-36.
- STEIN, R. M. 2006. Are the probabilities right? Dependent defaults and the number of observations required to test for default rate accuracy. *Journal of Investment Management*, 4, 61.
- STEIN, R. M. 2007. Benchmarking default prediction models: Pitfalls and remedies in model validation. *Journal of Risk Model Validation*, 1, 77-113.
- STUHR, D. P. & VAN WICKLEN, R. 1974. Rating the financial condition of banks: A statistical approach to aid bank supervision. Federal Reserve Bank of New York.
https://fraser.stlouisfed.org/docs/publications/frbnyreview/pages/1970-1974/67599_1970-1974.pdf [April 11, 2016].
- TABACHNICK, B. G. & FIDELL, L. S. 2007. *Using Multivariate Statistics* Boston, USA, Pearson Education Inc.
- TAM, K. Y. & KIANG, M. Y. 1992. Managerial applications of neural networks: the case of bank failure predictions. *Management Science*, 38, 926-947.
- THOMSON, J. B. 1991. Predicting bank failures in the 1980s. *Economic Review*, 27, 9-20.
- TRÜCK, S. & RACHEV, S. T. 2009. *Rating Based Modeling of Credit Risk: Theory and Application of Migration Matrices*, Norwell, USA, Academic Press.
- VITALE, R. & LAUX, J. A. 2012. The Economic Efficacy Of Banking Mergers: 2006-2008. *Journal of Applied Business Research (JABR)*, 28, 1211-1216.
- VUONG, Q. H. 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57, 307-333.
- WEST, R. C. 1985. A factor-analytic approach to bank condition. *Journal of Banking & Finance*, 9, 253-266.
- WHALEN, G. & THOMSON, J. B. 1988. Using financial data to identify changes in bank condition. *Economic Review*, 24, 17-26.
- ZMIJEWSKI, M. E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59-82.

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

Martin Gurny (contribution: 70%), Stefan Trück (contribution: 30%)

This paper was presented at:

- The Centre for Financial Risk PhD Workshop 2015, Sydney, Australia, October 2015.
- The 5th International Association for Energy Economics (IAEE) Asian Conference, Perth, Australia, February 2016 (paper is accepted for presentation).

Abstract

During the last decade, the renewable energy sector has exhibited significant growth rates in the global economy. At the same time, similar to the high-tech sector or venture capital, investments into renewable energy companies can be considered as being relatively risky; due to the nature of their business, there is often a significant gap between innovation, adoption, and a phase where the company really becomes established in the market. This paper examines the relationship between distress risk and returns of U.S. renewable energy companies. Using the *Expected Default Frequency* (EDF) obtained from *Moody's KMV*, we demonstrate that there is a positive cross-sectional relationship between default risk and equity returns in the renewable energy sector. We find that equally-weighted (EW) portfolios with higher default risk characteristics also yield higher returns. The positive relationship between distress risk and equity returns is also confirmed by investigating returns corrected for common Fama and French (1993) and Carhart (1997) risk factors. We further show that raw and risk-adjusted returns of EW portfolios that take a long position in the 10% most distressed stocks and a short position in the 10% safest stocks, outperform the S&P 500 index throughout our sample period.

Keywords: Renewable energy, Distress risk, EDF measure, Asset pricing models, Raw and risk-adjusted returns

4.1 Introduction

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), see, e.g., Inchauspe et al. (2015). At the same time, similar to technology stocks or venture capital, investments into renewable energy stocks can be considered as being relatively risky, see, for example, Henriques and Sadorsky (2008), Kumar et al. (2012), Sadorsky (2012a), or Managi and Okimoto (2013). In general, only a fraction of renewable energy companies become major players in the sector, 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 in the market. This gap is often referred to as the “Valley of Death”, see, for example, Weyant (2011). As a result, one may argue that, in particular, investors who buy shares in small and/or highly risky renewable energy companies, that is, stocks with typically higher volatility and possibly higher default risk, will also expect higher average returns for bearing this risk. Our paper aims to shed light on this important question and thoroughly examine 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 or realized 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 (2004) 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 (2010), 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, that is, 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, for example, Dichev (1998),

Griffin and Lemmon (2002), Campbell et al. (2008), Garlappi et al. (2008), or Avramov et al. (2009).

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, for example, Henriques and Sadorsky (2008), Kumar et al. (2012), Sadorsky (2012a), Bohl et al. (2013), or Managi and Okimoto (2013). 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.⁵⁶ However, none of these studies has examined how distress risk is priced in the renewable energy sector.

In this paper, 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. In particular, we provide the first empirical study that investigates the question of 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 (1993) and Carhart (1997) 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 (1974). One key advantage of this measure is its availability at a daily frequency; this 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, for example, a monthly basis. This also allows us to investigate the performance of

⁵⁶ 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.

the constructed portfolios on a monthly basis as it is typically done in the literature that motivates our study, see, for example, Fama and French (1993), Carhart (1997), Vassalou and Xing (2004), Boyer and Filion (2007), Campbell et al. (2008), Garlappi et al. (2008). Note that in comparison to most earlier studies focusing in particular on the renewable energy sector, for example, Henriques and Sadorsky (2008), Kumar et al. (2012), Sadorsky (2012a), Bohl et al. (2013), Managi and Okimoto (2013), we also significantly extend the time period considered by using a data set of monthly returns from 2002 up to 2013. 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 of the above-mentioned studies that typically look at renewable energy stock indices only, 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.

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 demonstrate that there is a positive relationship between equity returns of equally-weighted (EW) portfolios and default risk. Thus, distressed renewable energy companies earn on average higher expected returns than renewables with low default risk. Our results confirm a pattern also suggested by Vassalou and Xing (2004) and Chava and Purnanandam (2010). We find a significant difference between returns of EW portfolios consisting of the riskiest decile of stocks, and one consisting of the decile with the lowest failure risk.

We further examine a possible link between pricing factors such as the size effect, the BM effect and distress risk, and find 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 show 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 the observed default effect.

Positively priced distress risk in the U.S. stock market for the renewable energy sector is also confirmed by applying three major asset pricing models, namely, the Capital Asset Pricing Model (CAPM), the Fama and French (1993) three-factor model, and the Carhart (1997) 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 show that raw and risk-adjusted returns of EW portfolios that take a long position in the 10% most distressed renewable stocks and a short position in the 10% renewable companies with the lowest default risk, generally outperform the S&P 500 index throughout our sample period (2002–2014).

This paper is organised 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 4.2. Section 4.3 is devoted to a brief description of three well-known asset pricing models and Fama-MacBeth regression, 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 4.4. In Section 4.5, 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 4.6. Finally, we conclude in Section 4.7 with a summary of our results.

4.2 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, for example, Dichev (1998), Griffin and Lemmon (2002), or Campbell et al. (2008).

Dichev (1998) was among the first to demonstrate the negative cross-sectional relationship between default risk and future stock returns, measuring default risk by the Altman (1968) Z-score and Ohlson (1980) 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 (2002) confirm the findings of Dichev (1998), and by using Ohlson (1980) O-score also find a negative relationship between distress risk and realised 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 (1974) model, Vassalou and Xing (2004) 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 (2001), Campbell et al. (2008) 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 (1993) size and value factors. They also tend to do poorly when market-wide implied volatility increases. Interestingly, Campbell et al. (2008) also find evidence for the “distress effect” being most pronounced among small and illiquid stocks; however, this means that these stocks yield particularly low returns for these stocks. Garlappi et al. (2008) 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 et al. (2009) demonstrate that most of the negative returns for high default risk stocks are concentrated around rating downgrades. Consistent with Campbell et al. (2008), 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 (1991) and Fama and French (1992) argue that the size and book-to-market effects proxy for a priced distress risk factor, and Vassalou and Xing (2004) find evidence that the size and book-to-market factors contain some default-related information, their results are in line with Dichev (1998) and Campbell et al. (2008), who conclude that distress risk is unlikely to be systematic.

Apart from Vassalou and Xing (2004), the study by Chava and Purnanandam (2010) 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 realised 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 (2011) explicitly consider financial leverage and study the cross-sectional implications of potential shareholder recovery upon resolution of financial distress. Contrary to Griffin and Lemmon (2002) and Vassalou and Xing (2004), they argue 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 (1999) 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 (2001) finds positive effects of an increasing oil price on Canadian oil and gas stocks. His results are confirmed by Boyer and Filion (2007) 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 (2008) 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 that closely mirrored those of technology stock prices. Sadorsky (2012a) 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 (2012b) studies the determinants of systematic risk for U.S. listed renewable energy stocks between 2001 and 2007 and finds 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 et al. (2012) 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 (2008), 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 (2013) 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 (2008), 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 (2015), 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 et al. (2012) and Wen et al. (2014) are studies predominantly focusing on renewable energy markets in China. While Broadstock et al. (2012) 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 et al. (2014) 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 et al. (2014) 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 (2008), who use one global index, or Kumar et al. (2012), who use three specific global indices, their study uses 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 on nuclear and renewable energy stocks of the Fukushima Daiichi nuclear disaster in Japan in March 2011. Ferstl et al. (2012) 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 et al. (2013) 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 (2014) 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 et al. (2013) 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 (2013) 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 et al. (2015). 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 et al. (2015) 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, for example, Bohl et al. (2013), Inchauspe et al. (2015), have also included standard pricing factors such as market risk, Fama and French (1993) size and value factors or a Carhart (1997) 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, that is 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.

4.3 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 along with the two-stage Fama-Macbeth (1973) approach.

Building on the Markowitz framework (1952, 1959), Sharpe (1964), Lintner (1965), and Mossin (1966) 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 (1992) was published, summarising much of the earlier empirical work in the area. As a result, Fama and French (1993) introduced a new three-factor model where, in addition to a market risk factor, a size (market capitalisation) 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 (1997), 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.

4.3.1 The Capital Asset Pricing Model (CAPM)

The CAPM is defined as

$$R_{i,t} - R_{F,t} = \alpha_i^{CAPM} + \beta_i (R_{M,t} - R_{F,t}) + \varepsilon_{i,t}, \quad (1)$$

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 . α_i^{CAPM} and β_i are estimated coefficients, where α_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, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. The idea behind this model is that excess return of a particular company or portfolio $(R_{i,t} - R_{F,t})$ can be explained by their relationship with a market risk factor $(R_{M,t} - R_{F,t})$.

4.3.2 The Fama-French three-factor model

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

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

Here, SMB_t is the realisation on a capitalisation-based factor portfolio that buys small cap stocks and sells large cap stocks. Similarly, HML_t is the realisation 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.

4.3.3 The Carhart four-factor model

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

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

where 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 capitalisation and stocks with a big market capitalisation, 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 (1993), Carhart (1997), or Professor Kenneth French's website⁵⁷. 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.

4.3.4 Fama-MacBeth regression

The Fama-MacBeth regression (1973) is a two-stage approach used to estimate parameters for asset pricing models such as the CAPM, the Fama-French three-factor model, or the Carhart four-factor model. The method estimates the betas and risk premia for any risk factors that are expected to determine asset prices. This approach can be applied to a panel of multiple companies with return observations across several time periods.

Assume we have n assets or portfolio returns over T periods with a particular asset's excess return in a particular time denoted $R_{i,t}$. We want to test whether the m factors $F_{1,t}, F_{2,t}, \dots, F_{m,t}$ explain our asset and portfolio returns, as well as the premiums awarded to the exposure to each factor. To do so, we must run a two-stage regression. The first stage involves a set of regressions equal to the number of assets or portfolios being tested. The second stage is a set of regressions equal to the number of time periods. The first stage regressions are a set of time series regression for each individual asset or portfolio on the considered factors.

$$\begin{aligned}
 R_{1,t} &= \alpha_1 + \beta_{1,F_1} F_{1,t} + \beta_{1,F_2} F_{2,t} + \dots + \beta_{1,F_m} F_{m,t} + \varepsilon_{1,t} \\
 R_{2,t} &= \alpha_2 + \beta_{2,F_1} F_{1,t} + \beta_{2,F_2} F_{2,t} + \dots + \beta_{2,F_m} F_{m,t} + \varepsilon_{2,t} \\
 &\vdots \\
 R_{n,t} &= \alpha_n + \beta_{n,F_1} F_{1,t} + \beta_{n,F_2} F_{2,t} + \dots + \beta_{n,F_m} F_{m,t} + \varepsilon_{n,t}
 \end{aligned} \tag{4}$$

Based on the estimated regression models we then know to what extent each asset's or portfolio's return is affected by the considered factors. In the second set of regressions we

⁵⁷ http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

define $\hat{\beta}_{i,F_k}$ as the estimated β s for each asset or portfolio for factor F_k . To calculate the premiums for each factor, we can then run the following set of cross-sectional regressions:

$$\begin{aligned}
 R_{i,1} &= \alpha_1 + \gamma_{1,1}\hat{\beta}_{i,F_1} + \gamma_{2,1}\hat{\beta}_{i,F_2} + \cdots + \gamma_{m,1}\hat{\beta}_{i,F_m} + e_1 \\
 R_{i,2} &= \alpha_2 + \gamma_{1,2}\hat{\beta}_{i,F_1} + \gamma_{2,2}\hat{\beta}_{i,F_2} + \cdots + \gamma_{m,2}\hat{\beta}_{i,F_m} + e_2 \\
 &\vdots \\
 R_{i,T} &= \alpha_T + \gamma_{1,T}\hat{\beta}_{i,F_1} + \gamma_{2,T}\hat{\beta}_{i,F_2} + \cdots + \gamma_{m,T}\hat{\beta}_{i,F_m} + e_T
 \end{aligned} \tag{5}$$

Note that the independent variables β_{i,F_k} are always exactly the same for every regression. The only variation is in the dependent variables, which are different for each time period.

The $\gamma_{j,t}$ terms are the estimated regression coefficients of model (5). To calculate a single risk premium for each factor, one must average all the $\gamma_{j,t}$ terms into a single γ_j . To calculate standard errors for the γ_j terms, one must treat each $\gamma_{j,t}$ observation as an independent observation and calculate a t -statistic as follows

$$\frac{\gamma_j}{\sigma_{\gamma_j} / \sqrt{T}} \tag{6}$$

where σ_{γ_j} is the standard deviation of the $\gamma_{j,t}$ terms. The γ_k coefficient represents the factor premium for an exposure with a $\hat{\beta}_{i,F_k}$ of 1 to factor F_k . Note that instead of running multiple cross-sectional regressions, it is also possible to run a single model that uses the average returns of all assets over time.

4.4 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, these are 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.⁵⁸ 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 capitalisation for a majority of Clean Energy Index stocks is typically \$US200 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.⁵⁹ 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 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 capitalisation for a majority of NEX index stocks is typically \$US250 million and above.

⁵⁸ Source: <http://www.wildershires.com/>. Accessed: July 2015.

⁵⁹ Source: <http://www.nexindex.com/>. Accessed: July 2015.

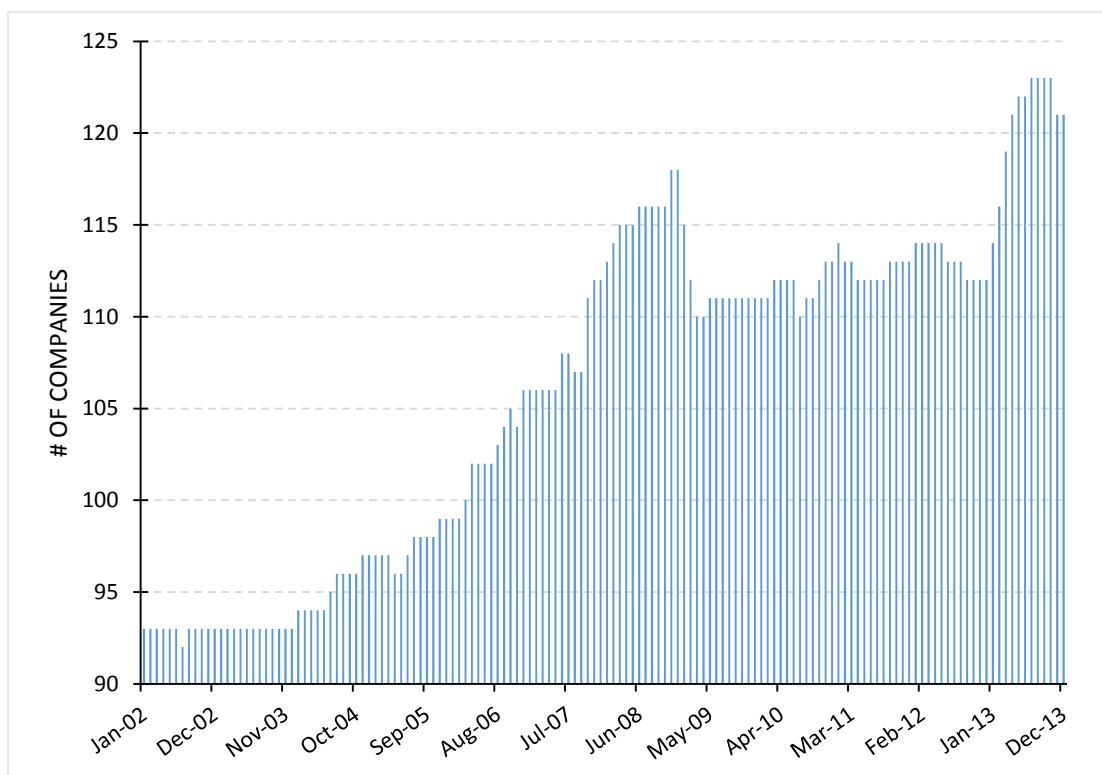
The **AGINA** index, as a part of the Ardour Global Alternative Energy Indices, merely focuses on North American renewable companies and tracked 55 companies as of June 2015. The largest company accounts for 2.02% and the top 5 holdings account for 9.75% of total AGINA investment. Companies included 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 must achieve at least 50% 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 capitalisation of \$US1.16 trillion. The **CELS** index is a modified market capitalisation-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 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 2013.

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, that is, 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

Figure 1

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 2013. We start with 93 companies in January 2002 and end with 121 companies in December 2013. Minimum number of companies is 92 (July 2002), maximum is then 123 (July-October 2013).

defaulted⁶⁰ (filed for bankruptcy protection under Chapter 11), 12 companies have been acquired⁶¹, 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. Note, that as it is typically done in asset pricing studies, we use monthly returns to measure the performance of the individual companies. Figure 1 provides a plot of the number of companies at each point in time throughout our sample period from January 2002 to December 2013.

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 discussion 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.

⁶⁰ We discuss defaulted companies, in terms of size and corresponding EDF values, in more detail in Appendix A.

⁶¹ We discuss acquired companies in more detail in Appendix B.

4.4.1 Returns

As mentioned above, returns are collected from the CRSP database.⁶² Because we focus on studying the returns on distressed stocks, we follow Campbell et al. (2008) 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 1. 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 15,379 monthly observations of returns for 141 companies, while there are 970,005 monthly observations of returns for 12,596 companies contained in CRSP. We can see that on average returns in our sample are slightly higher (1.2% compared to 1.0% for the U.S. market), although the median is lower (0.5% compared to 0.6%). Moreover, comparing the standard deviation of 0.179 in our sample to the lower standard deviation of 0.158 in the entire U.S. market, we conclude that our companies, in terms of returns, are typically more volatile. The fact that returns for our sample are slightly less skewed and leptokurtic suggests that there are more extreme return observations in the overall U.S. market.

Table 1
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.	15,379	970,005	# of obs.	141	12,596
Mean	0.012	0.010	Mean	0.171	0.144
Median	0.005	0.006	Median	0.164	0.121
Std.	0.179	0.158	Std.	0.072	0.110
Skewness	3.263	5.380	Skewness	0.850	4.429
Kurtosis	35.38	232.09	Kurtosis	4.76	77.19

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 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. Reported values are based on monthly returns (obtained from CRSP database) covering the period from January 2002 to December 2013.

⁶² We excluded returns for share prices below \$0.50.

The finding of our companies being more volatile 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 (0.171 compared to 0.144 for the U.S. market). The distribution of standard deviations is then more skewed and has a higher kurtosis for the overall U.S. market, as well as the variation of the estimated volatilities (standard deviation of 0.110 compared to 0.072 in our sample).

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.

4.4.2 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.⁶³ 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

Table 2

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
Full sample	15,270	2.96	6.93	0.01	35.00	0.39	0.12	1.78

Our sample spans a period from January 2002 to December 2013. 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 (2002-2013). EDF quantities are expressed in percent units. The number of companies in a "Full sample" denotes the total number of observations.

⁶³ This measure has 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.

originally by Merton (1974).⁶⁴ 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%.

The EDF measure is based on share prices and share price volatilities which may be subject to liquidity risk. Also, the literature has found liquidity risk as an important contributor to default (next to the profitability and capital structure of firms). Therefore, it would be interesting to investigate whether it is possible to decompose the EDF measure into a credit risk and a liquidity risk component, or alternatively, control for liquidity risk. Such an analysis is clearly beyond the scope of our work, however, it is a possible direction for future research.

For robustness check purposes within our analysis it would be worth to use alternative measures of default risk. For example, we could use the z-score introduced by Altman (1968) that is based on multiple discriminant analysis. This approach 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. Discriminant analysis 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 (2009)). 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. This approach derives the linear combinations from an equation that takes the form of $Z = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_n X_n$ where Z is the discriminant score (z-score), γ_0 is a constant, γ_i ($i = 1, 2, \dots, n$) the discriminant coefficients, and X_i ($i = 1, 2, \dots, n$) the independent variables, i.e. the financial ratios. Another possibility would be deriving a scoring model for the estimation of default probabilities, such

⁶⁴ See 2.3.1 or Appendix C for a more thorough discussion of this model.

as logit or probit model. Nevertheless, suggested measures would require obtaining additional accounting data for all companies in our sample.

Summary statistics for the EDF measure are reported in Table 2. The average EDF measure in our sample is 2.96% with a median of 0.39%. 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 1.8%.

Because the EDF measure is based on market prices, we follow Garlappi et al. (2008) 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

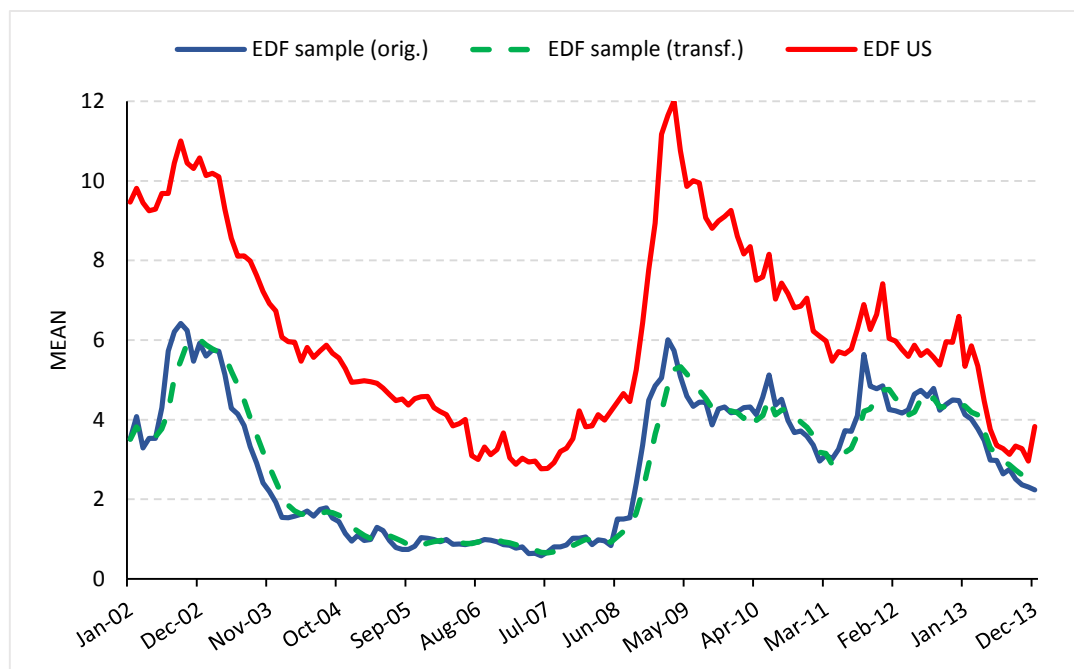
$$\overline{\text{EDF}}_t = \frac{\sum_{s=0}^5 e^{-s\nu} \text{EDF}_{t-s}}{\sum_{s=0}^5 e^{-s\nu}} \quad (7)$$

where ν is chosen to satisfy $e^{-5\nu} = 1/2$, 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 $\overline{\text{EDF}}_t$ 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 2. We also 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 3.

Figure 2 illustrates that there are only marginal differences between the average monthly original and transformed EDF measures. 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 3 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.

Figure 2

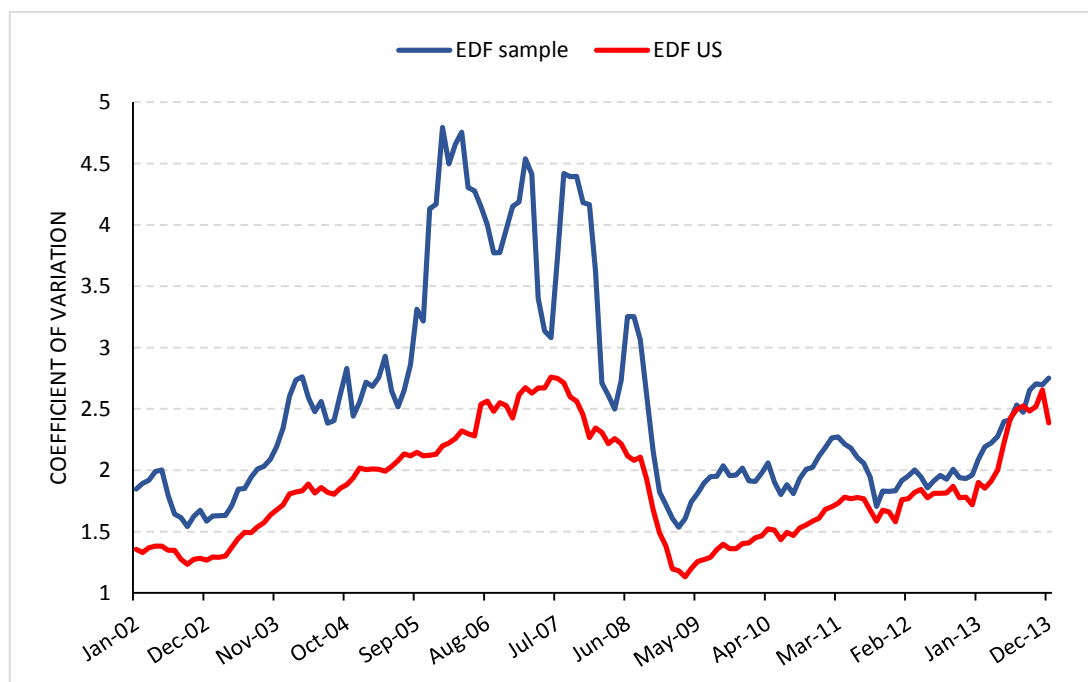
Mean of EDF measure



The figure plots the mean of original monthly EDF measure against the mean of transformed one, based on Equation (7) for our sample along with the average EDF for U.S. market over the period from January 2002 to December 2013.

Figure 3

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 2013.

4.5 Distress risk and equity returns

We start our analysis by investigating the relationship between distress risk (measured by the EDF) and equity returns. Specifically, we examine whether portfolios with different default risk characteristics provide significantly different returns. The results are reported in Table 3. At the end of each month t from January 2002 to November 2013, we form ten equally-weighted (EW) 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 9 and 12 companies. We then analyse returns of these EW portfolios in month $t + 1$. Portfolio 1 represents the portfolio of the 10% companies with the lowest distress risk, while portfolio 10 is the portfolio of the 10% 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 capitalisation, expressed in billions of \$US), and the average book-to-market (BM) ratio.

As illustrated in Table 3, we find a positive relationship between returns of EW portfolios and distress risk – the higher the EDF, typically the higher the corresponding return. This

Table 3

Raw returns on portfolios sorted on the basis of the EDF measure

Portfolios	EDF										High-Low
	Low 1	2	3	4	5	6	7	8	9	High 10	
Raw returns (EW)	0.72	1.12	1.52	0.94	1.42	0.67	0.84	1.15	1.44	2.98	2.26
t -stat (Newey-West)	(1.43)	(2.11)**	(2.32)**	(1.58)	(2.09)**	(0.95)	(1.14)	(1.33)	(1.48)	(2.41)**	(2.22)**
Average EDF	0.06	0.11	0.17	0.26	0.41	0.65	1.07	1.91	4.06	12.76	
Average size	9.42	6.03	3.59	2.58	2.15	1.70	1.49	1.33	0.66	0.62	
Average BM	0.37	0.45	0.48	0.51	0.51	0.58	0.60	0.63	0.62	0.90	

At the end of each month t from January 2002 to November 2013, we sort our companies into deciles based on their weighted EDF measures, as defined in (7). We then compute equally-weighted (EW) 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 10 with the highest one. The "High-Low" column is the difference between a quantity of the high EDF decile and that of the low EDF decile. 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 capitalisation) is expressed in billions of \$US.

positive relationship is consistent with findings of Vassalou and Xing (2004), who use their own “DLI” measure based on the Merton (1974) model as a measure of distress risk. The return difference between EW high default risk portfolio and low default risk portfolio is 2.26% per month (27.12% p.a.) and is statistically significant at the 5% level. Thus, similar to Vassalou and Xing (2004) we argue that the observed pattern is indicative of positively priced default risk.

Also note that small-capitalisation stocks have, on average, higher EDF scores, and as a result, they typically provide higher returns than big-capitalisation 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 (2004) and further investigate this possible link between the size and BM effects and default risk. We will focus on EW portfolios again, since this is the weighting scheme typically employed in studies that deal with the size and BM effects, see, for example, Fama and French (1992) and Vassalou and Xing (2004).

4.5.1 Size, BM, and distress risk

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

Table 4 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 1.61% per month (19.32% p.a.). This is about six times more than for a portfolio containing low-distress stocks (0.28% per month). We investigate to what extent we are truly

Table 4

Size effect controlled by default risk

Size	Small	Medium	Big	Small-Big	<i>t</i> -stat (Newey-West)
Panel A: Average Returns					
Low EDF	1.20	1.10	0.92	0.28	(0.54)
Medium EDF	1.60	1.20	0.61	0.99	(1.94)*
High EDF	2.80	1.20	1.19	1.61	(1.78)*
Panel B: Average Size (market capitalisation)					
Low EDF	0.592	2.237	14.474		
Medium EDF	0.188	0.726	4.914		
High EDF	0.060	0.237	2.444		
Panel C: Average EDF					
Low EDF	0.145	0.133	0.100		
Medium EDF	0.637	0.583	0.527		
High EDF	6.780	5.522	5.302		
Panel D: Average BM					
Low EDF	0.470	0.473	0.390		
Medium EDF	0.548	0.551	0.550		
High EDF	0.709	0.659	0.753		

At the end of each month from February 2002 to December 2013, 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 capitalisation 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 capitalisation) in Panel B is expressed in billions of \$US, while average EDF in Panel C is expressed in percent units, and average BM (book-to-market) in Panel D in decimal units.

capturing the size effect in Panel B. We can see that there really is substantial variation in the market capitalisation of stocks within the high EDF portfolio. However, it is not necessarily always the 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 \$US2.444 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 0.188 to 2.237). 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, the high EDF/small size portfolio does 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 eleven times riskier in terms of the applied EDF measure than small firms in medium EDF portfolios, and about 47 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 (2.478) compared to medium and low default risk groups (0.11 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 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 5 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 groups is sorted into three BM portfolios (high, medium, low). In the following discussion we will examine the BM effect within each of the EDF groups.

Panel A shows that the BM effect is statistically significant only for the constructed medium EDF portfolios, with a return differential of 0.91% per month (10.92% p.a.). This is five times more than the difference for a portfolio containing high-distress stocks (0.18% per month), and seven times more than for low EDF group (0.13%). 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 thus suggest that for

Table 5

BM effect controlled by default risk

BM	High	Medium	Low	High-Low	<i>t</i> -stat (Newey-West)
Panel A: Average Returns					
Low EDF	1.08	1.15	0.95	0.13	(0.30)
Medium EDF	1.50	1.28	0.59	0.91	(1.91)*
High EDF	2.27	0.86	2.09	0.18	(0.21)
Panel B: Average BM					
Low EDF	0.709	0.411	0.212		
Medium EDF	0.926	0.491	0.232		
High EDF	1.394	0.568	0.156		
Panel C: Average EDF					
Low EDF	0.141	0.124	0.114		
Medium EDF	0.623	0.568	0.556		
High EDF	6.812	4.104	6.615		
Panel D: Average Size (market capitalisation)					
Low EDF	5.051	5.526	7.008		
Medium EDF	2.105	2.552	1.216		
High EDF	0.873	1.263	0.653		

At the end of each month from February 2002 to December 2013, 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 capitalisation) in Panel D in billions of \$US.

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 and medium EDF category, that is, not for the portfolio with the highest default

risk where the relation is not monotonic. This is in contrast to our results from Panel C in Table 4, where we clearly find a monotonic relationship between default risk and size of portfolios.

Finally, Panel D shows that a portfolio with high distressed stocks contains mainly small firms. This time even the highest value of 1.263 is lower than any of the values from medium and high EDF groups (apart from low BM/medium EDF portfolio that is only marginally higher). And again, contrary to our findings from Table 4, 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 1.61%, and therefore about 1.4% larger than that of value stocks over growth stocks (0.18%).

4.5.2 The default effect

Tables 4 and 5 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 (2004) and define the default effect as a positive average return differential between high and low default risk firms.

In Table 6, we reverse the sorting procedure applied in Table 4 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 1.19% per month (14.28% 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 emphasises the substantially higher default risk for the high

Table 6

Default effect controlled by size

EDF	Low	Medium	High	High-Low	t-stat (Newey-West)
Panel A: Average Returns					
Small	1.62	1.21	2.81	1.19	(1.37)
Medium	0.78	0.79	1.27	0.49	(0.79)
Big	0.95	1.14	1.12	0.17	(0.30)
Panel B: Average Size (market capitalisation)					
Small	0.169	0.130	0.097		
Medium	0.791	0.733	0.634		
Big	11.221	6.636	5.142		
Panel C: Average EDF					
Small	0.448	1.961	9.424		
Medium	0.135	0.471	3.932		
Big	0.077	0.233	3.500		
Panel D: Average BM					
Small	0.553	0.633	0.733		
Medium	0.458	0.508	0.690		
Big	0.397	0.500	0.650		

At the end of each month from February 2002 to December 2013, stocks are first sorted into three portfolios based on their past month's market capitalisation (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 capitalisation) in Panel B is expressed in billions of \$US, while average EDF in Panel C is expressed in percent units, and average BM in Panel D in decimal units.

EDF categories, independent of the market capitalisation of the stocks. Note that within the small size portfolio, the average EDF varies between 9.42% (for the high distress risk category) and 0.45% (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 lower 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. 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 7. Stocks are first sorted into three groups according to their BM ratio (high, medium, low), and subsequently within each of these BM groups we sort stock into three default portfolios (low, medium, high).

Panel A shows that there is statistically significant default effect (at the 10% level) present in the low BM-sorted portfolios, with a return differential of 1.45% per month (17.4% p.a.). On the other hand, the difference between high and low EDF portfolios for medium BM groups actually yields negative return. Next, we can see that the return difference between

Table 7

Default effect controlled by BM

EDF	Low	Medium	High	High-Low	t-stat (Newey-West)
Panel A: Average Returns					
High BM	1.32	1.61	1.95	0.63	(0.83)
Medium BM	1.16	1.67	0.44	-0.72	(-1.27)
Low BM	0.70	0.72	2.15	1.45	(1.84)*
Panel B: Average Size (market capitalisation)					
High BM	4.078	1.915	0.932		
Medium BM	6.142	2.947	1.232		
Low BM	6.836	1.417	0.632		
Panel C: Average EDF					
High BM	0.260	1.186	7.885		
Medium BM	0.116	0.416	2.815		
Low BM	0.101	0.490	6.370		
Panel D: Average BM					
High BM	0.832	0.987	1.352		
Medium BM	0.467	0.479	0.485		
Low BM	0.221	0.218	0.141		

At the end of each month from February 2002 to December 2013, 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 capitalisation) in Panel B is expressed in billions of \$US, while average EDF in Panel C is expressed in percent units, and average BM in Panel D in decimal units.

high and low BM portfolios is relatively small (no more than 0.89% in case of the medium EDF portfolios).

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. The smallest firms are typically firms with the lowest BM ratios and are contained in high EDF/low BM portfolios.

4.6 Pricing of distress risk

In this section, using the asset pricing models described in Section 4.3, 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 3, at the end of each month from January 2002 to November 2013 we sort the companies in our sample into deciles based on their EDF measures and form ten equally-weighted (EW) portfolios. For each month, portfolio 1 contains the 10% of companies with the lowest distress risk, while portfolio 10 contains the 10% of companies with the highest distress risk. We also construct the long-short portfolio that takes a long position in the 10% of stocks with the highest distress risk (these stocks will provide higher returns), and a short position in the 10% of stocks with the lowest distress risk (stocks providing lower returns). A key question 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, yield abnormal or active returns beyond what would be suggested by standard asset pricing models.

In Table 8 we report the results from regressions using the excess returns of equally-weighted (EW) portfolios. Panel A reports monthly alphas expressed in annualised percent units with respect to the CAPM (1), the Fama-French three-factor model (2), and the Carhart four-factor model (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 57). 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 then graphically summarises the behaviour of alphas across particular portfolios, while Figure 5 shows the evolution of factor loadings from the four-factor model across the created distress risk portfolios. We also provide correlation

Table 8

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

	Low	EDF								High	
EW portfolios	1	2	3	4	5	6	7	8	9	10	High-Low
Panel A: Portfolio Alphas											
CAPM alpha	1.44	6.19	9.14	2.65	7.71	-2.12	0.18	2.03	5.52	20.72	19.28
(t-stat)	(0.39)	(1.55)	(1.88)*	(0.64)	(1.48)	(-0.42)	(0.03)	(0.28)	(0.57)	(1.97)**	(1.85)*
3-factor alpha	0.36	4.54	7.23	1.11	5.79	-4.37	-2.38	0.67	2.03	16.62	16.25
(t-stat)	(0.10)	(1.21)	(1.68)*	(0.28)	(1.22)	(-0.87)	(-0.47)	(0.10)	(0.24)	(1.76)*	(1.70)*
4-factor alpha	-0.46	4.34	7.48	0.89	5.18	-4.39	-1.53	1.29	2.64	18.73	19.19
(t-stat)	(-0.13)	(1.13)	(1.75)*	(0.23)	(1.11)	(-0.88)	(-0.30)	(0.17)	(0.30)	(1.87)*	(1.92)*
Panel B: CAPM Regression Coefficients											
RM	1.067	1.082	1.358	1.286	1.385	1.516	1.476	1.757	1.755	2.240	1.174
(t-stat)	(15.59)***	(14.26)***	(12.57)***	(16.75)***	(16.67)***	(17.48)***	(12.79)***	(13.79)***	(10.21)***	(12.01)***	(6.30)***
Panel C: Three-factor Regression Coefficients											
RM	0.988	0.966	1.215	1.177	1.243	1.359	1.293	1.653	1.501	1.953	0.964
(t-stat)	(14.89)***	(10.68)***	(11.03)***	(13.63)***	(14.27)***	(14.21)***	(10.09)***	(11.13)***	(10.21)***	(9.92)***	(4.81)***
SMB	0.504	0.571	1.024	0.579	0.964	0.758	1.009	0.828	1.571	1.405	0.901
(t-stat)	(3.93)***	(3.46)***	(5.61)***	(3.75)***	(5.69)***	(4.22)***	(4.76)***	(3.00)***	(5.48)***	(4.01)***	(2.39)**
HML	-0.195	0.015	-0.548	-0.055	-0.452	0.051	-0.166	-0.541	-0.530	0.067	0.261
(t-stat)	(-1.52)	(0.08)	(-3.01)***	(-0.42)	(-2.87)***	(0.32)	(-0.77)	(-2.21)**	(-1.84)*	(0.16)	(0.64)
Panel D: Four-factor Regression Coefficients											
RM	1.096	0.993	1.182	1.205	1.324	1.363	1.181	1.570	1.421	1.675	0.578
(t-stat)	(13.85)***	(10.42)***	(9.61)***	(11.29)***	(12.99)***	(12.52)***	(9.04)***	(11.68)***	(9.19)***	(8.10)***	(2.95)***
SMB	0.461	0.561	1.037	0.568	0.932	0.756	1.053	0.861	1.603	1.514	1.053
(t-stat)	(3.67)***	(3.34)***	(5.45)***	(3.53)***	(5.46)***	(4.19)***	(5.00)***	(3.05)***	(5.46)***	(4.06)***	(2.90)***
HML	-0.158	0.024	-0.559	-0.045	-0.424	0.053	-0.204	-0.569	-0.558	-0.027	0.131
(t-stat)	(-1.13)	(0.12)	(-3.33)***	(-0.35)	(-2.66)***	(0.32)	(-1.01)	(-2.37)**	(-1.94)*	(-0.07)	(0.39)
MOM	0.210	0.052	-0.064	0.055	0.156	0.007	-0.217	-0.160	-0.157	-0.539	-0.750
(t-stat)	(2.92)***	(0.67)	(-0.49)	(0.51)	(1.88)**	(0.09)	(-1.49)	(-1.47)	(-1.02)	(-2.09)**	(-2.69)***
Panel E: R-squared											
CAPM	0.6365	0.5833	0.6031	0.6611	0.5895	0.6504	0.5613	0.5843	0.4585	0.4635	0.1867
3-factor	0.6732	0.6226	0.7013	0.6929	0.6673	0.6903	0.6238	0.6255	0.5509	0.5078	0.2174
4-factor	0.6987	0.6240	0.7027	0.6942	0.6750	0.6903	0.6363	0.6305	0.5546	0.5355	0.2959

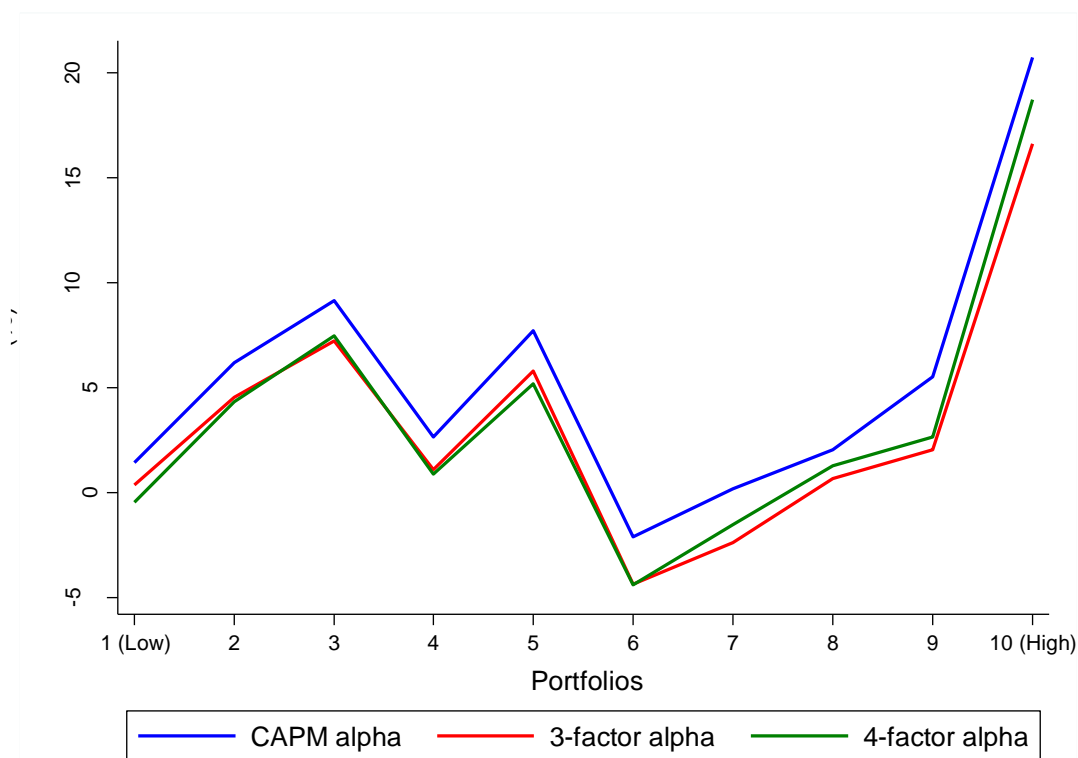
At the end of each month t from January 2002 to November 2013, we sort our companies into deciles based on their weighted EDF measures, as defined in (7). Portfolio 1 is the portfolio with the lowest default risk and portfolio 10 with the highest one. The “High-Low” column denotes a portfolio that takes a long position in the 10% stocks with high EDF and a short position in the 10% 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 2013) 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 annualised percent units) from these regressions and the corresponding values of t -statistics (in parenthesis). Panel B shows loadings on the market factor and the corresponding 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 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.

coefficients between raw returns and the applied factors in Table 9.

The risk-adjusted returns (alphas) corrected for given risk factors are reported in Panel A of Table 8. They are generally increasing across our portfolios, although this pattern is not monotonic. It is rather irregular between portfolios 1 to 6 (where they reach negative values), and they start substantially increasing and eventually reach relatively high values for portfolio 10, containing the 10% of renewable stocks with the highest EDF (see also in Figure 4). The “High-Low” strategy where we hold the riskiest decile of stocks and sell the decile of stocks with the lowest failure risk, provides positive returns from 16.25% to 19.28% p.a., depending on the applied asset pricing model. Note that, generally, results on alphas are also very stable with respect to the applied pricing models. The calculated annualised active returns are hardly affected by the choice of model. Overall, our results indicate again that the distress risk premium seems to be an effect that is mainly concentrated in the companies with substantial

Figure 4

Portfolio alphas from the regressions of excess returns



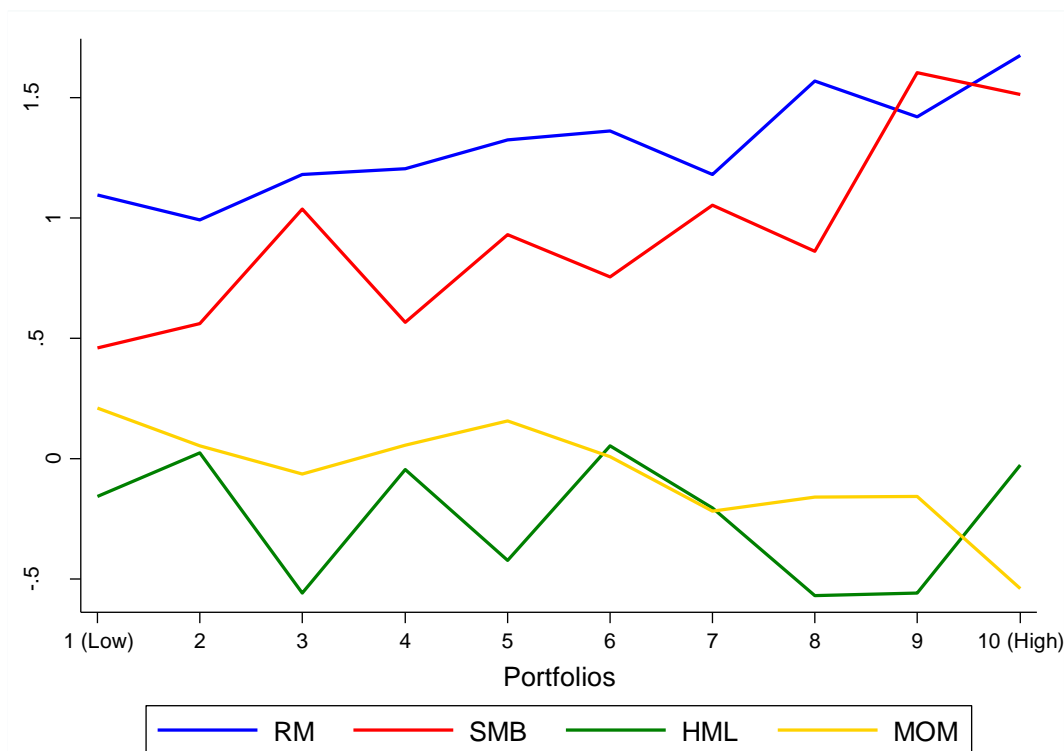
The figure plots monthly alphas (in annualised percent units) from the applied CAPM model, the Fama-French three-factor model, and the Carhart four-factor model for 10 distress risk-sorted equally-weighted (EW) portfolios from February 2002 to December 2013 (see panel A of Table 8). Portfolios are formed at the end of each month from January 2002 to November 2013, when we sort our companies into deciles based on their weighted EDF measures, as defined in (7).

default risk, that is, the highest decile. Only for these portfolios do we get active annualised returns of substantial magnitude (greater than 16%, depending on the applied model) that are at the same time statistically significant for all the models. Importantly, also note that the models yield statistically significant returns for our “High-Low” portfolio too.

Regarding the factor loadings that are reported in panels B, C, and D in Table 8, the market factor RM is increasing and statistically significant for all models. We can see that stocks in portfolio 10 (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 has irregular increasing pattern in both three-factor and four-factor models implying that the small companies prevail among distressed stocks. Also this factor is highly significant in all models. Finally, the value factor HML is irregular with no visible pattern, while the momentum factor MOM has a decreasing pattern with negative loadings on portfolios 3 and 7 to 10 (see

Figure 5

Factor loadings from the four-factor regression of excess returns



The figure plots loadings on excess market return (RM), size factor (SMB), value factor (HML), and momentum factor (MOM) from four-factor regression (see panel D of Tables 8) for 10 distress risk-sorted equally-weighted (EW) portfolios from February 2002 to December 2013. Portfolios are formed at the end of each month from January 2002 to November 2013, when we sort our companies into deciles based on their weighted EDF measures, as defined in (7).

also Figure 5). Both factors are significant only for some portfolios.

Reported R-squared measures from these regressions are relatively high, particularly for less risky portfolios, and generally decreasing with portfolios that hold more distressed stocks. The values start at 0.6365 – 0.6987 for portfolio 1 and end at 0.04635 – 0.5355 for portfolio 10. Note that for our “High-Low” strategy where we take a long position in the 10% of high-distress stocks and a short position in the 10% of safest stocks, the values drop significantly to 0.1867 – 0.2959. 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.

Thus, contrary to the findings of Campbell et al. (2008), 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 decile, that is, for companies with a relatively high probability of default.

From reported results in Table 9⁶⁵ we can see that the raw returns of our 10 portfolios are relatively highly correlated and exhibit correlations ranging from 0.59 to 0.80. Regarding the correlation between raw returns of our 10 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 the correlation is generally decreasing for portfolios that hold more distressed stocks. The correlation coefficient for portfolio 1 has a value of 0.80 compared to a value of 0.68 for portfolio 10. 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 around 0.5. Again, this confirms the relationship between distress risk and the size effect that we pointed out earlier. Finally, there is a very high correlation of 0.92 for raw returns of portfolio 10 (high risk) and our “High-Low” strategy,

⁶⁵ We provide a distribution plot that shows a linearity of the data (scatterplot matrix) in Appendix D.

Table 9

Correlation coefficients between raw returns and given factors

	Low			EDF						High						
	1	2	3	4	5	6	7	8	9	10	High-Low	RM	SMB	HML	MOM	
1 (Low)	1															
2	0.70	1														
3	0.74	0.71	1													
4	0.75	0.73	0.77	1												
5	0.76	0.70	0.75	0.77	1											
6	0.75	0.72	0.74	0.80	0.77	1										
7	0.71	0.68	0.77	0.74	0.75	0.79	1									
8	0.72	0.64	0.71	0.73	0.74	0.71	0.75	1								
9	0.73	0.66	0.71	0.69	0.72	0.71	0.70	0.70	1							
10 (High)	0.60	0.59	0.65	0.65	0.68	0.64	0.73	0.68	0.68	1						
High-Low	0.23	0.36	0.42	0.41	0.44	0.40	0.54	0.47	0.46	0.92	1					
RM	0.80	0.76	0.78	0.81	0.77	0.81	0.75	0.76	0.68	0.68	0.43	1				
SMB	0.46	0.47	0.54	0.47	0.52	0.49	0.51	0.44	0.52	0.45	0.32	0.38	1			
HML	0.12	0.19	0.04	0.17	0.07	0.21	0.14	0.06	0.07	0.18	0.16	0.21	0.18	1		
MOM	-0.18	-0.28	-0.34	-0.31	-0.23	-0.33	-0.40	-0.37	-0.32	-0.43	-0.43	-0.43	-0.09	-0.16	1	

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

confirming that it is predominantly the high distressed stocks that play a crucial role in our long-short strategy.

As another check of whether a distress risk is priced in the cross section of equity returns, we estimate additional multivariate regression models (using the Fama-MacBeth (1973) two-stage approach) where we consider a plausible empirical asset-pricing specification in which distress risk appears as a factor. First, we use time-series regressions to estimate the betas on the CAPM, the three-factor Fama-French (1993) model and the four-factor Carhart (1997), while adding the distress factor DS to all these specifications. In the second stage, we implement a cross-sectional regression of the average returns of all 141 stocks on their betas. The empirical asset-pricing specifications are given below:

$$R_{i,t} - R_{F,t} = \alpha_i^{Model1} + \beta_{1,i} (R_{M,t} - R_{F,t}) + \beta_{2,i} DS_t + \varepsilon_{i,t}, \quad (8)$$

$$R_{i,t} - R_{F,t} = \alpha_i^{Model2} + \beta_{1,i} (R_{M,t} - R_{F,t}) + \beta_{2,i} SMB_t + \beta_{3,i} HML_t + \beta_{4,i} DS_t + \varepsilon_{i,t}, \quad (9)$$

$$R_{i,t} - R_{F,t} = \alpha_i^{Model3} + \beta_{1,i} (R_{M,t} - R_{F,t}) + \beta_{2,i} SMB_t + \beta_{3,i} HML_t + \beta_{4,i} MOM_t + \beta_{5,i} DS_t + \varepsilon_{i,t}. \quad (10)$$

We use two different specifications for our distress risk factor DS . First, we follow standard conventions for constructing risk factors (see, for example, Fama and French (1993)) and use our “High-Low” strategy, that is, we use returns for the portfolio that is formed by taking a long position in the 10% of stocks with the highest distress risk and a short position in the 10% of stocks with the lowest distress risk. Second, we follow Vassalou and Xing (2004) and define the aggregate survival rate SV as $1 - \overline{EDF}_t$ where \overline{EDF}_t is defined as a simple

Table 10

The Fama-MacBeth regression with the distress risk factor

	Model 1		Model 2		Model 3	
alpha	8.63	7.36	14.52	14.45	18.39	22.44
(t-stat)	(1.35)	(1.19)	(2.30)**	(2.52)**	(3.43)***	(3.59)***
RM	0.0029	0.0013	0.0017	-0.0003	-0.0023	-0.0045
(t-stat)	(0.90)	(0.43)	(0.49)	(-0.10)	(-0.90)	(-1.43)
SMB			-0.0019	-0.0034	-0.0017	-0.0054
(t-stat)			(-0.94)	(-2.06)**	(-0.90)	(-2.79)***
HML			0.0033	0.0035	0.0051	0.0022
(t-stat)			(3.03)***	(3.12)***	(5.23)***	(1.70)*
MOM					-0.0105	-0.0057
(t-stat)					(-6.49)***	(-2.47)**
High-Low	-0.0066		-0.0124		-0.0063	
(t-stat)	(-0.88)		(-1.67)*		(-0.95)	
$\Delta(SV)$		0.0004		0.0005		0.0002
(t-stat)		(4.11)***		(5.81)***		(1.54)
R-squared	0.0147	0.1140	0.0975	0.2601	0.3113	0.2443

The table reports results from the second step of the Fama-MacBeth regression where the average returns of all 141 stocks are regressed on their betas (factor loadings). Model 1 represents the CAPM model, Model 2 the Fama-French three-factor model, and Model 3 the Carhart four-factor model, all including the additional distress risk factor. The distress risk factor is represented either by our “High-Low” strategy, or by the aggregate survival rate $\Delta(SV)_t$. Alphas are expressed in annualised percent units. The t -stat (in parenthesis) is the t -statistic of the corresponding estimated coefficient. *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

average of the transformed EDF measure (Equation 7) of all firms at given month t . Our distress risk factor is then defined as the change in aggregate survival rate $\Delta(SV)_t$ at time t which is given by $SV_t - SV_{t-1}$.

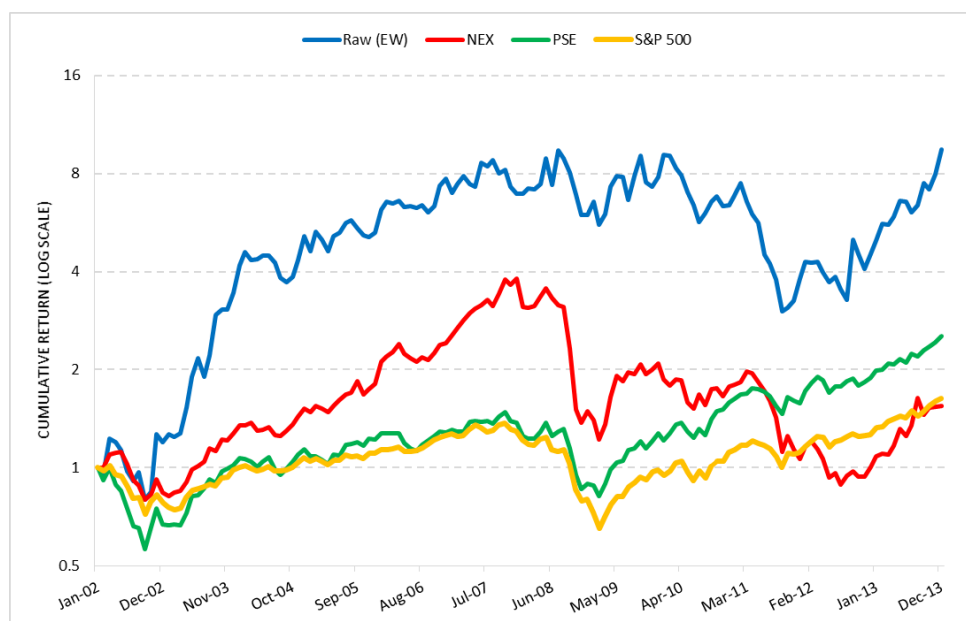
Results from the second step of the Fama-MacBeth regression are reported in Table 10. We can see that the market factor RM is insignificant in all three models, while the value factor HML and the momentum factor MOM are statistically significant. The size factor SMB appears to be significant only when the aggregate survival rate $\Delta(SV)$ is used as the distress factor. More importantly, the $\Delta(SV)$ factor is statistically significant for Models 1 and 2 at the 1% level. Note, however, that the ‘High-Low’ factor is only significant at the 10% level in Model 2. Overall, our results provide additional evidence for distress risk being priced in the cross-section of returns for renewable energy companies.

Finally, 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 6 with a graphical comparison between cumulative raw returns on the “High–Low” equally-weighted (EW) portfolio and cumulative returns on the NEX, PSE, and S&P 500 indices over the period from February 2002 to December 2013. We can see that raw returns from the created distress risk portfolios clearly outperform these three indices throughout the sample period.

We also report correlations between excess returns of EW “High–Low” portfolios, chosen indices (S&P 500, PSE, NEX), and the WTI crude oil price over the period from February 2002 to December 2013 in Table 11. 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 (0.75), 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 0.48 between EW and PSE). We also find that correlation between excess returns from our

Figure 6

Cumulative raw returns on “High–Low” EW portfolio and chosen indices



The figure plots cumulative raw returns on “High–Low” equally-weighted (EW) portfolio that takes a long position in the 10% most distressed stocks and a short position in the 10% 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 2013.

Table 11

Correlation coefficients between excess returns

	Raw (EW)	S&P 500	PSE	NEX	WTI
Raw (EW)	1				
S&P 500	0.42	1			
PSE	0.48	0.90	1		
NEX	0.41	0.75	0.73	1	
WTI	0.08	0.27	0.24	0.43	1

The table reports correlation coefficients between excess returns on equally-weighted (EW) "High-Low" portfolio, that takes a long position in the 10% most distressed stocks and a short position in the 10% safest stocks, excess returns of 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 2013.

long-short strategy and excess returns from WTI crude oil prices is very low (value of 0.08). While correlations between NEX or ECO index returns and returns from the WTI are typically quite pronounced (between 0.43 and 0.52), the identified distress risk premium for the renewable sector does not seem to be influenced 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 on cumulative returns and correlations using the ECO instead of the NEX for the period from August 2004 to December 2013 in Appendix E and F.⁶⁶ Overall, results are qualitatively the same as for the NEX.

Table 12

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

	Equally-weighted (EW) "High-Low" portfolio						
	Mean	Std.	Min	Max	Median	Quart 1	Quart 3
α -CAPM	0.0161	0.1094	-0.2406	0.4960	0.0024	-0.0528	0.0699
α -3F	0.0135	0.1073	-0.2595	0.4932	-0.0010	-0.0548	0.0670
α -4F	0.0160	0.1018	-0.2683	0.4974	0.0084	-0.0443	0.0691

The table reports descriptive statistics of monthly risk-adjusted returns (alphas) on "High-Low" equally-weighted (EW) portfolios. These portfolios take a long position in the 10% most distressed stocks and a short position in the 10% safest stocks over the period from February 2002 to December 2013. 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 decimal units.

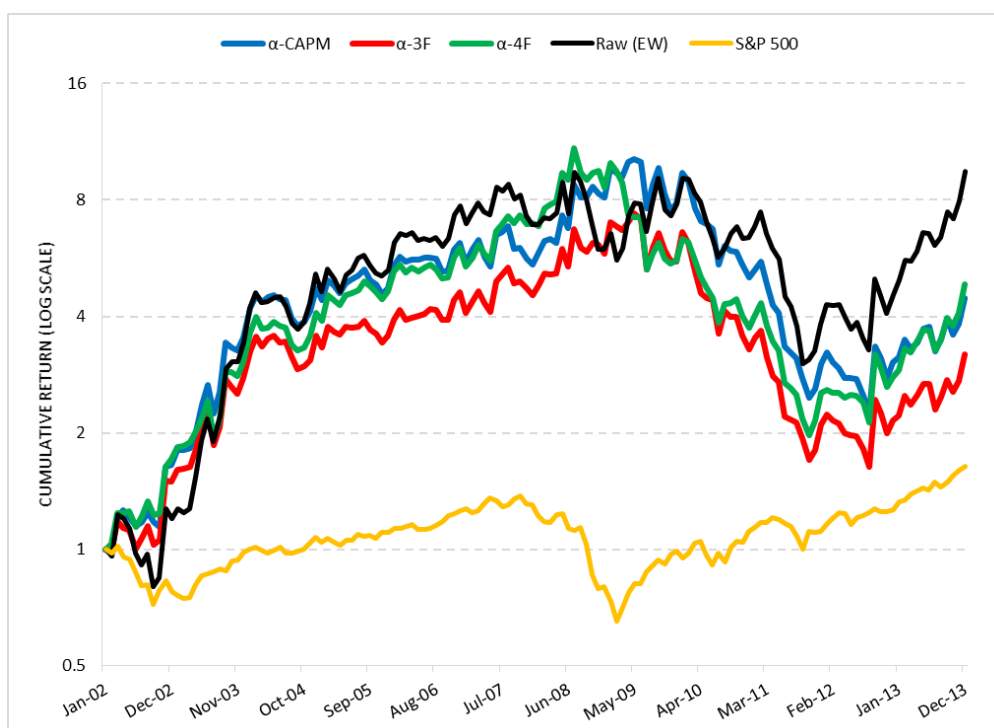
⁶⁶ The ECO index is only available from August 2004.

Furthermore, we report descriptive statistics of monthly risk-adjusted active returns (alphas) on our “High–Low” EW portfolio over the period February 2002 to December 2013 in Table 12. 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 1.35% to 1.61% 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 Table 8 (reported in annualised percent units). We observe that the active returns for the created portfolios are relatively volatile, with monthly standard deviations ranging from 10.18% to 10.94%.

Moreover, we illustrate the cumulative performance of these risk-adjusted returns for the EW long-short portfolio in Figure 7. For comparison purposes, we also plot cumulative

Figure 7

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 10% most distressed stocks and a short position in the 10% 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 2013. For comparison purposes, the figure also plots cumulative return on the S&P 500 index.

returns of the S&P 500 index. We find that cumulative raw returns of our “High–Low” EW 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 long-short portfolios. As illustrated by Figure 7, the performance of the cumulative risk-adjusted active returns is also above the cumulative performance of the S&P 500 index throughout our sample period (2002–2013). Note, however that the performance of the “High–Low” portfolio shows significant variations through time. Recall that earlier studies by, e.g. Bohl et al. (2013), emphasize high active returns in the renewable sector for the period up to 2008, and a relatively bad performance of the sector afterwards. Our findings for estimated alphas, i.e. active returns, of the examined “High–Low” portfolio, also suggest high active returns up to 2009, while during the subsequent period of low oil prices, the distress risk “High–Low” portfolio even yields negative returns. However, from mid-2012 onwards up to the end of the sample period in December 2013, active returns become highly positive again and point towards a significant pricing of distress risk in the renewable sector.

Overall, these results suggest that distress risk is particular priced during periods when the oil price is rising. Further, the excellent performance of the “High–Low” EW portfolio from a risk-adjusted perspective is also confirmed in this section. We conclude that in the renewable energy sector distress risk seems to be systematically priced and investors are compensated for taking on additional risks by investing in stocks with relatively high default probabilities.

4.7 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 achieved tremendous growth rates in the global economy, mostly due to the conjunction of rising oil prices, increasing market liquidity for investments in the 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), and the S&P Global

Clean Energy Index (SPGCE). At the same time, companies involved in the renewable energy business are relatively high risk firms with high profitability potential.

In this paper, 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 realised returns of U.S. renewable energy companies, and examine risk-adjusted returns corrected for common Fama and French (1993) and Carhart (1997) 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 realised equity returns of equally-weighted (EW) portfolios and distress risk in the renewable energy sector. Thus, we confirm findings of Vassalou and Xing (2004) and Chava and Purnanandam (2010) 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 EW portfolio consisting of the riskiest decile of stocks and those consisting of the decile 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 (1993) three-factor model, and the Carhart (1997) 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 (2004), 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 paper is particularly closely related to Bohl et al. (2013), 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 EW portfolio that take a long position in the 10% most distressed stocks, and a short position in the 10% safest stocks, outperforms the S&P 500 index throughout our sample period (2002–2013). 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 as are 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.

Bibliography

- ALTMAN, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23, 589-609.
- AVRAMOV, D., CHORDIA, T., JOSTOVA, G. & PHILIPOV, A. 2009. Credit ratings and the cross-section of stock returns. *Journal of Financial Markets*, 12, 469-499.
- BETZER, A., DOUMET, M. & RINNE, U. 2013. How policy changes affect shareholder wealth: the case of the Fukushima Dai-ichi nuclear disaster. *Applied Economics Letters*, 20, 799-803.
- BOHL, M. T., KAUFMANN, P. & SIKLOS, P. L. 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.
- BOHL, M. T., KAUFMANN, P. & STEPHAN, P. M. 2013. From hero to zero: Evidence of performance reversal and speculative bubbles in german renewable energy stocks. *Energy Economics*, 37, 40-51.
- BOYER, M. M. & FILION, D. 2007. Common and fundamental factors in stock returns of Canadian oil and gas companies. *Energy Economics*, 29, 428-453.

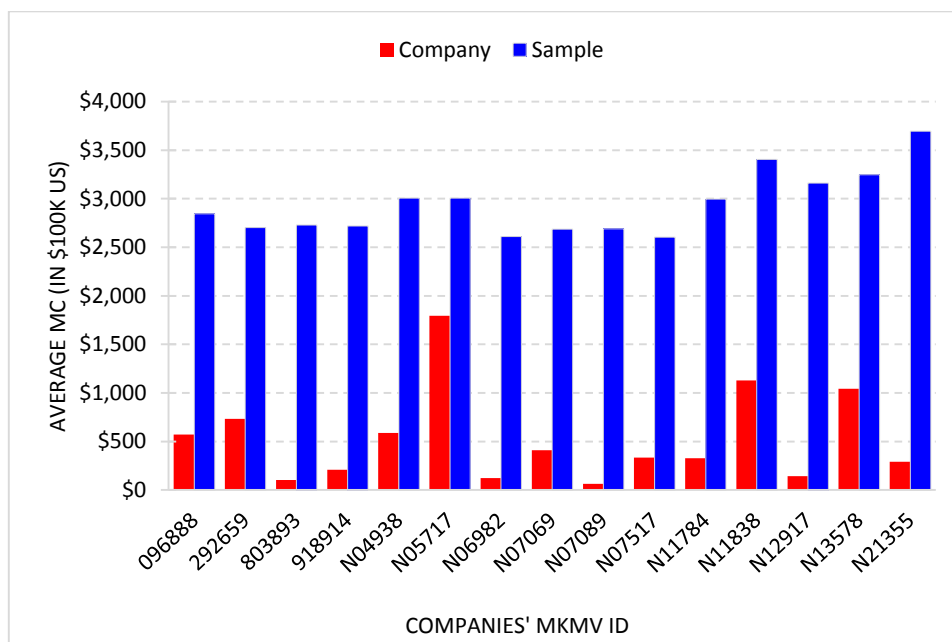
- BROADSTOCK, D. C., CAO, H. & ZHANG, D. 2012. Oil shocks and their impact on energy related stocks in China. *Energy Economics*, 34, 1888-1895.
- CAMPBELL, J. Y., HILSCHER, J. & SZILAGYI, J. 2008. In search of distress risk. *The Journal of Finance*, 63, 2899-2939.
- CARHART, M. M. 1997. On persistence in mutual fund performance. *The Journal of Finance*, 52, 57-82.
- CHAN, K. & CHEN, N.-F. 1991. Structural and return characteristics of small and large firms. *The Journal of Finance*, 46, 1467-1484.
- CHAVA, S. & PURNANANDAM, A. 2010. Is default risk negatively related to stock returns? *Review of Financial Studies*, 23, 2523-2559.
- CUMMINS, M., GARRY, O. & KEARNEY, C. 2014. Price discovery analysis of green equity indices using robust asymmetric vector autoregression. *International Review of Financial Analysis*, 35, 261-267.
- DICHEV, I. D. 1998. Is the risk of bankruptcy a systematic risk? *The Journal of Finance*, 53, 1131-1147.
- FAFF, R. W. & BRAILSFORD, T. J. 1999. Oil price risk and the Australian stock market. *Journal of Energy Finance & Development*, 4, 69-87.
- FAMA, E. F. & FRENCH, K. R. 1992. The cross-section of expected stock returns. *The Journal of Finance*, 47, 427-465.
- FAMA, E. F. & FRENCH, K. R. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.
- FAMA, E. F. & MACBETH, J. D. 1973. Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, 81, 607-636.
- FERSTL, R., UTZ, S. & WIMMER, M. 2012. The effect of the Japan 2011 disaster on nuclear and alternative energy stocks worldwide: an event study. *BuR-Business Research*, 5, 25-41.
- GARLAPPI, L., SHU, T. & YAN, H. 2008. Default risk, shareholder advantage, and stock returns. *Review of Financial Studies*, 21, 2743-2778.
- GARLAPPI, L. & YAN, H. 2011. Financial Distress and the Cross-section of Equity Returns. *The Journal of Finance*, 66, 789-822.
- GRIFFIN, J. M. & LEMMON, M. L. 2002. Book-to-market equity, distress risk, and stock returns. *The Journal of Finance*, 57, 2317-2336.
- HENRIQUES, I. & SADORSKY, P. 2008. Oil prices and the stock prices of alternative energy companies. *Energy Economics*, 30, 998-1010.
- INCHAUSPE, J., RIPPLE, R. D. & TRÜCK, S. 2015. The dynamics of returns on renewable energy companies: A state-space approach. *Energy Economics*, 48, 325-335.
- KUMAR, S., MANAGI, S. & MATSUDA, A. 2012. Stock prices of clean energy firms, oil and carbon markets: A vector autoregressive analysis. *Energy Economics*, 34, 215-226.
- 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, 13-37.
- LOPATTA, K. & 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.

- MANAGI, S. & 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.
- MARKOWITZ, H. 1952. Portfolio selection. *The Journal of Finance*, 7, 77-91.
- MARKOWITZ, H. 1959. *Portfolio Selection: Efficient Diversification of Investments*, New York, USA, John Wiley & Sons.
- MERTON, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29, 449-470.
- MOSSIN, J. 1966. Equilibrium in a capital asset market. *Econometrica: Journal of the Econometric Society*, 34, 768-783.
- OHLSON, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18, 109-131.
- ORTAS, E. & MONEVA, J. M. 2013. The Clean Techs equity indexes at stake: Risk and return dynamics analysis. *Energy*, 57, 259-269.
- REBOREDO, J. C. 2015. Is there dependence and systemic risk between oil and renewable energy stock prices? *Energy Economics*, 48, 32-45.
- SADORSKY, P. 2001. Risk factors in stock returns of Canadian oil and gas companies. *Energy Economics*, 23, 17-28.
- SADORSKY, P. 2012a. Correlations and volatility spillovers between oil prices and the stock prices of clean energy and technology companies. *Energy Economics*, 34, 248-255.
- SADORSKY, P. 2012b. Modeling renewable energy company risk. *Energy Policy*, 40, 39-48.
- SHARPE, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19, 425-442.
- SHUMWAY, T. 2001. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74, 101-124.
- VASSALOU, M. & XING, Y. 2004. Default risk in equity returns. *The Journal of Finance*, 59, 831-868.
- WEN, X., GUO, Y., WEI, Y. & HUANG, D. 2014. How do the stock prices of new energy and fossil fuel companies correlate? Evidence from China. *Energy Economics*, 41, 63-75.
- WEYANT, J. P. 2011. Accelerating the development and diffusion of new energy technologies: Beyond the "valley of death". *Energy Economics*, 33, 674-682.

Appendix A – Defaulted Companies

Figure A1

Defaulted companies (average size)



The figure compares the average size, expressed by market capitalisation (MC) in \$100K US, 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.

Table A1

Defaulted companies

MKMVID	Company name	Date of Chapter 11 filing	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 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.

In Figure A1, we can see that all 15 defaulted companies are significantly smaller compared to average size of the sample. Reported information from Table A1 show that average EDF values for these companies are considerably high with expected increasing trend in EDFs, and decreasing trend in stock prices towards default. In fact, most of the time, the first maximum EDF value (EDF = 35) reported in the MKMV database predate both the date of filing for bankruptcy protection under Chapter 11 and the date of delisted return in CRSP database.

Appendix B – Acquired Companies

Generally, there are two possible reasons why a given company may be acquired. 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 investigate acquired companies in our sample and probable reason for the acquisition in the following table (Table A2).

Table A2
Acquired companies

MKMVID	Company 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 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.

On 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 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 imply that the company was saved from likely going bankrupt by this acquisition by H.I.G. Capital.

Appendix C

The core concept of the Merton (1974) model is to treat company's equity and debt as a contingent claim written on a 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 . 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:

$$A_t = E_t + D_t. \quad (\text{A1})$$

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

$$E_T = \max[A_T - L, 0]. \quad (\text{A2})$$

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

$$dA_t = \mu A_t dt + \sigma A_t dW_t, \quad (\text{A3})$$

where μ is the expected return (drift coefficient), σ 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 (A3) as follows:

$$A_T = A_t \exp \left[\left(\mu - \frac{1}{2} \sigma^2 \right) (T - t) + \sigma \sqrt{(T - t)} W_t \right], \quad (\text{A4})$$

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

In accordance with the Black-Scholes option pricing theory (1973), the Merton model stipulates that the company's equity value satisfies the following equation for pricing the call option within a risk neutral framework

$$E_t = A_t \Phi(d_1) - L e^{-r(T-t)} \Phi(d_2), \quad (\text{A5})$$

where

$$d_1 = \frac{\ln\left(\frac{A_t}{L}\right) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{(T-t)}}, \quad (A6)$$

$$d_2 = d_1 - \sigma\sqrt{(T-t)}, \quad (A7)$$

and r denotes the risk-free interest rate and $\Phi(\cdot)$ the cumulative distribution function of the standard normal variable. Equation (A7) is referred to as the *distance-to-default* (DD) by Moody's KMV. The larger the DD, the lower is the probability that the company will default.

We can estimate PD by rearranging (A4) as follows:

$$\begin{aligned} PD_t &= P[A_t \leq L] \\ &= P\left[\ln(A_t) + \left(\mu - 0.5\sigma^2\right)(T-t) + \sigma\sqrt{(T-t)}W_t \leq \ln(L)\right] \\ &= \int_{-\infty}^{-y} \phi(x)dx, \end{aligned} \quad (A8)$$

where

$$y = \frac{\ln(A_t / L) + \left(\mu - 0.5\sigma^2\right)(T-t)}{\sigma\sqrt{(T-t)}}$$

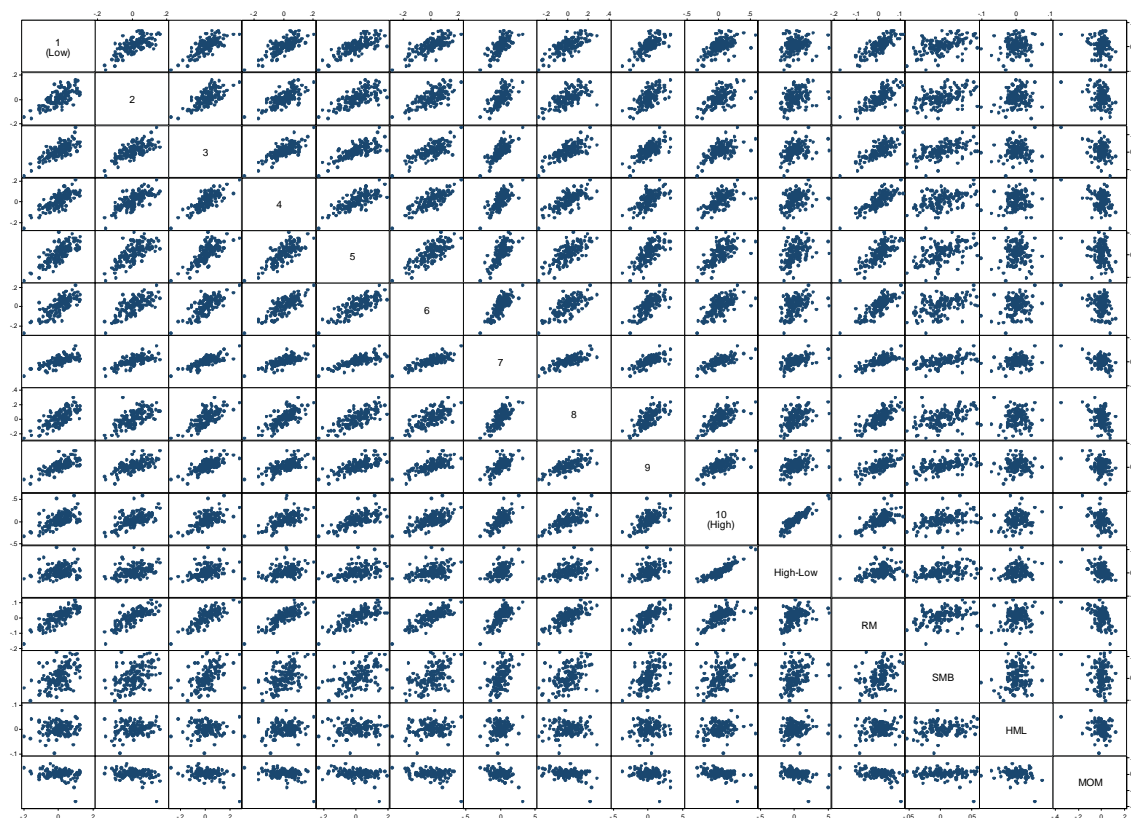
and ϕ is the probability density function of a standard normal variable. Note that unlike (A8), (A5) is not a function of μ but it is a function of r (we would get PD under the risk neutral probability measure). When we estimate PD, the risk-free interest rate r has to be replaced with real company drift μ 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

$$\begin{aligned} PD_t &= \Phi(-\hat{d}_2) \\ &= \Phi\left(-\frac{\ln\left(\frac{A_t}{L}\right) + \left(\mu - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{(T-t)}}\right). \end{aligned} \quad (A9)$$

Appendix D

Figure A2

Scatterplot matrix of equally-weighted (EW) raw returns and given factors

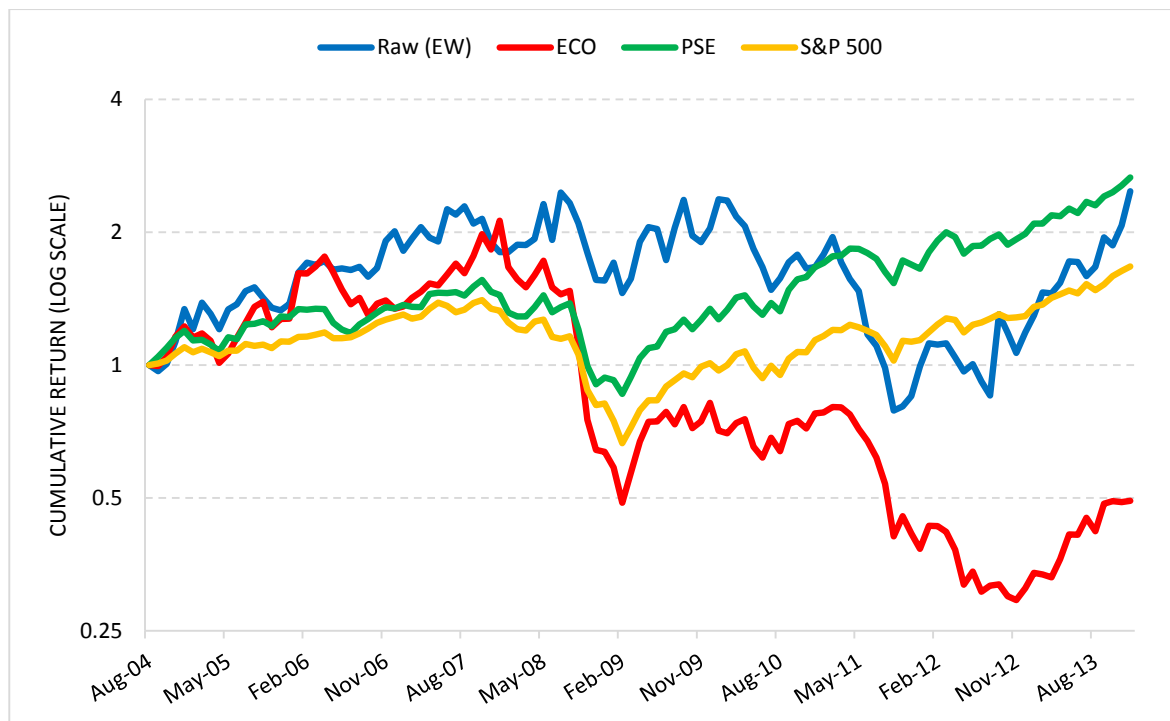


This scatterplot matrix identifies the type of relationship between raw returns of our created portfolios (1 to 10 and High-Low) and four factors from the correlation analysis conducted in Table 9. We can see that the relationship is close to being linear usually for our portfolio returns (portfolios 1 to 10) and the market factor RM. This is also confirmed by the estimated correlation coefficients in Table 9 that have relatively higher values. Note that for the remaining three factors (SMB, HML, MOM) and our High-Low portfolio we find a rather non-linear relationship with returns of the created decile portfolios and between each other. An exception is the linear relationship between portfolio 10 and the High-Low portfolio that yields a correlation coefficient of 0.92. This further confirms our finding that returns of our High-Low portfolio are mostly driven by returns in portfolio 10.

Appendix E

Figure A3

Cumulative raw returns on "High-Low" EW portfolio and chosen indices (Aug 04 – Dec 13)



The figure plots cumulative raw returns on "High-Low" equally-weighted (EW) portfolio that takes a long position in the 10% most distressed stocks and a short position in the 10% 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 2013.

Appendix F

Table A3

Correlation coefficients between excess returns (Sep 04 - Dec 13)

	Raw (EW)	S&P 500	PSE	ECO	WTI
Raw (EW)	1				
S&P 500	0.36	1			
PSE	0.40	0.91	1		
ECO	0.43	0.77	0.81	1	
WTI	0.10	0.43	0.44	0.52	1

The table reports correlation coefficients between excess returns on an equally-weighted (EW) "High-Low" portfolio, that takes a long position in the 10% most distressed stocks and a short position in the 10% safest stocks, excess returns of S&P 500, PSE (Arca Tech 100), and ECO indices, and excess returns on the U.S. WTI crude oil. Since the ECO index is only available from September 2004, considered period is September 2004 - December 2013.

5 Summary and Conclusions

This PhD thesis is addressing several dimensions for estimation and examination of default probabilities in credit risk management. This area of research has undergone substantial development in recent decades and has 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 behind the emergence of the recent global financial crises as undervaluation of default risk for mortgages and structured credit products caused significant stress on the financial system which had been extended through credit derivatives on global markets.

The thesis consists of three research papers. One of the most significant approaches for the estimation of default probabilities are structural credit risk models. This approach was originally introduced by Merton (1974) and is based on the idea of treating a company's equity and debt as a contingent claim written on the company's asset value. Significant attention has been given to this framework in the 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 the first paper (*Structural Credit Risk Models with Subordinated Processes*), by demonstrating that the distributional assumption of the Merton model (company value follows a log-normal distribution) is rejected, we have confirmed several empirical investigations that have shown that log-returns of equities typically follow skewed distributions with excess kurtosis, which leads to a greater density in the tails. 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 framework. In particular, we have proposed to use the Hurst et al. (1999) option

pricing model based on stable Paretian distributions, which generalises 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 model, and thus is a generalisation of the Merton one. The first relevant desirable property of the stable distributional assumption is that stable distributions have a domain of attraction. The generalised central limit theorem for the normalised 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 (1967), and stable laws then have been applied in several financial sectors. For these reasons, the stable Paretian law is the first candidate used as a subordinated model.

We have proposed two different methodologies for the parameter estimation: the first is to generalise the maximum likelihood parameter estimation proposed by Duan (1994); the second is a generalisation of the Moody's KMV methodology. Moreover, we have optimised 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 PDs are generally underestimated by the Merton model and that the stable Lévy model is substantially more sensitive to periods of financial crises. We have also referred to a study conducted by Brambilla et al. (2015) that extended our work and applied two alternative structural credit risk models based on well-known Lévy processes (the Variance Gamma (VG) process and the Normal Inverse Gaussian (NIG) process). These authors conclude that both models are able to capture the situation of

instability that affects each company in the 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 is that structural credit risk models based on the Merton framework are not appropriate for the estimation of PDs for financial institutions unless some adjustments are made. This is the reason why we have devoted our second paper (*Prediction of U.S. Commercial Bank Failures via Scoring Models: The FFIEC Database Case*) to the 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 paper was threefold. First, we have taken 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. A sufficient number of historical defaults is essential for estimating such models. In fact, our sample contained more than 7,000 U.S. commercial banks with 405 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 et al. (2005) worked with 40 privately owned Turkish commercial banks and 21 defaults; or Kolari et al. (2002) used over 1,000 large U.S. commercial banks in each year, with 55 defaults in total. Following a 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 the 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 practice, 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, a substantial part of this paper was devoted to the application of various model evaluation techniques, including techniques that have not previously been applied in the literature on credit scoring. 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 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 (2007), we have applied the Vuong 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 build 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 multiple comparison test to examine significant differences between the particular models in terms of bootstrapped ROC areas. The Tukey test has really 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 paper with all of the available information on U.S. commercial banks, and thus avoided choice-based samples within our 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 in the first two papers, where we have focused on the estimation of default probabilities, in our last paper (*Distress Risk and Stock Returns of U.S. Renewable Energy Companies*), we have taken advantage of the Moody's KMV database and used its structural-based default probability indicators (Expected Default Frequencies – EDFs) in an 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.

The renewable energy sector is considered to be a relatively risky sector with high profitability potential, similar to the 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

(2004) and Chava and Purnanandam (2010), found a positive cross-sectional relationship between distress risk and returns. Several other key studies (e.g. Dichev (1998), Campbell et al. (2008), or Garlappi et al. (2008)) suggest that distress risk is priced negatively - more distress stocks usually earn lower average returns (often referred to as a “distress risk puzzle”). Nevertheless, none of these studies was applied directly to the renewable energy sector.

We have combined two streams of the literature in this paper. Apart from studies that describe the relationship between distress risk and equity returns, we have also contributed to the literature that investigates returns on the 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 paper has provided the 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 2013 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 equally-weighted (EW) 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 findings from studies conducted by Vassalou and Xing (2004) and Chava and Purnanandam (2010). We have found a significant difference between returns of EW portfolios consisting of the riskiest decile of stocks, and one consisting of the decile 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 (1993) three-factor model, and the Carhart (1997) four-factor model. These models correct observed returns of the constructed portfolios for given risk factors (market return, size premium, value premium, momentum). We have concluded that, in the renewable energy sector, distress risk seems to be systematically priced and investors are compensated for taking on additional risks by investing in stocks with relatively high default probabilities. Finally, we have shown that raw and risk-adjusted returns of EW portfolios that take a long position in the 10% most distressed renewable stocks, and a short position in the 10% renewable companies with the lowest default risk, generally outperform the S&P 500 index throughout our sample period (2002–2013).

Bibliography

- BRAMBILLA, C., GURNY, M. & ORTOBELLI LOZZA, S. 2015. Structural Credit Risk Models with Lévy Processes: The VG and NIG Cases. *Far East Journal of Mathematical Sciences*, 97, 101-119.
- CAMPBELL, J. Y., HILSCHER, J. & SZILAGYI, J. 2008. In search of distress risk. *The Journal of Finance*, 63, 2899-2939.
- CANBAS, S., CABUK, A. & KILIC, S. B. 2005. Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The Turkish case. *European Journal of Operational Research*, 166, 528-546.
- CARHART, M. M. 1997. On persistence in mutual fund performance. *The Journal of Finance*, 52, 57-82.
- CHAVA, S. & PURNANANDAM, A. 2010. Is default risk negatively related to stock returns? *Review of Financial Studies*, 23, 2523-2559.

- DICHEV, I. D. 1998. Is the risk of bankruptcy a systematic risk? *The Journal of Finance*, 53, 1131-1147.
- DUAN, J. C. 1994. Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, 4, 155-167.
- FAMA, E. F. & FRENCH, K. R. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.
- GARLAPPI, L., SHU, T. & YAN, H. 2008. Default risk, shareholder advantage, and stock returns. *Review of Financial Studies*, 21, 2743-2778.
- HURST, S. R., PLATEN, E. & RACHEV, S. T. 1999. Option pricing for a logstable asset price model. *Mathematical and Computer Modelling*, 29, 105-119.
- KOLARI, J., GLENNON, D., SHIN, H. & CAPUTO, M. 2002. Predicting large US commercial bank failures. *Journal of Economics and Business*, 54, 361-387.
- MANDELBROT, B. & TAYLOR, H. M. 1967. On the distribution of stock price differences. *Operations Research*, 15, 1057-1062.
- MERTON, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29, 449-470.
- STEIN, R. M. 2007. Benchmarking default prediction models: Pitfalls and remedies in model validation. *Journal of Risk Model Validation*, 1, 77-113.
- VASSALOU, M. & XING, Y. 2004. Default risk in equity returns. *The Journal of Finance*, 59, 831-868.

Bibliography

- ACHARYA, V. V. & CARPENTER, J. N. 2002. Corporate bond valuation and hedging with stochastic interest rates and endogenous bankruptcy. *Review of Financial Studies*, 15, 1355-1383.
- ACHARYA, V. V., HUANG, J. Z., SUBRAHMANYAM, M. & SUNDARAM, R. K. 2006. When does strategic debt-service matter? *Economic Theory*, 29, 363-378.
- ALAM, P., BOOTH, D., LEE, K. & 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, 185-199.
- ALTMAN, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23, 589-609.
- ALTMAN, E. I., AVERY, R. B., EISENBEIS, R. A. & SINKEY, J. F. 1981. *Application of Classification Techniques in Business, Banking and Finance*, Greenwich, USA, JAI press.
- ALTMAN, E. I., HALDEMAN, R. G. & NARAYANAN, P. 1977. ZETA™ analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1, 29-54.
- ALTMAN, E. I., RIJKEN, H., BALAN, D., MINA, J., FORERO, J. & WATT, M. 2010. The Z-Metrics™ Methodology for Estimating Company Credit Ratings and Default Risk Probabilities. RiskMetrics Group. <https://www.fdic.gov/regulations/reform/altman2.pdf> [April 11, 2016].
- ANDERSON, R. W. & CARVERHILL, A. P. 2007. Liquidity and capital structure. Working paper No. 6044. Center for Economic Policy Research. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=997024 [April 11, 2016].
- ARORA, N., BOHN, J. R. & ZHU, F. 2005. Reduced form vs. structural models of credit risk: A case study of three models. *Journal of Investment Management*, 3, 43.
- ASVANUNT, A., BROADIE, M. & SUNDARESAN, S. 2011. Managing Corporate Liquidity: Strategies and Pricing Implications. *International Journal of Theoretical and Applied Finance*, 14, 369-406.
- AVRAMOV, D., CHORDIA, T., JOSTOVA, G. & PHILIPOV, A. 2009. Credit ratings and the cross-section of stock returns. *Journal of Financial Markets*, 12, 469-499.
- BAESENS, B., VAN GESTEL, T., VIAENE, S., STEPANOVA, M., SUYKENS, J. & VANTHIENEN, J. 2003. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54, 627-635.
- BALCAEN, S. & OOGHE, H. 2006. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38, 63-93.
- BEAVER, W. H. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71-111.
- BEAVER, W. H., MCNICHOLS, M. F. & 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, 93-122.
- BERGER, A. N. & BOUWMAN, C. H. 2009. Bank liquidity creation. *Review of Financial Studies*, 22, 3779-3837.

- BERGER, A. N. & BOUWMAN, C. H. 2013. How does capital affect bank performance during financial crises? *Journal of Financial Economics*, 109, 146-176.
- BETZER, A., DOUMET, M. & RINNE, U. 2013. How policy changes affect shareholder wealth: the case of the Fukushima Dai-ichi nuclear disaster. *Applied Economics Letters*, 20, 799-803.
- BHARATH, S. T. & SHUMWAY, T. 2008. Forecasting default with the Merton distance to default model. *Review of Financial Studies*, 21, 1339-1369.
- BIELECKI, T. R. & RUTKOWSKI, M. 2002. *Credit Risk: Modeling, Valuation and Hedging*, Berlin, Germany, Springer Science & Business Media.
- BLACK, F. & COX, J. C. 1976. Valuing corporate securities: Some effects of bond indenture provisions. *The Journal of Finance*, 31, 351-367.
- BLACK, F. & SCHOLES, M. 1973. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81, 637-654.
- BLUHM, C., OVERBECK, L. & WAGNER, C. 2003. *An Introduction to Credit Risk Modeling*, Boca Raton, USA, CRC Press.
- BOCHNER, S. 1955. *Harmonic Analysis and the Theory of Probability*, University of California Press, Berkeley, CA.
- BOHL, M. T., KAUFMANN, P. & SIKLOS, P. L. 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.
- BOHL, M. T., KAUFMANN, P. & STEPHAN, P. M. 2013. From hero to zero: Evidence of performance reversal and speculative bubbles in german renewable energy stocks. *Energy Economics*, 37, 40-51.
- BOURGEON, J.-M. & DIONNE, G. 2007. On debt service and renegotiation when debt-holders are more strategic. Working paper. CIRPEE.
http://www.cirpee.org/fileadmin/documents/Cahiers_2007/CIRPEE07-29.pdf [April 11, 2016].
- BOVENZI, J. F., MARINO, J. A. & MCFADDEN, F. E. 1983. Commercial bank failure prediction models. *Economic Review*, 68, 14-26.
- BOYACIOGLU, M. A., KARA, Y. & BAYKAN, Ö. K. 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, 3355-3366.
- BOYER, M. M. & FILION, D. 2007. Common and fundamental factors in stock returns of Canadian oil and gas companies. *Energy Economics*, 29, 428-453.
- BRAMBILLA, C., GURNY, M. & ORTOBELLI LOZZA, S. 2015. Structural Credit Risk Models with Lévy Processes: The VG and NIG Cases. *Far East Journal of Mathematical Sciences*, 97, 101-119.
- BROADSTOCK, D. C., CAO, H. & ZHANG, D. 2012. Oil shocks and their impact on energy related stocks in China. *Energy Economics*, 34, 1888-1895.
- BRUCHE, M. 2005. Estimating Structural Bond Pricing Models via Simulated Maximum Likelihood. Working paper. London School of Economics.
<http://eprints.lse.ac.uk/24647/1/dp534.pdf> [April 11, 2016].
- BYSTRÖM, H. N. 2006. Merton unraveled: A flexible way of modeling default risk. *The Journal of Alternative Investments*, 8, 39-47.
- CAMPBELL, J. Y., HILSCHER, J. & SZILAGYI, J. 2008. In search of distress risk. *The Journal of Finance*, 63, 2899-2939.

- CANBAS, S., CABUK, A. & KILIC, S. B. 2005. Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The Turkish case. *European Journal of Operational Research*, 166, 528-546.
- CAREY, M. & 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. <https://www.mnb.hu/letoltes/gordy.pdf> [April 11, 2016].
- CARHART, M. M. 1997. On persistence in mutual fund performance. *The Journal of Finance*, 52, 57-82.
- CHAN, K. & CHEN, N.-F. 1991. Structural and return characteristics of small and large firms. *The Journal of Finance*, 46, 1467-1484.
- CHAVA, S. & PURNANANDAM, A. 2010. Is default risk negatively related to stock returns? *Review of Financial Studies*, 23, 2523-2559.
- CHILDS, P. D., MAUER, D. C. & OTT, S. H. 2005. Interactions of corporate financing and investment decisions: The effects of agency conflicts. *Journal of Financial Economics*, 76, 667-690.
- CLARK, P. K. 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: Journal of the Econometric Society*, 135-155.
- COHEN, B. H. 2013. How have banks adjusted to higher capital requirements? BIS Quarterly Review. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2401569 [April 19, 2016].
- COLLIN-DUFRESNE, P., GOLDSTEIN, R. S. & MARTIN, J. S. 2001. The determinants of credit spread changes. *The Journal of Finance*, 56, 2177-2207.
- CROSBIE, P. & BOHN, J. 2003. Modeling default risk. Moody's KMV. http://www.defaultrisk.com/pp_model_35.htm [April 11, 2016].
- CUMMINS, M., GARRY, O. & KEARNEY, C. 2014. Price discovery analysis of green equity indices using robust asymmetric vector autoregression. *International Review of Financial Analysis*, 35, 261-267.
- DELIANEDIS, G. & GESKE, R. L. 2003. Credit risk and risk neutral default probabilities: Information about rating migrations and defaults. Working paper. EFA 2003. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=424301 [April 11, 2016].
- DELONG, E. R., DELONG, D. M. & CLARKE-PEARSON, D. L. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44, 837-845.
- DICHEV, I. D. 1998. Is the risk of bankruptcy a systematic risk? *The Journal of Finance*, 53, 1131-1147.
- DU, Y. & SUO, W. 2007. Assessing credit quality from equity markets: Is a structural approach a better approach? *Canadian Journal of Administrative Sciences*, 24, 212-218.
- DUAN, J. C. 1994. Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, 4, 155-167.
- DUAN, J. C., GAUTHIER, G. & SIMONATO, J. G. 2004. On the equivalence of the KMV and maximum likelihood methods for structural credit risk models. Working paper. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.121.2998> [April 11, 2016].
- DUFFIE, D. & LANDO, D. 2001. Term structures of credit spreads with incomplete accounting information. *Econometrica*, 69, 633-664.
- DUFFIE, D., SAITA, L. & WANG, K. 2007. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83, 635-665.

- DUFFIE, D. & SINGLETON, K. J. 2003. *Credit Risk: Pricing, Measurement, and Management*, Princeton, USA, Princeton University Press.
- DURAND, D. 1941. *Risk Elements in Consumer Instalment Financing*, New York, USA, National Bureau of Economic Research.
- EFRON, B. & TIBSHIRANI, R. J. 1994. *An Introduction to the Bootstrap*, Boca Raton, USA, Chapman & Hall/CRC.
- ENGELMANN, B., HAYDEN, E. & TASCHE, D. 2003. Testing rating accuracy. *Risk*, 16, 82-86.
- ERICSSON, J. & RENAULT, O. 2006. Liquidity and credit risk. *The Journal of Finance*, 61, 2219-2250.
- ERICSSON, J. & RENEBY, J. 2005. Estimating Structural Bond Pricing Models. *Journal of Business*, 78, 707-736.
- ESPAHBODI, P. 1991. Identification of problem banks and binary choice models. *Journal of Banking & Finance*, 15, 53-71.
- FAFF, R. W. & BRAILSFORD, T. J. 1999. Oil price risk and the Australian stock market. *Journal of Energy Finance & Development*, 4, 69-87.
- FAMA, E. F. 1963. Mandelbrot and the stable Paretian hypothesis. *The Journal of Business*, 36, 420-429.
- FAMA, E. F. 1965a. The behavior of stock-market prices. *The Journal of Business*, 38, 34-105.
- FAMA, E. F. 1965b. Portfolio analysis in a stable Paretian market. *Management Science*, 11, 404-419.
- FAMA, E. F. & FRENCH, K. R. 1992. The cross-section of expected stock returns. *The Journal of Finance*, 47, 427-465.
- FAMA, E. F. & FRENCH, K. R. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.
- FAMA, E. F. & MACBETH, J. D. 1973. Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, 81, 607-636.
- FELLER, W. 1966. *An Introduction to Probability Theory and Its Applications II*, New York, USA, John Wiley & Sons.
- FERSTL, R., UTZ, S. & WIMMER, M. 2012. The effect of the Japan 2011 disaster on nuclear and alternative energy stocks worldwide: an event study. *BuR-Business Research*, 5, 25-41.
- FISHER, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Human Genetics*, 7, 179-188.
- FOLLMER, H. & SCHWEIZER, M. 1993. A microeconomic approach to diffusion models for stock prices. *Mathematical Finance*, 3, 1-23.
- FOLLMER, H. & SONDERMANN, D. 1986. *Contributions to Mathematical Economics*, Amsterdam, The Netherlands, North Holland.
- GARLAPPI, L., SHU, T. & YAN, H. 2008. Default risk, shareholder advantage, and stock returns. *Review of Financial Studies*, 21, 2743-2778.
- GARLAPPI, L. & YAN, H. 2011. Financial Distress and the Cross-section of Equity Returns. *The Journal of Finance*, 66, 789-822.
- GESKE, R. 1977. The valuation of corporate liabilities as compound options. *Journal of Financial and Quantitative Analysis*, 12, 541-552.
- GIESECKE, K. & GOLDBERG, L. R. 2004. Sequential defaults and incomplete information. *Journal of Risk*, 7, 1-26.
- GREEN, D. M. & SWETS, J. A. 1966. *Signal Detection Theory and Psychophysics*, New York, USA, John Wiley & Sons.

- GRIFFIN, J. M. & LEMMON, M. L. 2002. Book-to-market equity, distress risk, and stock returns. *The Journal of Finance*, 57, 2317-2336.
- GURNY, M., ORTOBELLI LOZZA, S. & GIACOMETTI, R. 2013. Structural Credit Risk Models with Subordinated Processes. *Journal of Applied Mathematics*, 2013, 1-12.
- HACKBARTH, D., HENNESSY, C. A. & LELAND, H. E. 2007. Can the trade-off theory explain debt structure? *Review of Financial Studies*, 20, 1389-1428.
- HACKBARTH, D., MIAO, J. & MORELLEC, E. 2006. Capital structure, credit risk, and macroeconomic conditions. *Journal of Financial Economics*, 82, 519-550.
- HAND, D. J. & HENLEY, W. E. 1997. Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160, 523-541.
- HANLEY, J. A. 1989. Receiver operating characteristic (ROC) methodology: the state of the art. *Critical Reviews in Diagnostic Imaging*, 29, 307-35.
- HENRIQUES, I. & SADORSKY, P. 2008. Oil prices and the stock prices of alternative energy companies. *Energy Economics*, 30, 998-1010.
- HILLEGEIST, S. A., KEATING, E. K., CRAM, D. P. & LUNDSTEDT, K. G. 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9, 5-34.
- HOCHBERG, Y. & TAMHANE, A. C. 1987. *Multiple Comparison Procedures*, New York, USA, John Wiley & Sons.
- HOFMANN, N., PLATEN, E. & SCHWEIZER, M. 1992. Option pricing under incompleteness and stochastic volatility. *Mathematical Finance*, 2, 153-187.
- HOLLANDER, M. & WOLFE, D. 1999. *Nonparametric Statistical Methods*, New York, USA, John Wiley & Sons.
- HOSMER JR, D. W. & LEMESHOW, S. 2004. *Applied Logistic Regression*, John Wiley & Sons.
- HURST, S. R., PLATEN, E. & RACHEV, S. T. 1999. Option pricing for a logstable asset price model. *Mathematical and Computer Modelling*, 29, 105-119.
- INCHAUSPE, J., RIPPLE, R. D. & TRÜCK, S. 2015. The dynamics of returns on renewable energy companies: A state-space approach. *Energy Economics*, 48, 325-335.
- IOANNIDIS, C., PASIOURAS, F. & ZOPOUNIDIS, C. 2010. Assessing bank soundness with classification techniques. *Omega*, 38, 345-357.
- JARQUE, C. M. & BERA, A. K. 1987. A test for normality of observations and regression residuals. *International Statistical Review*, 55, 163-172.
- JO, H. & HAN, I. 1996. Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction. *Expert Systems with Applications*, 11, 415-422.
- JOVAN, M. 2010. The Merton structural model and IRB compliance. *Metodološki Zvezki*, 7, 39-57.
- JU, N., PARRINO, R., POTESHMAN, A. M. & WEISBACH, M. S. 2005. Horses and rabbits? Trade-off theory and optimal capital structure. *Journal of Financial and Quantitative Analysis*, 40, 259-281.
- KEALHOFER, S. & KURBAT, M. 2002. The Default Prediction Power of the Merton Approach, relative to Debt Ratings and Accounting Variables. KMV LLC.
<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=6936BCA6636D5133B152D3A5E1BB14CC?doi=10.1.1.101.7279&rep=rep1&type=pdf> [April 11, 2016].
- KIM, I. J., RAMASWAMY, K. & SUNDARESAN, S. 1993. Does default risk in coupons affect the valuation of corporate bonds?: A contingent claims model. *Financial Management*, 22, 117-131.

- KOLARI, J., CAPUTO, M. & WAGNER, D. 1996. Trait recognition: An alternative approach to early warning systems in commercial banking. *Journal of Business Finance & Accounting*, 23, 1415-1434.
- KOLARI, J., GLENNON, D., SHIN, H. & CAPUTO, M. 2002. Predicting large US commercial bank failures. *Journal of Economics and Business*, 54, 361-387.
- KOROBOW, L. & STUHR, D. P. 1975. Toward early warning of changes in banks' financial condition: A progress report. Federal Reserve Bank of New York. https://www.newyorkfed.org/medialibrary/media/research/monthly_review/1975_pdf/07_3_75.pdf [April 11, 2016].
- KOROBOW, L. & STUHR, D. P. 1985. Performance measurement of early warning models: Comments on West and other weakness/failure prediction models. *Journal of Banking & Finance*, 9, 267-273.
- KOROBOW, L., STUHR, D. P. & MARTIN, D. 1976. A probabilistic approach to early warning of changes in bank financial condition. Federal Reserve Bank of New York. https://www.newyorkfed.org/medialibrary/media/research/monthly_review/1976_pdf/07_3_76.pdf [April 11, 2016].
- KPMG 2014. Evolving Banking Regulation: Is the End in Sight? KPMG. <http://www.kpmg.com/KY/en/IssuesAndInsights/ArticlesPublications/PublishingImages/evolving-banking-regulation-ema-2014-v3-fs.pdf> [April 19, 2016].
- KUMAR, S., MANAGI, S. & MATSUDA, A. 2012. Stock prices of clean energy firms, oil and carbon markets: A vector autoregressive analysis. *Energy Economics*, 34, 215-226.
- LAAJIMI, S. 2012. Structural Credit Risk Models: A Review. *Insurance and Risk Management*, 80, 53-93.
- LAITINEN, E. K. 1991. Financial ratios and different failure processes. *Journal of Business Finance & Accounting*, 18, 649-673.
- LAM, K. F. & MOY, J. W. 2002. Combining discriminant methods in solving classification problems in two-group discriminant analysis. *European Journal of Operational Research*, 138, 294-301.
- LANDO, D. 2009. *Credit Risk Modeling: Theory and Applications*, Princeton, UK, Princeton University Press.
- LANE, W. R., LOONEY, S. W. & WANSLEY, J. W. 1986. An application of the Cox proportional hazards model to bank failure. *Journal of Banking & Finance*, 10, 511-531.
- LANINE, G. & VENNET, R. V. 2006. Failure prediction in the Russian bank sector with logit and trait recognition models. *Expert Systems with Applications*, 30, 463-478.
- 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, 13-37.
- LINTON, L. & HARDER, L. 2007. Biology 315—Quantitative Biology Lecture Notes. *Calgary, AB: University of Calgary*.
- LO, A. W. 1986. Logit versus discriminant analysis: A specification test and application to corporate bankruptcies. *Journal of Econometrics*, 31, 151-178.
- LOGAN, A. 2001. *The United Kingdom's small banks' crisis of the early 1990s: what were the leading indicators of failure?*, Bank of England.
- LONGSTAFF, F. A. 1995. How much can marketability affect security values? *The Journal of Finance*, 50, 1767-1774.
- LONGSTAFF, F. A. & SCHWARTZ, E. S. 1995. A simple approach to valuing risky fixed and floating rate debt. *The Journal of Finance*, 50, 789-819.

- LOPATTA, K. & 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.
- 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.
- MANAGI, S. & 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.
- MANDELBROT, B. 1963a. New Methods in Statistical Economics. *Journal of Political Economy*, 71, 421-440.
- MANDELBROT, B. 1963b. The variation of Certain Speculative Prices. *Journal of Business*, 36, 394-419.
- MANDELBROT, B. 1967. The variation of some other speculative prices. *The Journal of Business*, 40, 393-413.
- MANDELBROT, B. & TAYLOR, H. M. 1967. On the distribution of stock price differences. *Operations Research*, 15, 1057-1062.
- MARKOWITZ, H. 1952. Portfolio selection. *The Journal of Finance*, 7, 77-91.
- MARKOWITZ, H. 1959. *Portfolio Selection: Efficient Diversification of Investments*, New York, USA, John Wiley & Sons.
- MARTIN, D. 1977. Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance*, 1, 249-276.
- MAUER, D. C. & TRIANTIS, A. J. 1994. Interactions of corporate financing and investment decisions: A dynamic framework. *The Journal of Finance*, 49, 1253-1277.
- MCFADDEN, D. 1976. A comment on discriminant analysis" versus" logit analysis. *Annals of Economic and Social Measurement*, 5, 511-523.
- MECAJ, A. 2013. Structural and evolutionary patterns of financially distressed firms: Strategies to overcome decline. Universidad de Salamanca.
http://gredos.usal.es/jspui/bitstream/10366/124180/1/DAEE_ArjolaMecaj_Tesis.pdf [April 18, 2016].
- MERTON, R. C. 1973. Theory of rational option pricing. *The Bell Journal of economics and management science*, 4, 141-183.
- MERTON, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29, 449-470.
- MORELLEC, E. 2001. Asset liquidity, capital structure, and secured debt. *Journal of Financial Economics*, 61, 173-206.
- MOSSIN, J. 1966. Equilibrium in a capital asset market. *Econometrica: Journal of the Econometric Society*, 34, 768-783.
- NEAGU, R., KEENAN, S. & CHALERMKRAIVUTH, K. 2009. Internal credit rating systems: Methodology and economic value. *The Journal of Risk Model Validation*, 3, 11-34.
- NOLAN, J. P. 1997. Numerical calculation of stable densities and distribution functions. *Communications in Statistics - Stochastic models*, 13, 759-774.
- OHLSON, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18, 109-131.
- ORTAS, E. & MONEVA, J. M. 2013. The Clean Techs equity indexes at stake: Risk and return dynamics analysis. *Energy*, 57, 259-269.
- OSBORNE, M. M. 1959. Brownian motion in the stock market. *Operations Research*, 7, 145-173.

- QUEEN, M. & ROLL, R. 1987. Firm mortality: using market indicators to predict survival. *Financial Analysts Journal*, 43, 9-26.
- RACHEV, S. T. 2003. *Handbook of Heavy Tailed Distributions in Finance*, Amsterdam, The Netherlands, Elsevier.
- RACHEV, S. T. & MITTNIK, S. 2000. *Stable Paretian models in finance*, Chichester, UK, John Wiley & Sons.
- REBOREDO, J. C. 2015. Is there dependence and systemic risk between oil and renewable energy stock prices? *Energy Economics*, 48, 32-45.
- SADORSKY, P. 2001. Risk factors in stock returns of Canadian oil and gas companies. *Energy Economics*, 23, 17-28.
- SADORSKY, P. 2012a. Correlations and volatility spillovers between oil prices and the stock prices of clean energy and technology companies. *Energy Economics*, 34, 248-255.
- SADORSKY, P. 2012b. Modeling renewable energy company risk. *Energy Policy*, 40, 39-48.
- SAMORODNITSKY, G. & TAQQU, M. 1994. *Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*, New York, USA, Chapman & Hall.
- SANTOMERO, A. M. & VINSO, J. D. 1977. Estimating the probability of failure for commercial banks and the banking system. *Journal of Banking & Finance*, 1, 185-205.
- SARKAR, S. & ZAPATERO, F. 2003. The Trade-off Model with Mean Reverting Earnings: Theory and Empirical Tests. *The Economic Journal*, 113, 834-860.
- SHARPE, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19, 425-442.
- SHUMWAY, T. 2001. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74, 101-124.
- SINKEY, J. F. 1975. A multivariate statistical analysis of the characteristics of problem banks. *The Journal of Finance*, 30, 21-36.
- SOBEHART, J. R., KEENAN, S. C. & STEIN, R. M. 2000. Benchmarking Quantitative Default Risk Models: A Validation Methodology. Moody's Investors Service, Global Credit Research. <http://www.rogermstein.com/wp-content/uploads/53621.pdf> [April 11, 2016].
- STEIN, R. M. 2002. Benchmarking default prediction models: Pitfalls and remedies in model validation. Moody's KMV, Technical Report. https://riskcalc.moodysrms.com/us/research/crm/validation_tech_report_020305.pdf [April 11, 2016].
- STEIN, R. M. 2006. Are the probabilities right? Dependent defaults and the number of observations required to test for default rate accuracy. *Journal of Investment Management*, 4, 61.
- STEIN, R. M. 2007. Benchmarking default prediction models: Pitfalls and remedies in model validation. *Journal of Risk Model Validation*, 1, 77-113.
- STUHR, D. P. & VAN WICKLEN, R. 1974. Rating the financial condition of banks: A statistical approach to aid bank supervision. Federal Reserve Bank of New York. https://fraser.stlouisfed.org/docs/publications/frbnyreview/pages/1970-1974/67599_1970-1974.pdf [April 11, 2016].
- SUNDARESAN, S. & WANG, N. 2007. Investment under uncertainty with strategic debt service. *The American Economic Review*, 97, 256-261.
- TABACHNICK, B. G. & FIDELL, L. S. 2007. *Using Multivariate Statistics* Boston, USA, Pearson Education Inc.

- TAM, K. Y. & KIANG, M. Y. 1992. Managerial applications of neural networks: the case of bank failure predictions. *Management Science*, 38, 926-947.
- THOMSON, J. B. 1991. Predicting bank failures in the 1980s. *Economic Review*, 27, 9-20.
- TRÜCK, S. & RACHEV, S. T. 2009. *Rating Based Modeling of Credit Risk: Theory and Application of Migration Matrices*, Norwell, USA, Academic Press.
- UHRIG-HOMBURG, M. 2002. Valuation of Defaultable Claims - A Survey. *Schmalenbach Business Review*, 54, 24-57.
- VASICEK, O. A. 1984. Credit valuation. KMV Corporation.
http://www.defaultrisk.com/pp_model241.htm [April 11, 2016].
- VASSALOU, M. & XING, Y. 2004. Default risk in equity returns. *The Journal of Finance*, 59, 831-868.
- VITALE, R. & LAUX, J. A. 2012. The Economic Efficacy Of Banking Mergers: 2006-2008. *Journal of Applied Business Research (JABR)*, 28, 1211-1216.
- VUONG, Q. H. 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57, 307-333.
- WEN, X., GUO, Y., WEI, Y. & HUANG, D. 2014. How do the stock prices of new energy and fossil fuel companies correlate? Evidence from China. *Energy Economics*, 41, 63-75.
- WEST, R. C. 1985. A factor-analytic approach to bank condition. *Journal of Banking & Finance*, 9, 253-266.
- WEYANT, J. P. 2011. Accelerating the development and diffusion of new energy technologies: Beyond the "valley of death". *Energy Economics*, 33, 674-682.
- WHALEN, G. & THOMSON, J. B. 1988. Using financial data to identify changes in bank condition. *Economic Review*, 24, 17-26.
- ZHOU, C. 2001. The term structure of credit spreads with jump risk. *Journal of Banking & Finance*, 25, 2015-2040.
- ZMIJEWSKI, M. E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59-82.