

Corporate Default Prediction: Models, Drivers and Measurements

Submitted by Yangzhengxuan Wang to the University of Exeter as a Thesis for the Degree
of Doctor of Philosophy in Finance,
December 2011.

This thesis is available for Library use on the understanding that it is copyright material
and that no quotation from this thesis may be published without proper
acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and
that no material has previously been submitted and approved for the award of a degree by
this or any other University.

Signature:

Abstract

This thesis identifies the optimal set of corporate default drivers and examines the prediction performance of corporate default measurement tools, using a sample of companies in the United States from 1970 to 2009.

In the discussion of optimal default drivers, feature selection techniques including the t-test and stepwise methods are used to filter relevant default information collected from previous empirical studies. The optimal default driver information set consists of quantitative parameters from accounting ratios, market indices, macroeconomic indicators, default history, and firm age. While both accounting ratios and market information dominate the explanatory ability, followed by default history, macroeconomic indicators contribute additional explanation for default risk. Moreover, industry effects show significance across alternative models, with the retail industry presenting as the sector with highest risk. The results are robust in both traditional and advanced random models.

In investigating the optimal prediction method, two newly developed random models, mixed logit and frailty model, are tested for their theoretical superiority in capturing default clusters and unobservable information for default risk. The prediction ability of both models has been improved upon using the extended optimal set of default drivers. While the mixed logit model provides better prediction accuracy and shows stability in robustness checks, the frailty model benefits from computational efficiency and explains default clusters more thoroughly.

This thesis further compares the prediction performance of large dimensional models across five categories based on the default probabilities transferred from alternative results in different models. Besides the traditional assessment criteria - covering the

receiver operating characteristic curve, accuracy ratios, and classification error rates – this thesis thoroughly evaluates forecasting performance using innovative proxies including model stability under financial crisis, profitability and misclassification costs for creditors using alternative risk measurements. The practical superiority of the two advanced random models has been verified further in the comparative study.

Acknowledgements

First and foremost, I offer my sincerest gratitude to my supervisor, Richard D. F. Harris, who with his invaluable guidance has supported me throughout my thesis patiently and tolerantly, from preliminary data collection to the conclusions. This thesis would not have been possible without his effective input. One simply could not wish for a better, or kinder, supervisor.

My great thanks goes to my dearest friends, Peter, Debora, Diaz, William, Wenjing, Fiona, Yibo, Ian and Jane, who have helped me in every corner of my life in the UK. They have shared the highs and lows, and welcomed me as a family member.

Thanks to the University of Exeter Business School, which has provided the support and the friendly environment, thus enable me produce and complete my thesis. Thanks to the HOST UK, which helps international students, like myself, to make friends in this country, understand local culture, and gain insights into Britain. Throughout my studies I have benefited greatly from this charity, and it helped make my life easier and happier.

Thanks to AXA has generously funded my studies. Without which I could not have concentrated effectively on my work.

Finally, my appreciation goes to my parents, my grandmother, and my beloved Junxiong, for their understanding and their unwavering support throughout my studies.

Table of Contents

Abstract	2
Acknowledgements	4
List of Tables.....	10
List of Figures	12
Chapter 1: Introduction.....	14
1.1 Objectives and Motivations	14
1.1.1 The Concept of Corporate Default	14
1.1.2 The Importance of Corporate Default Prediction	16
1.1.3 Opportunity for Corporate Default Prediction.....	17
1.2 Principal Contributions	19
1.2.1 Default Drivers	19
1.2.2 Advanced Random Models.....	20
1.2.3 Optimal Predicting Models.....	22
1.3 Structure of the Thesis	23
Chapter 2: Literature Review	25
2.1 Introduction and Overview	25
2.2 Prediction Models.....	27
2.2.1 Discriminant Analysis	27
2.2.1.1 Univariate Discriminant Analysis	28
2.2.1.2 Multiple Discriminant Analysis.....	28
2.2.2 Conditional Probability Model	30
2.2.2.1 Linear Probability Model.....	31
2.2.2.2 Probit Model.....	31
2.2.2.3 Logit Model	31
2.2.3 Advanced Choice Models	34

2.2.3.1 Mixed Logit Model	34
2.2.3.2 Nested Logit Model.....	35
2.2.4 Market-Based Structure Model	36
2.2.5 Hazard Intensity Survival Model.....	37
2.2.6 Artificial Intelligent Techniques	40
2.2.6.1 Neural Network	40
2.2.6.2 Rough Set.....	41
2.2.6.3 Decision Tree	41
2.2.6.4 Case-Based Reasoning Model	42
2.3 Variables	43
2.3.1 Accounting Ratios	43
2.3.2 Market Variables	48
2.3.3 Macroeconomic Variables	51
2.3.4 Industry Effects.....	54
2.3.5 Other Significant Variables	56
2.4 Samples Used by Empirical Studies	57
2.5 Model Assessment Methods	58
2.6 Variable Combination and Selection.....	59
2.7 Conclusion	60
Chapter 3: Variable Selection for Default Prediction	72
3.1 Introduction.....	72
3.2 Methodology	76
3.2.1 Variable Selection Method	76
3.2.2 The T-Test	77
3.2.3 Stepwise Regression	78
3.2.4 Dynamic Logit Model.....	78
3.3 Data Description	81
3.3.1 The Corporate Default Database	81

3.3.2 Explanatory Variables.....	85
3.3.2.1 Accounting Variables.....	85
3.3.2.2 Firm-level Financial Market Variables.....	86
3.3.2.3 Other Firm-level Variables.....	87
3.3.2.4 Macro-level Variables.....	87
3.4 Results	89
3.4.1 Variable Selection	89
3.4.1.1 Mean Difference T-Test.....	89
3.4.1.2 Stepwise Tests.....	90
3.4.1.3 Feature Selection Methods Comparison	91
3.4.2 Variable Contributions.....	91
3.4.3 Regressions with Variables from Different Categories.....	95
3.4.4 Regressions of Different Horizons.....	96
3.4.5 Different Sub-periods.....	98
3.4.6 Industry Effects.....	99
3.4.7 Classification and Prediction Ability.....	100
3.4.7.1 Classification Rates.....	100
3.4.7.2 AUROC.....	101
3.5 Conclusion	102
Chapter 4: Corporate Default Prediction with Random Effects	133
4.1 Introduction and Background	133
4.2 Methodology	137
4.2.1 Mixed Logit Model	138
4.2.2 Frailty Model.....	140
4.3 Results	141
4.3.1 Comparison between Logit and Mixed Logit Models	142
4.3.2 Mixed Logit Models with Model Specific Tests	144
4.3.2.1 Halton Draws	145

4.3.2.2 Correlated Coefficients	145
4.3.3 Results from Cox Frailty and Non-frailty Models.....	147
4.3.3.1 Non-frailty Cox Model	148
4.3.3.2 Frailty Cox Model	149
4.4. Method Comparison.....	151
4.5 Conclusion	152
Chapter 5: Model Comparison.....	169
5.1 Introduction.....	169
5.2 Data and Methods	171
5.2.1 Model Sample.....	172
5.2.2 Default Prediction Methods.....	172
5.3 Model Comparison Approaches	174
5.3.1 AUROC Comparison	174
5.3.2 Economic Benefits and Costs	176
5.4 Results	178
5.4.1 Statistical Summary	178
5.4.2 Overall Prediction Accuracy	179
5.4.3 Misclassification Costs	181
5.4.3.1 Classification Errors.....	181
5.4.3.2 Economic Costs of Different Errors.....	182
5.6 Conclusion	184
Chapter 6: Conclusion and Recommendations.....	197
6.1 Summary of Research.....	197
6.2 Conclusions	199
6.2.1 Optimal Information Set	199
6.2.2 Optimal Prediction Methods	202
6.3 Research Limitations and Further Research Suggestions	205

Bibliography..... 207

List of Tables

CHAPTER TWO

Table 2.1 Lists of Previous Empirical Tests for Default Drivers Summary.....	61
Table 2.2 Accounting Variables Definitions.....	63
Table 2.3 Market Variables Definition.	64
Table 2.4 Macroeconomic Variables Definition.....	65

CHAPTER THREE

Table 3.1 Default Database Constructions.....	103
Table 3.2 Descriptive Statistics of Firm-level Variables.....	104
Table 3.3 Macroeconomic Variables Descriptive Statistics and T-test.....	105
Table 3.4 Stepwise Regressions with Alternative Approaches.....	106
Table 3.5 Comparisons of the T-test and the Stepwise Selection.....	108
Table 3.6 Regressions with Different Significant Tiers.....	109
Table 3.7 Regressions with the Best Variable Set in Previous Studies.....	111
Table 3.8 Regressions with Different Variable Categories.....	112
Table 3.9 Regressions with Different Horizons.....	114
Table 3.10 Regressions with Different Periods.....	116
Table 3.11 Regressions with Industry Effects.....	118
Table 3.12 Accuracy and Error Scenarios for Default Classifications.....	120
Table 3.13 In-sample and Out-of-sample Classification Rates.....	121
Table 3.14 AUROC Estimation of In-sample and Out-of-sample Tests.....	122

CHAPTER FOUR

Table 4.1 Standard Logit and Mixed Logit Results.....	154
Table 4.2 Mixed Logit Tests with Different Halton Replications.....	156
Table 4.3 Mixed Logit Models with Correlated Coefficients.....	158
Table 4.4 Covariance Matrix of Random Variables.....	160
Table 4.5 Cox Proportion Models on Different Information.....	161
Table 4.6 Frailty Cox Survival Model Results.....	163
Table 4.7 Out-of-sample Comparison for Advanced Random Models.....	165

CHAPTER FIVE

Table 5.1 Subsample Details.....	186
----------------------------------	-----

Table 5.2 Mathematical Methods for Comparison.....	187
Table 5.3 Economic Costs for Different Error Types.....	188
Table 5.4 Statistical Summaries of Default Predictions.....	189
Table 5.5 AUROC and AR Comparison over Different Periods.....	190
Table 5.6 Type I and Type II Errors.....	191
Table 5.7 Comparative Economic Costs of Different Default Models.....	192
Table 5.8 Economic Costs in a Dynamic Competitive Credit Market.....	193

List of Figures

CHAPTER TWO

Figure 2.1 Summaries of Default Prediction Models.....	67
Figure 2.2 Summaries of Significant Accounting Variables.....	68
Figure 2.3 Summaries of Significant Market Variables in Previous Studies.....	69
Figure 2.4 Summaries of Significant Macro Variables in Previous Studies.....	70
Figure 2.5 Industry Effects and Other Significant Variables.....	71

CHAPTER THREE

Figure 3.1 Feature Selection Procedures.	123
Figure 3.2 Shape of Logistic Function.....	124
Figure 3.3 Default Number Dropping Process.....	125
Figure 3.4 Yearly Default Numbers.....	126
Figure 3.5 Quarterly Default Rates.....	127
Figure 3.6 Default Information by Industries.....	128
Figure 3.7 Significance Levels of Selected Independent Variables.	129
Figure 3.8 Correlation Pairs Between Variables.....	130
Figure 3.9 Predicted vs. Actual Default Rates.....	131
Figure 3.10 AUROC of In-sample and Out-of-Sample Tests.....	132

CHAPTER FOUR

Figure 4.1 Ages at Defaults.	166
Figure 4.2 Kaplan-Meier Survival Estimation.....	167
Figure 4.3 Frailty by Different Categories.....	168

CHAPTER FIVE

Figure 5.1 The AUROC of Twelve Models for In-sample Tests.....	194
Figure 5.2 The AUROC of Twelve Models for Out-of-sample Tests.....	195
Figure 5.3 Predicting Stability with Rolling Out-of-sample AUROC.....	196

Abbreviations

ACM	Advance Choice Models
AR	Accuracy ratios
AUROC	Are Under the Receiver Operating Characteristic Curve
CPM	Conditional Probability Models
CPV	Credit Portfolio View
CRSP	Centre for Research in Security Prices
DA	Discriminant Analysis
FASB	Financial Accounting Standards Board
FRED	Federal Reserve Bank of St. Louis
HIM	Hazard Intensity Models
IIA	Independence of Irrelevant Alternatives
IID	Independent and Identically Distributed
IT	Information Techniques
MBS	Market-Based Structure models
MDA	Multivariate Discriminant Analysis
NN	Neural Network
O-score	Olson's (1980) measure of the probability of bankruptcy
OL	Opportunity Loss
OLOA	Opportunity Loss On Assets
OV	Original Variables
R&D	Research and Development
ROA	Return On Assets
RS	Rough Set technique
S.D.	Standard Deviation
S&P	Standard & Poor's
UDA	Univariant Discriminant Analysis
WRDS	Wharton Research Data Services
Z-score	Altman's (1968) measure of the probability of bankruptcy