Forecasting Financial Outcomes Using Variable Selection Techniques

By

Ping Zhang

Submitted in Fulfilment of the Requirements of the

Degree of Doctor of Philosophy

Adam Smith Business School

College of Social Science

University of Glasgow

February 2019

Abstract

Since the activities of market participants can be influenced by financial outcomes, providing accurate forecasts of these financial outcomes can help participants to reduce the risk of adjusting to any market change in the future. Predictions of financial outcomes have usually been obtained by conventional statistical models based on researchers' knowledge. With the development of data collection and storage, an extensive set of explanatory variables will be extracted from big data capturing more economic theories and then applied to predictive methods, which can increase the difficulty of model interpretation and produce biased estimation. This may further reduce predictive ability. To overcome these problems, variable selection techniques are frequently employed to simplify model selection and produce more accurate forecasts. In this PhD thesis, we aim to combine variable selection approaches with traditional reduced-form models to define and forecast the financial outcomes in question (market implied ratings, Initial Public Offering (IPO) decisions and the failure of companies). This provides benefits for market participants in detecting potential investment opportunities and reducing credit risk.

Making accurate predictions of corporate credit ratings is a crucial issue for both investors and rating agencies, since firms' credit ratings are associated with financial flexibility and debt or equity issuance. In Chapter 2, we attempt to determine market-implied credit ratings in relation to financial ratios, market-driven factors and macroeconomic indicators. We conclude that applying variable selection techniques, the least absolute shrinkage and selection operator (LASSO) and its extension (Elastic net) can improve predictive power. Moreover, the predictive ability of LASSO-selected models is clearly better than that of the benchmark ordered probit model in all out-of-sample predictions. Finally, fewer predictors can be selected into LASSO models controlled by BIC-type tuning parameter to produce more accurate out-of-sample prediction than its counterpart AIC-type selector.

Next, the LASSO technique is further applied to binary event prediction. A bank's decision to go public by issuing an Initial Public Offering (IPO) is the binary object in Chapter 3, which transforms the operations and capital structure of a bank. Much of the empirical investigation in this area focuses on the determinants of

the IPO decision, applying accounting ratios and other publicly available information in non-linear models. We mark a break with this literature by offering methodological extensions as well as an extensive and updated US dataset to predict bank IPOs. Combining the least absolute shrinkage and selection operator (LASSO) with a Cox proportional hazard, we uncover value in several financial factors as well as market-driven and macroeconomic variables, in predicting a bank's decision to go public. Importantly, we document a significant improvement in the model's predictive ability compared to standard frameworks used in the literature. Finally, we show that the sensitivity of a bank's IPO to financial characteristics is higher during periods of global financial crisis than in calmer times.

Moving to another line of variable selection techniques, Bayesian Model Averaging (BMA) is combined with reduced-form models in Chapter 4. The failure of companies is closely related to the health of the whole economy, since the beginning of the most recent global crisis was the bankruptcy of Lehman Brothers. In this chapter, we forecast the failure of UK private firms incorporating with financial ratios and macroeconomic variables. Since two important financial crises and firm heterogeneities are covered in our dataset, the predictive powers of candidate models in different periods and cross-sections are validated. We first detect that applying BMA to the discrete hazard models can improve the predictive performance in different sub-periods. However, comparing the results with classified models, it should be noted that the Naive Bayes (NB) classifier provides slightly higher predictive accuracy than BMA models of discrete hazard models. Moreover, the predictive performance of the discrete hazard model and its BMA version are more sensitive to adding time or industry dummy variables than other competing models. Considering financial crisis or firm heterogeneity can influence the predictive power of each candidate model in the out-of-sample prediction of failure.

Table of Contents

Abstract ii
List of Tables vi
Acknowledgements viii
Author's Declaration ix
Chapter 1 Introduction 1
1.1 General background 1
1.2 Structure 3
References 7
Chapter 2 Modelling market implied ratings using LASSO variable selection techniques 9
Abstract 9
2.1 Introduction 10
2.2 Literature 12
2.3 Data and summary statistics 18
2.3.1 Data sources 18
2.3.2 Choice of explanatory variables 20
2.3.3 Summary statistics 22
2.4 Methodology 27
2.4.1 OP 27
2.4.2 OP with LASSO 29
2.4.3 OP with Elastic net 30
2.4.4 CR with LASSO 32
2.4.5 CR with Elastic net 34
2.5 Empirical results 34
2.5.1 Accuracy 34
2.5.2 Statistical significance 36
2.5.3 Robustness tests 37
2.5.4 Discussion 44
2.6 Conclusion 45
References 47
Appendix 52
Chapter 3 What influences a bank's decision to go public? 68
Abstract 68
3.1 Introduction 69
3.2 Literature 71
3.3 Data and summary statistics 78
3.3.1 Data description 78
3.3.2 Choice of explanatory variables 79
3.3.3 Summary statistics 79
3.4 Methodology 82

3.4.1 Cox proportional hazard model (CPH) ... 82
3.4.2 Discrete hazard model (DH) ... 83
3.4.3 Logistic model ... 83
3.4.4 LASSO ... 83
3.4.5 L1 Penalized Semi-Parametric Cox Proportional Hazard Model (Penalized CPH model) ... 85
3.4.6 L1 Penalized Discrete Hazard Model (Penalized DH model) ... 85
3.4.7 L1 Penalized Logistic Model (Penalized Logistic Model) ... 86
3.5 Predictive ability ... 87
3.5.1 CPH and penalized CPH ... 87
3.5.2 Predictive Deciles ... 88
3.6 An empirical application using US data ... 89
3.6.1 The pre-crisis period ... 90
3.6.2 The crisis period ... 93
3.6.3 The post-crisis period ... 96
3.7 Conclusion ... 98
References ... 99
Appendix ... 104
Chapter 4 Predicting failure: evidence from UK firms ... 109
Abstract ... 109
4.1 Introduction ... 110
4.2 Literature ... 113
4.3 Data and summary statistics ... 119
4.3.1 Data sources ... 119
4.3.2 Choice of explanatory variables ... 121
4.3.3 Summary statistics ... 124
4.4 Methodology ... 127
4.4.1 Discrete Hazard (DH) model ... 127
4.4.2 The Bayesian Model Averaging (BMA) version of DH models ... 129
4.4.3 Naive Bayes (NB) Classifier ... 133
4.4.4 K-nearest neighbours (k-NN) classifier ... 135
4.5 Empirical results ... 136
4.5.1 Whole period (1991-2009) ... 137
4.5.2 The sub-sample periods ... 140
4.5.3 The sub-samples in cross sections ... 147
4.6 Conclusion ... 154
References ... 156
Chapter 5 Conclusion ... 163

List of Tables

Table 2-1 Rating categories 19
Table 2-2 CDSIRs of firms by year 23
Table 2-3 EQIRs of firms by year 23
Table 2-4 Descriptive statistics-CDSIRs 24
Table 2-5 Descriptive statistics-EQIRs 26
Table 2-6 Accuracy Ratios and selected variables 35
Table 2-7 Accuracy Ratios and selected variables (10-fold cross validation) 38
Table 2-8 PCA combined with OP 39
Table 2-9 Accuracy Ratios and selected variables (scaled LASSO) 40
Table 2-10 CP Ratios and selected variables 41
Table 2-11 Equal performance tests of out-of-sample prediction of CDSIRs 42
Table 2-12 Equal performance tests of out-of-sample prediction of EQIRs 42
Table 2-13 Accuracy Ratios and selected variables (investment-grade ratings) . 43
Table 2-14 Accuracy Ratios and selected variables (random effects) 44
Table 3-1 The distribution of banks 78
Table 3-2 Summary statistics 81
Table 3-3 Accuracy ratios and the number of surviving variables in the CPH model and its penalized versions 87
Table 3-4 IPO decision by out-of-sample prediction decile 89
Table 3-5 The estimates of candidate models in the pre-crisis period 91
Table 3-6 The estimates of candidate models in the crisis period 94
Table 3-7 The estimates of candidate models in the post-crisis period 96
Table 4-1 Expected signs and variables definition 122
Table 4-2 Firms by year 124
Table 4-3 Descriptive statistics 126
Table 4-4 Sample means 127
Table 4-5 AUC and brier scores for all competing models during 1991-2009 138
Table 4-6 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2009 140
Table 4-7 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2009 140
Table 4-8 AUC and brier scores for all competing models during 1994-2009 141
Table 4-9 Defaults by out-of-sample prediction decile among model related to the DH model during 1994-2009 141
Table 4-10 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1994-2009 142
Table 4-11 AUC and brier scores for all competing models during 1991-2007 ... 143
Table 4-12 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2007 143
Table 4-13 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2007 143
Table 4-14 AUC and brier scores for all competing models during 1994-2007 ... 145
Table 4-15 Defaults by out-of-sample prediction decile among model related to the DH model during 1994-2007 145
Table 4-16 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1994-2007 145
Table 4-17 AUC and brier scores for all competing models during 1991-2009_small firms 148
Table 4-18 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2009_small firms 148

Table 4-19 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2009_small firms 148
Table 4-20 AUC and brier scores for all competing models during 1991-2009_large firms 149
Table 4-21 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2009_large firms 150
Table 4-22 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2009_large firms..................................150
Table 4-23 AUC and brier scores for all competing models during 1991-2009_young firms 151
Table 4-24 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2009_young firms 151
Table 4-25 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2009_young firms..................................151
Table 4-26 AUC and brier scores for all competing models during 1991-2009_old firms 152
Table 4-27 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2009_old firms..152
Table 4-28 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2009_old firms 153

Acknowledgements

I would like to express my deepest gratitude to my principal supervisor Professor Serafeim Tsoukas. Without his continuous faith, encouragement and academic support, I would have had difficulty starting a PhD, let alone completing it. He is a dedicated mentor and I am also thankful for all the wonderful opportunities in academia he has given me. The PhD has been an unforgettable experience. Similarly, my profound thanks go to my second supervisor Professor Georgios Sermpinis for his guidance and support. I am deeply appreciative all discussions with him, which gave me valuable suggestions for my work. It will keep fond memories of my PhD time.

I am also appreciative to the editors and anonymous referees of the Journal of Empirical Finance for their comments and suggestions, which have greatly helped to improve my research. Additionally, I acknowledge the financial support provided by the China Scholarship Council (CSC) and the academic environment created by staff at the Adam Smith Business School.

Special mentions go to my friends and colleagues. In particular, I extend thanks to Li Pan and Jingxuan Zhang for all the experiences we have had together.

Last but not least, I am greatly indebted to my mum, dad and younger brother. Without their inspiration, love and support, I would not be the person I am today. My every success belongs to them as well.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Printed Name: Ping Zhang

Signature: __________

Chapter 1 Introduction

1.1 General background

Financial forecasting has significant ability to influence future activities for market participants. Providing accurate forecasts generally helps market participants to reduce uncertainty around costs, identify market tendencies, and manage efficient plans, which can frequently be achieved by statistical models in financial research. Such models can be explanatory or predictive, and serve to evaluate and develop theories in terms of causality or prediction (Shmueli 2010). Compared with explanatory models, predictive models have some attractive characteristics (Shmueli and Koppius 2011). First, more underlying patterns and relationships in a large, complex dataset can be captured by predictive analytics, and hence new causality would be suggested. As data storage and computing techniques advance, many individuals can be recorded and the substantial features of each individual can be tracked simultaneously in the dataset (Giraud 2015). Thus, an extensive set of potential variables capturing diverse economic theories can be extracted from the dataset and applied to building a model. Predictive models are able to reduce the interference from superabundant outliers in the dataset and simplify the constructed model. Moreover, predictive models can evaluate the predictive ability of predictors with high explanatory ability. Finally, more accurate forecasts can be provided by predictive models. In this thesis, we focus on exploring the properties of predictive models in empirical analysis to clarify important predictors and produce more accurate prediction for specific financial outcomes.

Considering bias-variance trade-off, two types of techniques are developed to provide more accurate forecasts and a sparser representation in predictive models (Shmueli and Koppius 2011). The first is called the shrinkage approach, which contains principal components regression, ridge regression and its extensions and so on. In this approach the bias is allowed, and the variance is reduced, and the major patterns are therefore captured to produce improved prediction. The second is related to ensemble technique in machine learning, for example bagging (Breiman 1996), boosting (Schapire 1999) and Bayesian derivatives. This method tends to combine several predictions from different models to generate more accurate prediction. We choose in particular in this thesis two variable selection techniques from these predictive models, where the least absolute shrinkage and selection operator (LASSO) is related to shrinkage approach and Bayesian model averaging (BMA) is derived from Bayesian theory.

The least absolute shrinkage and selection operator (LASSO), developed by Tibshirani (1996), can be regarded as a hybrid of variable selection and shrinkage estimators. It enables estimation and variable selection simultaneously in the non-orthogonal setting. If the tuning parameter exceeds a threshold value in LASSO, the coefficients of non-relevant independent variables are forced to zero in the model and the less shrinkage is allowed to be placed on the important predictors. Therefore, the multi-collinearity problem can be solved, and a more interpretable and sparser model can be generated after the LASSO estimator. Moreover, due to the smooth form of the penalty function, LASSO can select fewer independent variables and is a more stable model in comparison with best-subset and stepwise selection methods. Since LASSO can successfully be applied with an extensive set of predictors, the outstanding flexibility of LASSO can adjust to any changes to variables in the application. In addition, due to the exclusion of interfering information, the predictive performance will be better than in the conventional models.

Before LASSO was developed, best-subset (Hocking and Leslie 1967) and stepwise selection (Draper and Smith 1966) methods as representatives of variable selection techniques in predictive models were developed in the 1960s and generally implemented into scientific research due to their simple application and ease of interpretation. However, these methods have some disadvantages compared with LASSO. Tibshirani (1996) and Zou (2006) demonstrate that subset variable selection is not stable during the discrete process, even if the dataset changes slightly. When potential independent variables are large, the computation of subset selection is complex, which is always substituted by the stepwise subset selection (Tian et al. 2015). Fan and Li (2001) further mention that stochastic errors are omitted, and estimation is biased in the application of stepwise selection, since it is just based on correlation between latent variables.

Moving to the other type of variable selection techniques in this thesis, Bayesian Model Averaging (BMA) is one part of the averaging approach, which is extended from the usual Bayesian inference methods. It captures parameter uncertainty in one model through the prior distribution and solves model uncertainty by posterior parameter using Bayes' theorem (Fragoso *et al.* 2018). It should be noted that variable selection procedure in BMA is entirely different from LASSO. If researchers focus on variable or model selection, the variable or model with the highest posterior probability will be chosen and then this will be used to generate predictions. This compromise between selection and prediction can be easily completed by the Stochastic Search Variable Selection (SSVS) (Lamon and Clyde 2000) or the Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm (Jacobson and Karlsson 2004). In other words, it allows for direct model selection, combined estimation and prediction together, and further suggests that parameter and model uncertainty are taken into account in inferences and predictions. Considering the model evaluation, Madigan and Raftery (1994) confirm that the BMA approach can produce predictions with lower risk under a logarithmic scoring rule than using a single model.

In this thesis, we aim to apply these variable selection techniques to determine and forecast some corporate events such as Market Implied Ratings (MIRs), Initial Public Offering (IPO) decisions and the failure of a firm. It is known that even financial analysts or outsider investors cannot get the full information about companies from published information. This implies that these market participants are likely to miss investment chances or take more credit risks in the financial market. To reduce these risks and assess a company from various points of view, specific events in a firm's life (such as equity or debt issuance) are now considered as the primary signal of the firm's operating information available to outsiders (Sena 2002). An increasing number of researchers have emphasised specific corporate events to evaluate the financial health of companies (Eckbo 2009).

1.2 Structure

In order to better organize this thesis, we divide the systematically empirical forecast exercise into three main chapters. In Chapter 2[1], we investigate the determinants of market implied ratings (MIRs) in relation to firm-specific factors,

[1] This chapter is based on a research paper co-authored with my supervisors Serafeim Tsoukas and Georgios Sermpinis, where is published on Journal of Empirical Finance.

market-driven indicators and macroeconomic predictors, and then forecast them. Graham and Harvey (2001) demonstrate that credit ratings have considerable influence on debt issuance and capital structure. Making accurate predictions of corporate credit ratings is a crucial issue to both investors and rating agencies. Compared with conventional long-term credit ratings, MIRs can be published with high frequency, incorporating market information into data-intensive rating models, which provide more timely information about credit quality at short and medium term horizons (Rösch 2005, Tsoukas and Spaliara 2014). Hence, we employ the literature of credit ratings as foundations to evaluate the predictive performance of all candidate models that capture volatile market changes. Since MIRs are assigned into ordinal categories, ordered probit models are applied as the benchmark to define and forecast MIRs (Pasiouras *et al.* 2006, Hwang *et al.* 2010, Mizen and Tsoukas 2012, Hwang 2013). The variable selection technique (LASSO) and its extension (the Elastic net) are then added into the ordered probit model and continuation ratio model to determine and forecast MIRs.

From the conclusions of Chapter 2, we first indicate that MIRs can be affected by several firm-specific, market-driven and macroeconomic variables. Meanwhile, we confirm LASSO models can provide better predictive performance on the out-of-sample MIRs than the ordered probit model and simultaneously the most important predictors can be selected in LASSO models. We also demonstrate that the more accurate out-of-sample prediction and fewer predictors can be produced in the LASSO models with BIC-type tuning parameter selector than their LASSO counterparts with AIC-type selector. To validate our results, different robustness tests are applied, and our main results are robust to the above modifications. Thus, LASSO-selected models provide better predictive performance in ordinal financial outcomes.

In Chapter 3, a bank's IPO decision is chosen as a targeted event for empirical investigation. A bank is a financial intermediary whose core activity is to provide loans to borrowers and collect deposits from savers. The operating status of a bank is associated with the allocation of financial resources and the growth of investment in financial markets, which influence the cost of financial intermediation and the stability of the whole financial market. Since equity finance in the USA has become an important source of funding for banks, a bank's decision to go public has been attracting more attention from academics. We apply accounting ratios and other publicly available information in Cox proportional hazard model and its penalized model to identify the factors that influence a bank's IPO decision and predict the IPO decision. Since our dataset extends from 1996 to 2016, covering the key years of the banking crisis (2007-2009), we separate the dataset into pre-crisis, crisis and post-crisis periods and assess how market conditions influence the probability of banks going public.

According to the results, we first document that selected variables are indicators affecting the probability of a bank's IPO decision and further detect that a bank's IPO decision is more sensitive to the change in its financial health in extreme economic events than in non-crisis periods. Moving to the analysis of predictive performance, we observe a significant increase in the percentage of correct out-of-sample forecasts of IPO decisions for banks by adding the LASSO estimator into the Cox proportional hazard model. On the other hand, we indicate that the Cox proportional hazard model underperforms discrete hazard and logistic models and the Cox model with LASSO estimator outperforms discrete hazard, logistic models and their LASSO version. This highlights the effect of LASSO on our algorithms. In line with the conclusion in Chapter 2, the LASSO models controlled by BIC-type tuning parameter selector can provide a sparser model and higher predictive power in out-of-sample forecasts than their counterpart with AIC-type selector. Both Chapters 2 and 3 confirm that the LASSO technique can identify the most relevant predictors from an extensive set of candidate variables without considering a pre-selection of these potential variables (van de Geer 2008), enhance the predictive ability (Fan and Li 2001, Tian *et al.* 2015) and sidestep the problem of multi-collinearity (Tibshirani 1996).

Next, in Chapter 4 we forecast the failure of private firms by firm-specific and macroeconomic indicators in the UK. The UK as a developed European country entered into economic recession from 2008 with the signal that the real Gross Domestic Product (GDP) growth rate was extremely low. According to the US market performance after the financial crisis, increased credit risk for UK firms can be supposed. Therefore, providing precise forecasts of UK firms' failure is necessary for their managers, potential investors or policy makers, helping them reduce the probability of being exposed to credit risk and preventing economic depression. To distinguish this work from the literature, we choose private firms with small and medium-scale operating in several industries in the UK, because the growth of these firms has become the main power in economic development (Beck *et al.* 2005). The data period from 1991 to 2009 covers two important financial crises in UK economic history: the 1991–1993 European Exchange Rate Mechanism (ERM) currency crisis and the 2008-2009 global financial crisis (GFC). This provides an opportunity for us to detect whether the predictive powers of all candidate models are sensitive to financial crises. Meanwhile, we divide our dataset into two cross-sections, to capture two dimensions of firm heterogeneity (size and age). This could help us detect whether the cross-sectional differences affect the predictive ability of each model. To evaluate the predictive ability, we apply a combined model by discrete hazard model and BMA model to solve the parameter and model uncertainty and assess the predictive performance of firms' failure compared with the discrete hazard model and traditional classifiers, the Naive Bayes (NB) classifier and the k-nearest neighbours (k-NN) classifier.

From the main conclusions, the top two models providing higher predictive ability in firms' failure are the BMA of discrete hazard models and the NB classifier. The NB classifier outperforms the BMA version of discrete hazard model in some sub-samples while the predictive ability of the BMA version of discrete hazard model is comparable with that of the NB classifier. It should be noted that adding BMA into the discrete hazard models can increase the predictive accuracy in out-of-sample prediction and solve parameter and model uncertainty. The results also suggest that financial crisis and firm heterogeneity can affect the predictive power of each candidate model. We also observe that considering time effects or industry effects can improve the accuracy of out-of-sample prediction in different periods, especially for discrete hazard model and its BMA version.

In the last chapter, we conclude all empirical work and suggest some topics for future research.

References

Beck, T., Demirguc-Kunt, A. and Levine, R. (2005) 'SMEs, Growth, and Poverty: Cross-Country Evidence', *Journal of Economic Growth*, 10(3), 199-229.

Breiman, L. (1996) 'Bagging Predictors', *Machine Learning*, 24(2), 123-140.

Draper, N. R. and Smith, H. (1966) *Applied Regression Analysis*, New York: John Wiley & Sons.

Eckbo, B. E. (2009) *Handbook of Corporate Finance: Empirical Corporate Finance. Vol. 1,* Amsterdam: Elsevier North-Holland.

Fan, J. and Li, R. (2001) 'Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties', *Journal of the American Statistical Association*, 96(456), 1348-1360.

Fragoso, T. M., Bertoli, W. and Louzada, F. (2018) 'Bayesian Model Averaging: A Systematic Review and Conceptual Classification: BMA: A Systematic Review', *International Statistical Review*, 86(1), 1-28.

Giraud, C. (2015) *High-dimensional Statistics*, Boca Raton: CRC Press.

Graham, J. R. and Harvey, C. R. (2001) 'The Theory and Practice of Corporate Finance: Evidence from the Field', *Journal of Financial Economics*, 60(2), 187-243.

Hocking, R. R. and Leslie, R. N. (1967) 'Selection of the Best Subset in Regression Analysis', *Technometrics*, 9(4), 531-540.

Hwang, R.-C., Chung, H. and Chu, C. K. (2010) 'Predicting Issuer Credit Ratings Using a Semiparametric Method', *Journal of Empirical Finance*, 17(1), 120-137.

Hwang, R. C. (2013) 'Forecasting Credit Ratings with the Varying-Coefficient Model', *Quantitative Finance*, 13(12), 1947-1965.

Jacobson, T. and Karlsson, S. (2004) 'Finding Good Predictors for Inflation: A Bayesian Model Averaging Approach', *Journal of Forecasting*, 23(7), 479-496.

Lamon, E. C. and Clyde, M. A. (2000) 'Accounting for Model Uncertainty in Prediction of Chlorophyll in Lake Okeechobee', *Journal of Agricultural, Biological, and Environmental Statistics,* 5(3), 297-322.

Mizen, P. and Tsoukas, S. (2012) 'Forecasting US Bond Default Ratings Allowing for Previous and Initial State Dependence in an Ordered Probit Model', *International Journal of Forecasting,* 28(1), 273-287.

Pasiouras, F., Gaganis, C. and Zopounidis, C. (2006) 'The Impact of Bank Regulations, Supervision, Market Structure, and Bank Characteristics on Individual Bank Ratings: A Cross-Country Analysis', *Review of Quantitative Finance and Accounting*, 27(4), 403-438.

Rösch, D. (2005) 'An Empirical Comparison of Default Risk Forecasts from Alternative Credit Rating Philosophies', *International Journal of Forecasting*, 21(1), 37-51.

Schapire, R. E. (1999) 'A Brief Introduction to Boosting', *in Proceedings of the 16th international joint conference on Artificial intelligence - Volume 2*, Stockholm, Sweden, 1624417: Morgan Kaufmann Publishers Inc., 1401-1406.

Sena, V. (2002) 'Empirical Corporate Finance: Volumes I, II, III and IV. Edited by Brennan (Michael J.)', *112(Generic)*, F358-F360.

Shmueli, G. (2010) 'To Explain or to Predict?', *Statistical Science*, 25(3), 289-310.

Shmueli, G. and Koppius, O. R. (2011) 'Predictive Analytics in Information Systems Research', *MIS Quarterly*, 35(3), 553-572.

Tian, S., Yu, Y. and Guo, H. (2015) 'Variable Selection and Corporate Bankruptcy Forecasts', *Journal of Banking and Finance*, 52, 89-100.

Tibshirani, R. (1996) 'Regression Shrinkage and Selection via the Lasso', *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288.

Tsoukas, S. and Spaliara, M.-E. (2014) 'Market Implied Ratings and Financing Constraints: Evidence from US Firms: Market Implied Ratings and Financing Constraints', *Journal of Business Finance & Accounting,* 41(1-2), 242-269.

van de Geer, S. A. (2008) 'High-Dimensional Generalized Linear Models and the Lasso', *The Annals of Statistics*, 36(6), 614.

Zou, H. (2006) 'The Adaptive Lasso and Its Oracle Properties', *Journal of the American Statistical Association*, 101(476), 1418-1429.

Chapter 2 Modelling market implied ratings using LASSO variable selection techniques

Abstract

Both investors and rating agencies are interested in the accurate forecasting of corporate credit ratings to manage credit risk. In this chapter, financial factors, market-driven indicators and macroeconomic predictors are applied to determine market implied credit ratings. Adding a variable selection technique, the least absolute shrinkage and selection operator (LASSO) into reduced-form models can significantly improve predictive power. Moreover, when we compare our LASSO models with the benchmark ordered probit model, it can be shown that the former models have superior predictive ability and outperform the latter model in all out-of-sample predictions.

Key words: Market implied ratings, LASSO, Financial ratios, Forecasting

2.1 Introduction

Credit ratings are regarded as a measurement of the creditworthiness of a firm, and they are widely used to quantify the credit risk for the firm's external investors. They show the likelihood that a given borrower will default. They are generally issued by rating agencies such as Standard and Poor's, Moody's and Fitch according to the assessment of a firm's ability and willingness to fulfil debt servicing obligations in a specific period. This ability has significant potential to affect the pricing of credit risk and the allocation of investment strategies. According to the Financial Crisis Inquiry (2011), credit rating agencies as a key factor added to the most recent financial crisis in the US, since the updating of ratings was unable to adjust quickly to market changes. The users of credit ratings blindly rely on these ratings as early warning signals to identify credit risks in financial markets, which become the other element of this financial turmoil. Hence, rating accuracy and the process by which firms are assigned their ratings have been widely contested in the US. Fitch has recently developed a new model to derive Market Implied Ratings (MIRs) from bond and equity prices. The appealing advantage of these ratings in comparison with the conventional agency ratings is that they adjust instantaneously to market changes.

In this chapter, we aim to develop methodological extensions by adding a variable selection approach, the least absolute shrinkage and selection operator (LASSO), and its most promising derivation, the Elastic net, to ordered probit and continuation ratio models. The task in hand involves forecasting Fitch's CDS and Equity implied ratings (CDSIRs and EQIRs respectively hereafter). The major research contribution is to exploit LASSO properties and define the underlying structure of CDSIRs and EQIRs.[1] Firm-specific ratios from accounting reports and other publicly available information have been implemented in previous studies to predict credit ratings. The most important determinants for predicting bond ratings are identified by various techniques (OLS, multinomial and ordered logit/probit models) in this work (see for instance the early studies by Pogue and Soldofsky 1969, Pinches and Mingo 1973, Kaplan and Urwitz 1979, Kao and Wu 1990). The results suggest that a firm's financial situation and the economic

[1] With respect to the latter aim of this study, for reasons of space we do not report estimated coefficients of the prediction models. These results are available upon request.

environment can influence its ratings and the forecasting of default. Mizen and Tsoukas (2012) confirm the importance of capturing dynamic settings in ratings during model estimation and demonstrate that controlling state dependence can increase the percentage of correct prediction by the models.

In another line of credit ratings research, an increasing number of applied rating predictors as input in a model can be observed in literature, since researchers tend to capture more economic theories to define or forecast ratings. This gives us the opportunity to ask whether the predictors used are relevant in a piece of work. On the one hand, if only the subset of applied factors is related to ratings, it means that potentially important determinants of ratings are omitted, leading to a decrease in prediction accuracy. On the other hand, when the extensive set of predictors can explain rating, the multi-collinearity problem is likely to be met, resulting in biased estimation. If the multi-collinearity problem is not at work, a sparse representation cannot be provided due to a large number of applied predictors, which means that these models cannot be easily explained and readily used by market participants and rating agencies.

Our methodology carefully follows the literature that examines the determinants of credit ratings, but we add to it in two important ways. First, a methodological contribution is made by deriving a simple and more intuitive, yet innovative model, which is based on the variable selection technique, developed by Tibshirani (1996)—the least absolute shrinkage and selection operator (LASSO). Fan and Li (2001) and Tian *et al.* (2015) indicate that the most important predictors can be automatically selected from an extensive set of candidate variables and the predictive performance can simultaneously be improved in LASSO. Furthermore, LASSO does not depend on strict presumptions such as a pre-selection of the variables considered, and it is statistically consistent even under infinity observations (van de Geer 2008). It should be noted that LASSO can solve the problem of multi-collinearity that is fairly common in probit/logit models. In addition, LASSO technique is computationally efficient even when a large set of potential predictors are used. Our study is the first to provide a systematic empirical analysis of LASSO techniques in MIRs forecasts. In doing so, we investigate the relative importance of several time-varying covariates from an extensive set of firm-specific factors, market-specific predictors and macro-economic indicators applied to predict market implied ratings. After model estimation, a parsimonious set of predictors will be produced which can be readily employed by investors, managers and credit risk agencies.

Second, our panel dataset is constructed by market implied ratings instead of the standard long-term ratings. Compared with long-term ratings, MIRs represent an innovation to the ratings industry to address the issue of staleness in their long-term counterparts. Rösch (2005) and Tsoukas and Spaliara (2014) state that MIRs depend on proprietary and data-intensive rating models that incorporate market information into a model-based credit assessment. The most attractive characteristic of these ratings is that they can adjust immediately to market changes. Hence, we build on the foundations of the literature on implied ratings by investigating the forecasting power of models that capture volatile market changes.

To preview our findings, market implied ratings can be explained by several financial factors along with market-driven and macroeconomic indicators. Importantly, applying the LASSO techniques can considerably enhance the predictive power of our models in out-of-sample predictions compared to the benchmark (ordered probit model) which is commonly adopted in the literature. Moreover, the predictive performance of the LASSO models controlled by BIC-type tuning parameter selector are superior to their LASSO counterparts with AIC-type selector for the dataset and periods under our study. Thus, we suggest that LASSO-selected models can be implemented in future research to produce improved rating prediction.

The rest of the chapter is structured as follows. The literature of forecasting credit ratings and variable selection techniques is presented in section 2. We discuss the data and summary statistics in section 3. Following that, we describe our methodology in section 4. The empirical results and robustness tests are reported in section 5. The main conclusion is demonstrated in the final section.

2.2 Literature

The ways in which rating agencies use public information in setting quality ratings and researchers provide accurate prediction of ratings in the literature are contentious issues. Verifying the categories of credit ratings of companies is the extended work of identifying the default probabilities of companies in credit risk management. Horrigan (1966) first introduced the work of predicting bond ratings through accounting data and financial ratios from balance sheets and subordination. To provide a better set of financial ratios explaining credit ratings, Horrigan tested different combinations of predictor and kept the best one with the highest value of R-squared. This set was constructed by total assets, net worth to the book value of total debt, net operating profit to sales, working capital to sales adjusted industrial effects, and sales to net worth adjusted industrial effects. A dummy variable represented the subordination status of bond ratings, which produces about 60% correct out-of-sample prediction of newly issued and changed ratings. Unlike Horrigan's study (1966), Pogue and Soldofsky (1969) estimated the conditional probability that a bond would be assigned the higher rating by employing financial ratios and then converted the estimated probability into different rating categories to assess the predictive performance. In their conclusion, they demonstrate that leverage, the variation of earnings, the size of firm and profitability can influence bond ratings. Since bond ratings are related to the default risk, West (1970) adopted the four components of risk premiums (earnings volatility, capital structure, reliability and marketability) from Fisher's study (1959) into Horrigan's study to predict bond ratings again. He showed that the predictive accuracy of the Fisher model is higher than that of Horrigan's model.

With the development of econometrics, some researchers emphasised using multiple discriminant analysis (MDA) to categorize bonds into different rating levels. Based on the theory of MDA, there is no need to transform ordinal data into interval scale since the analysis of MDA focuses on differences in each category in dependent variables. In other words, MDA ignores the ordinal properties of ratings and just regards different rating categories as various segmentations of a single risk dimension. Pinches and Mingo (1973) implemented two methodological steps to categorize the ratings of bond issues. The factor analysis was used first to identify relatively important predictors and then these predictors were employed in multiple discriminant analysis to evaluate newly issued bond ratings. They concluded that about 70% correct predictions in actual ratings can be observed and around 60% correct forecasts in newly issued bonds. To find more potential relationships between credit ratings and financial ratios, Altman and Katz (1976) included 30 independent variables in MDA to determine the bond ratings of electric firms and confirmed that 14 predictors survived in the model and the predictive power in out-of-sample dataset increased to around 75%. Ang and Patel (1975) compared the predictive performance of methodologies applied in the studies of Horrigan (1966), Pogue and Soldofsky (1969), West (1970) and Pinches and Mingo (1973) in bond ratings. They confirmed that these models could be used in the prediction of short-term financial loss probability but they did not perform well in long-term prediction.

McKelvey and Zavoina (1975) developed a model known as the ordered probit model, for ordinal dependent variables, which assumed that the unobserved dependent variables were on an interval scale that could be measured by a linear model. They then divided these dependent variables with different cut-points to detect the ordinal dependent variable. This model provided a new opportunity for researchers to determine the ordinal ranks of credit ratings, and it is frequently used as a benchmark model to analyse credit ratings. Kaplan and Urwitz (1979) argued the disadvantages of using OLS and MDA models for analysing bond ratings and then implemented the probit models to estimate bond ratings. They point out that the results of financial ratios are similar to the conclusions reached by Pinches and Mingo (1973) and the market beta is one determinant of ratings. The predictive ability of the probit model is not significantly different from previous studies. Ederington (1985) further compared the application of four models on bond ratings, which contained a linear regression model, an ordered probit model, a linear discriminant model and a multinomial logit model. He confirmed that the ordered probit model clearly outperformed the regression model and the unordered logit model clearly dominated the discriminant model. Gentry *et al.* (1988) confirmed the predictive power of cash flow ratios during bond ratings categorization by using the probit model and suggested that involved inventories, other current liabilities, dividends, long-term financing and fixed coverage charges should be considered in the future prediction of bond ratings.

Previous research frequently emphasised analysing credit ratings under the assumption that rating standards are consistent. With an increasing number of downgrades in bond ratings, the performance of the debt market is gradually attracting researchers' attention. Blume *et al.* (1998) tried to confirm whether the credit quality of corporate debt had worsened over time. They used three-year averages of financial ratios in an ordered probit model and detected that rating standards had become stricter. In this situation, they reported that firms were less likely to be assigned higher ratings levels in the mid-1980s and early 1990s. Following the variable selection criterion in the work of Blume *et al.* (1998), Poon (2003) constructed the global sample dataset by averaging three-year financial ratios and then applying the ordered probit model to detect the indicators of credit ratings. He demonstrated that a government's credit risk is an important determinant of long-term ratings and it has a positive relationship with long-term ratings.

From the abovementioned literature, it can be seen that different statistical models capturing the characteristics of a dataset provide various ratings predictions and researchers cannot successfully suggest a preferred model. Kamstra *et al.* (2001) applied ordered logit regression (similar to the ordered probit model) combining methods to improve the predictive accuracy of bond ratings in the transportation and industrial sectors and confirmed that combined forecasts outperform their input forecasts. More recently, ordered probit methodologies have been commonly applied by for instance Amato and Furfine (2004) Hwang *et al.* (2009) and Hwang *et al.* (2010) to forecast credit ratings.

Amato and Furfine (2004) confirmed that credit ratings change based on the state of the business cycle, by analysing US firms in the ordered probit model. Moreover, they demonstrated procyclicality in the ratings of investment grade firms or newly assigned firms. Hwang *et al.* (2009) documented an analysis on the S&P's long-term issuer credit rating. They used a stepwise selection model on 24 latent independent variables to select relatively important ones before applying a mollified ordered probit model. In particular, there were two market-driven variables, three accounting variables and industry effect variables in the final list of independent variables. The estimated coefficients of final predictors met expectations. The predictive accuracy of long-term ratings can reach up to around 72.84% and 77.16% using different cut-off values in the ordered probit model. Hwang *et al.* (2010) constructed a prediction model by changing the linear regression function in the common ordered probit model into a semiparametric function and confirmed this developed model can outperform the original one in out-of-sample prediction. It was identified in both studies that some indicators are essential in forecasting credit ratings, such as the size of the company, balance sheet position, stock market performance and industry effects.

Altman and Kao (1992) report strong evidence that there was some serial correlation of ratings changes if the primary change was a downgrade, but no autocorrelation when the primary change was an upgrade, according to the two time periods examined, from 1970 to 1979, and from 1980 to 1985. As previously indicated, the balance sheet always shows the performance of a firm in a specific previous time period. This means that current credit ratings may be related to previous credit ratings and other previous accounting information about the corresponding firm. Parnes (2007) confirmed this changing tendency using several nonlinear models, which illustrate a clear positive autocorrelation among downgrade probabilities. By using the internal correlations model he also showed that time-series fluctuation is a more significant property in credit ratings than cross-ratings correlations.

Hwang (2013b) considered the autocorrelation of credit ratings in the dynamic ordered probit model (DOPM). In the empirical results, the out-of-sample total error rate was smaller and there was lower volatility compared with the simple DOPM under independence assumption. Hwang (2013a) further modified the DOPM by using smooth functions of macroeconomic variables to represent coefficients of firm-level independent variables in the DOPM, which is called the dynamic ordered varying-coefficient probit model (DOVPM). The latent coefficient functions in DOVPM can be measured by applying a local maximum likelihood method. The conclusion of this paper illustrated that the macroeconomic situation definitely influences the changing tendency of firm-level predictors. At the same time, the performance of DOVPM is obviously better than DOPM when it comes to predictive accuracy and total error rates of the prediction in the out-of-sample dataset. Both studies provide evidence that capturing the dynamic features of long-term ratings can increase predictive accuracy (see Hwang 2013b, Hwang 2013a). This conclusion is further supported by Mizen and Tsoukas (2012), who documented that the predictive performance in long-term ratings can be improved by considering the persistence of ratings. Tsoukas and Spaliara (2014) attempted to discover the relation between market implied ratings and financial constraints by adopting the ordered probit model. Their conclusion was that financial ratios are critical in ratings forecasts if a firm is likely to encounter financial difficulties.

Moving to the literature of implied ratings, almost all studies focus on detecting the differences between long-term agency ratings and market implied ratings. Breger *et al.* (2003) indicated that bond spreads have explanatory ability to define the cut-off points of categorized bonds and suggested that implied ratings perform better in identifying default probability compared with other ratings. After that, Rösch (2005) confirmed this finding and demonstrated that the default probability can be appropriately reflected by implied ratings in comparison with standard ratings. Castellano and Giacometti (2012) observed that the changes in credit ratings could be predicted by market implied ratings.

The variables selection techniques are not widely applied in determining or predicting credit ratings. Only a small amount of work can be found that studies bankruptcy. Amendola *et al.* (2012) first employed the LASSO approach to estimate the default probabilities of Italian companies in the limited liability sector. According to their conclusion, the LASSO approach can provide a superior predictive performance and more stable error rates than previous default models. Härdle and Prastyo (2013) aimed to forecast the default probability of Southeast Asian firms by adding LASSO and elastic net in the logit model and confirmed that adding these penalty functions into the model can clearly improve the predictive accuracy. To provide general evidence supporting the predictive ability of LASSO on default risk, Tian *et al.* (2015) studied a comprehensive U.S. bankruptcy database and concluded that accuracy in out-of-sample prediction is superior than in previous models (like reduced-form models with the distance to default) in estimating default by combining reduced model with the LASSO technique.

Importantly, the first study of LASSO was developed by Tibshirani (1996). He suggested two reasons why LASSO should be created on the basis of Ordinary Least Squares (OLS). First, OLS estimates have low bias but large variance. However, in LASSO, non-relevant variables in the regression would be forced to 0, which allows bias and reduces the variance of predicted value. It can improve the predictive power of OLS. The second reason is the power of interpretation. Faced with vast predictors in OLS estimates, there exist complex and important elements illustrating how predictors affect the dependent variable. In LASSO, due to zero coefficients of some predictors, the strongest effects can be determined objectively, which make the analysis more reasonable.

Zou and Hastie (2005) combined LASSO with ridge regression to produce the Elastic net, which is a good hybrid of sparsity and regulation. There exist some advantages of using the Elastic net. First, it can select correlated independent variables in the grouping effect. In addition, the penalty function is strictly convex, thus a unique global minimum would be calculated. Finally, it is useful to solve the problem that the number of predictors is bigger than the number of observations. Although the Elastic net is not very widely used in econometrics, it has been employed on microarray data and in gene selection.

From the above review, the determinants of credit ratings prediction are provided, which is helpful to identify relevant predictors in market implied ratings. Meanwhile, the development of econometric methods in credit ratings is shown, which gives us the opportunity to make a methodological contribution. We will discuss the applied dataset and estimation strategy in the following sections.

2.3 Data and summary statistics

2.3.1 Data sources

Fitch's database is employed to extract the data on market implied ratings, which refers to solicited ratings for all traded companies in the US. This database provides information on the CDS and Equity implied ratings assigned to each issuer as well as the date that the rating became available. From 2002 to 2008, both CDS and Equity implied ratings are reported on a monthly basis.[2] Following normal practice in the literature, we categorize our firms into rating buckets without consideration of notches (i.e. + or -). According to the studies of Amato and Furfine (2004) and Mizen and Tsoukas (2012), this classification considers large cumulative changes of ratings rather than small movements notch by notch and avoids very few observations in one rating category. Therefore, seven rating

[2] The research aims to study the structure and predictability of the implied ratings in the years preceding the recent global financial crisis. Our choice to focus on a time window ending in 2008 is motivated by two important considerations. First, the global financial crisis and the collapse of Lehman Brothers constituted a shock that may have acted as a confusing factor in the determination of credit ratings. In fact, it can be argued that the misinterpretation of the credit ratings by investors was one of the main contributors to the crisis. Second, the data were downloaded early in 2010 from a research project supported by Fitch: the coverage period is therefore 2002-2008.

categories are created, ranging from AAA to Below CCC, which are assigned numerical values in Table 2-1, starting with 1 to AAA, 2 to AA…, 7 to Below CCC.[3]

Table 2-1 Rating categories

Market Implied Ratings	Corresponding numerical values
AAA	1
AA	2
A	3
BBB	4
BB	5
B	6
Below CCC	7

Notes: The table presents the rating categories and the corresponding numerical values.

In our work, the independent variables can be divided into three parts: accounting variables, market-driven indicators and macroeconomic predictors. Firm-specific accounting data are taken from Fitch's Peer Analysis Tool. Quarterly corporate historical data for all companies rated by Fitch can be accessed in this database. MIRs of firms are linked with Fitch's balance sheet statements and profit and loss accounts of corresponding firms. Hence, we merge the monthly MIRs data and quarterly firm-level accounting data to construct our dataset.[4] For the final dataset, each firm with CDSIRs and EQIRs data and financial and market data can be observed every month. Following applied selection criteria in the literature, we keep companies with complete records on our explanatory variables and firm-months without negative sales and assets. To reduce the potential influence of outliers, the regression variables are winsorized at the 1st and 99th percentiles.

Monthly data on market indicators and macroeconomic variables are downloaded from Bloomberg. Our combined sample ultimately comprises data for 211 firms operating in all sectors of the US economy except agriculture, forestry and fishing and public administration. The number of observations on each firm in this unbalanced panel vary between 1 and 63. Two features of our dataset make our

[3] In EQIRs we do not observe any ratings in the last category, hence six groups are generated for this type of implied ratings.

[4] For each firm, the quarterly data is repeated every month during the same quarter if the monthly data is not observed.

work attractive. First, both investment and speculative grades ratings are included in our dataset, which covers the whole spectrum of firms in line with previous studies (see for instance Amato and Furfine 2004). Second, there is no significant difference between the distribution of standard long-term ratings in CDS data and the distribution of agency ratings in the general bond population (see Reyngold *et al*. 2007). Thus, our empirical analysis from both the CDSIRs and the EQIRs databases has a representative base.

2.3.2 Choice of explanatory variables

Both business and financial risks are incorporated into previous empirical research to determine credit ratings. The business risk is a measure of industry characteristics, firm size, management capability and organizational indicators. Financial risk is related to the quality of a firm's accounting procedures, profitability, cash flow situation and its overall financial policy. Market-related information is also considered in candidate models. Meanwhile, we check the material issued by rating agencies, especially Fitch, to verify what matters when assigning a market implied rating. In other words, the extensive selection of our explanatory variables is guided both from the existing empirical literature (see for example Kaplan and Urwitz 1979, Ederington 1985, Poon 2003, Chava and Jarrow 2004, Amendola *et al*. 2012, Mizen and Tsoukas 2012, Hwang 2013a, Creal *et al*. 2014, Doumpos *et al*. 2015, Tian *et al*. 2015), and the common practice of rating agencies (see Liu *et al*. 2007, Reyngold *et al*. 2007).[5]

2.3.2.1 Firm-specific variables

We apply 16 firm-specific accounting variables as potential predictors of ratings, which measure different aspects of firms' financial health such as size, leverage, coverage, cash flow, profitability and liquidity. Specifically, the firm size (DETA) is calculated by the natural logarithm of firms' real total assets, which indicates the scale of the firm. When firm size increases, the rating of the firm would be expected to improve. Next, we measure leverage using a number of ratios: The ratio of total assets over equity (AE), the ratio of long-term debt over total assets (LDA), the ratio of short-term debt to total assets (SDA), the ratio of total debt to

[5] The expected relationship between these variables and MIRs is presented in Table A1 in Appendix A. The Table also provides a detailed description of the variables used in this study.

total assets (TDA), and the ratio of total debt to earnings before interest, taxes, depreciation, amortization, and restructuring or rent costs (TDEBITDA). The negative relationship between leverage and MIRs would be expected, since higher values for these ratios are likely to increase financial risk and hence should worsen the rating. The next two measures proxy the creditworthiness of the firm, as they show the firm's ability to generate income to meet interest rate obligations: the ratio of earnings before interest and tax over interest expenses (EBITINT) and the ratio of earnings before interest, taxes, depreciation, amortization and restructuring or rent costs to interest expenses (EBITDAINT). If these ratios increase, this firm would have more ability to cover its obligations and reduce default risk, which would improve its rating. Further, cash flow is measured by the following ratios: cash flow from operating activities over total assets (CFOA), and cash and equivalent over total assets (CASHEQA). The positive relationship between these ratios and rating upgrades would be expected, since more cash flow would reduce liquidity risk and then improve ratings. The next five ratios explain the profitability of a firm: the ratio of operating income to net sales (OM), the ratio of net income without dividends over total capital (ROC), the ratio of net income over shareholders' equity (ROE), the ratio of net income over total assets (ROA) and the ratio of the funds from operations to total debt (FFD). With the improvement in operation performance, more profitability should be expected, which implies an improvement of ratings. Finally, liquidity is expressed by the ratio of cash from operations to liabilities (LIQ), which shows the ability of a firm to satisfy its short-term obligations as they become due. A firm operating with higher levels of liquidity is likely to be assigned the higher-level rating.

2.3.2.2 Market-driven indicators

Since stock prices have the ability to reflect publicly available information, it is reasonable that market-driven variables can affect the rating of a company as indicated previously. Thus, we apply the following market indicators: excess return (EXRET) as calculated by the monthly stock return on a firm minus the S&P 500 index return and the relative size of a firm in the market (RSIZE) measured by each firm's market equity value over the total market equity value. Both variables are expected to be positively associated with the improvement of ratings. Next, the volatility of stock return (STD) is expressed by the standard deviation of each company's monthly stock returns. The systematic risk of each firm (BETA) is employed, which is extracted from the Capital Asset Pricing Model for each firm. Finally, the 1-year and 5-year default probabilities (PD1 and PD5) are taken from Fitch's Peer Analysis Tool. An increase in these variables indicates the growth of default probability, which would worsen ratings.

2.3.2.3 Macroeconomic influences

An extensive set of macro-economic variables as potential predictors is applied to determine market implied ratings. Specifically, returns on the S&P 500 index (RLSP) are calculated as the evaluation of the stock market performance. Short-term interest rates are measured by the three-month commercial paper rate (CPFFM), three-month Treasury bill rate minus federal funds rate (TB3) and the one-year constant maturity treasury rate (GS1). The general price level is employed, as measured by the growth rate in the narrow money stock (MB) and inflation rate (INFL). Aggregate economic activity is proxied by the rate of change in industrial production (DLIP), the index of the growth rate of real GDP (DLGDP), the average of monthly Chicago Fed National Activity Index over the year (CFNA), the average monthly unemployment rate over the year (UNRATE) and the Chicago Board Options Exchange (CBOE) volatility index (VIX). All macro-variables, apart from VIX, are reported in percentages. The eleven macroeconomic variables measure different aspects of the aggregate economy's performance. Their relationship with the market implied ratings could be either positive or negative since ratings tend to improve during good times, but agencies have been observed to tighten their standards during these periods. Hence, the relationship between ratings and macro-economic variables is an issue that will be determined empirically.

2.3.3 Summary statistics

The distribution of firms by rating categories for CDSIRs and EQIRs are reported in Table 2-2 and Table 2-3 respectively. Based on these tables, there is no significant difference in the distribution of firms across the rating categories and most firms are assigned A and BBB ratings.

Table 2-2 CDSIRs of firms by year

	AAA	AA	A	BBB	BB	B	Below CCC	Number of Observations
2002	4	35	46	32	20	1	0	138
2003	10	44	68	65	32	7	0	226
2004	11	41	60	70	35	10	0	227
2005	9	34	57	71	29	10	1	211
2006	9	24	52	63	26	4	1	179
2007	14	45	54	63	28	13	6	223
2008	12	19	13	33	13	3	1	94
Number of Observations	69	242	350	397	183	48	9	1298

Note: This table presents the distribution of firms by rating category for CDSIRs by year.

Table 2-3 EQIRs of firms by year

	AAA	AA	A	BBB	BB	B	Below CCC	Number of Observations
2002	1	13	55	103	76	16	0	264
2003	2	19	91	106	43	5	0	266
2004	1	20	100	99	42	6	0	268
2005	0	10	70	87	38	4	0	209
2006	0	11	80	74	29	4	0	198
2007	1	19	91	80	41	8	0	240
2008	0	5	29	37	23	1	0	95
Number of Observations	5	97	516	586	292	44	0	1540

Note: This table presents the distribution of firms by rating category for EQIRs by year.

At the next stage, summary statistics for our explanatory variables are documented in Table 2-4 and Table 2-5. To capture any differences across ratings categories, the sample is separated into investment grades and non-investment grades and then statistics are presented. We report p-values for the tests of equality of means across the above-mentioned groups in the last columns of the tables. It can be observed that firms in the investment grade group experience better financial conditions, as measured by the accounting ratios from the balance sheet. This is consistent with our expectations. The significant differences between the two groups can be confirmed from tests, which suggest that there exists a link between better financial health and an improved rating. In other words, the cross-sectional variation can be observed in market implied ratings.

Moving to the market indicators, it can be confirmed that improved market conditions are associated with the rating changes, which also suggests a relation between the market climate and the ratings.

Table 2-4 Descriptive statistics-CDSIRs

Variable	Mean	Standard Deviation	Minimum	Maximum	p-value
	(1)	(2)	(3)	(4)	(5)
DETA					
Investment grade	9.6388	1.0132	7.2272	12.2084	
Non-investment grade	8.9536	0.9959	6.2539	12.2087	0.0000
AE					
Investment grade	3.2256	3.6336	1.3324	73.7340	
Non-investment grade	5.0224	8.8729	1.3283	123.5602	0.0000
LDA					
Investment grade	20.2416	11.2733	0.0000	79.3983	
Non-investment grade	30.1868	19.4641	0.0000	110.4453	0.0000
SDA					
Investment grade	3.5941	3.9838	0.0000	23.5410	
Non-investment grade	3.2961	4.1573	0.0000	23.4635	0.0000
TDA					
Investment grade	24.3546	11.9238	1.3017	87.0595	
Non-investment grade	37.4629	20.8746	0.6518	126.8760	0.0000
TDEBITDA					
Investment grade	2.3395	1.2107	0.3200	19.6400	
Non-investment grade	4.1208	2.8918	0.3200	23.0700	0.0000
EBITINT					
Investment grade	13.3805	19.0270	0.1541	209.3023	
Non-investment grade	7.2685	15.2730	0.1248	210.4054	0.0000
EBITDAINT					
Investment grade	16.0444	21.9950	0.6500	235.4000	
Non-investment grade	9.0872	14.5213	0.3100	196.7000	0.0000
CFOA					
Investment grade	6.9186	6.2625	-41.0623	38.4372	
Non-investment grade	7.1077	7.2097	-13.4106	65.9955	0.0000
CASHEQA					
Investment grade	8.6116	9.6531	0.0238	71.8277	
Non-investment grade	6.5258	8.1009	0.0030	64.0272	0.0000
OM					
Investment grade	13.9496	9.7304	-13.5155	53.0189	
Non-investment grade	10.3893	9.9617	-20.2276	52.6046	0.0000
ROC					
Investment grade	3.7919	5.3598	-34.1330	35.7771	
Non-investment grade	2.1389	7.0866	-36.3880	34.5209	0.0000
ROE					

Investment grade	12.5162	35.3349	-361.1511	452.2565	
Non-investment grade	9.5597	46.9441	-359.7868	516.7883	0.0000
ROA					
Investment grade	3.9430	4.4180	-26.4074	22.0518	
Non-investment grade	2.5165	5.5633	-23.8888	23.2336	0.0000
FFD					
Investment grade	40.7412	32.5483	-16.2800	267.1700	
Non-investment grade	24.4291	26.0776	-17.8000	225.3200	0.0000
LIQ					
Investment grades	12.02064	10.94679	-13.3151	59.8763	
Non-investment grade	11.55889	12.65565	-13.56745	62.18169	0.0000
EXRET					
Investment grade	0.0117	0.0646	-0.3525	0.4527	
Non-investment grade	0.0195	0.1108	-0.3526	0.4673	0.0149**
RSIZE					
Investment grade	0.2170	0.2641	0.0103	1.7570	
Non-investment grade	0.1107	0.1812	0.0041	1.7495	0.0000
STD					
Investment grade	0.0159	0.0079	0.0041	0.1134	
Non-investment grade	0.0256	0.0151	0.0041	0.1141	0.0000
BETA					
Investment grade	0.9366	0.6368	-0.8663	4.4144	
Non-investment grade	1.0795	0.9672	-0.8799	4.9196	0.0000
PD1					
Investment grade	24.0820	70.0007	2.0000	3000.0000	
Non-investment grade	162.7669	489.4410	2.0000	3000.0000	0.0000
PD5					
Investment grade	260.5300	310.3217	14.0000	4495.0000	
Non-investment grade	793.0404	907.6053	14.0000	5464.0000	0.0000

Notes: This Table reports the summary statistics of the explanatory variables used in the empirical models. Column 5 reports the p-value for the test of equality of means between the investment grade and non-investment grade categories. Investment grade refers to ratings from AAA to BBB. Non-investment grade refers to ratings BB and below. A detailed description of the variables used in this study is given in Table A1 in the online Appendix A.

Table 2-5 Descriptive statistics-EQIRs

Variable	Mean	Standard Deviation	Minimum	Maximum	p-value
	(1)	(2)	(3)	(4)	(5)
DETA					
Investment grade	9.5314	1.0515	7.0031	12.2084	
Non-investment grade	8.8701	0.9461	6.2539	12.2087	0.0000
AE					
Investment grade	3.4552	4.7315	1.3283	116.1204	
Non-investment grade	5.3173	9.3934	1.3283	123.5602	0.0000
LDA					
Investment grade	20.9162	12.5949	0.0000	110.1548	
Non-investment grade	32.5301	19.9743	0.0000	110.4453	0.0000
SDA					
Investment grade	3.7397	4.1512	0.0000	23.5410	
Non-investment grade	3.0151	3.9714	0.0000	23.4635	0.0000
TDA					
Investment grade	25.3924	13.5320	1.3017	126.8209	
Non-investment grade	40.3718	21.1606	0.6518	126.8760	0.0000
TDEBITDA					
Investment grade	2.5062	1.4809	0.3200	20.4000	
Non-investment grade	4.4812	3.0369	0.3200	23.0700	0.0000
EBITINT					
Investment grade	13.2588	20.3637	0.1601	210.4054	
Non-investment grade	5.3897	10.2101	0.1248	157.3792	0.0000
EBITDAINT					
Investment grade	15.7294	21.3681	0.4100	235.4000	
Non-investment grade	7.2437	11.9978	0.3100	171.8000	0.0000
CFOA					
Investment grade	7.8973	6.5051	-41.0623	40.5546	
Non-investment grade	5.9413	7.0742	-41.0623	65.9955	0.0000
CASHEQA					
Investment grade	8.6834	9.6845	0.0030	71.8277	
Non-investment grade	5.7724	7.3325	0.0033	58.4741	0.0000
OM					
Investment grade	13.0445	9.5560	-16.9133	53.0189	
Non-investment grade	10.3826	10.3917	-20.2276	52.5504	0.0000
ROC					
Investment grade	3.8570	5.6971	-35.0754	35.7771	
Non-investment grade	1.5077	7.1322	-36.3880	34.5140	0.0000
ROE					
Investment grade	13.3395	36.3731	-361.1511	473.0769	
Non-investment grade	7.5924	48.9666	-359.7868	516.7883	0.0000
ROA					
Investment grade	4.1096	4.6843	-26.4074	23.2190	
Non-investment grade	1.8536	5.4585	-26.4074	23.2336	0.0000
FFD					
Investment grade	38.8160	31.7571	-16.2800	267.1700	
Non-investment grade	21.6264	24.5682	-17.8000	224.0300	0.0000
LIQ					
Investment grade	13.5432	11.8937	-13.5675	62.1817	
Non-investment grade	9.495881	11.69661	-13.56745	57.22284	0.0000
EXRET					
Investment grade	0.0140	0.0723	-0.3525	0.4648	
Non-investment grade	0.0181	0.1089	-0.3526	0.4673	0.5831
RSIZE					
Investment grade	0.2047	0.2687	0.0088	1.7570	
Non-investment grade	0.0922	0.1307	0.0041	1.6261	0.0000
STD					
Investment grade	0.0172	0.0085	0.0041	0.1134	0.0000

Non-investment grade	0.0250	0.0155	0.0041	0.1141	
BETA					
Investment grade	0.9081	0.7224	-0.8757	4.8848	
Non-investment grade	1.1028	0.9284	-0.8799	4.9196	0.0000
PD1					
Investment grade	24.1369	54.1782	2.0000	3000.0000	
Non-investment grade	215.5352	567.0797	2.0000	3000.0000	0.0000
PD5					
Investment grade	250.8945	299.6934	14.0000	4495.0000	
Non-investment grade	1010.6400	970.2219	17.0000	5464.0000	0.0000

Notes: This Table reports the summary statistics of the explanatory variables used in the empirical models. Column 5 reports the p-value for the test of equality of means between the investment grade and non-investment grade categories. Investment grade refers to ratings from AAA to BBB. Non-investment grade refers to ratings BB and below. A detailed description of the variables used in this study is given in Table A1 in the online Appendix A.

2.4 Methodology

We predict the changes in market implied ratings with ordered probit (OP) and continuation ratio (CR) models combined with LASSO or the Elastic net. The proposed methodology aims to select the most important predictors and provide accurate MIRs forecasts. LASSO, originally proposed by Tibshirani (1996), is an extended form of an OLS regression which performs both variable selection and regularization through a shrinkage factor. It is capable of enhancing the accuracy and interpretability of classical regression methods (Tibshirani 1996). To maintain the properties of LASSO and capture the ordinal ranking of MIRs, penalty functions from LASSO or its variant (Elastic net) are added into OP or CR models. This helps us reveal the relation between the potential predictors (at the firm and macro level) and identify their significance in predicting MIRs. As a benchmark to our study, we rely on the standard OP model. A description of the empirical modelling strategy follows.

2.4.1 OP

MIRs as a branch of credit ratings are discrete-valued signs and have an ordinal ranking. To meet the ordinal property of MIRs, OP is applied naturally as a benchmark in the relevant literature (Kaplan and Urwitz 1979, Gentry *et al.* 1988, Blume *et al.* 1998, Amato and Furfine 2004, Hwang *et al.* 2009). OP takes into account both the existence of ordinal ranking and the difference between any two adjacent ratings.[6] We define the categorical variable $y_{it} = 1, 2..., 7$ according to

[6] For details on the exposition of the OP, see Maddala (2008), pp 47-48.

the rating assigned to each firm. We assume that there is an unobservable dependent variable y_{it}^* associated with y_{it}, which can be expressed as:

$$y_{it}^* = X_{it}\beta_1 + X_{it-1}\beta_2 + X_{it-2}\beta_3 + X_{it-3}\beta_4 + W_{it}\beta_5 + W_{it-1}\beta_6 + W_{it-2}\beta_7 + W_{it-3}\beta_8 + Z_{it}\beta_9 + Z_{it-1}\beta_{10} + Z_{it-2}\beta_{11} + Z_{it-3}\beta_{12} + y_{it-1}\beta_{13} + \varepsilon_{it}, \quad (1)$$

where $i = 1, 2, \dots, N$ represents firms, and $t = 1, 2, \dots, T$ represents different time periods. In this context, t is the month end for monthly data. β are vectors of unknown parameters to be estimated. X denotes a set containing 16 accounting variables, which can be divided into broad groups of size, leverage, coverage, cash flow, profitability and liquidity. W and Z contain 6 market-driven variables and 11 macroeconomic variables, respectively. Following the literature (Güttler and Wahrenburg 2007), all predictors in X, W and Z are lagged three periods denoted by $t-1, t-2$ and $t-3$ to mitigate potential time tendency. To capture potential non-linear influences, we allow for non-linear transformation of the variables and therefore the square of each predictor is considered and included.[7] Thus, the total number of firm-specific accounting, market-driven and macroeconomic variables is 264[8]. y_{it-1} is an indicator of the firm's rating in the previous time periods. We consider 4 lags to control for state dependence. The concern about persistency in ratings is an important dimension of time-series variation and the use of models with lagged rating categories is the standard way of addressing this issue.[9] The error term ε_{it} in equation (1) is assumed to be a normally distributed residual with a zero mean and unit variance. In our data y_{it}^* is not observed. Thus, what is observed are the market implied ratings assigned to firms, which can take M

[7] This approach is justified theoretically since some variables may have a positive effect up to a certain (turning) point and a negative one thereafter.

[8] We consider 16 accounting variables, 6 market-driven variables, 11 macroeconomic variables and their first three lags (in total, 132 variables). We also consider the square of these variables, thus in equation (1) we have in total 268 variables made up by 264 firm-specific accounting, market driven and macroeconomic variables and 4 state dependence variables. We use the same set of predictors throughout.

[9] State dependence captures previous rating state and indicates the realization of a rating in the previous time period. Following Contoyannis *et al.* (2004) and Mizen and Tsoukas (2012), state dependence is controlled for by applying dummy variables representing the first lags of each category in the dependent variable. Given that we observe limited observations in the extreme high and low rating categories, these ratings in the state dependence variables are merged into 5 groups, such as above AA ratings, A rating, BBB rating, BB rating and below B ratings. In addition, to avoid the dummy variable trap we omit one baseline rating category. Therefore, the lagged rating category BBB is not included in the models because it is regarded as the baseline category of lagged MIRs.

values. The relation between the observed variable y_{it} and the latent variable y_{it}^* it is assumed to be given by:

$$y_{it} = m \; if \; \alpha_{m-1} < y_{it}^* \leq \alpha_m \; for \; m = 1, \ldots, M \tag{2}$$

For a set of parameters α_0 to α_M, where $\alpha_0 < \alpha_1 < \ldots < \alpha_M$, $\alpha_0 = -\infty$ and $\alpha_M = \infty$. Assuming a standard normal distribution for ε_{it}, the conditional probabilities can be derived as:

$$Pr(y_{it} = m) = \Phi(\alpha_m - X_{it}\beta_1 - X_{it-1}\beta_2 - X_{it-2}\beta_3 - X_{it-3}\beta_4 - W_{it}\beta_5 - W_{it-1}\beta_6 - W_{it-2}\beta_7 - W_{it-3}\beta_8 - Z_{it}\beta_9 - Z_{it-1}\beta_{10} - Z_{it-2}\beta_{11} - Z_{it-3}\beta_{12} - y_{it-1}\beta_{13}) - \Phi(\alpha_{m-1} - X_{it}\beta_1 - X_{it-1}\beta_2 - X_{it-2}\beta_3 - X_{it-3}\beta_4 - W_{it}\beta_5 - W_{it-1}\beta_6 - W_{it-2}\beta_7 - W_{it-3}\beta_8 - Z_{it}\beta_9 - Z_{it-1}\beta_{10} - Z_{it-2}\beta_{11} - Z_{it-3}\beta_{12} - y_{it-1}\beta_{13}), \tag{3}$$

Where $\Phi(.)$ is the standard normal distribution function. We can evaluate the above probabilities for any combination of parameters in the vectors α and β.

2.4.2 OP with LASSO

According to Tibshirani (1996), LASSO is a method of regression that enables estimation and variable selection simultaneously in a non-orthogonal setting. By controlling penalty power, LASSO selects variables by forcing some coefficients to zero and shrinking others. This reduces the variance of the estimated value and increases the accuracy of the regression prediction. The LASSO estimator resolves the l_1-penalized OP problem of estimating β by maximizing a likelihood function. In particular, the maximization of the likelihood proceeds subject to the constraint $\sum_{q=1}^{p} |\beta_q| \leq s$, where s is a user-specified tuning parameter and $q = 1, 2 \ldots p$ indicates the number of surviving predictors with non-zero estimated coefficients. This penalty corresponds to the L1 norm, and therefore it is often referred to as the L1 penalized model. The OP with LASSO (L1 penalized OP model) can then be expressed as:

$$\hat{\beta} = argmax_\beta \left(\ell(\beta | y_{it}, A) - \lambda \sum_{q=1}^{p} |\beta_q| \right), \tag{4}$$

where $\ell(\beta|y_{it}, A)$ is the likelihood function of OP and A contains the pool of the potential predictors (X_{it}, X_{it-1}, X_{it-2}, X_{it-3}, W_{it}, W_{it-1}, W_{it-2}, W_{it-3}, Z_{it}, Z_{it-1}, Z_{it-2}, Z_{it-3}, y_{it-1}). All explanatory variables are standardized before applying the LASSO estimator. In equation (4), λ stands for the tuning parameter. As λ increases, the sum of absolute values of estimated coefficients is reduced, and shrinkage of coefficients is achieved. If λ exceeds a threshold value in the corresponding models, some estimated coefficients are set to zero ultimately. This “L1 norm penalty” or the constraint formulation in LASSO generates a more interpretable and sparser model.

As already noted, compared with other independent variable selection methods, LASSO can provide more stable and restricted models (Tibshirani 1996, Fan and Li 2001, Zou 2006, Tian *et al.* 2015). It is also a computationally simple and efficient method (Efron *et al.* 2004). Several approaches, such as cross validation and information criteria, have been proposed in selecting latent models with minimum prediction errors or maximum log-likelihood estimation. Zou *et al.* (2007) provided an efficient approach for obtaining the optimal LASSO fit with the Akaike information criterion (AIC) (Akaike 1974) and the Bayesian information criterion (BIC) (Schwarz 1978). Sun and Zhang (2012) noted that the computational cost of applying cross-validation in penalized models is considerable, while the theory of applying cross-validation is poorly understood. Therefore, AIC and BIC are used in the present study to select the tuning parameter and further detect the “best” model among a series of candidate models in OP with LASSO.[10] The “best” model in variable selection procedure will be the one that achieves the minimum AIC or BIC value. The exact algorithm behind this process is presented in Appendix C. These models are benchmarked with their LASSO 10-cross validation counterparts.

2.4.3 OP with Elastic net

Elastic net is a LASSO variant introduced by Zou and Hastie (2004) that can further improve the accuracy of the estimation in the presence of highly correlated predictors. The Elastic net allows “grouping” of variables in the model by adding a l_2-penalty from ridge regression. The OP with Elastic net resolves the l_1-

[10] It is well-known that AIC and BIC have different properties in model selection (for details see Shao 1997, Yang 2005, Zhang *et al.* 2010).

penalized and l_2-penalized OP problem of estimating β to maximize the OP likelihood function. The maximization of the likelihood proceeds subject to the following constraints (penalty functions) $\sum_{q=1}^{p}|\beta_q| \leq s_1$ and $\sum_{q=1}^{p}({\beta_q}^2) \leq s_2$, where s_1 and s_2 are user-specified tuning parameters. These penalty functions correspond to the L1 and L2 norm. The OP with Elastic net model is presented below:

$$\hat{\beta} = argmax_{\beta}\left(\ell(\beta|y_{it}, A) - \lambda_1 \sum_{q=1}^{p}|\beta_q| - \lambda_2 \sum_{q=1}^{p}({\beta_q}^2)\right), \tag{5}$$

where $\ell(\beta|y_{it}, A)$ is the likelihood function of the OP and A contains the pool of the potential predictors (X_{it}, X_{it-1}, X_{it-2}, X_{it-3}, W_{it}, W_{it-1}, W_{it-2}, W_{it-3}, Z_{it}, Z_{it-1}, Z_{it-2}, Z_{it-3}, y_{it-1}).

Equation (5) can be converted as follows:

$$\hat{\beta} = argmax_{\beta}\left(\ell(\beta|y_{it}, A) - \lambda \sum_{q=1}^{p}\left\{\alpha|\beta_q| + \frac{1}{2}(1-\alpha)({\beta_q}^2)\right\}\right), \tag{6}$$

where $0 \leq \alpha \leq 1$.

Equation (5) is the vanilla version of Elastic net. The factor $\lambda_2 \sum_{q=1}^{p}({\beta_q}^2)$ allows correlated variables in the corresponding models, which are drawn from ridge regression. If α is equal to 0, the Elastic net keeps the l_2-penalty in the model in equation (6) (ridge regression). Similarly, if α is equal to 1, the l_1-penalty will only be kept in the Elastic net and equation (6) reduces to a simple LASSO estimator. The l_2-norm constraint ensures a unique global minimum in the strictly convex loss function. As in the OP with LASSO estimator, the AIC-type and BIC-type tuning parameter selectors are employed for selecting the model with the minimum value (see Appendix C). As before, all predictors are standardized before applying the Elastic net estimator.

2.4.4 CR with LASSO

The CR, originally proposed by Fienberg (1980), was designed for ordinal outcomes in which the categories represent the progression of events or stages in some process. It estimates the probability of one particular category given the categories preceding or following it. More specifically, it is centred on the binary choice on each ordinal category, which provides the conditional probability of estimating categories. Fienberg (1980), Hardin and Hilbe (2018) and Long and Freese (2006) argued that the CR is superior compared to the binary logistic regression. It is applicable in multi-classification problems where an individual can jump to the discrete rating category without having to pass the intermediate rating categories.[11] Similar to the binary logistic regression, the CR creates binary choices on each ordinal category and these choices make it possible to calculate the relevant conditional probabilities. The conditional probability that an individual drops a level, given that this individual has been at a higher level, is based on "conditional incremental thresholds". The CR may be regarded as an advanced version of the proportional odds model (the ordered logistic model), which preserves the parsimony of the cumulative odds model and considers the ordinal categories of MIRs. These assigned integer values of categories in the CR can be controlled by users, implying that the estimated coefficients in the CR are influenced by the direction chosen for modelling the response variable. In our work, the backward formulation of CR of Archer and Williams (2012) is applied. The progression through the levels of MIRs from investment grade quality (AAA-BBB) to sub-investment grade quality (BB-Below CCC) is expressed by increasing integer values. This helps estimate the odds of lower MIRs rating compared with higher MIRs rating. The above can be expressed as follows:

$$logit\big(Pr(y_{it} = m | y_{it} \leq m, X = A\,)\big) = log\left(\frac{Pr(y_{it} = m | y_{it} \leq m,\ X = A)}{Pr(y_{it} < m | y_{it} \leq m,\ X = A)}\right)$$

$$= X_{it}\beta_1 + X_{it-1}\beta_2 + X_{it-2}\beta_3 + X_{it-3}\beta_4 + W_{it}\beta_5 + W_{it-1}\beta_6 + W_{it-2}\beta_7 + W_{it-3}\beta_8 + Z_{it}\beta_9 + Z_{it-1}\beta_{10} + Z_{it-2}\beta_{11} + Z_{it-3}\beta_{12} + y_{it-1}\beta_{13}\,, \quad (7)$$

[11] Market implied ratings share this property.

where $\mathrm{A} = X_{it}, X_{it-1}, X_{it-2}, X_{it-3}, W_{it}, W_{it-1}, W_{it-2}, W_{it-3}, Z_{it}, Z_{it-1}, Z_{it-2}, Z_{it-3}, y_{it-1}$. In equation (7), the dependent variable y_{it} belongs to one of the ordinal rating categories m described in equation (2). For each unit observation, rather than modelling the response y_{it} directly, each variable y_{it} is equal to 1 if the response falls in category m, and 0 otherwise. Thus, conditional likelihood is calculated as in multiple logistic regressions. The above equation can be transformed into the following version to derive the conditional probability:

$$Pr(y_{it} = m | y_{it} \leq m) = \frac{e^{a}}{1 + e^{a}}, \tag{8}$$

where $a = X_{it}\beta_1 + X_{it-1}\beta_2 + X_{it-2}\beta_3 + X_{it-3}\beta_4 + W_{it}\beta_5 + W_{it-1}\beta_6 + W_{it-2}\beta_7 + W_{it-3}\beta_8 + Z_{it}\beta_9 + Z_{it-1}\beta_{10} + Z_{it-2}\beta_{11} + Z_{it-3}\beta_{12} + y_{it-1}\beta_{13}$.

The parameters can be estimated with maximum likelihood. Similar to OP with LASSO, the maximization of the likelihood proceeds subject to the constraint $\sum_{q=1}^{p} |\beta_q| \leq s$, where s is a user-specified tuning parameter. This algorithm combined with LASSO can produce shrinkage coefficients that improve the model's predictive ability. The resultant model, a CR model, in which a l_1-penalized constraint is added to the corresponding likelihood function, is the L1-penalized continuation ratio model. The estimation is presented in equation (9) below:

$$\hat{\beta} = argmax_{\beta} \left(\mathcal{L}(\beta | y_{it}, A) - \lambda \sum_{q=1}^{p} |\beta_q| \right), \tag{9}$$

where $\mathcal{L}(\beta | y_{it}, A)$ is the likelihood function of CR and A contains the pool of the potential predictors (X_{it}, X_{it-1}, X_{it-2}, X_{it-3}, W_{it}, W_{it-1}, W_{it-2}, W_{it-3}, Z_{it}, Z_{it-1}, Z_{it-2}, Z_{it-3}, y_{it-1}). All explanatory variables are standardized before applying the LASSO estimator. In line with the abovementioned LASSO counterparts, the AIC-type and the BIC-type tuning parameter selectors assist with selecting the best model.[12] All predictors are standardized before applying the LASSO estimator.

[12] For more details, see Appendix C.

2.4.5 CR with Elastic net

To achieve CR with Elastic net, both l_1-penalty and l_2-penalty are added to maximum likelihood of CR to obtain estimated coefficients, are explained by equations (10) and (11) below:

$$\hat{\beta} = argmax_{\beta}\left(\mathcal{L}(\beta|y_{it}, A) - \lambda_1 \sum_{q=1}^{p} |\beta_q| - \lambda_2 \sum_{q=1}^{p} ({\beta_q}^2)\right). \tag{10}$$

Similar to the OP with Elastic net, equation (10) is converted as follows:

$$\hat{\beta} = argmax_{\beta}\left(\mathcal{L}(\beta|y_{it}, A) - \lambda \sum_{q=1}^{p} \left\{\alpha|\beta_q| + \frac{1}{2}(1-\alpha)({\beta_q}^2)\right\}\right), \tag{11}$$

where $0 \leq \alpha \leq 1$, $\mathcal{L}(\beta|y_{it}, A)$ is the likelihood function of CR and A contains the pool of the potential predictors (X_{it}, X_{it-1}, X_{it-2}, X_{it-3}, W_{it}, W_{it-1}, W_{it-2}, W_{it-3}, Z_{it}, Z_{it-1}, Z_{it-2}, Z_{it-3}, y_{it-1}). All explanatory variables are standardized before applying the Elastic net estimator. Similar to the aforementioned LASSO counterparts, the AIC-type and the BIC-type tuning parameter selectors assist with selecting the best model. Once again, before implementing the Elastic net estimator, all predictors are standardized.

2.5 Empirical results

2.5.1 Accuracy

Accuracy Ratios (ARs) are applied to evaluate the predictive ability of all candidates for firms' CDSIRs and EQIRs in Table 2-6. ARs can be calculated by the sum of all diagonal terms divided by the total number of observations in each contingency table (see Appendix B), which indicates the percentage of correct prediction. It can be expressed as $AR = \frac{1}{T}\sum_{t=1}^{T} 1(\hat{q}_t = q_t)$ where $\hat{q}_t$ is the predicted rating and q_t represents the actual outcome. We report statistics for all candidate models for both in- and out-of-sample predictions. The out-of-sample forecast of ratings is calculated by an expanding window method based on the past and current information available up to time T. This method allows successive observations to be included in the initial sample prior to forecast of the next step ahead prediction of the rating while keeping the start date of the sample fixed. By this method, we forecast future ratings $\hat{q}_{t+1}$, $\hat{q}_{t+2}$ etc. The initial estimation window is from 2002 to 2005 and the first prediction date is 2006. We then increase T by one each time (one month) until T reaches 2008. In addition, we report at the foot of each panel the number of surviving variables.[13]

Table 2-6 Accuracy Ratios and selected variables

		OP	OP_LASSO		OP_ELASTIC NET		CR_LASSO		CR_ELASTIC NET	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CDSIRs	In-sample prediction	90.23%	90.30%*	89.39%***	90.30%	89.39%***	89.80%***	89.48%***	89.63%***	89.48%***
	Out-of-sample prediction	22.02%	31.56%***	84.53%***	31.56%***	84.53%***	58.80%***	83.73%***	60.92%***	84.26%***
	Surviving variables	268	143	45	144	45	75	48	78	51
EQIRs	In-sample prediction	95.05%	95.00%	94.69%*	95.05%	94.52%**	95.19%	94.77%	95.16%	94.73%
	Out-of-sample prediction	48.98%	84.77%***	90.95%***	80.88%***	90.78%***	85.45%***	91.03%***	80.80%***	90.95%***
	Surviving variables	268	167	95	181	87	152	83	187	91

Notes: This Table reports the Accuracy Ratios and the number of surviving variables for each model under study. “OP” stands for the ordered probit model. “OP_LASSO” refers to the ordered probit model with LASSO estimator. “OP_ELASTIC NET” stands for the ordered probit model with Elastic net estimator. “CR_LASSO” indicates the continuation ratio model with LASSO estimator. “CR_ELASTIC NET” is the continuation ratio model with Elastic net estimator. “AIC” is the AIC-type tuning parameter selector. “BIC” is the BIC-type tuning parameter selector. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

In Table 2-6, no significant differences are indicated between all competing models in in-sample predictions, since they provide a similar in-sample performance for both types of market implied ratings. With reference to CDSIRs, the accuracy ratio of in-sample predictions can reach approximately 90% and for EQIRs about 95% correct predictions can be found across competing models in Table 2-6. This tendency cannot be observed in out-of-sample predictions. We find that LASSO models are able to clearly produce more accurate out-of-sample prediction than the benchmark model (OP) in both types of market implied ratings. For CDSIRs, our results indicate that the accuracy ratio in out-of-sample prediction rises from 22% in the OP model to 84% in the LASSO models. This increasing

[13] The surviving variables are defined as predictors with non-zero estimated coefficients after the penalized procedure. In the benchmark model, we do not drop any variables since OP does not penalise regression coefficients.

tendency in the forecast of EQIRs also can be found. The ratio increases from 49% in the OP model to 91% in the LASSO models.

Next, the within performance of the LASSO models are compared by considering different tuning parameter selectors. The predictive performance in the out-of-sample for the LASSO models with BIC-type tuning parameter selector are superior to their counterparts with AIC selector in CDSIRs. For EQIRs, there exists little difference in the correct predictions among LASSO candidate models. It is interesting to note that the LASSO models with BIC-type tuning parameter selector consistently choose a smaller number of predictors than their AIC-type counterparts. This does not seem to affect their forecasting performance for EQIRs but leads to more accurate predictions for CDSIRs.[14]

2.5.2 Statistical significance

To evaluate the relative performance of the models presented in the previous sub-section, we employ the Stuart–Maxwell test. This approach will help us formally test for the statistical significance of difference between forecasts and further validate our main findings. The Stuart–Maxwell test (Stuart 1955, Maxwell 1970) is a generalized version of McNemar's test (McNemar 1947), which is associated with multiple (k) categories and tests whether the difference between two related samples from an ordinal field is statistically different from zero. The Stuart–Maxwell tests the null hypothesis of equal marginal proportion for each category between the forecasts of two models (model A vs model B). Under the null hypothesis, the statistic is distributed as chi-square with $k-1$ degrees of freedom. A statistically significant Stuart–Maxwell test statistic indicates that the forecasts of the first model (A) are different from those of the second model (B).

For the CDSIRs, our results indicate a statistically significant difference in out-of-sample predictions between the OP model and other competing models in Table 2-6. This statistically significant difference can also be confirmed for the EQIRs. However, this tendency cannot be clearly observed in the in-sample forecasts of all competing models for both CDSIRs and EQIRs. Combining these statistics with the accuracy ratios, it can be suggested that adding LASSO or the Elastic net

[14] Tables B1 to B36 in Appendix B illustrate the contingency tables of the predicted against the actual outcome for both in- and out-of-sample results for the various models presented in Table 2-6.

estimator to the OP model or the CR model can produce different out-of-sample predictions that are more accurate than those generated by the OP.

2.5.3 Robustness tests

The findings of the previous section are further validated by carrying out several robustness tests. In the first test, an alternative choice of tuning parameters selector, namely cross validation, is applied to LASSO models. Next, we employ an alternative benchmark for forecasting, namely the Principal Component Analysis. In a further test a tuning-free version of the LASSO estimator is considered. Following that, alternative measures of predictive ability are employed. Furthermore, we limit our data to investment grade ratings. Finally, random effects in panel data are considered.

2.5.3.1 Cross validation

It is well accepted that the finite-sample performance of estimators such as LASSO or Elastic net are controlled by the choice of tuning parameter selector. Motivated by this consideration, we present forecasts where cross validation, one of the most commonly used model selection criteria, is employed (see for instance Stone 1977, Yang 2007). Following the relevant literature on model selection (see for example, Kohavi 1995), the ten-fold stratified cross validation (10-fold CV) is applied as a comparable application of tuning parameter selector.

Our main findings are supported by the Accuracy Ratios and Stuart–Maxwell test statistics, in Table 2-7. In the in-sample forecast exercises, similar predictive power is observed in all candidates. In the out-of-sample prediction, however, for the CDSIRs, the percentages of correct predictions in all competing LASSO models are higher than in the benchmark model. This tendency can also be observed in the EQIRs. To sum up, there are significant gains in predictive ability once the LASSO is applied even when cross validation is employed. We conclude that our main results are robust for alternative tuning choices.

Table 2-7 Accuracy Ratios and selected variables (10-fold cross validation)

		OP	OP_LASSO	OP_ELASTIC NET	CR_LASSO	CR_ELASTIC NET
CDSIRs	In-sample prediction	90.23%	90.32%**	90.33%**	89.53%***	89.48%***
	Out-of-sample prediction	22.02%	35.10%***	35.28%***	84.53%***	84.26%***
	Surviving variables	268	130	134	48	51
EQIRs	In-sample prediction	95.05%	94.88%	94.90%	95.13%	95.08%
	Out-of-sample prediction	48.98%	90.86%***	90.86%***	89.68%***	90.36%***
	Surviving variables	268	135	136	137	145

Notes: This Table reports the Accuracy Ratios and the number of surviving variables for each model under study. “OP” stands for the ordered probit model. “OP_LASSO” refers to the ordered probit model with LASSO estimator. “OP_ELASTIC NET” stands for the ordered probit model with Elastic net estimator. “CR_LASSO” indicates the continuation ratio model with LASSO estimator. “CR_ELASTIC NET” is the continuation ratio model with Elastic net estimator. “AIC” is the AIC-type tuning parameter selector. “BIC” is the BIC-type tuning parameter selector. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

2.5.3.2 Principal Component Analysis with OP

One of the most popular statistical procedures for variable selection is the Principal Component Analysis (PCA). PCA converts a set of possibly correlated variables to a smaller set of uncorrelated variables called Principal Components (PC). The first PC accounts for as much of the variability in the dataset as possible and each succeeding component turn attains the highest possible variance under the constraint that it is orthogonal to the preceding components. PCA is probably the most popular dimension reduction procedure in Economics and Finance and has been applied successfully to a series of forecasting problems (see Stock and Watson 2002, Ludvigson and Ng 2009, Bailey *et al*. 2016). In our application, we are dealing with a large set of possibly correlated variables and thus a natural candidate to benchmark our procedure is the PCA. We select the PCs that account for 70%, 80% and 90% of the variability in our dataset and combine them with OP.

Comparing Table 2-8 and Table 2-6 (where the accuracy ratios of OP and OP with LASSO are presented), it can be noted that PCA has the ability to improve the out-of-sample predictive accuracy of the OP model. For our OP combined with LASSO models, there is the single case that the predictive performance provided by PCA is better than LASSO models, in terms of the out-of-sample for CDSIRs. In that case, OP based on the PCs explaining 70% and 90% of the variability, indicates higher accuracy ratios compared to the LASSO model with the AIC-type tuning parameter selector (but not under the BIC-type selector). It is also important to note that PCA with OP presents lower in-sample accuracy ratios in all cases. These results make it possible to argue that our LASSO formulations are robust in relation to an alternative benchmark.

Table 2-8 PCA combined with OP

		PCA with OP		
		(1)	(2)	(3)
CDSIRs	In-sample prediction	45.94%	51.36%	57.39%
	Out-of-sample prediction	45.71%	32.01%	46.33%
	Selected principal components	16	25	45
	PCs cumulative percentage of total variation	70.92%	80.00%	90.31%
EQIRs	In-sample prediction	63.61%	72.00%	76.93%
	Out-of-sample prediction	63.37%	54.74%	51.69%
	Selected principal components	15	23	41
	PCs cumulative percentage of total variation	71.23%	80.43%	90.20%

Notes: This Table reports the Accuracy Ratios and the number of PCs. “OP” stands for the ordered probit model. “PC” refers to the principal components and “PCA” stands for principal component analysis.

2.5.3.3 A tuning-free version of the LASSO

While the results presented so far are robust for different tuning choices, including cross validation, it is important to note that the latter is computationally costly and theoretically less well developed, especially for the purpose of variable selection and the estimation of regression coefficients (see Sun and Zhang 2012). Thus, to further alleviate potential concerns regarding the choice of the tuning parameter, we employ the scaled LASSO, developed by Sun and Zhang (2012), without depending on model selection criteria such as AIC, BIC or CV.[15][16]

Table 2-9 presents the relevant Accuracy Ratios and the results of the Stuart–Maxwell statistical tests. All candidate models provide similar accuracy ratios in in-sample predictions. In the out-of-sample evidence, the percentage of correct predictions in the scaled LASSO is higher than in the OP for the CDSIRs and EQIRs. Once again, this finding is consistent with our main results, indicating that our findings are robust when using the scaled LASSO.

[15] Another tuning-free version of the LASSO estimator is the square-root LASSO of Belloni et al. (2011).

[16] For the scaled LASSO, the authors generated the gradient descent algorithm in a convex minimization of a penalized joint loss function.

Table 2-9 Accuracy Ratios and selected variables (scaled LASSO)

		OP	SCALED_LASSO
CDSIRs	In-sample prediction	90.23%	87.98%***
	Out-of-sample prediction	22.02%	39.26%***
	Surviving variables	268	124
EQIRs	In-sample prediction	95.05%	93.93%***
	Out-of-sample prediction	48.98%	89.26%***
	Surviving variables	268	121

Notes: This Table reports the Accuracy Ratios and the number of surviving variables for each model under study. “OP” stands for the ordered probit model. “SCALED_LASSO” refers to the scaled LASSO for tuning-free parameter. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

2.5.3.4 Alternative measures of predictive ability

2.5.3.4.1

Thus far, the relative performance of the estimated models has been evaluated in terms of an informal goodness of fit indicator, by comparing predicted and observed ratings. It is possible, however, to give a more quantitative measure of the predictive ability of our models. We therefore check the robustness of our measure of predictive power by using a measure based on a technique proposed by Merton (1981) and used in Henriksson and Merton (1981), Pesaran and Timmermann (1994), Kim *et al.* (2008) and Mizen and Tsoukas (2012). Specifically, let CP_j be the proportion of the correct predictions made by $\hat{q}_t$ when the true state is given by $q_t = j$. From the definition of conditional probability, CP is computed as $CP_j = \frac{\frac{1}{T}\sum_{t=1}^{T} 1(\hat{q}_t=j)(q_t=j)}{\frac{1}{T}\sum_{t=1}^{T} 1(q_t=j)}$and Merton’s correct measure, expressed CP, is given by $CP = \frac{1}{J-1}\left[\sum_{j=1}^{J} CP_j - 1\right]$ where J is the number of categories, and $-\frac{1}{J-1} < CP < 1$. In the contingency table (see Appendix B) CP is the unweighted average of CP_js minus one (to correct for the phenomenon that certain categories are over-represented). The CP_j s are calculated as the proportion of correct predictions divided by the total of each row. This modifies the measure of predictive ability to discount the influence of the dominant outcome. A high CP score indicates that the predictor is accurate for all rating categories.

The Accuracy Ratios when we account for the influence of the dominant outcome by reporting the Merton’s correct predictions are shown in Table 2-10. The corresponding Stuart-Maxwell statistical tests are also documented. The test produces CP ratios that confirm our main findings. In the in-sample exercise, there is little difference between the OP and the LASSO models. In contrast, the predictive ability of the out-of-sample predictions is superior when penalty functions are applied.

Table 2-10 CP Ratios and selected variables

		OP	OP_LASSO		OP_ELASTIC NET		CR_LASSO		CR_ELASTIC NET	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CDSIRs	In-sample prediction	69.29%	66.89%*	61.04%***	66.95%	61.05%***	63.29%***	61.84%***	62.74%***	61.84%***
	Out-of-sample prediction	18.22%	25.49%***	61.01%***	25.49%***	61.01%***	45.70%***	64.16%***	47.94%***	63.42%***
	Surviving variables	268	143	45	144	45	75	48	78	51
EQIRs	In-sample prediction	78.30%	77.16%	74.46%*	77.19%	74.25%**	78.58%	74.51%	78.45%	74.39%
	Out-of-sample prediction	35.49%	66.50%***	70.10%***	63.42%***	69.93%***	67.29%***	70.23%***	62.39%***	70.10%***
	Surviving variables	268	167	95	181	87	152	83	187	91

Notes: This Table reports the CP ratios and the number of surviving variables for each model under study. "OP" stands for the ordered probit model. "OP_LASSO" refers to the ordered probit model with LASSO estimator. "OP_ELASTIC NET" stands for the ordered probit model with Elastic net estimator. "CR_LASSO" indicates the continuation ratio model with LASSO estimator. "CR_ELASTIC NET" is the continuation ratio model with Elastic net estimator. "AIC" is the AIC-type tuning parameter selector. "BIC" is the BIC-type tuning parameter selector. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

2.5.3.4.2

An alternative measure for our forecasts can be constructed based on the misclassification rate of Hastie *et al.* (2009): $T^{-1}\sum_{t=1}^{T} 1\{\widehat{q_t} \neq q_t\}$.[17] In the spirit of Diebold and Mariano (1995), we estimate the misclassification rate for each point and model and then we test if the mean difference of these rates ($1\{\widehat{q_t}^{Model\ A} \neq q_t\} - 1\{\widehat{q_t}^{Model\ B} \neq q_t\}$) between two models is zero. If this is statistically different from zero it indicates that the two models generate different forecasts. Table 2-11 and Table 2-12 present the relevant p-values for our out-of-sample forecasts.

[17] Using the notation from the previous sub-section, let stand for the target variable and denote the forecast.

Table 2-11 Equal performance tests of out-of-sample prediction of CDSIRs

		OP	OP_LASSO		OP_ELASTIC NET		CR_LASSO		CR_ELASTIC NET	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
OP		~	7.60***	38.38***	7.60***	38.38***	23.66***	38.57***	24.79***	38.75***
OP_LASSO	AIC		~	32.58***	~	32.58***	19.31***	32.75***	20.40***	32.90***
	BIC			~	-32.58***	~	-17.08***	-1.62	-16.01***	-0.69
OP_ELASTIC NET	AIC				~	32.58***	19.31***	32.75***	20.40***	32.90***
	BIC					~	-17.08***	-1.62	-16.01***	-0.69
CR_LASSO	AIC						~	17.36***	4.95***	17.43***
	BIC							~	-16.27***	1.73*
CR_ELASTIC NET	AIC								~	16.35***
	BIC									~

Note: This table reports the test of equality of the mean difference in losses for two models. *** denotes that the null hypothesis of equal performance of two models is rejected at the 1% significant level. ** denotes that the null hypothesis of equal performance of two models is rejected at the 5% significant level and * denotes that the null hypothesis of equal performance of two models is rejected at the 10% significant level. ~ indicates that the two models generate the same set of forecasts.

Table 2-12 Equal performance tests of out-of-sample prediction of EQIRs

		OP	OP_LASSO		OP_ELASTIC NET		CR_LASSO		CR_ELASTIC NET	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
OP		~	23.65***	26.97***	21.38***	26.80***	23.72***	27.02***	21.26***	26.97***
OP_LASSO	AIC		~	6.03***	-5.83***	5.82***	1.79*	6.09***	-5.82***	6.03***
	BIC			~	-8.79***	-1.41	-5.46***	1.00	-8.66***	~
OP_ELASTIC NET	AIC				~	8.64***	6.29***	8.85***	-0.26	8.79***
	BIC					~	-5.25***	1.73*	-8.51***	1.41
CR_LASSO	AIC						~	5.57***	-6.37***	5.46***
	BIC							~	-8.71***	-1.00
CR_ELASTIC NET	AIC								~	8.66***
	BIC									~

Note: This table reports the test of equality of the mean difference in losses for two models. *** denotes that the null hypothesis of equal performance of two models is rejected at the 1% significance level. ** denotes that the null hypothesis of equal performance of two models is rejected at the 5% significance level, * denotes that the null hypothesis of equal performance of two models is rejected at the 10% significance level. ~ indicates that the two models generate the same set of forecasts.

We note that in almost all cases, our forecasts are statistically different. These results complete the picture by further demonstrating the superiority of LASSO as a variable selection technique and the effectiveness of the BIC criterion in tuning the LASSO parameters. To sum up, the results are robust when carrying out an alternative test to evaluate the forecasting performance based on the proportion of correct predictions for each of the various rating categories.

2.5.3.5 Investment grade ratings

Much of the previous related literature studied employs data with investment grade ratings. However, Amato and Furfine (2004) report that selection bias is like to be generated if researchers emphasise analysing one category. On the other hand, it is well accepted that firms with varying levels of financial health have different ability to withstand the effects of changes in financial and business risk on their creditworthiness. This implies that pooling together both categories may lead to misspecification of our model. Therefore, all speculative grade ratings are removed, and our models are re-estimated.

The results in Table 2-13 validate our main findings. First, similar accuracy ratios can be observed in all models considering the in-sample forecast evaluations. Moving to the out-of-sample, for the CDSIRs, both versions of the CR with LASSO provide more accurate forecasts. On the other hand, for EQIRs all LASSO models display better predictive performance than their OP benchmark. In conclusion, even when limiting our sample to investment grade ratings only, the out-of-sample predictions demonstrate that the LASSO models outperform the benchmark model.

Table 2-13 Accuracy Ratios and selected variables (investment-grade ratings)

		OP	OP_LASSOO		OP_ELASTICT NET		CR_LASSO		CR_EALSTIC NET	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CDSIRs	In-sample prediction	90.76%	90.45%**	89.38%***	90.14%***	89.42%***	90.56%**	89.44%***	90.74%	89.40%***
	Out-of-sample prediction	49.18%	77.04%***	84.33%***	84.22%***	84.87%***	82.81%***	85.53%***	80.41%***	84.66%***
	Surviving variables	266	143	19	94	43	98	39	122	28
EQIRs	In-sample prediction	95.80%	95.61%*	95.02%**	95.49%	94.95%**	95.59%	95.05%	95.72%	94.97%*
	Out-of-sample prediction	57.82%	87.72%***	92.28%***	90.99%***	92.28%***	90.69%***	92.28%***	86.44%***	92.28%***
	Surviving variables	266	137	74	143	77	135	61	153	68

Notes: This Table reports the Accuracy Ratios and the number of surviving variables for each model under study. “OP” stands for the ordered probit model. “OP_LASSO” refers to the ordered probit model with LASSO estimator. “OP_ELASTIC NET” stands for the ordered probit model with Elastic net estimator. “CR_LASSO” indicates the continuation ratio model with LASSO estimator. “CR_ELASTIC NET” is the continuation ratio model with Elastic net estimator. “AIC” is the AIC-type tuning parameter selector. “BIC” is the BIC-type tuning parameter selector. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

2.5.3.6 Accounting for the panel data dimension

As a final robustness test, a random-effects version of the ordered probit model is considered to capture the panel data dimension of the dataset. The Accuracy Ratios and the corresponding Stuart–Maxwell statistical tests are reported in Table 2-14. The conclusion is consistent with other robustness tests which show that the main findings remain unchanged: when we apply the LASSO or Elastic net estimator the models have superior predictive ability compared to the ordered probit model, even when random effects are included. We conclude that our findings are robust in estimating the models with random effects to deal with the panel data nature of the sample.

Table 2-14 Accuracy Ratios and selected variables (random effects)

		OP_re	OP_re_LASSO		OP_re_ELASTIC NET	
			AIC	BIC	AIC	BIC
CDSIRs	In-sample prediction	83.62%	84.10%***	83.00%***	84.03%**	83.00%***
	Out-of-sample prediction	21.84%	25.29%***	65.08%***	25.46%***	65.08%***
	Surviving variables	268	143	45	144	45
EQIRs	In-sample prediction	93.17%	94.27%***	93.71%***	93.31%***	93.98%***
	Out-of-sample prediction	48.98%	82.99%***	89.59%***	12.44%***	89.26%***
	Surviving variables	268	167	95	181	87

Notes: This Table reports the Accuracy Ratios and the number of surviving variables for each model under study. “OP_re” stands for the ordered probit model with random effects. “OP_re_LASSO” is the ordered probit model with LASSO estimator and random effects. “OP_re_ELASTIC NET” stands for the ordered probit model with Elastic net estimator and random effects. “AIC” is the AIC-type tuning parameter selector. “BIC” is the BIC-type tuning parameter selector. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

2.5.4 Discussion

In the previous sections, a forecasting exercise on CDSIRs and EQIRs prediction was presented. For both types of market implied ratings all models provide similar in-sample predictive performance. In the out-of-sample evidence, for CDSIRs, we note that a better predictive performance can be produced by the LASSO models with BIC-type tuning parameter selector than their benchmarks. Meanwhile, these LASSO models controlled by BIC-type selector tend to choose a smaller set of surviving variables than their counterparts with the AIC-type selector. This lends support to the argument that the models with BIC-type tuning parameter selector make better use of the available information. Moving to the EQIRs, a similar pattern can be observed. The models with the BIC-type tuning parameter selector outperform their counterparts with the AIC-type selector with fewer predictors in terms of accuracy.

To conclude, we note that the LASSO models are able to provide more accurate out-of-sample forecasts on the CDSIRs and EQIRs ratings and they outperform the OP model. This is of particular interest given that the OP model dominates the related literature in predicting credit ratings. From the LASSO models under study, the optimized models with BIC-type tuning parameter selector seem able to provide better forecasts while at the same time using fewer predictors. These results are robust when modifying the tuning parameters, considering a tuning-free version of LASSO, evaluating the predictive performance of the models using a different statistical measure and restricting the dataset to investment grade ratings.

2.6 Conclusion

Providing a reasonable accuracy ratio in credit ratings prediction is critically important for both market participants and rating agencies since these predictions will be used as a reference for evaluating credit risk. Following the most recent financial crisis, the performance of conventional credit ratings has been heavily criticized because of their out of date nature. In this work, market implied ratings have been chosen as the target since they can immediately adjust to market change compared with traditional credit ratings. To achieve more accurate forecasts of market implied ratings, a variable selection technique, the least absolute shrinkage and selection operator (LASSO), and its most promising derivation, the Elastic net, are applied to ordered probit and continuation ratio models modelling market implied ratings. All LASSO models select the most relevant predictors from a set of 268 variables and forecast the MIRs for a period of six years (2002 to 2008). This marks a break with the existing literature which typically depends on discrete limited dependent variable models.

Our results using monthly data from the US are interesting in several respects. First, market implied ratings can be explained by accounting variables along with market-driven and macroeconomic indicators. Second, the predictive performance produced by the LASSO models are clearly better than that produced by ordered probit models, mostly adopted in previous studies. Finally, the LASSO models controlling by BIC-type tuning parameter selector can provide more accurate out-of-sample predictions than their counterparts with AIC-type selector for the dataset and periods under study. Hence, LASSO-selected models attain an improved forecasting power.

These results suggest risk managers and academics should further explore the properties of variable selection models during credit risk assessment. Researchers apply an explanatory variables set in limited dependent variable models based on their a priori knowledge, which can cause misspecifications. Meanwhile, the unknown factors affecting credit ratings will change over time or under different regulations. On the other hand, variable selection approaches such as LASSO are more flexible and can disclose the underlying structure of the problem, resulting in a sparse representative and improved predictive ability.

References

Akaike, H. (1974) ' A New Look at the Statistical Model Identification', *IEEE Transactions on Automatic Control*, 19(6), 716-723.

Altman, E. I. and Kao, D. L. (1992) 'The Implications of Corporate Bond Ratings Drift', *Financial Analysts Journal*, 48(3), 64-75.

Altman, E. I. and Katz, S. (1976) *Statistical Bond Rating Classification Using Financial and Accounting Data* [press release], available: [Accessed. 21th July 2015]

Amato, J. D. and Furfine, C. H. (2004) ' Are Credit Ratings Procyclical?', *Journal of Banking and Finance*, 28(11), 2641-2677.

Amendola, A., Restaino, M. and Sensini, L. (2012) 'Variable Selection in Forecasting Models for Default Risk' in Perna, C. and Sibillo, M., eds., *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, Milano: Springer Milan, 11-18.

Ang, J. S. and Patel, K. A. (1975) 'Bond Rating Methods: Comparison and Validation', *The Journal of Finance*, 30(2), 631-640.

Archer, K. J. and Williams, A. A. A. (2012) 'L1 Penalized Continuation Ratio Models for Ordinal Response Prediction Using High-Dimensional Datasets', *Statistics in Medicine*, 31(14), 1464-1474.

Bailey, N., Holly, S. and Pesaran, M. H. (2016) 'A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence: A Two-Stage Approach to Spatio-Temporal Analysis', *Journal of Applied Econometrics*, 31(1), 249-280.

Belloni, A., Chernozhukov, V. and Wang, L. (2011) 'Square-root Lasso: Pivotal Recovery of Sparse Signals via Conic Programming', *Biometrika*, 98(4), 791-806.

Blume, M. E., Lim, F. and Mackinlay, A. C. (1998) 'The Declining Credit Quality of U.S. Corporate Debt: Myth or Reality?', *The Journal of Finance*, 53(4), 1389-1413.

Breger, L. L., Goldberg, L. R. and Cheyette, O. (2003) 'Market Implied Ratings', *SSRN Electronic Journal*.

Castellano, R. and Giacometti, R. (2012) 'Credit Default Swaps: Implied Ratings versus Official Ones', *4OR*, 10(2), 163-180.

Chava, S. and Jarrow, R. A. (2004) 'Bankruptcy Prediction with Industry Effects', *European Finance Review*, 8(4), 537-569.

Contoyannis, P., Jones, A. M. and Rice, N. (2004) 'The Dynamics of Health in the British Household Panel Survey', *Journal of Applied Econometrics*, 19(4), 473-503.

Creal, D. D., Gramacy, R. B. and Tsay, R. S. (2014) 'Market-Based Credit Ratings', *Journal of Business & Economic Statistics*, 32(3), 430-444.

Diebold, F. X. and Mariano, R. S. (1995) 'Comparing Predictive Accuracy', *Journal of Business & Economic Statistics*, 13(3), 253-263.

Doumpos, M., Niklis, D., Zopounidis, C. and Andriosopoulos, K. (2015) 'Combining Accounting Data and a Structural Model for Predicting Credit Ratings: Empirical Evidence from European Listed Firms', *Journal of Banking and Finance*, 50, 599-607.

Ederington, L. H. (1985) 'Classification Models and Bond Ratings', *The Financial Review*, 20(4), 237.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) 'Least Angle Regression', *The Annals of Statistics*, 32(2), 407-451.

Fan, J. and Li, R. (2001) 'Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties', *Journal of the American Statistical Association*, 96(456), 1348-1360.

Fienberg, S. E. (1980) *The Analysis of Cross-Classified Categorical Data*, Cambridge, Mass.: MIT Press.

Financial Crisis Inquiry, C. (2011) *The Financial Crisis Inquiry Report : Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States*, New York, NY: Public Affairs.

Fisher, L. (1959) 'Determinants of Risk Premiums on Corporate Bonds', *Journal of Political Economy*, 67(3), 217-237.

Gentry, J. A., Whitford, D. T. and Newbold, P. (1988) 'Predicting Industrial Bond Ratings with a Probit Model and Funds Flow Components', *Financial Review*, 23(3), 269-286.

Güttler, A. and Wahrenburg, M. (2007) 'The Adjustment of Credit Ratings in Advance of Defaults', *Journal of Banking and Finance*, 31(3), 751-767.

Hardin, J. W. and Hilbe, J. M. (2018) *Generalized Linear Models and Extensions*, College Station (Tex.): Stata Press.

Härdle, W. K. and Prastyo, D. D. (2013) 'Default Risk Calculation Based on Predictor Selection for the Southeast Asian Industry'. Berlin, SFB 649, Economic Risk. Available from: http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2013-037.pdf. [Accessed: 21th July 2015].

Hastie, T., Tibshirani, R., Friedman, J. H. and SpringerLink (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., New York: Springer.

Henriksson, R. D. and Merton, R. C. (1981) 'On Market Timing and Investment Performance: II. Statistical Procedures for Evaluating Forecasting Skills', *The Journal of Business*, 54(4), 513.

Horrigan, J. O. (1966) 'The Determination of Long-Term Credit Standing with Financial Ratios', *Journal of Accounting Research*, 4(2), 44-62.

Hwang, R.-C. (2013a) 'Forecasting Credit Ratings with the Varying-Coefficient Model', *Quantitative Finance*, 13(12), 1947-1965.

Hwang, R.-C. (2013b) 'Predicting Issuer Credit Ratings Using Generalized Estimating Equations', *Quantitative Finance*, 13(3), 383-398.

Hwang, R.-C., Cheng, K. F. and Lee, C.-F. (2009) 'On Multiple-class Prediction of Issuer Credit Ratings', *Applied Stochastic Models in Business and Industry*, 25(5), 535-550.

Hwang, R.-C., Chung, H. and Chu, C. K. (2010) 'Predicting Issuer Credit Ratings Using a Semiparametric Method', *Journal of Empirical Finance*, 17(1), 120-137.

Kamstra, M., Kennedy, P. and Suan, T. K. (2001) 'Combining Bond Rating Forecasts Using Logit', *Financial Review*, 36(2), 75-96.

Kao, C. and Wu, C. (1990) 'Two-Step Estimation of Linear Models with Ordinal Unobserved Variables: The Case of Corporate Bonds', *Journal of Business & Economic Statistics*, 8(3), 317-325.

Kaplan, R. S. and Urwitz, G. (1979) 'Statistical Models of Bond Ratings: A Methodological Inquiry', *The Journal of Business*, 52(2), 231.

Kim, T. H., Mizen, P. and Chevapatrakul, T. (2008) 'Forecasting Changes in UK Interest Rates', *Journal of Forecasting*, 27(1), 53-74.

Kohavi, R. (1995) 'A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection', *IJCAI ... : proceedings of the ... International Joint Conference on Artificial Intelligence*, 1137.

Liu, B., Kocagil, A. E. and Gupton, G. M. (2007) *Fitch Equity Implied Rating and Probability of Default Model* [online], available: [Accessed. 25th July 2015]

Long, J. S. and Freese, J. (2006) *Regression Models for Categorical Dependent Variables Using Stata*, 2nd ed., College Station, Tex: Stata Press.

Ludvigson, S. C. and Ng, S. (2009) 'Macro Factors in Bond Risk Premia', *The Review of Financial Studies*, 22(12), 5027-5067.

Maddala, G. S. (2008) *Limited-dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press.

Maxwell, A. E. (1970) 'Comparing the Classification of Subjects by Two Independent Judges', *The British Journal of Psychiatry*, 116(535), 651-655.

McKelvey, R. D. and Zavoina, W. (1975) 'Statistical-Model for Analysis of Ordinal Level Dependent Variables', *Journal of Mathematical Sociology*, 4(1), 103-120.

McNemar, Q. (1947) 'Note on the Sampling Error of the Difference between Correlated Proportions or Percentages', *Psychometrika*, 12(2), 153-157.

Merton, R. C. (1981) 'On Market Timing and Investment Performance. I. an Equilibrium Theory of Value for Market Forecasts: I. Introduction', *The Journal of Business (pre-1986)*, 54(3), 363.

Mizen, P. and Tsoukas, S. (2012) 'Forecasting US Bond Default Ratings Allowing for Previous and Initial State Dependence in an Ordered Probit Model', *International Journal of Forecasting*, 28(1), 273-287.

Parnes, D. (2007) 'Applying Credit Score Models to Multiple States of Nature', *The Journal of Fixed Income*, 17(3), 57-71.

Pesaran, M. H. and Timmermann, A. G. (1994) 'A Generalization of the Non-Parametric Henriksson-Merton Test of Market Timing', *Economics Letters*, 44(1), 1-7.

Pinches, G. E. and Mingo, K. A. (1973) 'A Multivariate Analysis of Industrial Bond Ratings', *The Journal of Finance*, 28(1), 1-18.

Pogue, T. F. and Soldofsky, R. M. (1969) 'What's in a Bond Rating', *Journal of Financial and Quantitative Analysis*, 4(2), 201-228.

Poon, W. P. H. (2003) 'Are Unsolicited Credit Ratings Biased Downward?', *Journal of Banking and Finance*, 27(4), 593-614.

Reyngold, A., Kocagil, A. E. and Gupton, G. M. (2007) *Fitch CDS Implied Ratings (CDS-IR) Model* [online], available: [Accessed. 25th September 2015]

Rösch, D. (2005) 'An Empirical Comparison of Default Risk Forecasts from Alternative Credit Rating Philosophies', *International Journal of Forecasting*, 21(1), 37-51.

Schwarz, G. (1978) 'Estimating the Dimension of a Model', *The Annals of Statistics*, 6(2), 461-464.

Shao, J. (1997) 'An Asymptotic Theory for Linear Model Selection', *Statisticasinica Statistica Sinica*, 7(2), 221-242.

Stock, J. H. and Watson, M. W. (2002) 'Forecasting Using Principal Components from a Large Number of Predictors', *Journal of the American Statistical Association*, 97(460), 1167.

Stone, M. (1977) 'An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion', *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 44-47.

Stuart, A. (1955) 'A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification', *Biometrika*, 42(3/4), 412-416.

Sun, T. and Zhang, C.-H. (2012) 'Scaled Sparse Linear Regression', *Biometrika*, 99(4), 879-898.

Tian, S., Yu, Y. and Guo, H. (2015) 'Variable Selection and Corporate Bankruptcy Forecasts', *Journal of Banking and Finance*, 52, 89-100.

Tibshirani, R. (1996) 'Regression Shrinkage and Selection via the Lasso', *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288.

Tsoukas, S. and Spaliara, M.-E. (2014) 'Market Implied Ratings and Financing Constraints: Evidence from US Firms: Market Implied Ratings and Financing Constraints', *Journal of Business Finance & Accounting*, 41(1-2), 242-269.

van de Geer, S. A. (2008) 'High-Dimensional Generalized Linear Models and the Lasso', *The Annals of Statistics.*, 36(6), 614.

West, R. R. (1970) 'An Alternative Approach to Predicting Corporate Bond Ratings', *Journal of Accounting Research*, 8(1), 118-125.

Yang, Y. (2005) 'Can the Strengths of AIC and BIC Be Shared? A Conflict between Model Indentification and Regression Estimation', *Biometrika*, 92(4), 937-950.

Yang, Y. (2007) 'Prediction/Estimation with Simple Linear Models: Is It Really That Simple?', *Econometric Theory*, 23(1), 1-36.

Zhang, Y., Li, R. and Tsai, C.-L. (2010) 'Regularization Parameter Selections via Generalized Information Criterion', *Journal of the American Statistical Association*, 105(489), 312.

Zou, H. (2006) 'The Adaptive Lasso and Its Oracle Properties', *Journal of the American Statistical Association*, 101(476), 1418-1429.

Zou, H. and Hastie, T. (2005) 'Regularization and Variable Selection via the Elastic Net', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301-320.

Zou, H., Hastie, T. and Tibshirani, R. (2007) 'On the "Degrees of Freedom" of the Lasso', *The Annals of Statistics*, 35(5), 2173-2192.

Appendix

Appendix A

Table A1: Expected signs and variables definition

Covariates	Predicted relationship	Definition
Firm-specific variables (16)		
Size (1)		
DETA	+	Logarithm of real total assets
Leverage (5)		
AE	-	Total assets/Equity
LDA	-	Long-term debt/Total assets
SDA	-	Short-term debt/Total assets
TDA	-	Total debt/Total assets
TDEBITDA	-	Total debt/Earnings before interest, taxes, depreciation, amortization, and restructuring or rent costs
Coverage (2)		
EBITINT	+	Earnings before interest and tax/Interest expenses
EBITDAINT	+	Eearnings before interest, taxes, depreciation, amortization, and restructuring or rent costs/Interest expenses
Cash flow (2)		
CFOA	+	Cash flow from operating activities/Total assets
CASHEQA	+	Cash and equivalent/Total assets
Profitability (5)		
OM	+	Operating income/Net sales
ROC	+	Net income less dividends/Total capital
ROE	+	Net income/Shareholders' equity
ROA	+	Net income/Total assets
FFD	+	Funds from operations/Total debt
Liquidity (1)		
LIQ	+	Cash from operations/Liabilities
Market-driven Variables (6)		
EXRET	+	Monthly stock return-the S&P 500 index return
RSIZE	+	Firm equity value/Total market equity value
STA	-	The standard deviation of a company's monthly stock returns
BETA	-	Systematic risk in the Capital Asset Pricing Model
PD1	-	1-year default probability
PD5	-	5-year default probability
Macroeconomic Variables (11)		
RLSP	~	Return on S&P 500 index
CPFFM	~	3-month commercial paper rate
TB3	~	3-month Treasury bill rate minus federal funds rate
GS1	~	1-year constant maturity treasury rate
MB	~	Growth rate in the narrow money stock
INFL	~	Inflation rate
DLIP	~	Rate of change in industrial production
DLGDP	~	Real GDP growth
CFNA	~	Average Chicago Fed National Activity Index
UNRATE	~	Average unemployment rate
VIX	~	The Chicago Board Options Exchange volatility index

Notes: "+" indicates that the Market Implied Ratings would improve if the covariates rose. "-" indicates that the Market Implied Ratings would worsen if the covariates rose. "~" indicates uncertainty in the sign.

Appendix B

We cross tabulate predicted against observed CDSIRs outcomes in contingency Table B1 to B12 for the in-sample prediction.

Table B1: In-sample Prediction in Ordered Probit Model in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	64	169	1	0	0	0	0	234
AA	34	879	71	11	0	0	0	995
A	0	45	1591	69	0	0	0	1705
BBB	0	1	67	1804	23	0	0	1895
BB	0	0	0	50	783	8	0	841
B	0	0	0	0	11	133	4	148
Below CCC	0	0	0	0	0	5	2	7
Total	98	1094	1730	1934	817	146	6	5825

AR= 90.23%
CP= 69.29%

Table B2: In-sample Prediction in Ordered Probit Model with LASSO by AIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	59	174	1	0	0	0	0	234
AA	26	887	70	12	0	0	0	995
A	0	45	1592	68	0	0	0	1705
BBB	0	1	67	1805	22	0	0	1895
BB	0	0	0	53	781	7	0	841
B	0	0	0	0	11	135	2	148
Below CCC	0	0	0	0	0	6	1	7
Total	85	1107	1730	1938	814	148	3	5825

AR= 90.30%
CP= 66.89%

Table B3: In-sample Prediction in Ordered Probit Model with LASSO by BIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	6	227	1	0	0	0	0	234
AA	6	905	56	28	0	0	0	995
A	0	41	1567	97	0	0	0	1705
BBB	0	1	60	1813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	12	1174	1684	1993	812	151	0	5825

AR= 89.39%
CP= 61.04%

Table B4: In-sample Prediction in Ordered Probit Model with Elastic Net by AIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	60	173	1	0	0	0	0	234
AA	27	886	70	12	0	0	0	995
A	0	45	1592	68	0	0	0	1705
BBB	0	1	67	1805	22	0	0	1895
BB	0	0	0	53	781	7	0	841
B	0	0	0	0	11	135	2	148
Below CCC	0	0	0	0	0	6	1	7
Total	87	1105	1730	1938	814	148	3	5825

AR= 90.30%
CP= 66.95%

Table B5: In-sample Prediction in Ordered Probit Model with Elastic Net by BIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	6	227	1	0	0	0	0	234
AA	6	905	56	28	0	0	0	995
A	0	41	1567	97	0	0	0	1705
BBB	0	0	60	1813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	12	1173	1684	1993	812	151	0	5825

AR= 89.39%
CP= 61.05%

Table B6: In-sample Prediction in Continuation Ratio Model with LASSO by AIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	40	193	1	0	0	0	0	234
AA	18	893	61	23	0	0	0	995
A	0	42	1569	94	0	0	0	1705
BBB	0	0	60	1813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	58	1128	1691	1985	812	151	0	5825

AR= 89.80%
CP= 63.29%

Table B7: In-sample Prediction in Continuation Ratio Model with LASSO by BIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	19	214	1	0	0	0	0	234
AA	14	897	55	29	0	0	0	995
A	0	41	1567	97	0	0	0	1705
BBB	0	0	60	1813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	33	1152	1683	1994	812	151	0	5825

AR= 89.48%
CP= 61.84%

Table B8: In-sample Prediction in Continuation Ratio Model with Elastic Net by AIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	33	200	1	0	0	0	0	234
AA	21	890	61	23	0	0	0	995
A	0	41	1569	95	0	0	0	1705
BBB	0	0	60	1813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	54	1131	1691	1986	812	151	0	5825

AR= 89.63%
CP= 62.74%

Table B9: In-sample Prediction in Continuation Ratio Model with Elastic Net by BIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	19	214	1	0	0	0	0	234
AA	14	897	56	28	0	0	0	995
A	0	41	1567	97	0	0	0	1705
BBB	0	0	60	1813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	33	1152	1684	1993	812	151	0	5825

AR= 89.48%
CP= 61.84%

Table B10: In-sample Prediction in Principal Component Analysis_16 PCs with Ordered Probit model in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	37	156	40	1	0	0	0	234
AA	35	271	446	243	0	0	0	995
A	3	200	774	721	7	0	0	1705
BBB	5	33	536	1,255	64	2	0	1895
BB	0	1	55	433	304	48	0	841
B	0	0	5	39	66	35	3	148
Below CCC	0	0	0	1	3	3	0	7
Total	80	661	1856	2693	444	88	3	5825

AR= 45.94%
CP= 19.08%

Table B11: In-sample Prediction in Principal Component Analysis_25 PCs with Ordered Probit model in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	31	186	17	0	0	0	0	234
AA	42	440	420	93	0	0	0	995
A	1	184	873	626	21	0	0	1705
BBB	0	9	619	1188	74	5	0	1895
BB	0	0	9	356	417	57	2	841
B	0	0	0	19	81	43	5	148
Below CCC	0	0	0	0	4	3	0	7
Total	74	819	1938	2282	597	108	7	5,825

AR= 51.36%
CP= 25.00%

Table B12: In-sample Prediction in Principal Component Analysis_45 PCs with Ordered Probit model in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	47	184	3	0	0	0	0	234
AA	70	483	397	45	0	0	0	995
A	2	109	1027	550	17	0	0	1705
BBB	0	17	582	1240	54	2	0	1895
BB	0	0	10	245	514	68	4	841
B	0	0	0	37	77	32	2	148
Below CCC	0	0	0	0	4	3	0	7
Total	119	793	2019	2117	666	105	6	5825

AR= 57.39%
CP= 29.51%

We cross tabulate predicted against observed CDSIRs outcomes in contingency Table B13 to B24 for the out-of-sample prediction.

Table B13: Out-of-sample Prediction in Ordered Probit Model in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	9	14	4	1	23	22	11	84
AA	24	38	51	8	43	48	22	234
A	1	35	50	65	8	42	45	246
BBB	0	0	36	82	58	40	139	355
BB	0	0	0	10	37	14	76	137
B	0	0	0	0	5	13	34	52
Below CCC	0	0	0	0	0	3	20	23
Total	34	87	141	166	174	182	347	1131

AR= 22.02%
CP= 18.22%

Table B14: Out-of-sample Prediction in Ordered Probit Model with LASSO by AIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	1	20	15	22	20	6	0	84
AA	0	62	64	49	39	20	0	234
A	0	1	94	71	45	33	2	246
BBB	0	0	1	107	109	75	63	355
BB	0	0	0	0	57	20	60	137
B	0	0	0	0	1	17	34	52
Below CCC	0	0	0	0	0	4	19	23
Total	1	83	174	249	271	175	178	1131

AR= 31.56%
CP= 25.49%

Table B15: Out-of-sample Prediction in Ordered Probit Model with LASSO by BIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	84	0	0	0	0	0	84
AA	0	212	22	0	0	0	0	234
A	0	12	224	10	0	0	0	246
BBB	0	0	12	340	3	0	0	355
BB	0	0	0	5	132	0	0	137
B	0	0	0	0	4	48	0	52
Below CCC	0	0	0	0	0	23	0	23
Total	0	308	258	355	139	71	0	1131

AR= 84.53%
CP= 61.01%

Table B16: Out-of-sample Prediction in Ordered Probit Model with Elastic Net by AIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	1	20	15	22	20	6	0	84
AA	0	62	64	49	39	20	0	234
A	0	1	94	71	45	33	2	246
BBB	0	0	1	107	109	75	63	355
BB	0	0	0	0	57	20	60	137
B	0	0	0	0	1	17	34	52
Below CCC	0	0	0	0	0	4	19	23
Total	1	83	174	249	271	175	178	1131

AR= 31.56%
CP= 25.49%

Table B17: Out-of-sample Prediction in Ordered Probit Model with Elastic Net by BIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	84	0	0	0	0	0	84
AA	0	212	22	0	0	0	0	234
A	0	12	224	10	0	0	0	246
BBB	0	0	12	340	3	0	0	355
BB	0	0	0	5	132	0	0	137
B	0	0	0	0	4	48	0	52
Below CCC	0	0	0	0	0	23	0	23
Total	0	308	258	355	139	71	0	1131

AR= 84.53%
CP= 61.01%

Table B18: Out-of-sample Prediction in Continuation Ratio Model with LASSO by AIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	3	42	39	0	0	0	0	84
AA	0	137	89	8	0	0	0	234
A	0	2	170	73	1	0	0	246
BBB	0	0	6	218	131	0	0	355
BB	0	0	0	2	93	42	0	137
B	0	0	0	0	1	32	19	52
Below CCC	0	0	0	0	0	11	12	23
Total	3	181	304	301	226	85	31	1131

AR=58.80%
CP=45.70%

Table B19: Out-of-sample Prediction in Continuation Ratio Model with LASSO by BIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	2	81	1	0	0	0	0	84
AA	0	206	28	0	0	0	0	234
A	0	12	215	19	0	0	0	246
BBB	0	0	10	339	6	0	0	355
BB	0	0	0	4	133	0	0	137
B	0	0	0	0	4	46	2	52
Below CCC	0	0	0	0	0	17	6	23
Total	2	299	254	362	143	63	8	1131

AR= 83.73%
CP= 64.16%

Table B20: Out-of-sample Prediction in Continuation Ratio Model with Elastic Net by AIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	3	46	35	0	0	0	0	84
AA	0	142	87	5	0	0	0	234
A	0	2	173	70	1	0	0	246
BBB	0	0	6	227	122	0	0	355
BB	0	0	0	2	98	37	0	137
B	0	0	0	0	1	34	17	52
Below CCC	0	0	0	0	0	11	12	23
Total	3	190	301	304	222	82	29	1131

AR= 60.92%
CP= 47.94%

Table B21: Out-of-sample Prediction in Continuation Ratio Model with Elastic Net by BIC-typing Tuning Parameter Selector in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	2	81	1	0	0	0	0	84
AA	0	209	25	0	0	0	0	234
A	0	12	218	16	0	0	0	246
BBB	0	0	10	341	4	0	0	355
BB	0	0	0	5	132	0	0	137
B	0	0	0	0	4	47	1	52
Below CCC	0	0	0	0	0	19	4	23
Total	2	302	254	362	140	66	5	1131

AR= 84.26%
CP= 63.42%

Table B22: Out-of-sample Prediction in Principal Component Analysis_16 PCs with Ordered Probit model in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	15	47	21	1	0	0	0	84
AA	2	85	122	25	0	0	0	234
A	4	24	145	72	1	0	0	246
BBB	0	10	124	211	10	0	0	355
BB	0	0	12	70	53	2	0	137
B	0	0	2	11	31	8	0	52
Below CCC	0	0	0	3	13	7	0	23
Total	21	166	426	393	108	17	0	1131

AR= 45.71%
CP= 21.11%

Table B23: Out-of-sample Prediction in Principal Component Analysis_25 PCs with Ordered Probit model in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	67	17	0	0	0	0	0	84
AA	89	130	15	0	0	0	0	234
A	45	98	81	22	0	0	0	246
BBB	46	114	126	64	5	0	0	355
BB	4	29	21	64	18	1	0	137
B	0	5	5	17	22	2	1	52
Below CCC	0	0	0	5	10	8	0	23
Total	251	393	248	172	55	11	1	1131

AR= 32.01%
CP= 17.21%

Table B24: Out-of-sample Prediction in Principal Component Analysis_45 PCs with Ordered Probit model in CDSIRs

	Predicted CDSIRs							
Actual CDSIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	6	42	29	7	0	0	0	84
AA	0	77	108	49	0	0	0	234
A	0	12	99	131	4	0	0	246
BBB	0	2	44	237	70	2	0	355
BB	0	0	0	29	87	18	3	137
B	0	0	0	6	31	11	4	52
Below CCC	0	0	0	0	8	8	7	23
Total	6	133	280	459	200	39	14	1131

AR= 46.33%
CP= 27.02%

We cross tabulate predicted against observed EQIRs outcomes in contingency Table B25 to B36 for the in-sample prediction.

Table B25: In-sample Prediction in Ordered Probit Model in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	3	12	0	0	0	0	0	15
AA	0	409	32	0	0	0	0	441
A	0	19	2492	115	0	0	0	2626
BBB	0	0	70	2915	56	0	0	3041
BB	0	0	0	46	1273	14	0	1333
B	0	0	0	0	14	173	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	3	440	2594	3076	1343	187	0	7643

AR= 95.05%
CP= 78.30%

Table B26: In-sample Prediction in Ordered Probit Model with LASSO by AIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	2	13	0	0	0	0	0	15
AA	0	410	31	0	0	0	0	441
A	0	16	2495	115	0	0	0	2626
BBB	0	0	72	2910	59	0	0	3041
BB	0	0	0	50	1269	14	0	1333
B	0	0	0	0	12	175	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	2	439	2598	3075	1340	189	0	7643

AR=95.00%
CP=77.16%

Table B27: In-sample Prediction in Ordered Probit Model with LASSO by BIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	15	0	0	0	0	0	15
AA	0	412	29	0	0	0	0	441
A	0	18	2492	116	0	0	0	2626
BBB	0	0	84	2895	62	0	0	3041
BB	0	0	0	55	1262	16	0	1333
B	0	0	0	0	11	176	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	0	445	2605	3066	1335	192	0	7643

AR=94.69%
CP=74.46%

Table B28: In-sample Prediction in Ordered Probit Model with Elastic Net by AIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	2	13	0	0	0	0	0	15
AA	0	410	31	0	0	0	0	441
A	0	16	2498	112	0	0	0	2626
BBB	0	0	71	2911	59	0	0	3041
BB	0	0	0	49	1269	15	0	1333
B	0	0	0	0	12	175	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	2	439	2600	3072	1340	190	0	7643

AR=95.05%
CP=77.19%

Table B29: In-sample Prediction in Ordered Probit Model with Elastic Net by BIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	15	0	0	0	0	0	15
AA	0	412	29	0	0	0	0	441
A	0	20	2490	116	0	0	0	2626
BBB	0	0	92	2888	61	0	0	3041
BB	0	0	0	58	1259	16	0	1333
B	0	0	0	0	12	175	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	0	447	2611	3062	1332	191	0	7643

AR=94.52%
CP=74.25%

Table B30: In-sample Prediction in Continuation Ratio Model with LASSO by AIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	3	12	0	0	0	0	0	15
AA	0	409	32	0	0	0	0	441
A	0	15	2501	110	0	0	0	2626
BBB	0	0	72	2914	55	0	0	3041
BB	0	0	0	47	1273	13	0	1333
B	0	0	0	0	12	175	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	3	436	2605	3071	1340	188	0	7643

AR=95.19%
CP=78.58%

Table B31: In-sample Prediction in Continuation Ratio Model with LASSO by BIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	15	0	0	0	0	0	15
AA	0	412	29	0	0	0	0	441
A	0	15	2495	116	0	0	0	2626
BBB	0	0	82	2897	62	0	0	3041
BB	0	0	0	55	1263	15	0	1333
B	0	0	0	0	11	176	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	0	442	2606	3068	1336	191	0	7643

AR=94.77%
CP=74.51%

Table B32: In-sample Prediction in Continuation Ratio Model with Elastic Net by AIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	3	12	0	0	0	0	0	15
AA	0	409	32	0	0	0	0	441
A	0	16	2497	113	0	0	0	2626
BBB	0	0	70	2918	53	0	0	3041
BB	0	0	0	49	1272	12	0	1333
B	0	0	0	0	13	174	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	3	437	2599	3080	1338	186	0	7643

AR=95.16%
CP=78.45%

Table B33: In-sample Prediction in Continuation Ratio Model with Elastic Net by BIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	15	0	0	0	0	0	15
AA	0	412	29	0	0	0	0	441
A	0	16	2494	116	0	0	0	2626
BBB	0	0	83	2896	62	0	0	3041
BB	0	0	0	56	1263	14	0	1333
B	0	0	0	0	12	175	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	0	443	2606	3068	1337	189	0	7643

AR=94.73%
CP=74.39%

Table B34: In-sample Prediction in Principal Component Analysis_15 PCs with Ordered Probit model in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	14	1	0	0	0	0	15
AA	3	152	278	8	0	0	0	441
A	0	148	1651	817	10	0	0	2626
BBB	0	4	559	2318	160	0	0	3041
BB	0	0	0	612	674	47	0	1,333
B	0	0	0	7	113	67	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	3	318	2489	3762	957	114	0	7643

AR= 63.61%
CP= 31.99%

Table B35: In-sample Prediction in Principal Component Analysis_23 PCs with Ordered Probit model in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	15	0	0	0	0	0	15
AA	0	174	258	9	0	0	0	441
A	1	102	1929	589	5	0	0	2626
BBB	0	1	433	2440	164	3	0	3041
BB	0	0	0	408	864	61	0	1333
B	0	0	0	5	86	96	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	1	292	2620	3451	1119	160	0	7643

AR= 72.00%
CP= 41.86%

Table B36: In-sample Prediction in Principal Component Analysis_41 PCs with Ordered Probit model in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	15	0	0	0	0	0	15
AA	0	286	154	1	0	0	0	441
A	0	74	2052	500	0	0	0	2626
BBB	0	0	393	2465	183	0	0	3041
BB	0	0	0	272	990	71	0	1333
B	0	0	0	9	91	87	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	0	375	2599	3247	1264	158	0	7643

AR= 76.93%
CP= 48.97%

We cross tabulate predicted against observed EQIRs outcomes in contingency Table B37 to B48 for the out-of-sample prediction.

Table B37: Out-of-sample Prediction in Ordered Probit Model in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	7	51	45	11	0	0	0	114
A	28	13	248	172	38	0	0	499
BBB	11	26	17	185	123	34	0	396
BB	0	9	14	11	78	40	0	152
B	0	0	0	1	2	17	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	46	100	324	380	241	91	0	1182

AR= 48.98%
CP= 35.49%

Table B38: Out-of-sample Prediction in Ordered Probit Model with LASSO by AIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	0	93	21	0	0	0	0	114
A	0	7	434	58	0	0	0	499
BBB	0	0	21	334	41	0	0	396
BB	0	0	0	17	121	14	0	152
B	0	0	0	0	0	20	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	101	476	409	162	34	0	1182

AR=84.77%
CP=66.50%

Table B39: Out-of-sample Prediction in Ordered Probit Model with LASSO by BIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	0	109	5	0	0	0	0	114
A	0	8	475	16	0	0	0	499
BBB	0	0	39	349	8	0	0	396
BB	0	0	0	26	124	2	0	152
B	0	0	0	0	2	18	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	118	519	391	134	20	0	1182

AR=90.95%
CP=70.10%

Table B40: Out-of-sample Prediction in Ordered Probit Model with Elastic Net by AIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	0	85	29	0	0	0	0	114
A	0	5	411	82	1	0	0	499
BBB	0	0	18	319	59	0	0	396
BB	0	0	0	16	121	15	0	152
B	0	0	0	0	0	20	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	91	458	417	181	35	0	1182

AR=80.88%
CP=63.42%

Table B41: Out-of-sample Prediction in Ordered Probit Model with Elastic Net by BIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	0	109	5	0	0	0	0	114
A	0	9	474	16	0	0	0	499
BBB	0	0	39	349	8	0	0	396
BB	0	0	0	27	123	2	0	152
B	0	0	0	0	2	18	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	119	518	392	133	20	0	1182

AR=90.78%
CP=69.93%

Table B42: Out-of-sample Prediction in Continuation Ratio Model with LASSO by AIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	0	93	21	0	0	0	0	114
A	0	6	436	57	0	0	0	499
BBB	0	0	21	335	40	0	0	396
BB	0	0	0	15	126	11	0	152
B	0	0	0	0	0	20	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	100	478	407	166	31	0	1182

AR=85.45%
CP=67.29%

Table B43: Out-of-sample Prediction in Continuation Ratio Model with LASSO by BIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	0	109	5	0	0	0	0	114
A	0	8	475	16	0	0	0	499
BBB	0	0	39	349	8	0	0	396
BB	0	0	0	25	125	2	0	152
B	0	0	0	0	2	18	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	118	519	390	135	20	0	1182

AR=91.03%
CP=70.23%

Table B44: Out-of-sample Prediction in Continuation Ratio Model with Elastic Net by AIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	0	86	28	0	0	0	0	114
A	0	7	410	82	0	0	0	499
BBB	0	0	18	321	57	0	0	396
BB	0	0	0	18	119	15	0	152
B	0	0	0	0	1	19	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	94	456	421	177	34	0	1182

AR=80.80%
CP=62.39%

Table B45: Out-of-sample Prediction in Continuation Ratio Model with Elastic Net by BIC-typing Tuning Parameter Selector in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	0	109	5	0	0	0	0	114
A	0	8	475	16	0	0	0	499
BBB	0	0	39	349	8	0	0	396
BB	0	0	0	26	124	2	0	152
B	0	0	0	0	2	18	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	118	519	391	134	20	0	1182

AR=90.95%
CP=70.10%

Table B46: Out-of-sample Prediction in Principal Component Analysis_15 PCs with Ordered Probit model in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	0	51	61	2	0	0	0	114
A	0	18	357	119	5	0	0	499
BBB	0	0	84	269	43	0	0	396
BB	0	0	7	80	65	0	0	152
B	0	0	0	1	12	7	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	70	509	471	125	7	0	1182

AR= 63.37%
CP= 32.39%

Table B47: Out-of-sample Prediction in Principal Component Analysis_23 PCs with Ordered Probit model in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	4	76	34	0	0	0	0	114
A	2	120	355	22	0	0	0	499
BBB	0	16	206	173	1	0	0	396
BB	0	0	32	83	36	1	0	152
B	0	0	0	1	12	7	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	6	213	627	279	49	8	0	1182

AR= 54.74%
CP= 28.04%

Table B48: Out-of-sample Prediction in Principal Component Analysis_41 PCs with Ordered Probit model in EQIRs

	Predicted EQIRs							
Actual EQIRs	AAA	AA	A	BBB	BB	B	Below CCC	Total
AAA	0	1	0	0	0	0	0	1
AA	10	80	24	0	0	0	0	114
A	2	150	323	24	0	0	0	499
BBB	0	14	210	171	1	0	0	396
BB	0	0	25	92	35	0	0	152
B	0	0	0	4	14	2	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	12	245	582	291	50	2	0	1182

AR= 51.69%
CP= 22.22%

Appendix C

In this section, we report the procedure for selecting tuning parameters.

In our study, we select the tuning parameter with the aid of the AIC (Akaike 1974) and the BIC (Schwartz 1978). These are presented below:

$$AIC = -2 * \log(maximum\ likelihood) + 2 * \ (the\ number\ of\ used\ parameters\ wthin\ the\ model)$$

and,

$$BIC = -2 * \log(maximum\ likelihood) + 2 * \ (the\ number\ of\ used\ parameters\ wthin\ the\ model) * \log(the\ number\ of\ observations)$$

The number of used parameters within the model is the number of non-zero estimated coefficients of used parameters (Efron et al. 2004 and Zou et al. 2007). We follow the parametrization procedure of Wurm et al. (2017). The exact algorithm for the LASSO tuning parameter based on the AIC and the BIC criteria is presented below:

1. Estimate the intercept-only model by maximum likelihood (see equation (4) for the OP with LASSO or equation (9) for the CR with LASSO).

2. Select a sequence of tuning parameters λ_i $(i = 1\ to\ n)$.

   For each tuning parameter λ_i, there exists a set of selected independent variables and corresponding estimated coefficients. In other words, different potential models including different selected predictors controlled by the sequence of tuning parameters $\lambda_1, \lambda_2, \dots, \lambda_n$ are constructed.

3. Calculate the AIC or BIC for each potential model.

4. Choose the models with minimum value of AIC or BIC from the aforementioned series. These two models (one based on AIC-type tuning parameter selector and one based on BIC-type tuning parameter selector) are applied in our study.

Similarly, for the Elastic net estimator we follow Wurm et al. (2017):

1. Estimate the intercept-only model by maximum likelihood (see equation (6) for the OP with Elastic net or equation (11) for the CR with Elastic net).

2. Select a sequence of tuning parameters $\alpha_1, \alpha_2, \dots, \alpha_{10}$.

3. For each α_i $(i = 1\ to\ 10)$, a sequence of tuning parameters λ_i $(i = 1\ to\ n)$ is selected. For each tuning parameter λ_i, there is a set of selected independent variables and the corresponding estimated coefficients. For each of these models, we generate an AIC and BIC value.

4. The minimum AIC and BIC values are assigned to the corresponding $\alpha_{i.}$

5. The α_i $(i = 1\ to\ 10)$ with the minimum AIC and the one with the minimum BIC value are selected. These two models are applied in our study.

Chapter 3 What influences a bank's decision to go public?

Abstract

A bank's decision to go public by issuing an Initial Public Offering (IPO) transforms its operations and capital structure. Much of the empirical investigation in this area focuses on the determinants of the IPO decision, applying accounting ratios and other publicly available information in non-linear models. We mark a break with this literature by offering methodological extensions as well as an extensive and updated US dataset to predict bank IPOs. Combining the least absolute shrinkage and selection operator (LASSO) with a Cox proportional hazard, we uncover value in several financial factors as well as market-driven and macroeconomic variables, in predicting a bank's decision to go public. Importantly, we document a significant improvement in the model's predictive ability compared to standard frameworks used in the literature. Finally, we show that the sensitivity of a bank's IPO to financial characteristics is higher during periods of global financial crisis than in calmer times.

Key words: Equity financing, US banks, financial ratios, LASSO, forecasting

3.1 Introduction

Market finance in the USA has become an important source of funding for banks. According to the Federal Reserve Board, over the period 1996 to 2016 the net new issuance of US financial corporate equities outstanding more than tripled, from less than \$50 billion to over \$150 billion. The same body reports that the market value of total US corporate equity issues rose from about \$8 trillion in 1996 to around \$36 trillion in 2016. This implies that market participants have taken advantage of economic conditions as interest rates fell to historic lows. But not all banks were in a position to benefit from these unusual conditions. Using new estimation techniques over an extensive sample that covers both periods of crisis and calmer times, the present study aims to identify the factors that influence a bank's desire to issue an equity IPO.

Our study considers the influence of bank-level financial information as well as market-level indicators; we ask how these explicators influence the decision at the level of the bank to issue stocks for the first time. The focus is on the decision of a bank to go public by issuing an Initial Public Offering (IPO), which is a financially significant step for a bank and provides new opportunities for financial flexibility, increased liquidity, better diversification, and attracting potential investors (Amihud and Mendelson 1988, Pagano 1993, Lowry 2003, Bodnaruk *et al.* 2008, Kim and Weisbach 2008, Lowry *et al.* 2017). In addition, Houge and Loughran (1999) demonstrate that a bank's IPO decision can help managers to satisfy regulatory capital requirements, sell overvalued stock, and take advantage of better growth opportunities. After going public, Harris and Raviv (2014) indicate that the conditions of underlying market discipline and capital markets have more considerable influence on a public bank's ability to take risk than on a private bank. Samet *et al.* (2018) further clarify that public banks are able to take less credit risk during non-crisis periods compared to private banks. Moreover, if banks go public, market discipline can improve credibility and transparency in the banking industry and force public banks to maintain operational quality because of regular announcements of their financial health (Delis *et al.* 2011).

In this chapter, we extend the literature methodologically, by developing a series of Cox proportional hazard, discrete hazard and logistic models combined with a more intuitive, yet innovative model, which is based on the variable selection technique, pioneered by Tibshirani (1996)—the least absolute shrinkage and selection operator (LASSO). Our study, as far as we know, is the first to apply this methodology to analyse the timing of a bank's decision to issue for the first time in the public market. This model, also known as L1 norm penalty, has proved very useful in identifying the most relevant predictors from an extensive set of candidate variables, without considering a pre-selection of these potential variables (van de Geer 2008). The LASSO selection approach has a number of appealing characteristics: it not only helps identify the most relevant predictors from an extensive set of candidate variables, but it also improves the predictive power (Fan and Li 2001, Tian *et al.* 2015). In addition, LASSO does not require strict assumptions such as a pre-selection of the variables considered, and it is consistent statistically, as the number of observations approaches infinity (van de Geer 2008). Importantly, LASSO can potentially sidestep the problem of multi-collinearity, which is fairly common in reduced-form models, and it is computationally efficient even when considering a large set of potential predictors.

An additional important contribution of the present chapter is that we test our preferred estimator with superior predictive ability utilising a panel of US banks over an extensive time period. This approach not only allows us to compare our results with previous research, but also consider different time periods. Intuitively, banks respond in a different manner to extreme economic events as opposed to non-crisis periods, when they time their IPOs. Our sample covers the most recent global financial crisis as well as calmer (pre- and post-crisis) periods. We argue that across time periods, there is a differential sensitivity to bank and market information when it comes to the probability of banks going public.

To preview our findings, we discover value in several bank-specific financial factors as well as market-driven and macroeconomic variables in predicting the decision of banks to go public. In terms of the models' predictive ability, when we apply the LASSO estimator in a Cox proportional hazard model, we note a significant improvement in predicting a bank's IPO and the penalized Cox proportional hazard model outperforms other candidates. Specifically, we note improvements compared to a Cox proportional hazard, discrete hazard and logistic models with or without LASSO. On the other hand, we show that the Cox proportional hazard model underperforms discrete hazard and logistic models, which highlights the effect of LASSO on our algorithms. Our L1 penalized models are tuned through the AIC and the BIC criteria. We observe increased predictability on our dataset when the latter criterion is applied. Finally, when we apply the model with superior predictive ability to the data and split our sample into crisis and non-crisis periods, we find that the above variables become more potent in determining banks' IPOs. This finding signifies the ability of banks to time their IPOs relative to the economic conditions.

The rest of this work is laid out as follows. In section 2, we present an overview of the relevant literature. The data statistics and methodologies are introduced in Sections 3 and 4 respectively. Section 5 explains the empirical results of the forecasting simulation and Section 6 presents the econometric results of an empirical application. Section 7 provides conclusions.

3.2 Literature

The literature of determining IPO decisions contains two parts: one is associated with the evaluation of the motivations of a firm making an IPO decision and the other is related to empirical works to validate these motivations. The reasons for a firm to make an IPO decision are documented first. These are related to how a private firm assesses the costs and benefits of being a public firm. They can be categorized into six groups: capital structure; diversification; control consideration; fixed cost and loss confidentiality; adverse selection; and potential investors and customers.

The preliminary and intuitive reason for a firm going public is to access different sources of capital, which can support the subsequent growth of this firm (Lowry *et al.* 2017). Röell (1996) confirms that intensifying competition among financial suppliers in the public market makes for more efficient investment and means that private capital can be negotiated at better rates. This can encourage private firms to go public. Lowry (2003) and Kim and Weisbach (2008) also suggest that a private firm is more likely to decide IPO issuance in order to access more current and future investment, reduce financial constraints and increase its value. However, as the cost of capital in the public market increases, more companies are likely to stay in the private market, since producing efficient information in the public market (which may be not liquid) is costly (Modigliani and Miller 1963, Grullon *et al.* 2015).

When it comes to diversification, becoming a public firm can provide diversified opportunities for the owners and reduce their unsystematic risk due to trading holding shares in the secondary market. Pagano (1993) applies a simple illustrative model to analyse the importance of diversification in the decision to go public. He concludes that a firm has greater likelihood of going public to diversify portfolios if borrowing constraints and a lack liquidity have been experienced. Bodnaruk *et al.* (2008) use all detailed IPO data in Sweden from 1995 to 2001 to estimate the influence of controlling diversification of owners on the IPO process. They indicate that shareholders with less diversification are more likely to obtain positive profits after going public and that they then tend to sell their shares during the IPO process. Their findings further support the argument that diversification is a major element in the decision to go public.

Moving to control consideration, owners of private firms going into the public market may find it easy to transfer control. Zingales (1995) posits that the decision about an IPO can be regarded as the first step in selling a firm with maximized profits, since initial holders of a private firm can transform the scale of cash flow rights and manage rights maintained through issuing IPOs. Brau and Fawcett (2006) examine the principal motivation of a firm for going public, using surveys for 336 chief financial officers (CFOs). Their main conclusion is that the significant consideration of a firm going public is to promote future acquisitions. Minimizing the cost of capital cannot be considered to be an important motivation. Hsieh *et al.* (2011) advise that the decision regarding an IPO enables potential buyers to evaluate the true value of a targeted firm and choose the optimal form of restructuring. Thus, the IPO process can lower the uncertainty of valuation of a firm in the public and help create a more appropriate acquisition strategy, which can improve the value of the firm.

The fixed cost and loss confidentiality referring to regulation in the public market also influences the IPO decision. High explicit fixed costs have to be spent by a firm during IPO issuance, including various initial fees to pay for achieving certification in the public market and other variable expenses have to be made to keep this certification every year (Ritter 1987). Pagano and Roell (1998) confirm that large direct costs with fixed components are an important factor in a firm's decision to go public, since these costs are treated as necessary expenses to get certification in the public market. The decision to issue IPOs may negatively affect the operating performance of the company in the long run, since the fixed costs are a burden for a company with poor profitability (Brau *et al.* 2003). After going public, public firms are obliged to disclose some inside or sensitive information to maintain the efficiency of scrutiny, which causes loss of confidentiality. Campbell (1979) demonstrates that owners of a firm can utilize inside information to improve profits and thus a firm is less able to finance in the public market with an increased degree of confidentiality. Since the disclosed information may be related to scheduled or ongoing Research & Development (R&D) projects or an investment plan, it lessens the proportion of tax elusion and evasion (Pagano *et al.* 1998), leads to fierce competition in the entire industry and reduces the potential revenue of this listed firm (Maksimovic and Pichler 2001). To keep more potential profits from suffering from confidentiality issues, private firms are likely to stay in the private market.

We turn now to explaining the correlation between adverse selection and IPO issuance. It is well accepted that the process of IPO issuance can generate information asymmetry, since managers or initial owners normally have more information about the true value of targeted firms which are public than outside investors. Potential investors prefer to use a lower price to purchase IPOs in order to protect their potential profits, and this becomes a factor associated with the under-pricing of IPOs (Rock 1986). This under-pricing of IPOs prevents small and young firms from going public and as such firms are unwilling to release efficient information into the financial market, this leads to adverse selection. Lowry (2003) concludes that adverse selection costs negatively affect the decision regarding IPOs. Brau *et al.* (2003) also demonstrate that the adverse selection cost becomes a relatively serious obstacle for young and small companies going public since they have a limited tracking record and poor visibility.

Investors and customers as outsiders of a firm is the other motivation for an IPO decision. It is possible for investors to miss the investment opportunities offered by a company due to financial fraction. Going public through IPO issuance becomes the appropriate market strategy for a firm to show its major influence to other firms in its sector and hence this decision can attract more attention from latent investors and customers in the public market (Lowry *et al.* 2017). This issue of IPOs also can be regarded as advertising to expand client groups, improve client loyalty and increase stock liquidity (Lowry *et al.* 2017). When clients buy shares in this firm, the benefits of issuing IPOs can be further extended, which can improve income and lower the cost of capital. As the operation of private firms going public improves, more investors have more confidence investing in them.

To confirm these motivations, there is a large amount of empirical work about determining IPO decisions. Pagano *et al.* (1998) present the first systematic study on the determinants of firms' IPOs and suggest determinants of IPOs can be found through ex ante and ex post features of IPOs for Italian companies. According to the reliable dataset extracted from three databases, they measure the performance of a firm based on size, capital expenses, future investment opportunities in the corresponding industry, leverage, profitability and the level of concentration. They then employ the probit model to analyse the decision of firms to go public, looking at each year from 1982 to 1992. Through hypothesis tests, it can be seen that size, growth rate, profitability and future investment opportunities are statistically significant in their work, which indicates a positive relationship with the probability of IPOs. Therefore, they conclude that larger firms, those with higher growth rates, or improved future investment opportunities, are more likely to go public. Following a similar objective, Fischer (2000) extracts balance sheet information from technology-based German companies between 1993 and 1997 as a dataset to analyse the decision about IPOs in a logistic model. In his conclusion, a technology firm operating with a higher proportion of intangible assets or R&D intensity is more likely to go public, which does not prove that firms with a greater level of confidentiality are less likely to be public. This may be caused by considerations which have been overlooked, such as the degree of concentration or market competition. The second finding of Fischer (2000) is that managers or initial owners prefer to list their firm in the public when other stock market segments are booming.

Since the financial information of private firms is not easy to collect, Boehmer and Ljungqvist (2004) employ social media to collect data and improve the data quality of 330 privately-held German firms, and then determine the probability of IPOs from 1984 to 1995. Despite the extremely original source of the dataset, the major distinction of this work in comparison to the previous literature is the application of the hazard model, which is a branch of survival analysis. It focuses on analysing the segment of time before a targeted event occurs, which can be linked with potential elements mentioned by relevant hypothesis. The major conclusion is that firms operating with larger future investment opportunities and higher valuations were more likely to complete IPOs. Several other studies confirm the importance of financial health in determining access to the public market in the UK and India (Gill de Albornoz and Pope 2004, Mayur and Kumar 2013). One other conclusion of these works that should be mentioned is that a firm operating with higher probability was less likely to go public from 1990 to 2000 in the UK (Albornoz and Pope, 2004). In a slightly different setting, Helwege and Packer (2003) exploited the requirements of the Securities and Exchange Commission to obtain information about US public firms. The authors show that variables measuring size, profitability, leverage, interest coverage, R&D investment, capital structure, growth rate, future investment opportunities, ownership information and riskiness all have an important role in influencing the decision to issue an IPO.[1]

As well as evaluating the importance of financial information, Chemmanur *et al.* (2010) apply annual data drawn from the Longitudinal Research Database (LRD) containing all private and public enterprises related to manufacturing industry during the period 1972 – 2000. Following previous research, they include both a dynamic probit model and a Cox proportional hazard model to define the elements of the decisions regarding IPOs. It is interesting that there is no significant difference in conclusions between the probit model and a Cox proportional hazard model. They detect that total factor productivity is a key contributor to the probability that a firm will issue IPOs. They find that a private firm which is larger, with a better growth rate and higher total factor productivity is more likely to go public compared to its counterparts. Moreover, they show that if a private firm operates in an industry which is facing a higher degree of information asymmetry or costly evaluation of projects for outsiders, it is less likely to go public. Combined with the analysis of post-IPO performance, the authors conclude that a firm is more likely to issue an IPO at the peak of its productivity cycle.

[1] This lends support to the finding of Pagano et al. (1998) and further demonstrates that issuing IPOs can be regarded as a primary mechanism to raise outside equity.

Taken from a different perspective, researchers have sought to explain the influence of the market environment on firms' IPO decisions. Subrahmanyam and Titman (1999) first demonstrated that public financing is more attractive than private financing if valuable information can be cheaply obtained and then confirmed that companies benefit more by issuing IPOs in a large, liquid public market. These benefits can encourage private companies to go public by issuing IPOs. Pástor and Veronesi (2005) further indicated that private firms are more likely to make IPO decisions when market conditions improve, or stock prices increase. In contrast to this, Helwege and Liang (2004) indicated that clusters of IPOs in stock markets are positively associated with investor optimism and are not related to the characteristics of industries such as profitability or growth opportunities.

Some work has also been done on distinguishing private firms' choices between IPO issuance and takeover. Brau *et al.* (2003) applied logistic regression for US companies, 1984-1998. They conclude that the level of concentration, the relation with the technology industry, hot issue period of IPOs, the cost of debt, company size and the proportion of insider ownership can result in a higher likelihood of private companies issuing IPOs. Nevertheless, firms operating in industries with high future investment opportunities, financial services, highly leverage or greater liquidity for trading insiders are more likely candidates for takeover. Adjei *et al.* (2008) collected the accounting information of firms going public by reverse merge or IPOs from 1990 to 2002 and then determined the decision of these firms to choose IPO. They suggested that private firms which were larger, had a longer history and better operating performance were on average more likely to issue IPOs to go public and more likely to fulfil the requirements of targeted stock exchange.

Contrary to the above literature about the determinants of IPO decisions, the opposite direction, explaining why firms delisted from the public market, is similarly an interesting one to explore. The related studies look at the reasons for firms going public to determine why they go private. Bharath and Dittmar (2010) examined the decision of public companies to go private between 1980 and 2004 in the US based on the reverse motivations of private firms going public. To construct the dataset, the legal definition of going private according to the Securities and Exchange Commission was applied, keeping the unique criteria in data collection. Following this, similar predictors were drawn from the Securities Data Corporation (SDC) database and Securities and Exchange Commission to measure size, growth rate and leverage and so on in the Cox proportional hazard model. In their conclusion, information and liquidity were proved to be important elements for private firms going public, while the ability to gain capital and controlling power were regarded as crucial determinants of public firms going private. At the same time, they confirm the predictive ability of the Cox proportional hazard model, which can achieve around 80% correct prediction of firms going private.

While the literature on firms' IPOs is vast, the decision of banks to go public is less well studied. Kashian and Ahmad (2010) used quarterly data of twenty-three credit unions provided by SNL Financial Services in a Cox proportional hazard model, capturing duration dependence, and they demonstrated that the quality of both assets and loans was linked to the IPO decision for credit unions that had converted into mutual savings institutions. They also observed that banks with either low or very high ratios of equity to assets were less likely to make IPO decisions. On the other hand, the return on equity, the ratio of total loans to total assets and the size of the institution were not important determinants of a bank's decision to go public. Francis *et al.* (2009) used 272 US bank IPOs and 440 bank mergers and acquisitions from the Securities Data Company (SDC) Global New Issues database in the logistic regression to distinguish IPO decisions from mergers and acquisitions from 1985 to 1999. They demonstrated that a bank is less likely to go public during difficult economic times. In addition, Geyfman (2014) employed a cross-sectional dataset which included 208 large commercial banks with more than $200 million in assets in 20 transition economies in Central and Eastern Europe and the former Commonwealth of Independent States in 2010. The author found that banks operating in advanced and mature markets are more inclined to go public, highlighting the role of financial architecture.

The above review suggests the relevant predictors and econometric methods in IPO decisions. It provides us with a chance to make a methodological contribution on this topic. The dataset used and the estimation strategy will be discussed in the following sections.

3.3 Data and summary statistics

3.3.1 Data description

Our dataset is drawn from the quarterly accounting reports taken from the Orbis Bank Focus database, published by Bureau van Dijk Electronic Publishing (BvDEP). The Bank Focus database provides information on almost 40,000 institutions across the globe, with detailed coverage in the US. Moreover, the accounting information for each bank can be compared directly in the BvDEP since it has adjusted the differences in accounting and reporting criteria and converted into standardized format. High-quality information can be offered in the BvDEP since it is constantly updated with the latest accounting and regulatory disclosures. We rely on Orbis Bank Focus to identify banks' IPO date over the period 1996–2016. The distribution of public and private banks studied is presented in Table 3-1.

Table 3-1 The distribution of banks

Year	Public Banks	Private Banks	Public banks percentage	Total
1996	176	345	33.78%	521
1997	194	406	32.33%	600
1998	203	467	30.30%	670
1999	222	603	26.91%	825
2000	229	677	25.28%	906
2001	242	742	24.59%	984
2002	265	6007	4.23%	6272
2003	276	6177	4.28%	6453
2004	285	6265	4.35%	6550
2005	301	6477	4.44%	6778
2006	251	5705	4.21%	5956
2007	256	5824	4.21%	6080
2008	237	5377	4.22%	5614
2009	257	5832	4.22%	6089
2010	275	5847	4.49%	6122
2011	283	5755	4.69%	6038
2012	301	6278	4.58%	6579
2013	307	6290	4.65%	6597
2014	319	6265	4.85%	6584
2015	298	5966	4.76%	6264
2016	292	5766	4.82%	6058

Note: This table presents the distribution of banks by year.

Data on market indicators and macroeconomic variables are sourced from Bloomberg. These data items are reported quarterly. Following commonly used selection criteria in the literature, we exclude banks that do not have complete records on our explanatory variables and bank quarters with negative sales and assets. To control for the potential influence of outliers, we winsorize the regression variables at the 1st and 99th percentiles.

3.3.2 Choice of explanatory variables

Our models are supplied with forty-two potential explanatory variables, which can be divided into the following broad categories: bank-specific indicators, industry-specific predictors and macroeconomic variables. The choice of the explicators is based on a series of related studies (see for instance Pagano *et al*. 1998, Brau *et al*. 2003, Helwege and Packer 2003, Pástor and Veronesi 2005, Adjei *et al*. 2008, Tregenna 2009, Chemmanur *et al*. 2010, Kashian and Ahmad 2010, Geyfman 2014). To begin with the bank accounting variables, which measure various aspects of banks' health, these potential predictors are related to the determinants of CAMELS ratings. Specifically, they are aimed at assessing the overall safety and soundness of banks, covering capital adequacy; asset quality; management quality; earnings; liquidity; and sensitivity to market risk. Next, our industry-specific variables capture market concentration. Finally, we allow for fourteen macro-economic covariates that are likely to influence the timing of a bank's IPO.[2]

3.3.3 Summary statistics

We report summary statistics of the variables used in the empirical models in Table 3-2. We also present p-values for the tests of equality of means across the public and private banks in column 6 of Table 3-2. We observe, as expected, that public banks' size, growth rate, market share and income diversification are higher compared to private banks. On the other hand, capital, leverage and deposits in public banks are lower than in private banks. These statistics imply that public banks may absorb more growth opportunities from the stock market to

[2] For detailed definitions and abbreviations of all variables see Table A.1 in the Appendix. Table A.2 presents the cross-correlations between bank-specific variables. It is generally observed that some variables exhibit relatively high correlation with each other, with some exceptions for variables that measure similar dimensions (e.g. banks' profitability using ROAA and ROAE). We note, however, that our preferred empirical methodology will carefully address this issue.

enhance their performance and reduce risk. Overall, the tests point to significant differences between the two groups, which indicate that there is a correlation between banking activities and the decision about IPOs. Moving to the industry-specific indicators, we find significant differences between public and private banks, suggesting a link between the market climate and a bank's likelihood of going public.

Table 3-2 Summary statistics

Variable	Status	Mean	Standard Deviation	Minimum	Maximum	p-value
	(1)	(2)	(3)	(4)	(5)	(6)
Bank-specific						
LNDETAS	Public	14.334	1.475	10.896	19.800	0.000
	Private	12.120	1.188	9.328	16.760	
LNDETAS2	Public	24.063	2.949	17.187	34.996	0.000
	Private	19.635	2.375	14.052	28.915	
GROAS	Public	2.212	4.303	-7.052	34.731	0.000
	Private	1.553	4.037	-9.679	28.596	
LOAAS	Public	65.881	10.779	18.460	86.760	0.000
	Private	62.668	14.066	12.921	89.681	
EQAS	Public	9.965	2.349	4.376	28.822	0.000
	Private	10.766	3.154	4.473	52.197	
LIQASTAS	Public	5.336	3.890	1.018	35.805	0.000
	Private	9.039	7.126	1.055	51.237	
NETLOADEPSTFUN	Public	78.688	13.754	25.987	126.787	0.000
	Private	72.917	16.755	16.893	113.334	
NETLOATAS	Public	65.881	10.779	18.460	86.760	0.000
	Private	62.668	14.066	12.921	89.681	
DEPSTFUNTAS	Public	84.054	5.154	38.180	92.784	0.000
	Private	86.148	4.462	21.859	93.437	
LIQASDEPSTFUN	Public	6.341	4.663	1.291	50.953	0.000
	Private	10.474	8.235	1.350	60.950	
ROAA	Public	0.966	0.565	-5.249	2.890	0.012
	Private	0.950	0.741	-5.681	5.072	
ROAE	Public	10.011	6.114	-52.624	26.988	0.000
	Private	9.309	7.562	-52.168	35.841	
NETINTMAR	Public	3.960	0.762	1.500	7.350	0.000
	Private	4.057	0.825	1.381	7.999	
TCAPTAS	Public	10.354	2.059	5.432	29.297	0.000
	Private	11.101	3.035	5.526	50.690	
TIER1CAPTAS	Public	9.286	2.080	4.311	28.643	0.000
	Private	10.314	3.063	4.643	50.146	
LOALOSPROLOA	Public	0.108	0.155	-0.129	1.442	0.000
	Private	0.090	0.159	-0.144	1.671	
PROGRO	Public	0.720	9.249	-44.202	64.424	0.369
	Private	0.647	10.513	-49.428	66.672	
OPEXPTAS	Public	0.754	0.209	0.313	2.256	0.211
	Private	0.755	0.267	0.270	6.032	
COSINC	Public	64.386	12.073	35.777	158.992	0.000
	Private	68.609	15.566	31.229	178.738	
OVHTAS	Public	0.754	0.209	0.313	2.256	0.211
	Private	0.755	0.267	0.270	6.032	
MSAS	Public	0.112	0.415	0.001	6.978	0.000
	Private	0.004	0.011	0.000	0.249	
DEPLOA	Public	127.984	29.081	71.429	372.021	0.000
	Private	144.767	46.532	80.322	450.640	
DEPLOAGRO	Public	-0.108	3.774	-13.184	16.239	0.000
	Private	0.179	5.210	-17.411	21.725	
INCDIV	Public	22.592	11.542	-2.613	81.469	0.000
	Private	16.400	10.111	-5.800	95.082	
Industry-specific						
HHI3	Public	8.439	2.124	4.610	12.368	0.000
	Private	3.153	1.579	1.739	94.565	
HHI5	Public	9.415	1.827	6.209	12.972	0.000
	Private	3.731	1.584	2.292	94.568	
CON3	Public	48.917	6.158	36.667	60.279	0.000
	Private	29.781	4.725	22.743	98.826	
CON5	Public	62.483	4.487	54.539	70.307	0.000
	Private	39.698	4.855	32.906	99.603	

Notes: This Table reports summary statistics of the explanatory variables used in the empirical models. Column 6 reports the p-value for the test of equality of means between the public and private groups. A detailed description of the variables used in this study is given in Table A.1 in the Appendix.

3.4 Methodology

3.4.1 Cox proportional hazard model (CPH)

The advantage of hazard analysis compared to simple regressions is its ability to explicitly account for time and handle censored observations and time-varying covariates (Guo and Brooks 2009). In the Cox hazard model (Cox 1972), the dependent variable is constructed by the time spent by a bank in the first group (has not issued an IPO or 0) and the IPO status for a bank (1). When firms leave this group for any other reason than by issuing an IPO, they are considered censored. Simple regression models consider these banks to still be in the first group. In hazard models, information from censored and uncensored observations is combined in order to estimate consistent parameter estimates and provide accurate forecasts.

The CPH studies the effect of variables upon the time a specified event (IPO issuance) takes to happen. More specifically in the CPH model, the baseline function is not pre-specified and can take any form. Thus, unrealistic assumptions or approximations on the form of the dataset are not necessary. In our context, the CPH model is the likelihood of a bank issuing its equity IPO in a given quarter (last quarter) conditional on the fact that this bank did not undertake IPO in any of the previous quarters. Thus, the probability that a bank will issue an IPO takes the form:

$$Pr\left(Y_{i,t} = 1 \middle| Y_{i,t-1} = 0, X_{i,t}\right) = h\left(t, X_{i,t}\right) = h_0(t, 0)\exp\left(\beta' X_{i,t}\right), \qquad (1)$$

where $Y_{i,t}$ is equal to 1 if a bank issues IPO in the public market and 0 otherwise; $h\left(t, X_{i,t}\right)$ is the hazard rate at time t for a bank controlling by a set $X_{i,t}$ of time-varying indicators including bank-specific, industry-specific and macroeconomic variables; β is a vector of unknown parameters to be estimated and $h_0(t, 0)$ is the baseline hazard function without any restriction. The model is estimated by maximizing a partial-likelihood function.

3.4.2 Discrete hazard model (DH)

A bank can issue an IPO at any time within a quarter, but this event can only be observed when the information is released at the end of the corresponding quarter. The DH model is a discrete-time extension of the CPH that can capture this characteristic of our dataset and its estimation can be applied by the complementary log-log (cloglog) model (Grilli 2005, Jenkins 2005, Rabe-Hesketh and Skrondal 2012). It can contain time-varying explanatory variables in a discrete-time estimation without any correction (unlike CPH) while it is computationally efficient for large datasets (Allison 1982, Rabe-Hesketh and Skrondal 2012). Our model takes the form:

$$cloglog(h(t)^D) = ln\{-ln(1 - h(t)^D)\} = \gamma_0 + \gamma' X_{i,t}, \quad (2)$$

where γ_0 is the baseline hazard rate without any assumption; γ are estimated coefficient vectors and $X_{i,t}$ is our dataset.

3.4.3 Logistic model

Our dependent variable is binary and thus we also consider the logistic model, which is commonly used in the literature. The probability that a bank will issue an IPO based on the logistic model is:

$$Pr\left(Y_{i,t} = 1 \middle| X_{i,t}\right) = \frac{e^{\delta_0 + \delta' X_{i,t}}}{1 + e^{\delta_0 + \delta' X_{i,t}}}, \quad (3)$$

where δ_0 is the intercept to be estimated; δ are the estimated coefficient vectors in the logistic model and $X_{i,t}$ is our dataset.

3.4.4 LASSO

LASSO is a method of regression that enables estimation and variable selection simultaneously in a non-orthogonal setting (Tibshirani 1996). Based on a shrinkage factor, LASSO selects variables by forcing some coefficients to zero and shrinking others. The variance of the estimated value is decreased while the accuracy of the regression prediction is increased. Given a linear regression with standardized predictors and centred response values, LASSO resolves the l_1 -penalized regression problem of estimating B to minimize:

$$\sum_{i=1}^{N}\left(Y_{i,t}-\mathrm{B}'X_{i,t}\right)^{2}, subject\ to\ \sum_{q=1}^{p}\left|\mathrm{B}_{q}\right|\leq s. \tag{4}$$

The above can be written in Lagrangian form as:

$$\widehat{\mathrm{B}}=\arg min_{\mathrm{B}}\left(\sum_{i=1}^{N}\left(Y_{i,t}-\mathrm{B}'X_{i,t}\right)^{2}+\lambda\sum_{q=1}^{p}\left|\mathrm{B}_{q}\right|\right). \tag{5}$$

where $i=1,2,\ldots,N$ represents banks, $q=1,2...p$ indicates the surviving number of predictors with non-zero estimated coefficients and $t=1,2,\ldots,T$ represents different time periods. In equation (5), λ is the tuning parameter. The process of controlling different values of λ can be regarded as the procedure for selecting the number of independent variables in LASSO. As λ increases, the sum of absolute values of estimated coefficients is reduced, and shrinkage of coefficients is achieved. If λ exceeds a threshold value in the corresponding model, some estimated coefficients are ultimately set to zero. This procedure, the L1 norm penalty, generates a more interpretable and sparser model. Several approaches, such as cross-validation and information criteria, have been proposed in selecting the shrinkage factor λ. Zou *et al.* (2007) provided an algorithm to obtain the optimal LASSO fit with the Akaike information criterion (AIC) (Akaike 1974) and the Bayesian information criterion (BIC) (Schwarz 1978).[3] Sun and Zhang (2012) noted that the computational cost of applying cross-validation in penalized models is considerable, while the theory of applying cross-validation is poorly understood. Therefore, the AIC and the BIC are used in selecting the tuning parameter λ in the LASSO and CPH, the DH and the logistic model combinations presented below.

As discussed above, LASSO provides more stable and restricted models (Tibshirani 1996, Fan and Li 2001). In addition, it is a computationally simple and efficient

[3] It is well-known that AIC and BIC have different properties in model selection (for details see Yang, 2005; Shao, 1997 and Zhang et al, 2010).

method (Efron *et al.* 2004). Hence, these elements can lead to superior predictability for its outputs (Tibshirani 1996, Zou 2006).

3.4.5 L1 Penalized Semi-Parametric Cox Proportional Hazard Model (Penalized CPH model)

Tibshirani (1997) added the LASSO constraint form into the estimation of the CPH regression parameter and derived the L1 Penalized Semi-Parametric Cox Proportional Hazard Model. The LASSO estimator of the estimated coefficient β in the semi-parametric Cox proportional hazard model is:

$$\hat{\beta} = argmax\ l(\beta)\ subject\ to\ \sum_{q=1}^{p} |\beta_q| \leq s, \tag{6}$$

where the likelihood function $l(\beta)$ is $l(\beta) = \sum_{Y_i\, uncensored} \left\{\beta' X_{i,t} - \log\left(\sum_{Y_j \geq Y_i} \exp(\beta' X_{j,t})\right)\right\}$ in the semi-parametric Cox proportional hazard model and $X_{j,t}$ contains the pool of the potential predictors in equation (1).

The above can be written in Lagrangian form as:

$$\hat{\beta} = argmin \left\{ -l(\beta) + \lambda \sum_{q=1}^{p} |\beta_q| \right\}. \tag{7}$$

In equation (7), as λ increases, the sum of absolute values of estimated coefficients is decreased, and shrinkage of coefficients is achieved. If λ exceeds a threshold value in the corresponding models, some estimates are ultimately shrunk to zero. This “L1 norm penalty” generates a more interpretable and sparser Cox model. All explanatory variables are standardized before applying the LASSO estimator.

3.4.6 L1 Penalized Discrete Hazard Model (Penalized DH model)

In the L1 Penalized Discrete Hazard Model, the LASSO parameter of the coefficient γ is estimated by maximizing the log-likelihood function with a L1-norm penalty placed on the sum of the absolute value of the covariate parameters. The model can be expressed as:

$$\hat{\gamma} = argmax\ l(\gamma)\ subject\ to \sum_{q=1}^{p} |\gamma_q| \leq s, \tag{8}$$

where the log-likelihood function $l(\gamma)$ is equal to $\sum_{i \in Q} w_{i,t} \ln\{F(\gamma_0 + \gamma' X_{i,t})\} + \sum_{i \notin Q} w_{i,t} \ln\{1 - F(\gamma_0 + \gamma' X_{i,t})\}$ where Q is the set of all observations that $Y_{i,t} = 1$ and $F(\gamma_0 + \gamma' X_{i,t}) = 1 - \exp\{-exp(\gamma_0 + \gamma' X_{i,t})\}$ and $w_{i,t}$ represents the optional weights in the discrete hazard model and $X_{i,t}$ is the same used in equation (2).

Or alternatively as:

$$\hat{\gamma} = argmin \left\{ -l(\gamma) + \lambda \sum_{q=1}^{p} |\gamma_q| \right\}. \tag{9}$$

In line with this, all predictors are standardized before applying the LASSO estimator in this model.

3.4.7 L1 Penalized Logistic Model (Penalized Logistic Model)

The logistic model can be combined with LASSO as:

$$\hat{\delta} = argmaxl(\delta)\ subject\ to \sum_{q=1}^{p} |\delta_q| \leq s, \tag{10}$$

where $l(\delta)$ is $\sum \left(Y_{i,t} \left(\log \left(\frac{e^{\delta_0 + \delta' X_{i,t}}}{1 + e^{\delta_0 + \delta' X_{i,t}}} \right) \right) + (1 - Y_{i,t}) \left(\log \left(\frac{1}{1 + e^{\delta_0 + \delta' X_{i,t}}} \right) \right) \right)$ in the corresponding logistic model and $X_{i,t}$ is the same in equation (3). All independent variables are standardized before implementing the LASSO estimator in the logistic model.

The Lagrangian form is determined as:

$$\hat{\delta} = argmin\{-l(\delta) + \lambda \sum_{q=1}^{p} |\delta_q|. \quad (11)$$

3.5 Predictive ability

We begin our analysis by presenting a forecasting simulation exercise to determine which model has the superior predictive ability. To measure the predictive performance of all competing models, we calculate the area under receiver operating characteristic curve (AUC), the accuracy ratio, and the brier score (see Duffie *et al.* 2007, Tian *et al.* 2015).[4]

3.5.1 CPH and penalized CPH

Table 3-3 reports the AUC, the accuracy ratios and the brier scores for CPH and the penalized CPH. We benchmark their performance with a DH model, a logistic model and their two penalized variants. For the out-of-sample predictions of IPO decisions for banks, we use the past and current information and roll forward one step ahead of the prediction of the IPO decision. The initial estimation window is from 1996 to 2009.

Table 3-3 Accuracy ratios and the number of surviving variables in the CPH model and its penalized versions

Model		CPH model	Penalized CPH model		DH model	Penalized DH model		Logistic model	Penalized Logistic model	
			AIC	BIC		AIC	BIC		AIC	BIC
AUC	In-sample	0.238	0.779	0.779	0.746	0.745	0.745	0.749	0.748	0.689
	Out-of-sample	0.210	0.793	0.797	0.450	0.533	0.533	0.466	0.455	0.577
AR	In-sample	-0.524	0.557	0.557	0.492	0.489	0.489	0.497	0.497	0.379
	Out-of-sample	-0.581	0.586	0.594	-0.101	0.067	0.067	-0.068	-0.090	0.153
BS	In-sample	0.756	0.112	0.112	0.101	0.105	0.105	0.101	0.102	0.111
	Out-of-sample	0.707	0.109	0.109	0.205	0.195	0.195	0.203	0.202	0.182
Surviving variables		42	23	21	42	35	5	42	39	8

Notes: CPH model represents the Cox proportional hazard model and DH model refers to the discrete hazard model. “AUC” refers to the area under receiver operating characteristic curve. “AR” stands for accuracy ratio. “BS” represents the brier score. “AIC” is the AIC-type tuning parameter selector. “BIC” is the BIC-type tuning parameter selector.

[4] For a detailed description of the tests see section B in the Appendix.

In the in-sample, we note that the CPH presents an AUC of 24% and the penalized CPH lies at 78%, which is three times larger. Applying penalty function from LASSO seems to significantly improve the accuracy of the CPH model. The same trends can be observed from the accuracy ratios. For the brier scores, the value of about 0.76 reduces to about 0.11 in penalized CPH models, which confirms that the penalized CPH model can outperform the CPH. In the out-of-sample, we note a similar improvement in terms of accuracy for the penalized CPH model compared to its simple CPH counterpart, since the predictive ratio grows from 20% in the CPH to approximately 80% in the penalized CPH model. The penalized CPH models are tuned, based on the AIC and BIC criteria. We note that the BIC models present slightly better accuracy in the out-of-sample. The BIC models also select a lower number of predictors compared to the AIC.

Concerning our benchmarks, we note that the DH model and the logistic model can provide more accurate in-sample and out-of-sample predictions than the simple CPH model. Adding the LASSO estimator in the DH model can improve the proportion of correct out-of-sample predictions from 45 percent to 53 percent. This increase in predictive performance by adding the LASSO estimator can also be noted in the logistic model, which confirms that predictive ability can be improved by adding LASSO. In general, we note that in the out-of-sample the penalized CPH model has the more accurate forecasts for the measures retained.

3.5.2 Predictive Deciles

To confirm the above-mentioned results, IPO decisions by out-of-sample prediction decile is reported in Table 3-4. The decile method is frequently implemented in the default prediction (Shumway 2001, Chava and Jarrow 2004, Bharath and Shumway 2008, Tian *et al.* 2015, Traczynski 2017). The small changes in the predicted probabilities of IPO decisions do not have considerable influence on the decile in which a firm quarter lies in the distribution. The lowest probability of IPO decisions for banks would be included in the tenth decile and the highest would be in the first. Thus, the high proportion of banks appearing in the high probability for IPO decisions decile suggests high out-of-sample accuracy.

Table 3-4 IPO decision by out-of-sample prediction decile

Decile	CPH model	Penalized CPH model		DH model	Penalized DH model		Logistic model	Penalized Logistic model	
		AIC	BIC		AIC	BIC		AIC	BIC
1	0	64.22%	64.22%	11.01%	15.60%	15.60%	11.01%	13.76%	17.43%
2	0	4.59%	4.59%	5.50%	11.01%	11.01%	5.50%	4.59%	9.17%
3	0	4.59%	4.59%	10.09%	10.09%	10.09%	11.01%	11.93%	14.68%
4	0	5.50%	6.42%	8.26%	8.26%	8.26%	8.26%	7.34%	8.26%
5	0	2.75%	1.83%	8.26%	6.42%	6.42%	11.01%	7.34%	7.34%
6-10	100%	18.35%	18.34%	56.88%	48.62%	48.62%	53.21%	55.03%	43.13%
AUC	0.210	0.793	0.797	0.450	0.533	0.533	0.466	0.455	0.577

Notes: CPH model represents the Cox proportional hazard model. DH model refers to the discrete hazard model. “AUC” refers to the area under receiver operating characteristic curve. “AR” stands for accuracy ratio. “BS” represents the brier score. “AIC” is the AIC-type tuning parameter selector. “BIC” is the BIC-type tuning parameter selector.

From Table 3-4, we note that there is no observation in the first five percentiles in the CPH model, which suggests the lowest percentage of correct out-of-sample prediction among all candidate models (which is consistent with the AUC values in the previous sections). The highest percentage of correct out-of-sample prediction is about 80%, which can be observed from the first five deciles in the penalized CPH model. This confirms the main conclusion from the above-mentioned evaluation methods (AUC, accuracy ratio and brier score) and demonstrates that the penalized CPH model outperforms all models studied in out-of-sample predictability.

3.6 An empirical application using US data

Our findings thus far show that the penalized CPH model has substantial predictive ability compared to other models. We now present empirical evidence using data for US banks. Our extensive sample period covers the global financial crisis (2007-2009) which coincided with the collapse of the sub-prime mortgage lending market (Bekaert *et al*. 2014, Acharya and Mora 2015, Dungey and Gajurel 2015, Ramcharan *et al*. 2016). We therefore have a unique opportunity to examine the sensitivity of our findings to different economic conditions. Motivated by this consideration, we split our sample into three parts: the pre-crisis period (1996-2006), the crisis period (2007-2009) and the post-crisis period (2010-2016). In the sub-sections below, we discuss our findings for each sub-sample separately. Table

3-5 to Table 3-7 report the estimates of predictors, employing the CPH model and its corresponding penalized versions[5] for each sub-period.[6] A positive coefficient indicates that an increase in that explanatory variable will improve the likelihood of the IPO issue for a bank in any given quarter for a bank.

3.6.1 The pre-crisis period

To begin with the analysis of the CPH model, as shown in column 1 of Table 3-5, we observe that most bank-specific determinants behave according to our expectations. Specifically, an increase in the bank's size (LNDETAS) reduces the likelihood of the IPO issue in any given quarter and the estimate of LNDETAS2 illustrates a non-linear effect. These findings can be interpreted as follows. As banks grow in size they are less likely to issue IPOs, but once they attain a certain size threshold, the probability of issuing an IPO is positively associated with the bank's size.[7] This finding is not only statistically significant, but also economically important. A unit increase in LNDETAS is associated with a reduction of 76% in the likelihood of IPO issuance. As for banks' profitability (NETINTMAR), we find that it is negatively related to the probability of issuance. TIER1CAPTAS measures a bank's leverage and its estimated coefficient is positive and highly significant. A unit increase in this indicator (TIER1CAPTAS) improves the chances of an IPO issuance by 25%. This finding illustrates that banks with higher leverage are more likely to go public. The above findings on leverage and profitability suggest that banks with lower profitability and higher leverage are likely to make an IPO issue to diversify the credit risk (Gill de Albornoz and Pope 2004, Kim and Weisbach 2008). Finally, a decrease in capital (TCAPTAS) is likely to increase the probability of a bank going public. This is linked with the preliminary and intuitive consideration of going public, which is to tap into different sources of capital (Lowry et al, 2017).

---

[5] As an additional test, we replaced all macroeconomic determinants with time-fixed effects. Our results remain unaffected in all models.

[6] We opt for estimated coefficients instead of hazard ratios, since the direction of effects is more important than their magnitude.

[7] There is a line of thinking that argues the idea of "too big to fail" (TBTF) in the banking industry. Boyd and Heitz (2016) note that larger banks suffer more costs than benefits from TBTF in comparison with small and medium-sized banks. Therefore, larger banks may not go public because of the burden of the TBTF cost.

Table 3-5 The estimates of candidate models in the pre-crisis period

Variable	CPH model	Penalized CPH model_AIC	Penalized CPH model_BIC
	(1)	(2)	(3)
LNDETAS	-1.4340***	-2.607E-07***	
	(0.0000)	(0.0000)	
LNDETAS2	0.7176***		
	(0.0000)		
GROAS	0.0098	0.0111***	
	(0.2758)	(0.0000)	
LOAAS	-0.0434	-1.129E-06***	
	(0.4133)	(0.0000)	
LIQASTAS	-0.0220		
	(0.7749)		
NETLOADEPSTFUN	0.0373	-0.0039	-0.0035***
	(0.3979)	(1.0000)	(0.0000)
NETLOATAS	NA		
DEPSTFUNTAS	0.0356	-0.0135	-0.0135***
	(0.4663)	(1.0000)	(0.0000)
LIQASDEPSTFUN	0.0292		
	(0.6115)		
ROAA	0.3717		
	(0.3247)		
ROAE	0.0055		
	(0.8340)		
NETINTMAR	-0.3862***	-0.1439	-0.1567
	(0.0141)	(1.0000)	(1.0000)
TCAPTAS	-0.2128**		-0.0330***
	(0.0499)		(0.0000)
EQAS	0.0665		
	(0.2318)		
TIER1CAPTAS	0.2208**	0.0480	0.0840***
	(0.0490)	(1.0000)	(0.0000)
LOALOSPROLOA	0.4510	-0.2152***	
	(0.5053)	(0.0000)	
PROGRO	-0.0056	-0.0028***	
	(0.3047)	(0.0000)	
OPEXPTAS	1.2440	0.4969***	0.4868***
	(0.1140)	(0.0000)	(0.0000)
COSINC	0.0037		
	(0.7937)		
OVHTAS	NA	0.0300	0.0386
		(1.0000)	(1.0000)
MSAS	-1.4630**	-0.9438***	-0.9142***
	(0.0338)	(0.0000)	(0.0000)
DEPLOA	0.0009		
	(0.7583)		
DEPLOAGRO	0.0041		
	(0.7011)		
INCDIV	-0.0133		
	(0.2292)		
HHI3	3.8430	-0.0246***	
	(0.5143)	(0.0000)	
HHI5	-4.0980		
	(0.4938)		
CON3	-0.4539		
	(0.4614)		
CON5	0.5354		
	(0.4179)		

RSP500	0.0034	0.0085***	0.0074***
	(0.7727)	(0.0000)	(0.0000)
CPI	0.1336	0.2122***	0.1632***
	(0.5310)	(0.0000)	(0.0000)
GRGDP	0.1943	0.0456***	0.0580
	(0.2679)	(0.0000)	(1.0000)
LNGDPCAP	-0.4085	-1.8429***	-2.2406***
	(0.8069)	(0.0000)	(0.0000)
GRGNP	-0.5861	-0.2704***	-0.3669
	(0.3543)	(0.0000)	(1.0000)
INTR_10Y	1.1540		
	(0.4249)		
INTR_10Y2	-11.6200		
	(0.3525)		
INTR_3M	-0.0796		
	(0.8650)		
INTR_3M2	4.4950		
	(0.3156)		
SLYC	NA	-0.1343***	
		(0.0000)	
SLYC2	3.0640	3.4246***	
	(0.6633)	(0.0000)	
GRM1	0.0239		-0.0169***
	(0.7645)		(0.0000)
GRM2	-0.1683	-0.1844***	-0.1637
	(0.3716)	(0.0000)	(1.0000)
HPI	0.1482	0.0656	0.0926
	(0.1351)	(1.0000)	(1.0000)

Notes: CPH model represents the Cox proportional hazard model. “AIC” is the AIC-type tuning parameter selector. “BIC” is the BIC-type tuning parameter selector. P-values related to z-statistics reported in the parentheses are Huber-White robust estimates, clustered at the firm level. *** denotes significance at the 1% level. ** denotes significance at the 5% level. * denotes significance at the 10% level.

At the next stage, we add the penalized function into the CPH model. It should be noted that all surviving predictors after the penalty estimation are efficient variables that have predictive ability regarding the banks’ decision to issue an IPO. Thus, p-values are calculated under post-selection after fitting the LASSO with a fixed value of tuning parameter, since the estimated coefficients are shrunk in LASSO estimation to select the “best” model. Compared to the CPH model, there exist 22 surviving explanatory variables in the L1 penalized CPH model with AIC-type and 16 in the BIC-type tuning parameter selector. All surviving determinants in the L1 penalized model contain bank-specific, industry-specific and macroeconomic factors.

In the L1 penalized CPH model with AIC-type tuning parameter selector, as shown in column 2, LNDETAS, GROAS and MASA are selected, and they are statistically significant. The sign of LNDETAS and MASA is negative, which is consistent with the findings in the baseline model (CPH). The estimate of GROAS shows that an increase in asset growth rate can lead to an increased likelihood of banks going public. This is in line with previous work which notes that banks with more investment opportunities are more likely to raise external finance (Pagano *et al.* 1998). PROGRO measures the productivity growth rate and its estimate demonstrates that a bank with a higher productivity growth rate is less likely to go public. This suggests that banks can operate efficiently using internal funds and therefore may be less inclined to source external finance. As for operating expenses, OPEXPTAS is positive and statistically significant, which implies that a bank with lower management efficiency is more likely to go public.

With respect to industry and macroeconomic indicators, HHI3 is the only industry-specific variable that is kept in the model and that is statistically significant under post-selection. The negative estimate of this predictor illustrates the high degree of concentration in the banking industry that may prevent private banks going public. This confirms the findings of Grullon *et al.* (2015) that in industries with a relatively high concentration level in the US, firms can acquire more profits from mergers and acquisitions than from IPOs. Almost all macro-economic predictors are selected after the penalty. Overall, it appears that banks time their decision to go public and the probability of issuing is positively correlated with booming economic conditions.

In the L1 penalized CPH model with BIC-type tuning parameter selector, as reported in column 3, the selected predictors are slightly different from those in the model with the AIC-selector. In particular, no industry-related variables are included, while bank-specific variables such as liquidity, profitability, capital, leverage, operating expenses management and market share of a bank are found to be important determinants of the bank's IPO. Finally, several macroeconomic variables such as RSP500, CPI, LNGDPCAP and GRM1 are chosen in the model and are statistically significant under post-selection in line with the model that uses the AIC-type selector.

3.6.2 The crisis period

Starting with the analysis of the CPH model, as reported in column 1 of Table 3-6, most bank-specific variables are statistically significant. Importantly, the absolute value of these variables is higher compared to the pre-crisis period. This is a key finding which suggests that bank-specific variables are quantitatively more important predictors of IPOs during extreme economic conditions.[8] For example, the estimate of GROAS is only 0.0098 and not statistically significant in the pre-crisis period, while it increases to 0.1764 and becomes statistically significant at the 1% level in the crisis period.

Table 3-6 The estimates of candidate models in the crisis period

Variable	CPH model	Penalized CPH model_AIC	Penalized CPH model_BIC
	(1)	(2)	(3)
LNDETAS	-1.737E+03***	-0.0062*	-0.0072*
	(0.0000)	(0.0940)	(0.0560)
LNDETAS2	NA		
GROAS	0.1764***	-0.0244	
	(0.0000)	(0.9060)	
LOAAS	-1.1290***	-0.0191*	
	(0.0000)	(0.0980)	
LIQASTAS	-31.0100***	-0.1885*	-0.1905**
	(0.0000)	(0.0940)	(0.0390)
NETLOADEPSTFUN	1.1890***	-0.0016	
	(0.0000)	(0.9030)	
NETLOATAS	NA	-0.0008	
		(0.8690)	
DEPSTFUNTAS	1.6790***	-0.0715	-0.1137*
	(0.0000)	(0.9020)	(0.0530)
LIQASDEPSTFUN	25.1600***		
	(0.0000)		
ROAA	-21.1700***		
	(0.0000)		
ROAE	1.9640***		
	(0.0000)		
NETINTMAR	3.6150***		
	(0.0000)		
TCAPTAS	0.2698**	0.1632*	
	(0.0228)	(0.0990)	
EQAS	1.9360***		
	(0.0000)		
TIER1CAPTAS	0.1859		
	(0.1663)		
LOALOSPROLOA	-5.6620***		
	(0.0000)		
PROGRO	-0.1182***	0.0143	
	(0.0000)	(0.6290)	
OPEXPTAS	19.7500***		
	(0.0000)		
COSINC	0.0707		
	(0.2069)		
OVHTAS	NA		
MSAS	11.5500***		
	(0.0000)		
DEPLOA	NA		

[8] We report formal tests for the equality of coefficients across the sample periods in Table A.3 in the Appendix.

DEPLOAGRO	-0.1145***		
	(0.0000)		
INCDIV	-0.1053**	0.0426*	0.0556*
	(0.0188)	(0.0970)	(0.0590)
HHI3	1.306E+05***		
	(0.0000)		
HHI5	NA		
CON3	NA		
CON5	NA		-0.0171
			(0.9390)
RSP500	5.011E+04***		
	(0.0000)		
CPI	-1.589E+04***	-0.4462	
	(0.0000)	(0.1530)	
GRGDP	-2.026E+05***	0.4019*	
	(0.0000)	(0.0920)	
LNGDPCAP	NA		
GRGNP	NA		3.1577
			(0.1270)
INTR_10Y	NA		
INTR_10Y2	-2.9830		
	(0.6816)		
INTR_3M	NA		
INTR_3M2	NA		
SLYC	NA		
SLYC2	NA	-19.6895	-31.4148
		(0.1870)	(0.7370)
GRM1	-2.388E+04***		
	(0.0000)		
GRM2	-1.916E+04***		
	(0.0000)		
HPI	NA		

Notes: CPH model represents the Cox proportional hazard model. "AIC" is the AIC-type tuning parameter selector. "BIC" is the BIC-type tuning parameter selector. P-values related to z-statistics reported in the parentheses are Huber-White robust estimates, clustered at the firm level. *** denotes significance at the 1% level. ** denotes significance at the 5% level. * denotes significance at the 10% level.

Next, we find that in the L1 penalized CPH model with the AIC-type selector, LNDETAS, LOAAS, LIQASTAS, TCAPTAS, INCDIV and GRGDP are all statistically significant under post-selection from 13 surviving predictors in column 2. For the BIC-type selector counterpart, only four bank-specific variables are statistically significant among all seven selected variables, namely LNDETAS, LIQASTAS, DEPSTFUNTAS and INCDIV. The smaller set of industry-specific and macroeconomic variables can be kept after the LASSO estimator in the crisis period than those in the pre-crisis period (compared in Table 3-5 and Table 3-6), which indicates that bank-specific conditions are more important than the economic environment for banks' IPO decisions in the crisis period.

3.6.3 The post-crisis period

We now focus on the aftermath of the crisis. To begin with the analysis of the CPH model, as reported in column 1 of Table 3-7, most bank-specific variables enter with expected sign and retain their significance. However, the coefficients are significantly smaller compared to their counterparts in the crisis period. Regarding the industry-specific and macroeconomic variables, they are both statistically significant and the absolute value of the above-mentioned variables in the post-crisis period is lower than in the crisis period. This indicates that while bank information and market conditions are important in affecting the probability of an IPO, they are less important than during the crisis period.

Table 3-7 The estimates of candidate models in the post-crisis period

Variable	CPH model	Penalized CPH model_AIC	Penalized CPH model_BIC
	(1)	(2)	(3)
LNDETAS	6.3560***		-0.0007
	(0.0002)		(0.3240)
LNDETAS2	-3.1730***		
	(0.0002)		
GROAS	0.0002		
	(0.9897)		
LOAAS	-0.3259***	-0.0143***	
	(0.0065)	(0.0000)	
LIQASTAS	-2.3900***		
	(0.0000)		
NETLOADEPSTFUN	0.2474***	-1.062E-07	
	(0.0078)	(1.0000)	
NETLOATAS	NA		
DEPSTFUNTAS	0.4191***	-0.0279	-0.0369*
	(0.0002)	(1.0000)	(0.0640)
LIQASDEPSTFUN	2.1610***	0.0229	
	(0.0000)	(1.0000)	
ROAA	1.5470***		
	(0.0002)		
ROAE	-0.2509***		
	(0.0000)		
NETINTMAR	0.8035**		
	(0.0367)		
TCAPTAS	0.9610***	0.0325***	
	(0.0000)	(0.0000)	
EQAS	-0.3103***	-0.0078	
	(0.0026)	(1.0000)	
TIER1CAPTAS	-0.6686***		
	(0.0003)		
LOALOSPROLOA	-2.8330***		

	(0.0005)		
PROGRO	-0.0314***	-0.0120***	
	(0.0000)	(0.0000)	
OPEXPTAS	-2.1830		
	(0.2147)		
COSINC	0.0125		
	(0.4718)		
OVHTAS	NA		
MSAS	-6.2670***	-0.4111	
	(0.0000)	(1.0000)	
DEPLOA	-0.0193*	-0.0095***	-0.0023
	(0.0755)	(0.0000)	(0.5630)
DEPLOAGRO	-0.0525**		
	(0.0147)		
INCDIV	0.0497***		
	(0.0094)		
HHI3	19.9700		
	(0.4986)		
HHI5	74.1200***		
	(0.0001)		
CON3	-10.7400**		-0.0442
	(0.0484)		(0.5500)
CON5	-17.7200***	-0.1094***	-0.0600
	(0.0000)	(0.0000)	(0.4740)
RSP500	0.3170***		
	(0.0008)		
CPI	0.3504	-0.1486***	
	(0.2414)	(0.0010)	
GRGDP	-4.7710***		
	(0.0008)		
LNGDPCAP	-78.1600**	-0.7421***	-0.8477
	(0.0235)	(0.0000)	(0.7980)
GRGNP	19.6100***		
	(0.0008)		
INTR_10Y	0.0011		0.0950**
	(0.9998)		(0.0450)
INTR_10Y2	-514.5000		
	(0.1355)		
INTR_3M	-26.4500***	-0.3214	
	(0.0092)	(0.0000)	
INTR_3M2	3.091E+03***	49.3428***	
	(0.0002)	(0.0000)	
SLYC	NA		
SLYC2	465.0000	3.1383	
	(0.1384)	(1.0000)	
GRM1	-1.4570***	-0.0738***	
	(0.0002)	(0.0000)	
GRM2	2.3200***	-0.0242	-0.1523*
	(0.0016)	(1.0000)	(0.0910)
HPI	-1.1830**	0.0120	
	(0.0359)	(0.9960)	

Notes: CPH model represents the Cox proportional hazard model. "AIC" is the AIC-type tuning parameter selector. "BIC" is the BIC-type tuning parameter selector. P-values related to z-statistics reported in the parentheses are Huber-White robust estimates, clustered at the firm level. *** denotes significance at the 1% level. ** denotes significance at the 5% level. * denotes significance at the 10% level.

Moving to the analysis of penalized models, in the L1 penalized CPH model with the AIC-type selector, as shown in column 2, nine indicators of 18 surviving variables are statistically significant under post-selection. On the other hand, only three variables are statistically significant under post-selection from 8 selected variables in the penalized model with BIC-type selector. It should be noted that the number of selected industry-specific and macroeconomic predictors increases after financial crisis compared Table 3-6 and Table 3-7. Comparing the magnitudes of the selected variables, we find, once again, that they are higher in the crisis period than in the post-crisis period.

3.7 Conclusion

The decision of a bank to go public by issuing an IPO is an important operational threshold event, which can lead to various investment and development plans for market participants. This chapter uses quarterly data for US banks as original input in benchmark models and all competing models. Our study, as far as we know, is the first to apply an innovative methodology to analyse the timing of a bank's decision to issue for the first time in the public market.

We find that several bank-specific financial factors, market-driven and macroeconomic variables are important in predicting the decision of banks to go public. In terms of the models' predictive ability, when we apply the LASSO estimator in a Cox proportional hazard model, we note a significant improvement in predicting a bank's IPO. The L1 penalized semi-parametric Cox proportional hazard model provides the most accurate out-of-sample prediction among all candidate models. On the other hand, we show that the Cox proportional hazard model underperforms discrete hazard and logistic models, which highlights the effect of LASSO on our algorithms. Our L1 penalized models are tuned through the AIC and the BIC criteria. We observe increased predictability on our dataset when the latter criterion is applied. Finally, when we split our sample into crisis and non-crisis periods, we find that bank-specific and macro variables become more potent in determining banks' IPOs, which signifies the ability of banks to time their IPOs relative to the economic conditions.

References

Acharya, V. V. and Mora, N. (2015) 'A Crisis of Banks as Liquidity Providers', *The Journal of Finance*, 70(1), 1-43.

Adjei, F., Cyree, K. B. and Walker, M. M. (2008) 'The Determinants and Survival of Reverse Mergers vs IPOs', *Journal of Economics and Finance*, 32(2), 176-194.

Akaike, H. (1974) 'A New Look at the Statistical Model Identification', *IEEE Transactions on Automatic Control*, 19(6), 716-723.

Allison, P. D. (1982) 'Discrete-Time Methods for the Analysis of Event Histories', *Sociological Methodology*, 13, 61-98.

Amihud, Y. and Mendelson, H. (1988) 'Liquidity and Asset Prices: Financial Management Implications', *Financial Management*, 17(1), 5-15.

Bekaert, G., Ehrmann, M., Fratzscher, M. and Mehl, A. (2014) 'The Global Crisis and Equity Market Contagion', *The Journal of Finance*, 69(6), 2597-2649.

Bharath, S. T. and Dittmar, A. K. (2010) 'Why Do Firms Use Private Equity to Opt Out of Public Markets?', *The Review of Financial Studies*, 23(5), 1771-1818.

Bharath, S. T. and Shumway, T. (2008) 'Forecasting Default with the Merton Distance to Default Model', *The Review of Financial Studies*, 21(3), 1339-1369.

Bodnaruk, A., Kandel, E., Massa, M. and Simonov, A. (2008) 'Shareholder Diversification and the Decision to Go Public', *The Review of Financial Studies*, 21(6), 2779-2824.

Boehmer, E. and Ljungqvist, A. (2004) 'On the Decision to Go Public: Evidence from Privately-held Firms', *SSRN Journal SSRN Electronic Journal*.

Boyd, J. H. and Heitz, A. (2016) 'The Social Costs and Benefits of Too-Big-To-Fail Banks: A “Bounding” Exercise', *Journal of Banking and Finance*, 68, 251-265.

Brau, J. C. and Fawcett, S. E. (2006) 'Initial Public Offerings: An Analysis of Theory and Practice', *The Journal of Finance*, 61(1), 399-436.

Brau, J. C., Francis, B. and Kohers, N. (2003) 'The Choice of IPO versus Takeover: Empirical Evidence', *The Journal of Business*, 76(4), 583.

Campbell, T. S. (1979) 'Optimal Investment Financing Decisions and the Value of Confidentiality', *Journal of Financial and Quantitative Analysis*, 14(5), 913-924.

Chava, S. and Jarrow, R. A. (2004) 'Bankruptcy Prediction with Industry Effects', *European Finance Review*, 8(4), 537-569.

Chemmanur, T. J., He, S. and Nandy, D. K. (2010) 'The Going-Public Decision and the Product Market', *The Review of Financial Studies*, 23(5), 1855-1908.

Cox, D. R. (1972) 'Regression Models and Life-Tables', *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2), 187-220.

Delis, M. D., Molyneux, P. and Pasiouras, F. (2011) 'Regulations and Productivity Growth in Banking: Evidence from Transition Economies', *Journal of Money, Credit and Banking*, 43(4), 735-764.

Duffie, D., Saita, L. and Wang, K. (2007) 'Multi-period Corporate Default Prediction with Stochastic Covariates', *Journal of Financial Economics*, 83(3), 635-665.

Dungey, M. and Gajurel, D. (2015) 'Contagion and Banking Crisis - International Evidence for 2007–2009', *Journal of Banking and Finance*, 60, 271-283.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) 'Least Angle Regression', *The Annals of Statistics*, 32(2), 407-451.

Fan, J. and Li, R. (2001) 'Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties', *Journal of the American Statistical Association*, 96(456), 1348-1360.

Fischer, C. N. M. I. (2000) 'Why Do Companies Go Public? Empirical Evidence from Germany's Neuer Markt', *SSRN Journal SSRN Electronic Journal*.

Francis, B., Hasan, I. and Siregar, D. (2009) 'The Choice of IPO versus M&A: Evidence from Banking Industry', *Applied Financial Economics*, 19(24), 1987-2007.

Geyfman, V. (2014) 'The Effect of Economic and Financial System Development on Banks' Listing Decisions: Evidence from Transition Economies', *Emerging Markets Finance and Trade*, 50(6), 174-192.

Gill de Albornoz, B. and Pope, P. F. (2004) *The Determinants of the Going Public Decision: Evidence from The U.K.,* Instituto Valenciano de Investigaciones Económicas, S.A. (Ivie), Available from: https://EconPapers.repec.org/RePEc:ivi:wpasad:2004-22 [Accessed. 17th May, 2016]

Grilli, L. (2005) 'The Random-effects Proportional Hazards Model with Grouped Survival Data: A Comparison between the Grouped Continuous and Continuation Ratio Versions', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1), 83-94.

Grullon, G., Larkin, Y. and Michaely, R. (2015) 'The Disappearance of Public Firms and the Changing Nature of U.S. Industries', *SSRN Electronic Journal*.

Guo, H. and Brooks, R. (2009) 'Duration of IPOs between offering and listing: Cox proportional hazard models—Evidence for Chinese A-share IPOs', *International Review of Financial Analysis*, 18(5), 239-249.

Harris, M. and Raviv, A. (2014) 'How to Get Banks to Take Less Risk and Disclose Bad News', *Journal of Financial Intermediation*, 23(4), 437-470.

Helwege, J. and Liang, N. (2004) 'Initial Public Offerings in Hot and Cold Markets', *Journal of Financial and Quantitative Analysis*, 39(3), 541-569.

Helwege, J. and Packer, F. (2003) *The Decision to Go Public: Evidence from Mandatory SEC filings of Private Firms*. Columbus, Ohio, Charles A. Dice Centre for Research in Financial Economics, Available from: http://www.cob.ohio-state.edu/fin/dice/papers/2003/2003-18.pdf. [Accessed. 17th March, 2016]

Houge, T. and Loughran, T. (1999) 'Growth fixation and the performance of bank initial public offerings, 1983–1991', *Journal of Banking and Finance*, 23(8), 1277-1301.

Hsieh, J., Lyandres, E. and Zhdanov, A. (2011) 'A Theory of Merger-Driven IPOs', *Journal of Financial and Quantitative Analysis*, 46(5), 1367-1405.

Jenkins, S. (2005) 'Survival Analysis', Unpublished Lecture Notes manuscript.

Kashian, R. D. and Ahmad, Y. S. (2010) 'Modeling the Time to an Initial Public Offering: When Does the Fruit Ripen?', *Journal of Economics and Finance*, 34(4), 391-414.

Kim, W. and Weisbach, M. S. (2008) 'Motivations for Public Equity Offers: An International Perspective', *Journal of Financial Economics*, 87(2), 281-307.

Lowry, M. (2003) 'Why Does IPO Volume Fluctuate So Much?', *Journal of Financial Economics*, 67(1), 3-40.

Lowry, M., Michaely, R. and Volkova, E. (2017) 'Initial Public Offerings: A Synthesis of the Literature and Directions for Future Research', *Foundations and Trends in Finance*, 11(3-4), 154-320.

Maksimovic, V. and Pichler, P. (2001) 'Technological Innovation and Initial Public Offerings', *The Review of Financial Studies*, 14(2), 459-494.

Mayur, M. and Kumar, M. (2013) 'Determinants of Going-Public Decision in an Emerging Market: Evidence from India', *Vikalpa*, 38(1), 65-86.

Modigliani, F. and Miller, M. H. (1963) 'Corporate Income Taxes and the Cost of Capital: A Correction', *The American Economic Review*, 53(3), 433-443.

Pagano, M. (1993) 'The Flotation of Companies on the Stock Market: A Coordination Failure Model', *European Economic Review*, 37(5), 1101-1125.

Pagano, M., Panetta, F. and Zingales, L. (1998) 'Why Do Companies Go Public? An Empirical Analysis', *The Journal of Finance*, 53(1), 27-64.

Pagano, M. and Roell, A. (1998) 'The Choice of Stock Ownership Structure: Agency Costs, Monitoring, and the Decision to go Public', *The Quarterly Journal of Economics*, 113(1), 187-225.

Pástor, L. u. and Veronesi, P. (2005) 'Rational IPO Waves', *The Journal of Finance*, 60(4), 1713-1757.

Rabe-Hesketh, S. and Skrondal, A. (2012) *Multilevel and Longitudinal Modeling Using Stata*, College Station, Tex.: Stata Press.

Ramcharan, R., Verani, S. and Van Den Heuvel, S. J. (2016) 'From Wall Street to Main Street: The Impact of the Financial Crisis on Consumer Credit Supply', *The Journal of Finance*, 71(3), 1323-1356.

Ritter, J. R. (1987) 'The Costs of Going Public', *Journal of Financial Economics*, 19(2), 269-281.

Rock, K. (1986) 'Why New Issues Are Underpriced', *Journal of Financial Economics*, 15(1), 187-212.

Röell, A. (1996) 'The Decision to Go Public: An Overview', *European Economic Review*, 40(3), 1071-1081.

Samet, A., Boubakri, N. and Boubaker, S. (2018) 'Does Public–private Status Affect Bank Risk Taking? Worldwide Evidence', *Journal of International Financial Markets, Institutions & Money*, 53, 287-306.

Schwarz, G. (1978) 'Estimating the Dimension of a Model', *The Annals of Statistics*, 6(2), 461-464.

Shumway, T. (2001) 'Forecasting Bankruptcy More Accurately: A Simple Hazard Model', *The Journal of Business*, 74(1), 101-124.

Subrahmanyam, A. and Titman, S. (1999) 'The Going-Public Decision and the Development of Financial Markets', *The Journal of Finance*, 54(3), 1045-1082.

Sun, T. and Zhang, C.-H. (2012) 'Scaled Sparse Linear Regression', *Biometrika*, 99(4), 879-898.

Tian, S., Yu, Y. and Guo, H. (2015) 'Variable Selection and Corporate Bankruptcy Forecasts', *Journal of Banking and Finance*, 52, 89-100.

Tibshirani, R. (1996) 'Regression Shrinkage and Selection via the Lasso', *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288.

Tibshirani, R. (1997) 'The Lasso Method for Variable Selection in the Cox Model', *Statistics in Medicine*, 16(4), 385-395.

Traczynski, J. (2017) 'Firm Default Prediction: A Bayesian Model-Averaging Approach', *Journal of Financial and Quantitative Analysis*, 52(3), 1211-1245.

Tregenna, F. (2009) 'The Fat Years: The Structure and Profitability of the US Banking Sector in the Pre-Crisis Period', *Cambridge Journal of Economics*, 33(4), 609-632.

van de Geer, S. A. (2008) 'High-dimensional Generalized Linear Models and the Lasso', *The Annals of Statistics.*, 36(6), 614.

Zingales, L. (1995) 'Insider Ownership and the Decision to Go Public', *The Review of Economic Studies*, 62(3), 425-448.

Zou, H. (2006) 'The Adaptive Lasso and Its Oracle Properties', *Journal of the American Statistical Association*, 101(476), 1418-1429.

Zou, H., Hastie, T. and Tibshirani, R. (2007) 'On the "Degrees of Freedom" of the Lasso', *The Annals of Statistics*, 35(5), 2173-2192.

Appendix

Section A

Table A.1 Variables definition and expected relationship

Variable	Definition	Label	Expected sign
Bank-specific (24)			
Size	The logarithm of total real assets	LNDETAS	+
	The logarithm of the square of real total assets	LNDETAS2	~
	The rate of growth of real assets	GROAS	+
Liquidity	Loans / assets	LOAAS	+
	Liquid asset / total assets	LIQASTAS	-
	Net loans / deposits and short-term funding	NETLOADEPSTFUN	+
	Net loans / total assets	NETLOATAS	+
	Deposits and short-term funding / total assets	DEPSTFUNTAS	-
	Liquid assets / deposits & short-term funding	LIQASDEPSTFUN	-
Profitability	The average return on equity	ROAA	-
	The average return on assets	ROAE	-
	Net interest margin	NETINTMAR	-
Capital	Capital to assets ratio	TCAPTAS	~
	Equity / assets	EQAS	~
Leverage	Tier 1 ratio	TIER1CAPTAS	+
Credit risk	Loan loss provisions / loans	LOALOSPROLOA	+
Productivity growth	Rate of change in inflation-adjusted gross total revenue / the number of employees	PROGRO	-
Operating expenses management	Operating expenses / total assets	OPEXPTAS	+
	Operating costs / Operating income ratio	COSINC	+
	Overheads to total assets	OVHTAS	+
Market share	Market share (in terms of assets) of individual banks	MSAS	~
Deposit	Total deposits / total loans	DEPLOA	~
	The growth rate of deposits	DEPLOAGRO	~
Income diversification	Non-interest income to total operating revenue	INCDIV	-
Industry-specific (4)			
Concentration	The three-firm Herfindahl-Hirschman index	HHI3	-
	The five-firm Herfindahl-Hirschman index	HHI5	-
	The assets of the three largest banks / the assets of all banks in the same dataset	CON3	-

	The assets of the five largest banks / the assets of all banks in the same dataset	CON5	-
Macroeconomic (14)			
Stock market performance	The return of S&P500	RSP500	+
Inflation rate	Current period inflation	CPI	~
GDP growth rate	The real gross domestic product (GDP) growth rate	GRGDP	~
GDP per capita	The logarithm of GDP per capita	LNGDPCAP	~
GNP growth rate	The GNP growth rate	GRGNP	~
Interest rate	10-year government bond yield	INTR_10Y	~
	The square of 10-year government bond yield	INTR_10Y2	~
	3-month interbank rate	INTR_3M	~
	The square of 3-month interbank rate	INTR_3M2	~
Slope of the yield curve	The difference between the 10-year government bond yield and the three-month interbank rate	SLYC	~
	The square of the abovementioned yield curve	SLYC2	~
Market growth	The growth rate in money supply (M1)	GRM1	~
	The growth rate in money supply (M2)	GRM2	~
House price growth rate	All-Transactions House Price Index for the United States	HPI	+

Notes: "+" indicates that the probability of a bank going public would improve if the covariates rose. "-" indicates that the probability of a bank going public would reduce if the covariates rose. "~" indicates uncertainty in the sign.

Table A.2 Cross-Correlations

Variable	LNDETAS	LNDETAS2	GROAS	LOAAS	LIQASTAS	NETLOADEPSTFUN	NETLOATAS	DEPSTFUNTAS	LIQASDEPSTFUN	ROAA	ROAE	NETINTMAR	TCAPTAS	EQAS	TIER1CAPTAS	LOALOSPROLOA	PROGRO	OPEXPTAS	COSINC	OVHTAS	MSAS	DEPLOA	DEPLOAGRO	INCDIV
LNDETAS	1.00																							
LNDETAS2	1.00	1.00																						
GROAS	-0.06	-0.06	1.00																					
LOAAS	-0.16	-0.16	0.03	1.00																				
LIQASTAS	0.07	0.07	0.16	-0.19	1.00																			
NETLOADEPSTFUN	-0.04	-0.04	0.01	0.92	-0.20	1.00																		
NETLOATAS	-0.16	-0.16	0.03	1.00	-0.19	0.92	1.00																	
DEPSTFUNTAS	-0.34	-0.34	0.05	-0.01	0.06	-0.39	-0.01	1.00																
LIQASDEPSTFUN	0.09	0.09	0.15	-0.20	0.99	-0.16	-0.20	-0.06	1.00															
ROAA	0.04	0.04	-0.11	-0.13	-0.09	-0.14	-0.13	0.07	-0.10	1.00														
ROAE	0.04	0.04	-0.04	-0.06	-0.07	-0.11	-0.06	0.16	-0.08	0.87	1.00													
NETINTMAR	-0.35	-0.35	0.00	0.03	-0.11	-0.05	0.03	0.23	-0.13	0.42	0.31	1.00												
TCAPTAS	-0.11	-0.11	-0.05	-0.05	0.05	0.07	-0.05	-0.35	0.08	0.00	-0.26	0.06	1.00											
EQAS	0.07	0.07	-0.05	-0.17	0.02	-0.02	-0.17	-0.40	0.06	0.01	-0.31	0.09	0.83	1.00										
TIER1CAPTAS	-0.22	-0.22	-0.04	-0.09	0.04	0.02	-0.09	-0.30	0.07	0.00	-0.26	0.09	0.97	0.83	1.00									
LOALOSPROLOA	0.13	0.13	-0.02	0.07	0.00	0.10	0.07	-0.12	0.01	-0.52	-0.52	-0.10	0.04	0.01	0.00	1.00								
PROGRO	0.00	0.00	0.07	0.02	0.02	0.02	0.02	-0.01	0.02	0.03	0.04	0.01	0.00	0.02	0.00	0.07	1.00							
OPEXPTAS	-0.02	-0.02	-0.02	-0.02	0.15	-0.01	-0.02	-0.01	0.17	-0.12	-0.09	0.35	-0.01	-0.03	-0.03	0.05	0.02	1.00						
COSINC	-0.11	-0.11	0.12	0.02	0.16	0.02	0.02	-0.01	0.17	-0.74	-0.63	-0.19	0.03	0.01	0.03	0.12	-0.04	0.56	1.00					
OVHTAS	-0.02	-0.02	-0.02	-0.02	0.15	-0.01	-0.02	-0.01	0.17	-0.12	-0.09	0.35	-0.01	-0.03	-0.03	0.05	0.02	1.00	0.56	1.00				
MSAS	0.73	0.73	-0.04	-0.13	0.06	-0.04	-0.13	-0.23	0.07	0.10	0.09	-0.20	-0.09	0.04	-0.19	0.05	0.03	0.05	-0.10	0.05	1.00			
DEPLOA	0.00	0.00	0.01	-0.88	0.20	-0.92	-0.88	0.31	0.18	0.09	0.06	0.03	-0.04	0.04	0.01	-0.07	-0.01	0.00	0.02	0.00	-0.02	1.00		
DEPLOAGRO	0.03	0.03	0.31	-0.02	0.20	-0.02	-0.02	0.01	0.20	-0.04	-0.04	-0.05	-0.04	-0.06	-0.04	0.11	-0.02	-0.04	-0.02	-0.04	-0.01	0.05	1.00	
INCDIV	0.51	0.51	-0.06	-0.13	0.26	-0.04	-0.13	-0.22	0.29	0.11	0.15	-0.26	-0.09	-0.10	-0.17	0.07	0.02	0.52	0.09	0.52	0.44	0.00	0.04	1.00

Note: All bank-specific variables are as defined in Table A.1. The number in each cell indicates the correlation between the row and column variables.

Table A.3 Tests of equality of estimated coefficients

	CPH in the pre-crisis period	CPH in the crisis period	CPH in the post-crisis period
	(1)	(2)	(3)
LNDETAS	0.0000	0.0000	0.0312
LNDETAS2	NA	NA	NA
GROAS	0.0000	0.0000	0.0000
LOAAS	0.2648	0.1121	0.4351
LIQASTAS	0.0000	0.0000	0.0000
NETLOADEPSTFUN	0.2005	0.2395	0.0842
NETLOATAS	NA	NA	NA
DEPSTFUNTAS	0.0004	0.0001	0.0230
LIQASDEPSTFUN	0.0000	0.0000	0.0000
ROAA	0.0039	0.0009	0.0011
ROAE	0.0000	0.0000	0.0000
NETINTMAR	0.0319	0.0735	0.4098
TCAPTAS	0.0000	0.0030	0.0000
EQAS	0.0002	0.0008	0.0000
TIER1CAPTAS	0.0000	0.0000	0.0000
LOALOSPROLOA	0.0483	0.0654	0.0215
PROGRO	0.0000	0.0000	0.0000
OPEXPTAS	0.2972	0.7541	0.3845
COSINC	0.6639	0.3804	0.3694
OVHTAS	NA	NA	NA
MSAS	0.1303	0.1434	0.0479
DEPLOA	NA	NA	NA
DEPLOAGRO	0.1029	0.0336	0.0461
INCDIV	0.2599	0.8992	0.3605
HHI3	0.0000	0.0000	0.0000
HHI5	NA	NA	NA
CON3	NA	NA	NA
CON5	NA	NA	NA
RSP500	0.0016	0.0332	0.0006
CPI	0.0000	0.0000	0.0000
GRGDP	0.0006	0.3827	0.0003
LNGDPCAP	NA	NA	NA
GRGNP	NA	NA	NA
INTR_10Y	NA	NA	NA
INTR_10Y2	0.2820	0.3028	0.2114
INTR_3M	NA	NA	NA
INTR_3M2	NA	NA	NA
SLYC	NA	NA	NA
SLYC2	NA	NA	NA
GRM1	0.0070	0.0000	0.0017
GRM2	0.0265	0.0871	0.0130
HPI	NA	NA	NA

Notes: Column (1) refers to the test of the coefficient equality for one variable in three sub-periods. Column (2), reports the test of the coefficient equality for one variable between pre-crisis and crisis periods. Column (3) shows the test of the coefficient equality of one variable between crisis and post-crisis periods.

Section B

Accuracy ratios

AUC is a non-parametric measure generated from the receiver operating characteristic curve, which is commonly employed to assess the ability of a model to discriminate between binary events. It is already applied in related studies to evaluate the predictive ability to identify a default event. The receiver operating characteristic curve is the plot of the likelihood of verifying true-positive (in practice, a bank issues IPO and the model classifies it as an expected event) and false-positive (in practice, a bank issues IPO but the model classifies it as an expected non-event) for a whole range of probable threshold points of probability values. If AUC is equal to 1, this represents a perfect prediction. If AUC is equal to or less than 0.5, it means that the corresponding model had no predictability. If the value of AUC is above 0.8, the predictive ability may be considered to be accurate (Hosmer Jr et al. 2013). The accuracy ratio is defined as the double difference between the value of AUC and 0.5, which is a frequently applied measure for corporate bankruptcy model evaluation. Thus, a value of 1 for accuracy ratio illustrates a perfect forecast, while a value of 0 for this shows a random forecast. To confirm the conclusions from AUC and the accuracy ratio, the brier score is included, to measure how close the predicted probability of a bank issuing IPOs in order to go public is to a bank staying in the private market. It is equal to the average of the squared differences between the forecast probabilities and the actual outcomes (1 if a bank issues IPO and 0 if a bank does not issue it). The brier score can be expressed as $\frac{1}{N}\int_{t=1}^{N}(p_t - o_t)^2$, where p_t is the forecast probability of a bank issuing IPO and o_t is the corresponding actual event. The lower the brier score is for a series of predictions, the better the predictions are deemed to be.

Chapter 4 Predicting failure: evidence from UK firms

Abstract

The accurate prediction of failure for a private firm can be considered an early signal which enables market participants to manage their credit risks and reduce potential loss. Due to the importance of failure prediction, firm-specific and macroeconomic indicators are implemented in this empirical exercise. We observed that adding Bayesian Model Averaging (BMA) can improve the predictive performance of a discrete hazard model in out-of-sample predictions of firms' failure. In addition, to confirm the accuracy of simple classifiers in failure prediction, two classifiers (the Naive Bayes classifier and the k-nearest neighbours classifier) are implemented compared to hazard models. According to the results, the Naive Bayes classifier frequently outperforms other models in failure prediction. Moreover, the predictive power of each candidate model can be influenced by financial crisis or firm heterogeneity in the out-of-sample prediction of failure.

Key words: Failure prediction, UK firms, Financial ratios, Bayesian Model Averaging, Classifier

4.1 Introduction

It is well accepted that the failure of a firm can bring with the unimaginable loss of wealth, in some cases potentially constituting a financial crisis. The timely detection and accurate prediction of a firm's failure are essential for the actions of market participants. Firms' managers, as insiders, can use reliable and efficient failure prediction of their firm in their internal performance evaluations to check management performance, and they can also use this failure prediction as an early warning mechanism, implementing remedial actions within their companies (Geng *et al.* 2015). Moreover, accurate failure prediction for firms can lower the probability of firms' outsiders (for example investors and creditors) being exposed to default risks and losses. Such prediction can also encourage policymakers to announce new regulations or policies adapted to market changes which can stabilize the financial market. As the Gross Domestic Product (GDP) growth rate dropped to below zero in 2008, the UK economy entered recession, which implies that more UK firms were likely to face higher default risks. Lang and Stulz (1992) indicate that the failure of a firm has a contagion effect on other firms with similar cash flow features. Hence, market participants can utilize the correct failure prediction of companies to seek to address the critical situation. This will prevent more similar companies from facing insolvency in the future. Thus, providing accurate failure prediction regarding firms has once again become important.

According to a comprehensive review of bankruptcy prediction techniques (Ravi Kumar and Ravi 2007), numerous studies have been dedicated to the study of failure prediction for companies. Such studies apply firm-specific financial ratios and other publicly available information in reduced-form models, structure models and machine learning models. The methodologies used in this chapter are in line with the literature on failure prediction for companies. The objective for this work is to detect a model which provides more accurate predictions of private firms' failure in the UK, incorporating several time-varying covariates from a set of firm-specific and macro-economic predictors. In this chapter we aim to make several contributions to existing literature. First, a Bayesian technique, namely Bayesian Model Averaging (BMA), is added to the discrete hazard model in order to solve the parameter and model uncertainty in a straightforward and formal way. In reduced-form models, variable selection depends on the background of researchers, which may not include the "true" model which reflects reality.

Ignoring parameter and model uncertainty can lead to overconfidence in predictions and reduce out-of-sample predictive ability. BMA is an advanced version of Bayesian inference, which directly allows for model combination, combined parameter estimation and prediction (Roberts 1965) with reduced overconfidence in the parameter estimation, fewer omissions of explanatory variables and increased applicability. Meanwhile, Madigan and Raftery (1994) confirm that BMA can produce more reliable results than a single model if scoring rules are employed to validate predictions. Thus, to validate the efficiency and performance of BMA in failure prediction, a combined model, known as the BMA version of discrete hazard models, was implemented in comparison with the benchmark model, the discrete hazard (DH) model. This benchmark model contains the timing of non-interested events compared to simple logit or probit models.

Second, the basic classifiers, namely the Naive Bayes (NB) classifier and the k-nearest neighbours (k-NN) classifier, are also applied as competing models in this work. We aim to test the accuracy of these simple classifiers in corporate failure prediction by comparing with the DH model and its BMA version. The NB classifier is also developed from Bayes' theorem, like BMA, and the k-NN classifier is exploited from pattern recognition. The studies of Henley and Hand (1996) and Sarkar and Sriram (2001) indicate that both models can produce reliable failure predictions of companies. As for the NB classifier, it can successfully handle missing values and irrelevant predictors in datasets and produce the minimum error rate during estimation compared with other classifiers such as decision trees and neural network classifiers (Han *et al.* 2011). Furthermore, the computation of the NB classifier is simple due to its class conditional independence assumption and it is still able to perform without satisfying this assumption (Domingos and Pazzani 1997). As for k-NN classifier, it is a non-parametric method in machine learning models, which can easily be implemented. There are no strict assumptions in the k-NN classifier (Murphy 2012) and hence this classifier is more flexible. Moreover, the k-NN classifier is robust when it comes to a noisy dataset and efficient in large datasets (Kuramochi and Karypis 2005).

The dataset (1991-2009) covers two important financial crises in the UK's economic history, the 1991–1993 European Exchange Rate Mechanism (ERM) currency crisis and the 2008-2009 global financial crisis (GFC). The causes and scope of these two financial crises are different, which may affect the predictive performance of candidate models in different ways. Therefore, to investigate the influence of these financial crises on failure prediction, our dataset is separated into three sub-samples: the post-ERM currency crisis period, the pre-global crisis period and the non-crisis period.

Meanwhile, this dataset is composed mostly of unlisted UK firms with different economic scales operating in several industries. These companies are generally younger and smaller than listed counterparts, which implies that asymmetric information can easily be observed. In other words, compared to public firms, unlisted firms are more likely to face financial constraints and suffer higher liquidity risks because of more asymmetric information. Due to the existence of asymmetric information, outsiders to private firms have difficulty accessing available information in order to evaluate a firm's operation, which suggests that it might be more difficult to predict the failure of private firms than that of public firms. Beck *et al.* (2005) and Ayyagari *et al.* (2007) suggest that the operation of private firms, in particular small and medium enterprises, can influence employment and economic development in a country. According to the special role of private firms in the market, our work on developing appropriate credit risk models to produce correct failure prediction of these firms is very beneficial. This study can help to close the existing gap in the literature concerning failure prediction, where evaluations of UK private firms and of the two major financial crises in UK's recent history are still lacking.

Furthermore, since our panel data contains firm heterogeneity, it would be expected that the various internal and external causes of a firm's failure can be observed (Ropega 2011). Ultimately, the probability of small (young) firms failing will be distinct from that of old (large) firms. Therefore, we choose two cross-sectional dimensions to split our dataset. With respect to firm size, the dataset is separated into small and large firms. The dataset is then divided into old and young firms based on their age. To further analyse whether the position of a firm can lead to failure, time or industry dummy variables are included, in line with the studies of Chava and Jarrow (2004), Hillegeist *et al.* (2004) and Traczynski (2017).

In general, our results show that adding BMA into the DH model can improve the predictive accuracy in out-of-sample failure prediction compared to the DH model. However, when comparing all candidate models, the NB classifier outperforms the others in failure predictions in most samples. The BMA version of DH models is the only model that takes parameter and model uncertainty into account, which means it could provide comparable predictive ability compared to the NB classifier in certain situations. Our results also indicate that the two financial crises and firm heterogeneity can influence the predictive power of each candidate model. Controlling time effects or industry effects increases the out-of-sample prediction performance in different sub-samples, especially for the benchmark model and its BMA version.

The rest of the work is organized as follows. We document the relevant literature on failure prediction for firms in section 2. Following that, data and summary statistics are discussed in section 3. Section 4 introduces in detail the methodologies used. In Section 5 the empirical results are reported, with section 6 concluding this work.

4.2 Literature

Before applying statistical methodologies to identify a company's failure, since the 1930s researchers have been studying the difference in financial ratios between "successful" and "failed" firms (FitzPatrick 1932, Smith and Winakor 1935). This provides the foundation for later researchers using financial ratios or other information to determine and forecast the failure of a firm.

The use of accounting information to predict bankruptcy is derived from Beaver (1966). Altman (1968) extended this analysis to multiple discriminant analysis and forecasts the bankruptcy of a firm based on several accounting-based financial ratios. This method is called Z-score analysis and became the classical model to group failed and non-failed companies. The predictive accuracy of the Z-score model[1] is more accurate than Beaver's study (1966). The Z-score model is modified to the Zeta credit risk model, providing more accurate prediction of bankruptcy

[1] In conclusion of this study, it is recommended that five ratios be used in the Z-score model: working capital to total assets, retained earnings to total assets, earnings before interest and taxes to total assets, market value equity to book value of total liabilities, and sales to total assets.

up to five years prior to failure (Altman *et al*. 1977). Edmister (1972) applied 19 financial ratios and transforms them into zero-one in the step-wise multivariate discriminant analysis to reduce the biased estimation by correlation between variables and then predict the failure of a firm. For the same purpose of lowering the influence of correlation in discriminant analysis, Libby (1975) employed principal components analysis to generate five variables from fourteen financial factors to predict the failure of a company, and shows the usefulness of accounting information in the prediction of companies' failure. Scott (1981) compared existing research (see for instance Beaver 1966, Altman 1968, Deakin 1972 and Wilcox 1973) and concluded that these works can be applied empirically and explained theoretically for bankruptcy prediction.

In comparison with the ordinal output of discriminant analysis, Ohlson (1980) combined the logit model with different financial ratios to develop the O-score model and achieved the default probability of a firm. In this work, he concluded that the predictive power of this model was robust for large sample estimation and suggested that adding variables to the model could increase the predictive power of this model to a certain degree. Zmijewski (1984) utilised the probit model to examine the biased estimation from non-random samples. Zavgren (1985) applied seven financial ratios to both logit and probit models to calculate bankruptcy probability and then indicated that these models more accurately identify the financial risk than dichotomous classification from a discriminant analysis. Due to the simple application and easy explanation, logit or probit models are generally used to predict the probability of bankruptcy (see Keasey and McGuinness 1990, Tennyson *et al*. 1990, Kolari *et al*. 2002, Jones and Hensher 2004, Canbas *et al*. 2005).

These studies confirm that accounting information from the balance sheets of a company can be applied in one model to predict the failure probability of the corresponding company. It has long been known that the balance sheet reflects the previous information of a firm and may not give enough support to show its current performance. Thus, market-driven variables drawn from stock price as a complement of accounting information are gradually included in the reduced-form models to improve the accuracy of the out-of-sample test. Queen and Roll (1987) mentioned a dynamic forecasting model to address the bias in the static model solely dependent on market information to predict the bankruptcy of a firm, and they illustrated the importance of market information in bankruptcy prediction. Dichev (1998) and Griffin and Lemmon (2002) confirmed the correlation between stock market information and bankruptcy risk of insolvent and risky companies by using the Z-score model (Altman 1968) and conditional logit model (Ohlson 1980).

To produce more consistent estimates and more efficient out-of-sample prediction, hazard models are widely used in the study of forecasting bankruptcy, since these can capture the timing of alternative outcomes in the work. Shumway (2001) employed market-driven variables and accounting variables extracted from previous studies and confirmed that the predictive ability of discrete hazard outperforms discriminant analysis and logit models. Since a company is faced with various competitive pressures from peers in an industry, Chava and Jarrow (2004) investigated the importance of industry effects in the prediction of bankruptcy, by adding industry effects to Shumway's model. In this study, a relatively higher forecasting accuracy of bankruptcy is achieved than in previous studies. Despite industry effects, Pesaran *et al.* (2006) described the existence of a fundamental relationship between the default probabilities of a firm and the corresponding internal and external business cycles, especially in times of severe financial turmoil. Furthermore, business cycles are related to countries. Bhattacharjee *et al.* (2009) directly detected the relationship between the failure of a UK listed firm and the macroeconomic environment by applying hazard models. They explicitly stated that unstable economic conditions, especially in the US business cycle, can improve the likelihood of a firm going bankruptcy in the UK. The survival analysis has gradually become another important methodology for predicting failed events in finance (Beaver *et al.* 2005, Duffie *et al.* 2007, Bharath and Shumway 2008, Campbell *et al.* 2008, Ding *et al.* 2012, Bauer and Agarwal 2014).

The selection of explanation variables and statistical methodologies depends on researchers' subjective knowledge so these may not be the "true" models to solve the targeted issue. Ignoring parameter and model uncertainty may cause overconfidence in forecasts form models and lower out-of-sample predictive ability. To solve parameter and model uncertainty, Bayesian Model Averaging (BMA) is developed as a method of selecting a subset of potential regressors, but it allows that all inference is averaged over models, using the corresponding posterior model probabilities as weights. It captures parameter uncertainty in a model through prior distribution and then model uncertainty is obtained by posterior parameter using Bayes' theorem (Fragoso *et al.* 2018). BMA is not widely applied to failure prediction. Traczynski (2017) employed BMA with discrete hazard model to explain model uncertainty in the default of US firms and detected that the ratio of total liabilities to total assets and the volatility of market returns have a role to play in failure prediction. He also suggested that considering industry-specific effects in the BMA model may provide more accuracy in out-of-sample prediction.

In another approach to failure prediction (structural models), Black and Scholes (1973) and Merton (1974) applied the option pricing theory to calculate the probability of default from a company's market value if the value of a company cannot meet the required payments under the strong assumption of a frictionless market. Jarrow and Turnbull (1995) and Duffie and Singleton (1999) further indicated that default intensity can be identified from financial securities in the market regardless of whether default events are experienced by issuers. Thus, financial participants start to evaluate potential loss and credit risk using market-based measures and some financial agencies issue the default probability of a company on the basis of option-pricing structural models such as KMV (Crosbie and Bohn 2003). Since structural models rely on an assumption that all information can be reflected in stock prices, this limits the application of these models to listed firms.

Unlike reduced-form and structural models relying on certain assumptions, another dimension of modelling firm failure is related to machine learning models, which can deal with a huge number of factors in the model. Most of these models can be categorised into data mining fields such as intelligent techniques. It can be exemplified by the Naive Bayes (NB) classifier, the k-nearest neighbours (k-NN) classifier, neural network (NN), decision trees, case-based reasoning, evolutionary approaches, data envelopment analysis (DEA) and quadratic programming (QP), other intelligent techniques such as support vector machine, fuzzy logic techniques and so on. These studies do not focus on detecting the relationship between potential factors and firms' bankruptcy but they tend to improve the predictive performance.

Frydman *et al.* (1985) introduced Recursive Partitioning Algorithm (RPA) into the prediction of financial distress for companies and compared it with discriminant analysis. They concluded that the RPA model could produce more accurate in-sample and out-of-sample forecasts than discriminant analysis. Messier and Hansen (1988) employed the induction algorithm in loan default and bankruptcy forecast and obtained better performances in comparison with discriminant analysis. Cronan *et al.* (1991) further studied the results of RPA compared to those of discriminant analysis, logit, probit, and ID3 decision tree algorithm when it came to mortgage, commercial, and consumer lending problems, and confirmed the outperformance of RPA in comparison to other models.

Tam (1991), Coats and Fant (1993), Udo (1993), Wilson and Sharda (1994), Lacher *et al.* (1995), Lee *et al.* (1996), Etheridge and Sriram (1997) and Yang *et al.* (1999) demonstrated neural network models, which outperformed other previous models in terms of differently corresponding dimensions of verification such as prediction accuracy, adaptability and robustness. In contrast to these conclusions, a significant difference in predictive ability of bankruptcy between neural network and discriminant analysis for Italian companies (Altman *et al.* 1994) cannot be observed. Bryant (1997) developed a Case-based Reasoning (CBR) system in the prediction of bankruptcy, which did not outperform logit regression. A similar conclusion was also reached by Jo *et al.* (1997).

Meanwhile, genetic algorithms (GA) have been gradually implemented in bankruptcy prediction. Varetto (1998) demonstrated that GA could not consistently produce better performances than discriminant analysis during the evaluation of insolvency risk and he demonstrated that discriminant analysis is relatively stable and has generalized ability in the estimated procedure of risk. Nanda and Pendharkar (2001) aimed to clarify how the misclassification cost matrix is applied for bankruptcy prediction and suggested that GA was the best model, providing the highest percentage of correct predictions compared to statistical linear discriminant analysis and goal programming. They further demonstrated that taking into account asymmetric misclassification costs can increase the percentage of correct bankruptcy prediction. Shin and Lee (2002) further confirmed the usefulness of GA in bankruptcy prediction.

During recent decades, the support vector machine (SVM) as a specific algorithm of the supervised learning model with the maximum margin hyperplane has attracted more attention from researchers and it is frequently employed in forecasting firms' failure. Min and Lee (2005) suggested that the predictive accuracy of SVM for firms' bankruptcy is better than that of multiple discriminant analysis, logistic regression analysis and three-layer fully connected back-propagation neural networks. An increasing number of researchers tend to combine SVM with other algorithm techniques to improve its predictive performance when it comes to firms' failure since the model form, parameter setting, and features selection can significantly affect the predictive ability of SVM and ultimately confirm the efficiency of SVM in bankruptcy prediction (Zhou *et al.* 2014).

From this literature review of intelligent techniques, it can be observed that the applied algorithm is gradually complex and abstruse, which implies that it may not be easy for market participants to use. According to the results obtained in the study of Barboza *et al.* (2017), the predictive performance of reduced-form methodologies is not constantly worse than in other machine learning models. In other words, the complicated machine learning models do not imply better predictive performance when it comes to the failure of firms, since more outliers may be captured in the dataset. Thus, some research still emphasises and promotes the application of a simple method in bankruptcy prediction. The Naive Bayes (NB) classifier and the k-nearest neighbours (k-NN) classifier as the basic and simple machine learning models are now attracting more attention from researchers.

Sarkar and Sriram (2001) indicated that the predictive performance of banks' failure provided by the Naive Bayes classifier and composite attribute model are comparable, which is similar to the prediction provided by decision tree classification algorithm C4.5. Twala (2010) investigated the predictive performance of five classifiers containing both the NB classifier and the k-NN classifier for credit risk. He demonstrated that the NB classifier is the most effective classifier in credit risk prediction and its performance can be improved by applying classifier ensembles. Henley and Hand (1996) implemented k-NN with an adjust Euclidian distance metric to solve the credit scoring problem and suggested k-NN is capable of solving this problem with a lower risk rate in comparison with linear regression, logit, decision trees and decision graphs. Park and Han (2002) employed an Analytic Hierarchy Process weighted k-NN model to forecast bankruptcy again and demonstrated that this model was the best one considering classification accuracy among regression, logit, weighted k-NN and the pure k-NN classifier. In order to compare the NB classifier and the k-NN classifier in detail, Islam *et al*. (2010) already reviewed the theory of these classifiers and then applied them to a dataset credit card approval prediction, which indicated the advantages and disadvantages of these models in detail.

This previous work provides us with a foundation for choosing related variables to predict the failure of firms and indicates a methodological gap in failure prediction. We will discuss our data and methodologies in the following sections.

4.3 Data and summary statistics

4.3.1 Data sources

In this study, firms are defined as having failed in a particular year if the company is in receivership, liquidation or has been dissolved, and its last accounts were reported in the previous year (Bunn and Redwood 2003, Bridges and Guariglia 2008, Guariglia *et al*. 2016). A firm which has been taken over is not included as a failure in this work since takeovers can be considered as evidence of continuing success (i.e. the firm will continue to operate) rather than failure.

The FAME database provided by Bureau van Dijk Electronic Publishing is applied to extract firm-specific accounting data from the balance sheet and profit and loss statements of companies in UK and Ireland. In this database, the turnover, pre-tax profits and shareholders' funds in a company are at least £1.5m, £150,000 and £1.5m respectively. Meanwhile, these firms should be recorded within the last five years. If a firm has been inactive for more than four years, its record cannot be observed from FAME. This implies that a firm that failed before 2006 may be omitted if only the 2010 version of FAME is used. To solve this potential problem, Javorcik and Li (2013), Guariglia *et al*. (2016) and Görg and Spaliara (2018) suggest that different versions of the database be included in order to check the status of a company. Thus, the FAME editions of October 2010, October 2008 and February

2005 and archived editions of 1998 and 1994 are used to track the status of a firm continuously.

Almost all firms in this database are not in the public market or alternative exchanges such as the Alternative Investment Market (AIM) and the Off-Exchange (OFEX) market. It is an attractive characteristic of our dataset that considerable diversity in the relationship between financial conditions and failed probability can be observed there, because of a high degree of information asymmetry in these private firms. Importantly, the operation of these unlisted firms is highly correlated with the development of the economy (Beck *et al*. 2005, Ayyagari *et al*. 2007). The special role of private firms in the economy makes this study appealing. Following common selection criteria in the literature, companies that do not have complete records on our explanatory variables, and firm-years with negative sales and assets are excluded.

In addition, the data period spans from 1991 to 2009 and thus two important financial crises in UK economic history are covered in the applied dataset: the 1991–1993 European Exchange Rate Mechanism (ERM) currency crisis and the 2008-2009 global financial crisis (GFC). In the early 1990s, the UK as the member of the ERM had to comply with its restrictions, which caused high interest rates and the devaluation of the pound. Meanwhile, the activity of currency speculators further increased the pressure on the pound. Both these factors forced the UK to exit the ERM and the country entered into recession in the early 1990s. The cause of the GFC was different from the causes of the ERM currency crisis, since the GFC was set off by the crisis in the subprime mortgage market in the US and the bankruptcy of Lehman Brothers. As these two financial crises have different causes and scope, our dataset is separated into sub-samples to make clear the influence of financial crises on failure prediction of private firms in the UK.

To control for the potential influence of outliers, we winsorize the regression variables at the 5th and 95th percentiles. Consolidated firms are only kept in the dataset to prevent double-counting firms and subsidiaries or operations abroad. Since private firms are incorporated in the dataset, some market-driven factors cannot be applied in this work. Thus, macroeconomic variables as another part of explanatory variables are sourced from Bloomberg. Our combined panel has an unbalanced structure containing a total of 80,585 annual observations (firm-years) on 18,744 UK and Irish firms. We have an entry for each firm-year with financial and market data.

4.3.2 Choice of explanatory variables

Previous research on the failure prediction of firms has accounted for both business and financial risks. In this work, business risk is related to macro-economic conditions and industry characteristics, while financial risk includes the assessment of a firm's accounting procedures, its profit and loss situation and its overall financial policy. Thus, the selection of independent variables is guided by the existing failure prediction literature (see for instance Altman 1968, Altman *et al*. 1977, Ohlson 1980, Altman 1993, Shumway 2001, Griffin and Lemmon 2002, Chava and Jarrow 2004, Jiménez and Saurina 2004, Duffie *et al*. 2007, Campbell *et al*. 2008, Bhattacharjee *et al*. 2009, Bonfim 2009, Margaritis and Psillaki 2010, Traczynski 2017). The expected relationship between applied predictors and firms' failure is presented in Table 4-1, which provides a detailed description of the variables used in this study. All firm-specific ratios are presented as percentages and the macroeconomic variables are de-trended by the Hodrick-Prescott filter.

Table 4-1 Expected signs and variables definition

Variable	Definition	Expected relationship
Firm-specific variables (10)		
AGE	Firm ages	-
Size (2)		
LNTA	The logarithm of the ratio of total assets to GDP deflator	-
LNSALE	The logarithm of total turnover over GDP deflator	-
Liquidity (1)		
WCTA (%)	Working Capital / Total Assets	-
Leverage (2)		
TLTA (%)	Total Liabilities/ Total Assets	+
CLTA (%)	Current Liabilities / Total Assets	+
Profitability (3)		
RETPROTA (%)	Retained profits / Total Assets	-
GROPROTA (%)	Gross Profits / Total Assets	-
SALETA (%)	Total Turnover / Total Assets	-
Collateral (1)		
TANTA	The tangible assets / total assets	+
Macroeconomic Variables (8)		
RGDPGR_UK	The growth rate of real GDP in UK.	~
RGDPGR_US	The growth rate of real GDP in US.	~
RINTR	The yield on 10-year Treasury bonds in UK less the annual rate of inflation (CPI)	~
REER	The real effective exchange rate in UK	~
REER_VOL	The volatility of the real effective exchange rate	~
LNRET	The logarithm of FTSE 100 return	~
VOL	The volatility of Stock Price Index for United Kingdom	~
CIEA	The coincident indicator for economic activity in UK	~

Notes: "+" indicates that the probability of a firm failing would increase if the covariates rise. "-" indicates that the probability of a firm failing would reduce if the covariates rise. "~" indicates uncertainty in the sign.

4.3.2.1 Firm-specific variables

There are 10 firm-specific accounting variables measuring the different financial conditions of firms, as predictors of the probability of failure. These are related to age, size, liquidity, leverage, profitability and collateral. The firm age (AGE) is defined as the difference between the current year and the date of incorporation. Young firms in general are not have efficient enough to compete with existing firms, which would lead us to expect them to face more default risk. The firm size is measured by two ratios, the natural logarithm of firms' real total assets (LNTA) and the natural logarithm of firms' turnover (LNSALE), which explain the scale of a firm. A large firm would be expected to have lower probability of failure. Next, the liquidity of firms is specified by the ratio of working capital over total assets (WCTA), which indicates the percentage of remaining liquid assets in comparison with total assets. Lower levels of liquidity should increase the probability of failure of a company. The following two measures capture the leverage information of a firm: ratio of total liabilities to total assets (TLTA) and the ratio of current liabilities over total assets (CLTA). This metric enables comparisons of leverage to be made across different companies. The higher the ratio is, the higher the degree of leverage, and consequently, financial risk and the probability of failure increase. Further, we employ profitability as measured by the ratio of retained profits over total assets (RETPROTA), the ratio of gross profits to total assets (GROPROTA) and the ratio of total turnover over total assets (SALETA). These ratios can be regarded as indicators of the efficiency with which a firm arranges its assets to generate revenue. It is expected that firms with higher profitability are not likely to fail. Finally, the collateral condition of a firm is captured by the ratio of tangible assets over total assets (TANTA). Since tangibles are potentially good collateral and they are easy to monitor, firms with more tangibles are likely to use them as collateral to get riskier investment, which is in turn likely to increase the probability of failure.

4.3.2.2 Macroeconomic indicators

A list of macroeconomic factors, which measure different aspects of the aggregate economy’s performance, are considered as a potential influence on the probability of failure. Specifically, the growth rate of real GDP (RGDPGR_UK and RGDPGR_US) captures the aggregate business cycle in the UK and US respectively. The interest rate (RINTR) is measured by the yield on 10-year Treasury bonds in the UK minus the annual rate of inflation (CPI). Since the European Exchange Rate Mechanism (ERM) currency crisis is covered in the data period, the real effective exchange rate in the UK (REER) and the volatility of the real effective exchange rate (REER_VOL) are included in this study. The stock market performance is evaluated by the FTSE 100 return, which calculates logarithm returns on the FTSE 100 index (LNRET). The volatility of the stock price index for the United Kingdom is represented by VOL. Aggregate economic activity is captured by a coincident indicator in the UK (CIEA). Their relationship with the probability of failure could be either positive or negative since probability would be lower during economic prosperity, but firms may change their capital structure to achieve more profits and ultimately face more credit risk in the market. Hence, the relationship between the probability of failure and macroeconomic variables is an issue that will be determined empirically.

In addition to these explanatory variables, industry dummy variables are constructed based on the four-digit UK SIC code information provided from FAME to classify firms. Time dummy variables are also considered to capture the specific feature of the probability of failure in each year.

4.3.3 Summary statistics

Table 4-2 reports the distribution of the percentage of surviving and failed firms annually in the full sample. It can be observed that the distribution of failed firms during the financial crisis is different from in other periods, which implies financial crisis affects the probability of a firm failing.

Table 4-2 Firms by year

Year	The percentage of non-failed firms	The percentage of failed firms	Year	The percentage of non-failed firms	The percentage of failed firms
1991	95.60%	4.40%	2001	83.71%	16.29%
1992	95.66%	4.34%	2002	83.54%	16.46%
1993	95.67%	4.33%	2003	83.56%	16.44%
1994	91.23%	8.77%	2004	83.74%	16.26%
1995	91.11%	8.89%	2005	88.35%	11.65%
1996	91.14%	8.86%	2006	88.01%	11.99%
1997	91.19%	8.81%	2007	87.82%	12.18%
1998	91.28%	8.72%	2008	87.44%	12.56%
1999	84.70%	15.30%	2009	87.93%	12.07%
2000	83.68%	16.32%			

Note: This table presents the distribution of percentage of non-failed and failed firms by year.

At the next stage, summary statistics related to the full sample for explanatory variables is reported in Table 4-3. The statistics splitting the sample between non-failed firms and failed firms are presented to measure any differences across operating statuses. We test the equality of means across the above-mentioned groups and corresponding p-values are reported in the final columns of the tables. It can be observed, as expected, that surviving firms have better financial characteristics, as measured by the balance sheet indicators. The results from equity tests suggest that significant differences between the two groups can be observed, which indicates that there is a correlation between better financial health and a reduced probability of failure. In other words, there is considerable cross-sectional variation in the probability of a firm failing. This gives us the opportunity to consider the influence of firm heterogeneity on failure predictions for candidate models.

Table 4-3 Descriptive statistics

Variable	Mean	Standard Deviation	Minimum	Maximum	P-value
	(1)	(2)	(3)	(4)	(5)
AGE					
Non-failed firms	22.6570	16.0314	0.0000	68.0000	
Failed firms	19.9702	15.0966	0.0000	68.0000	0.0000
Size (2)					
LNTA					
Non-failed firms	3.9043	1.1595	0.8311	6.3460	
Failed firms	3.6045	1.0473	0.8326	6.3439	0.0000
LNSALE					
Non-failed firms	4.3554	1.1061	1.7038	6.9179	
Failed firms	4.0980	0.9947	1.7057	6.9117	0.0000
Liquidity (1)					
WCTA (%)					
Non-failed firms	29.6667	14.5518	-1.0695	61.9880	
Failed firms	27.8332	14.4563	-1.0080	61.8729	0.0000
Leverage (2)					
TLTA (%)					
Non-failed firms	73.2779	27.0327	13.3021	167.3826	
Failed firms	81.9776	26.3960	13.9838	167.2278	0.0000
CLTA (%)					
Non-failed firms	49.8662	19.6798	13.1116	97.5595	
Failed firms	57.1058	19.4512	13.2031	97.4340	0.0000
Profitability (3)					
RETPROTA (%)					
Non-failed firms	2.7953	5.9244	-18.2854	17.1717	
Failed firms	1.2631	6.4274	-18.2788	17.1046	0.0000
GROPROTA (%)					
Non-failed firms	48.1921	25.4957	4.7934	131.1709	
Failed firms	47.4295	25.6442	4.8336	130.7475	0.0359
SALETA (%)					
Non-failed firms	169.1523	64.9279	48.0219	392.1053	
Failed firms	176.8552	68.1927	48.1886	390.8996	0.0000
Collateral (1)					
TANTA (%)					
Non-failed firms	28.2724	16.5645	2.1207	66.3787	
Failed firms	29.9559	16.9569	2.1269	66.3717	0.0000

Notes: This Table reports the summary statistics of the explanatory variables used in the empirical models. Column 5 reports the p-value for the test of equality of means between the failed and non-failed groups.

Moving to compare the ERM currency crisis and GFC periods, failed firms with opposite financial characteristics can be observed in Table 4-4. During the ERM currency crisis, failed firms were in the main young and small, with worse ways of operating, such as higher leverage and lower profitability. However, under a tightened credit supply, old and bigger firms, with lower leverage and higher profitability, tended to fail during the GFC. This tendency suggests the whole sample period should be separated into sub-sample periods to check the predictive ability of the models implemented.

Table 4-4 Sample means

Variable	ERM currency crisis	Non-ERM currency crisis	P-value	Global financial crisis	Non-global financial crisis	P-value
	(1)	(2)	(3)	(4)	(5)	(6)
AGE	21.2714	22.6081	0.0000	24.7052	22.3243	0.0000
LNTA	3.6121	3.9133	0.0000	4.0309	3.8744	0.0000
LNSALE	4.0687	4.3669	0.0000	4.5032	4.3270	0.0000
WCTA	30.6969	29.4263	0.0000	28.8288	29.5975	0.0002
TLTA	81.1101	73.0857	0.0000	58.7676	74.9115	0.0000
CLTA	51.4288	50.2254	0.0000	46.5853	50.6047	0.0000
RETPROTA	1.9204	2.7762	0.0000	3.0687	2.6687	0.0000
GROPROTA	52.5805	47.6772	0.0000	48.4509	48.1204	0.3613
SALETA	168.6858	169.7576	0.1714	172.9032	169.4266	0.0002
TANTA	31.0116	28.1070	0.0000	23.4429	28.7312	0.0000

Notes: This table reports the sample means of the explanatory variables in different periods. The period from 1991 to 1993 is defined as the ERM currency crisis and otherwise non-ERM currency crisis. The global financial crisis is between 2008 and 2009 and otherwise non-global financial crisis. Column 3 reports the p-value for the test of equality of means between firms during the ERM currency crisis and during the non-ERM currency crisis. Column 6 reports the p-value for the test of equality of means between firms during the global financial crisis and during the non-global financial crisis.

4.4 Methodology

4.4.1 Discrete Hazard (DH) model

In line with previous studies of default prediction, a DH model as the discrete-time version of survival models is used as the benchmark model in this work (Shumway 2001, Chava and Jarrow 2004, Beaver et al. 2005, Duffie et al. 2007, Campbell et al. 2008, Ding et al. 2012, Tian et al. 2015, Traczynski 2017). It captures the discrete-time characteristics of the dataset when rough timescales are applied in data collection; for example when the time of an event is expressed in weeks, months or years (Allison 1982, Shumway 2001, Rabe-Hesketh and Skrondal 2012). The DH model can also explain the clustered property of panel data. In the work, dependent variable $(Y_{i,t})$ is the failure, which is set to 1 in year t if a firm fails; otherwise it is equal to 0. Hence, this DH model for one-year-ahead prediction of a firm failing is given by the following equation:

$$Pr\left(Y_{i,t} = 1 \middle| Y_{i,t-1} = 0, X_{i,t}\right) = h\left(t, X_{i,t}\right)$$

$$\frac{h\left(t, X_{i,t}\right)}{1 - h\left(t, X_{i,t}\right)} = \frac{h_0(t,0)}{1 - h_0(t,0)} \exp(\beta_0 + \beta' X_{i,t}), \tag{1}$$

where $h\left(t, X_{i,t}\right)$ is the hazard rate at time t for a firm, controlling by a set $X_{i,t}$ of time-varying indicators including firm-specific and macroeconomic variables , and $h_0(t,0)$ is the baseline hazard function without any restriction. The hazard rate $h\left(t, X_{i,t}\right)$ indicates the interval hazard for the period between the beginning and the end of the jth year after the first appearance of failure for a firm. In other words, it is the likelihood that firms fail at time t conditional on the fact that they have survived in $t - 1$, which takes the generalized linear model with logistic link (Shumway 2001, Chava and Jarrow 2004), written as:

$$logit\left(h\left(t, X_{i,t}\right)\right) = \beta_0 + \beta' X_{i,t}, \tag{2}$$

where β_0 can be regarded as the logit of the baseline hazard rate without any assumption, β is a vector of unknown parameters to be estimated and $X_{i,t}$ is the same dataset used in equation (1). According to equation (2), the parameter estimates can be achieved by maximizing the log-likelihood function:

$$l(\hat{\beta}) = \sum \left(Y_{i,t} * \ln\left(\frac{1}{1 + \exp\left(-\left(\beta_0 + \beta' X_{i,t}\right)\right)}\right) + \left(1 - Y_{i,t}\right) * \ln\left(\frac{\exp\left(-\left(\beta_0 + \beta' X_{i,t}\right)\right)}{1 + \exp\left(-\left(\beta_0 + \beta' X_{i,t}\right)\right)}\right)\right). \tag{3}$$

To capture the changes in the relationship between the potential position of a firm and probability of a firm failing, the time dummy factors and the industry dummy factors respectively are included in equation (2). Thus, we can construct another two DH models, which can be written as:

$$logit\left(h(t, X_{i,t})\right) = \beta_0 + \beta_1{}'X_{i,t} + \beta_2{}'Time, \tag{4}$$

and

$$logit\left(h(t, X_{i,t})\right) = \beta_0 + \beta_1{}'X_{i,t} + \beta_2{}'Industry, \tag{5}$$

where $Time$ and $Industry$ represent the time dummy and industry dummy variables respectively. Therefore, the corresponding estimated coefficients can be calculated by maximizing the log-likelihood functions:

$$l(\hat{\beta}) = \sum\left(Y_{i,t} * \ln\left(\frac{1}{1 + \exp\left(-(\beta_0 + \beta_1{}'X_{i,t} + \beta_2{}'Time)\right)}\right) + (1 - Y_{i,t}) * \ln\left(\frac{\exp\left(-(\beta_0 + \beta_1{}'X_{i,t} + \beta_2{}'Time)\right)}{1 + \exp\left(-(\beta_0 + \beta_1{}'X_{i,t} + \beta_2{}'Time)\right)}\right)\right), \tag{6}$$

and

$$l(\hat{\beta}) = \sum\left(Y_{i,t} * \ln\left(\frac{1}{1 + \exp\left(-(\beta_0 + \beta_1{}'X_{i,t} + \beta_2{}'Industry)\right)}\right) + (1 - Y_{i,t}) * \ln\left(\frac{\exp\left(-(\beta_0 + \beta_1{}'X_{i,t} + \beta_2{}'Industry)\right)}{1 + \exp\left(-(\beta_0 + \beta_1{}'X_{i,t} + \beta_2{}'Industry)\right)}\right)\right). \tag{7}$$

4.4.2 The Bayesian Model Averaging (BMA) version of DH models

4.4.2.1 Bayesian Model Averaging (BMA)

Bayesian Model Averaging (BMA) is a method related to identifying a subset of potential regressors, but it allows that all inference is averaged over models, using the corresponding posterior model probabilities as weights. Fragoso *et al.* (2018) state that BMA is derived from the usual Bayesian inference methods. It captures parameter uncertainty in a model through the prior distribution, and then model uncertainty is obtained by posterior parameter using Bayes' theorem. In BMA, we suppose that $M = (M_1, \ldots, M_m)$ is our collection of candidate models with different predictors from vector $X_{i,t}$ and $\boldsymbol{y}$ is data information. $Y_{i,t}$ is a failed event, presumed to be well-determined for every model. The $\hat{\beta}_M$ for all candidate models M in BMA can be estimated:

$$\hat{\beta}_M = \sum_{i=1}^{m} \hat{\beta}_{M_i} * p(M_i|\boldsymbol{y}), \tag{8}$$

where $p(M_i|\boldsymbol{y})$ is the posterior probability of the model M_i and $\hat{\beta}_{M_i}$ is the estimated parameter in the model M_i. $p(M_i|\boldsymbol{y})$ can be computed by Bayes' rule as:

$$p(M_i|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|M_i)p(M_i)}{\sum_{j=1}^{m} p(\boldsymbol{y}|M_j)p(M_j)}, \tag{9}$$

where $p(\boldsymbol{y}|M_i)$ is the marginal distribution of the data under the ith model, M_i, given previously and $p(M_m)$ is the prior probability for the model, M_i. $p(\boldsymbol{y}|M_i)$ is calculated by the integral:

$$p(\boldsymbol{y}|M_i) = \int f(\boldsymbol{y}|\beta_{M_i}, M_i) * f(\beta_{M_i}|M_i) d\beta_{M_i}, \tag{10}$$

where $f(\boldsymbol{y}|\beta_{M_i}, M_i)$ is the likelihood of the data conditional on the model M_i, $f(\beta_{M_i}|M_i)$ is the prior distribution of β_{M_i}, and β_{M_i} is the estimated parameter vector from the model M_i.

4.4.2.2 The BMA version of DH models

From equation (3), the coefficients β can only be estimated from one DH model. In order to generate the BMA version of DH models, equations (3) and (8) have to be combined and modified to fit the average situation. In line with the study of Traczynski (2017), a constraint has to be added to control the scale of the coefficients, which is that the variance of the potential variable $Y_{i,t}{}^*$, measuring the financial conditions of firms in the log function, is set to 1 in every model. This can ensure that the interpretation of a predictor's coefficient would be the change in standard deviations of a variable with a 1-unit change in that variable. Thus, the log-likelihood function of a model M_i in the BMA version of DH models without considering time dummy and industry dummy variables can be written as:

$$\ln\left(f\left(\boldsymbol{y}\middle|\beta_{0,M_i},\beta_{M_i},M_i\right)\right)$$
$$= \sum\left(Y_{i,t} * \ln\left(\frac{1}{1+\exp\left(-\left(\beta_{0,M_i}+\beta'_{M_i}X_{i,t,M_i}\right)\right)}\right) + \left(1-Y_{i,t}\right)\right.$$
$$\left. * \ln\left(\frac{\exp\left(-\left(\beta_{0,M_i}+\beta'_{M_i}X_{i,t,M_i}\right)\right)}{1+\exp\left(-\left(\beta_{0,M_i}+\beta'_{M_i}X_{i,t,M_i}\right)\right)}\right)\right),$$

$$Y_{i,t}{}^{*} = \beta_{0,M_i} + \beta'_{M_i}X_{i,t,M_i} + \varepsilon_{i,t,M_i},$$

$$Y_{i,t} = \mathbf{1}\left[Y_{i,t}{}^{*} > 0\right],$$

$$var\left(Y_{i,t}{}^{*}\right) = 1, \tag{11}$$

where X_{i,t,M_i} contains the predictors in model M_i, which is part of $X_{i,t}$ in equation (1), β_{0,M_i} and β_{M_i} are the corresponding estimated parameter vectors in model M_i, ε_{i,t,M_i}, is the error term and $\mathbf{1}\left[Y_{i,t}{}^{*} > 0\right]$ is an indicator function. According to the work of Duffie *et al.* (2009), β_{0,M_i} and β_{M_i} are estimated by different prior formulations. The prior $f\left(\beta_{0,M_i}\right) \propto 1$ is used for the estimated coefficients of baseline hazard rates and the g-prior proposed by Zellner (1986) is assigned for the estimation of coefficients on the covariates, $f\left(\beta_{M_i}\middle|M_i\right) = N\left(0, g\left(X'_{i,t,M_i} * X_{i,t,M_i}\right)^{-1}\right)$, where N is a multivariate normal distribution of the same dimension as $\beta_{M_i}$and g is a scalar parameter controlling the weights of a prior in each posterior distribution of X_{i,t,M_i}. The unit information prior, $g = \frac{1}{n}$, is used following the suggestion by Fernández et al. (2001), where n is the number of observations in our sample. If the prior is set to 0, all posterior model parameter estimation of used predictors would be shrunk to 0, which implies that some applied predictors cannot be used to explain the failed probability of a firm.

Thus, the prior overall parameters of model M_i can be generated by the two priors conditional on $f\left(\beta_{0,M_i},\beta_{M_i}\middle|M_i\right) = f\left(\beta_{0,M_i}\right) * f\left(\beta_{M_i}\middle|M_i\right) \propto N\left(0, g\left(X'_{i,t,M_i} * X_{i,t,M_i}\right)^{-1}\right)$ and then the estimated coefficients $\hat{\beta}_{p,M_i} = \beta_{0,M_i},\beta_{M_i}$ in the BMA version of DH models are posterior modes, which can be calculated by maximizing the posterior log-likelihood function, written as:

$$l(\hat{\beta}_{p,M_i}) = \ln\left(f(\hat{\beta}_{p,M_i}|\boldsymbol{y}, M_i)\right) = \ln\left(f(\hat{\beta}_{p,M_i}|M_i)\right) + \ln\left(f(\boldsymbol{y}|\hat{\beta}_{p,M_i}, M_i)\right). \tag{12}$$

In line with the structure of the benchmark model, time dummy and industry dummy variables are also considered in the BMA version of the DH models. Thus, the log-likelihood function of the BMA version of the DH models including time dummy and industry dummy variables can be written respectively as:

$$\begin{aligned}
&\ln\left(f(\boldsymbol{y}|\beta_{0,M_i}, \beta_{1,M_i}, \beta_{2,M_i}, M_i)\right) \\
&\qquad = \sum\left(Y_{i,t} * \ln\left(\frac{1}{1+\exp\left(-(\beta_{0,M_i} + \beta'_{1,M_i}X_{i,t,M_i} + \beta'_{2,M_i}Time)\right)}\right)\right. \\
&\qquad\quad \left. + (1 - Y_{i,t}) * \ln\left(\frac{\exp\left(-(\beta_{0,M_i} + \beta'_{1,M_i}X_{i,t,M_i} + \beta'_{2,M_i}Time)\right)}{1+\exp\left(-(\beta_{0,M_i} + \beta'_{1,M_i}X_{i,t,M_i} + \beta'_{2,M_i}Time)\right)}\right)\right),
\end{aligned}$$

$$Y_{i,t}{}^{*} = \beta_{0,M_i} + \beta'_{1,M_i}X_{i,t,M_i} + \beta'_{2,M_i}Time + \varepsilon_{i,t},$$

$$Y_{i,t} = \mathbf{1}[Y_{i,t}{}^{*} > 0],$$

$$var(Y_{i,t}{}^{*}) = 1, \tag{13}$$

and

$$\begin{aligned}
&\ln\left(f(\boldsymbol{y}|\beta_{0,M_i}, \beta_{1,M_i}, \beta_{2,M_i}, M_i)\right) \\
&\qquad = \sum\left(Y_{i,t} * \ln\left(\frac{1}{1+\exp\left(-(\beta_{0,M_i} + \beta'_{1,M_i}X_{i,t,M_i} + \beta'_{2,M_i}Industry)\right)}\right)\right. \\
&\qquad\quad \left. + (1 - Y_{i,t}) * \ln\left(\frac{\exp\left(-(\beta_{0,M_i} + \beta'_{1,M_i}X_{i,t,M_i} + \beta'_{2,M_i}Industry)\right)}{1+\exp\left(-(\beta_{0,M_i} + \beta'_{1,M_i}X_{i,t,M_i} + \beta'_{2,M_i}Industry)\right)}\right)\right),
\end{aligned}$$

$$Y_{i,t}{}^{*} = \beta_{0,M_i} + \beta'_{1,M_i}X_{i,t,M_i} + \beta'_{2,M_i}Industry + \varepsilon_{i,t},$$

$$Y_{i,t} = \mathbf{1}[Y_{i,t}{}^{*} > 0],$$

$$var(Y_{i,t}{}^{*}) = 1. \tag{14}$$

Estimated coefficient parameters in the BMA version of DH models can then be achieved by maximizing the corresponding posterior log-likelihood function, as in equation (12).

4.4.3 Naive Bayes (NB) Classifier

The NB classifier as a probabilistic learning method has been applied in the default prediction of firms (Sarkar and Sriram 2001), which can produce probabilistic predictions such as DH models. There exist some appealing features of the NB classifier. First, it is a simple generative classified approach derived from Bayesian theorem, which can be easy to implement, even for users without a specific background in statistics. Second, the computation of evaluating the NB classifier is simple and the missing values in the dataset can easily be handled in the NB classifier when an important presumption known as the class conditional independence is satisfied in the NB classifier (Han *et al*. 2011). This presumption is that all characteristics should be independent given a specific class. In reality, this assumption is easily broken. However, it does not bring considerable bias into the probability estimation and the NB classifier is still able to work well (Domingos and Pazzani 1997). This becomes the other important reason why this classifier can be widely applied. Moreover, Han *et al*. (2011) demonstrate that the NB classifier is able to produce the minimum error rate during estimation compared with decision tree and neural network classifiers, and it is also robust if irrelevant features are included.

In our work, the NB classifier indicates how to classify the vector $X_{i,t}$ of various characteristics (explanatory variables) $K_1, K_2 \dots, K_n$ in equation (1). The interested event, $Y_{i,t}$ only has two classes $C_{i,t}$, non-failed class, 0 and failed class, 1 in this work. Thus, the NB classifier can classify an object to a specific class having the maximum posterior probability given potential features $X_{i,t} = \{K_1, K_2 \dots, K_n\}$. The posterior probability can be calculated by the Bayes rule. Based on the statement of the NB classifier (Han et al. 2011), this procedure can be explained by the following:

$$Y_{i,t} = C_{i,t} = argmax\left(p\left(C_{i,t}\middle|X_{i,t}\right)\right) = argmax\left(p\left(X_{i,t}\middle|C_{i,t}\right) * \frac{p\left(C_{i,t}\right)}{p\left(X_{i,t}\right)}\right), \quad (15)$$

where $p(C_{i,t}|X_{i,t})$ is the conditional probability given the vector $X_{i,t}$, $p(X_{i,t}|C_{i,t})$ is equal to the posterior probability of the vector $X_{i,t}$ conditioned on a specific class, $C_{i,t}$, $p(C_{i,t})$ is the prior probability of $C_{i,t}$ and $p(X_{i,t})$ is the prior probability of the vector $X_{i,t}$. Since $p(X_{i,t})$ is constant in equation (15), according to class conditional independence, $argmax\left(p(X_{i,t}|C_{i,t}) * \frac{p(C_{i,t})}{p(X_{i,t})}\right)$ in equation (15) can be converted into $argmax\left(p(X_{i,t}|C_{i,t}) * p(C_{i,t})\right)$. Hence, equation (15) can be written as:

$$Y_{i,t} = C_{i,t} = argmax\left(p(X_{i,t}|C_{i,t}) * p(C_{i,t})\right) \propto argmax\left(p(K_1, \dots, K_n|C_{i,t}) * p(C_{i,t})\right)$$
$$= argmax\left(\prod_{j=1}^{n} p(K_1|C_{i,t}) * p(C_{i,t})\right), \tag{16}$$

where $C_{i,t}$ is equal to 0 or 1, and $X_{i,t}$ contains all potential factors influencing the probability of a firm failing in equation (1).

To capture the influence of different years on the probability of failure, time dummy variables are included as the characteristics in the NB classifier to identify the failed event. The probability of a class can be obtained, conditional on potential features $W_{i,t} = \{K_1, K_2 \dots, K_n, Time\ dummy\}$ and then the posterior probability of this class can be calculated. The procedure can be calculated by the following:

$$Y_{i,t} = C_{i,t} = argmax\left(p(C_{i,t}|W_{i,t})\right) = argmax\left(p(W_{i,t}|C_{i,t}) * \frac{p(C_{i,t})}{p(W_{i,t})}\right)$$

$$= argmax\left(p(W_{i,t}|C_{i,t}) * p(C_{i,t})\right) \propto p(K_1, \dots, K_n, Time\ dummy|C_i) * p(C_i)$$
$$= \prod p(K_n, Time\ dummy|C_i) * p(C_i). \tag{17}$$

If the industry effects are also considered in the NB classifier, the potential characteristics would be changed to $Z_{i,t} = \{K_1, K_2 \dots, K_n, Industy\ dummy\}$. Thus, the process of getting a specific classification can be written as:

$$Y_{i,t} = C_{i,t} = argmax\left(p\left(C_{i,t}\middle|Z_{i,t}\right)\right) = argmax\left(p\left(Z_{i,t}\middle|C_{i,t}\right) * \frac{p\left(C_{i,t}\right)}{p\left(Z_{i,t}\right)}\right)$$

$$= argmax\left(p\left(Z_{i,t}\middle|C_{i,t}\right) * p\left(C_{i,t}\right)\right) \propto p(K_1, \dots, K_n, Industry\ dummy|C_i) * p(C_i)$$

$$= \prod p(K_n, Industry\ dummy|C_i) * p(C_i). \tag{18}$$

4.4.4 K-nearest neighbours (k-NN) classifier

The k-NN classifier is a non-parametric method, which classifies the targeted objects into some groups based on closest training objects in the feature space. It has been implemented in credit risk identification (Davis *et al.* 1992, Henley and Hand 1996, Hand and Vinciotti 2003, Islam *et al.* 2010). The k-NN classifier does not have strict assumption and removes the functions used in the parametric method; hence, the flexibility of the k-NN classifier is relatively high and it is robust for a noisy dataset (Kuramochi and Karypis 2005). The logit of this classifier is distributing an unclassified object into the classification of the nearest of a series of already classified objects and then the whole training set is stored in the memory. Thus, it can be categorized in a memory-based learning or an instance-based learning algorithm to generate a probabilistic framework (Aha 1997). In this work, following the description in Murphy (2012), the probability of a specific classification for an object in the k-NN classifier can be formally written as:

$$p\left(Y_{i,t} = C_i\middle|X_{i,t}, D, k\right) = \frac{1}{k}\sum_{i\epsilon N_k\left(X_{i,t}, D\right)} \mathbf{1}\left(Y_{i,t} = C_i\right), \tag{19}$$

where $N_k\left(X_{i,t}, D\right)$ is a specific set where “k” observations in the training data are closest to $X_{i,t}$. The set, D, is the distance measure for “closeness”, $X_{i,t}$ is already defined in equation (1) and $\mathbf{1}(e)$ is the indicator function defined as follows:

$$\mathbf{1}(e) = \begin{cases} 1\ if\ e\ is\ true \\ 0\ if\ e\ is\ false \end{cases}. \tag{20}$$

In equation (19), how to choose a suitable distance measure gauging “closeness” and how to decide the number of neighbourhoods “k” become statistical issues for researchers. For the first issue, Euclidean distance is frequently employed as a distance metric to define “closeness” (Han *et al.* 2011). It is expressed as

$d(A,B) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$, where $A = (a_1, a_2, \dots, a_n)$ and $B = (b_1, b_2, \dots, b_n)$ are two observations in the dataset. For the second, most researchers prefer to choose the value of "k" by themselves to fit their purpose and application. Rodriguez *et al.* (2010) recommend that 10-fold cross-validation can be used in the k-NN classifier to objectively select "k", since it can produce less bias and have less computational cost than other folds. In our work, the 10-fold cross-validation k-NN classifier with Euclidian metric distance are applied. The "k" value can be automatically selected in the optimal model with the highest accuracy through 10-fold cross-validation.

In order to consider time and industry influences on the probability of failure, the probability of a specific classification for an object in the k-NN classifier can be written as

$$p\left(Y_{i,t} = C_i \middle| W_{i,t}, D, k\right) = \frac{1}{k} \sum_{i \epsilon N_k(W_{i,t}, D)} \mathbf{1}\left(Y_{i,t} = C_i\right), \tag{21}$$

and

$$p\left(Y_{i,t} = C_i \middle| Z_{i,t}, D, k\right) = \frac{1}{k} \sum_{i \epsilon N_k(Z_{i,t}, D)} \mathbf{1}\left(Y_{i,t} = C_i\right), \tag{22}$$

where $W_{i,t}$ is equal to $\{X_{i,t}, Time\ dummy\}$ and $Z_{i,t}$ is $\{X_{i,t}, Industry\ dummy\}$.

4.5 Empirical results

To measure the predictive performance of all competing models, the area under receiver operating characteristic curve (ROC curve) known as AUC and the brier score are calculated. AUC is a non-parametric measure generated from the receiver operating characteristic curve, which is frequently applied to evaluate the ability of a model to discriminate between binary events. It has already been used in relevant studies to evaluate the predictive ability to identify a default event (Shumway 2001, Chava and Jarrow 2004, Bharath and Shumway 2008, Tian *et al.* 2015, Traczynski 2017). The receiver operating characteristic curve is the plot of the likelihood of verifying true-positive (in practice, a firm fails, and the model classifies it as an interested event) and false-positive (in practice, a firm fails but the model classifies it as an interested non-event) for a whole range of probable cut-points of probability values. If AUC is equal to 1, it represents a perfect prediction. If AUC is equal to or less than 0.5, it means that the corresponding model had no predictability. If the value of AUC is above 0.8, the predictive ability may be considered accurate (Hosmer Jr et al. 2013).

In order to confirm the conclusions from AUC ratios, the brier score is applied, to measure how close the predicted probability of a firm's failure is to its survival. It is equal to the average of the squared differences between the forecast probabilities and the actual outcomes (1 if a firm fails and 0 if a firm does not fail). The brier score can be calculated as $\frac{1}{N}\int_{t=1}^{N}(p_t - o_t)^2$, where p_t is the predictive probability of a firm's failure and o_t is the corresponding actual event. The lower the brier score is for a series of predictions, the higher the accuracy ratio in out-of-sample predictions is deemed to be.

4.5.1 Whole period (1991-2009)

4.5.1.1 AUC and brier scores

The statistics of out-of-sample predictions for all candidate models is reported. For the out-of-sample predictions of a failed event, an expanding window method is employed based on the past and current information available up to time T. It allows successive observations to be included in the initial sample prior to forecast of the next one-step ahead prediction of the failed event, while keeping the start date of the sample fixed. By this method, we forecast future failure $\hat{f}_{t+1}$, $\hat{f}_{t+2}$ etc. The initial estimation window is 1991 to 2005 and the first prediction date is 2006. We then increase T by one each time until T reaches 2009.

Table 4-5 indicates all AUC and brier score values among all candidates with or without considering time or industry effects. Starting with the analysis of AUC measure, the out-of-sample prediction results suggest that the NB classifier outperforms other competing models. In comparison with results of models related to DH models, this suggests that the percentage of correct predictions increases from 67% in the DH model to 68% in the BMA version of DH models without considering time or industry effects. This difference is not large, but it can be considered a significant improvement in failure prediction, since Campbell *et al.* (2008) and Traczynski (2017) argue that a 1% difference of predicted default probabilities is considerable for a firm, affecting its performance in the stock market. This implies that adding BMA into the DH model can improve the predictive performance regarding the failure of a firm. Considering two classifiers exclusively, the NB classifier produces the best predictive ability of a firm's failure, 74% correct predictions, while the k-NN classifier provides the worst, only 64% correct prediction. The brier scores further support the conclusion from AUC since a model with higher AUC value has a lower brier score and the brier scores of the NB classifier are the smallest of all competing models.

Moving to consideration of time effects or industry effects, models referred to DH models are able to provide better predictive performance than the same model without time or industry effects in Table 4-5. The value of AUC grows from 69% in the DH model to 73% in the BMA version of DH models when time effects are controlled. The improvement of prediction is about 1% when comparing the BMA version DH models with the DH model considering industry effects. This means that between 1991 and 2009 controlling time effects is more important in out-of-sample prediction than controlling industry effects. However, this trend cannot be observed in the results of the NB classifier and k-NN classifier. In particular, the NB classifier is still the best model for forecasting the failure of a firm. The same conclusion can be reached by comparing brier scores of all competing models.

Table 4-5 AUC and brier scores for all competing models during 1991-2009

	AUC	Brier scores		AUC	Brier scores
DH	66.74%	0.0314	NB	74.06%	0.0218
DH + time dummy	69.39%	0.1304	NB + time dummy	74.06%	0.0285
DH + industry dummy	68.52%	0.0308	NB + industry dummy	72.95%	0.0322
BMA_DH	67.88%	0.0288	KNN	64.00%	0.0240
BMA_DH + time dummy	72.98%	0.0246	KNN + time dummy	63.90%	0.0241
BMA_DH + industry dummy	69.33%	0.0286	KNN + industry dummy	63.97%	0.0240

Notes: "DH" means the discrete hazard model. "BMA_DH" represents the BMA version of DH models. The Naive Bayes classifier is assigned as "NB". The k-nearest neighbours classifier is represented as "KNN". "time dummy" model means models with time dummy variables. "industry dummy" represents that dummy variables are applied to capture industry effects.

According to these results, we note that improved predictive ability can be observed directly by adding BMA into the benchmark model (DH model). The NB classifier as a simple machine learning model can provide the best predictive performance of firms' failure compared with other candidate models. Controlling time or industry effects can improve the predictive accuracy of the DH model and BMA version of DH models, especially time effects. However, adding time or industry effects cannot significantly influence the predictive ability of the NB classifier and the k-NN classifier.

4.5.1.2 The decile method

In order to confirm these measures of accuracy, the percentage of failed events occurring in each decile of the distribution of predicted values is also calculated (Shumway 2001, Chava and Jarrow 2004, Bharath and Shumway 2008, Tian et al. 2015, Traczynski 2017). The decile method depending on the rank order of firm-years is not significantly affected by small changes in the predicted probabilities of failure, since they are unlikely to change the decile in which a firm-year lies in the distribution. The lowest probabilities of failure would be included in the tenth decile and the highest probabilities would be in the first decile. Thus, the high proportion in high probability deciles suggests high accuracy of out-of-sample prediction.

Table 4-6 and Table 4-7 report the percentage of failed events occurring in each decile of the predicted distribution for probability of failure and the corresponding AUC for each candidate model. Overall, all models can provide the highest percentage in the first decile compared with the rest of the deciles, and, among all models, the NB classifier provides the highest predictive accuracy in the first decile, 38%. Almost all failed events can be predicted in the first five deciles for all competing models. This suggests that candidate models have predictive accuracy in the failure of firms. It also can be confirmed that adding time or industry effects can improve the out-of-sample forecast accuracy of the DH model and BMA version of DH models in each decile prediction, but this influence cannot be observed clearly when it comes to the NB classifier and the k-NN classifier. These results also verify that simple machine learning models can provide better out-of-sample forecast performance than the benchmark model.

Table 4-6 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2009

Decile	DH	DH + time dummy	DH + industry dummy	BMA_DH	BMA_DH + time dummy	BMA_DH + industry dummy
1	23.95	23.19	27.76	25.86	30.80	28.90
2	13.69	9.51	12.93	12.17	19.77	12.55
3	11.79	16.35	11.03	13.69	11.79	11.41
4	11.03	18.25	12.17	11.79	10.65	13.31
5	9.89	9.51	9.13	9.51	6.84	8.37
6-10	29.65	23.19	26.99	26.99	20.14	25.47
AUC	66.74%	69.39%	68.52%	67.88%	72.98%	69.33%

Notes: "DH" means the discrete hazard model. "BMA_DH" represents the BMA version of DH models. "time dummy" model means models with time dummy variables. "industry dummy" represents that dummy variables are applied to capture industry effects.

Table 4-7 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2009

Decile	NB	NB + time dummy	NB + industry dummy	KNN	KNN + time dummy	KNN + industry dummy
1	38.02	38.02	31.94	19.77	19.39	19.77
2	14.45	14.45	13.69	14.83	15.21	14.83
3	8.75	8.75	13.69	13.31	12.93	13.31
4	9.13	9.13	14.83	7.60	7.60	7.60
5	11.03	11.03	8.75	8.37	8.75	8.37
6-10	18.62	18.62	17.10	36.11	36.11	36.11
AUC	74.06%	74.06%	72.95%	64.00%	63.90%	63.97%

Notes: The Naive Bayes classifier is assigned as "NB". The k-nearest neighbours classifier is represented as "KNN". "time dummy" model means models with time dummy variables. "industry dummy" represents that dummy variables are applied to capture industry effects.

4.5.2 The sub-sample periods

The preceding analysis employed the full-time period (1991 to 2009) which spans the onset of the ERM currency crisis and the global financial crisis. It could on the one hand be argued that the failure probability for a firm reaches a high level during financial crisis and on the other that financial turmoil can provide new opportunities for a firm to survive. To detect the influence of crisis-related events on probability of failure, we drop the years 1991, 1992 and 1993 to construct the post-ERM currency crisis period, remove the years 2008 and 2009 to construct the pre-global crisis period and then omit both the ERM currency crisis and the global financial crisis to generate the non-crisis period.

4.5.2.1 The post-ERM currency crisis period (1994-2009)

In this exercise, the in-sample spans from 1994 to 2005 while the years between 2006 and 2009 act as the out-of-sample period. Table 4-8 presents the AUC values, while brier scores and the percentage of failed events occurring in each decile for each candidate model in the restricted dataset are reported in Table 4-9 and Table 4-10 respectively.

Table 4-8 AUC and brier scores for all competing models during 1994-2009

	AUC	Brier score		AUC	Brier score
DH	68.85%	0.0272	NB	72.61%	0.0247
DH + time dummy	70.16%	0.0263	NB + time dummy	72.61%	0.0253
DH + industry dummy	70.39%	0.0269	NB + industry dummy	69.86%	0.0457
BMA_DH	73.55%	0.0243	KNN	64.97%	0.0243
BMA_DH + time dummy	72.93%	0.0244	KNN + time dummy	64.92%	0.0243
BMA_DH + industry dummy	74.26%	0.0242	KNN + industry dummy	64.98%	0.0243

Notes: “DH” means the discrete hazard model. “BMA_DH” represents the BMA version of DH models. The Naive Bayes classifier is assigned as “NB”. The k-nearest neighbours classifier is represented as “KNN”. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

Table 4-9 Defaults by out-of-sample prediction decile among model related to the DH model during 1994-2009

Decile	DH	DH + time dummy	DH + industry dummy	BMA_DH	BMA_DH + time dummy	BMA_DH + industry dummy
1	27.00	25.48	28.90	30.42	32.32	32.70
2	14.45	17.49	14.07	23.57	18.25	20.15
3	12.17	14.83	14.45	9.13	11.03	13.69
4	11.41	9.13	8.75	9.13	12.93	6.46
5	7.98	7.98	9.51	6.84	5.32	7.22
6-10	27.00	25.10	24.32	20.91	20.14	19.76
AUC	68.85%	70.16%	70.39%	73.55%	72.93%	74.26%

Notes: “DH” means the discrete hazard model. “BMA_DH” represents the BMA version of DH models. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

Table 4-10 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1994-2009

Decile	NB	NB + time dummy	NB + industry dummy	KNN	KNN + time dummy	KNN + industry dummy
1	31.18	31.18	28.14	24.33	24.71	24.71
2	21.29	21.29	19.39	11.79	11.03	11.03
3	10.27	10.27	10.27	8.75	9.13	9.13
4	9.13	9.13	10.65	17.87	17.87	17.87
5	7.98	7.98	6.46	6.46	6.46	6.46
6-10	20.14	20.14	25.08	30.79	30.79	30.79
AUC	72.61%	72.61%	69.86%	64.97%	64.92%	64.98%

Notes: The Naive Bayes classifier is assigned as “NB”. The k-nearest neighbours classifier is represented as “KNN”. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

Overall, the results in Table 4-8 compared to Table 4-5 suggest that removing the ERM currency crisis period can improve the predictive accuracy of all candidate models except the NB classifier. We observe that approximately 74% of predictions are correct in the BMA version of DH models, which provides the best predictive performance among all competing models. Adding time or industry dummy variables cannot improve the AUC value in the DH models. It is interesting that the AUC value in the BMA version of DH models is reduced by including time effects and increased by considering industry effects. This means that the industry position of a firm becomes an important factor influencing the probability of firms failing after the ERM currency crisis period. Similar to the previous results, adding time or industry effects does not enhance the correct out-of-sample predictions in the NB classifier and the k-NN classifier. The results of brier scores are consistent with those extracted from AUC values. Moving to the analysis in the decile method, results in Table 4-9 and Table 4-10 also support our findings in this period. The highest percentage (32%) of failed events can be observed in the first decile in the BMA version of DH models and it still outperforms others in out-of-sample prediction measured by decile method.

4.5.2.2 The pre-global crisis period (1991-2007)

In this section, the years between 1991 and 2003 are chosen as the in-sample period and the remaining years constitute the out-of-sample period. Table 4-11 reports the AUC and brier scores for all models and Table 4-12 and Table 4-13 indicate the percentage information of failed events occurring in each decile of the distribution of predicted values for all models.

Table 4-11 AUC and brier scores for all competing models during 1991-2007

	AUC	Brier score		AUC	Brier score
DH	53.03%	0.0545	NB	69.92%	0.0475
DH + time dummy	52.05%	0.1384	NB + time dummy	69.92%	0.0622
DH + industry dummy	54.23%	0.0541	NB + industry dummy	67.90%	0.0617
BMA_DH	55.37%	0.0517	KNN	63.01%	0.0476
BMA_DH + time dummy	61.14%	0.0488	KNN + time dummy	63.01%	0.0477
BMA_DH + industry dummy	57.48%	0.0518	KNN + industry dummy	63.00%	0.0477

Notes: "DH" means the discrete hazard model. "BMA_DH" represents the BMA version of DH models. The Naive Bayes classifier is assigned as "NB". The k-nearest neighbours classifier is represented as "KNN". "time dummy" model means models with time dummy variables. "industry dummy" represents dummy variables are applied to capture industry effects.

Table 4-12 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2007

Decile	DH	DH + time dummy	DH + industry dummy	BMA_DH	BMA_DH + time dummy	BMA_DH + industry dummy
1	10.28	9.66	12.31	11.06	16.98	12.15
2	8.26	4.83	7.48	7.94	14.49	9.97
3	9.81	12.15	10.28	13.55	12.93	11.37
4	11.53	14.02	11.68	14.02	11.06	14.64
5	15.58	13.40	14.33	11.06	9.35	12.62
6-10	44.55	45.95	43.93	42.37	35.21	39.26
AUC	53.03%	52.05%	54.23%	55.37%	61.14%	57.48%

Notes: "DH" means the discrete hazard model. "BMA_DH" represents the BMA version of DH models. "time dummy" model means models with time dummy variables. "industry dummy" represents that dummy variables are applied to capture industry effects.

Table 4-13 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2007

Decile	NB	NB + time dummy	NB + industry dummy	KNN	KNN + time dummy	KNN + industry dummy
1	25.23	25.23	25.39	19.63	19.47	19.78
2	19.31	19.31	15.89	13.24	13.40	13.08
3	13.08	13.08	9.81	11.99	11.99	12.15
4	9.81	9.81	10.44	11.84	11.53	11.53
5	8.26	8.26	10.12	5.30	5.76	5.45
6-10	24.30	24.30	28.35	38.01	37.84	38.01
AUC	69.92%	69.92%	67.90%	63.01%	63.01%	63.00%

Notes: The Naive Bayes classifier is assigned as “NB” model. The k-nearest neighbours classifier is represented as “KNN”. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

From Table 4-11, it can be noted that, compared to Table 4-8, the predictive accuracy among all competing models is lower than in the post-ERM currency crisis period. The AUC results in Table 4-11 suggest that the NB classifier clearly outperforms the benchmark model, since the percentage of correct predictions increases from 53% in the DH model to 70% in the NB classifier. It is attractive that the BMA version of DH models produces about 10% more correct predictions than the DH model through controlling time effects. This trend cannot be achieved by considering industry effects in corresponding models. This indicates that the BMA version of the DH model can capture more features of the dataset and then produce more accurate out-of-sample predictions than the DH model under the consideration of time effects in the pre-global crisis period. The failure prediction provided by the NB classifier and k-NN classifier is still not sensitive to the consideration of time or industry effects. The results of brier scores and the decile method for candidate models still confirm the AUC results. In the decile method, if the AUC value is greater than 60% in a model, the highest percentage of failed events can be obtained in the first decile.

4.5.2.3 The non-crisis period (1994-2007)

In this section, the years 1991, 1992, 1993, 2008 and 2009 are removed to reduce the effect of two financial crises on out-of-sample predictions in failed events for firms. The in-sample spans from 1994 to 2005 while the years between 2006 and 2007 act as the out-of-sample period. Table 4-14, Table 4-15 and Table 4-16 respectively illustrate the corresponding AUC values, brier scores and the percentage of failed events occurring in each decile for all competing models in this dataset.

Table 4-14 AUC and brier scores for all competing models during 1994-2007

	AUC	Brier score		AUC	Brier score
DH	62.71%	0.0403	NB	70.65%	0.0367
DH + time dummy	59.74%	0.0414	NB + time dummy	70.65%	0.0371
DH + industry dummy	64.33%	0.0399	NB + industry dummy	68.29%	0.0568
BMA_DH	65.22%	0.0380	KNN	66.55%	0.0360
BMA_DH + time dummy	65.22%	0.0380	KNN + time dummy	66.49%	0.0360
BMA_DH + industry dummy	66.68%	0.0378	KNN + industry dummy	66.55%	0.0360

Notes: “DH” means the discrete hazard model. “BMA_DH” represents the BMA version of DH models. The Naive Bayes classifier is assigned as “NB”. The k-nearest neighbours classifier is represented as “KNN”. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

Table 4-15 Defaults by out-of-sample prediction decile among model related to the DH model during 1994-2007

Decile	DH	DH + time dummy	DH + industry dummy	BMA_DH	BMA_DH + time dummy	BMA_DH + industry dummy
1	20.59	20.59	24.37	22.27	22.27	26.05
2	14.71	9.66	13.03	15.97	15.97	15.97
3	10.92	12.18	12.61	15.97	15.97	13.03
4	12.61	9.24	11.34	7.98	7.98	11.76
5	8.40	11.76	7.14	5.88	5.88	5.04
6-10	32.76	36.54	31.50	31.92	31.92	28.14
AUC	62.71%	59.74%	64.33%	65.22%	65.22%	66.68%

Notes: “DH” means the discrete hazard model. “BMA_DH” represents the BMA version of DH models. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

Table 4-16 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1994-2007

Decile	NB	NB + time dummy	NB + industry dummy	KNN	KNN + time dummy	KNN + industry dummy
1	27.31	27.31	26.47	24.79	24.79	24.79
2	19.33	19.33	18.49	16.81	16.39	16.81
3	14.29	14.29	12.18	11.76	12.18	12.18
4	10.50	10.50	8.40	7.98	7.98	7.56
5	7.56	7.56	6.72	6.30	6.30	6.30
6-10	21.00	21.00	27.72	32.34	32.34	32.34
AUC	70.65%	70.65%	68.29%	66.55%	66.49%	66.55%

Notes: The Naive Bayes classifier is assigned as “NB”. The k-nearest neighbours classifier is represented as “KNN”. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

The results in Table 4-14 demonstrate that competing models can provide more accurate predictions in this period compared with the pre-global crisis period and less accurate predictions compared with post-ERM currency crisis period. This suggests that the ERM currency crisis more heavily affected the prediction of failure for private firms than did the global financial crisis in the UK. The highest predictive accuracy in a firm's failure is produced by the NB classifier and the lowest predictive performance is provided by the DH model. It should be noted that considering time effects does not improve the correct prediction in all candidate models. There exists about a 2% increase in the percentage of correct predictions in the DH model and the BMA version of DH models after controlling industry effects, while this tendency cannot be observed in the NB classifier and the k-NN classifier in Table 4-14. This suggests that controlling industry effects can provide better prediction in the non-crisis period. The brier scores meet the expectation that a model with higher AUC value has lower brier scores. For the decile method, the results in Table 4-15 and Table 4-16 support these conclusions. The highest percentage of failed events can be identified in the first decile of the distribution of predicted values among all competing models and the NB classifier provides the best performance in the decile measure.

4.5.2.4 Discussion

To sum up, the predictive performance in prediction of failure of all candidate models is influenced by financial crises, according to the out-of-sample evaluations. Among three sub-period samples, candidate models provide the best predictive performance during the post-ERM currency crisis period (1994-2009) and the worst predictive performance during the pre-global crisis period (1991-2007). This suggests that the ERM currency crisis heavily influenced the operations of UK private firms and significantly reduced the predictive ability of candidate models compared with the global crisis in the UK. In other words, a financial crisis that begins in the UK has a greater effect on a private firm's failure in the UK compared to a crisis that started elsewhere. It also can be observed that the NB classifier outperforms other models in failure prediction apart from the post-ERM currency crisis period. This indicates that the simple classifiers in machine learning techniques still have relatively high predictive accuracy regarding a firm's failure. Due to the simple application, the NB classifier can be recommended for future prediction of failure. Similar to the results over the whole period, adding

BMA can improve the predictive power of the DH model when it comes to firms' failure, since the BMA version of the DH model considers parameter and model uncertainty in prediction. Controlling time or industry effects can significantly improve accuracy of out-of-sample prediction in the DH models and its BMA version while it does not work on the NB classifier and k-NN classifier. In particular, industry effects have better predictive ability than time effects in the post-ERM currency crisis period and non-crisis period while time effects are more important in the pre-global crisis period and over the whole sample.

4.5.3 The sub-samples in cross sections

4.5.3.1 The sub-samples in firm size

Since firm heterogeneity was identified in Section 3, we first separate firms into small and large firms to control cross-sectional difference.[2] With respect to small firms, the AUC values, brier scores and the percentage of failed events occurring in each decile for all competing models are reported in Table 4-17, Table 4-18 and Table 4-19 respectively. For large firms, all information is documented in Table 4-20, Table 4-21 and Table 4-22.

Starting with small firms, we note that, among all candidate models in Table 4-17, the best predictive performance is provided by the BMA version of DH models. The AUC improves from about 69% in the DH model to around 73% in the BMA version of DH models without considering either time or industry effects. It confirms previous conclusions that adding BMA to the DH model can improve predictive ability. It is interesting to note that the NB classifier does not outperform the BMA version of DH models in terms of out-of-sample accuracy, even when time or industry dummy variables are included. For both the DH model and the BMA version of DH models, adding industry effects increases by about 2% in AUC. It means that controlling industry effects for small firms can improve the predictive ability of the DH and its BMA version. The predictive performance of the NB classifier and the k-NN classifier is not sensitive to adding time or industry effects. These results can be validated by brier scores (Table 4-17) and decile method

[2] In these sub-samples, the data period spanning from 1991 to 2005 is chosen as the in-sample period, while the remaining years between 2006 and 2009 are considered as the out-of-sample period.

(Table 4-18 and Table 4-19). The highest percentage of failures can be observed in the first decile of the distribution of predicted values in the BMA version of DH models and its first five deciles are about 80%.

Table 4-17 AUC and brier scores for all competing models during 1991-2009_small firms

	AUC	Brier scores		AUC	Brier scores
DH	69.12%	0.0438	NB	70.23%	0.0309
DH + time dummy	57.12%	0.4624	NB + time dummy	70.23%	0.0382
DH + industry dummy	70.57%	0.0434	NB + industry dummy	68.85%	0.0424
BMA_DH	72.53%	0.0375	KNN	64.31%	0.0331
BMA_DH + time dummy	73.13%	0.0344	KNN + time dummy	64.23%	0.0332
BMA_DH + industry dummy	73.14%	0.0373	KNN + industry dummy	64.33%	0.0331

Notes: “DH” means the discrete hazard model. “BMA_DH” represents the BMA version of DH models. The Naive Bayes classifier is assigned as “NB”. The k-nearest neighbours classifier is represented as “KNN”. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

Table 4-18 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2009_small firms

Decile	DH	DH + time dummy	DH + industry dummy	BMA_DH	BMA_DH + time dummy	BMA_DH + industry dummy
1	24.24	6.06	24.85	25.45	28.48	27.27
2	11.52	0.61	16.36	14.55	21.21	17.58
3	19.39	21.21	16.97	21.21	12.12	17.58
4	13.33	22.42	10.30	12.12	10.91	12.73
5	6.67	11.52	9.70	9.70	9.70	7.88
6-10	24.85	38.19	21.82	16.98	17.57	16.97
AUC	69.12%	57.12%	70.57%	72.53%	73.13%	73.14%

Notes: “DH” means the discrete hazard model. “BMA_DH” represents the BMA version of DH models. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

Table 4-19 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2009_small firms

Decile	NB	NB + time dummy	NB + industry dummy	KNN	KNN + time dummy	KNN + industry dummy
1	32.12	32.12	26.06	16.97	16.97	16.97
2	14.55	14.55	19.39	18.79	18.79	18.79
3	10.91	10.91	9.70	6.06	5.45	7.27
4	6.06	6.06	7.88	13.94	14.55	12.73
5	7.27	7.27	12.12	13.33	13.33	13.94
6-10	29.10	29.10	24.85	30.90	30.90	30.30
AUC	70.23%	70.23%	68.85%	64.31%	64.23%	64.33%

Notes: The Naive Bayes classifier is assigned as “NB”. The k-nearest neighbours classifier is represented as “KNN”. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

Moving to large firms, comparing Table 4-17 and Table 4-20 shows that the predictive power of applied models is on average lower than in small firms except the NB classifier. The NB classifier provides the higher percentage of correct prediction in firms’ failure compared with other candidates in Table 4-20. It still confirms that adding BMA can improve predictive performance, since the AUC value increases from about 60% in DH models to about 67% in the BMA version of DH models. It is interesting to point out that the predictive powers of all models do not significantly change when considering time or industry effects in large firms. This implies that predictions of failure of large firms produced by all models are not sensitive to year and industry positions. The brier scores suggest a similar conclusion in Table 4-20. In Table 4-21 and Table 4-22, in terms of decile method, the highest predictive power can be observed in the first decile of the NB classifier, which is consistent with the above results.

Table 4-20 AUC and brier scores for all competing models during 1991-2009_large firms

	AUC	Brier scores		AUC	Brier scores
DH	59.36%	0.0220	NB	72.37%	0.0147
DH + time dummy	56.04%	0.5369	NB + time dummy	72.37%	0.0243
DH + industry dummy	62.30%	0.0215	NB + industry dummy	73.37%	0.0222
BMA_DH	67.10%	0.0203	KNN	62.35%	0.0163
BMA_DH + time dummy	67.10%	0.0203	KNN + time dummy	62.90%	0.0163
BMA_DH + industry dummy	67.68%	0.0201	KNN + industry dummy	62.67%	0.0163

Notes: “DH” means the discrete hazard model. “BMA_DH” represents the BMA version of DH models. The Naive Bayes classifier is assigned as “NB”. The k-nearest neighbours classifier is represented as “KNN”. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

Table 4-21 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2009_large firms

Decile	DH	DH + time dummy	DH + industry dummy	BMA_DH	BMA_DH + time dummy	BMA_DH + industry dummy
1	15.31	4.08	18.37	12.24	12.24	15.31
2	8.16	5.1	10.20	15.31	15.31	16.33
3	10.20	19.39	9.18	19.39	19.39	17.35
4	15.31	27.55	16.33	19.39	19.39	16.33
5	8.16	16.33	10.20	12.24	12.24	14.29
6-10	42.86	27.54	35.71	21.42	21.42	20.40
AUC	59.36%	56.04%	62.30%	67.10%	67.10%	67.68%

Notes: "DH" means the discrete hazard model. "BMA_DH" represents the BMA version of DH models. "time dummy" model means models with time dummy variables. "industry dummy" represents that dummy variables are applied to capture industry effects.

Table 4-22 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2009_large firms

Decile	NB	NB + time dummy	NB + industry dummy	KNN	KNN + time dummy	KNN + industry dummy
1	36.73	36.73	31.63	20.41	20.41	20.41
2	15.31	15.31	19.39	16.33	17.35	18.37
3	10.20	10.20	11.22	4.08	5.10	4.08
4	6.12	6.12	11.22	14.29	9.18	8.16
5	9.18	9.18	10.20	7.14	9.18	10.20
6-10	22.44	22.44	16.32	37.75	38.76	38.76
AUC	72.37%	72.37%	73.37%	62.35%	62.90%	62.67%

Notes: The Naive Bayes classifier is assigned as "NB". The k-nearest neighbours classifier is represented as "KNN". "time dummy" model means models with time dummy variables. "industry dummy" represents that dummy variables are applied to capture industry effects.

4.5.3.2 The sub-samples in firm age

In this section, we differentiate young from old firms in the same period. The in-sample and out-of-sample periods are also the same as in the previous section. Table 4-23, Table 4-24 and Table 4-25 report the AUC values, brier scores and the percentage of failed events for young companies respectively.

Table 4-23 AUC and brier scores for all competing models during 1991-2009_young firms

	AUC	Brier scores		AUC	Brier scores
DH	61.86%	0.0404	NB	70.87%	0.0288
DH + time dummy	56.79%	0.4620	NB + time dummy	70.87%	0.0431
DH + industry dummy	64.86%	0.0394	NB + industry dummy	72.72%	0.0390
BMA_DH	63.72%	0.0396	KNN	61.47%	0.0315
BMA_DH + time dummy	69.98%	0.0321	KNN + time dummy	61.39%	0.0315
BMA_DH + industry dummy	69.93%	0.0362	KNN + industry dummy	61.45%	0.0315

Notes: "DH" means the discrete hazard model. "BMA_DH" represents the BMA version of DH models. The Naive Bayes classifier is assigned as "NB". The k-nearest neighbours classifier is represented as "KNN". "time dummy" model means models with time dummy variables. "industry dummy" represents that dummy variables are applied to capture industry effects.

Table 4-24 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2009_young firms

Decile	DH	DH + time dummy	DH + industry dummy	BMA_DH	BMA_DH + time dummy	BMA_DH + industry dummy
1	22.88	5.88	24.18	22.22	25.49	21.57
2	9.80	1.96	12.42	10.46	17.65	20.26
3	7.84	20.26	9.80	10.46	16.99	10.46
4	9.80	22.88	11.11	11.11	10.46	15.03
5	10.46	12.42	10.46	13.73	5.88	10.46
6-10	39.21	36.59	32.02	32.03	23.53	22.22
AUC	61.86%	56.79%	64.86%	63.72%	69.98%	69.93%

Notes: "DH" means the discrete hazard model. "BMA_DH" represents the BMA version of DH models. "time dummy" model means models with time dummy variables. "industry dummy" represents that dummy variables are applied to capture industry effects.

Table 4-25 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2009_young firms

Decile	NB	NB + time dummy	NB + industry dummy	KNN	KNN + time dummy	KNN + industry dummy
1	38.56	38.56	31.37	13.07	13.07	13.07
2	11.11	11.11	13.07	13.73	13.73	13.73
3	5.88	5.88	17.65	11.76	11.76	11.76
4	8.50	8.50	8.50	20.26	19.61	19.61
5	10.46	10.46	9.80	9.15	9.80	9.80
6-10	25.49	25.49	19.62	32.03	32.03	32.03
AUC	70.87%	70.87%	72.72%	61.47%	61.39%	61.45%

Notes: The Naive Bayes classifier is assigned as "NB". The k-nearest neighbours classifier is represented as "KNN". "time dummy" model means models with time dummy variables. "industry dummy" represents that dummy variables are applied to capture industry effects.

Based on Table 4-23, adding BMA to the DH model improves the AUC value from about 62% to 64%. The predictive performance of the BMA version of DH models cannot outperform that of the NB classifier, but it is better than that of the k-NN classifier in Table 4-23 according to both AUC and brier scores. Similar to sub-samples of firm size, adding time effects does not increase the AUC values or decrease brier scores in the DH model, NB classifier and k-NN classifier. In this sample, industry effects have more power than time effects to increase AUC values for all candidates except the k-NN classifier. This suggests that industry effects are important in prediction of failure in the case of young firms. These results are also supported by decile methods in Table 4-24 and Table 4-25. In the NB classifier, the highest predictive ability can be observed in the first decile and about 80% correct prediction rate in the first five deciles.

Table 4-26 AUC and brier scores for all competing models during 1991-2009_old firms

	AUC	Brier scores		AUC	Brier scores
DH	69.91%	0.0247	NB	74.69%	0.0163
DH + time dummy	58.03%	0.4830	NB + time dummy	74.69%	0.0187
DH + industry dummy	70.19%	0.0247	NB + industry dummy	70.67%	0.0288
BMA_DH	73.56%	0.0203	KNN	64.94%	0.0179
BMA_DH + time dummy	73.56%	0.0203	KNN + time dummy	64.06%	0.0182
BMA_DH + industry dummy	72.87%	0.0203	KNN + industry dummy	64.19%	0.0182

Notes: "DH" means the discrete hazard model. "BMA_DH" represents the BMA version of DH models. The Naive Bayes classifier is assigned as "NB". The k-nearest neighbours classifier is represented as "KNN". "time dummy" model means models with time dummy variables. "industry dummy" represents that dummy variables are applied to capture industry effects.

Table 4-27 Defaults by out-of-sample prediction decile among model related to the DH model during 1991-2009_old firms

Decile	DH	DH + time dummy	DH + industry dummy	BMA_DH	BMA_DH + time dummy	BMA_DH with industry dummy
1	28.18	4.55	29.09	24.55	24.55	24.55
2	12.73	2.73	13.64	17.27	17.27	19.09
3	10.91	19.09	11.82	16.36	16.36	15.45
4	12.73	23.64	10.91	16.36	16.36	13.64
5	10.00	15.45	10.91	10.91	10.91	10.00
6-10	25.46	34.54	23.63	14.55	14.55	17.27
AUC	69.91%	58.03%	70.19%	73.56%	73.56%	72.87%

Notes: "DH" means the discrete hazard model. "BMA_DH" represents the BMA version of DH models. "time dummy" model means models with time dummy variables. "industry dummy" represents that dummy variables are applied to capture industry effects.

Table 4-28 Defaults by out-of-sample prediction decile in the NB classifier and the k-NN classifier during 1991-2009_old firms

Decile	NB	NB + time dummy	NB + industry dummy	KNN	KNN + time dummy	KNN + industry dummy
1	38.18	38.18	28.18	27.27	26.36	26.36
2	12.73	12.73	15.45	11.82	14.55	14.55
3	10.91	10.91	17.27	10.91	3.64	2.73
4	10.00	10.00	13.64	4.55	7.27	12.73
5	10.00	10.00	7.27	13.64	12.73	8.18
6-10	18.19	18.19	18.19	31.81	35.46	35.46
AUC	74.69%	74.69%	70.67%	64.94%	64.06%	64.19%

Notes: The Naive Bayes classifier is assigned as “NB”. The k-nearest neighbours classifier is represented as “KNN”. “time dummy” model means models with time dummy variables. “industry dummy” represents that dummy variables are applied to capture industry effects.

Moving to old firms, it should be noted that the AUC values are generally higher than for young firms when we compare Table 4-23 and Table 4-26. This means that all candidates can more efficiently provide out-of-sample predictions for old firms. The NB classifier consistently provides the highest AUC value among all models. The predictive power of the DH model can be improved by adding BMA. It is interesting to note that adding time or industry effects does not improve predictive ability for all models with reference to old firms. Table 4-27 and Table 4-28 report the percentage of correct prediction in each decile. It can be seen that the highest value of this correct prediction rate in the first decile is in the NB classifier and that in the top five deciles it is in the BMA version of DH models. This is not a very unexpected result since the AUC values are comparable in these two models.

4.5.3.3 Discussion

Considering firm heterogeneities in the dataset, our results indicate that all candidate models can provide higher predictive ability for small or old firms compared to large or young firms. It has been well known that the causes of firms’ failure are different between small (young) and large (old) firms. From these internal and external causes, there exists a similar cause for both small and old firms’ failure, which is the change in market conditions. Sipa *et al.* (2015) suggest that the operation of small firms is sensitive to the change in market conditions, since they do not have effective control and plentiful cash flow planning compared with large firms (Charitou *et al.* 2004). Thornhill and Amit (2003) suggest that the failure of old firms is associated with an inability to adjust to market changes compared with younger firms, because old firms already have stable resources and competitive capabilities in an industry and they do not find it easy to change business direction. During the period when the data was applied, when two important financial crises occurred in the UK, small or old firms were more likely to fail than large or young firms because of the change in market conditions. According to our results, the candidate models can provide better predictive performance for small or old firms' failure, which implies that the models applied can be used in future to detect the failure of small or old firms in periods of financial turmoil. Based on the prediction of failure, market participants can attempt rescue strategies in advance to keep economic vitality. In addition, for old and large firms, the predictions of failure are not be clearly affected by considering time or industry effects compared with small and young firms. This confirms that old and large firms have relatively more stable operating ability in comparison to young and small firms.

4.6 Conclusion

The failure of a firm should be carefully forecasted since it is an event which can bring significant wealth losses for market participants and even lead to economic depression. Thus, a reasonable margin of accuracy in failure prediction for firms can bring many benefits for the public. The prediction of failure for private firms in the UK is modelled by the discrete hazard (DH) model, the Bayesian Model Averaging (BMA) version of DH models, the Naive Bayes Classifier (NB) and the K-nearest neighbour (k-NN) classifier in this chapter. Annual data of firm-specific factors and macroeconomic variables for a period of about twenty years (1991 to 2009) is employed as the input in the benchmark model (the DH model) and all competing models. This model selection not only follows the literature of binary dependent variable models, but also compares predictive performance of the reduce-form models and simple machine learning models.

First, the explanatory ability of financial ratios and macroeconomic predictors in the prediction of failure for the UK firms is confirmed. Second, the predictive performance of the DH model in prediction can be significantly improved by adding the BMA technique. Furthermore, the NB classifier produces higher predictive accuracy than other competing models in the out-of-sample prediction for a firm's failure. The non-strict assumption leads to the greater flexibility of the NB and k-NN classifiers compared to reduced-form models when it comes to capturing the characteristics of the dataset. These classifiers do not however consider parameter uncertainty and model uncertainty. Among all candidate models, the BMA version of DH models is the only one in this study which considers parameter and model uncertainty. The BMA version of DH models can also produce comparable predictions with the NB classifier in some samples. In addition, controlling time dummy or industry dummy variables can provide better predictive performance than without adding them in different periods, especially for the DH model and its BMA version. Moreover, due to the two financial crises covered in the data period, according to the empirical results in different time periods, it can be confirmed that the predictive ability of all candidate models is affected by financial crises, especially the ERM currency crisis. Finally, to capture heterogeneities between firms, we distinguish the dataset into cross-sectional sub-samples based on firm size and age. We observe that the failure of a small or old private firm can be predicted more accurately than that of large or young private firms.

These results suggest that the simple classifier, the NB classifier, still can be widely applied to the prediction of failure, since it does not require a priori knowledge of this method and does not need to satisfy assumptions carefully. Over time, the reasons causing the failure of firms will change, which implies that the "best" model would also change. To solve the parameter and model uncertainty, the BMA version of the DH model is a reasonable model selection to improve predictive accuracy in further research.

References

Aha, D. W. (1997) *Lazy Learning*, Dordrecht: Springer Netherlands.

Allison, P. D. (1982) 'Discrete-Time Methods for the Analysis of Event Histories', *Sociological Methodology*, 13, 61-98.

Altman, E. I. (1968) 'Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy', *The Journal of Finance*, 23(4), 589-609.

Altman, E. I. (1993) *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy*, 2nd ed., New York; Chichester: Wiley.

Altman, E. I., Haldeman, R. G. and Narayanan, P. (1977) 'ZETA TM Analysis. A New Model to Identify Bankruptcy Risk of Corporations', *Journal of Banking and Finance*, 1(1), 29-54.

Altman, E. I., Marco, G. and Varetto, F. (1994) 'Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (The Italian Experience)', *Journal of Banking and Finance*, 18(3), 505-529.

Altman, E. I., Sabato, G. and Wilson, N. (2010) 'The Value of Non-Financial Information in Small and Medium-Sized Enterprise Risk Management', *Journal of Credit Risk*, 6(2), 95-127.

Ayyagari, M., Beck, T. and Demirguc-Kunt, A. (2007) 'Small and Medium Enterprises across the Globe', *Small Business Economics*, 29(4), 415-434.

Barboza, F., Kimura, H. and Altman, E. (2017) 'Machine Learning Models and Bankruptcy Prediction', *Expert Systems with Applications*, 83, 405-417.

Bauer, J. and Agarwal, V. (2014) 'Are Hazard Models Superior to Traditional Bankruptcy Prediction Approaches? A Comprehensive Test', *Journal of Banking and Finance*, 40, 432-442.

Beaver, W. H. (1966) 'Financial Ratios As Predictors of Failure', *Journal of Accounting Research*, 4, 71-111.

Beaver, W. H., McNichols, M. F. and Rhie, J.-W. (2005) 'Have Financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy', *Review of Accounting Studies*, 10(1), 93-122.

Beck, T., Demirguc-Kunt, A. and Levine, R. (2005) 'SMEs, Growth, and Poverty: Cross-Country Evidence', *Journal of Economic Growth*, 10(3), 199-229.

Bharath, S. T. and Shumway, T. (2008) 'Forecasting Default with the Merton Distance to Default Model', *The Review of Financial Studies*, 21(3), 1339-1369.

Bhattacharjee, A., Higson, C., Holly, S. and Kattuman, P. (2009) 'Macroeconomic Instability and Business Exit: Determinants of Failures and Acquisitions of UK Firms', *Economica*, 76(301), 108-131.

Black, F. and Scholes, M. (1973) 'The Pricing of Options and Corporate Liabilities', *Journal of Political Economy*, 81(3), 637-654.

Bonfim, D. (2009) 'Credit Risk Drivers: Evaluating the Contribution of Firm Level Information and of Macroeconomic Dynamics', *Journal of Banking and Finance*, 33(2), 281-299.

Bridges, S. and Guariglia, A. (2008) 'Financial Constraints, Global Engagement, and Firm Survival in the United Kingdom: Evidence from Micro Data', *Scottish Journal of Political Economy*, 55(4), 444-464.

Bryant, S. M. (1997) 'A Case-Based Reasoning Approach to Bankruptcy Prediction Modeling', *International Journal of Intelligent Systems in Accounting, Finance & Management*, 6(3), 195-214.

Bunn, P. and Redwood, V. (2003) 'Company Accounts-Based Modelling of Business Failures and the Implications for Financial Stability', *SSRN Electronic Journal*.

Campbell, J. Y., Hilscher, J. and Szilagyi, J. (2008) 'In Search of Distress Risk', *The Journal of Finance*, 63(6), 2899-2939.

Canbas, S., Cabuk, A. and 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(2), 528-546.

Charitou, A., Neophytou, E. and Charalambous, C. (2004) 'Predicting Corporate Failure: Empirical Evidence for the UK', *European Accounting Review*, 13(3), 465-497.

Chava, S. and Jarrow, R. A. (2004) 'Bankruptcy Prediction with Industry Effects', *European Finance Review*, 8(4), 537-569.

Coats, P. K. and Fant, L. F. (1993) 'Recognizing Financial Distress Patterns Using a Neural Network Tool', *Financial Management*, 22(3), 142-155.

Cronan, T. P., Glorfeld, L. W. and Perry, L. G. (1991) 'Production System Development for Expert Systems Using a Recursive Partitioning Induction Approach: An Application to Mortgage, Commercial, and Consumer Lending', *Decision Sciences*, 22(4), 812-845.

Crosbie, P. J. and Bohn, J. R. (2003) *Modeling Default Risk* [online], available: https://business.illinois.edu/gpennacc/MoodysKMV.pdf [Accessed. 27th May, 2017]

Davis, R. H., Edelman, D. B. and Gammerman, A. J. (1992) 'Machine-Learning Algorithms for Credit-Card Applications', *IMA Journal of Management Mathematics*, 4(1), 43-51.

Deakin, E. B. (1972) 'A Discriminant Analysis of Predictors of Business Failure', *Journal of Accounting Research*, 10(1), 167-179.

Dichev, I. D. (1998) 'Is the Risk of Bankruptcy a Systematic Risk?', *The Journal of Finance*, 53(3), 1131-1147.

Ding, A. A., Tian, S., Yu, Y. and Guo, H. (2012) 'A Class of Discrete Transformation Survival Models with Application to Default Probability Prediction', *Journal of the American Statistical Association*, 107(499), 990-1003.

Domingos, P. and Pazzani, M. (1997) 'On the Optimality of the Simple Bayesian Classifier under Zero-One Loss', *Mach. Learn.*, 29(2-3), 103-130.

Duffie, D., Eckner, A., Horel, G. and Saita, L. (2009) 'Frailty Correlated Default', *The Journal of Finance*, 64(5), 2089-2123.

Duffie, D., Saita, L. and Wang, K. (2007) 'Multi-Period Corporate Default Prediction with Stochastic Covariates', *Journal of Financial Economics*, 83(3), 635-665.

Duffie, D. and Singleton, K. J. (1999) 'Modeling Term Structures of Defaultable Bonds', *The Review of Financial Studies*, 12(4), 687-720.

Edmister, R. O. (1972) 'An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction', *Journal of Financial and Quantitative Analysis*, 7(2), 1477-1493.

Etheridge, H. L. and Sriram, R. S. (1997) 'A Comparison of the Relative Costs of Financial Distress Models: Artificial Neural Networks, Logit and Multivariate Discriminant Analysis', *International Journal of Intelligent Systems in Accounting, Finance & Management*, 6(3), 235-248.

Fernández, C., Ley, E. and Steel, M. F. J. (2001) 'Benchmark Priors for Bayesian Model Averaging', *Journal of Econometrics*, 100(2), 381-427.

FitzPatrick, P. J. (1932) *A Comparison of the Ratios of Successful Industrial Enterprises with Those of Failed Companies*, [Washington].

Fragoso, T. M., Bertoli, W. and Louzada, F. (2018) 'Bayesian Model Averaging: A Systematic Review and Conceptual Classification: BMA: A Systematic Review', *International Statistical Review*, 86(1), 1-28.

Frydman, H., Altman, E. I. and Kao, D.-L. (1985) 'Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress', *The Journal of Finance*, 40(1), 269-291.

Geng, R., Bose, I. and Chen, X. (2015) 'Prediction of Financial Distress: An Empirical Study of Listed Chinese Companies Using Data Mining', *European Journal of Operational Research*, 241(1), 236-247.

Görg, H. and Spaliara, M.-E. (2018) 'Export Market Exit and Financial Health in Crises Periods', *Journal of Banking and Finance*, 87, 150-163.

Griffin, J. M. and Lemmon, M. L. (2002) 'Book-to-Market Equity, Distress Risk, and Stock Returns', *The Journal of Finance*, 57(5), 2317-2336.

Guariglia, A., Spaliara, M. E. and Tsoukas, S. (2016) 'To What Extent Does the Interest Burden Affect Firm Survival? Evidence from a Panel of UK Firms during the Recent Financial Crisis', *Oxford Bulletin of Economics and Statistics*, 78(4), 576-594.

Han, J., Kamber, M. and Pei, J. (2011) *Data Mining: Concepts and Techniques*, 3rd ed., Burlington: Elsevier Science.

Hand, D. J. and Vinciotti, V. (2003) 'Choosing K for Two-Class Nearest Neighbour Classifiers with Unbalanced Classes', *Pattern Recognition Letters*, 24(9-10), 1555-1562.

Henley, W. E. and Hand, D. J. (1996) 'A k-Nearest-Neighbour Classifier for Assessing Consumer Credit Risk', *Journal of the Royal Statistical Society. Series D (The Statistician)*, 45(1), 77-95.

Hillegeist, S. A., Keating, E. K., Cram, D. P. and Lundstedt, K. G. (2004) 'Assessing the Probability of Bankruptcy', *Review of Accounting Studies*, 9(1), 5-34.

Islam, M. J., Wu, Q. M. J., Ahmadi, M. and SidAhmed, M. A. (2010) 'Investigating the Performance of Naive-Bayes Classifiers and K- Nearest Neighbor Classifiers', *Journal of Convergence Information Technology*, 5(2), 133-137.

Jarrow, R. A. and Turnbull, S. M. (1995) 'Pricing Derivatives on Financial Securities Subject to Credit Risk', *The Journal of Finance*, 50(1), 53-85.

Javorcik, B. S. and Li, Y. (2013) 'Do the Biggest Aisles Serve a Brighter Future? Global Retail Chains and Their Implications for Romania', *Journal of International Economics*, 90(2), 348-363.

Jiménez, G. and Saurina, J. (2004) 'Collateral, Type of Lender and Relationship Banking as Determinants of Credit Risk', *Journal of Banking and Finance*, 28(9), 2191-2212.

Jo, H., Han, I. and Lee, H. (1997) 'Bankruptcy Prediction Using Case-Based Reasoning, Neural Networks, and Discriminant Analysis', *Expert Systems with Applications*, 13(2), 97-108.

Jones, S. and Hensher, D. A. (2004) 'Predicting Firm Financial Distress: A Mixed Logit Model', *The Accounting Review*, 79(4), 1011-1038.

Keasey, K. and McGuinness, P. (1990) 'The Failure of UK Industrial Firms for the Period 1976-1984, Logistic Analysis and Entropy Measures', *Journal of Business Finance and Accounting*, 17(1), 119-135.

Kolari, J., Glennon, D., Shin, H. and Caputo, M. (2002) 'Predicting Large US Commercial Bank Failures', *Journal of Economics and Business*, 54(4), 361-387.

Kuramochi, M. and Karypis, G. (2005) 'Gene Classification Using Expression Profiles: A Feasibility Study', *International Journal on Artificial Intelligence Tools*, 14(4), 641-660.

Lacher, R. C., Coats, P. K., Sharma, S. C. and Fant, L. F. (1995) 'A Neural Network for Classifying the Financial Health of a Firm', *European Journal of Operational Research*, 85(1), 53-65.

Lang, L. H. P. and Stulz, R. (1992) 'Contagion and Competitive Intra-Industry Effects of Bankruptcy Announcements: An Empirical Analysis', *Journal of Financial Economics*, 32(1), 45-60.

Lee, K. C., Han, I. and Kwon, Y. (1996) 'Hybrid Neural Network Models for Bankruptcy Predictions', *Decision Support Systems*, 18(1), 63.

Libby, R. (1975) 'Accounting Ratios and the Prediction of Failure: Some Behavioral Evidence', *Journal of Accounting Research*, 13(1), 150-161.

Madigan, D. and Raftery, A. E. (1994) 'Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window', *Journal of the American Statistical Association*, 89(428), 1535-1546.

Margaritis, D. and Psillaki, M. (2010) 'Capital Structure, Equity Ownership and Firm Performance', *Journal of Banking and Finance*, 34(3), 621-632.

Merton, R. C. (1974) 'On the Pricing of Corporate Debt: The Risk Structure of Interest Rates', *The Journal of Finance*, 29(2), 449.

Messier, W. F., Jr. and Hansen, J. V. (1988) 'Inducing Rules for Expert System Development: An Example Using Default and Bankruptcy Data', *Management Science*, 34(12), 1403-1415.

Min, J. H. and Lee, Y.-C. (2005) 'Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters', *Expert Systems with Applications*, 28(4), 603-614.

Murphy, K. P. (2012) *Machine Learning: A Probabilistic Perspective*, London; Cambridge, Mass: MIT Press.

Nanda, S. and Pendharkar, P. (2001) 'Linear Models for Minimizing Misclassification Costs In Bankruptcy Prediction', *Intelligent Systems in Accounting, Finance & Management*, 10(3), 155-168.

Ohlson, J. A. (1980) 'Financial Ratios and the Probabilistic Prediction of Bankruptcy', *Journal of Accounting Research*, 18(1), 109-131.

Park, C.-S. and Han, I. (2002) 'A Case-Based Reasoning with the Feature Weights Derived by Analytic Hierarchy Process for Bankruptcy Prediction', *Expert Systems with Applications*, 23(3), 255-264.

Pesaran, M. H., Schuermann, T., Treutler, B.-J. and Weiner, S. M. (2006) 'Macroeconomic Dynamics and Credit Risk: A Global Perspective', *Journal of Money, Credit and Banking*, 38(5), 1211-1261.

Queen, M. and Roll, R. (1987) 'Firm Mortality: Using Market Indicators to Predict Survival', *Financial Analysts Journal*, 43(3), 9-26.

Rabe-Hesketh, S. and Skrondal, A. (2012) *Multilevel And Longitudinal Modeling Using Stata*, College Station, Tex.: Stata Press.

Ravi Kumar, P. and Ravi, V. (2007) 'Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques – A Review', *European Journal of Operational Research*, 180(1), 1-28.

Roberts, H. V. (1965) 'Probabilistic Prediction', *Journal of the American Statistical Association*, 60(309), 50-62.

Rodriguez, J. D., Perez, A. and Lozano, J. A. (2010) 'Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), 569-575.

Ropega, J. (2011) 'The Reasons and Symptoms of Failure in SME', *International Advances in Economic Research*, 17(4), 476-483.

Sarkar, S. and Sriram, R. S. (2001) 'Bayesian Models for Early Warning of Bank Failures', *Management Science*, 47(11), 1457-1475.

Scott, J. (1981) 'The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models', *Journal of Banking & Finance*, 5(3), 317-344.

Shin, K.-S. and Lee, Y.-J. (2002) 'A Genetic Algorithm Application in Bankruptcy Prediction Modeling', *Expert Systems with Applications*, 23(3), 321-328.

Shumway, T. (2001) 'Forecasting Bankruptcy More Accurately: A Simple Hazard Model', *The Journal of Business*, 74(1), 101-124.

Sipa, M., Gorzeń-Mitka, I. and Skibiński, A. (2015) 'Determinants of Competitiveness of Small Enterprises: Polish Perspective', *Procedia Economics and Finance Procedia Economics and Finance*, 27, 445-453.

Smith, R. F. and Winakor, A. H. (1935) *Changes in the Financial Structure of Unsuccessful Industrial Corporations*, Urbana: University of Illinois.

Tam, K. Y. (1991) 'Neural Network Models and the Prediction of Bank Bankruptcy', *Omega*, 19(5), 429-445.

Tennyson, B. M., Ingram, R. W. and Dugan, M. T. (1990) 'Assessing the Information Content of Narrative Disclosures in Explaining Bankruptcy', *Journal of Business Finance & Accounting*, 17(3), 391-410.

Thornhill, S. and Amit, R. (2003) 'Learning About Failure: Bankruptcy, Firm Age, and the Resource-Based View', *Organization Science*, 14(5), 497-509.

Tian, S., Yu, Y. and Guo, H. (2015) 'Variable Selection and Corporate Bankruptcy Forecasts', *Journal of Banking and Finance*, 52, 89-100.

Traczynski, J. (2017) 'Firm Default Prediction: A Bayesian Model-Averaging Approach', *Journal of Financial and Quantitative Analysis*, 52(3), 1211-1245.

Twala, B. (2010) 'Multiple Classifier Application to Credit Risk Assessment', *Expert Systems with Applications*, 37(4), 3326-3336.

Udo, G. (1993) 'Neural Network Performance on the Bankruptcy Classification Problem', *Computers & Industrial Engineering*, 25(1), 377-380.

Varetto, F. (1998) 'Genetic Algorithms Applications in the Analysis of Insolvency Risk', *Journal of Banking and Finance*, 22(10), 1421-1439.

Wilcox, J. W. (1973) 'A Prediction of Business Failure Using Accounting Data', *Journal of Accounting Research*, 11, 163-179.

Wilson, R. L. and Sharda, R. (1994) 'Bankruptcy Prediction Using Neural Networks', *Decision Support Systems*, 11(5), 545.

Yang, Z. R., Platt, H. D. and Platt, M. B. (1999) 'Probabilistic Neural Networks in Bankruptcy Prediction', *Journal of Business Research*, 44(2), 67-74.

Zavgren, C. V. (1985) 'Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis', *Journal of Business Finance & Accounting*, 12(1), 19-45.

Zhou, L. G., Lai, K. K. and Yen, J. (2014) 'Bankruptcy Prediction Using SVM Models with a New Approach to Combine Features Selection and Parameter Optimisation', *International Journal of Systems Science*, 45(3), 241-253.

Zmijewski, M. E. (1984) 'Methodological Issues Related to the Estimation of Financial Distress Prediction Models', *Journal of Accounting Research*, 22(2), 59-82.

Chapter 5 Conclusion

The market implied rating, IPO decisions and failure are important events in a firm's operation, which are associated with its financial health and the economic environment. The accurate prediction of such events heavily influences the identification of potential investment opportunities and the reduction of credit risk for market participants, which are already achieved by reduced-form models in finance. To achieve more flexibility and more accurate prediction, the properties of variable selection techniques (LASSO techniques and model averaging approaches) are employed in this thesis.

In Chapter 2, we made a methodological contribution with regards to the improvement of predictive power through applying variable selection techniques (the least absolute shrinkage and selection operator, LASSO, and its derivation, the Elastic net) to reduced-form models in order to forecast market implied ratings. Since market implied ratings updated in a timely fashion can reduce the outdated nature of ratings compared with long-term ratings, better forecasts of market implied ratings can be considered as more accurate and early signals of the change in credit risk. During the forecast exercise in Chapter 2, an extensive dataset containing firm-specific indicators, market-driven variables and macroeconomic factors were considered as potential predictors to determine and predict the market implied ratings extracted from Fitch's database from 2002 to 2008. Our results corroborate that LASSO models are able to select a small set of relevant predictors from the vast potential predictors, which suggests that market implied ratings are related to financial and business risks. It also implies that the sparse representation generated can in future be applied directly. Next, the predictive performances in the out-of-sample for LASSO models are clearly superior to those of the ordered probit model, which has mostly been adopted in the literature. This provides sufficient evidence that LASSO models can produce improved predictive ability, which can be implemented widely in subsequent research. Finally, we also confirm that the BIC-type tuning parameter selector can successfully use information from fewer predictors and then provide more accurate out-of-sample predictions than their counterparts with AIC-type selector.

Chapter 3 focused on applying LASSO to model binary events such as IPO decisions. The operating status of banks is linked to the entire economic efficiency. Making correct predictions of IPO decisions can be regarded as a way of assessing the health of banks. To investigate the probability of banks deciding to go public by IPO issuance, we applied bank-specific accounting ratios and other publicly available information in the semi-parametric Cox proportional hazard model, which is fundamental in the literature. We then extended this Cox model to its L1 penalized versions by adding LASSO. Our results illustrate that a significant improvement in the predictive ability can be observed in the L1 penalized semi-parametric Cox proportional hazard model incorporating fewer predictors compared to the standard frameworks used in the literature. To validate our conclusions, discrete hazard model, logistic model and their corresponding L1 penalized models were implemented. They were however unable to perform better than the L1 penalized semi-parametric Cox proportional hazard model in out-of-sample predictions of IPO decisions for banks. This evidence further supports the argument that adding LASSO into reduce-form models can improve predictive ability. To explore the sensitivity of our findings to different economic conditions, we divided our sample into three parts: the pre-crisis period (1996-2006), the crisis period (2007-2009) and the post-crisis period (2010-2016). It should be noted that the sensitivity of a bank's IPO to financial characteristics was higher during the global financial crisis. Chapters 2 and 3 suggest that LASSO models can consistently produce improved predictive performance and simultaneously identify the most important predictors, which can be recommended in the forecast of both ordinal and binary financial outcomes in the future.

In another attempt to produce more accurate predictions, in Chapter 4 we applied model averaging technique to predict the failure of private firms in the UK. Compared with listed firms, these unlisted firms have become a new powerhouse to boost the economy. Providing reasonable accuracy of prediction can help market participants identify potential credit risk and reduce loss. We employed accounting ratios and macroeconomic indicators in the discrete hazard (DH) model, the Bayesian Model Averaging (BMA) version of discrete hazard models, Naive Bayes (NB) classifier and k-nearest neighbour (k-NN) classifier to forecast the failure of private firms. Based on our results, we suggest that adding the BMA approach can improve the predictive ability of the discrete hazard model. This means that the BMA version of DH models can perform better than the discrete hazard model and solve the parameter and model uncertainty at the same time. If time dummy or industry dummy variables are added, the accuracy ratios in out-of-sample prediction are increased in our candidate models, especially for discrete hazard models and their BMA version. This means that time and industry effects should be considered in the forecast of failure. Furthermore, financial crisis can affect the predictive ability of all candidate models. Our results also indicate that firm heterogeneity is another factor affecting the predictive power for each candidate. We finally suggested that the BMA version of discrete hazard models and Naive Bayes Classifier can produce a comparably higher percentage of correct failure prediction. Both can be applied in failure prediction, while only the BMA version of discrete hazard models solves parameter and model uncertainty.

In consequence, variable selection techniques such as LASSO or Bayesian Model Averaging can be combined in the reduced-form models to improve predictive performance and identify the most relevant predictors when an extensive set of predictors are used. Due to the flexibility of these variable selection techniques, the changed structure of financial outcomes can be captured over time. In future research, this work may be developed through several topics. One approach is to identify the important indicators by LASSO or Bayesian Model Averaging in a structural model and then predict an important event in the course of firms' operation such as the issuance of bonds. In addition, it is possible to apply different types of variable selection techniques in other reduced-form models with continuous dependent variables and then examine the predictive performance or investigate the important determinants.