Macroeconomic Conditions, Corporate Defaults, and Economic Recessions

Kai Xing

MSc Newcastle University, United Kingdom
BSc Jilin Huaqiao University of Foreign Languages, People's Republic of China

Thesis submitted to The University of Nottingham

for the degree of Doctor of Philosophy

September 2016

Acknowledgements

First and foremost, I would like to thank my two supervisors: Professor David P. Newton and Professor Michael Tretyakov. Their supervision styles have kept me calm and able to carry on my research work. They never gave up on me and always supported me, consciously and unconsciously teaching me how good research is done. I greatly appreciate all their contributions of time, ideas, and sources of learning to make my Ph.D. experience productive and inspiring.

In particular, I recognise my poor background from my bachelors degree. It seemed really hard to get an offer from our internationally recognised university, even though I had an MSc with distinction in Banking and Finance from Newcastle University, UK. However, David thought that I had some potential in research, and he gave me an offer. He made one of my dreams, which was to study in a world top 1% university, come true. He also helped our research group to buy a very expensive Moody's default database, which enabled all the research ideas to be implemented. Otherwise, I might have been unable to work in the area of credit risk management. I thank him for his strong support in finishing my first working paper during my Ph.D. life.

It has been an honour to be Michael's first Chinese Ph.D. student. I could not even have imagined that I would have this great opportunity to work with as successful a mathematician and professor. He led me to acquire good programming skills and good knowledge of R languages and he helped me to access the Shiny server in the School of Mathematical Sciences in order to let me feel free to develop my own applications in support of my research. Otherwise, I could not have completed my current achievements from the IT perspective. Thank you to David and Michael for being my PhD supervisors.

I am also deeply indebted to Dr Wenfeng Duan and Dr Hanxiong Zhang. I met Wenfeng when I was a postgraduate student at Newcastle University Business School. He gave selfless support and helped with my Ph.D. application. I cannot even count how much time he spent

on patiently revising my application documents and sharing his research experience with me. He was my first mentor leading up to my getting the Ph.D. offer from Nottingham University Business School. Hanxiong, a lecturer at Lincoln University Business School has been supportive since I started my Ph.D. study. He helped me come up with the structure of my first working paper and guided me to revise and polish it. Even now, both Wenfeng and Hanxiong still regularly give me moral and academic support. I very much appreciate what they have done for me; so thank you Wenfeng and Hanxiong, who became big brothers in my life.

Dr James Water has been a mentor during my Ph.D. study. He has been really helpful with my research. I remember how he would sometimes come to my office to ask how my research was going. He never seemed bored when I shared my research ideas with him. Even now, when he is a post-doctoral fellow at the Warwick University Business School, he still emails me from time to time about recent progress on my work. I feel very thankful to him indeed.

Two special people are Dr Shamim Ahmed and Dr Thanaset Chevapatrakul. They figured out the mistakes in the results of my second empirical chapter. They also gave me valuable suggestions about how to verify whether or not my current results are right and how to polish the chapter in order to get it published in economic journals. Thank you to both of you.

My time at Nottingham has been made enjoyable in large part due to the many friends and groups that became a part of my life. I wish to thank my friends, Miss Silvia Cinque, Mr Shukhrat Nasirov, Mr Aristeidis Dadoukis, Miss Claire Jane Ingram, Miss Xiaozhao Wang, Mr Zhongxiang Xu, Miss Dan Li, and my previous flatmates, Mr Xiao Ding, Mr Zilin Chen, Mr Hao Sun, Mr Xiaoping Jiang, Mr Xianxu Hou. I am grateful for time spent with you guys. Thanks to everyone again for your selfless ways, your encouragement and for helping me during times when I needed it. My time at Nottingham was also enriched by my office mates (B44 in the Business School South Building). Meanwhile, my appreciation also goes to my colleagues in the Business School, for their help in my research and daily life.

I would like to express my gratitude to two visiting senior academics from Shanxi University to our Business School, Professor Xindong Zhang and Professor Weiqi Liu; I met them a year ago. During this period, they have been interested in my work and have also given me

a big picture of the current development of the academic world in China. I would like to thank Xindong for being one of my referees when I applied to become a post-doctoral fellow at the Academy of Mathematical and Systems Sciences in the Chinese Academy of Sciences.

Lastly, I would like to thank my family for all their love and encouragement. They raised me with a love of science and supported me in all my pursuits. I promise that I will keep being myself and being a good man to society. I also believe that my efforts will let you continually feel proud for my future achievements. Thank you.

Abstract

In this thesis, I investigate effective predictors for corporate defaults and measurement of economic recessions. I use corporate default probabilities in US industrial firms from April 1980 to December 2014 and effective predictors extracted from 92 macroeconomic and financial variables and I propose a framework to determine whether there is a dynamic in effective predictors for US corporate defaults.

I apply LASSO, an advanced variable selection technique, and I find that there are no macro factors that can consistently explain default risk over time, suggesting that default risk has been dynamic during the 35-year period. These dynamics persist even during non-recession periods. Interestingly, the strong predictive powers of macro factors over shorter periods are related to monetary policy indicators and tools, which provides empirical support for prior theoretical theories that emphasize the unique role of monetary policies in corporate defaults.

Another interesting finding is that institutional market funding, as the liquidity source provided by non-bank groups, has gradually impacted on corporate defaults since the 2001 recession. In the financial crisis occuring in 2007 this money market funds (MMFs) have strongest impact on corproate defaults. This implies that MMFs paly a crucial role of destabilizing credit markets in recent years, which is consistent with current studies of whether MMFs can result in financial stability.

I study US economic recessions by introducing a cutting-edged technique from the Natural Sciences in order to capture the critical transitions in the macroeconomic system, using hundreds of macroeconomic and financial variables covering the period January 1980 to December 2014. Based on this method, I construct macro indicators to measure the interactions among these variables in order to capture the critical transitions in the macroeconomic system. Then I employ a standard logit model to study whether these proposed indicators offer a prediction one-month ahead of US economic recessions from September 1980 to December 2014.

Abstract

I find that the interactions among macroeconomic and financial variables measured by covariance among these variables can provide one-month ahead prediction for US economic recessions. In particular, the best predictive macro indicators are constructed by employing procyclical factors and factors from six economic groups based on results from both in-sample and out-of-sample analysis. Regarding the standard ideal indicator defined by Shiskin and Moore (1967), the indicator constructed by procyclical factors is preferable over that constructed by the factors from six economic groups since the former is smoother. I also find that the threshold based on 25% for classifying the recessions can generate better estimation results than using 50%, consistent with prior studies. In forecasting economic recessions in the US, the indicator generated by using broad macro factors is able to provide predictive power. The implication of these results is to provide a new quantitative approach for central bankers and policy makers to predict economic recessions one month ahead.

Table of Contents

Acknowledgements **2**

Abstract **5**

List of Tables **11**

List of Figures **13**

Abbreviations **16**

CHAPTER 1

1. Introduction **17**

1.1 Motivations 20

1.1.1 Investigating Effective Predictors (EPs) for Corporate Defaults 20

1.1.2 Investigating the Relationship between the Dynamic of Various Macro Indicators and Economic Recessions 22

1.2 Research Questions 22

1.3 Thesis Structure 23

CHAPTER 2

2. Literature Review **25**

2.1 Introduction 25

2.2 Review of the Literature on Default Risk Modelling 26

2.2.1 Background of Research on Default Risk Management 26

2.2.2 Overview of Models 29

2.2.3 Reduced Form Models 30

2.2.4 Structural Models 34

2.2.5 Comparisons between Reduced Form Models and Structural Models 36

2.3 Review of the Literature on Critical Transitions in the Complex System.....38

2.3.1 Theoretical Perspective 38

2.3.2 Empirical Perspective.....41

2.4 Review of the Literature on Measurement of Economic Recession.....43

2.4.1 Capturing Business Cycle by Business Cycle Indicators 43

2.4.2 Predicting Economic Recessions.....44

2.5 Conclusion 48

CHAPTER 3

3. Methodologies and New Developed Analytical Tools 49

3.1 Introduction.....49

3.2 Methodologies.....50

3.2.1 Hodrick-Prescott (HP) Filter 50

3.2.2 Least Absolute Shrinkage and Selection Operator (LASSO).....51

3.2.3 Logistic Regression.....54

3.2.4 LASSO based on Logistic Regression 59

3.3 New Developed Analytical Tools 61

3.3.1 R Package "Kai.R.Tools".....61

3.3.2 Shiny Web Applications.....64

3.4 Conclusion 65

CHAPTER 4

4. Data 66

4.1 Introduction.....66

4.2 Corporate Default Database 69

4.2.1 Definition of Defaults in Moody's 69

4.2.2 Method of Calculation of IDPs 70

4.2.3 Cleaning Methods 70

4.3 Macroeconomic Database 72

4.4 Microeconomic Database.....115

4.5 Conclusion 125

CHAPTER 5

5. Investigating Effective Predictors (EPs) for Corporate Defaults 126

5.1 Introduction 126

5.2 Data 131

5.3 Empirical Analysis and Discussions 139

5.3.1 Identification of Dynamic in EPs 139

5.3.2 Exploration of Key EPs 143

5.4 Robustness Testing 171

5.5 Conclusion 175

CHAPTER 6

6. Investigating the Relationship between the Dynamic of Various Macro Indicators and Economic Recessions 177

6.1 Introduction 177

6.2 Data 181

6.2.1 Construction of Six Macro Indicators 186

6.3 Empirical Analysis 193

6.3.1 Graphical Analysis 193

6.3.2 Logistic Regression Analysis 199

6.3.3 Robustness Testing 203

6.4 Discussions 213

6.5 Conclusion 220

CHAPTER 7

7. Conclusion 222

7.1 Conclusions and Implications 222

7.1.1 First Empirical Chapter 222

7.1.2 Second Empirical Chapter 223

7.2 Limitations 224

7.3 Future Research 225

7.3.1 First Direction 225

7.3.2 Second Direction 226

8. References 227

9. Appendix 237

9.1 Supplemented Document for Chapter 5 237

9.2 Supplemented Document for Second Empirical Chapter 239

9.2.1 In-sample Forecasts for Economic Recessions 239

9.2.2 Out-of-sample Forecasts for Economic Recessions 242

9.3 User's Manual of R Package "Kai.R.Tools" 244

9.3.1 Functionalities of Construction of Issuer-weighted Default Probabilities (IDP) 247

9.3.2 Algorithm of IDP and Data File Description 248

9.3.3 Functionalities of Construction of Value-weighted Default Probabilities (VDP) 264

9.3.4 Functionalities of Construction of Internal Risk Drivers 280

9.3.5 Other key functions for empirical analysis 289

9.4 User's Manual of Ph.D. Project Web Applications: KAI.APP.PhD/KAI.APP.EIA 293

9.4.1 1. GENERAL INFORMATION 297

9.4.2 2. APPLICATIONS SUMMARY 300

9.4.3 3. GETTING STARTED 303

List of Tables

Table 4-1 Summary statistics for 114 macro factors....74

Table 4-2 Detail description of macro factors....80

Table 4-3 Summary of missing values for each variable....116

Table 4-4 Data description for micro factors....121

Table 4-5 Summary statistics for 35 micro factors....122

Table 5-1 Summary of previous studies about modelling corporate default rates with various risk factors....129

Table 5-2 Data description of macro factors....133

Table 5-3 The strength of explanation power of EPs....143

Table 5-4 Occurrence of key EPs....145

Table 5-5 Complete results of EPs in Table 5-4....147

Table 5-6 Key EPs without merging microeconomic database as previous studies....173

Table 6-1 Historical recession from 1980 to 2014 in the US....181

Table 6-2 Data description for macro factors....182

Table 6-3 Effective factors for economic recession....191

Table 6-4 Summary statistics of 6 macro indicators....192

Table 6-5 Logistic regression of economic recessions in the full sample from September 1980 to December 2014 - Odds ratios....204

Table 6-6 Measures of goodness of fit tests for the models in Table 6-4....205

Table 6-7 In-sample estimation results for various probit models: percentage of incorrect prediction of recession....206

Table 6-8 Logistic regression of economic recessions in the first half of full sample from September 1980 to October 1997 - Odds ratios....208

Table 6-9 Measures of goodness of fit tests for the models in Table 6-7....209

Table 6-10 Out-of-sample estimation results for various probit models: Percentage of incorrect prediction of recession....210

Table 6-11 Iterative out-of-sample estimation results for various probit models: percentage of incorrect prediction of recession....212

Table 9-1 EPs in different subsamples from April 1980 to December 2014....238

Table 9-2 Relevant files for calculating IDP in DRD....248

Table 9-3 Relevant files for calculating VDP in DRD.....265

Table 9-4 Data description for 35 internal risk drivers 281

.....

List of Figures

Figure 1-1 Risks Interconnection Map 2011 illustrating systemic interdependencies in the hyper-connected world we are living in 18

Figure 1-2 Time series of monthly default data in the US industrial firms from April 1980 to December 2014 21

Figure 2-1 Diagram of default risk 28

Figure 2-2 Different classes of generic observations that can be used to indicate the potential for critical transitions in a complex system 41

Figure 2-3 Correlations and variance in crisis 41

Figure 3-1 Diagram of the methodologies 60

Figure 3-2 Structure of R package titled “Kai.R.Tools” 63

Figure 4-1 Relation graph among the three databases 68

Figure 4-2 Correlation heat map for 114 macro factors 79

Figure 4-3 Plot for the missing values 117

Figure 4-4 Correlation heat map for 35 Micro factors 124

Figure 5-1 Time series of monthly default data in the US industrial firms from April 1980 to December 2014 136

Figure 5-2 Changes in the EPs extracted in 54 intervals 141

Figure 5-3 Changing ratios of EPs extracted in 54 intervals 142

Figure 5-4 Dynamic change of specific EP on explaining IDPs (1) 154

Figure 5-5 Dynamic change of specific EP on explaining IDPs (2) 157

Figure 5-6 Dynamic change of specific EP on explaining IDPs (3) 161

Figure 5-7 Dynamic change of specific EP on explaining IDPs (4) 163

Figure 5-8 Dynamic change of specific EP on explaining IDPs (5) 169

Figure 5-9 Dynamic change of specific EP on explaining IDPs (6) 171

Figure 6-1 Dynamic change of recession indicators in red lines and economic recessions in green bar from 1980 to 2014 197

Figure 6-2 ROC curve for the logistic model as classifiers on the data from September 1980 to December 2014 206

Figure 6-3 In-sample recession probabilities implied by the probit model, with selective macro indicators based on classification rate 25% 207

Figure 6-4 ROC curve for the logistic model as classifiers on the data from September 1980 to October 1997 210

Figure 6-5 Out-of-sample recession probabilities implied by the probit model, with selective macro indicators based on classification rate 25% 211

Figure 9-1 In-sample forecasts by using various macro indicators 239

Figure 9-2 Out-of-sample forecasts by using various macro indicators. The forecast starts from January 1998 to December 2014 242

Figure 9-3 Flow chat of web apps 300

Figure 9-4 Structure of Inputs for the apps 301

Figure 9-5 Structure of the KAI.APP.PhD 304

Figure 9-6 Appearance of the KAI.APP.PhD 305

Figure 9-7 Detail description of "Area: Uploading the file" in Figure 9-6 306

Figure 9-8 Uploading the file and reading the file 306

Figure 9-9 Structure of Tab 1 and Tab 2 307

Figure 9-10 Appearance and outputs of Tab 2.1 308

Figure 9-11 Selection of the group of factors 309

Figure 9-12 Structure of Tab 3 311

Figure 9-13 Appearance and outputs of Tab 3.1 312

Figure 9-14 Construction of the indicator and adjustment of both indicator and IDP 313

Figure 9-15 Work with Tab 3.2 (1) 314

Figure 9-16 Work with Tab 3.2 (2) 315

Figure 9-17 Work with Tab 3.2 (3) 316

Figure 9-18 Work with Tab 3.2 (4) 317

Figure 9-19 Work with Tab 3.2 (5) 318

Figure 9-20 Work with Tab 3.3 (1) 319

Figure 9-21 Work with Tab 3.3 (2) 320

Figure 9-22 Work with Tab 3.3 (3) 321

Figure 9-23 Work with Tab 3.3 (4) 322

Figure 9-24 Work with Tab 3.3 (5) 323

Figure 9-25 Structure of Tab 4 324

Figure 9-26 Implementation of LASSO 324

Figure 9-27 Structure of the KAI.APP.EIA 326

Figure 9-28 Appearance of the KAI.APP.EIA 327

Figure 9-29 Using LASSO to extract effective factors 330

Figure 9-30 Structure of Tab 4.1 – 4.3.....331

Figure 9-31 Investigation of dynamic in the effective factors in the full sample by Tab 4.1 (1).....332

Figure 9-32 Investigation of dynamic in the effective factors in the full sample by Tab 4.1 (2).....333

Figure 9-33 Investigation of dynamic in the effective factors in the full sample by Tab 4.1 (3).....334

Figure 9-34 Investigation of dynamic in the effective factors in the 8 sub-samples by Tab 4.2 335

Figure 9-35 Investigation of dynamic in the spcific factor in the 8 sub-samples by Tab 4.3 336

Abbreviations

The following table describes various abbreviations and acronyms used throughout this thesis.

Abbreviation	Explanation
DPs	Default probabilities
IDPs	Issuer-weighted default probabilities
VDPs	Value-weighted default probabilities
FRED	Federal Reserve Economic Database
EPs	Effective predictors
NBER	National Bureau of Economic Research
ECRI	Economic Cycle Research Institute
BCBS	Basel Committee on Banking Supervision
IRB	Internal ratings-based
DTD	Distance to default
RBC	Real Business Cycle
AMH	Adaptive market hypothesis
CB	Conference Board
CEI	Index of coincident economic indicators
LEI	Index of leading economic indicators
EMH	Efficient market hypothesis
AMH	Adaptive market hypothesis
VAR	Vector autoregression
VECM	Vector error correction model
BSM	Black Scholes Merton
CEI	Index of coincident economic indicators
LEI	Index of leading economic indicators
PMI	Purchasing manager's index
RBC	Real business cycle
CCM	Compustat Merged Database
DRD	Default Decovery Database
CRSP	Centre for Research in Security Prices
FISD	Fixed Income Securities Database
LASSO	Least absolute shrinkage and selection operator
HP	Hodrick-Prescott
CRAs	Credit rating agencies
MMFs	Money market funds
MSE	Mean square error
OLS	Ordinary least squares
KPSS	Kwiatkowski-Phillips-Schmidt-Shin test
ADF	Augmented Dickey–Fuller test
ML	Maximum liklihood
ROC	Receiver operating characteristic curve
AUC	Area under the curve
OTC	Over-the-counter
CDs	Certificates of deposit

CHAPTER 1

1. Introduction

My thesis focuses on two main topics: exploration of the dynamics in effective default risk predictors of corporate defaults and measurement of US economic recessions by using macro indicators.

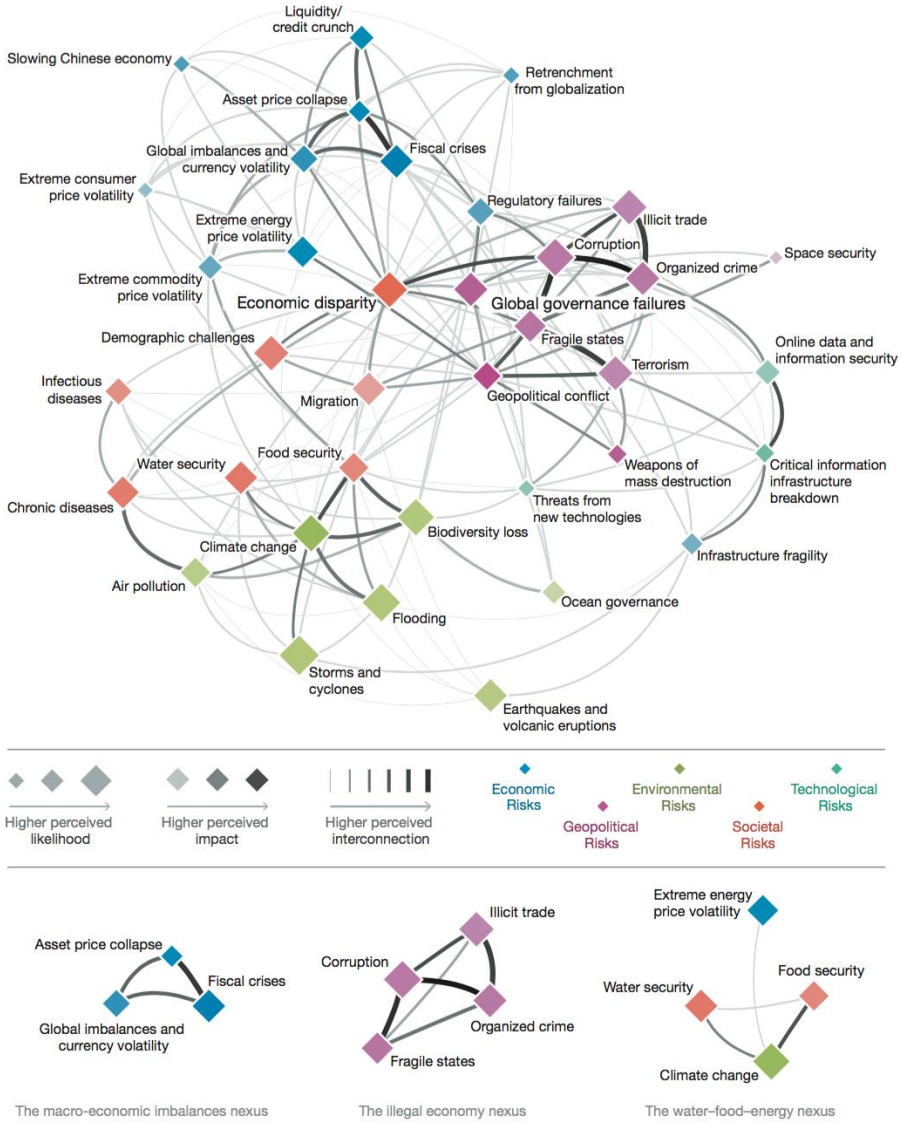
My study is motivated by Helbing (2013)'s paper *Globally Networked Risk and How to Respond*. He notes that:

> *Today's strongly connected, global networks have produced highly interdependent systems that we do not understand and cannot control well. These systems are vulnerable to failure at all scales, posing serious threats to society, even when external shocks are absent. As the complexity and interaction strengths in our networked world increase, man-made systems can become unstable, creating uncontrollable situations even when decision-makers are well-skilled, have all data and technology at their disposal, and do their best. To make these systems manageable, a fundamental redesign is needed. A 'Global Systems Science' might create the required knowledge and paradigm shift in thinking.*

His study figures why numerous advanced and sophisticated techniques fail to predict unwanted collapse in the various systems such as financial systems, and macroeconomic systems. The main reason is that we are living in a hyper-connected world and we cannot understand it very well. Specifically, there are many sub-systems in this world which are highly connected with each other. Figure 1-1 shows how each component in the hyper-connected world is connected with each other. He argue that the underlying networks result

in increased systemic risks. These systemic risks mean that there are strong cascade effects existing in the networks, implying that the dangerous and damaging events can spread rapidly and globally along different paths. As he explained, global air traffic leads to today's rapid spreading of emergent epidemics, which may have negative effects on our global health, social and economic systems. The major implication of systemic risks is that there are unstable correlations among the system components.

Figure 1-1 Risks Interconnection Map 2011 illustrating systemic interdependencies in the hyper-connected world we are living in



Source: this figure is taken from Helbing (2013)

In practice, current techniques for predicting risks are mainly based on the strict assumption of a well-behaved system without considering the reality of high interactions among systems. Thus, the models used cannot work well. For example, copula methodology cannot provide a reasonable modelling framework for the modern financial derivatives if there is a dynamic in the correlations between these derivatives (Helbing, 2013).

The credit market and the macroeconomy are surely complex systems. Systemic risks in them interact with each other as indicated by Figure 1-1. For example, Bernanke (1993) argues for a link between the credit market and macroeconomy, citing the 1966 credit crunch and the 1980 experiment with selective credit controls. For the credit market, systemic risk is reflected in default clustering, which results in dynamic default risk. A large amount of literature has discussed the macroeconomic environment as one of the main sources generating systemic risk in the credit market (Azizpour et al., 2014).

Helbing (2013) declares that "making progress towards a better understanding of complex systems and systemic risks also depends crucially on the collection of 'big data' (massive amounts of data) and the development of powerful machine learning techniques that allow one to develop and validate realistic explanatory models of interdependent systems." It is fortunate that changes in the macroeconomic environment are reflected by numerous macroeconomic factors from the macroeconomic database. Thus I can use these macroeconomic factors to investigate how the system of the credit market dynamically changes.

For the second topic of predicting economic recessions, a study by Scheffer et al. (2012) provides a way of measuring changes in the complex system through capturing critical transitions via 'big data'. This new approach to measuring a complex macroeconomic system may improve the predictive performance of economic recessions in practice for policymakers and practitioners.

The structure of this chapter is as follows. Section 1.1 shows the motivations for researching the two topics mentioned above, which is followed by the research questions in Section 1.2. Section 1.3 provides the structure of this thesis.

1.1 Motivations

Motivations of researching the two topics discussed in this thesis may be classified as follows:

1.1.1 Investigating Effective Predictors (EPs) for Corporate Defaults

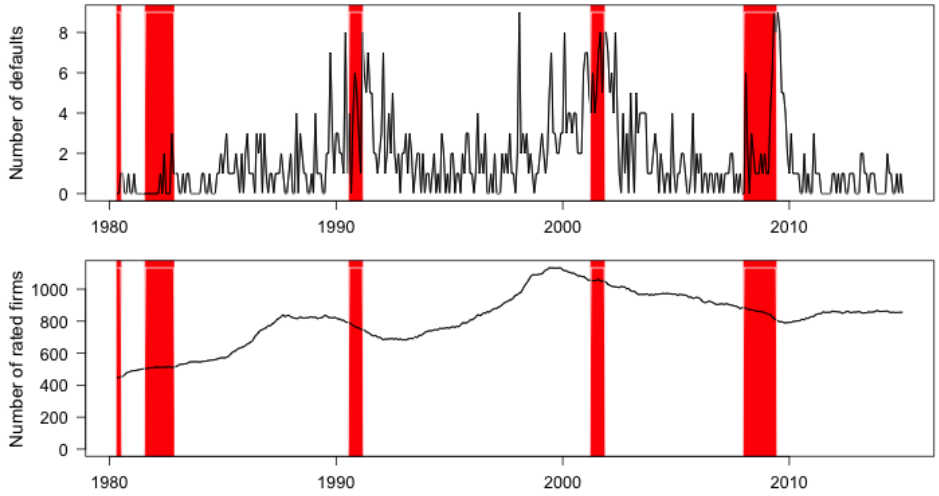
Previous research finds that default clustering implies correlation among defaults events (Das et al., 2007, Duffie et al., 2009, Azizpour et al., 2013). In other words, default correlation is reflected in the default clustering. Figure 1-2 shows the number of monthly defaulters and rated firms and four recession periods in the US industrial firms from April 1980 to December 2014. Red shading represents historical recession periods. It is clear that there are three clusterings of defaulters following three recession periods (excepting the savings and loan crisis of the 1980s), the junk bond crisis of the 1990s, the bursting of the dotcom bubble in 2001 and the subprime crisis of 2007-2009, which were followed by clusters of firm defaults. Default clustering started earlier than the dotcom bubble crisis since an Asian financial crisis occurred in 1997.

Recent developments in the field have led to a renewed interest in investigation of sources of defaults and understanding the determinants of default risk. The literature reports that a firm's financial health may have multiple influences over time (Duffie, 2011). These influences on default clustering are, in general terms, the business cycle, default contagion and default frailties. The business cycle has been a focus of interest for scholars since it results in systemic risk around the economic sectors (Koopman and Lucas, 2005, Koopman et al., 2009). Firms are exposed to common or correlated factors, such as stock return, industrial production growth, GDP growth and many other macroeconomic variables. Their co-movements produce correlated changes in the conditional default rates of firms.[1] Azizpour

[1] Several empirical studies document a large time variation in default frequencies and link this variation to, among other variables, business cycle indicators (Shumway, 2001, Das et al., 2007, Koopman and Lucas, 2005, Koopman et al., 2011, Duffie et al., 2007). Since such indicators simultaneously affect the DPs of many firms, their variations induce correlation between default events just as variation of common factors in asset return models induce correlation between returns.

et al. (2014) argue that strong economic growth reduces the likelihood of default across the board.

Figure 1-2 Time series of monthly default data in the US industrial firms from April 1980 to December 2014



This figure shows time series of the number of defaults (total defaults), number of rated month (total exposures), and IDPs in the US industrial firms from April 1980 to December 2014. The first panel presents time series plots of the total defaults in the end of each month d_t; the second panel is the time series of number of rated firms in the beginning of each month n_t. This figure is generated by using Moody's Default Recovery Database.

There are two key issues for understanding and modelling the common dynamics of default probabilities (DPs). One is how to choose an appropriate model for modelling corporate defaults from two broad categories of models (including reduced form models and structural models). The other is how to understand the determinants of default risk in order to model DPs. Generally, research focuses on the first issue and how to create a new sophisticated model. However, there is no agreement that a specific group of macroeconomic factors can consistently explain corporate defaults. Thus, I was motivated to investigate any evolution of effective factors for predicting corporate defaults in the US industrial firms.

Meanwhile, an increasing number of papers have focused on the relationship between monetary policies and credit market stability (Bernanke, 1993, Bernanke and Gertler, 1995, Curdia and Woodford, 2010, Bhamra et al., 2011). The main stream of these studies argues

that monetary tools used by central banks lead to instability of the credit market and monetary policies result in deterioration of corporate defaults. In this context, my study is motivated to explore whether macroeconomic factors that are related to monetary tools and monetary policies impact on corporate defaults over time.

1.1.2 Investigating the Relationship between the Dynamic of Various Macro Indicators and Economic Recessions

The issue of capability in early and correct prediction of recessions is of general concern to society. Consumers need to know how salaries and employment may change, firms want to know how market supply and demand are going to shrink or grow, and governments are interested in when it is appropriate to make fiscal and monetary policy decisions. Especially for central bankers and policy makers, a more precise and reliable predication of the state of the economy facilitates the implementation of appropriate and preemptive policies.

Since the 1930s, the forecasting of business cycles has been considered by central banks and policy makers and extensive research has studied model estimation with this purpose. Since 1990, studies have moved from forecasting business cycles to economic recessions (Estrella and Mishkin, 1998). The literature has mostly focused on two issues: one is to find which group of macroeconomic and financial variables have stronger predictive power by using different forecast horizons; the other is to develop more sophisticated models for improving predictive performance, with conventional proposed recession factors.

However, for the first issue, there has been a little improvement in finding effective and efficient recession factors for predicting recession months since the recession factors used cannot reflect the dynamically evolving macroeconomic conditions. Techniques exist, outside the finance and economics literature, for anticipating critical transitions in complex systems and these provide a possible solution. I was motivated to introduce such techniques to the prediction of economic recessions.

1.2 Research Questions

Research questions follow from the two main topics described above.

For the first topic, they are:

1) Do Effective Predictors (EPs) change over time for predicting corporate defaults?
2) Do monetary policies such as reserve rate, borrowing rate, and money supply have consistent impact on corporate defaults over time?

For the second topic, they are:

1) Do interactions among macroeconomic and financial variables signal the occurrence of NBER economic recessions in the US?
2) Can I build macro indicators by measuring interaction among these aforementioned factors to predict economic recessions one-month ahead?
3) Does the macro indicator constructed by broader macro factors outperform the others constructed by less macro factors in terms of predicting US economic recessions?

1.3 Thesis Structure

This thesis is organized as follows. Chapter 2 provides a literature review covering three main issues: default risk modeling, critical transitions in a complex system and prediction of economic recessions. Each section is discussed from both empirical and theoretical perspectives. In particular, the first section on the literature of default risk modeling leads to the first empirical chapter, Chapter 5. The remainder of Chapter 2 supports the second empirical chapter in Chapter 6. Chapter 3 presents a description of methodologies and newly developed analytical tools. This chapter only briefly explains two analytical tools I developed myself which include R package "Kai.R.Tools" and two web applications, "KAI.APP.PhD" and "KAI.APP.EIA". A detailed description of how to use them is shown in the corresponding manuals in Appendices 9.3 and 9.4. Chapter 4 describes the data. The descriptions of the corporate default database, macroeconomic database, and microeconomic database are given in sections 4.2, 4.3, and 4.4, respectively. Chapter 5 presents the results of the first empirical chapter, entitled Investigating Effective Predictors (EPs) for Corporate Defaults, which is followed by the second empirical results in Chapter 6 Investigating the Relationship between Dynamic of Various Macro Indicators and Economic Recessions.

Chapter 7 concludes this thesis, also covering implications for policy makers and central bankers and future research directions.[2]

.

[2] Based on my Ph.D. study, I have done some other work. For example, I have written a paper: *Macroeconomic Conditions, Corporate Defaults, and the Adaptive Market Hypothesis: Principal Indicator and Early Warning Signal*. This paper was accepted by two international conferences including the 2016 IFABS Asian Conference and 2016 Paris Financial Management Conference, respectively. I presented it in Brunei August 2016. For the second conference, I will present it in Paris in December 2016. In particular, the acceptance rate in Paris Financial Management Conference is only 30%. This year this conference has received more than 600 submissions. Another work is entitled *Exploring Effective Default Risk Predictors for Modelling Default Probabilities*. I presented this work in 2015 Finance and Risk Doctoral Colloquium organised by Industrial Economics and Finance division, NUBS (Nottingham University Business School) in May 2015, and I got Best Presentation Award (2 out of 25 candidates were awarded).

CHAPTER 2

2. Literature Review

2.1 Introduction

This chapter covers three main topics: default risk modelling, literature review on capturing critical transitions in a complex system, and anticipating economic recessions.

For the first, default risk modelling, I start with a brief description of research on default risk modelling, which is followed by the overview of models used for predicting corporate defaults. Then two main types of models including reduced form models and structural models are described in detail. I then figure the comparisons between two models, which motivates Chapter 5 Investigating Effective Predictors (EPs) for Corporate Defaults. The second, on how to capture critical transitions in a complex system, is described from both theoretical and empirical perspectives. For the third, on prediction of economic recessions, I review the literature from prediction of the business cycle to measurement of economic recessions in order to give a 'big picture' of how the research about the forecasts of economic recessions has developed.

The chapter is structured as follows. Section 2.2 describes the literature on default risk modeling. Section 2.3 introduces the literature on critical transitions in a complex system. Section 2.4 reviews the literature on prediction of economic recessions. Section 2.5 concludes.

2.2 Review of the Literature on Default Risk Modelling

2.2.1 Background of Research on Default Risk Management

Understanding the determinants of financial risk is a major issue for financial stability, and measurement and control are of clear concern around the world. According to Duffie and Singleton (2012), from the perspective of financial institutions, financial risk can be divided into the following broad categories: market risk, liquidity risk, operational risk, systemic risk and credit risk. Credit risk has been of particular interest since the subprime crisis started in 2007 (Turner, 2009).

(1) Types of credit risk

Credit risk is *the risk of either default or reduction in market value caused by changes in the credit quality of issuers or counter-parties* (Duffie and Singleton, 2012: 3). Anson et al. (2004) propose a detailed classification of credit risk and they state that credit risk can be divided into three categories of risk including credit default risk, downgrade risk, and credit spread risk. Their definitions of three types of credit risk are as follows:

> Credit default risk is the risk that the issuer of a bond or the debtor on a loan will not repay the outstanding debt in full. Default risk can be complete in that no amount of the bond or loan will be repaid, or it can be partial in that some portion of the original debt will be recovered.
>
> Downgrade risk is the risk that a nationally recognized statistical rating organization, such as Standard & Poor's, Moody's Investors Services, or Fitch Ratings reduces its outstanding credit rating for an issuer based on an evaluation of that issuer's current earning power versus its capacity to pay its debt obligations as they become due.
>
> Credit spread risk is the risk that the spread over a reference rate will increase for an outstanding debt obligation. Credit spread risk and downgrade risk differ in that the latter pertain to a specific, formal credit review by an independent

rating agency, while the former is the financial market's reaction to a perceived credit deterioration.

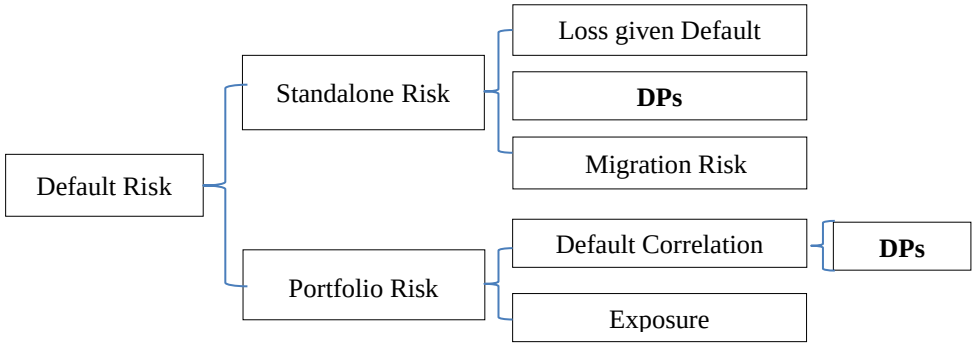
Two recent economic crises, the Dot-com crisis and the Subprime crisis, indicate that default risk is the key risk among three aforementioned risks for financial institutions, especially rating agencies.[3] For example, despite growing questions about Enron's ties to the private partnerships that proved its undoing, the agencies kept it at an investment-grade rating until just four days before it filed for bankruptcy in December 2001. Likewise, during the financial crisis of 2007–2009, rating agencies failed to detect the higher default risk among investment banking groups such as Bear Stearns, AIG, Fannie Mae, Freddie Mac, Lehman Brothers, and Merrill Lynch (Gurný and Gurný, 2013). Nowadays, predicting default risk correctly and in timely fashion is an especial concern in comparison with downgrade risk and spread risk. Thus, this research project concentrates on default risk rather than the other two risks.

(2) Insights into default risk

Duffie and Singleton (2012) state that default risk management used by financial firms may involve: (1) measuring the extent and sources of exposure; (2) charging each position a cost of capital appropriate to its risk; (3) providing information on a firm's financial integrity to outside parties. A significant target that applies specifically to default risk is to assign and enforce counterparty default exposure limits, which is reflected in Figure 2-1. This figure illustrates the structure of management of default risk in practice. It is clear that the management of default risk contains two risk management activities including standalone risk and portfolio risk. Actually, standalone risk aims to measure the default loss of a firm or sovereign over some time horizon such as one year and portfolio risk aims to measure relative credit quality among firms or loans.

[3] Default risk is short for credit default risk; spread risk is short for credit spread risk in this project.

Figure 2-1 Diagram of default risk



Source: Crosbie, P. & Bohn, J. 2002. Modelling Default Risk. KMV working paper.

DPs are the key element for measuring both standalone risk and portfolio risk. Crosbie and Bohn (2002) argue that each of the items in Figure 2-1 is critical to the management of credit portfolios, such as bond valuation and risk management of loan portfolios; however, the calculation of DPs is the most significant element, which leads to the question of whether the probability of extreme default losses on corporate and collateralized debts can be estimated correctly in financial practice. Indeed, management of default risk is essentially concerned with how to calculate DPs correctly.

An example is the international bank capital accord designed by the Basel Committee on Banking Supervision (BCBS). The first bank capital accord, from the late 1990s, is known as Basel II and proposes the use of credit risk models to determine banks' regulatory capital requirements. Based on credit exposure calculated by DPs, capital requirements are determined in order to identify an optimal level of capital once credit contagion happens. Under Basel II, it becomes crucial to measure credit risk accurately. BCBS (2005) states "The area of validation of […] DPs will prove to be a key challenge for banking institutions in the foreseeable future." The default forecast expressed in terms of DPs is then a key input parameter thereby determining the credit exposure in credit risk management (Blöchlinger, 2012).[4]

[4] Two issues arise following implementation of the Basel II capital accord. On the one hand, banks must hold enough capital to limit risks for depositors and to reduce insolvency risks. On the other hand, holding excessive capital is costly and limits efficiency. BCBS permits banks to use internal (or external) rating models (IRA) to classify borrowers according to their risk, which means they can use their own models to calculate the micro-level DPs. Thus, financial institutions have incentives to prefer

In order to handle weaknesses in the calculation of DPs, in 2010, BCBS published a new regulatory framework for banks, known as Basel III (BCBS, 2010). Compared with Basel II, Basel III has made some modifications in terms of strengthening the definition of capital and raising capital requirements but has not resolved the fundamental problem posed by reference to risk-weighted assets (RWA) measured by the banks with flawed internal models (Aikman et al., 2014). Under the internal ratings-based (IRB) approach, the banks are required to estimate only the probability of default while these advanced internal model systems estimate all the parameters (probability of default, loss given default and the correlation between the portfolio of loans and the macro-economic risk factor) (Dewatripont et al., 2010). Thus, the issue of how to calculate DPs correctly has received considerable critical attention.

In this context, DPs should be given priority in the present research. There are two key issues in terms of understanding and modelling the common dynamics of DPs for practitioners and academics in default risk modelling research. One is about how to choose an appropriate model for modelling corporate defaults from two broad categories of models, including reduced form models and structural models; the other is about how to understand the determinants of default risk for financial stability in order to model DPs. Generally, prior research mainly focuses on the former, which is about how to create a new sophisticated model, and an increasing number of researchers have been interested in statistically analysing the determinants of default risk, such as key predictors required to model DPs (Couderc et al., 2008, Azizpour et al., 2014, Tian et al., 2015).

The main purpose of this section is to give an overview the two types of models and the understanding of determinants of default risk, in support of Chapter 5. First of all, the two broad categories of credit risk models will be described individually in detail.

2.2.2 Overview of Models

The literature on default risk modelling shows that the major models are generally divided into two categories: structural (also known as firm-value) models and reduced form models,

models that engineer low default rates, so as to reduce the level of capital they hold, which consequently causes serious bank defaults in the financial crisis.

(McNeil et al., 2010). The approach of Black and Scholes (1973) and Merton (1974) is the progenitor of all structural models in which the process driving default is the value of the firm. In most situations, default time in these structural models is predictable. Reduced form models originate with Jarrow and Turnbull (1992). In contrast with structural models, defaults are regarded as an exogenously unspecified process rather than as a predictable process (Gregoriou and Hoppe, 2009).

According to Jarrow and Protter (2004), another difference between the two model types is not whether the default time is predictable or inaccessible, but whether the information set is observed or not. A modeller who designs a structural model has the same information set as the firm's manager, which means the modeller can also know all the firm's assets and liabilities. However, a designer of reduced form models has the market information set. Based on this viewpoint, the first generation of reduced forms dates back to Beaver (1968) and Altman (1968).

2.2.3 Reduced Form Models

Early research analyses the data of DPs via ratio analysis and most models employ reduced form models. The seminal concept in statistical modelling of DPs is Edward Altman's 1968 Z-score model that relates to probability of firm default to firm-specific accounting ratios—the ratio to total assets of working capital, retained earnings, earnings before interest and taxes, and the market value of equity (Allen, 2012).

The second generation of reduced form models is based on qualitative-response models. For instance, Ohlson (1980) and Zmijewski (1984) propose a conditional logit and probit model to calculate DPs with financial ratios.[5] One of the most striking characteristics of these reduced form models is the covariates they take account of. The common feature for these early reduced form models in these studies is the similar risk factors that are all firm-specific covariates (Wilson, 1997a, Wilson, 1997b, Dichev, 1998, Lennox, 1999, Griffin and Lemmon, 2002).

[5] The model proposed by Ohlson (1980) is known as the O-score model.

These early stage models are all static and only one set of explanatory variables for each firm is considered; thus, static models ignore the fact that the characteristics of most firms change from year to year. Also, these models neglect the scenario in which healthy firms can go bankrupt. In this context, researchers who use static models introduce a selection bias into their estimation of parameters by choosing arbitrarily when they observe characteristics of each firm (Shumway, 2001). Campbell et al. (2008) design a multiple logit model and attempt to forecast bankruptcy for different time horizons.

More advanced and sophisticated reduced form models are then developed which are also called intensity-based models. These models of DPs and timing are based on the notion of the arrival intensity of default. According to Duffie and Singleton (2012), the simplest version of such a model default is defined as the first arrival time t of a Poisson process with some constant mean arrival rate, called intensity, often called λ. They note that both hazard rate and intensity are used interchangeably in certain cases. In order to identify two items, they emphasize that the definition of default intensity λ_t is the arrival rate of default rate at t, conditioning on all information available at t. The hazard rate $f(t)$, however, is the mean arrival rate of default at t, conditioning only on survival to t. If the only information available is survival, then the hazard rate and intensity coincide (Duffie and Singleton, 2012).

For the information used in default intensities (in addition to financial ratios employed by early models), extended reduced form model take external covariates into consideration since some researchers find that there is correlation between macroeconomic conditions and defaults (Lo, 1986, Lennox, 1999, McDonald and Van de Gucht, 1999, Duffie, 2005). Numerous macroeconomic variables (such as 1-year S&P 500 index, US 3-month Treasury rate, growth rate in industrial production, spread between 10-year Treasury rate and 1-year Treasury rate) are added into reduced form model.

Based on the issue of whether firms' default intensities are correlated, these default intensity models can be divided into two groups including models with doubly-stochastic assumption and models without doubly-stochastic assumption.[6] In the doubly-stochastic model setting,

[6] Duffie et al. (2007) define the doubly-stochastic assumption: given the path of the state vector process, X, the merger and failure times of different firms are conditionally independent. The doubly-stochastic

the default correlation between firms arises from common dependence of default intensities on macroeconomic variables and correlation across firms of firm-specific covariates (Duffie, 2011). Moreover, compared with counterparty models without doubly-stochastic assumption, they cannot handle additional unobservable sources of default correlation. The next part will show the two groups of models with different assumptions.

(1) Models with doubly-stochastic assumption

The recent generation of risk models estimates single firm intensity (hazard) models that ignore the effects of default correlation among firms, and they are all based on a doubly-stochastic assumption. These models are dominated by duration analysis. Typically, these duration models apply a Cox proportional hazard model. In particular, Lane et al. (1986), as early researchers, use time-independent covariates to analyse bank default. Lee and Urrutia (1996) further develop duration models which are based on Weibull distribution of default time. They find that there are more significant explanatory variables in duration models in comparison with a logit model. Generally, duration models are based on time-varying covariates. Studies of duration analysis on predication of bankruptcy are done by Shumway (2001), Chava and Jarrow (2004), and Hillegeist et al. (2004). Duffie et al. (2007) provide maximum likelihood estimates of term structures of DPs by using a joint model for default intensities and the dynamics of the underlying time-varying covariates.

(2) Models without doubly-stochastic assumption

Duffie et al. (2007, p.637) argue that "the doubly-stochastic assumption is overly restrictive to the extent that default or another form of exit by one firm could have an important direct influence on the default or other-exit intensity of another firm." Das et al. (2007) evaluate this assumption by a time-change test and conclude the joint hypothesis of well-specified intensities and the doubly-stochastic assumption should be rejected because of dependence of firms' intensities on firm-specific and macroeconomic covariates. Research by Chava and

assumption is overly restrictive to the extent that default or another form of exit by one firm could have an important direct influence on the default or other-exit intensity of another firm.

Jarrow (2004), Duffie et al. (2007) and Figlewski et al. (2012) fit bottom-up models of correlated default intensities for individual firms.

Azizpour et al. (2013) run Poisson regressions of the monthly defaults on the explanatory variables and find that strong positive correlation between defaults and residuals exists and residuals are significantly auto-correlated. Azizpour et al. (2013, p.7) suggest "the existence of unmeasured or missing explanatory variables ("frailties") influencing default timing beyond the variables we have included as regressors."

Duffie et al. (2009) report that there is the possibility of "contagion", by which the default of one firm can have a direct influence on the revenues (or expenses or capital-raising opportunities) of another firm. Azizpour and Giesecke (2008) investigate default clustering and find that both contagion and frailty generate "self-exciting" phenomena, in that a default tends to increase the likelihood of other firms to fail. Many research studies have analysed existence of "frailties" and "contagion" (Azizpour and Giesecke, 2008, Koopman et al., 2011).

Azizpour and Giesecke (2008, p.3) point out that "contagion and frailty phenomena are found to be roughly equally important for explaining the default clustering in the data that is not captured by a traditional doubly stochastic model, in which firms' default times are correlated only as implied by the correlation of observable risk factors determining their default intensities."

Duffie (2011) finds that models that can handle these unobservable risk factors are able to correctly capture the default intensities. Prior studies develop frailty models and document that they have a better performance than the models with doubly-stochastic assumption (Giesecke, 2004, Collin-Dufresne et al., 2010, Koopman et al., 2011, Nielsen, 2011). Although Lando and Nielsen (2010) determine that the doubly-stochastic assumption cannot be rejected by using a different set of covariates, numerous studies prove that the models that are able to deal with unobservable frailty factors are more appropriate than the others without catching unobservable frailty factors.

2.2.4 Structural Models

The first structural model comes from Merton (1974) and is based on option pricing theory. His model describes (non-observable) firms' asset values as varying according to geometric Brownian motion, with a single debt: a bond paying only principal at maturity and no coupons. The firm defaults at maturity if its assets move lower than its outstanding debt. Equity is viewed as a long European call option on the value of the firm's assets and debt as a short European put option. Subsequently, more sophisticated conditions have been added but in the same spirit as the original model. The firm files for bankruptcy and ownership is assumed to be transferred costlessly to the debt holders while the payoff for equity holders is zero (Hillegeist et al., 2004). However, it is widely recognised that the underlying assumptions are quite unrealistic. According to Elizalde (2006) and Charitou et al. (2013), these models have been extended in terms of relaxing the assumptions of early structural models, for example, allowing for default before maturity, stochastic interest rate, various transaction costs, and many others. Five key inputs in the structural model, which correspond to these relaxed assumptions, are listed as follows.

(1) Asset value

Asset value is modelled as geometric Brownian motion. Recent studies determine an asset value process with both a diffusion and a jump component (Zhou, 2001, Delianedis and Geske, 2001). Duffie and Lando (2001) find a case where asset value is observed only with white noise, giving results similar to models with jumps. Goldstein et al. (2001) employ current cash flow as the state variable rather than asset value. However, asset value is fixed times of cash flow in these alternative methods, then the outcomes of analysis are identical. A recent notable extension by Charitou et al. (2013) demonstrates that the firm value can be assumed as delta-geometric random walk instead of geometric Brownian motion.

Another distinctive approach is introduced by Bharath and Shumway (2008). They modify the Black Scholes Merton (BSM) model by employing market-based estimates of firm value instead of solving the simultaneous equations of the theoretical BSM model and this extended model results in improved predictions.

(2) Debt

The Merton model considers only zero-coupon bonds. Geske (1977) extends the model to consider characteristics such as sinking funds, safety covenants, debt subordination, and payout restrictions. He considers the debt structure of the firm as a coupon bond, in which each coupon payment is viewed as a compound option and a possible cause of default. The other extended models can consider callable and or convertible debt (Fischer et al., 1989, Goldstein et al., 2001, Collin‐Dufresne et al., 2001).

(3) Default boundary

According to Leland (2004), a "default boundary" can be defined as a level of asset value which may vary over time. If the value of a firm's asset falls below this level, then this firm will default on its debt. Explicit formal assumptions regarding the nature of the default trigger appear most commonly in structural models of credit risk.

In modified Merton models, default can happen before debt maturity. Black and Cox (1976) introduce the first-passage-time model that specifies default as an event of the first time that the firm's asset value hits the default boundary. This approach was extended by Leland (1994) and Leland and Toft (1996). The level of the boundary may be specified exogenously by existing covenants, typically in relation to the face value of debt. The assumption of an exogenous value-based boundary appears in, among others Longstaff and Schwartz (1995) and Collin‐Dufresne et al. (2001).

(4) Transaction costs

Costs of default include not only the direct costs of restructuring and/or bankruptcy but also the loss of value resulting from employees leaving, customers directing business to other firms, trade credit being curtailed and the possible loss of growth options. These 'indirect' costs may be several times greater than the direct costs. Several researchers have included corporate taxes and bankruptcy costs in practice (Turnbull, 1979, Leland and Toft, 1996).

(5) Default-free interest rate

The assumption of a flat term structure of interest rates is other major criticism of the early models. Kim et al. (1993) consider the riskless interest rate following a square root process correlated with firm value. They suppose that default risk is insensitive to the volatility of interest rates but is sensitive to interest rate expectations. Meanwhile, Longstaff and Schwartz (1995) also model stochastic interest rate that is negatively correlated with the firm process, an exogenous early default, and an exogenous recovery rate. Jones et al. (1984, p.624) declare that "there exists evidence that introducing stochastic interest rates, as well as taxes, would improve the model's performance." Kim et al. (1993) and Longstaff and Schwartz (1995) allow for stochastic default-free interest rates. Bennin (1980) introduces stochastic interest rates in an endogenous bankruptcy model.

2.2.5 Comparisons between Reduced Form Models and Structural Models

The major issue in modelling DPs is to handle the default clustering/correlation among various firms. Therefore, a reasonable standard for choosing a model is to evaluate their abilities to capture default correlations. This section is to describe further how models in the literature handle default correlation.

(1) Structural models in capturing default correlation

Default correlation is caused by three resources including business cycle, default contagious, and default frailties. Structural models focus on the stochastically varying assets of a single firm or on the investigation of the correlations of two firms (Zhou, 2001, Löffler, 2004, Hull et al., 2005, Erlenmaier and Gersbach, 2013, Liu et al., 2013). The major components in structural models are asset, debt, equity, and default boundary and any dependence of one component on another can generate default correlation. Thus the various default correlations can be categorised by the channels of correlation (Liu et al., 2013). Asset values are treated as a function of common factors and firm-specific factors, where the common factors dictate the asset return correlation between firms (Finger, 1999).

Since asset value process and default boundary are unobservable, there is a limitation in implementing this type of model to relate asset return correlation to default correlation. Past

studies address this problem by assuming that asset return correlation is equal to equity return correlation. However, research suggests that equity return correlation is an imperfect proxy for asset return correlation (De Servigny and Renault, 2002, Zeng and Zhang, 2002). Further evidence comes from Liu et al. (2013) who investigate the relationship between equity correlation and default correlation. The modified structural model they propose uses this finding to estimate the default correlations between two firms. Surulescu (2010) states that "a crucial weakness of these structural models is that the empirical evidence from time series of equity prices contradicts the Merton model. The observation goes back at least to Black (1976), who discussed the fat tail characteristics of return distributions. Moreover, the Merton model also predicts that the implied volatility of options is constant through time, which is false."

More importantly, the main reason why structural models cannot handle default correlation is essentially due to structural models heavy reliance on stock market components which are noisy and contain bubbles (Janosi et al., 2002). The final criticism of this type of model is that it does not include external risk drivers, in contrast with reduced form models (Duffie, 2011). Duffie et al. (2007) propose that the performance of the extended structural models can be improved in prediction of DPs by developing such structural models with additional state variables, such as macroeconomic variables, which means the business cycle can be included.

Bharath and Shumway (2008) note that the Merton distance to default (DTD)[7] model is not a typical econometric model and it is not clear how this structural model can be extended to consider additional default predictors or how its parameters may be estimated with alternative techniques. They suggest that research on combinations of structural models and econometric methods in practice will become a promising topic.

(2) Reduced form models in capturing default correlation

Since reduced form models are a kind of econometric model, they are more flexible in terms of coping with default correlation and can include for various risk predictors, such as

[7] KMV is a distance to default model.

economic predictors, firm-specific predictors, and even unobservable risk predictors. In this context, reduced form modelling can, in principle, capture the business cycle by using economic predictors, capture default contagion by using firm-specific predictors, and handle default frailties by using unobservable risk predictors.

The idea of default frailties is first proposed by Duffie et al. (2009) who present a reduced form model allowing for unobservable predictors. Specifically, statistcs literature regards an unobserved covariate as "frailty" in the event forecasting. They explain that some defaults are highly correlated and these defaults are caused by same missing risk factors in the corporate debt arena. For example, in 2001 both Enron and WorldCom defaulted because of accounting scandles. In fact, the other firms also used faulty accounting practices as what Enron and WorldCom did. Thus this acoounting violation have had influence on the conditional default probablities of other firms, and therefore on portfolio losses. In this context, the uncertainty caused by these frailty factors is an additional source of clustering. Duffie et al. (2009) employ a method, which is to use "Bayes's Rule to update the posterior distribution of unobserved risk factors whenever defaults arrive with a timing that is more or less clustered than would be expected based on the observable risk factors alone." Azizpour et al. (2014) divide unobservable predictors into two unobservable internal predictors and unobservable external predictors in order to capture default clustering in default risk management.

2.3 Review of the Literature on Critical Transitions in the Complex System

2.3.1 Theoretical Perspective

A definition of critical transitions is given by Scheffer et al. (2012, p.344):

> *Sharp regime shifts that punctuate the usual fluctuations around trends in ecosystems or societies may often be simply the result of an unpredictable external shock. However, another possibility is that such a shift represents a so-called critical transition. The likelihood of such transitions may gradually increase as a system approaches a 'tipping point' [i.e., a catastrophic*

> *bifurcation], where a minor trigger can invoke a self-propagating shift to a contrasting state.*

We may conclude that critical transitions show the process of changes in the system. Meanwhile, the accumulation of such transitions leads to tipping points, which may indicate high probability of unwanted collapse in the system, or opportunities or positive change.

Scheffer et al. (2012) state that in a complex system the occurrence of critical transitions implies that there are interactions among the system components. The interactions result in cascade effects (or domino effects). Helbing (2013) uses catastrophe theory to explain how the critical transitions lead to cascade effects. Specifically, discontinuous transitions may result in disasters if there are gradual changes in parameters. The system starts shifting at certain "tipping points" which are critical parameter values. Then these shifts result in different system properties. He further explains that "the theory of critical phenomena has shown that, at such tipping points, power-law (or other heavily skewed) distributions of event sizes are typical. They relate to cascade effects, which may have any size." He argues that these "extreme events" can be caused by the inherent system instead of unexpected external events, which can be explained by the theory of self-organized criticality, for example, piles of particles can lead to avalanches.

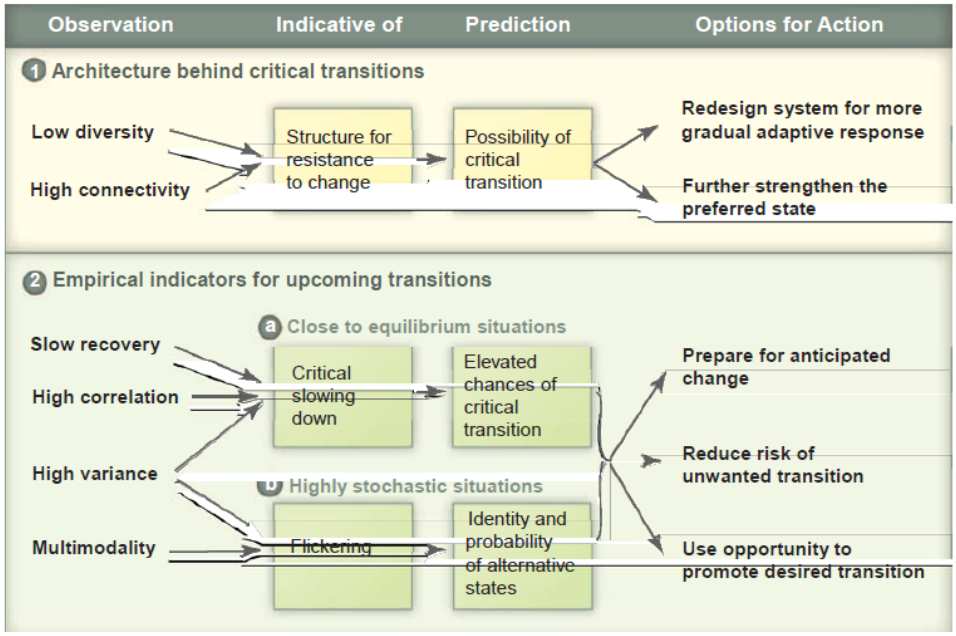
For these cascade effects, Helbing (2013) figures that they are closely related to the strong interactions. He explains that if the interactions among system components become strong, the behaviour of these components may vary greatly or influence the functionality or operation of other components. Helbing (2013, p.53) adds that "One event can trigger further events, thereby creating amplification and cascade effects, which implies a large vulnerability to perturbations, variations or random failures. Cascade effects come along with highly correlated transitions of many system components or variables from a stable to an unstable state, thereby driving the system out of equilibrium."

Scheffer et al. (2012) conclude that the empirical indicators based on different classes of generic observations can be used to indicate the potential for critical transitions in a complex system, such as correlation among system components and their variance, which is shown in Figure 2-2. Specifically, the high connectivity and low diversity in the network lead to critical

slowing down which means an increased probability of a sudden transition to a new unknown state; that is, the probability of critical transition occurring is quite high. This network can then be shaped into a new one after gradually adapting the response. To detect the upcoming transitions, empirical indicators constructed by information of correlation or variance enable the capture of tipping points resulting from critical transitions. They determine that tipping points that result from such transitions may provide an early warning signal for potential crisis even before obvious symptoms of crisis appear, and building the empirical indicators can capture unwanted change in the complex system. They also argue that, in order to detect the incipient transitions from the system, long, high-resolution data should be used. That is, the more data about system components we use, the better performance our empirical indicators can have. Meanwhile, there is an advantage of using these indicators, which is that "Perhaps the most exciting aspect of this work is that it is uncovering generic features that should in principle hold for any complex system. This implies that we may use these approaches even if we do not understand all details of the underlying mechanisms that drive any particular system. This is the rule rather the exception, as we are far from being able to construct accurate predictive mechanistic models for most, if not all, complex models" (Scheffer et al., 2012, p.347). In other words, this method can be used in any system since the correlation and variance among system components can reflect the transitions.

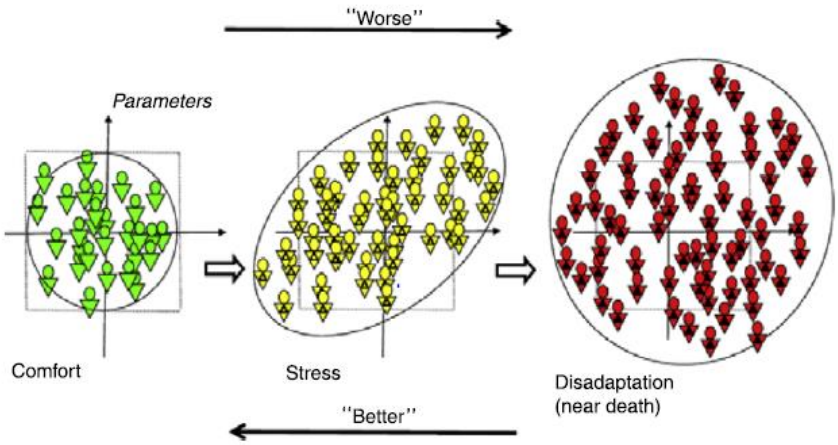
Meanwhile, Gorban et al. (2010) claim that correlations between individual systems are better indicators than using value of attributes to construct the indicator. They argue that the correlation increases during a crisis, even before the occurrence of that crisis, and the variance (volatility) also increases, as in Figure 2-3. After the crisis reaches its bottom, some systems can develop into two outcomes. One is that both correlation and the variance decrease and the system is recovering; another is that the correlations decrease but variance continually increases which means that fatal catastrophe occurs, as indicated in red on the right-hand graph in Figure 2-3. Thus, theoretically, both correlation and variance can be used to construct indicators for measuring transitions in the system and even for predicting the risk of unwanted collapse.

Figure 2-2 Different classes of generic observations that can be used to indicate the potential for critical transitions in a complex system



Source: this figure is taken from Scheffer et al. (2012, p.347)

Figure 2-3 Correlations and variance in crisis



The typical picture: Cor ↑; Var ↑ — stress; Cor ↓; Var ↓ — recovering; Cor ↓; Var ↓ — approaching the disadaptation catastrophe after the bottom of the crisis. In this schematic picture, axes correspond to attributes, normalized to the unite variance in the comfort state.

Source: this figure is taken from Gorban et al. (2010, p.3194).

2.3.2 Empirical Perspective

Scheffer et al. (2012) find an increasing number of studies across areas such as ecology, physiology, epilepsy, engineering, and climate have worked on building empirical indicators.

The majority of studies focus on ecology and climate science. They claim that social sciences and medicine might well be particularly rich fields for exploration.

Recently, the idea of constructing empirical indicators has been introduced in Finance. Gorban et al. (2010) investigate both correlation and variance in financial systems by using daily stock price from January 3, 2006 to November 20, 2008 for companies that are registered on the FTSE 100 index. The FTSE 100 index represents the performance of the 100 largest firms, accounting for approximately 88% of total market capitalization in the UK. They choose thirty companies with the highest value from different economic sectors. They use several methods (from simple weight of the correlation graphs to more sophisticated characteristics based on principal component analysis) to construct indicators by collecting the correlations among these thirty companies' stocks. They find that a simple weight of correlation graph performs well by using scatter plot to compare the behaviours of different indicators.[8] Then they compare indicators based on either correlation or variance to the value of the FTSE 100 index. Variances and correlations among these stocks increases when the market decreases and decrease when it recovers.

The pioneering study of May et al. (2008, p.893) proposes a potential application to capture systemic risk in banking 'ecosystems', arguing that "there is a common ground in analysing financial systems and ecosystems, especially in the need to identify conditions that dispose a system to be knocked from seeming stability into another, less happy state." They suppose that systemic risk is generated under the increasingly complex and globally interlinked financial networks, which can be interpreted as cascade effects in the ecosystem. They declare that highly connected balance-sheets cover the potential systemic risk in the financial system, which could provide useful information in analysing the effects of asset-pricing bubbles, credit crises and the poorly understood but potentially concerning effects of the wide use of financial derivatives including futures and options and dynamic hedging by investment banks to manage risk on the fly. Haldane and May (2011) further investigate systemic risk in the banking ecosystems, and find that this is reflected by cascade effects in the financial system. Specifically, external shocks to the banking system, such as recessions, major wars,

[8] This study also mentions the other possible methods for comparisons of behaviour of different indicators, such as correlation analysis or other analytical tools

civil unrest or environmental catastrophes, can lower the value of a bank's assets so seriously that the system fails. Additionally, they determine that interbank loans could be used for measuring the systemic risk among banks. Neither study engages in empirical analysis, though they propose the possible application of building empirical indicators in the financial market in order to detect potential changes in the system.

2.4 Review of the Literature on Measurement of Economic Recession

Studies focus on the exploration of business cycle indicators and use these variables to identify the turning points or cyclical points in the business cycle in order to determine the state of the economy. The literature on prediction of economic recessions has concentrated on application of probit models by employing macroeconomic indicators for predicting economic recessions. This section begins with a review of research on the business cycle, followed by a review of the literature on prediction of economic recessions.

2.4.1 Capturing Business Cycle by Business Cycle Indicators

There is a crucial group of papers on business cycles that study the development of cyclical indicators in forecasting, measuring, and interpreting short-run changes in aggregate economic activity (Shiskin and Moore, 1967).

The pioneering study on developing the cyclical indicator is by Mitchell and Burns (1938, p.4). They investigate almost 500 monthly or quarterly series and identify that "71 indicators have been tolerably consistent in their timing in relation to business-cycle revivals and that at the same time are of general interest to warrant some attention by students of current economic conditions". Their study further explores these indicators and finds 21 as the most trustworthy indicators of cyclical revivals. They classify them as three groups of indicators including leading, coincident, and lagging indicators by comparing the dates of the cyclical turns in each series with the corresponding reference. As the first study on the identification of cyclical indicators, Mitchell and Burns (1938) contribute much in terms of analysing alternating sequences of business cycles. Moore (1961) extends the sample size and uses about 800 various macro factors from 1938, classifying 21 selected series from 800 series

into three categories based on their timing at business cycle peaks and troughs including leading, coincident, and lagging indicators. After the 1960s, the US Department of Commerce published business cycle indicator data and composite indices based on leading, coincident, and lagging indicators. This list of core factors has been refined and further developed several times. Shiskin and Moore (1967) present a scoring method for evaluating and selecting the indicators and they figure 1966 list of 25 cyclical indicators covering 12 leading, 7 roughly coincident, and 6 lagging indicators. Moreover, they also state the economic process that is regarded as a secondary principal of classification for analysing the business cycle. All 88 economic factors have been filed into 8 major economic groups. Before the Business Cycle Indicators programme was taken over by the Conference Board (CB), this list was revised three times, in 1975, 1989, and 1996. CB now maintains the database and publishes a monthly report. Meanwhile, CB has determined the final list of the most reliable factors, which includes ten leading indicators, four coincident indicators, and seven lagging indicators.[9]

2.4.2 Predicting Economic Recessions

Based on the cyclical indicators, economists normally apply mathematical models to anticipate economic activity and likelihood of recession. Stock and Watson (1989) develop three monthly indices including an index of coincident economic indicators (CEI), an index of leading economic indicators (LEI), and a Recession Index. These indices are constructed by using models with leading and coincident indicators in order to improve the predictive ability. In contrast with earlier literature, which focuses on quantitative measures of future economic activity, the Recession Index is uniquely used to estimate the probability that the economy will be in a recession in the next six months.

From the 1990s, studies turned to the investigation of forecasting economic recessions rather than future economic activity. The seminal study in predicting economic recessions is by

[9] The Business Cycle Indicator handbook details the historical changes of the list of most reliable macro factors and it also explains why some factors are removed and the other factors are added to this list. For the latest list of the factors and their corresponding classifications, see the Business Cycle Indicators Handbook 2001.

Estrella and Mishkin (1998) using simple financial variables to predict the future recessions, such as interest rates, interest rate spreads, stock price indices, and monetary aggregates, benchmarking against predictive performance using traditional macroeconomic indicators, such as LEI and another experimental leading index developed by Stock and Watson (1993). A distinctive feature in this study is that they apply a simple probit regression model for binary outcomes in the occurrence of recessions instead of using sophisticated models. Based on quarterly US data from the first quarter of 1959 to the first quarter of 1995, they find that financial variables produce more accurate predictions of recessions than the complex leading indicators in out-of-sample analysis. Stock price is able to predict recession with one to three-quarter horizons, which is the same case as these complex leading indicators. Yield spread is the best predictor in terms of predicting future changes of real economic activity and recession periods, providing a signal of recessions four quarters ahead. Its performance is better than in conjunction with other variables.

Dueker (1997) uses two econometric models including static probit model and its extension with Markov switching to examine the predictive power of yield spread, stock price and the other well-known macroeconomic indicators, such as the Commerce Department's index of leading indicators. Similar results are found by Estrella and Mishkin (1998). The yield-curve slope is still the single best predictor, even under two different probit models.

Chauvet and Potter (2005) use yield spread as the predictor for forecasting US recessions with the sample from January 1954 to March 2001, whereas the business cycle data runs over the period of January 1955 to December 2000, since they are interested in investigation of predicting performance by using different specifications of the probit models, such as a time invariant conditionally independent version; a time varying conditionally independent version that takes into account multiple breakpoints across business cycles; a model with constant parameters and autocorrelated latent variable; and a time varying specification with autocorrelated latent variable. The main motivation of applying these sophisticated models is that the standard probit model fails to allow for the evolving dynamics of the relationship between the yield curves. They extend the probit specification of Estrella and Mishkin (1998) to obtain four specifications of the probit models mentioned above in order to improve the predictive performance of the yield curve on the recessions. They find that the standard probit

specification has a tendency to over-forecast recession events whereas the more sophisticated models, by using the yield curve, are able to provide some warning signal for recessions. However, the signal is not clear. In particular, the models have a poor performance in capturing recession in 2001; the terrorist attacks on September 11th 2001 result in the increase in the probability of occurrence of a recession, and the yield curve data up to March 2001 do not provide predictive content on this event.

Kauppi and Saikkonen (2008) extend the sample used by early studies to the period from 1955 to 2005 by using a single recession predictor, interest spread. They apply a new out-of-sample method which is able to compute "iterated" multiperiod-ahead forecasts. They find that dynamic probit models outperform their static counterparts. If the different dynamic specifications of models are compared, the models with lagged values of the binary response generate more precise predictions of US recessions than models where dynamics enter only through a lagged probit probability. Also, iterated multiperiod-ahead forecasts have better performance than direct forecasts based on horizon-specific model specifications used by prior studies. However, they still cannot predict recessions in 1990 and 2001.

Nyberg (2010) further develops a dynamic probit model based on the models proposed by Kauppi and Saikkonen (2008) in order to predict recessions in both the US and German economies. Four financial variables including the domestic and foreign term spreads, stock market returns and the interest rate differential between the USA and Germany are used as predictive variables in those models. Nyberg (2010) investigates the period from January 1972 to December 2007. He uses the recession periods referenced from the National Bureau of Economic Research (NBER) for the US and Economic Cycle Research Institute (ECRI) for Germany. He demonstrates that dynamic probit models have better performance than the standard static recession prediction models in terms of both in-sample and out-of-sample predictions. The best predictor for both the US and German recession periods is the term spread as in many previous studies. In addition, he also determines that additional predictive power for explaining both countries' recessions is from stock returns.

Ng (2012) uses the advanced dynamic probit models to generate an accurate and reliable forecast of the recessions. More significantly, he points out that the models in the one-variable approach used by earlier papers have limitations in providing a whole assessment of

forthcoming recession risks. Then he incorporates five recession predictors to improve the predicting performance of these dynamic probit models. These predictors cover financial market expectations of a gloomy economic outlook, credit or liquidity risks in the general economy, the risks of negative wealth effects resulting from the bursting of asset price bubbles, and signs of deteriorating macroeconomic fundamentals. He uses monthly data from January 1969 to December 2009. He finds that various probit models with proposed risk factors outperform the conventional static models with only single recession predictors, such as only yield spread and stock market return in terms of forecasting the duration of recessions precisely. Additionally, for predicting the duration of recessions, the advanced dynamic probit models have better performance than the static probit model. Static models with proposed risk predictors have similar predicting performance as the dynamic probit model in anticipating the peaks of business cycles. This study suggests that both static and dynamic probit models with the proposed recession predictors are able to complement each other thereby providing a better assessment on the business cycle in the future.

A recent study by Christiansen et al. (2014) differs from the previous studies in terms of allowing for the sentiment variables as the recession predictors for US recessions. Specifically, this study concerns two variables including the Index of Consumer Sentiment and the Purchasing Manager's Index (PMI) which are used to measure US business and consumer confidence. Two sets of control variables are used for determining the extent to which two sentiment variables cover independent incremental explanatory power not captured by other economic predictors. The first set of control variables contains three conventional recession predictors: the interest spread, the short rate, and the return on the SP500 index. The second set of control variables consists of 176 macroeconomic and financial data that are used to construct common factors. This study focuses on the US economy and uses monthly data from January 1978 to December 2011. They find that business sentiment measured by PMI is the single best recession predictor among the other predictors. In addition, they find that sentiment variables combined with classical recession predictors are able to provide stronger forecasting performance for recessions in the future in in-sample and out-of-sample analysis.

2.5 Conclusion

This chapter reviews literature relevant to the thesis. The review of default risk modelling in Section 2.2 supports the first empirical chapter (Chapter 5) entitled Investigating Effective Predictors (EPs) for Corporate Defaults. Both the second (Section 2.3) and third (Section 2.4) reviews, on capturing critical transitions in a complex system and measuring economic recessions, support the second empirical chapter (Chapter 6) entitled Investigating the Relationship between Dynamic of Various Macro Indicators and Economic Recessions.

CHAPTER 3

3. Methodologies and New Developed Analytical Tools

3.1 Introduction

This chapter describes the methods used in this thesis. The first section presents econometric and statistical methods, such as filtering time series data, variable selection, and statistical inference methods including linear regression and logistic regression. The second section describes new analytical tools developed and designed by the thesis author, which includes R package known as toolbox, entitled "Kai.R.Tools" and two Shiny web applications including "KAI.APP.PhD" and "KAI.APP.EIA". Manuals for these can be found in Appendices 9.3 and 9.4.

The chapter is organized as follows. In Section 3.2 there are 4 statistical/econometric methods for supporting empirical analysis. Section 3.3 presents the description of analytical tools that have been developed and designed by the author during my Ph.D. study. Section 3.4 concludes.

3.2 Methodologies

This section only describes the statistical and econometric methods that are used in my Ph.D. study. The first method is Hodrick-Prescott (HP) Filter, which is used to remove the cyclical component of a time series from raw data. The second main method is an advanced variable selection method called LASSO, which is one of the most popular statistical learning methods (James et al., 2013). In this study, based on linear regression and logit regression, LASSO is applied to select key variables from numerous variables. The next group of methods are all econometric techniques, and they are used for doing causality analysis, which includes VAR and VECM. The two statistical inference methods are linear regression and logit model, respectively.

3.2.1 Hodrick-Prescott (HP) Filter

In recent years, the HP filter, developed by Hodrick and Prescott (1997), has become a popular method for removing the growth component from data exhibiting long run trends over time. The HP filter has been much applied in the Real Business Cycle (RBC) literature as a method of isolating business cycle frequencies from the long run trend component in both actual and simulated data.

The HP filter is based on the assumption that a time-series y_t can be modelled as the sum of a growth component (g_t) and a cyclical component (c_t), which can be written as follows:

$$y_t = g_t + c_t, \qquad \text{where } i = 1, \dots, T. \tag{1}$$

In the real business cycle literature, it is desirable to remove the growth component, whether it be a stochastic or deterministic trend, in order to study the behaviour of the cyclical component and to compare that behaviour between different series. The HP filter achieves this by defining the cyclical component as:

$$c_t = y_t - g_t, \tag{2}$$

where c_t is the desired detrended series. The HP filter minimizes the variance of Equation (1) while penalizing for excessive changes in the growth component. This can be expressed in terms of the following minimization problem:

$$\underset{\{g_t\}_{t=-1}^{T}}{\text{Min}} \left\{ \sum_{t=1}^{T} c_t^2 + \gamma \sum_{t=1}^{T} [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \right\} \quad (3)$$

$$\underset{\{g_t\}_{t=-1}^{T}}{\text{Min}} \left\{ \sum_{t=1}^{T} (y_t - g_t)^2 + \gamma \sum_{t=1}^{T} [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \right\} \quad (4)$$

where γ is termed the 'smoothing parameter', since its value determines the penalty applied to excessive changes in the growth rate of the series. The larger the value of γ, the smoother is the solution series. The smoothing parameter can be viewed as the ratio of the variance of the cyclical component to that of the change in the growth component $\gamma = \sigma_c^2 / \sigma_g^2$.

The value chosen for γ will differ depending on the sampling frequency of the data and represents the trade-off between smoothness and goodness of fit in the resulting series. Prior studies normally fix γ by using $100 * (number\ of\ periods\ in\ a\ year)^2$. I then find that:

1) Annual data: 100;
2) Quarterly data: 1600;
3) Monthly data: 14400.

A smoothing constant of 100/1600/14400 for annual/quarterly/monthly data has been taken as a standard by many researchers in macroeconomics (Hodrick and Prescott, 1997). However, the more γ is used, more smoothness transferred data I have, which means that some information hidden in the data may be deleted. It can be seen that using 14400 seriously can make these time series remarkable smooth and the key information is lost. Therefore, it should be cautious to choose γ in order for filtering cyclical component in the time series.

3.2.2 Least Absolute Shrinkage and Selection Operator (LASSO)

This study uses one of the state-of-art variable selection techniques called LASSO proposed by Tibshirani (1996) in order to determine the effective predictors (EPs) for explaining IDPs and economic recessions. This part only presents the model of LASSO used for extracting effective factors for IDPs, and the application of LASSO in the logistic regression for choosing effective factors for recession will be discussed in the next part. LASSO penalizes regression coefficients through a shrinkage method and thus provides a sparse variable-set solution. Specifically, this model is to shrink some coefficients and set others to zero and

then attempt to retain the good features of subset selection and ridge regression.[10] LASSO enjoys the easy interpretability, as does the traditional subset variable selection, but has additional advantages of: (1) the stability of model selection; and (2) potential improvement in prediction accuracy.

Suppose that we have data $(x_i, y_i), i = 1, 2, \dots, N$, where $x_i = (x_{i1}, \dots, x_{ip})$ are the predictor variables and y_i is the response. The basic assumption is that the observations for both x_i and y_i are independent and follow normal distribution. Since this method is a special case of general linear regression and LASSO only adds a penalty term called ℓ_1 norm (see below), then this method should have similar assumptions including independent observations and normal distribution of the explanatory variables and dependent variable.

In order to allow for the influence from scaling of the predictors, all the predictors should be standardized before application of this model. The formula is shown below:

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}}. \tag{5}$$

The predictors are then all on the same scale after applying Equation (5). Here j means the jth predictor and the denominator is the sample standard deviation of the jth predictor. Therefore, all of the standardized predictors will have standard deviation of one. The final fit will not then depend on the scale of with the predictors.

The LASSO coefficient, $\hat{\beta}_\lambda^L$, minimizes the quantity

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}|\beta_j| = \text{RSS} + \lambda\sum_{j=1}^{p}|\beta_j|. \tag{6}$$

RSS is the least squares fitting procedure for the linear regression. There is only one difference between Equation (6) of LASSO and least squares, which is that the coefficients are estimated by minimizing a slightly different quantity. λ is turning parameter which is

[10] The ridge regression method is another shrink method. The main difference between Lasso and this method is that ridge regression cannot set any coefficients to zero but keeps all the coefficients. Consequently, ridge regression does give an easily interpretable model. Ridge regression is not used in this research (James et al., 2013).

larger than 0. The penalty term for shrinking the coefficients is $\sum_{j=1}^{p}|\beta_j|$, which is also called ℓ_1 norm of the vector β.

In order to obtain a reliable result for λ, two major issues should be considered in practice. One is to select an effective resampling method for training model and testing the model performance by minimizing out-of-sample mean square error (MSE) to determine the value of λ. In this study, on a resampling method, we use the k-fold cross validation set method.[11] Following the statistical literature, k being equal to 10 can balance the trade-off between and bias and variance in the model estimation (James et al., 2013). Another issue is about how to set appropriate proportion of training and testing sample for applying resampling method. It is clear that reliable results should not be influenced by using different proportions such as 60%, 65%, 70%, 75%, 80% and 85% for determining training observations. Therefore, in this study, I use them to identify λ thereby obtaining set of EPs.

This study uses non-filtered IDPs and 92 macroeconomic factors for extracting effective factors. It should be noted that growth rates of most of factors are stationary with roughly normal distribution. IDPs is not normally distributed. Two reasons lead to using IDPs instead of using transformed IDPs (growth rate of filtered IDPs).

First, I use growth rate of filtering IDP as response thereby converting to normality. I find that there are only 8 factors extracted under the 70% and 60% level of training sample. If the other training sample sizes are used, the results are 4 effective factors in using 75%, 80%, 85% of training sample and 23 effective factors in using 65%. Then it is not stable. In addition, the results of performance of predicting higher default events are almost same as the performance of using indicator constructed by 29 effective factors from using non-filtered IDPs and the indicators constructed by 8 effective factors can still explain IDPs. However, R square is smaller than using 29 factors. Second, although the series of levels of IDP is not

[11] The common point between the validation set approach and k- k-fold cross validation is that they all need to split the sample into two parts including the sample used to estimate the coefficients in the model and the sample used to do model assessment, depending on the fix percentage. Then they calculate MSE thereby obtaining the minimum value of λ. The main difference is that how many MSE are calculated in order to identify λ. Specifically, k-fold cross validation method randomly select k groups of subsample and λ MSE values are obtained. In another words, the procedure is repeated k times; each time, a different group of observations is treated as a validation set. This process results in k estimates of the test error, MSE_1, $\mathrm{MSE}_2, \ldots, \mathrm{MSE}_\lambda$. The k-fold CV (cross validation) is computed by averaging these values, $\mathrm{CV}_{(k)=1/k}\sum_{i=1}^{k}\mathrm{MSE}_i$.

normality, the results are almost stable even using different percentage levels of training samples from 60% to 85%. Therefore, I use non-filtered levels of IDPs as response.

3.2.3 Logistic Regression

(1) Model description

The purpose of using logistic regression is to verify whether the recession index can measure U.S economic recession. The logistic regression can predict the probability that an observation falls into two categories of dichotomous dependent variables by using one or more independent variables that can be either continuous or categorical. Economic recession is a type of dichotomous variable. In the econometric analysis, the occurance of recession is regarded as 1; otherwise, it is zero. Actually, this is one of the main assumptions of implementaion of the logistic regression model.

In Chapter 6, macro indicators measuring the changes of economic environment are regarded as the independent variable and this is a continuous variable. Then the second assumption of one or more independent variables being either continuous or categorical in using logistic regression is also satisfied. The third assumption is that the observation should be independent and the dependent variable should have mutually exclusive and exhaustive categories. These assumptions are also fulfilled. In this study, each macro indicator is individually investigated with economic recession rather than model them together.

Although this model does not need a linear relationship between the dependent variable and independent variables, it requires that there should be a linear relationship between continuous independent variables and the logit transformation (also titled log odds) of the dependent variables. The assumption should be tested after modelling logistic regression. In addition, there is no strict assumption of normality in the independent variables and error terms in the estimated logistic regression models. This is the same case as the homoscedasticity and logistic regression does not require the homoscedasticity in the level of independent variables. However, the error terms need to be independent.

Logistic regression model for explaining recession in relation to macro indcaior is derived as follows:

$$P_t = E(Y = 1|\, X_t) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_t)}}, \tag{7}$$

where the dependent variable Y_t is a recession indictor, which equals one if there is recession at date t or equals zero if there is no recession. X_t is is an independent variable, macro indicator at date t. β_1 is a scalar parameter. This equation can be also written as:

$$P_t = \frac{1}{1 + e^{-Z_t}} = \frac{e^{Z_t}}{1 + e^{Z_t}}, \tag{8}$$

where $Z_t = \beta_1 + \beta_2 X_t$. Equation (7) is the logistic distribution function. If P_t is the probabilty of recession occuring, then $(1 - P_t)$, the probability of no recession, is:

$$1 - P_t = \frac{1}{1 + e^{Z_t}}. \tag{9}$$

Equation (9) can be further written as:

$$\frac{P_t}{1 - P_t} = \frac{1 + e^{Z_t}}{1 + e^{-Z_t}} = e^{Z_t}. \tag{10}$$

Then $P_t/(1 - P_t)$ is the odds ratio in favour of occurance of the recession – the ratio of the probabilty that a recession will occur to the probability that it will not occur. If both sides in Equation (10) are log transformed, then the new equation is:

$$L_t = \ln\left(\frac{P_t}{1 - P_t}\right) = Z_t = \beta_1 + \beta_2 X_t. \tag{11}$$

Equation (11) is the logit model. For the interpretation of the variables, if L_t, the logit, is positive, meaning that when the value of the regressor increases, the odds that the regressand equals 1 (meaning recession happens) increases; otherwise, it decreases as the value of X increases. For the slope, β_2, measures the change in L for a unit change in X, that is, it tells how the log-odds in favour of recession occuring change as macro indicator changes by a unit. The intercept β_1 is the value of the log-odds in favour of recession if indicator is zero. Like most interpretations of intercepts, this interpretation may not have any physical meaning (Gujarati, 2003).

For estimation purpose, Equation (11) is as follows:

$$L_t = \ln\left(\frac{P_t}{1-P_t}\right) = Z_t = \beta_1 + \beta_2 X_t + u_t, \quad (12)$$

where u_t is the stochastic error term. Since this model focuses on individual level, then OLS cannot be used for estimating the parameters but maximum likelihood (ML).

(2) Measures of Fit for Logistic Regression

There are two major issues in this section: measures of goodness of fit tests and predictive power.

a. Measures of goodness of fit tests

The first issue of goodness of fit tests is to evluate model fit. Three methods are used in this study. The first is based on deviance and Pearson residuals by using chi-square test in order to test whether the null hypothesis of the fitted model is correct. The p-value is higher suggesting that the null hypothesis cannot be rejected and logit model fits correctly.

Suppose we have just fitted a model and want to assess how well it fits the data. A measure of discrepancy between observed and fitted values is the deviance statistic, which is given by:

$$D = 2\sum\left\{y_t \log\left(\frac{y_t}{\hat{u}_t}\right) + (n_t - y_t)\log\left(\frac{n_t - y_t}{n_t - \hat{u}_t}\right)\right\}, \quad (13)$$

where y_t is the observed and $\hat{u}_t$ is the fitted value for the t-th observation (date). n_t is the number of observations at date t. Note that this statistic is two times a sum of 'observed times log of observed over expected', where the sum is over both successes and failures (i.e. we compare both y_t and $n_t - y_t$ with their expected values). In a perfect fit the ratio observed over expected is one and its logarithm is zero, so the deviance is zero. That is, the smaller the deviance, the closer the fitted value is to the saturated model. The larger the deviance, the poorer the fit. The distribution of this value converges to a chi-squared distribution with $n - p$ d.f., where n is the number of groups and p is the number of parameters in the model, including the constant.

For reasonably large groups n, the deviance provides a goodness of fit test for the model (Gujarati, 2003). With individual data the distribution of the deviance does not converge to

a chi-squared distribution, and cannot be used as a goodness of fit test. An alternative measure of goodness of fit is Pearson's chi-squared statistic, which for binomial data is written as:

$$\chi_P^2 = \sum_t \frac{n_t(y_t - \hat{u}_t)^2}{\hat{u}_t(n_t - \hat{u}_t)}. \tag{14}$$

Each term in the sum is the squared difference between observed and fitted values y_t and $\hat{u}_t$, divided by the variance of y_t, which is $u_t(n_t - u_t)/n_t$, estimated using $\hat{u}_t$ for u_t. Both approaches use the same standards since they are all based on chi-squared distribution. If the value is larger than 5% significance level, then the model fits correctly; otherwise, it does not fit well.

Secondly, I use a specification test (IDRE, 2016). This test aims to find whether amongst all the relevant variables that we have not included any variables that should not be in the model, and the logit function is a linear combination of the predictors. This model only has an independent variable, so the second purpose of using this test is that the logit function, as the link function, is not the correct choice, or alternatively the relationship between the logit of outcome variable and the independent variables is not linear.

The test is based on the idea that if the model is properly specified, one should not be able to find any additional predictors that are statistically significant except by chance. The test is to build the relationship between linear predicted value (hat) and linear predicted value squared (hatsq) as the predictors and the dependent variable, recession by logit model. The variable hat should be a statistically significant predictor, since it is the predicted value from the model. This will be the case unless the model is completely misspecified. On the other hand, if our model is properly specified, variable hatsq shouldn't have much predictive power except by chance. Therefore, if hatsq is significant, then the test is significant.

b. Measures of predictive power

The second issue of predictive power is the approach of finding a statistic that measures how well the model can predict the dependent variable based on the independent variable. Two types methods are used in this study, pseudo R-squared (like R-squared in the OLS estimation) and the area under the receiver operating characteristic curve (ROC) curve.

The method for calculating pseudo R-squared is based on the approach proposed by McFadden (1973). The method is defined as follows. Let L_0 be the value of the likelihood function for a model with no predictors, and let L_M be the likelihood for the model being estimated. McFadden's R^2 is defined as

$$R^2_{McF} = 1 - \ln\left(\frac{L_M}{L_0}\right). \quad (15)$$

The rationale for this formula is that $\ln(L_0)$ plays a role analogous to the residual sum of squares in linear regression. Consequently, this formula corresponds to a proportional reduction in "error variance". Following prior studies, this value is regarded as a pseudo R-squared. There are two main interpretation for pseudo R-squared. One is that pseudo R-squared is regarded as explained variabilty. The denominator of the ratio can be considered as the total variability in the dependent variables, or how much a dependent variable varies from its mean value. The numerator of this ratio is the variablity in the dependent varibale that is not predicted by the model. Then this ratio represents the percentage of the total variability unexplained by the model. After substracting this ratio by one, the results mean that the proportion of the total variability explained by the model. Then the more variabiltiy explained, the better the model. Another is that pseudo R-squared can be thought of as improvement from null model[12] to fitted model. Specifically, the denominator of this ratio can be considered as the sum of squared errors from the null model. In the null model, each value in the dependent variable is explained to be the mean of the values of dependent variables, assuming that we predict a dependent variable without having any additional information about what we were predicting. The mean of the dependent variables would be the best result if the aim is to minimise the squared difference between the predicted value and the real values of dependent variables. Then the numerator of the ratio should be the sum of squared errors of the fittted model. This ratio is the indicator of how much the model parameters improve upon the prediciton of the null model. If this ratio is smaller, it means that there is a great improvement in the model with a higher pseudo R-squared. According to

[12] Null model represents the model predicting the dependent variable without any independent variables.

Domencich and McFadden (1975), if the McFadden's pseudo R-squared is from 0.2 to 0.4, then the model fit data well.

ROC is used for quantitfying whether the predictive performance is good or not. ROC is a standard technique for summarizing classifier performance over a range of trade-offs between true positive and false positive error rates (Swets, 1988). ROC curve is a plot of sensitivity (the ability of the model to predict an event correctly) versus 1-specificity for the possible cut-off classification probability values π_0. Here we need to use the area under the curve (AUC) and it is an accepted traditional performance metric for a ROC curve. The higher the area under the curve, the better prediction power the model has; 0.8 can be interpreted to mean that a randomly selected individual from the positive group has a test value larger than that for a randomly chosen individual from the negative group 80 percent of the time.

3.2.4 LASSO based on Logistic Regression

For the logit model, LASSO aims to estimate the parameters by minimising the negative log-likelihood function with a roughness penalty placed on the sum of absolute value of the covariate parameters, which is called penalty form l_1 penalty (James et al., 2013). Equation (11) is a specific case for logistic regression since it only covers one variable. In this part, LASSO should select the key variables for the recession based on the logistic regression. Then we extend Equation (11) by allowing for more than one hundred factors. Thus extended logistic regression is defined as follows:

$$\pi_t = P(Y_t = 1|X_t', \beta) = \frac{1}{1 + \exp(-\beta^T X_t')}, t = 1, \dots, n \tag{16}$$

where X_t' is the vector $(1, X_t)^T$ and β is the vector $(\beta_0, \beta_1, \dots, \beta_t)^T$. For the OLS regression model, the joint likelihood across all t is expressed as:

$$\begin{aligned} L(\beta|y_1, \dots, y_n) &= \prod_{t=1}^{n} (\pi_t)^{y_t}(1 - \pi_t)^{1-y_t} \\ &= \prod_{t=1}^{n} \left(\frac{1}{1 + \exp(-\beta^T X_t')}\right)^{y_t} \left(\frac{\exp(-\beta^T X_t')}{1 + \exp(-\beta^T X_t')}\right)^{1-y_t}, \end{aligned} \tag{17}$$

and so the joint log-likelihood for the vector β is:

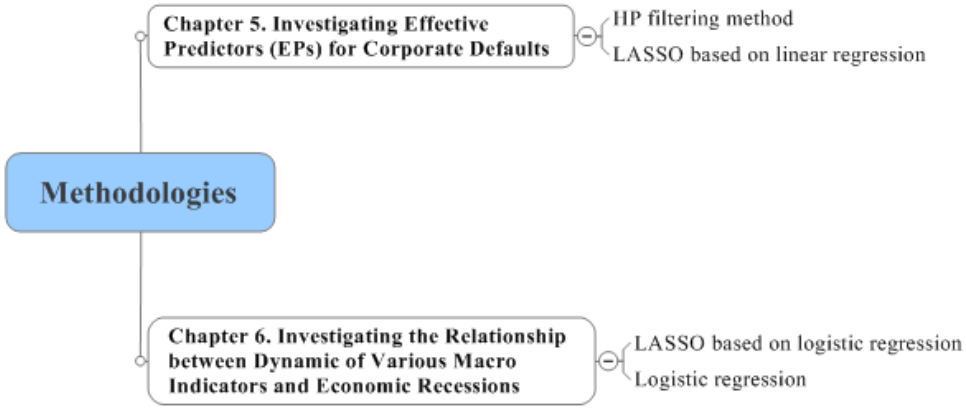
$$\begin{aligned} l(\beta) &= \ln\big(L(\beta\,|y_1,\dots,y_n)\big) \\ &= \sum_{t=1}^{n}[y_t \ln\left(\frac{1}{1+\exp(-\beta^T X_t')}\right) + (1 \\ &\quad - y_t)\ln\left(\frac{\exp(-\beta^T X_t')}{1+\exp(-\beta^T X_t')}\right)] \\ &= -\sum_{t=1}^{n}[(1-y_i)\beta^T X_t' + \ln(1+\exp(-\beta^T X_t'))] \end{aligned} \tag{18}$$

In order to perform logistic likelihood regression, the Lagrangian multiplier is added. By adding a Lagrangian penalty to the joint log-likelihood in Equation (18), we have the following constrained maximization equation:

$$l(\beta)_c = \left\{ = -\sum_{t=1}^{n}[(1-y_i)\beta^T X_t' + \ln(1+\exp(-\beta^T X_t'))] - \lambda\sum_{j=1}^{p}|\beta_j| \right\}. \tag{19}$$

where $l(\beta)_c$ is defined as the constrained log-likelihood, $\lambda \geq 0$ is the Lagrangian multiplier and the intention remains to find the vector $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p{}^T)$ that maximizes $l(\beta)_c$. Finally, Figure 3-1 provides the diagram to explain where each of the techniques is used in each following chapters.

Figure 3-1 Diagram of the methodologies



3.3 New Developed Analytical Tools

3.3.1 R Package "Kai.R.Tools"

This package named "Kai.R.Tools" is developed by the thesis author using the R programming language, and it is used in R statistical software. The manual is given in Appendix 9.2. This package can generate data which can be used in both firm level and industrial level of studies on corporate default risk.

The main purposes of this package are twofold. One is data management. This package can manage and merge various databases, such as Moody's DRD, daily/monthly CRSP, and quarterly/annual Compustat (or CCM) in order to extract core information including corporate defaults, bond defaults, and corresponding default probabilities, and various firm-specific risk drivers consisting of accounting ratios and two types of distance to default (DTD) etc. Another is that this package is able to execute a set of empirical analyses such as network analysis and construction of various indicators.

Figure 3-2 shows the structure of R package "Kai.R.Tools". There are three modules that have different roles in this package. They can be used either independently or dependently. For example, if extracting key micro factors (known as firm-specific risk drivers) for forecasting corporate defaults probabilities is the main interest, then both Module A and B are used to get industrial level of micro factors and DPs, such as IDPs and VDPs. Specifically, Module A can help me to get all the micro factors, such as accounting ratios, market-driven variables, market size, past stock returns, the idiosyncratic standard deviation of stock returns, and various DTDs. Module B can help me identify the DPs. Then I can use variable selection techniques to find the key variables for explaining historical DPs.

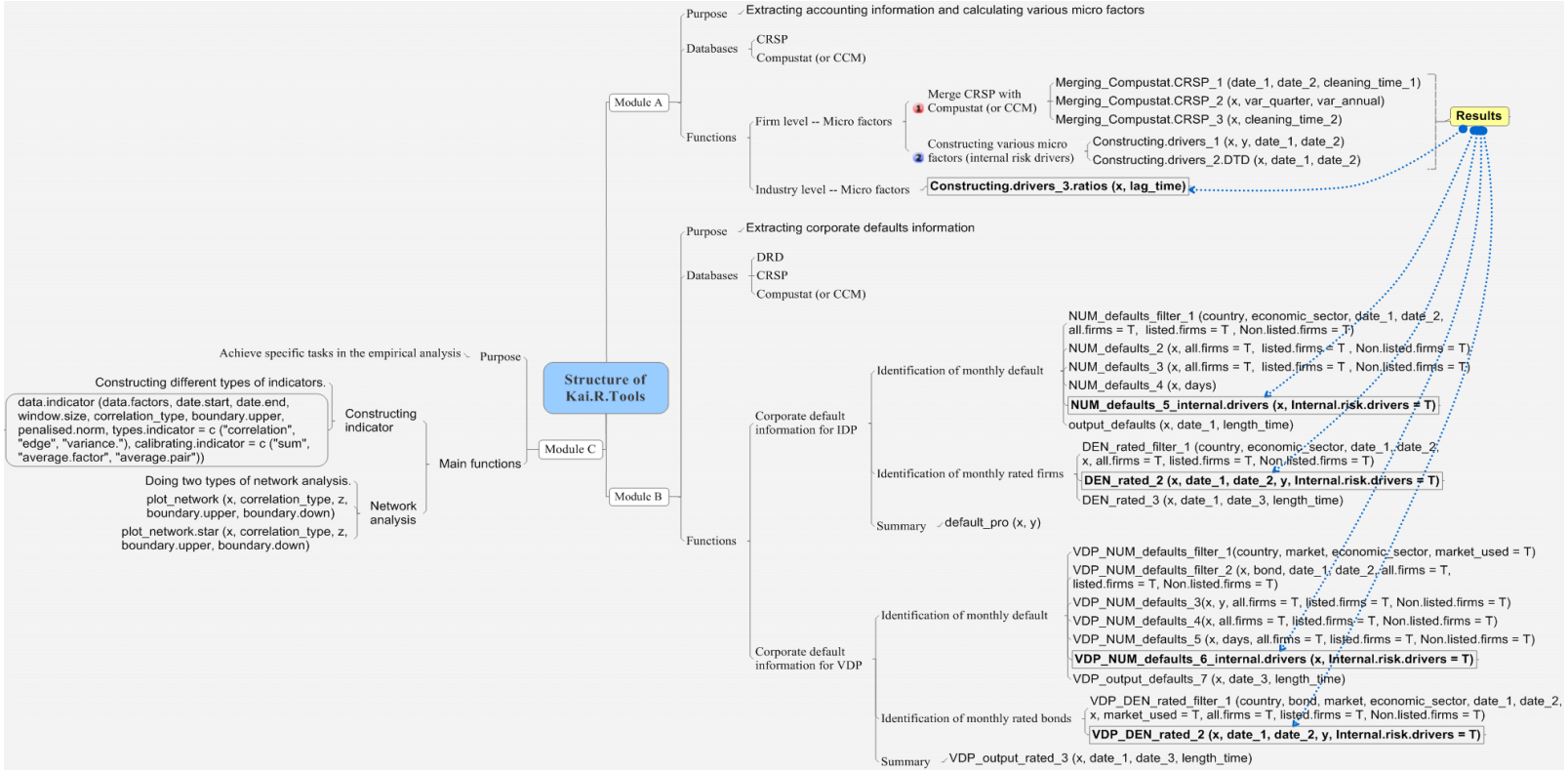
If I am interested in the firm-level of study on corporate defaults, I need to use the results from functions for Firm level–Micro factors in Figure 3-2. The result is the panel data for different firms' firm-specific variables and dummy variable representing corporate defaults. Based on this result, I can do the same analysis done by Tian et al. (2015) and Campbell et al. (2008).

Module C can be used in different scenarios or databases, which depends on the research object. The function entitled 'data.indicator' can be used to construct indicators based on the

results of Industry level–Micro factors in Figure 3-2. Results can also be used in the network analysis in order to find the dynamic correlation among various micro factors by the function 'plot_network'. In fact, there are some other useful functions in this toolbox, which have not been mentioned in the manual. Since they are more relevant to the specific research topic, I did not add them to the manual. In particular, if future studies are interested in how firm-specific variables impact the other dependent variables, then the panel results from using functions in the Firm level-Micro factors in Figure 3-2 can be directly merged with another database that has the other dependent variables. That is, this package is helpful to do the research in areas other than credit risk management.

Thus, this package has several salient benefits. First, based on different combinations of various parameters, such as country code, industry code, time span, and screening standards for the defaults, it can generate different types of corporate and bond default database by only using several main functions without changing the codes in the do file. Second, using these functions is flexible and easy since the functions can be combined in different ways in order to consider whether the relevant screening assumptions are used in practice. Third, this package can generate the database which almost satisfies the majority of studies in credit risk management. Fourth, this package can do extra empirical analysis, for example it can execute network analysis based on correlation method.

Figure 3-2 Structure of R package titled "Kai.R.Tools"



Note: the detailed description of each function can be found in the manual for "Kai.R.Tools" in Appendix 9.3. The connection among these functions and their implementation are also shown in the manual.

3.3.2 Shiny Web Applications

In order to do interactive exploration, I designed and developed two web applications including KAI.APP.PhD and KAI.APP.EIA by Shiny based on R programming languages. Both apps (short for applications) are intended to do basic empirical analysis in support of my Ph.D. project. Basically, they have different roles in my Ph.D. project which can achieve their specific purposes.

Specifically, KAI.APP.PhD has three main purposes. The first purpose is to find whether there is a relationship between the correlation among specific group of factors (known as risk drivers) and economic recession and IDPs by using network analysis. The second is to construct various indicators based on different algorithms[13] and analyse the relationship between each indicator and IDPs by using various statistical methods, such as correlation analysis, causality analysis, and statistical inference analysis. In particular, a set of relevant model diagnostics tests can be done in this app in order to verify whether the data fit the statistical models well. The third purpose is to investigate whether these indicators can explain economic recession by using logistic regression. By using LASSO, this app can do variable selection for dummy variable representing economic recession. A group of effective factors are extracted which can strongly explain historical economic recession. Based on this group of factors, this app can construct the indicator for further determining whether this indicator can explain economic recession.

'EIA' in the name of the application KAI.APP.EIA represents 'Effective Indicator Analysis'. Compared with KAI.APP.PhD, this app has some common features. For example, KAI.APP.EIA uses similar empirical techniques to KAI.APP.PhD. In contrast, there are two major differences in the research purposes. First, KAI.APP.EIA focuses on the selected group of effective factors that can explain historical IDPs by using LASSO. Second, this app is able to investigate dynamic change in the group of effective factors for explaining IDPs. KAI.APP.EIA can provide empirical results in the chapter in my thesis which is related to

[13] The detailed description of how to construct various indicators is shown in the manual called "Kai.R.Tools".

investigation of dynamic change in the EPs. This is why this application is named 'Effective Indicator Analysis' since this app focuses on the effective factors rather than all the macro factors in the different groups in KAI.APP.PhD.

It should be noted that KAI.APP.PhD provides basic empirical results in my first two papers called *Macroeconomic Conditions, Corporate Defaults, and the Adaptive Market Hypothesis: Principal Indicator and Early Warning Signal* and *A New Approach to Measuring Economic Recessions*, respectively.

Neither application is restricted to my current databases. If I am interested in the other countries' data, I am still able to use two apps in practice. Then there is a flexibility of using them. The details of using these apps in practice can be found in Appendix 9.3.

3.4 Conclusion

This chapter describes the methodologies and new IT developed techniques which are used to do empirical analysis in support of the contributions in this thesis. In the first methodology section, three main econometric methods are presented: HP filtering method, VAR, and VECM. The rest including LASSO, linear regression, and logistic regression models are used to do statistical inference tests.

As for the new analytical tools, they can also be used widely in the relevant areas. For instance, by using the "Kai.R.Tools" package, I can obtain various types of database. This package can help me construct firm-level default databases and both microeconomic and macroeconomic databases, enabling me to do the research on default risk prediction at the firm level. Based on a country's identifier, industry sector identifier, and time span in Moody's DRD, I can finalise different countries' default databases for future potential studies. Meanwhile, the web applications can also be used in the other empirical research if I can obtain relevant data inputs following the requirements in the manual. More importantly, based on the current structure of applications, I can develop more applications for doing more sophisticated interactive exploration, which can achieve a specific research purpose.

CHAPTER 4

4. Data

4.1 Introduction

This chapter describes the data used in this Ph.D. thesis. It also shows the sources of data, management of the data including cleaning methods and extracting methods, interpretation of the data and calculations.

My Ph.D. project uses four major databases: Moody's Default Recovery Database (DRD), Federal Reserve Economic Database (FRED), Center for Research in Security Prices (CRSP), and Compustat Merged Database (CCM), respectively. The description for each database is presented below:

1) DRD provides rating transition and default history for all rated US industrial firms. In particular, the firms rated by Moody's can be either listed or non-listed firms. The non-listed firms do not want to release their company's information to the public, and they do not intend to be financed by equity and controlled by the shareholders. However, these firms need to finance by borrowing debts from banks, issuing debts through other financial derivatives in the financial market. Ratings are a critical element in pricing securities and are often used as a benchmark for setting investment guidelines. For example, debt issuers with higher-rated ratings can reduce funding cost for financing and further influence the prices of debt traded in the market.

2) FRED is a database maintained by the research division of the Federal Reserve Bank of St. Louis that has more than 247,000 economic time series from 79 sources. They

cover banking, business/fiscal, consumer price indices, employment and population, exchange rates, gross domestic product, interest rates, monetary aggregates, producer price indices, reserves and monetary base, US trade and international transactions, and US financial data. The time series are compiled by the Federal Reserve and collected from government agencies, such as the US Census and the Bureau of Labour Statistics.

3) CRSP only provides stock prices, returns and additional 'event' information on securities for the NYSE, AMEX, NASDAQ, and Arca exchanges and include CRSP broad market indices. Bennin (1980) figures that CRSP covers the NYSE stocks in early stage rather than containing AMEX and NASDAQ.

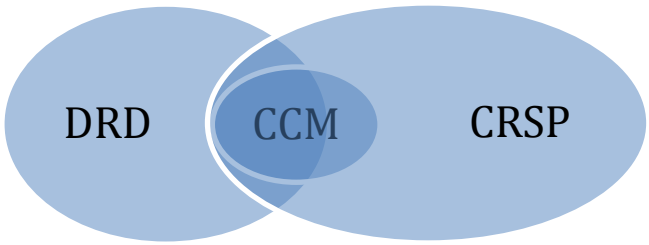
4) CCM is provided by the CRSP database. This database is constructed by merging CRSP with Compustat. Compustat is a database of financial, statistical and market information on active and inactive global companies throughout the world. The service began in 1962 and it provides more than 300 annual and 100 quarterly income statements, balance sheets, statements of cash flow, and supplemental data items. According to Ulbricht and Weiner (2005), the particular file called Compustat North America is the most complete database of US and Canadian accounting data with 10,000 actively traded US firms and 10,900 inactive firms as well as 1,100 Canadian firms. This database consists of most public listed firms in the U.S.

The relation between three databases is shown in Figure 4-1. CRSP contains all the firms in CCM. The firms in DRD do not exist in CRSP and CCM as these firms are not listed in the NYSE, AMEX, or NASDAQ; that is, the interconnected area with darkest blue colour in Figure 4-1 for three databases is the listed firms existing in three databases. I use them to construct a corporate defaults database, a macroeconomic database, and a microeconomic database.

DRD is used to extract corporate defaults information in the corporate defaults database, such as corporate (or bond) defaults, exposures (known as rated firms or rated bonds), and corporate (or bond) defaults probabilities. If DRD is merged with CCM and CRSP, then the corporate default information can be further categorised into two main groups including

listed firms and non-listed firms. The macroeconomic database is built by FRED, which provides 114 macroeconomic and financial time series in this research project. These variables are regarded as macro factors. Both CRSP and CCM are used for providing the equity and accounting information in order to construct a microeconomic database that contains 34 firm-specific drivers (known as micro factors).

Figure 4-1 Relation graph among the three databases



The time span for all the data is from January 1980 to December 2014. Three reasons of why the initial date can be January 1980 are presented as follows. Although Moody's started to record the default events in 1898, the quality of default records is not reliable from 1898 to 1980. Another reason is that there is no available defaulted data before January 1, 1980 that can match CRSP data (or CCM) based on CUSIP value.[14] Lando and Nielsen (2010) identify that the material is sparse until 1982 in terms of corporate default data in DRD. Third, in order to cover more macro factors, I can only start from January 1980.

The remainder of this chapter proceeds as follows. Section 4.2 is to describe the corporate default database, which is followed by the macroeconomic database. Section 4.3 is about the description of the microeconomic database. Section 4.4 is the conclusion for this chapter.

[14] Most securities are identified by a nine-character (letters and numbers) CUSIP stands for Committee on Uniform Security Identification Procedures. The CUSIP number is not determined randomly but is assigned in such a way so as to identify an issue's key differentiating characteristics within a common structure. Specifically, the first six characters identify the issuer: the corporation, government, agency, or municipality. The next two characters identify whether the issue is debt or equity and issuer of the issue. The last character is simply a check character that allows for accuracy checking and is sometimes truncated or ignored; that is, only the first characters are listed.

4.2 Corporate Default Database

The corporate default database draws elements from DRD, and both CRSP and CCM are used to identify the listed and privately-held firms. Based on different parameters, such as country identifier, industrial sector identifier, and time span etc., I can use the package "Kai.R.Tools" to obtain various default information for the bonds or firms. In this chapter, I show a case of getting the time series of IDPs, which counts the number of monthly defaults over the rated firms in the beginning of each month. The time span is from January 1980 to December 2013. Next I will show the definition of defaults, method of measuring IDPs, and cleaning methods.

4.2.1 Definition of Defaults in Moody's

According to Moody (2015), default is applicable only to debt or debt-like obligations (e.g., swap agreements). Four events constitute a debt default under Moody's definition:

1) A missed or delayed disbursement of a contractually-obligated interest or principal payment (excluding missed payments cured within a contractually allowed grace period), as defined in credit agreements and indentures;

2) A bankruptcy filing or legal receivership by the debt issuer or obligor that will likely cause a miss or delay in future contractually-obligated debt service payments;

3) A distressed exchange whereby: (1) an obligor offers creditors a new or restructured debt, or a new package of securities, cash or assets that amount to a diminished financial obligation relative to the original obligation and (2) the exchange has the effect of allowing the obligor to avoid a bankruptcy or payment default in the future;

4) A change in the payment terms of a credit agreement or indenture imposed by the sovereign that results in a diminished financial obligation, such as a forced currency re-denomination (imposed by 38 MOODY'S INVESTORS SERVICE RATING SYMBOLS AND DEFINITIONS the debtor, himself, or his sovereign) or a forced change in some other aspect of the original promise, such as indexation or maturity.

4.2.2 Method of Calculation of IDPs

IDPs are fractions in which the numerator represents the number of issuers that defaulted in a particular time period and the denominator represents the number of issuers that could have defaulted in that time period. The formula is shown below:

$$IDPs = \frac{d_t^{IDPs}}{n_t^{IDPs}} \tag{54}$$

Under this method, this research defines the outstanding number of credits (n_t) recursively as:

$$n_t^{IDPs} = n_{t-1}^{IDPs} - d_{t-1}^{IDPs} - w_t^{IDPs}. \tag{55}$$

The numerators d_t^{IDPs} and d_{t-1}^{IDPs} are the numbers of issuers defaulting at period t and $t-1$. The denominators n_t^{IDPs} and n_{t-1}^{IDPs} are the numbers of issuers that potentially could have defaulted at date t and $t-1$. w_t^{IDP} denotes the number of credits which withdraw between periods t and $t+1$ (Acharya and Carpenter, 2002, Gansecki, 2010, Moody, 2015).

For example, if 25 issuers defaulted in December 2013, and that static pool consisted of 2,000 issuer ratings, the resulting monthly default rate for all ratings in December would be 1.3% (25/2000). While the calculation of static pool allows for the credits that withdraw and default, it is the number of issuers in November 2013 (2020 issuer ratings) minus 10 issuers which defaulted in November 2013 minus 10 issuers which withdraw between November 2013 and December 2013.

4.2.3 Cleaning Methods

This section is to illustrate the procedures of cleaning default data, such as how to identify the defaulters and rated firms. Basically, this thesis uses the cleaning methods used by three papers including *Multi-period corporate default prediction with stochastic covariates* written by Duffie et al. (2007), *Correlation in corporate defaults: Contagion or conditional independence?* written by Lando and Nielsen (2010), and *When do firms default? A study of the default boundary* written by Davydenko (2012).

Two cleaning methods are used for screening default data:

1) Cleaning the family structure including clearing firms that changed their parent firm and then coping with dependent default events. In DRD, there are some firms that changed their parent firm in the past since they experienced restructure events, such as expansion, acquisition, and merger. For this case, this thesis should identify these special firms in order to do the next stage of handing dependent default events. Actually, this procedure does not change the number of firms after implementation of this method.

   As Davydenko (2012) notes, the default events in Moody's default database are not independent. Two approaches are then applied to clean family structure for handling the dependent events. One is related to the case of both parent company and its subsidiaries default simultaneously, since firms often default together with their wholly owned subsidiaries. If both the parent company and its subsidiaries issue Moody's-rated bonds, several entries may be registered in DRD for different legal entities, even if the defaults all result from one default decision by the parent company. This scenario is especially common for financial subsidiaries of industrial firms. To handle this scenario, the parent's consolidated financial information is used to study the default decision for the whole group of bond issuers (Davydenko, 2012). Another is to handle the multiple default events within a short period. In this scenario, this study follows prior studies and combines defaults for each company if they happen within two years of the first event (Duffie et al., 2007, Davydenko et al., 2012).[15]

2) Deleting the firms that have not enough stock price data. This procedure is used by Duffie et al. (2007) and Lando and Nielsen (2010). In practice, this study only keeps the firms that have more than 6 months' stock price in CRSP; otherwise, they should

[15] Lando and Nielsen (2010) use one year instead of two years to combine default events for each compay. Since using one year could generate more default events in comparision with using two years, then this could potentially change the results. However, majoriy of studies concern multiple default events based on two years. In this study, I use the approach used by most of prior studies.

be deleted. It should be noted that this standard is used based on whether the study considers listed firms or not.

These adjustments reduce the number of defaults by industrial U.S. firms in DRD between 1980 and 2014 from 2210 (including 1942 firms) to 543 (including 459 firms).

4.3 Macroeconomic Database

A second database is constructed from FRED, which contains 114 macroeconomic factors in Table 4-1. This thesis uses the approach of categorising the variables which is based on Koopman et al. (2011). Then 114 variables can be grouped into five broad categories: (a) bank lending conditions; (b) macroeconomic and business cycle indicators, including labour market conditions and monetary policy indicators; (c) open economy macroeconomic indicators; (d) micro-level business conditions such as wage rates, cost of capital, and cost of resources; and (e) stock market returns and volatilities.

Based on whether the variable contains current, past, or forward looking information, these variables, known as cyclical indicators, are further classified into three main groups: leading indicators, lagging indicators, and coincident indicators. They are shown in the column called Frequency/Classification in Table 4-1. According to The Conference Board (2001) and Shiskin and Moore (1967), cyclical indicators are classified into three different groups including leading, coincident, and lagging indicators based on the timing of their movements. Leading indicators tend to move direction before the business cycle occurs and the timing comparisons between the cyclical turn in the indicators and business cycle turn is one-month or more, such as new orders, housing starts, and consumer sentiment. Lagging indicators tend to shift direction after the business cycle occurs or coincident series occur, such as interest rate, unit labour cost, and cost of resources. For the timing comparisons, these indicators follow the business cycle turn by one month or more. Coincident indicators are the series that measure aggregate economic activity which are used to define the business cycle. For example, these indicators are employment, production, and personal income etc. Coincident indicators are special in terms of timing comparisons since they may follow or lead turning points in the business cycle by three months or less, including that they completely coincide

with the business cycle. Then there are 42, 38, and 34 macro factors in the groups of leading indicators, lagging indicators, and coincident indicators, respectively.

The macro factors are monthly time series from January 1980 to December 2014. 89 drivers are all directly downloaded from FRED without any transformation from the other frequency to the monthly type. The rest of them (25 variables) are quarterly and annually, such as gross domestic product (GDP), gross national product (GNP), and Federal debt: total public debt etc. In practice, cubic spline interpolation is used for converting quarterly and annual data to monthly data.[16] This method is to use piecewise polynomial function model to smooth the trend in each space (if there are n quarterly data points, there are $n-1$ spaces). All variables except exchange rates, interest rates, and industrial material price index are seasonally adjusted.

Table 4-1 provides a summary statistics for 114 macro factors. For some varibales, their mean value is equal to the median value, which means that these variables follow normal distribution. The others are either postively or negatively skewed. Figure 4-2 shows the heat map of correlation matrix in the lower triangle for 114 macro factors. Since I have 114 macro factors, it is thus not effective and efficient to use the table to illustrate the results.[17] If the colour tends to be darker, the correlation between two variables becomes higher; otherwise, they become lower. White colour represents weak correlation. Blue colour represents positive strong or modertate correlation. Brown colour represents negative strong or moderate correlation. It can be seen the correlation among these macro factors are not really high since the white joins the main colour. Also the higher correlation happens in certain variables, such as TCU, CBI, and FSDP tend to be highly correlated with the other variables.

[16] The article in the following link shows that cubic spline interpolation has been widely used for economists to convert quarterly data to monthly data (http://columbiaeconomics.com/2010/01/20/how-economists-convert-quarterly-data-into-monthly-cubic-spline-interpolation/).

[17] The results can be provided upon request.

Table 4-1 Summary statistics for 114 macro factors

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
LOANS	420	0.005	0.006	-0.017	0.002	0.006	0.008	0.047
REALLN	420	0.006	0.006	-0.016	0.002	0.007	0.01	0.046
TOTALSL	420	0.005	0.005	-0.008	0.003	0.005	0.008	0.044
BUSLOANS	420	0.004	0.009	-0.029	-0.001	0.005	0.01	0.035
CONSUMER	420	0.004	0.015	-0.029	-0.0005	0.003	0.007	0.271
BOWACBM027SBOG	420	0.005	0.02	-0.066	-0.006	0.005	0.018	0.116
IBLACBM027SBOG	420	0.0002	0.046	-0.286	-0.018	0.004	0.025	0.233
TDSP	420	-0.0001	0.004	-0.013	-0.003	0.0001	0.002	0.016
FODSP	420	-0.00002	0.004	-0.012	-0.002	-0.00001	0.002	0.016
GFDEBTN	420	0.007	0.005	-0.006	0.004	0.007	0.01	0.024
INDPRO	420	0.002	0.007	-0.044	-0.002	0.002	0.005	0.02
RECPROUSM156N	420	-0.005	0.873	-3.839	-0.405	0	0.423	3.091
IPMANSICS	420	0.002	0.008	-0.035	-0.002	0.002	0.006	0.027
IPMINE	420	0.0005	0.015	-0.124	-0.006	0.001	0.008	0.084
IPUTIL	420	0.001	0.02	-0.087	-0.011	0.002	0.013	0.07
IPMAT	420	0.002	0.008	-0.065	-0.002	0.003	0.007	0.028
IPN213111S	420	-0.001	0.044	-0.223	-0.02	0.002	0.02	0.153
CMRMTSPL	420	0.002	0.009	-0.032	-0.004	0.002	0.008	0.035
NAPMNOI	420	-0.0003	0.079	-0.314	-0.039	-0.001	0.038	0.405
NAPMII	420	0.00003	0.07	-0.308	-0.04	0	0.044	0.243
NAPMSDI	420	-0.00004	0.043	-0.214	-0.024	0.002	0.022	0.145
M12MTVUSM227NFWA	420	0.002	0.002	-0.004	0.001	0.002	0.003	0.01

NAPM	420	-0.0002	0.046	-0.241	-0.023	0	0.021	0.262
DSPIC96	420	0.002	0.008	-0.066	0.0001	0.002	0.005	0.047
PI	420	0.005	0.006	-0.055	0.003	0.004	0.006	0.045
PCE	420	0.005	0.005	-0.02	0.002	0.004	0.007	0.027
PCEDG	420	0.004	0.027	-0.159	-0.009	0.004	0.018	0.135
USPHCI	420	0.002	0.002	-0.005	0.001	0.002	0.003	0.006
PCEPI	420	0.002	0.002	-0.011	0.001	0.002	0.003	0.012
GDP	420	0.004	0.003	-0.007	0.003	0.004	0.006	0.018
A001RP1Q027SBEA	420	-0.033	0.62	-7.254	-0.135	-0.005	0.126	2.85
W068RCQ027SBEA	420	0.005	0.005	-0.014	0.002	0.004	0.007	0.029
UMCSENT	420	0.001	0.051	-0.199	-0.026	-0.001	0.029	0.22
PNFI	420	0.004	0.007	-0.028	-0.0005	0.005	0.009	0.021
GPDI	420	0.004	0.013	-0.045	-0.002	0.004	0.012	0.051
UNRATE	420	-0.0002	0.027	-0.089	-0.018	0	0.017	0.091
HOHWMN02USM065S	420	0.0001	0.006	-0.05	-0.002	0	0.002	0.06
FRBLMCI	420	0.323	4.025	-13.646	-0.495	0.069	0.868	15.676
EMRATIO	420	-0.00004	0.003	-0.008	-0.002	0	0.002	0.009
LNU03025703	420	0.004	0.077	-0.264	-0.041	0.006	0.053	0.271
CE16OV	420	0.001	0.003	-0.009	-0.001	0.001	0.002	0.015
UEMP15OV	420	0.003	0.046	-0.183	-0.024	0.0001	0.03	0.151
MANEMP	420	-0.001	0.004	-0.023	-0.002	-0.0004	0.001	0.008
NAPMEI	420	-0.0001	0.056	-0.223	-0.031	0.002	0.031	0.252
PAYEMS	420	0.001	0.002	-0.006	0.0001	0.001	0.002	0.012
UEMPMEAN	420	0.003	0.03	-0.087	-0.017	0	0.023	0.086

LNS12032198	420	0.002	0.053	-0.228	-0.027	0.0004	0.03	0.309
HSN1F	420	-0.0004	0.075	-0.41	-0.048	0	0.046	0.237
MCUMFN	420	-0.0002	0.008	-0.035	-0.004	0	0.004	0.025
TCU	420	-0.0002	0.007	-0.045	-0.004	0	0.004	0.019
ALTSALES	420	0.0004	0.071	-0.444	-0.025	0.002	0.033	0.302
HOUST	420	-0.001	0.08	-0.307	-0.047	-0.003	0.048	0.257
PERMIT	420	-0.001	0.058	-0.274	-0.029	-0.001	0.03	0.261
MSACSR	420	-0.001	0.082	-0.248	-0.049	0	0.048	0.405
A190RL1Q225SBEA	420	-0.072	0.954	-6.303	-0.292	-0.029	0.227	3.647
CBI	420	0.273	1.428	-3.803	-0.212	0.039	0.303	6.444
FSDP	420	0.004	0.003	-0.01	0.003	0.004	0.006	0.014
PRFI	420	0.003	0.016	-0.069	-0.002	0.006	0.012	0.058
M2SL	420	0.005	0.004	-0.005	0.003	0.005	0.007	0.028
M1SL	420	0.005	0.008	-0.034	0.0001	0.004	0.009	0.056
MICH	420	-0.003	0.148	-1.03	-0.046	0	0.057	1.504
PMSAVE	420	0.003	0.141	-0.933	-0.043	0.003	0.049	0.871
REQRESNS	420	0.003	0.032	-0.163	-0.013	0.005	0.021	0.169
CPILFESL	420	0.003	0.002	-0.002	0.002	0.002	0.003	0.014
CPIENGSL	420	0.002	0.025	-0.199	-0.006	0.002	0.011	0.126
BOGMBASE	420	0.008	0.023	-0.092	0.001	0.006	0.012	0.238
RMFSL	420	0.006	0.021	-0.083	-0.005	0.005	0.015	0.14
MABMM301USM189S	420	0.005	0.004	-0.005	0.003	0.005	0.007	0.028
MZMSL	420	0.006	0.008	-0.022	0.003	0.006	0.009	0.093
CPITRNSL	420	0.002	0.012	-0.108	-0.002	0.003	0.007	0.057

CPIHOSSL	420	0.003	0.003	-0.008	0.001	0.002	0.003	0.018
CPIAUCSL	420	0.003	0.003	-0.018	0.001	0.002	0.004	0.014
GPSAVE	420	0.004	0.012	-0.041	-0.002	0.004	0.01	0.055
GDPDEF	420	0.002	0.001	-0.001	0.001	0.002	0.003	0.009
GSAVE	420	0.004	0.012	-0.032	-0.004	0.005	0.012	0.04
B056RC1A027NBEA	420	0.006	0.009	-0.033	0.004	0.007	0.011	0.026
CNCF	420	0.005	0.014	-0.051	-0.003	0.006	0.014	0.058
CP	420	0.005	0.029	-0.237	-0.005	0.007	0.016	0.199
TWEXBMTH	420	0.003	0.013	-0.035	-0.006	0.003	0.01	0.064
TWEXMMTH	420	-0.0001	0.018	-0.054	-0.012	0.001	0.012	0.065
IMFSL	420	0.012	0.031	-0.06	-0.004	0.008	0.023	0.214
NETEXP	420	0.006	0.105	-0.698	-0.011	0.012	0.036	0.888
EXPGSC1	420	0.004	0.008	-0.031	0.001	0.005	0.009	0.023
IMPGSC96	420	0.004	0.009	-0.036	0.001	0.005	0.01	0.028
BAA10YM	420	0.003	0.084	-0.3	-0.041	0	0.039	0.47
MORTG	420	-0.003	0.033	-0.136	-0.023	-0.007	0.014	0.159
FEDFUNDS	420	-0.011	0.111	-0.911	-0.033	0	0.024	0.383
AAA	420	-0.003	0.031	-0.192	-0.021	-0.003	0.015	0.106
BAA	420	-0.003	0.027	-0.088	-0.018	-0.004	0.013	0.195
AAA10YM	420	0.005	0.203	-1.543	-0.063	0	0.057	1.211
MPRIME	420	-0.003	0.039	-0.272	0	0	0	0.237
TB3MS	420	-0.014	0.217	-1.846	-0.035	0	0.03	1.466
GS1	420	-0.01	0.096	-0.781	-0.049	-0.004	0.033	0.48
GS10	420	-0.004	0.054	-0.378	-0.034	-0.006	0.028	0.178

T10YFF	420	1.184	1.79	-9.57	0.16	1.47	2.476	4.04
SPREAD.GS	420	0.005	0.074	-0.532	-0.022	-0.001	0.026	0.403
SPREAD.MOODY.1	420	0.0004	0.016	-0.085	-0.006	-0.0004	0.006	0.104
SPREAD.MOODY.2	420	-0.002	0.152	-1.239	-0.037	-0.005	0.029	1.143
PPIACO	420	0.002	0.009	-0.055	-0.002	0.002	0.006	0.029
PPIIEG	420	0.002	0.029	-0.163	-0.011	0.002	0.016	0.096
PPIFGS	420	0.002	0.006	-0.031	-0.001	0.002	0.005	0.019
PPIIDC	420	0.002	0.01	-0.059	-0.002	0.002	0.006	0.032
PPICEM	420	0.002	0.073	-0.366	-0.026	0.003	0.033	0.275
PPIITM	420	0.002	0.008	-0.053	-0.001	0.002	0.005	0.032
ULCNFB	420	0.002	0.004	-0.013	-0.0002	0.002	0.004	0.016
COMPRNFB	420	0.001	0.003	-0.012	-0.001	0.001	0.003	0.014
COMPNFB	420	0.003	0.003	-0.011	0.002	0.003	0.005	0.014
OPHNFB	420	0.002	0.002	-0.006	0.0001	0.002	0.003	0.008
OPHPBS	420	0.002	0.002	-0.006	0.0002	0.002	0.003	0.009
ICSA	420	-0.001	0.064	-0.324	-0.036	-0.003	0.034	0.336
SP500	420	0.007	0.044	-0.245	-0.017	0.011	0.035	0.124
NASDAQ	420	0.007	0.066	-0.318	-0.026	0.017	0.045	0.199
RU2000	420	0.008	0.064	-0.367	-0.029	0.016	0.043	0.482
WILL5000	420	0.009	0.047	-0.259	-0.017	0.014	0.039	0.121

Figure 4-2 Correlation heat map for 114 macro factors

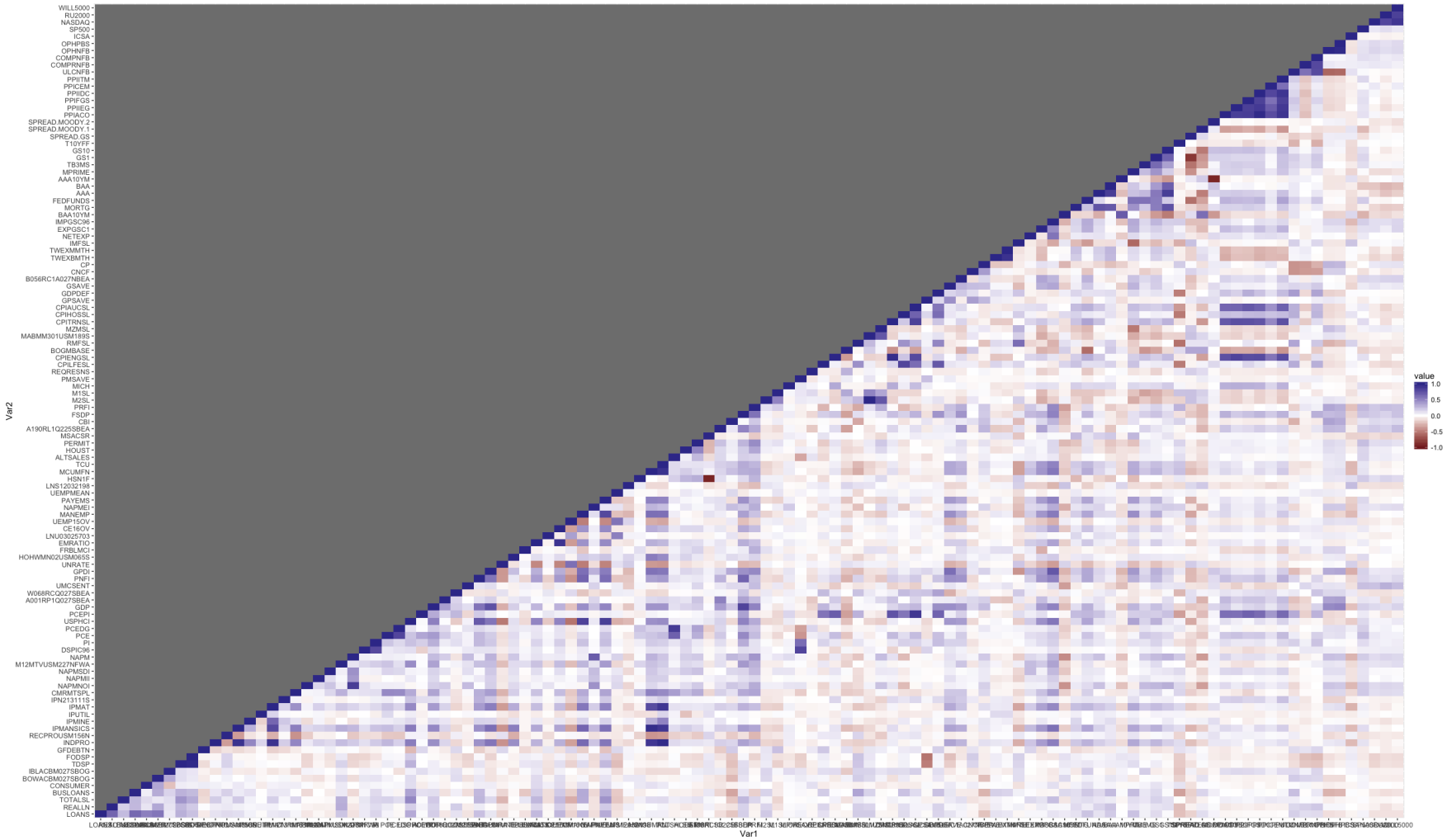


Table 4-2 Detail description of macro factors

Abbreviation	Full name/Data type	Frequency/Classification	Detail description
1. Macro and BC conditions			
1.1 Bank lending conditions			
LOANS	Loans and leases in bank credit, all commercial banks (Seasonally adjusted)	Monthly lagging indicator	This series covers the volume of loans and leases held by commercial banks. This factor is pro-cyclical with the business cycle. In boom the federal fund rate is lower which can encourage two main parities: one is the banking sector and banks have more capital to issue the loans; another is the investments by individuals and businesses. This enables banks to make loans at cheaper rates, which encourages investors to take fresh loans from the banks. During this period, some unqualified borrowers may get loans easily without strict examination. The potential default risk (long-term default risk) on these loans and leases is increasing. Once the economy turns to recession state, the interest rate will increase after the central bank tightens monetary policy. A large number of borrowers will then face a higher borrowing cost from the banks, especially borrowers with poor credit quality. However, default risk in the short term is decreasing since businesses and individuals can get liquidity from the banks. We can expect that this indicator has negative impact on default risk.
REALLN	Real estate loans, all commercial banks (Seasonally adjusted)	Monthly lagging indicator	This series measures the volume of real estate loans held by banks. This indicator has a tendency for increase in boom and a tendency for decrease in recession. In the expansion period, the interest rate and mortgage rate are lower, which booms the growth of investment in the real estate sector. Estate loans are increasingly issued to investors.

			Once the interest rates rise, mortgage rates also rise. This impacts the demand for housing and real estate loans. It also impacts refinancing activity. Investors may then face illiquidity problems thereby increasing the probabilities of default in the long-term perspective. However, in the short term, the default risk is going down. We expect that this factor should negatively correlate with default risk.
TOTALNS	Total consumer credit owned and securitized, outstanding (Seasonally adjusted)	Monthly lagging indicator	This series consists of the outstanding credit extended to individuals for household, family, and other personal expenditures, excluding loans secured by real estate and securitized loans. Likewise, this is a pro-cyclical indicator. We expect that this indicator is negatively associated with default risk.
BUSLOANS	Commercial and industrial loans, all commercial banks (Seasonally adjusted)	Monthly lagging indicator	This factor covers any type of loan issued to a business or corporation and not to an individual. This factor can be made in order to provide either working capital or to finance major capital expenditures. The condition of credit loan market is the sign of whether the future operation cost goes up or not, which results in the deteriation of corporate default crisis. In fact, if domestic credit markets are imperfect, firms may face credit constraints or higher financing costs as creditors raise interest rates to compensate for the increased risk of default. In addition, as the net worth of firms falls, so does the quality of the portfolios of their domestic creditors. Banks and financial intermediaries may be forced to reduce their exposure by cutting their loans—or they may simply go bankrupt (Serven and Solimano, 1992). This variable is pro-cyclical with business cycle. In the expansion period, banks tend to ease the lending standards and issue more loans. In the recession period, they tend to tighten the standards thereby reducing the number of loans issued to businesses or corporations. More loans issued by banks means that banks use looser policies, which may increase the number of loans that are issued to borrowers with poor credit quality.

			Consequently, the default risk in the long term is increasing. However, in the short term default risk is declining. It can then be concluded that this factor has a negatively impact on default risk.	
CONSUMER	Consumer loans at all commercial banks (Seasonally adjusted)	Monthly lagging indicator	This factor measures various loans issued to an individual. As for BUSLOANS mentioned above, this variable is pro-cyclical with business cycle. This factor negatively correlates with default risk.	
GFDEBTN	Federal debt: total public debt (Seasonally adjusted)	Quarterly lagging indicator	This series measures the amount of public debts (including treasury bills (T-bills), notes and bonds, as well as U.S. savings bonds) issued by the federal reserve. This is a counter-cyclical indicator for the business cycle. When this variable increases, the federal reserve tends to reduce the excess liquidity in the market by repurchasing the debt from the market. As a result, firms may face insufficient liquidity in the monetary market which increases the probabilities of defaults. Then we can expect that this variable positively impacts default risk.	
TDSP	Household debt/income-ratio (Seasonally adjusted)	Quarterly lagging indicator	This series measures the ratio between lenders' house debt value and their income. This is a counter-cyclical indicator.	A rise in both ratios implies lower income growth and higher debt or obligations. It is clear that this leads to default risk increasing. We can then expect that this variable positively associates with default risk.
FODSP	Household obligations/income	Monthly lagging indicator	This series measures the ratio between lenders' house obligations value and their income. This is a counter-cyclical indicator.	

	(Seasonally adjusted)			
IBLACBM027SBOG	Interbank loans, all commercial banks (Seasonally adjusted)	Monthly lagging indicator	This macro factor measures the amount of short-term loans made between banks. This indicator tends to increase in boom and decrease in bust. We can expect that this indicator has negative correlation with default risk.	
BOWACBM027SBOG	Borrowings, all commercial banks (Seasonally adjusted)	Monthly lagging indicator	This series measures the amount that depository institutions have borrowed from the discount window at the central bank. This indicator positively fluctuates with the business cycle. If there is an increase in borrowing, it means that banks have more owned capital available. This encourages banks to issue more loans to individuals or businesses. We expect that this variable negatively influences default risk.	
REQRESNS	Required Reserves of Depository Institutions (Not seasonally adjusted)	Monthly lagging indicator	This series measures the value of minimum balances that must be maintained by the depository institutions following the requirements from the central bank. This factor is pro-cyclical with the business cycle. In the expansion period, this factor tends to decrease and it tends to increase in the recession period. Increased variable means that the central bank tights monetary policy by suppressing bank lending to borrowers. Borrowers may then face financial distress since there is not sufficient liquidity. The default risk goes up. We can expect that this variable positively affects default risk.	

1.2 General market indicators

USPHCI	Coincident economic activity index (Seasonally adjusted)	Monthly coincident indicator	This index measures the current economic activities based on four indicators including nonfarm payroll employment, the unemployment rate, average hours worked in manufacturing and wages and salaries. If the index is increasing, the current economy tends to be expansion; otherwise, it tends to be in a recession period. It can then be concluded that this indicator is negatively associated with default risk.	
INDPRO	Industrial production Index (Seasonally adjusted)	Monthly coincident indicator	This is an economic indicator that measures the amount of real output from manufacturing, mining, and electric, and gas utilities industries.	All these factors are all pro-cyclical with business cycle. Then we expect that increasing variables drive the increase in corporate profits, which reduces the default risk.
IPN213111S	Industrial production: mining: drilling oil and gas wells (Seasonally adjusted)	Monthly coincident indicator	This series measures the amount of output from mining: drilling oil and gas wells.	
IPMANSICS	Industrial production-manufacturing (Seasonally adjusted)	Monthly coincident indicator	This series measures the amount of output from the manufacturing sector.	

IPMINE	Industrial production: mining (Seasonally adjusted)	Monthly coincident indicator	This series measures the amount of output from the mining industry.	
IPUTIL	Industrial production: electric and gas utilities (Seasonally adjusted)	Monthly coincident indicator	This series measures the amount of output from the electric and gas utilities industries.	
IPMAT	Industrial production: materials (Seasonally adjusted)	Monthly coincident indicator	This series measures the amount of output related to industrial materials.	
PI	Personal income (Seasonally adjusted)	Monthly coincident indicator	This series measures the income received by people from all sources such as participation in production, transfer payments from government, transfer payments from business, interest, dividends, and rental income. This indicator conveys slightly lagged information on the previous economic conditions; however, economists still regard it as a tool for reflecting the current state of the economy. Basically, if this value is rising, the personal income is increasing which means that the firms increase the income for their staff and people who invest in the equity market might have high earnings. This indicator is then pro-cyclical with the	

			economic conditions. In contrast, higher personal income shows that the firm default risk is quite lower, which has a negative impact of default risk.
DSPIC96	Real disposable personal income (Seasonally adjusted)	Monthly coincident indicator	This indicator is different from the personal income since this is to exclude the elements of personal current taxes and allow for inflation, which can show the amount of spending or saving that people can use. This dollar series also exhibits the current stage of economic conditions. This value has a positive relationship with business cycle. Lower disposable income might reflect that firms cannot offer a high salary for their employees since the corporate profits are lower with high production and operation cost in the depressed economic environment. Therefore, we can expect that this indicator negatively impacts on default risk.
PCE	Personal consumption expenditure (Seasonally adjusted)	Monthly coincident indicator	Personal consumption expenditures are used to measure the price changes in consumer goods and services, which contain the actual and imputed expenditures of households; the measure includes data pertaining to durables, non-durables and services. It is essentially a measure of goods and services targeted toward individuals and consumed by individuals. Personal consumption is the largest component of GDP. Then consumption behaviour is crucial to the movements of the business cycle. For example, the increasing indicator means that consumers tend to spend more on purchasing goods and services, which suggests that there is a high demand in the market. This is a good sign for suppliers since they can make more profit in this scenario. In this context, firms might have good financial conditions since net profits are increasing. If supply is smaller than demand, goods' and services' prices are going up. Then the firms can get more profits thereby decline the default risk.
PCEPI	Personal consumption expenditures: chain-type price index	Monthly coincident indicator	This indicator as PCE is also part of a report (actually a part of the personal income report) published by the Bureau of Economic Analysis of the Department of Commerce. This index is based on a chain index, which takes the changes of consumers' consumption behaviour due to prices into consideration. Compared with the PCE, this indicator can further clarify consumers' behaviour such as the substitution of goods will be sold

	(Seasonally adjusted)		well when prices rise. Then this indicator is pro-cyclical with business conditions. For the influence on default risk, it can be expected that the overall default risk might decline since there is an increase in this index.
W068RCQ027SBEA	Government expenditure (Seasonally adjusted)	Quarterly coincident indicator	This indicator is also a key element of calculating the government spending in GDP, which contains net purchases by federal, state, local governments of goods and services, payments by general government to households in the form of compensation of employees and the consumption of general government fixed capital (including the value of current services of fixed assets of general government). This indicator tends to increase in the recession period and decrease in the expansion period. The government actively invests in the market which might boom the market in order to increase demand (number of orders) for firms. If this is the case, it means that firms might be in financial distress. Therefore, it can be seen that the default risk might be positively associated with government expenditures.
GDP	Gross domestic product (Seasonally adjusted)	Quarterly coincident indicator	GDP is the most widely used indicator for measuring the nation's product, which counts the market value of the goods and services produced by the labour and property located in the U.S. This indicator is regarded as a signal of the current state of economy. It is clear that this variable is positively correlated with the business cycle and negatively associated with the default risk.
GPDI	Gross private domestic investment (Seasonally adjusted)	Quarterly coincident indicator	This dollar series covers non-residential fixed investment, residential fixed investment, and change in private inventories. This macro factor can give a signal of the current business condition in the private sector. If there is an increase for this indicator, the current economy is growing. Then this factor positively fluctuates with business conditions. It can be concluded that this factor should have negative impact factor on default risk.

PNFI	Private non-residential fixed Investment (Seasonally adjusted)	Quarterly coincident indicator	This series is part of GPDI, which measures the value of private domestic investment in non-residential structures and producers' durable equipment that are purchased by private business and non-profit institutions. The decrease of this factor means that the value of private investment depreciates, suggesting that the current economy is in a downturn period. Then this factor should have negative impact on default risk.
CBI	Change in private inventories (Seasonally adjusted)	Quarterly Leading indicator	This series is a key input of calculation of GDP and it is the most volatile component. This factor measures the value of the change in the physical volume of goods purchased by businesses for use in the production of other commodities, or for resale. The movement of this factor covers the future information about the timing, duration, and magnitude of business cycles since high value indicates that the future production should be reduced and low value means that the shortages in inventories should be handled by increasing production. Basically, in economic expansion periods, this factor increases; otherwise it decreases. Then the higher inventories might negatively affect default risk.
PRFI	Private residential fixed investment (Seasonally adjusted)	Quarterly coincident indicator	This macro factor measures purchases of housekeeping and non-housekeeping residential buildings by businesses and individuals and improvements to these buildings. Higher value of this factor indicates that there is a high demand of housekeeping and non-housekeeping residential buildings and the firms that provide these products might have higher profit. Then this indicator positively correlates with business conditions and it should negatively affect default risk.
A001RP1Q027SBEA	Gross national product (Seasonally adjusted)	Quarterly coincident indicator	GNP is the market value of the goods and services produced by labour and the property supplied by U.S. residents or in a foreign country. The main difference between GDP and GNP is that GDP does not consider the goods and services produced abroad by using labour and property supplied by the U.S. but GNP does consider it. The higher the factor, the more vigorous the economy. Then this factor should be pro-cyclical with business conditions and it should be negatively associated with default risk.

CMRMTSPL	Real manufacturing and trade industries sales (Seasonally adjusted)	Monthly coincident indicator	This series counts the monthly sales volume of manufacturing, merchant wholesalers, and retail establishments. This indicator has the same trend with business cycle. When there is strong economic growth, the demand is high which leads to firms increasing productivity. The velocity of trading activities increases in this context. Firms can get a high income from increasingly selling their products to the market. Then we can expect that this indicator should negatively influence default risk.	
RECPROUSM156N	Smoothed recession probabilities (Not seasonally adjusted)	Monthly leading indicator	This series measures the recession probabilities. Higher value in this indicator means that the economy tends to be in a recession; otherwise, it means low probability of recession. Then it is clear that this factor should be positively related with default risk.	
UMCSENT	University of Michigan consumer sentiment (Not seasonally adjusted)	Monthly leading indicator	This series measures public opinion about personal finances, business conditions, and buying conditions. Since this indicator is based on the consumer's expectation, then it has a predicting power for future economic conditions. We can expect that this factor should negatively impact default risk.	
A190RL1Q225SBEA	Real final sales of domestic product (Seasonally adjusted)	Quarterly coincident indicator	This indicator is real GDP less change in private inventories.	Both variables increase and fall with business cycle. We expect that they negatively associate with default risk.

FSDP	Final sales to domestic purchasers (Seasonally adjusted)	Quarterly coincident indicator	This is equal to GDP minus net exports and inventory investment, which measures demand for goods and services from U.S. households, businesses and government, regardless of whether those goods and services are imported or domestically produced.	
PCEDG	Expenditure durable goods (Seasonally adjusted)	Monthly leading indicator	This series measures the amount of personal consumption expenditure on buying durable goods. This indicator tends to rise in bull market and fall in bear market. We expect that this indicator negatively correlates with default risk.	
HSN1F	New one family houses sold (Seasonally adjusted)	Monthly leading indicator	This series measures the number of new houses sold. This is a pro-cyclical indicator. We expect that this indicator negatively impacts on default risk in the short term.	
MCUMFN	Capacity utilization: manufacturing (NAICS) (Seasonally adjusted)	Monthly leading indicator	This series measures how much the nation's capital is being used in production in the manufacturing sector.	Both variables rise and fall with business cycles. They are negatively associated with default risk.

TCU	Capacity utilization: total industry (Seasonally adjusted)	Monthly leading indicator	This variable measures the extent to which the nation's capital is being used in the production of goods in total industry.	
M12MTVUSM227NFWA	Moving 12-month total vehicle miles travelled (Not seasonally adjusted)	Monthly leading indicator	This series measures the number of miles travelled. This indicator tends to increase in boom and tends to decrease in bust. We can expect that this variable negatively affects default risk.	
ALTSALES	Light weight vehicle sales: autos & light trucks (Seasonally adjusted)	Monthly leading indicator	This series measures the number of lightweight vehicle sold. This is a pro-cyclical indicator. We expect that this variable negatively correlates with default risk.	
HOUST	Housing starts (Seasonally adjusted)	Monthly leading indicator	This series measures the number of new residential construction projects that have begun during any particular month, which is considered to be a leading indicator of economic strength. This is a pro-cyclical indicator. In a strong economy, people tend to purchase houses; in contrast, in a weak economy, they are not likely to buy new properties. Then we can expect that this indicator negatively correlates with default risk in the short term.	

PERMIT	Building permits (Seasonally adjusted)	Monthly leading indicator	This series measures the number of new residential construction projects granted by a government or other regulatory body, which is also a pro-cyclical indicator. The value of this variable reflects the attitude of the government towards the market. We can expect that this variable is negatively related to default risk.
NAPMNOI	ISM manufacturing: new orders index (Seasonally adjusted)	Monthly leading indicator	This macro factor is created by National Association of Purchasing Management (NAPM), now called the ISM, which is based on a survey of over 250 companies within twenty-one industries covering all 50 states. Its importance derives from the fact that it is the first indicator each month of manufacturing performance during the past 30-day period and it represents attitudes of business toward future profitability. This survey investigates six areas: production, orders, commodity prices, inventories, employment, and vendor performance. This study covers five indices for measuring economic activities. This one is used to measure the number of new orders over 250 companies. Increased variable indicates that there is a high expectation of new orders in the future, which means that the firms' profits might increase. Then we expect that the default risk should be negatively correlated with this factor.
NAPMII	ISM manufacturing: inventories index (Not seasonally adjusted)	Monthly leading indicator	This series measures the expectation of number of inventories. An extreme increase in inventories may signal that aggregate demand is slowing down and that firms will soon cut back production and output responding to such a fall in the demand for goods and services. In this context, firms' profits may go down which means that default risk is going up. However, in the short term, this increased variable implies that there is a high demand in the bull market and firms have more profits during this period. Therefore, we expect that this variable has negative impact on the default risk in the short period.
NAPMSDI	ISM manufacturing: supplier deliveries Index	Monthly leading indicator	This series is also a part of NAPM, which is to measure the amount of supplier deliveries. Higher supplier deliveries suggests that those firms have a higher future demand from the market and the corporate profits will increase. Then firm' default risk is going down.

	(Seasonally adjusted)		
NAPM	ISM manufacturing: PMI Composite Index (Seasonally adjusted)	Monthly leading indicator	This index is to measure whether the manufacturing economy is expanding or not based on the benchmark of 50 percent. The variable above 50 percent indicates that there is an expansion in this manufacturing sector; otherwise, there is a decline. It can be expected that this variable is negatively associated with default risk.
MSACSR	Monthly supply of houses (Seasonally adjusted)	Monthly leading indicator	The factor is the ratio of houses for sale to houses sold, which is counter-cyclical with business cycle. This factor can provide a predictive sign of the future market. If this value is increasing, there is more house supply than houses sold. Consequently, this can reduce the corporate profits for the firms investing in real estate sectors and profits in the construction sector will be influenced since the house supply will reduce responding to a relative reduction in the demand of houses from the market. Then we can expect that this factor positively impact default risk.

1.3 Labour market indicators

The following variables are used to measure labour market conditions, which can also reflect labour uneampoyment risk. Basically, Agrawal and Matsa (2013) study the impact of labor unemployment risk on corporate financing in U.S. and they find that there is a strong relation between unemployment risk and financial policy among firms facing tight financial constraints. They state that reductions in labor unemployment risk are associated with increases in corporate leverage and reducing leverage decreases the probability that a firm will encounter financial distress. Meanwhile, Jaggia and Thakor (1994) also find that firms can reduce the probability

of distress by taking less risky projects, or reduce workers' losses in distress by redesigning job tasks to require fewer firm-specific skills. This is why I emply these varibales for investigation of corporate defaults.			
ICSA	Initial claims (Seasonally adjusted)	Weekly leading indicator	This monthly indicator is for measuring unemployment insurance by recording the number of people who seek to receive state jobless benefits per week under an unemployment insurance programme in the 50 states and the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. Higher initial claims suggest that there is a weakening economy; otherwise, the economy is strong. Then it is a counter-cyclical indicator which can give a predicting sign for the economic conditions. In practice, the Conference Board's composite indices use this indicator for measuring economic conditions. Then it can be seen that there is a negative correlation with default risk. Specifically, more people claiming unemployment insurance means that a company tends to reduce the number of employees thereby reducing the operation cost. Then this company has a higher default risk now. This is weekly data and it is calculated based on the sum up weekly data in each month.
HOHWMN02USM065S	Weekly hours worked: Manufacturing (Seasonally adjusted)	Weekly leading indicator	This is used to measure the average number of hours paid per worker working only in the manufacturing sector (such as production or factory-type work). This hours-worked series is recorded per week during the survey week. This indicator can provide insight into the direction of the economy. For example, when this indicator goes up, it is indicated that the firm starts expanding the volume of current productions in order to satisfy the high future and present demand from the market. Then this manufacturing firm is able to earn more profits which avoids the default risk to a large extent. Therefore, this is a pro-cyclical indicator with the economic environment and a counter-cyclical indicator for default risk. Likewise, the composite indices constructed by the Conference Board also employ this data for measuring economic conditions.
LNS12032198	Employment level: part-time for economic reasons, slack	Monthly leading indicator	This is also a leading indicator since it can give a sign of changes of labour market conditions. This indicator shows the number of people who worked 1 to 34 hours during the reference week for an economic reason such as slack work or unfavourable business conditions, or seasonal declines in demand which is provided by the U.S. Bureau of Labour Statistics. It can be seen that if there is a favourable economic condition, people

	work or business conditions, nonagricultural industries (Seasonally adjusted)		tend to work rather than slacking work. Then this indicator is a pro-cyclical indicator with economic environment. Favourable economic condition means that there is an active market with a lot of demand. Therefore, the firms have high opportunity to earn more money by expanding production through employing more people working for them. The default risk is clearly small, which means that this indicator is counter-cyclical with default risk. Stock and Watson (1989) initially employ this indicator for constructing an experimental leading index for measuring business conditions.
NAPMEI	ISM manufacturing: employment index (Seasonally adjusted)	Monthly leading indicator	This series is a part of NAPM that measures the expectations of employment. Higher value suggests that the firms have a higher future expectation on the economy and they anticipate that there will be a vigorous economy. Then more employees should be employed in order to fulfil the increasing demands of production. Firms may make more profits after the selling increases. Then we can expect that default risk has negative correlation with this indicator.
UNRATE	Civilian unemployment rate (Seasonally adjusted)	Monthly coincident indicator	This coincident indicator is the ratio of the number of unemployed as a percentage of the civilian labour force. In particular, labour force data are restricted to the civilian noninstitutional population 16 years of age and older. This series is negatively correlated with the change of economic conditions. The higher the indicator is, the worse weakening economy we face. Then this is counter-cyclical with the economy. It can be seen that the higher ratio means higher default risk among the firms, suggesting that it is a pro-cyclical indicator with default risk. Gilchrist and Zakrajšek (2011) construct a new indicator based on corporate bond credit spreads as a robust predictor of future economic activity and in their study they use this ratio to measure the state of economic conditions.
FRBLMCI	Labour market conditions index (LMCI)	Monthly coincident indicator	This indicator was introduced in 2014, which was used to construct a single indicator for measuring overall labour market conditions by employing a dynamic factor model concerning 19 labour market indicators covering the broad categories of unemployment and underemployment, employment, workweeks, wages, vacancies, hiring, layoffs, quits, and surveys of consumers and businesses. The LMCI is able to mirror the

	(Seasonally adjusted)		complicated nature of the current labour market. If this index increases, it means that the economic conditions tend to be more active and default risk tends to go down; otherwise, the economic activity decreases and default risk rises. Then this is a pro-cyclical indicator with business cycle and counter-cyclical with default risk.
EMRATIO	Civilian employment-population ratio (Seasonally adjusted)	Monthly coincident indicator	The employment-population ratio represents the proportion of the civilian noninstitutional population that is employed. This series positively correlates with broad movements the aggregate economic activity. In a bull market, this indicator is high meaning that firms need more employees to produce more products for satisfying big demand from the market. Then default risk for each firm is small. Therefore, this indicator is pro-cyclical with the business cycle and counter-cyclical with the default cycle.
CE16OV	Civilian employment (Seasonally adjusted)	Monthly coincident indicator	This indicator is actually the numerator in the EMRATIO. It is clear that this value has same trend with market conditions and adverse trend with default risk.
PAYEMS	All employees: total nonfarm payrolls (Seasonally adjusted)	Monthly coincident indicator	This series shows the number of paid U.S. workers in any business which excludes proprietors, private household employees, unpaid volunteers, farm employees, and the unincorporated self-employed. This measure accounts for approximately 80 percent of the workers producing the entire GDP. This is why this indicator can provide useful information from the current state of the economy. If this data increases, it might indicate that businesses are hiring more workers which might also mean that businesses are rising. This case reflects the expansion of the economy. The default risk for each firm decreases in this context. Therefore, this is a pro-cyclical indicator for the business cycle and a counter-cyclical indicator for the default cycle. This series is used by Conference Board for calculating recession probabilities.

MANEMP	All employees: manufacturing (Seasonally adjusted)	Monthly coincident indicator	Compared with the PAYEMS mentioned above, this indicator only counts the number of paid U.S. workers in the manufacturing sector. Increases in employment in this sector suggest that the manufacture business might boom with increase of a large amount of demand. These types of firms are able to earn more in order to have healthier financial conditions which reduce the probabilities of defaults. Thus this indicator has the same direction as economic fluctuations and it has an inverse trend with the default risk.
UEMPMEAN	Average (mean) duration of unemployment (Seasonally adjusted)	Monthly lagging indicator	This indicator is to measure the average number of weeks, including the survey's reference week, during which people recorded as unemployed had been looking for work or, in the case of people on layoff, the number of weeks since their layoff. As the duration increases, people should spend more time finding a job since there are quite a few job opportunities during economic downturn. Early studies have identified that weak labour demand plays a key role in prolonged unemployment duration (Valletta and Kuang, 2012). Then this indicator is a lagging indicator for measuring the state of the economy, which also has negative correlation with business cycle. Then we can see that this indicator might have a positive impact on the default risk since firms in poor financial condition cannot offer sufficient jobs in the job market.
LNU03025703	Of total unemployed, percent unemployed 27 weeks and over (Seasonally adjusted)	Monthly lagging indicator	This series is to measure the ratio of the number of people who have been unemployed for 27 weeks or more expressed as a percent of the civilian labour force. This indicator is not able to reflect changing conditions in the labour market early but shows the changes in the lagging period. If there is an economic recession happening, then this indicator starts increasing. Even though the recession ends, this percent series might still increase since in the recovery period the market condition is still poor with less economic activities. This series is negatively related to business cycle. Higher value in this indicator shows that there are not so many work opportunities in the market. Then firms have higher default risk since they might face financial illiquidity in this period and less demand from the market.

UEMP15OV	Number of civilians unemployed for 15 weeks & over (Seasonally adjusted)	Monthly lagging indicator	This is the number of people who have been unemployed for 15 weeks. Compared with the indicator mentioned above, this value has similar trend. However, this one only considers the number of unemployed people with short unemployment duration. We also expect that this factor positively impacts default risk.

1.4 Monetary policy indicators

GSAVE	Gross saving (Seasonally adjusted)	Quarterly coincident indicator	This series measures gross private savings including personal savings, gross business savings and gross government savings.	Three types of saving rates are able to reflect the current state of the economy. They have same trend as the business cycle. That is, they become higher in boom and lower in recession. Thus the default risk should be negatively associated with three saving rates.
GPSAVE	Gross private saving (Seasonally adjusted)	Quarterly coincident indicator	This series covers two types of savings: one is personal savings; another is the sum of undistributed corporate profits with inventory valuation and capital consumption adjustments, and the assumption of fixed capital by business.	
PSAVERT	Personal saving (Seasonally adjusted)	Monthly coincident indicator	This series measures personal income less personal outlays and personal tax and nontax payments.	

GDPDEF	GDP deflator, implicit (Seasonally adjusted)	Quarterly coincident indicator	This series is used to measure the price level in the current economy, which accounts for inflation by converting output measured at current prices into constant-dollar GDP. This factor presents how much a change in the base year's GDP relies upon changes in the price level. This indicator can show how the change of consumption patterns or the introduction of new goods and services have influence on the overall economy. Barro (1989) identifies that GDP deflator is systematically associated with the level of economic development. This variable tends to go up in boom and fall in recession periods. Then we expect that this factor has negative effect on the default risk.	
MICH	University of Michigan inflation expectation (Not seasonally adjusted)	Monthly leading indicator	This factor measures future inflation provided by the University of Michigan's 'Survey of Consumers'. This value can provide predictive content of future inflation since this survey is based on the inflation expectation of U.S. residents.	Sarel (1996) finds that there is a non-linear relationship between inflation and economic growth. However, in most historical periods, the inflation rates are pro-cyclical and lag the business cycle except MICH (it tends to build up during an expansion and fall after the cyclical peak). For the default risk, early studies find that deflation can lead to high default risk (Fisher, 1933, Bhamra et al., 2011); however, recent studies determine that higher inflation can also lead to higher default risk among the firms (Wadhwani, 1986). However, it should be noted that in the short term, higher inflation reflects economic
CPILFESL	Consumer price index for all urban consumers: all Items less food & energy (Seasonally adjusted)	Monthly lagging indicator	This series contains all items in CPIUCAL except for food and energy components.	
CPIENGSL	Consumer price index for all urban	Monthly lagging indicator	This series only includes energy items in CPIUCAL.	

	consumers: energy (Seasonally adjusted)			expansion. Therefore, the ultimate impact on default risk is negative.
CPITRNSL	Consumer price index for all urban consumers: transportation (Seasonally adjusted)	Monthly lagging indicator	This series contains all items in transportation in CPIUCAL.	
CPIHOSSL	Consumer price index for all urban consumers: housing (Seasonally adjusted)	Monthly lagging indicator	This series contains all items in CPIUCAL which are only related to housing sectors.	
CPIAUCSL	Consumer price index for all urban consumers: all Items	Monthly lagging indicator	This series is the broadest among these CPI, which represents all buying habits of all urban consumers (this is around 80 percent of the non-institutional population).	

	(Seasonally adjusted)			
BOGMBASE	Monetary base (Not seasonally adjusted)	Monthly leading indicator	This macro factor measures total balances maintained and currency in circulation, which is the most liquid currencies.	Basically, the growth of money supply injects more liquidity into the market and business and investors and consumers find it easy to get credit from banks, which is able to boom the economy. This factor is pro-cyclical with the business cycle. In this context, firms can easily obtain loans from banks and financial distress can be avoided. Then this factor should have negative impact on default risk.
M1SL	M1 money stock (Seasonally adjusted)	Monthly leading indicator	The series consists of almost all of the currency portion of the monetary base such as travelelrs' cheques of non-bank issuers, and demand deposits at all commercial banks (rather than in the central bank).	
M2SL	M2 money stock (Seasonally adjusted)	Monthly leading indicator	M2 includes more monetary sources in comparison with M1. They are savings deposits, time deposits in amounts of less than $100,000 and retail money market mutual funds.	
MABMM301USM189S	M3 money stock (Seasonally adjusted)	Monthly leading indicator	M3 is the third measure of money supply which covers M1, M2 and the other 4 sources including time deposits of $100,000 or more at all depositary institutions, balances in money market funds, term repurchase agreements and Eurodollars.	
MZMSL	MZM money stock (Seasonally adjusted)	Monthly leading indicator	This series is money with zero maturity, which contains M2 and institution money fund excluding total small denomination time deposits. This indicator is a preferable money supply measure	

			since it better shows money readily available within the economy for consumption.	

1.5 Liquidity from non-banks sectors

RMFSL	Retail money funds (Seasonally adjusted)	Monthly leading indicator	This series measures the volume of the fund that is registered with the Securities and Exchange Commission (SEC) and is sold to individual investors through investment dealers and in open market transactions. Retail funds are often categorized as mutual funds, and carry lower initial investments and management expense ratios than non-retail funds.	Both variables tend to increase in bull and decrease in bust. This is a pro-cyclical indicator. It should be noted that the market funds play a key role in the monetary market since they inject a large amount of liquidity into the market, which is a kind of channel of money supply. Then we regard two factors as the leading indicator. We expect that they negatively correlate with default risk.
IMFSL	Institutional money funds (Seasonally adjusted)	Monthly leading indicator	This series measures the volume of funds that target high value investors with low management fees, but very high minimum investing requirements, for example, pension or endowment funds (managed by large institutional investors).	

1.6 Firm profitability

CP	Corporate profits after tax (Seasonally adjusted)	Quarterly lagging indicator	This series counts profits before tax less the sum of federal, state, and local income taxes, which includes dividends and undistributed corporate profits.	Basically, three corporate profits indicators are very pro-cyclical as they tend to increase during booms and strongly fall during recessions. It is very clear that high defaults occur when the firms face financial distress. In other words, an increase of these indicators means higher profits, indicating that there is a quite low default risk.
B056RC1A027NBEA	Net corporate dividends (Not seasonally adjusted)	Annually lagging indicator	This factor includes payments in cash or other assets, excluding the corporation's own stock, made by corporations located in the U.S. and abroad to stockholders who are U.S. residents. The payments are netted against dividends received by U.S. firms, thereby providing a measure of the dividends paid by U.S. firms to other sectors.	
CNCF	Corporate net cash flow (Seasonally adjusted)	Quarterly lagging indicator	This indicator covers undistributed corporate profits, deflated by the implicit price deflator for non-residential fixed investment, and corporate capital consumption allowances, deflated by the chain-weighted price index for range of capital stock items.	

1.7 Balance of payments

NETEXP	Net exports of goods and services	Quarterly leading indicator	The factor measures the difference between exports and imports. Net exports negatively fluctuate with the business cycle since the trade balance tends to improve in recessions and worsen in periods of high growth. The changes of this variable have two side effects on the firms including exporters and importers, which can increase profits for either exporters or importers. Therefore, the ultimate influence on default risk is not clear.

	(Seasonally adjusted)		
EXPGSC1	Real exports goods, services (Seasonally adjusted)	Quarterly leading indicator	This series measures goods and services sold outside the United States by U.S. businesses or individuals. Basically, this is counter-cyclical with the business cycle. In a recession period, the exchange rate depreciates, which can increase the competitiveness of products for U.S. exporters with lower price. In this context, the default risk in the export sector is quite small. In contrast, this is not the same case for the importers since they face a high importing cost. Therefore, it is unclear whether this factor positively influences default risk or not.
IMPGSC96	Real imports goods & services (Seasonally adjusted)	Quarterly leading indicator	This series measures goods and services bought outside the United States by U.S. businesses or individuals. Compared with EXPGSC1, this is pro-cyclical with the business cycle. In boom, the interest rate is normally high which can increase the U.S. exchange rate. Importers can get the benefits from the increasing value of U.S. dollars and the purchasing power is rising; however, there is a bad sign for exporters since the selling price outside of the U.S. is going up. Therefore, it is unclear whether this factor positively influences default risk or not.

1.8 Terms of trade

TWEXBMTH	Trade weighted U.S. dollar index: broad	Monthly leading indicator	This series measures the value of the dollar against the currencies of a broad group of major U.S. trading partners.	Basically, in favourable economic conditions, the US dollar tends to be strong against other currencies. An increasing number of investors want to invest in the U.S. market in order to get highly valuable assets. Meanwhile, the majority of resources such as natural resources and commodities are priced in dollars. If the dollar increases, it can boom the internal market. Also the appreciation

	(Not seasonally adjusted)			of the dollar can attract more capital from outside in order to boom the U.S. market and the import sector can increase since the purchasing cost is lower. It is uncertain to judge whether an increased U.S. dollar positively influences default risk or not. Although the appreciation can attract more capital and more investors to enter into the US market and importers can get more profits from reducing import costs, this is not factorable for exporters. For exporters, there are high exporting costs and they may face decreasing profits thereby increasing the default risk. Therefore, the ultimate impact on default risk is uncertain.
TWEXMMTH	Trade weighted U.S. dollar index: major currencies (Not seasonally adjusted)	Monthly leading indicator	This series measures the value of the dollar against seven major currencies.	

2. Micro-level conditions

2.1 Labour cost/wages

ULCNFB	Unit labour cost: nonfarm business	Quarterly coincident indicator	This business indicator measures the labour compensation cost in the nonfarm business sector, which is required to produce one unit of output. There are two main elements for constructing this indicator. Specifically, one is to measure compensation per hour; another is to calculate the output per hour. This increased factor can affect the trend of inflation in the future since the labour union might argue against the

	(Seasonally adjusted)		firm owner for increasing the income responding to the increase of unit labour cost. This variable basically becomes higher in boom and falls in recession. It should be noted that an increasing unit labour cost can cause a higher financial problem to firms thereby making firms tend to default.	
COMPRNFB	Nonfarm business sector: real compensation per hour (Seasonally adjusted)	Quarterly coincident indicator	This series measures the changes in the spending power of average hourly compensation of employees in the nonfarm business sector.	The only difference between the two indices is whether I use the output from real gross domestic product or not. The movements of both indices correspond to the movement of the business cycle. Increased compensation cost may lead to high production cost, which may cause firms to default. Thus we expect that they positively affect default risk.
COMPNFB	Nonfarm business sector: compensation per hour (Seasonally adjusted)	Quarterly coincident indicator	This series measures the changes in average hourly compensation of employees in the nonfarm business sector.	
OPHPBS	Business sector: real output per hour of all persons (Seasonally adjusted)	Quarterly coincident indicator	This index reflects the changes in labour productivity in the business sector.	Both indices are pro-cyclical with business conditions. In boom, they tend to increase since there is high demand from the market and firms need to expand their productivity to fulfil this

OPHNFB	Nonfarm business sector: real output per hour of all persons (Seasonally adjusted)	Quarterly coincident indicator	This index reflects the changes in labour productivity in the nonfarm business sector.	increasing demand. Therefore, firms can earn more in this period which may lead to income rising thereby reducing probabilities of default. We expect that both indices are negatively related to default risk.

2.2 Cost of capital

FEDFUNDS	Effective federal funds rate (Not seasonally adjusted)	Monthly lagging indicator	This series is the interest rate banks charge other banks for very short-term loans, often just overnight. Since banks must close each day with a minimum required reserve related to the amount of money loaned out, a bank with surplus funds may lend the extra to a bank that comes up short so both banks can meet their capital quotas for the day. This rate is the central rate, which can influence the following rates such as mortgage rate, discount rate, and prime rate. This monetary policy is normally used after recession happens. Then there is a strong lagging effect compared with the business cycle. Therefore, the following interest rates are also the lagging indicators.	Higher interest rate implies higher cost of financing or borrowing. This scenario normally occurs in the expansion period since the policy makers need to cool down the hot market. It is clear that increased interest rates positively influence default risk.
MORTG	30 year mortgage rate	Monthly lagging indicator	This interest rate is for contract interest rates on commitments for fixed-rate first mortgages.	

	(Not seasonally adjusted)		
AAA	AAA corporate bond yield (Not seasonally adjusted)	Monthly lagging indicator	This interest rate measures the yield of Moody's AAA corporate bonds.
BAA	BAA corporate bond yield (Not seasonally adjusted)	Monthly lagging indicator	This interest rate measures the yield of Moody's BAA corporate bonds.
AAA10YM	Treasury bond yield, 10 years (Aaa) (Not seasonally adjusted)	Monthly lagging indicator	This interest rate measures the yield of 10-year Moody's Aaa corporate bonds.
MPRIME	Bank prime loan rate (Not seasonally adjusted)	Monthly lagging indicator	This interest rate measures the prime rate used by banks to price short-term business loans.

GS1	Year treasury constant maturity Rate (Not seasonally adjusted)	Monthly lagging indicator	This interest rate measures the yield of 1-year treasury bonds.
GS10	10-year treasury constant maturity Rate (Not seasonally adjusted)	Monthly lagging indicator	This interest rate measures the yield of 10-year treasury bonds.
BAA10YM	Treasury bond yield, 10 years (Baa) (Not seasonally adjusted)	Monthly lagging indicator	This interest rate measures the yield of 10-year Moody's Baa corporate bonds.
TB3MS	3-month treasury bill (Not seasonally adjusted)	Monthly lagging indicator	This interest rate measures the yield of 3-month treasury bills.

T10YFF	10-year treasury constant maturity Minus Federal Funds Rate (Not seasonally adjusted)	Monthly leading indicator	This series reflect the default risk premium by calculating the spread between 10-year treasury constant maturity and federal fund rate.	All these yield spreads based on the difference between different long term yield ratios and short term yield ratios or between more risky bonds and less risky bonds in the same maturity have the same function, which can imply the future trend in the growth of the money supply. Basically, these spread ratio are also regarded as the yield curve. If the yield curve steepens, the long term bonds have a higher yield than the shorter term bonds since the risk positively correlates with time. If the yield curve becomes inverted, the short term bonds' yield is higher than the long term bonds' yield, meaning that there might be a sign of forthcoming recession. For the flat curve, it means that the economy is in a transition stage. Actually, the yield curve is initially used for measuring the default risk in the bonds market. These yield spreads have been widely used for measuring default
SPREAD. GS	Interest spead: Difference between 10-year Treasury constant maturity rate and 1-year Treasury constant maturity rate (Not seasonally adjusted)	Monthly leading indicator	This business factor measures the spread between 10-year Treasury constant maturity rate and 1-year Treasury constant maturity rate.	
SPREAD. MOODY.1	Interest spread: Difference between 1-year BAA yield and 1-year AAA yield	Monthly leading indicator	This business factor measures the spread between difference between 1-year BAA yield and 1-year AAA yield.	

	(Not seasonally adjusted)			risk in the macro and micro level research (Azizpour et al., 2014, Giesecke et al., 2011). Increased spread ratio indicates that there is a higher default risk. Therefore, this value should be positively associated with default risk.
SPREAD. MOODY.2	Interest spread: Difference between 10-year BAA yield 10-year and AAA yield (Not seasonally adjusted)	Monthly leading indicator	This business factor measures the spread between 10-year BAA yield 10-year and AAA yield.	

2.3 Cost of resources

PPIACO	PPI all commodities (Not seasonally adjusted)	Monthly lagging indicator	This index is a weighted index of prices calculated at the wholesale, or producer level, which is developed by the Bureau of Labor Statistics (BLS). This index monitors prices paid by urban consumers for a generally fixed market basket of goods and services. Based on different portions in the market, this index can be changed to different sub-indices, which are shown below. All of the physical goods-producing industries that make up the U.S. economy are included, but imports are not. Compared with CPI, this variable only focuses on the prices received by producers rather than consumers. In addition, CPI are highly influenced by	These variables are the sign of whether there is an inflation in the economic environment. Wadhwani (1986) intially investiages how a high level of inflation has adverse effects on the real economy by using a theoretical framework and he explores some of the possibilities, for example, higher inflation implies higher liquidation rates and default premia and it can also depress share values.

			both service sector inflation and an input of the housing component. This is not the case for PPI.	Consequently, he figures out that why higher inflation lead to more bankruptcies by using theoretical framework. Thus I use these variables for investigating corporate default. These PPI indices have a tendency to increase in economic expansion and decrease in economic depression. Then this is a pro-cyclical indicator. The relation among them is quite sensitive. In boom, an increase demand occurs in the market, which raises the demand of crude energy materials. Consequently, their prices increase and then the firms obtain profits from the high demand in boom. It should be noted that an increase of these materials can push a high pressure on the producers since the materials used for producing the goods are increasing and transportation cost will also increase simultaneously. Thus, all these indices are increasing in the end
PPIIEG	PPI intermediate energy goods (Not seasonally adjusted)	Monthly lagging indicator	This PPI index only monitors the prices of energy commodities.	
PPIFGS	PPI finished goods (Not seasonally adjusted)	Monthly lagging indicator	This index is the price index for all finished goods.	
PPIIDC	PPI industrial commodities (Not seasonally adjusted)	Monthly lagging indicator	This index is the price index for all industrial commodities.	
PPICEM	PPI crude energy materials (Not seasonally adjusted)	Monthly lagging indicator	This index is the price index for all crude energy materials.	

PPIITM	PPI intermediate materials (Not seasonally adjusted)	Monthly lagging indicator	This index is the price index for all the intermediate materials that should be further processed before they become finished goods.	thereby suppressing the corporate defaults in the long term. This study focuses on the short term. That is, higher PPI implies that the economy is in an expansion stage and firms have healthy financial condition. Therefore, these indices negatively influence default risk.

2.4 Equity market conditions

SP500	S&P500 index (Not seasonally adjusted)	Daily leading indicator	The S&P 500 is considered as a gauge of the large cap U.S. equities market which covers 500 leading companies in leading industries of the U.S. economy. These firms are publicly held on either the NYSE or NASDAQ, and they cover 75% of U.S. equities. A great amount of studies use this index to measure the economic activities since the stock market can provide the predictive content of future growth in the economy (Duffie et al., 2007, Duffie et al., 2009).	From the macro perspective, higher stock price implies that there is a high economic growth with active economic activities. In this context, firms' profits can increase thereby reducing the firms facing financial distress. From the micro perspective, an increase in equity prices can lead to increase in firm assets, which tends to reduce firm leverage and pushes down default rates. It is clear that there is negative relation between default risk and stock price. We average daily values for calculating monthly values.
NASDAQ	NASDAQ composite index (Not seasonally adjusted)	Daily leading indicator	This index is a market capitalization weighted index with more than 3000 common equities listed on the NASDAQ Stock Market.	

RU2000	Russell 2000 index (Not seasonally adjusted)	Daily leading indicator	This series measures the performance of the small-cap segment of the U.S. equity universe.
WILL5000	Wilshire 5000 total market full cap index (Not seasonally adjusted)	Daily leading indicator	This is to measure total market returns from the top 750 companies, which consists of reinvested dividends.

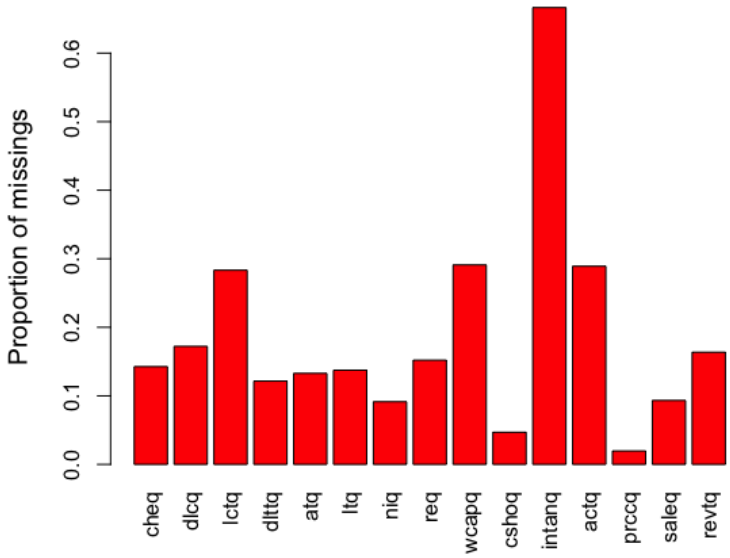
Note: the method of classifying these macroeconomic and financial time series into three groups including leading, cocindent, and lagging indicator follows the study done by The Conference Board (2001) and Shiskin and Moore (1967).

4.4 Microeconomic Database

Table 4-2 shows the third database, entitled microeconomic database, which consists of 35 micro factors. These variables are classified into five groups: 10 profitability ratios, 5 liquidity ratios, 6 leverage ratios, 4 activity ratios, and 7 other variables. All these factors are constructed using the CRSP and CCM data items. In particular, in order to fill in the missing values in each accounting variable, I use daily, monthly, annually CRSP and quarterly and annually CCM.

For the methods of how to handle missing values, I use the same approach as Duffie et al. (2007). Specifically, if the firm does not have at least 6 months of available data for stock information, this firm should be deleted. Additionally, for annual data, the corresponding records in CCM are DATA34, DATA5, DATA9, and DATA25, respectively. For cases with missing debt data, if the missing value corresponds to year Y, quarter Q, then I complete the data whenever possible as follows. Before presenting the procedure of handling missing values, I initially shows the summary of the scale of missing values in these variables mentioned above, which is shown in Table 4-3. This table shows the number of missing values in each variable and their corresponding percentage in the database.

Figure 4-3 Plot for the missing values



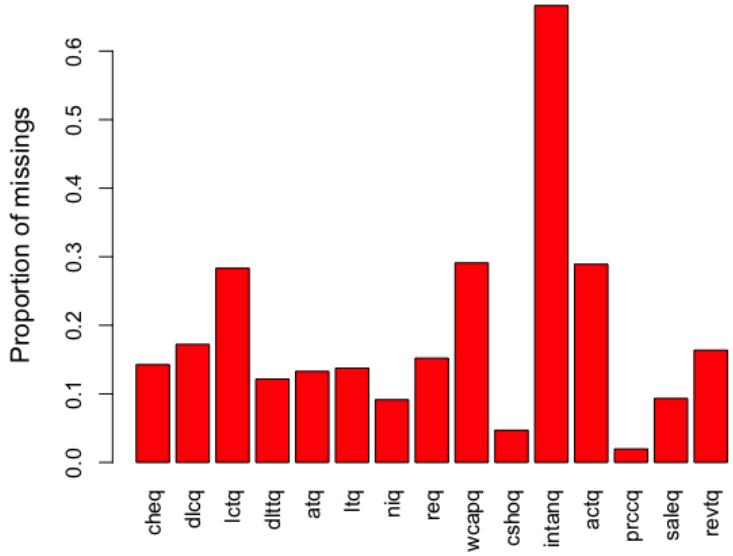
Note: the full name for each variabel in the horizontal axis is shown in Table 4-3

Figure 4-3 plots the proporation of missing values for these variables. For certain variables, they have a large amount of missing values in the database, For example, the missing values of intanq accounts for more than 66%. I thus need to handle the missing values first.

Table 4-3 Summary of missing values for each variable

Abbreviation	**Full name**	**Missing**	**Total**	**Percent**
cheq	Cash and Short-Term Investments	132581	929536	14.26%
dlcq	Debt in Current Liabilities	160074	929536	17.22%
lctq	Current Liabilities - Total	263316	929536	28.33%
dlttq	Long-Term Debt - Total	112970	929536	12.15%
atq	Assets - Total	123486	929536	13.28%
ltq	Liabilities - Total	127765	929536	13.75%
niq	Net Income	85005	929536	9.14%
req	Retained Earnings	141362	929536	15.21%
wcapq	Working Capital	270508	929536	29.10%
cshoq	Common Shares Outstanding	43622	929536	4.69%
intanq	Intangible Assets - Total	619437	929536	66.64%
actq	Current assets	268733	929536	28.91%
prccq	Price - Calendar Year - Close	18269	929536	1.97%
saleq	Sales/Turnover (Net)	86730	929536	9.33%
revq	Revenue	152154	929536	16.37%

Figure 4-3 Plot for the missing values



Note: the full name for each variabel in the horizontal axis is shown in Table 4-3

First, I only consider there to be missing data for firm F at year Y, quarter Q if there is debt data for firm F before and after Y, Q. Given that, if there are no missing quarterly debt data for F in year Y, I take the closest observation to quarter Q, before quarter Q, to complete the data set. If all quarterly debt observations for F in year Y are missing, I look in the yearly debt file for firm F, year Y. If this observation is not missing I take it to be the debt level for year Y, quarter Q, firm F. In all other cases of missing observations I take the data to be missing. Actually, when this study considers the other accounting variables, the method mentioned above is also used for identifying the missing values and filling in the missing values. The high frequency of accounting data is quarterly and this study uses the quarterly data to cover corresponding months' information. For example, following the guidance from Duffie et al. (2007), if there is quarter Q, then each month's data is the same as this quarter's data.

Winsorizing each variable is used to limit the influence of outliers. I replace any observation below the 5^{th} percentile with the 5^{th} percentile, and any observation above the 95^{th} percentile with the 95^{th} percentile. Below I provide details for 35 micro factors in Table 4-2. In order to

convert panel data to time series data for each micro factor, I average all the firms' micro factors in each month.

For the methods of how to calculate micro factors, this thesis follows the approaches used by Shumway (2001), Duffie et al. (2007), Lando and Nielsen (2010), and Campbell et al. (2008). This thesis shows a detailed algorithm of how to calculate DTD.ANNUALLY and DTD.MONTHLY in the following part.

Table 4-5 provides the summary statistics for 35 micro factors. For some variables, their mean values are equal to the median values, which means that these variables follow normal distribution. The others are either postively or negatively skewed.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
NITA	420	0.01	0.003	-0.0004	0.009	0.01	0.012	0.017
CACL	420	1.986	0.078	1.844	1.926	1.979	2.041	2.167
RETA	420	0.174	0.061	0.096	0.135	0.145	0.204	0.312
EBITTA	420	0.091	0.011	0.073	0.085	0.09	0.095	0.123
NIMTA	420	0.006	0.003	-0.002	0.004	0.006	0.007	0.015
WCTA	420	0.191	0.03	0.151	0.166	0.184	0.219	0.254
CSHTA	420	0.087	0.014	0.058	0.077	0.084	0.1	0.12
TLTA	420	0.583	0.024	0.529	0.576	0.586	0.599	0.62
TLMTA	420	0.458	0.044	0.338	0.429	0.46	0.486	0.569
MB	420	3.41	0.662	1.619	2.947	3.553	3.911	4.771
S.debt	420	0.165	0.037	0.093	0.132	0.169	0.201	0.222
BA	420	6.994	0.636	6.255	6.424	6.77	7.571	8.2
SIZE	420	7.35	0.728	6.357	6.677	7.209	8.051	8.746
METL	420	2.077	0.406	1.1	1.822	2.095	2.319	3.834
EXRET	420	0.001	0.037	-0.184	-0.016	0.004	0.021	0.127
NISALE	420	0.037	0.014	-0.017	0.029	0.038	0.048	0.068
EBITSALE	420	0.389	0.058	0.283	0.345	0.378	0.432	0.522
REMTA	420	0.125	0.065	0.052	0.083	0.093	0.148	0.304
CSHMTA	420	0.058	0.009	0.039	0.051	0.057	0.063	0.085
WCMTA	420	0.142	0.039	0.076	0.112	0.128	0.169	0.251
CLTA	420	0.23	0.016	0.2	0.218	0.231	0.243	0.266
CLMTA	420	0.175	0.033	0.106	0.148	0.17	0.195	0.271
REVWC	420	1.696	0.231	1.297	1.586	1.683	1.773	3.328
CLSALE	420	0.951	0.055	0.801	0.919	0.955	0.993	1.074
SALETA	420	0.558	0.077	0.421	0.496	0.57	0.6	0.726
SALEMTA	420	0.218	0.047	0.147	0.182	0.208	0.238	0.356
SA	420	5.58	0.529	4.913	5.126	5.37	6.132	6.627
PRICE	420	3.01	0.195	2.465	2.867	2.995	3.153	3.515
EBITMTA	420	0.06	0.013	0.044	0.053	0.057	0.063	0.112
Return.MONTHLY	420	0.007	0.037	-0.162	-0.011	0.007	0.025	0.141

Return.ANNUAL	420	0.007	0.173	-0.575	-0.086	0.028	0.113	0.493
SIGMA.DAILY	420	0.008	0.175	-0.36	-0.112	0.029	0.138	0.298
DTD.MONTHLY	420	12.473	3.542	3.111	10.274	12.298	14.473	25.85
DTD.ANNUAL	420	50.671	18.81	6.553	39.036	48.199	62.517	96.664

Figure 4-4 shows the heat map of correlation matrix in the lower triangle for 35 micro factors.[18] If the colour tends to be darker, the correlation between two variables becomes higher; otherwise, they become lower. White colour represents weak correlation. Blue colour represents positive strong or modertate correlation. Brown colour represents negative strong or moderate correlation. Compared with correlation matrix for macro factors in Figure 4-2, the correlation among these micro factors are higher since there are quite a few areas with white colour.

[18] The results can be provided upon requent.

Table 4-4 Data description for micro factors

Main category	Predictors	Description	Total
(a) Profitability ratios	NITA	Net income over total assets	8
	NIMTA	Net income over market value of total assets	
	EBITTA	Earnings before interest and taxes over total assets	
	EBITMTA	Earnings before interest and taxes over market value of total assets	
	RETA	Retained earnings over total assets	
	REMTA	Retained earnings over market value of total assets	
	NISALE	Net income over sales	
	EBITSALE	Earnings before interest and taxes over sales	
(b) Liquidity ratios	CACL	Current asset over current liabilities	6
	METL	Market equity to total liabilities	
	WCTA	Working capital over total assets	
	WCMTA	Working capital over market value of total assets	
	CSHTA	Cash and short-term investments over total assets	
	CSHMTA	Cash and short-term investments over market value of total assets	
(c) Leverage ratios	TLTA	Total liabilities over total assets	6
	TLMTA	Total liabilities over market value of total assets	
	CLTA	Current liabilities over total assets	
	CLMTA	Current liabilities over market value of total assets	
	RECL	Retained earnings over current liabilities	
	MB	Market-valued total assets over book value of firm[19]	
(d) Activity ratios	REVWC	Revenue over working capital	4
	CLSALE	Current liabilities over sales	
	SALETA	Sales over total assets	
	SALEMTA	Sales over market value of total assets	
(e) Other variables	S.DEBT	'short term' debt over total debt	11
	BA	log(book asset value)	
	SIZE	Log(market capitalization)	
	SA	Log(sale)	
	PRICE	Log (equity price)	
	EXRET	log of gross excess return over value-weighted S&P 500 return (EXRET)[20]	
	RETURN.MONTHLLY	The firm's trailing one-month stock return	
	RETURN.ANNUAL	The firm's trailing one-year stock return	
	SIGMA.DAILY	Standard deviation of each firm's daily stock return over the past 3 months.	
	DTD.ANNUALLY	Merton distance to default (MDTD) (Annually)	
	DTD.MONTHLY	Market value proxy distance to default (MVPDTD) (Monthly)*	
		Total	35

Note: The methods of calculating DTD.ANNUALLY and DTD.MONTHLY are shown in the manual of the Package "Kai.R.Tools" in Appendix 9.3.

[19] Book value of firm is calculated by Total assets- Intangible Assets- Total liabilities)

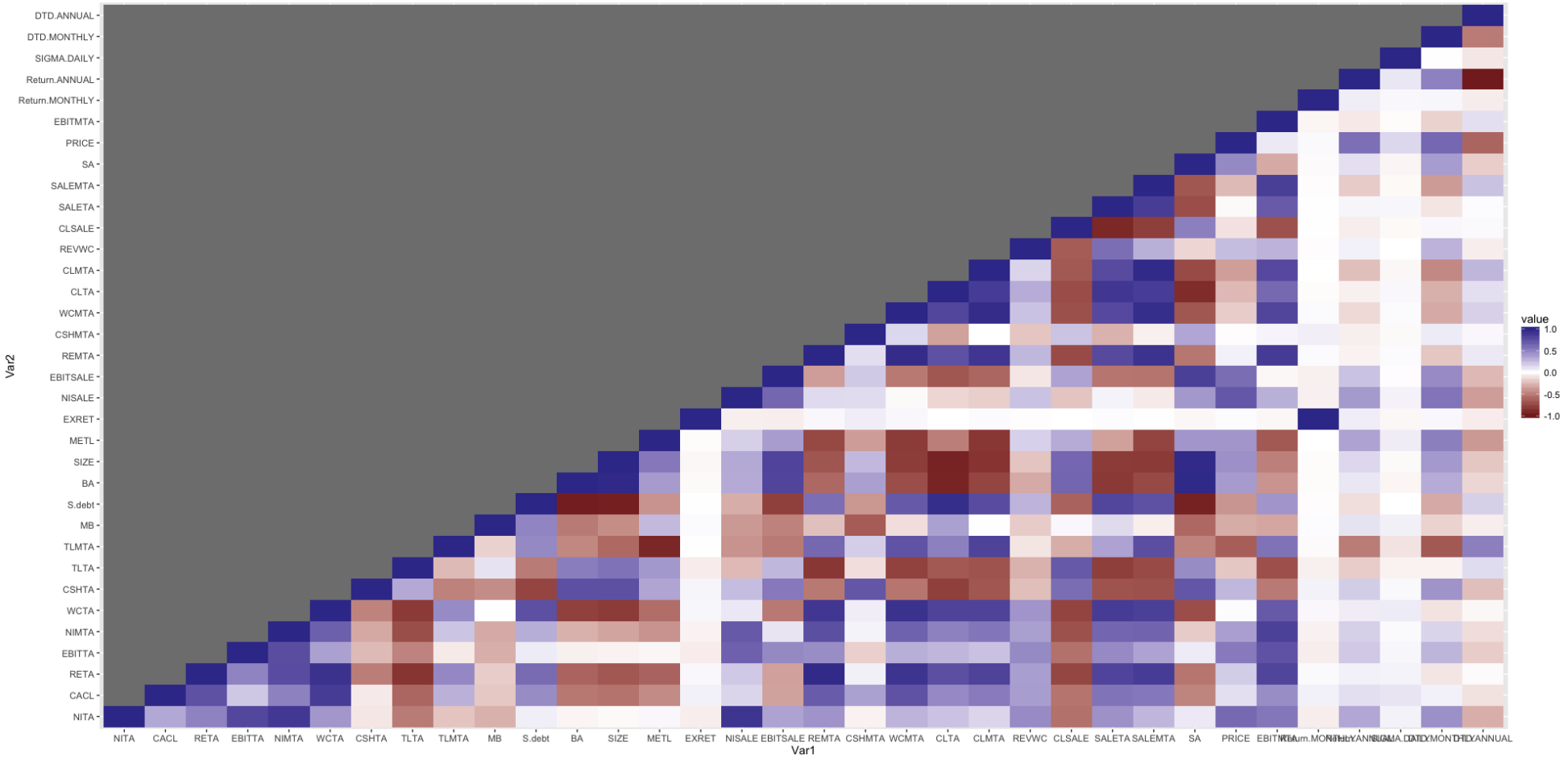
[20] The formula for calculating $EXRET_{i,t}$is $EXRET_{i,t} = \log(1 + R_{i,t}) - \log(1 + R_{S\&P500,t})$

Table 4-5 Summary statistics for 35 micro factors

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
NITA	420	0.01	0.003	-0.0004	0.009	0.01	0.012	0.017
CACL	420	1.986	0.078	1.844	1.926	1.979	2.041	2.167
RETA	420	0.174	0.061	0.096	0.135	0.145	0.204	0.312
EBITTA	420	0.091	0.011	0.073	0.085	0.09	0.095	0.123
NIMTA	420	0.006	0.003	-0.002	0.004	0.006	0.007	0.015
WCTA	420	0.191	0.03	0.151	0.166	0.184	0.219	0.254
CSHTA	420	0.087	0.014	0.058	0.077	0.084	0.1	0.12
TLTA	420	0.583	0.024	0.529	0.576	0.586	0.599	0.62
TLMTA	420	0.458	0.044	0.338	0.429	0.46	0.486	0.569
MB	420	3.41	0.662	1.619	2.947	3.553	3.911	4.771
S.debt	420	0.165	0.037	0.093	0.132	0.169	0.201	0.222
BA	420	6.994	0.636	6.255	6.424	6.77	7.571	8.2
SIZE	420	7.35	0.728	6.357	6.677	7.209	8.051	8.746
METL	420	2.077	0.406	1.1	1.822	2.095	2.319	3.834
EXRET	420	0.001	0.037	-0.184	-0.016	0.004	0.021	0.127
NISALE	420	0.037	0.014	-0.017	0.029	0.038	0.048	0.068
EBITSALE	420	0.389	0.058	0.283	0.345	0.378	0.432	0.522
REMTA	420	0.125	0.065	0.052	0.083	0.093	0.148	0.304
CSHMTA	420	0.058	0.009	0.039	0.051	0.057	0.063	0.085
WCMTA	420	0.142	0.039	0.076	0.112	0.128	0.169	0.251
CLTA	420	0.23	0.016	0.2	0.218	0.231	0.243	0.266
CLMTA	420	0.175	0.033	0.106	0.148	0.17	0.195	0.271
REVWC	420	1.696	0.231	1.297	1.586	1.683	1.773	3.328
CLSALE	420	0.951	0.055	0.801	0.919	0.955	0.993	1.074

SALETA	420	0.558	0.077	0.421	0.496	0.57	0.6	0.726
SALEMTA	420	0.218	0.047	0.147	0.182	0.208	0.238	0.356
SA	420	5.58	0.529	4.913	5.126	5.37	6.132	6.627
PRICE	420	3.01	0.195	2.465	2.867	2.995	3.153	3.515
EBITMTA	420	0.06	0.013	0.044	0.053	0.057	0.063	0.112
Return.MONTHLY	420	0.007	0.037	-0.162	-0.011	0.007	0.025	0.141
Return.ANNUAL	420	0.007	0.173	-0.575	-0.086	0.028	0.113	0.493
SIGMA.DAILY	420	0.008	0.175	-0.36	-0.112	0.029	0.138	0.298
DTD.MONTHLY	420	12.473	3.542	3.111	10.274	12.298	14.473	25.85
DTD.ANNUAL	420	50.671	18.81	6.553	39.036	48.199	62.517	96.664

Figure 4-4 Correlation heat map for 35 Micro factors



4.5 Conclusion

This chapter gives a detailed description of how to construct three main databases including the corporate default databases, the macroeconomic database, and the microeconomic database by using four main databases, such as Moody's DRD database, FRED, CRSP, and CCM.

By using the functions in the R package "Kai.R.Tools", I can use relevant parameters to construct different types of corporate default databases for different countries in different economic sectors. In order to obtain listed and nonlisted firms in the corporate default database, both CRSP and CCM can be used to achieve this purpose. The basic information in this database consists of monthly defaults for the firms or bonds, number of monthly rated issuers or bonds, and monthly corporate or bonds default probabilities.

The second database is the macroeconomic database, which provides 114 macro factors in the US economy. They are classified into 5 major groups based on economic process. Meanwhile, I further classify them into three different groups of factors, depending on the timing movement with the business cycle.

The third database is the microeconomic database, and it covers 35 micro factors. They cover the majority of popular firm-specific variables used by prior studies, such as different types of accounting ratios, market-driven variables, DTDs, and the idiosyncratic standard deviation of stock returns etc. In addition, this chapter also details the information about how to extract relevant default information to calculate two types of indicators – IDPs and VDPs, respectively.

CHAPTER 5

5. Investigating Effective Predictors (EPs) for Corporate Defaults

5.1 Introduction

In this chapter, I use almost one hundred macroeconomic and financial variables to investigate dynamic change in the effective predictors for corporate default probabilities (DPs) in US industrial firms from April 1980 to December 2014. To provide additional perspective, I focus on how the monetary policies implemented by central banks impact corporate defaults over time.

Two perspectives motivate me to carry out this study. First, although using firm-level data to predict corporate defaults in the US has been the focus of many studies, relatively little attention has been given to predict corporate defaults by using industrial level data in the literature.[21] The primary reason for this phenomenon may result from two characteristics of historical data on corporate DPs based on issuer-weighted measurement. One is that default is a deceptively rare event.[22] Certain firms rated by Moody's may not have been observed in the past. Consequently, the value of issuer-weighted default probabilities (IDPs) was quite small and always equal to zero in the past in the industry since there were no default events in certain months. This nature of IDPs leads to poor performance of models for prediction

[21] Industrial level data means time series of macro factors, such as GDP, CPI, and PPI. These data can only reflect macro environment for these industrial firms rather than micro environment.

[22] Crosbie and Bohn (2002) state that the typical firm has a default probability of around 2% in any year.

(Zmijewski, 1984);[23] another issue is that there are not enough default events recorded by rating agencies. Lando and Nielsen (2010) argue that the material is sparse until 1982 in terms of corporate default data in Moody's DRD. In this study, I hope to provide a mechanism of allowing for special features of corporate defaults (numerous months with zero value in the sample) in order to find which corporate risk factors influence corporate defaults in industrial firms and how they impact on the corporate defaults over time.

Second, numerous theoratical and empirical studies have identified that many macroeconomic elements have influence on corporate defaults. However, a few studies investigate whether these macroeconoic elements that can explain corporate defaults are stable rather than dynamic; that is, I am interested in whether the explanatory power of these macroeconomic elements on corporate defaults is related to the time period or not.

More specifically, Table 5-1 summarises prior studies employing models to calculate corporate DPs. Based on different time periods and the same sample, the outcomes of credit risk covariates that can explain DPs are not the same in the research done by Azizpour et al. (2014), Lando and Nielsen (2010), Duffie et al. (2009), and Das et al. (2007). For example, GDP has different performance in different sample periods. Azizpour et al. (2014) investigate the period from January 1970 to December 2013 and find that GDP has predictive power for corporate defaults. However, this is not the case in the study done by Das et al. (2007). They use a sample from January 1979 to October 2004 and identify that GDP has no predictive power on corporate defaults. Likewise, the other predictors have different performance in different periods.

For this scenario, Bernanke (1993) explains that default risk is not constant but dynamic. It is indicated that the predictors for explaining default risk should not be stable over time. He argues that the predictive power for the predictors should reflect the market's perception of default risk. It is suggested that when the market's perception towards default risk changes,

[23] Zmijewski (1984) states that estimating financial distress models based on non-random samples can cause biased parameter and probability estimates. This bias result is produced by two issues. One is from "oversampling" distressed firms (also called topic of choice-based sample biases. Another is from using a "complete data" sample selection criterion (called sample selection biases). The two issues results from sample selection and data collection constraints. The first constraint is the tremendously low frequency rate of firms exhibiting financial distress characteristics in the population. The second is that data for financial distressed firms are often unavailable.

the predictor should change as well. As a result, it is interesting to investigate further how default risk changes over time in the past thereby helping decision makers to understand the dynamic change of the credit market in order to reduce the default risk, expecially for understanding the interaction between implementation of monetary policies and stability of the credit market thereby helping them to make an appropirate decision to use monetary tools to stablise the economy and the credit market simultaneously.

This chapter proceeds as follows. In Section 5.2, I describe the data in this study. The empirical analysis and discussions are presented in Section 5.3, which covers two main issues: identification of the phenomenon of dynamics in EPs and exploration of key EPs. Section 5.4 shows the results from robustness testing. Section 5.5 concludes.

Table 5-1 Summary of previous studies about modelling corporate default rates with various risk factors

Authors/Year	Data	Observed Frailties *	Unobserved Frailties*	Results
Azizpour et al. (2014)	US, **Industrial and Financial issures**, 1/1/1970-12/31/2012	**External Risk Drviers:**1, Industrial Production; 2, GDP; 3, 1-year S&P 500 index/its Volatility; 4, 3-month Treasury bill; 5, Spread (10-year Treasury note & 1-year Treasury bill); 6, Spread (the yields of BAA & AAA Moody's rated corporate bonds).	The constant coefficient of the regression	**INFLUENCE:** 1, GDP; 2, Unmeasured frailties; **NO INFLUENCE:** Corporate bond yield spread; 2, S&P 500 index/its volatility; 3, Industrial production; **SUMMARY:** the frailty and contagion channels of clustering IMPROVE calculation of Default Rates.
Lando and Nielsen (2010)	US,**Non-Financial issuers**,1/1982-12/2005	**External Risk Drivers:** 1, 1-year S&P 500 index; 2, 3-month treasury rate; 3, Industrial production; 4, Spread (10-year treasury rate &1-year treasury rate); **Internal Risk Drivers:** 1, 1-year equity return; 2, Distance-to-default; 3, Quick ratio; 4 Percentage short term debt; 4, Book asset value; 5, Firm size; 6, Decay rate.	The constant coefficient (Contagion effect)	**INFLUENCE:** 1, 1-year S&P 500 index; 2, US 3-month treasury rate; 3, Industrial production; 4, Distance-to-default; 5,Book asset value; **NO INFLUENCE:** 1, Unemployment rate; 2, Wages of US production workers; 3, US consumer price index; 4, GDP(real & nominal); 5, price of crude oil; 5,Spread (Moody's Aaa- & Baa-rated corporate bonds); 6, Some of the empirical default predictors proposed by Altman (1968) and Zmijewski (1984).
Duffie et al. (2009)	US, **Non-financial firms**, 01/1979-03/2004	**External Risk Drivers:** 1, 1-year S&P 500 index; 2, 3-month treasury rate; 3, Industrial production; 4, Spread (Average BBB-AAA corporate bond yield); 5, Industry average distance to default; 6, Firm size; **Internal Risk Drivers**: 1, Distance-to-default; 2, Trailing stock return.	**1, Latent-factor volatility** (Unobservable firm-specific covariate); **2, Latent-factor mean reversion**(Unobservable macroeconomic covariate);	**INFLUENCE:** 1, 1-year S&P 500 index; 2, US 3-month treasury rate; 3, Distance to default; 4, Trailing returns of the firm; **NO INFLUENCE:** 1, Firm size; 2,Industry average distance to default; 3, Industrial production; 4, Spread (Average BBB-AAA corporate bond yield).

Das et al. (2007)	US, **Non-Financial Issuers, Non-utilities**, 1/1979-10/2004	**External Risk Drivers:** 1,1-year S&P 500 index; 2, 3-month treasury rate; 3, GDP; **Internal Risk Drivers:** 1, 1-year equity return; 2, Distance-to-default.	Doubt there could be contagion or frailties (Unobserved covariates that are correlated across firms)	**INFLUENCE:** 1, Industrial Production; 2, US 3-month treasury rate; **NO INFLUENCE:** 1, GDP.

Note: * Frailties can be divided into two parts, observable and unobservable ones. Two types of frailties include external and internal risk drivers. Unobservable drivers cannot obtained in the real world but abstracted from the model, and normally the error term cover this information.

5.2 Data

I focus on corporate defaults in the industrial economic sector in the US. There are three reasons for this. First, the other economic sectors, such as media, financial sectors, gas and electric utilities, the return on equity, revenues, and thus the risk of the default, are strongly influenced by regulators (Eom et al., 2004). Economists consider that a strong industrial sector is a sign of a well-functioning economy with a high GDP and high quality of life. That is, the performance of the industrial sector is associated with macroeconomic conditions. This selection can ensure that the results of using changes of macroeconomic conditions to predict the dynamic of defaults is not biased, indicating that it can avoid influence from government interventions. Second, the data in this study are almost consistent with the data in the previous studies, such as Duffie et al. (2007), Duffie et al. (2009), Lando and Nielsen (2010), Davydenko (2012) in order to determine that the results of the dynamic in the EPs in using industrial level data are reliable. Third, all the US industrial firms with a debt issue registered in Moody's DRD can match with stock market data from CRSP and accounting information from Compustat. Two main databases including corporate default database and macroeconomic factors database are constructed in this study, which is described in Chapter 4.

Time span for the constructed data is from April 1980 to December 2014. Although Moody's started to record the default in 1898, the quality of default records is not reliable from 1898 to 1980. There are no available default data before January 1980 that can match with CRSP data (or Compustat) based on CUSIP value for merging. Lando and Nielsen (2010) note that the material is sparse until 1982 in terms of corporate default data in Moody's DRD. Also, there is the question of availability of macro factors that can be used after lagging macro factors by three months for prediction of corporate DPs.

The first database is corporate default database which draws elements from three main data sources DRD, and CRSP and Compustat. The methods used to extract the corporate default data set from those three data sources corresponds to the previous studies (Duffie et al., 2007, Lando and Nielsen, 2010, Davydenko, 2012). Then I clean dependent default events in

Moody's default database. Specifically, I first remove the influence of family structures, and the parent's consolidated financial information is used to study the default decision for the whole group of bond issuers. In addition, if a firm consecutively defaults more than once within two years, all these multiple default events are counted as a single default event and the first default date is regarded as the default date. Then, the redundant default events can be handled. Consequently, this study chooses the US firms classified as 'Industrial' group in DRD from April 1980 to December 2014. This leaves me with a total of 2655 firms comprising 614 defaults and these are all public listed firms. The method of measuring default risk is based on IDPs, which counts the number of monthly defaults over the rated firms in the beginning of each month.

The second database is constructed from Federal Reserve Economic Database (FRED). There are 92 macro factors in Table 5-2, which are grouped into five broad categories: (a) bank lending conditions; (b) macroeconomic and business cycle indicators, including labour market conditions and monetary policy indicators; (c) open economy macroeconomic indicators; (d) micro-level business conditions such as wage rates, cost of capital, and cost of resources; and (e) stock market returns and volatilities.

Table 5-2 Data description of macro factors

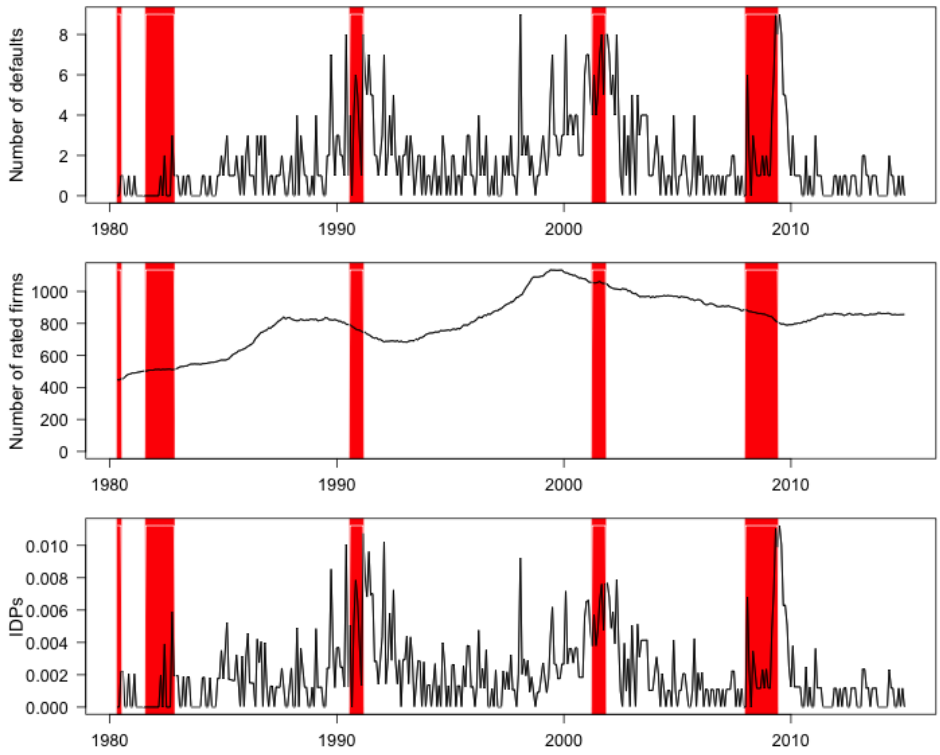
Main category	Summary listing	Factors				Total
(a) Bank lending conditions (9 factors)	Size of overall lending	Total loans and leases, all banks	LOANS	Total borrowings from fed reserve	BORROW	9
		Total real estate loans	REALLNNSA	Household debt/income-ratio	TDSP	
		Total consumer credit outst.	TOTALNS	Household obligations/income	FODSP	
		Commercial & industrial loans	BUSLOANSNSA	Total public debt	GFDEBTN	
		Consumer loans	CONSUMER			
(b) Macro and BC conditions (50 factors)	General macro indicators	Industr. production index	INDPRO	Personal consumption expenditures: chain-type price index	PCEPI	23
		Smoothed recession probabilities	RECPROUSM156N	Expenditure durable goods	PCEDG	
		Industr. production-manufacture	IPMANSICS	Economic activity index	USPHCI	
		Industr. production: mining	IPMINE	Gross domestic product	GDP	
		Industr. production: electric and gas utilities	IPUTIL	Gross national product	A001RP1Q027SBEA	
		Manuf. sector productivity	IPMAT	Government total expenditures	W068RCQ027SBEA	
		ISM manufacturing: new orders index	NAPMNOI	University of Michigan: consumer sentiment	UMCSENT	
		ISM manufacturing: inventories index	NAPMII	Real disposable personal income	A067RL1Q156SBEA	
		moving 12-month total vehicle miles travelled	M12MTVUSM227NFWA	Real gross domestic product	GDPC1	
		ISM manufacturing index	NAPM	Gross domestic product: implicit price deflator	PCEC96	
		Personal income	PI	Gross private domestic investment	GPDI	
		Consumption expenditure	PCE			
	Labour market conditions	Civilian unemployment rate	UNRATE	Civilian employment	CE16OV	7
		Weekly hours worked	HOHWMN02USM065S	Unemployed, more than 15 weeks	UEMP15OV	
		Employment/population-ratio	EMRATIO	All Employees: manufacturing	MANEMP	
		total no unemployed	LNU03025703			

	Business cycle leading/coinciding indicators	New houses sold	HSN1F	Housing starts	HOUST	8
		Capacity util. manufacturing	MCUMFN	New building permits	PERMIT	
		Capacity util. total industry	TCU	Real final sales of domestic product	A190RL1Q225SBEA	
		Light weight vehicle sales	ALTSALES	Final sales to domestic buyers	FSDP	
	Monetary policy indicators	M2 money stock	M2SL	St. Louis adjusted monetary base	AMBSL	10
		UMich infl. Expectations	MICH	CPI: energy index	CPIENGSL	
		Personal savings	PMSAVE	CPI: all items less food	CPILFESL	
		Required reserves of depository institutions	REQRESNS	Gross private saving	GPSAVE	
		Personal savings rate	PSAVERT	Gross domestic product: implicit price deflator	GDPDEF	
	Money stock from non-banks	Institutional money funds	IMFSL			1
	Firm profitability	Net corporate dividend payments	B056RC1A027NBEA	Corporate net cash flow	CNCF	2
(c) International competitiveness (6 factors)	Terms of trade	Trade weighted U.S. dollar index: broad	TWEXBMTH	Trade weighted U.S. dollar index: major currencies	TWEXMMTH	2
	Balance of payments	Balance on merchandise trade (discontinued series)	BOPBM	Real exports of goods & services	EXPGSC1	3
		Real imports of goods & services	IMPGSC96			
(d) Micro-level conditions (21 factors)	Cost of capital	Bank prime loan rate	MPRIME	BAA corporate bond yield*	BAA/BAA.v	14
		30 year mortgage rate	MORTG	Treasury bond yield, 10 years(Aaa)	AAA10YM	
		Effective federal funds rate*	FEDFUNDS/ FEDFUNDS.v	Treasury bond yield, 10 years(Baa)	BAA10YM	
		AAA corporate bond yield*	AAA/ AAA.v	BofA Merrill Lynch US corp master total return index value	BAMLCC0A0CMTRIV	
		1-year treasury constant maturity rate*	GS1/GS1.v	10-Year treasury constant maturity rate*	GS10/GS10.v	
		3-month treasury bill: secondary market rate*	TB3MS/ TB3MS.v	Interest spread: difference between 10-year treasury constant maturity rate and 1-year one	SPREAD.GS	

		Interest spread: difference between BAA yield and AAA yield)	SPREAD.MOODY.1	Interest spread: (Difference between 10-year BAA yield and 10-year AAA yield)	SPREAD.MOODY.2	
	Cost of resources/Labour cost	PPI all commodities PPI interm. energy goods PPI crude energy materials PPI finished goods	PPIACO PPIIEG PPICEM PPIFGS	PPI industrial commodities PPI intermediate materials Unit labour cost: nonfarm business	PPIIDC PPIITM ULCNFB	7
(e) Equity market conditions (1 factor)	Equity indices and respective volatilities	Annual return of S&P 500	SP500.v			1
Total number of risk drivers						86

Note: All the factors are transformed to growth rate except annual return of S&P 500, SPREAD.GS, SPREAD.MOODY.1, and SPREAD.MOODY.2. * means this factor has two types of data. In addition to use growth rate, raw data is applied in this study. Considering the non-transformed factors with "*", there are 92 macro factors in the macroeconomic database.

Figure 5-1 Time series of monthly default data in the US industrial firms from April 1980 to December 2014



This figure shows time series of the number of defaults (total defaults), number of rated month (total exposures), and IDPs in the US industrial firms from April 1980 to December 2014. The first panel presents time series plots of the total defaults in the end of each month d_t; the second panel is the time series of number of rated firms in the beginning of each month n_t; the final panel is the time series of monthly IDPs. IDPs are fractions in which the numerator represents the number of issuers that defaulted in a particular time period in the first graph and the denominator represents the number of issuers that could have defaulted in that time period in the second graph. The formula of IDP in the third graph is shown below: $IDPs = d_t^{IDPs}/n_t^{IDPs}$, where $n_t^{IDPs} = n_{t-1}^{IDPs} - d_{t-1}^{IDPs} - w_t^{IDPs}$. The numerators$d_t^{IDPs}$ and d_{t-1}^{IDPs} are the numbers of issuers defaulting at period t and $t-1$. The denominators n_t^{IDPs} and n_{t-1}^{IDPs} are the numbers of issuers that potentially could have defaulted at date t and $t-1$. w_t^{IDPs} denotes the number of credits which withdraw between periods t and $t+1$ (Acharya and Carpenter, 2002, Gansecki, 2010, Moody, 2015).

Figure 5-1 presents the total number of defaulters, rated firms, and IDPs. It can be seen that default clustering around the recession years of 1990, 2001, and the financial crisis of 2007-2009 in each panel except the middle one is very clear. Before doing empirical analysis, all the factors and IDPs should be filtered in order to reduce the strong short-term fluctuation in these time series. Especially, for the IDPs, default is a kind of rare event which means that its number can either be quite small (even no defaults) in stable economic environment or quite large in non-stable environment in Figure 5-1. In particular, the number of rated firms

as denominator in IDPs is normally quite large. Then the value of IDPs is quite small. Hodrick-Prescott (HP) filter method can balance this situation by extracting the trend component in order to remove the influence from many zero values in the data. More importantly, after using filtering method, the results of investigation of effective predictors by using variable selection technique called LASSO are stable, which allows me to correctly understand how macro factors explain IDPs.[24] In recent years, HP filter, developed by Hodrick and Prescott (1997), has become a popular method for removing the growth component from data exhibiting long run trends over time. The HP filter has been much applied in the RBC literature as a method of isolating business cycle frequencies from the long run trend component in both actual and simulated data. Therefore, the HP filter based on gamma (γ) equal to 5 is used in this study.

As noted, to make macro factors have predictive effects, this study lags each macro factor by three months following previous study (Duffie et al., 2007).[25] This study only considers

[24] I also considered the other methods for doing variable selection, such as principal component analysis (PCA), subset selection methods (best subset selection and stepwise selection: forward and backward), even more sophisticated methods (random forest, support vector approach). They are not prefereable in practice. For example, the results obtained by PCA is not stable. Specifically, PCA is unsupervised approach. The response (dependent) variable is not used to determine the component direction. Then the results are quite hard to be interpreted. In addtion, the standard for selecting the number of components in PCA is not fix. The number of components is decided by how much proportion of variance can be explained. However, there is no fix method to determine the proportion of variance. Different components lead to quite different results. Then PCA is not reliable in practice. For the subset selection methods, although they can provide interpretable results, this method can be extremely variable since it is a discrete process. If the other more advanced approaches are used, the results are quite hardly interpretted. Compared with these methods, LASSO has many advantages. Specifically, the basic understanding of this model is to shrink some coefficients and sets others to zero and then attempt to retain the good features of subset selection and ridge regression. This method's results are easily interpreted and they are stable. Additionally, the major advantage of using Shrinkage method is rooted in the bias-variance trade-off. That is, LASSO approach can generate a reduction in variance at the expense of a small increase in bias, and consequently can yield a smaller expected test MSE in comparison of the subset selection based on method of least square estimation (James et al., 2013).

[25] According to The Conference Board (2001), cyclical indicators are classified into three different groups including leading, coincident, and lagging indicators based on the timing of their movements. Leading indicators tend to move direction before the business cycle occurs and the timing comparisons between the cyclical turn in the indicators and business cycle turn is one months or more, such as, new orders, housing starts, and consumer sentiment. Lagging indicators tend to shit direction after the business cycle occurs or coincident series occur, such as interest rate, unit labour cost, and cost of resources. For the timing comparisons, these indicators follow the business cycle turn by one month or more. Coincident indicators are the series that measure aggregate economic activity which are used to define the business cycle. For example, these indicators are employment, production, and personal income etc. Coincident indicators are special in terms of timing comparisons since they may follow or lead turning points in the business cycle by three months or less, including that they completely coincide with the business cycle. Therefore, lagging factors do not mean that these factors are lagged by specific months compared with the other factors such as coincident indicators and leading indicators.

macro factors rather than micro factors for the listed firms in the US.[26] In other words, I am interested in how effective macro predictors dynamically change over time. The motivations for doing this are fourfold. First, the micro factors are firm-specific variables, such as accounting ratios, DTD, and stock return etc. According to Duffie et al. (2007), the quarterly data are converted to monthly data by using the same value of corresponding data and three months within a quarter are the same. Actually, this method is not reliable and has poor performance in correctly showing monthly financial condition in each firm.

Second, Duffie et al. (2007) also argue that "the best way to derive historical correlations may be through a mutual dependence on macroeconomic factors, such as growth rates in the economy." It is implied that the research about the macroeconomic effects among firms should be given priority in terms of calculation of default rates since macroeconomic factors are able to capture the correlation among default events.

Third, micro factors are panel data. There are thousands of monthly data from corresponding firms in each month. In order to transfer cross-sectional data to a single value for each month, using average value to transform the data is not appropriate since it neglects the influence from firm size. Even with firm size using weighted average, the issue of 'too big to fail' is also introduced in this case, which leads to poor performance in measuring real corporate financial performance.

Fourth, the most important reason is that accounting information is not sufficiently reliable, as identified by prior studies. In practice, credit rating agencies (CRAs) mainly rely on accounting information to give the rating to the issuers. They failed to correctly rate the firms, such as Enron and WorldCom in 2001. Meanwhile, several micro factors are constructed by stock price. These variables are only reliable under the consumption of efficient market hypothesis (EMH). This consumption is not appropriate and they have poor performance as accounting ratios. Last but not least, from an economic perspective, firms' accounting ratios as microeconomic information cannot sufficiently reflect changes of macroeconomic economic conditions.

[26] These micro factors are described in detail in Chapter 4, which covers the majority of popular micro factors used by prior studies such as, Shumway (2001), Chava and Jarrow (2004), Duffie et al. (2007), Campbell et al. (2008), Lando and Nielsen (2010), and Tian et al. (2015).

5.3 Empirical Analysis and Discussions

This section is to investigate EPs for corporate defaults from three perspectives. First, this study is initially to identify the phenomenon of dynamic in EPs from April 1980 to December 2014. Second, this study is to further explore key EPs for historical corporate defaults within that period. Third, based on these results, this study attempts to interpret these results through financial understanding.

5.3.1 Identification of Dynamic in EPs

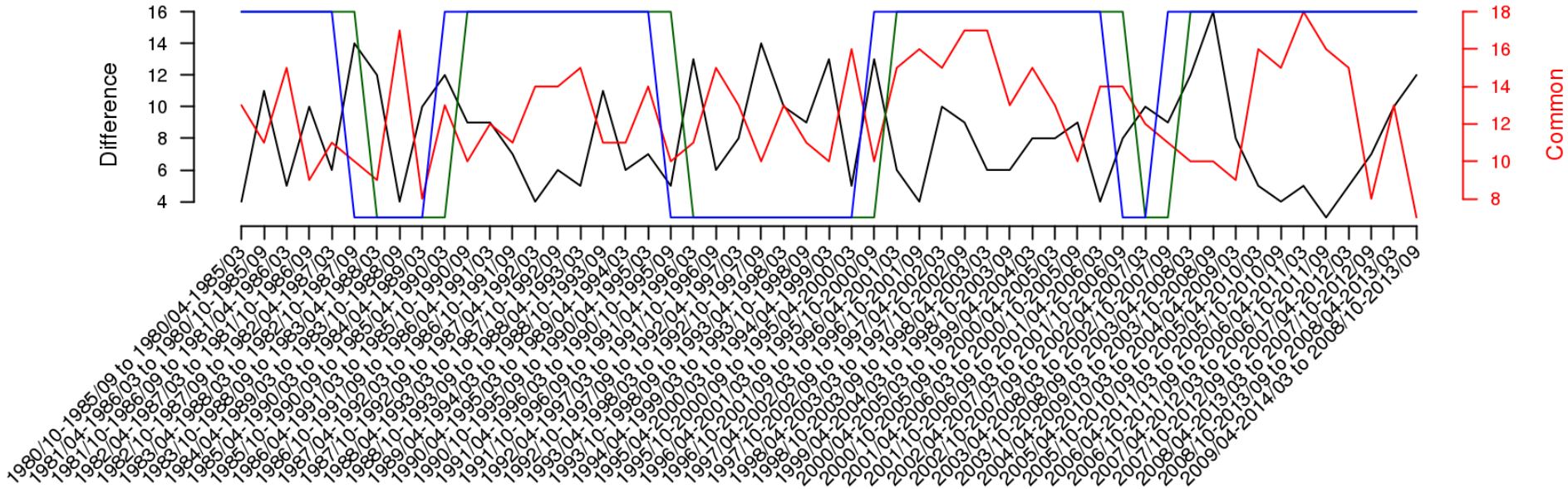
Since this study has its sample from April 1980 to December 2014, in order to cover the majority of the sample period, the size of time span for shifting the sub-sample is from 0 month to 54 months. Each sub-sample has 6 intervals and each interval covers a 5-years period with 60 monthly observations. Then there are 360 monthly observations in each sub-sample. LASSO is used to extract the EPs in each interval. I find that the minimum size of each interval for extracting EPs should be 5 years. Otherwise, LASSO cannot generate results.[27]

Meanwhile, I use 6 months to shift the subsamples with 360 observations. Then there are 9 sub-samples and 54 groups of EPs extracted from 54 intervals. This approach is actually a sliding window analysis based on a 6-month length window. The results are displayed in Figure 5-2 and Figure 5-3. Both figures show how the EPs change dynamically over time. The former shows the change in the EPs extracted in 54 intervals between two consecutive intervals; the latter shows their changing ratios. The trend of black line representing difference of EPs is almost opposite that of the red line representing common EPs in both figures. That is, as the proportion of common EPs increases, the proportion of different EPs decreases. As the proportion of common EPs decreases, the proportion of different EPs increases. It should be noted that the changing behaviour of EPs is not influenced by recession periods. EPs still dynamically change over time even in non-recession periods.

[27] I also use a larger sample size for each interval, for example, 6 years, 9 years, and 12 years. There are no big differences in these results. Therefore, I only use the minimum size 5 years for extracting effective predictors.

Therefore, the phenomenon of dynamic change in EPs is clear. Next is to further explore what key EPs have more impact on IDPs.

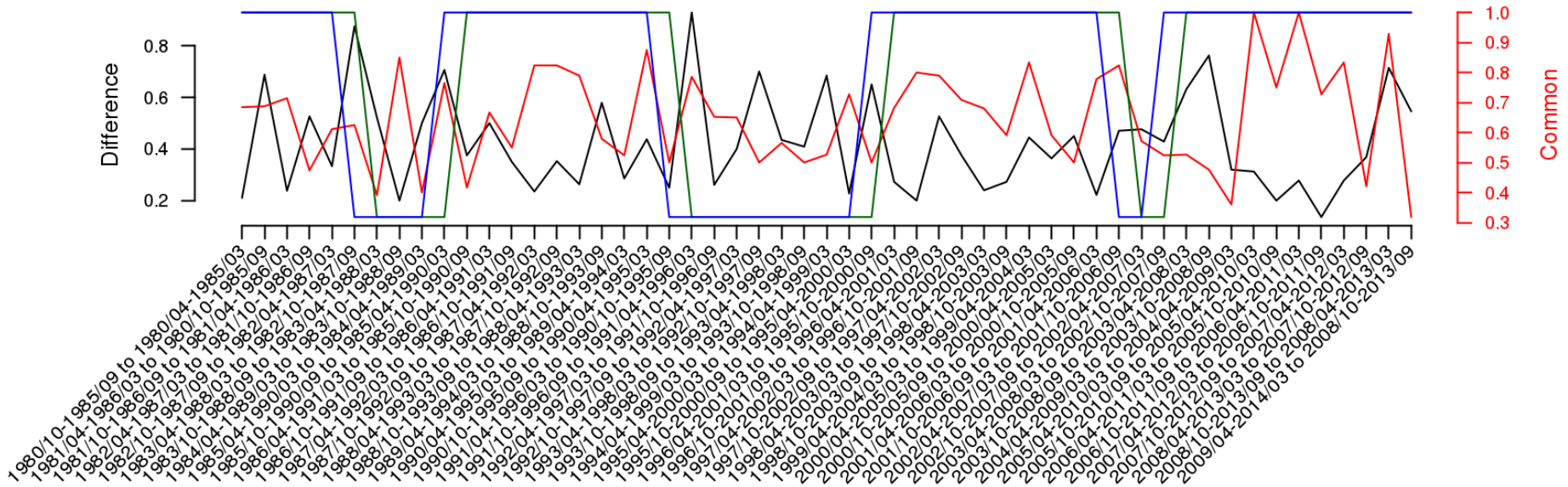
Figure 5-2 Changes in the EPs extracted in 54 intervals



Note: The total sample is from April 1980 to March 2014. 5-year as the size of each internal is used to extract a set of EPs. Each sub-sample has 6 intervals. For example, if the starting point in the first sub-sample is April 1980, then the ending point is March 2014. This paper starts from the first sub-sample to the final sub-sample through shifting the interval (5 years' window) by 6 months, which means that this study moves the first sub-sample covering the period of April 1980 to March 2010 to the last sub-sample covering the period of January 1985 to March 2014. Then there are 9 sub-samples and 54 groups/intervals of EPs. Difference in black colour means that the number of new EPs added in the current interval compared with the previous interval. Common in green line means that number of EPs in the current interval being same as the EPs in the previous internal. The horizontal points without having both green line and blue line mean that these intervals do not cover recession period.

The reasons are why I use 6 months are shown as follows. First, 6 months are frequently used for analysing the performance of the portfolio's return in asset pricing research. Second, the majority of macro factors are released later in each year. 6 months can allow for the lagging effects in the macro factors. Third, it is computational effects compared with moving the sample by one month in this part, and I can draw the graph for showing this dynamic phenomenon in the EPs. I also use 12 and 18 months for shifting the sub-samples and the results are still consistent. This study only reports the results of using 6 months. This approach is a sliding window analysis based on a 6-month length window.

Figure 5-3 Changing ratios of EPs extracted in 54 intervals



Note: The total sample is from April 1980 to March 2014. 5-year as the size of each internal is used to extract a set of EPs. Each sub-sample has 6 intervals. For example, if the starting point in the first sub-sample is April 1980, then the ending point is March 2014. This paper starts from the first sub-sample to the final sub-sample through shifting the interval (5 years' window) by 6 months, which means that this study moves the first sub-sample covering the period of April 1980 to March 2010 to the last sub-sample covering the period of January 1985 to March 2014. Then there are 9 sub-samples and 54 groups/intervals of EPs. Difference in black colour means that the number of new EPs added in the current interval compared with the previous interval is divided by the number of EPs extracted in the previous interval. Common in green line means that number of EPs in the current interval being same as the EPs in the previous internal is divided by the number of EPs in the previous interval. The horizontal points without having both green line and blue line mean that these intervals do not cover recession period.

The reasons are why I use 6 months are shown as follows. First, 6 months are frequently used for analysing the performance of the portfolio's return in asset pricing research. Second, the majority of macro factors are released later in each year. 6 months can allow for the lagging effects in the macro factors. Third, it is computational effects compared with moving the sample by one month in this part, and I can draw the graph for showing this dynamic phenomenon in the EPs. I also use 12 and 18 months for shifting the sub-samples and the results are still consistent. This study only reports the results of using 6 months. This approach is a sliding window analysis based on a 6-month length window.

5.3.2 Exploration of Key EPs

The method used in the previous part cannot give a full picture of how EPs dynamically change over time since each sub-sample is shifted by 6 months. In this part, this study shifts each interval with 5 years' observations by one month, which is actually a rolling window analysis. Then I have 357 groups of EPs rather than only 54 groups in the previous method.

The EPs are only selected based on value of frequency that is at least more than one-third of all groups. For example, in this case, I have 357 groups of EPs. Then one-third of 357 means that I choose the EPs appearing more than (equal to) 107 out of 357 groups of EPs. Thus I get the results in Table 5-4.

Table 5-4 shows the results of key EPs for explaining IDPs from April 1980 to December 2014 in US industrial firms. It is clear that there are no EPs that can continually and strongly explain IDPs. The strongest EP is BORROW (Total Borrowings of Depository Institutions from the Federal Reserve) which can only explain IDPs in 54% of 357 intervals. This study introduces a method used in the correlation study done by Cohen (1988) to classify the strength of explanatory power of EPs, which is shown in Table 5-3. There are three classifications of predictive power depending on percentage groups (p) that EP can explain IDPs. If an EP can explain IDPs in more than 50% of 348 intervals ($p \geq 50\%$), then it is regarded as a strong EP. If p is in the range of between 0.3 and 0.5, the EPs are in the group of moderate EPs. For $p < 0.3$, these EPs are considered as weak EPs.

Table 5-3 The strength of explanation power of EPs

p	Interpretation
[0.00, 0.30)	Weak
[0.30, 0.50)	Moderate
[0.50, 1.0]	Strong

The first group of EPs, called strong EPs, includes BORROW (Total Borrowings of Depository Institutions from the Federal Reserve) and CPILFESL (Consumer Price Index for All Urban Consumers: All Items Less Food & Energy). Both predictors have significant influence on prediction of IDPs in more than half of all intervals.

The second group of EPs is named moderate EPs which covers 22 EPs in different categories of macroeconomic and financial factors. Specifically, there are: (1) 3 monetary policy indicators including AMBSL (St. Louis Adjusted Monetary Base), GPSAVE (Gross Private Saving), and MICH (University of Michigan: Inflation Expectation); (2) 5 general market indicators, such as IPUTIL (Industrial Production: Electric and Gas Utilities), UMCSENT (Consumer Sentiment), GPDI (Gross Private Domestic Investment), USPHCI (Coincident Economic Activity Index), and W068RCQ027SBEA (Government Total Expenditure); (3) 3 EPs from banking lending conditions including TOTALNS (Total Consumer Credit Owned and Securitized, Outstanding), REQERSNS (Required Reserves of Depository Institutions), and REALLNNSA (Total real estate loans); (4) 2 Business cycle leading/coinciding indicators including M12MTVUSM227NFWA (Moving 12-Month Total Vehicle Miles Traveled) and ALTSALES (Light Weight Vehicle Sales: Autos and Light Trucks); (5) 2 Balance of Payments indicators including IMPGSC96 (Real Imports of Goods and Services) and EXPGSC1 (Real Exports of Goods & Services); (6) 2 labour market conditions indicators including MANEMP (All Employees: Manufacturing) and CE16OV (Civilian employment); (7) the other 5 EPs are related to International competiveness – TWEXMMTH (Trade Weighted U.S. Dollar Index: Major Currencies), Cost of capital – SPREAD.MOODY.1 (Spread between Moody's Seasoned Baa Corporate Bond Yield and Moody's Seasoned Aaa – Corporate Bond Yield), Cost of labour – ULCNFB (Nonfarm Business Sector: Unit Labour Cost), Firm profitability – CNCF (Corporate net cash flow). The other EPs that can explain IDPs in less than 30% of all intervals belong to the group of weak EPs.

In particular, the majority key EPs are from both bank lending conditions and the monetary policy indicators. It is interesting that this issue of how monetary policies impact corporate default has been theoretically and empirically discussed in prior studies. This study provides evidence of how monetary policy transmissions are closely related to the credit role in the macroeconomy.

Table 5-4 Occurrence of key EPs

	Predictors			Category
	Abbreviation	Frequency %	Full name	
1	BORROW	54%	Total Borrowings of Depository Institutions from the Federal Reserve	Bank lending condition
2	CPILFESL	53%	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy	Monetary policy indicator
3	NAPMII	48%	ISM Manufacturing: Inventories Index	Business cycle leading indicator
4	GPSAVE	47%	Gross Private Saving	Monetary policy indicator
5	AMBSL	45%	St. Louis Adjusted Monetary Base	Monetary policy indicator
6	M12MTVUSM227NFWA	41%	Moving 12-Month Total Vehicle Miles Traveled	Business cycle leading indicator
7	IPUTIL	41%	Industrial Production: Electric and Gas Utilities	General market indicator
8	IMPGSC96	40%	Real Imports of Goods and Services	Balance of payments
9	MANEMP	39%	All Employees: Manufacturing	Labour market conditions
10	UMCSENT	38%	Consumer Sentiment	General market indicator
11	TOTALNS	38%	Total Consumer Credit Owned and Securitized, Outstanding	Bank lending condition
12	MICH	37%	University of Michigan: Inflation Expectation	Monetary policy indicator
13	SPREAD.MOODY.1	37%	Spread between Moody's Seasoned Baa Corporate Bond Yield and Moody's Seasoned Aaa -Corporate Bond Yield	Cost of capital
14	TWEXMMTH	37%	Trade Weighted U.S. Dollar Index: Major Currencies	International competiveness
15	EXPGSC1	36%	Real Exports of Goods & Services	Balance of payments
16	REQRESNS	35%	Required Reserves of Depository Institutions	Bank lending condition
17	ULCNFB	35%	Nonfarm Business Sector: Unit Labour Cost	Cost of labour
18	CE16OV	34%	Civilian employment	Labour market conditions
19	GPDI	33%	Gross Private Domestic Investment	General market indicator

20	ALTSALES	32%	Light Weight Vehicle Sales: Autos and Light Trucks	Business cycle leading indicator
21	CNCF	32%	Corporate net cash flow	Firm profitability
22	USPHCI	32%	Coincident Economic Activity Index	General market indicator
23	REALLNNSA	31%	Total real estate loans	Bank lending condition
24	W068RCQ027SBEA	30%	Government Total Expenditure	General market indicator

Note: The total sample is from April 1980 to December 2014, which contatins 417 observations. I split them into 357 intervals by using rolling window. There is 5-year data in the window with 60 observations, and LASSO is used to extract a set of EPs in the window. I move the first window from April 1980 to March 1985 by a month. The second interval from May 1980 to April 1985 is created. This procedure is iterated until the window ends in January 2010. The final window is then from January 2010 to December 2014. Based on the 357 intervals, I eventually get 357 groups of EPs by using LASSO. Frequency % means that the percentage of frequency of occurrence in the 357 groups of EPs. If the number is more than 50%, it means that this predictor can explain more than half of sample period. For example, an EP entitled CPILFESL occurs in 189 out of 357 groups of EPs, and the frequency for this EP is 189. Then the value of Frequency % is calculated by 189 divided by 357, which is 53%. The complete results is shown in Table 5-5. In Appendix 9.1, Table 9-1 shows the precedure of how to calcuate frequency and its % in detail.

Table 5-5 Complete results of EPs in Table 5-4

Predictors	Frequency	Frequency%	Predictors	Frequency	Frequency%
BORROW	191	54.00%	IPMANSICS	60	17.00%
CPILFESL	188	53.00%	M2SL	60	17.00%
NAPMII	170	48.00%	BOPBM	58	16.00%
GPSAVE	167	47.00%	AAA	57	16.00%
AMBSL	159	45.00%	CONSUMER	56	16.00%
M12MTVUSM227NFWA	148	41.00%	LNU03025703	56	16.00%
IPUTIL	146	41.00%	PPICEM	55	15.00%
IMPGSC96	144	40.00%	GDP	54	15.00%
MANEMP	141	39.00%	IPMAT	54	15.00%
UMCSENT	136	38.00%	TWEXBMTH	51	14.00%
TOTALNS	134	38.00%	CPIENGSL	48	13.00%
MICH	131	37.00%	LOANS	47	13.00%
SPREAD.MOODY.1	131	37.00%	PCEDG	46	13.00%
TWEXMMTH	131	37.00%	PPIFGS	46	13.00%
EXPGSC1	129	36.00%	A001RP1Q027SBEA	45	13.00%
REQRESNS	126	35.00%	PI	45	13.00%
ULCNFB	124	35.00%	MPRIME	44	12.00%
CE16OV	121	34.00%	BAA	43	12.00%
GPDI	117	33.00%	INDPRO	43	12.00%
ALTSALES	116	32.00%	EMRATIO	40	11.00%
CNCF	114	32.00%	A191RL1Q225SBEA	36	10.00%
USPHCI	113	32.00%	TB3MS.v	33	9.00%
REALLNNSA	112	31.00%	NAPM	31	9.00%
W068RCQ027SBEA	108	30.00%	BUSLOANSNSA	30	8.00%
RECPROUSM156N	105	29.00%	GS10	28	8.00%
GDPDEF	102	29.00%	BAMLCC0A0CMTRIV	27	8.00%
HOUST	99	28.00%	SPREAD.GS	25	7.00%
UEMP15OV	98	27.00%	FODSP	24	7.00%
IPMINE	94	26.00%	MCUMFN	24	7.00%
PSAVERT	93	26.00%	TDSP	24	7.00%
IMFSL	92	26.00%	UNRATE	24	7.00%
A067RL1Q156SBEA	89	25.00%	GS1.v	23	6.00%
AAA10YM	70	20.00%	FSDP	21	6.00%
BAA10YM	70	20.00%	PPIIEG	21	6.00%
PCE	70	20.00%	TCU	19	5.00%
SP500.v	70	20.00%	PPIACO	17	5.00%
HOHWMN02USM065S	67	19.00%	FEDFUNDS	14	4.00%
PPIITM	66	18.00%	PPIIDC	13	4.00%
HSN1F	64	18.00%	MORTG	10	3.00%
PERMIT	64	18.00%	TB3MS	9	3.00%
AAA.v	62	17.00%	GS1	1	0.00%
GS10.v	62	17.00%			

The total sample is from April 1980 to December 2014, which contatins 417 observations. I split them into 357 intervals by using rolling window. There is 5-year data in the window with 60 observations, and LASSO is used to extract a set of EPs in the window. I move the first window from April 1980 to March 1985 by a month. The second interval from May 1980 to April 1985 is created. This procedure is iterated until the window ends in January 2010. The final window is then from January 2010 to December 2014. Based on the 357 intervals, I eventually get 357 groups of EPs by using LASSO. Frequency % means that the percentage of frequency of occurrence in the 357 groups of EPs. If the number is more than 50%, it means that this predictor can explain more than half of sample period. For example, an EP entitled CPILFESL occurs in 189 out of 357 groups of EPs, and the frequency for this EP is 189. Then the value of Frequency % is calculated by 189 divided by 357, which is 53%.

(1) Explanation of the dynamic of EPs

The phenomenon of no explanatory variables that can strongly and continually explain IDPs means that default risk is not constant. With the change in default risk, the set of EPs also changes. This result corresponds to theoretical and empirical findings. Bernanke (1993) explains this phenomenon from different perspectives. Firstly, he proposes that the mix of default risk is not constant over time. For example, even the central bank tightens monetary policy by increasing interest rate; however, the number of bank loans may increase in this case. Since the true price of a bank loan is multidimensional, involving not only the contractual interest rate but a variety of other terms and conditions (such as covenants, collateral requirements and so on), then default risk can still increase. Actually, this example reflects the idea of a "flight to quality" within tight monetary policy period (Lang and Nakamura, 1995, Gertler and Gilchrist, 1991). Secondly, Bernanke (1993) argues that the predictive power for the predictors should reflect the market's perception of default risk. It is suggested that when the market's perception to default risk changes, the predictor should change as well. He states that the current financial market changes rapidly because of financial innovation, deregulation, and new forms of financial competition. Bernanke further clarifies that it is obviously easier to find good indicators retrospectively than prospectively; that is, the predictor that worked well in a previous specific period may not work well in the current period. He uses several cases in his paper in order to prove this point. He argues that although many studies, such as Romer et al. (1990) and Ramey (1993), find that M2 velocity has the strongest forecasting power of real economic activity, it is not sure to determine M2 can still keep this predictive ability since previously nobody would know in advance that M1 velocity would collapse.

The dynamics of default risk mentioned above could be explained by the interaction among complex behaviours in the credit market. According to Duffie and Singleton (2012), default risk acts through most of the same channels followed by other forms of market risk and also operates via other mechanisms. The channels can be reflected by the circumstance of adverse selection, risk concentration, and moral hazard in the imperfect market. The case of "flight to quality" may happen because of adverse selection. In reality, the business lenders as banks cannot know more than the borrowers in terms of their default risk. For example, the bank

makes profit in the tight period to attempt to compensate for default risk by considering whether the borrowers can provide high quality of collateral. However, the collateral provided by borrowers may not correspond to the real value shown in the contract. In this situation, the bank allows borrowers to select the sizes of their own loans without restriction.

In the case of risk concentration, the default clustering occurring in the credit market can be understood. Specifically, broker-dealers in the banks and OTC measure and limit the default risk, relying on two standards including individual counterparties and industry groups, geographic regions and sometimes other classifications (Duffie and Singleton, 2012). Assuming that various banks have different knowledge about the credit risk in the real estate industry, if some banks set a naïve rate depending on their own best estimates of the expected rate of losses owing to default compared with the other banks setting a high rate, many borrowers prefer to choose banks offering attractive rates even if these banks have no private information about their own credit quality (Duffie and Singleton, 2012). In this context, a large amount of loans from these banks flow to unqualified borrowers, which suggests that these banks will therefore suffer an expected loss on their real estate loans. Default cluster can then happen in the banking industry if there is a high default happening among borrowers in a recession period.

Moral hazard also exists in the credit market. Duffie and Singleton (2012) declare that since small loans are less risky than large loans, other things being equal, borrowers borrowing large loans have strong motivation to undertake riskier behaviour than borrowers borrowing small loans. This case can be reflected by the US savings-and-loan crisis which occurred in the 1980s. Federal deposit insurance provides extensive credit to savings-and-loan institutions without enforcing sufficient limits on the riskiness of savings-and-loan investments. Many savings-and-loan owners are motivated to engage in highly levered and risky portfolios of long-term loans, mortgage-backed securities, and other risky assets. It is clear that if these investments are successful, the owners can gain a lot; otherwise, they can lose a lot. Additionally, if large borrowers default, they often have a stronger bargaining power in order to obtain more preferable terms in bankruptcy or restructuring than small borrowers. This phenomenon could lead to the moral hazard problem among borrowers.

In fact, these different channels are generated by various and complex behaviour among participants, such as borrowers and lenders in the imperfect credit market. Fortunately, a new and influential theory called adaptive market hypothesis (AMH), proposed by Lo (2004), explains these phenomena by using evolutionary theory. More specifically, AMH reconciles market efficiency with behavioural alternatives by using the principles of evolution including competition, adaptation, and natural selection to explain financial interactions. For example, AMH attempts to handle the weakness of EMH in terms of failing to understand irrational behaviours in the financial market. Labelling 'irrational behaviour' is not appropriate and AMH argues that this behaviour should be recognized as 'maladaptive'. It is interesting that these channels often happen in the imperfect market with less market efficiency rather than in the perfect market, which means that they are generated by irrational behaviour in behaviour finance and maladaptive behaviour in AMH.

Consequently, one of the most important implications of AMH in the credit market is that default risk is not constant through time but varies according to the recent path of the credit market and the demographics of borrowers and lenders over that path. Based on AMH, another possible explanation for why there is a dynamic of EPs over time is that credit market is not stable but dynamic. Different behaviours are from the individuals' heuristics responding to changes in environment. The heuristics can generate approximately optimal solutions in stable environments. However, when the environment changes, the heuristics of the old environment are not necessarily suited to the new (Lo, 2004). The maladaptive behaviour is regarded as suboptimal behaviour, which fails to generate optimal solutions in the new environment. Thus, the evolution reflecting the change and complex behaviour in response to the change of environment suggests that the credit market is dynamic and EPs are not stable.

(2) Financial interpretation of EPs

a. Banking Lending Conditions

Three are 4 EPs which belong to bank lending conditions: BORROW (Total Borrowings of Depository Institutions from the Federal Reserve), TOTALNS (Total Consumer Credit Owned and Securitized, Outstanding), REQERSNS (Required Reserves of Depository

Institutions), and REALLNNSA (Total Real Estate Loans). In Table 5-4 they can explain IDPs in the 54%, 38%, 35%, and 31% of all the intervals, respectively. BORROW is the most significant predictor compared with the others, indicating that the ability of borrowing from the central bank is closely related to default risk in industrial firms from 1980 to 2014.

These EPs are all relevant to the size of overall lending in the bank lending conditions, which are monetary tools used by central bank for managing the money supply in the US market. In particular, REQERSNS as the reserve requirements cannot directly influence the liquidity of money in the market. Central bank increases this value thereby changing the amount of funds available for banks to make loans with, which has further impact on the excess reserves for the deposit institutions.

The condition of the credit loan market can be a sign of whether future operation cost goes up or not, which results in the deterioration of corporate defaults. Suppose the central bank tightens the monetary policy by either increasing REQERSNS or decreasing BORROW thereby impacting the willingness of commercial banks releasing loans to the market, and both TOTALNS and REALLNNSA will reduce later following this trend. As discussed earlier, there is an increasing trend in REQERSNS and a decreasing trend in BORROW; commercial banks may still have a tendency to lend money to borrowers following the idea of "flight to quality". Since these depositary institutions should obtain profits and reduce costs in tightened periods, then they are highly likely to lend to borrowers that have good high quality of collateral. In this context, both TOTALNS and REALLNNSA can still increase, which may lead to potential corporate defaults.

Then the market suffers from an illiquidity problem. If domestic credit markets are imperfect, firms may face credit constraints or higher financing costs as creditors raise interest rates to compensate for the increased risk of default. In addition, as the net worth of firms falls, so does the quality of the portfolios of their domestic creditors. Banks and financial intermediaries may be forced to reduce their exposure by cutting their loans – or they may simply go bankrupt (Serven and Solimano, 1992). Schwaab et al. (2015) suppose that banking lending conditions are strongly related to the decoupling of systematic default risk from macro fundamentals.

Figure 5-4 shows how these four specific predictors from banking lending conditions and IMFSL dynamically explain IDPs over time. Interestingly, there is a clear trend that BORROW gradually loses explanatory power on IDPs after the 2001 recession. Before this turning point, it has quite a highly explanatory power on IDPs. REQERSNS only has higher explanatory ability in the period from the end of the 2001 recession to before the beginning of the 2007 recession. In addition, the trend of TOTALNS is opposite to REALLNNSA. As mentioned earlier, the number of loans may not be related to macroeconomic conditions. Even in poor situations, the issued loans and credits are also quite high. They are not really comparable in this case as BORROW and REQERSNS.

In contrast, IMFSL (Institutional Money Funds) as the main money market funds (MMFs) has experienced a different trend in comparison with BORROW and REQERSNS. Specfically, both BORROW and REQERSNS have significantly explanatory power on corporate defaults in the historical recession periods except the one occuring in 2007. IMFSL can only have significant impact on corporate default in two recessions occuring in 1990 and 2007. Particularly, in the 2007 recession period, this predictor reaches the highest level with a value of around 80%. That is, there are about 34 out of 42 intervals where this predictor can highly explain IDPs. However, this is not the case for the lending conditions indicators. This suggests that IMFSL gradually plays a significant role in impacting corporate defaults.

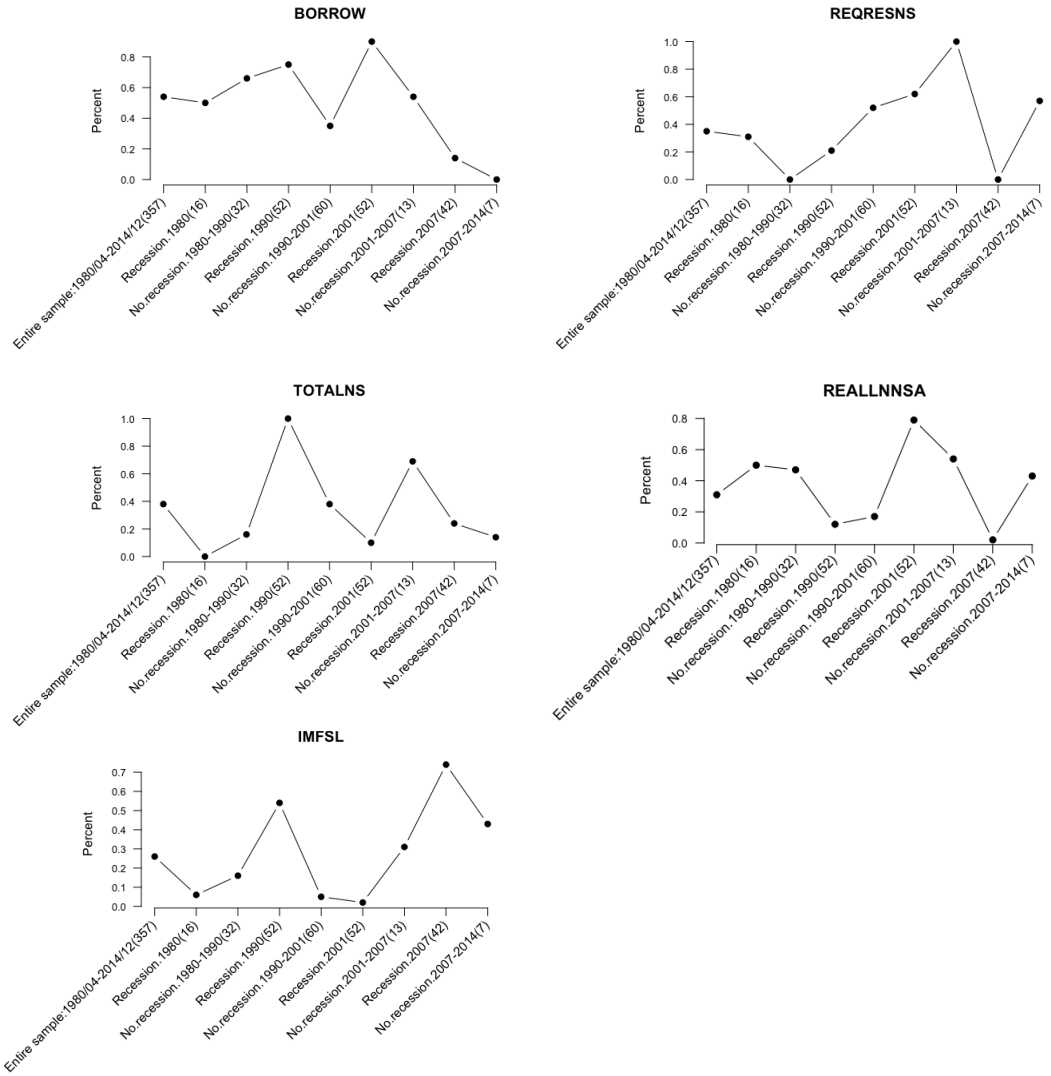
This phenomeon may be explained as follows. Specifically, the financial innovation such as MMFs generates new credit channels. Bernanke (1993) determines that both money market mutual funds and bond funds that cannot be controlled by the Federal Reserve are able to provide liquidity to the economy. MMFs as the supplemented liquidity provider supply payroll and inventory financing to a lot of businesses which can reduce the influence of conventional monetary policy tools, such as required reserve, quantitative easing, and interest rate on excess reserve balances. Sellon and Weiner (1996) suggest that many new financial instruments, such as development of certificates of deposit (CDs), Eurodollar borrowing, repurchase agreements, and ‘sweep accounts’ have been used to avoid money control from changing reserve requirements and borrowing rates. Firms can thus avoid defaults by relying on these new financial instruments to handle the problem of illiquidity.

Meanwhile, this finding is consistent with recent argument of whether MMFs play a critical role in the amplification of financial problems in recent years. Rosengren, as the president and Chief Executive Officer in Federal Reserve Bank of Boston, gave a speech titled *Money Market Mutual Funds and Financial Stability* at the Federal Reserve Bank of Atlanta's 2012 Financial Markets Conference in 2012. He argues that "The amount of credit risk taken by some money market funds was highlighted during the financial crisis by the loss associated with the $785 million of Lehman securities held by the Reserve Primary Fund." He claims that money market fund leads to high credit risk in the financial market.[28] FSOC (2012) reports that "The 2007–2008 financial crisis demonstrated that MMFs are susceptible to runs that can have destabilizing implications for financial markets and the economy." FSOC explains that MMFs' size, scale, and concentration result in the amplification of the transmission of risk of signficant liqudity and credit problems in the financial system. Therefore, this finding provides an emprical evidence of how MMFs have impact on the credit market particually in recent financial crisis.

[28] This finding may be also explained as follows. Money market fund can influence the effects of monetary policies. Thus, economic policy makers need to be concerned with how to achieve the policy execution objectives by allowing for the other elements, for example, now a sufficiently wide set of nonbank counterparties has access to the capital market against government collateral. Nowadays there is a discussion about whether central bank loses the control in the monetary market since the non-banking group provide a lot of liquidity in the market. For example, Simon Potter (Executive Vice President) gave a speech called Money Markets and Monetary Policy Normalization in 2015. He declared that policymakers should test ON RRP and use the other supplementary tools to maintain control over interest rates. He also stated that "An important determinant of the effectiveness of an ON RRP (overnight reverse repurchase agreement) facility is whether a sufficiently wide set of nonbank counterparties has access to it". Now the New York Fed executes RRPs with 163 counterparties comprising many of the major cash lenders in secured and unsecured money markets—including money market funds that manage about 70 percent of the total U.S. money market fund assets under management and account for more than a quarter of total tri-party repo lending against government collateral. This provides a high degree of confidence that we are transacting with an adequate set of counterparties to achieve our policy execution objectives.

Figure 5-4 Dynamic change of specific EP on explaining IDPs (1)

The total sample is from April 1980 to December 2014. 5-year as the size of each internal is used to extract a set of effective predictors. I use rolling window by a month based 1980/04on the unit of 5-year interval. Then there are 357 sub-intervals from the first interval from April 1980 to March 1985 to the last interval from January 2010 to December 2014. According to the classification of NBER recession periods, I split the 357 intervals into 9 samples. The first sample covers all the intervals. The second sample covers the 16 intervals that can fully cover the 1980s recession periods. The third sample has 32 intervals which do not cover any recession periods. The fourth sample includes 52 intervals which cover the 1990s recession period. The fifth sample has no recession period which has 160 intervals. The sixth sample contains 52 intervals with covering 2001s recession. The seventh sample covers no recession period with 13 intervals. The eighth sample includes 42 intervals which considering the 2007-2009 recession period. The last sample does not contain any recession period with 7 intervals inside. Then I track each predictor in explain default risk in each interval and count the frequency of them. BORROW is the Total Borrowings of Depository Institutions from the Federal Reserve. BORROW is total borrowings of depository institutions from the Federal Reserve. TOTALNS is total consumer credit owned and securitized, outstanding. REQERSNS is required reserves of depository institutions. REALLNNSA is total real estate loans. IMFSL is institutional money funds.



b. Monetary Policy Indicators

There are 4 EPs from monetary policy indicators: CPILFESL (Consumer Price Index for All Urban Consumers: All Items Less Food & Energy), AMBSL (St. Louis Adjusted Monetary

Base), GPSAVE (Gross Private Saving), and MICH (University of Michigan: Inflation Expectation). In Table 5-4 they can explain IDPs in the 53%, 47%, 45%, and 37% of all the intervals, respectively. Figure 5-5 shows how these monetary policy indicators have dynamic influence on IDPs over time. It is interesting that for the subsample with 56 intervals related to the 2001 economic recession, there is influence from neither CPIFESL nor GPSAVE. Actually, this economic crisis is not caused by monetary policy, and this finding corresponds to the intuition. However, the other crises in US history are highly related to monetary policies, which is reflected in this figure. Both variables have strong power of explaining corporate defaults in both the 1990s crisis and the subprime crisis of 2007-2009.

CPILFESL is the second most significant EP in predicting IDPs, which corresponds to the prior theoretical findings. For example, Bhamra et al. (2011) determine that a decrease in expected inflation, due to a monetary policy shock, increases the incentives of a corporation to default; that is, deflation can lead to companies going to default. Kang and Pflueger (2015) investigate the relation between corporate bond yields and deflation. They figure that unexpectedly deflation increases real liabilities and default risk. Fisher (1933) as the pioneer in this area initially proposes a "debt defilation" theory. It is suggested that when a surprise drop in inflation leads to increases in real liabilities and corporate default risk. In contrast, Wadhwani (1986) identifies that higher inflation implies higher liquidation rates and default premia, which means that higher inflation may lead to more bankruptcies. It should be noted that CPILFESL is the core inflation, which has been widely used by economists since this index considers the influence from the high volatility from energy and food. By contrast, my finding shows that change of CPIFESL fails to explain IDPs over time. It is indicated that although the consumer price index can provide a signal of future movements of default risk, this index does not have any predicting effects in certain periods. There is no serious inflation and deflation in these periods that consumer price cannot predict default risk, since the rapid change of consumer price often happens in the brunt of an economic downturn.

Early empirical studies rarely employ this indicator for predicting default risk. Current researchers are starting to use it; however, they mainly employ CPIAUCSL (Consumer Price Index for All Urban Consumers: All Items) to predict the default risk instead of core inflation. They find that this type of macro factor cannot provide predictive contents for corporate

defaults, such as Giesecke et al. (2011) who employ CPIAUCSL to predict the default risk and determine that this consumer index cannot predict default rates. Three possible reasons are shown below. The first reason is that they investigate the default risk in the bond market; the second is that core CPI is not employed for prediction; the final reason is that the consumer price index can only explain specific periods from our current findings.

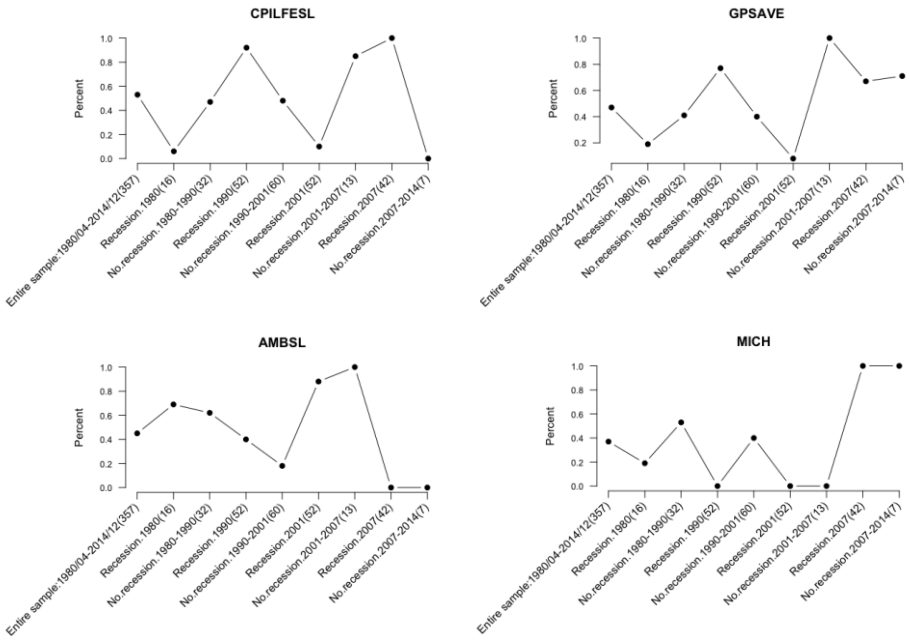
AMBSL measures the currency amount held outside the Treasury and the Federal Reserve Banks (referred to as currency in circulation) plus deposits held by depository institutions at the Federal Reserve Banks. The Federal Reserve controls the demand from the private sector for these liabilities to affect money market interest rates. In fact, this monetary policy acts in the same role as REQERSNS and BORlROW, and they are used for controlling the liquidity of money in the market. Then it is clear that AMBSL also fails to manipulate the behaviours of commercial banks to some extent.

Tightened monetary policies including increasing reserve ratios, reducing borrowing rate from central bank, and reducing monetary base are able to influence the size of overall lending in the credit market. Consequently, the liquidity of money becomes more volatile thereby increasing the uncertainties of financial distress. Once the tightened monetary policies are implemented, then the financial market becomes more illiquid for the firms. The banks reduce the credit to firms and firms face a higher default probability. This may explain why there is a lag effect for the default cycle compared with the business cycle, since the central bank often adjusts monetary policies after the economy starts experiencing economic recession. In addition, there is also a lag effect between the implementation of monetary policies and reaction to these policies from the market.

It can be concluded that the interaction between monetary policies and credit market stability should be given priority when decision makers plan to use monetary policies to stabilize the economy. Inappropriate monetary policies can result in a depressed credit market with high default events, and appropriate monetary policies can stabilize the credit market with rare default events. This is also the major concern of how to stabilize the economy with decision makers, in reality, concerned by both monetary policies and financial policies in the market.

Figure 5-5 Dynamic change of specific EP on explaining IDPs (2)

The total sample is from April 1980 to December 2014. 5-year as the size of each internal is used to extract a set of effective predictors. I use rolling window by a month based 1980/04on the unit of 5-year interval. Then there are 357 sub-intervals from the first interval from April 1980 to March 1985 to the last interval from January 2010 to December 2014. According to the classification of NBER recession periods, I split the 357 intervals into 9 samples. The first sample covers all the intervals. The second sample covers the 16 intervals that can fully cover the 1980s recession periods. The third sample has 32 intervals which do not cover any recession periods. The fourth sample includes 52 intervals which cover the 1990s recession period. The fifth sample has no recession period which has 160 intervals. The sixth sample contains 52 intervals with covering 2001s recession. The seventh sample covers no recession period with 13 intervals. The eighth sample includes 42 intervals which considering the 2007-2009 recession period. The last sample does not contain any recession period with 7 intervals inside. Then I track each predictor in explain default risk in each interval and count the frequency of them. CPILFESL is the consumer price index for all urban consumers: all Items less food & energy. AMBSL is the St. Louis adjusted monetary base. GPSAVE is gross private saving. MICH is the University of Michigan: inflation expectation.



c. Business Cycle Leading Indicators

Business cycle leading indicators as the third important type of predictors have weak performance of forecasting corporate default compared with bank lending conditions and monetary policy indicators. There are three business cycle leading/coinciding indicators: NAPMII (ISM Manufacturing: Inventories Index), M12MTVUSM227NFWA (Moving 12-Month Total Vehicle Miles Travelled) and ALTSALES (Light Weight Vehicle Sales: Autos and Light Trucks). They can explain IDPs in 48%, 41%, and 32% of all the intervals, respectively in Table 5-4. Figure 5-6 provides detailed information about dynamic impact of each EP on IDPs. The trends of M12MTVUSM227NFWA and ALTSALES are similar. This finding corresponds to intuition since the increase in car sales is dependent on vehicle miles travelled.

Basically, three predictors belong to leading indicators following the definition of timing classification of indicators. Leading indicators can be used to forecast whether the business cycle is going to occur or not. They tend to move earlier than the other cyclical indicators. In addition, prior studies have identified that default clustering tends to follow the business cycle (Couderc et al., 2008, Azizpour et al., 2014). It is clear then that leading indicators have the ability to forecast corporate defaults.

NAPMII, as the most predictive among three predictors, is an index based on surveys of more than 300 manufacturing firms by the Institute of Supply Management, which monitors production inventories. This finding can be explained from both theoretical and empirical perspectives. Previous studies identify that this index has a measurable impact on the default risk of any given firm. According to Richardson et al. (1998), theoretically, suppose there is a shock due to economic recessions, which means that the business activity experiences a sharp decrease in the short term. Meanwhile, the monetary and fiscal constraints are often coincident with the occurrence of recessions. There is then a rapid increase in the cost of capital and a decrease in the availability of credit. The decrease in sales and rising costs in this period can jointly result in reduced profitability or even loss for many businesses and industries, which is accompanied by increase of aggregate business inventories and capital spending outlays. It is clear then that inventory can be a sign predicting default risk for firms.

Default clustering always appears in the recession period and higher defaults seemingly has a lag effect in comparison with recession periods. Previous studies identify that movements of inventory play a significant role in the fluctuations of the business cycle. Blinder and Maccini (1991) determine that the decline in inventory investment occupies an average of 87% of the total peak-to-trough movement in GNP in postwar US recessions. I can conclude that the movement of inventory may coincide with default cycle or even be ahead of default cycle.

However, this index cannot perform well over time, which may be understood as follows. Kashyap et al. (1992) state that recessions usually follow a period of tight credit. They further declare that a tightening monetary policy has great influence on inventory behaviour and this impact is transmitted through two cost-of-financing channels: internal and external finance. Since the firm has a bad position in the balance sheet of net worth, then internal sources

cannot be an effective and efficient tool for financing in the recession period. Consequently, the firm can only use external sources of funds for handling the illiquidity problem caused by higher inventories, lower sales, and reduced profitability. Richardson et al. (1998) conclude that both liquidity and the presence of adequate opportunities have been theoretically regarded as the significant elements in firm survival. Thus a firm that can successfully obtain external finance thereby reducing the illiquidity problem caused by high inventory can be free from default.

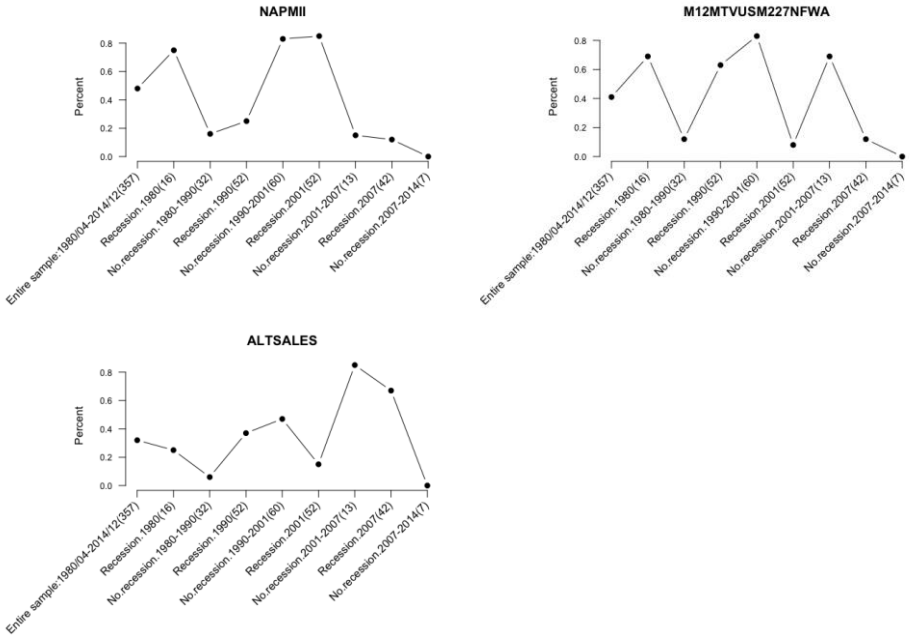
Gertler and Gilchrist (1991) investigate the behaviour of small and large manufacturing firms by collecting data from the Quarterly Financial Reports. Specifically, they find that the growth rate of inventories for large firms picks up slightly just prior to recessions, except in the last recession. Inventory growth for small firms declines steadily over recessionary periods, and it is generally at a faster pace for the larger firms. After examination of the inventory/sales ratio, they identify that large firms can easily and heavily borrow funds to smooth the impact of declining sales and carry the inventories compared with small firms. This phenomenon is actually reflected by the "flight to quality" theory (Bernanke et al., 1994). Since imperfect information exists in the credit market, lenders tend to provide short-term loans to large borrowers rather than small borrowers following an adverse macroeconomic shock. Therefore, inventory index is not always a good predictor for forecasting default risk in this case.

The next important leading indicator is vehicle miles travelled which is calculated by the Institute of Supply Management. This is used to measure vehicle activity in transportation. Former research has identified that transportation output has significant leading value to economy-wide business cycles. According to Lahiri (2010), transportation demand can reflect producers' expectation toward future profits. In addition, monetary policy tightening and oil price shocks remarkably impact on transportation output since this item highly relies on the usage of capital requirements and fuel consumption. It is clear then that if there is a shock caused by economic recession, this value decreases since the firm relying on transportation faces financial distress and fails to do this business. However, this predictor fails to provide a sign of default risk over time. Pozdena (2009) finds that the relationship between vehicle miles travelled on US public roads and the vigour of the economy becomes

particularly ambiguous. He figures that with an increasing amount of new technologies used in transportation and the identification of climate change, producers and consumers have an incentive, at all times, to reduce energy use for vehicles thereby disturbing that relationship. In fact, this is the same case for vehicles sold, ALTSALES. Interestingly, vehicles sold is less predictive for IDPs compared with vehicle miles travelled.

Figure 5-6 Dynamic change of specific EP on explaining IDPs (3)

The total sample is from April 1980 to December 2014. 5-year as the size of each internal is used to extract a set of effective predictors. I use rolling window by a month based 1980/04on the unit of 5-year interval. Then there are 357 sub-intervals from the first interval from April 1980 to March 1985 to the last interval from January 2010 to December 2014. According to the classification of NBER recession periods, I split the 357 intervals into 9 samples. The first sample covers all the intervals. The second sample covers the 16 intervals that can fully cover the 1980s recession periods. The third sample has 32 intervals which do not cover any recession periods. The fourth sample includes 52 intervals which cover the 1990s recession period. The fifth sample has no recession period which has 160 intervals. The sixth sample contains 52 intervals with covering 2001s recession. The seventh sample covers no recession period with 13 intervals. The eighth sample includes 42 intervals which considering the 2007-2009 recession period. The last sample does not contain any recession period with 7 intervals inside. Then I track each predictor in explain default risk in each interval and count the frequency of them. NAPMII is the ISM manufacturing: inventories index; M12MTVUSM227NFWA is the moving 12-Month Total Vehicle Miles Travelled, and ALTSALES is the light weight vehicle sales: autos and light trucks.



d. General market indicators

There are five EPs in this group: IPUTIL (Industrial Production: Electric and Gas Utilities), UMCSENT (Consumer Sentiment), GPDI (Gross Private Domestic Investment), USPHCI (Coincident Economic Activity Index), and W068RCQ027SBEA (Government Total Expenditure). Although there are more predictors from this group, the most predictive one, IPUTIL, can only explain 41% of all the intervals in Table 5-4. Figure 5-7 illustrates that each EP dynamically predicts IDPs over time.

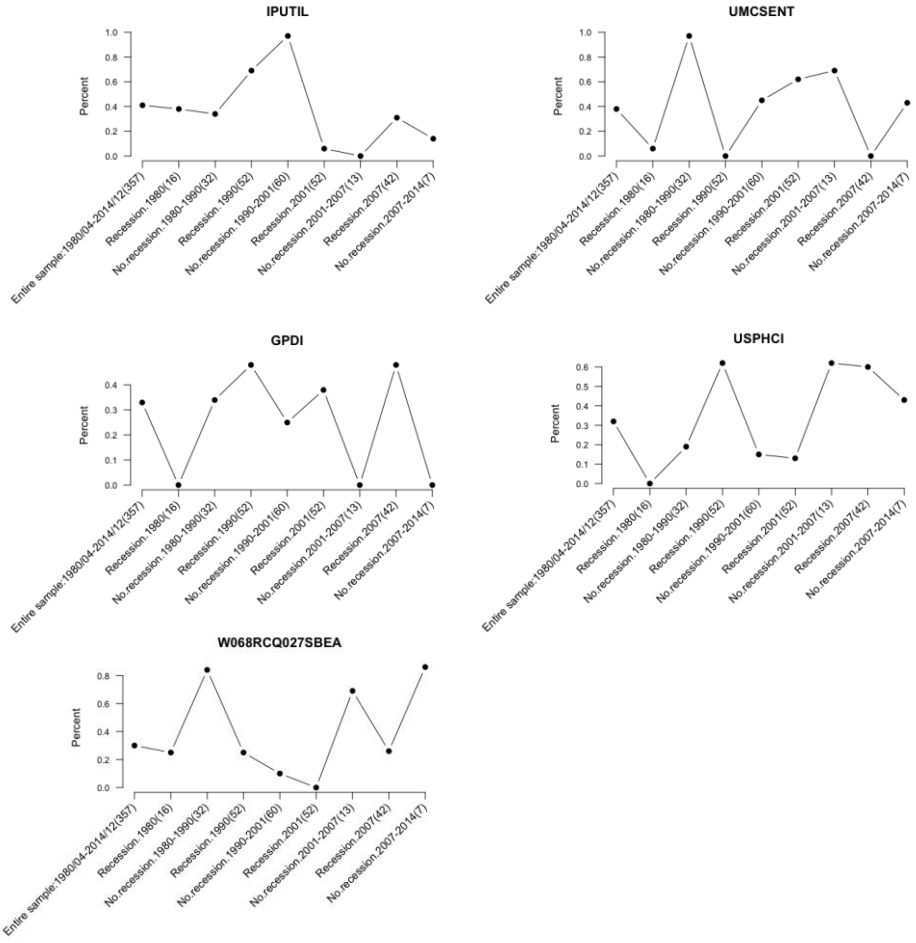
IPUTIL is related to industrial production in electric and gas utilities, which are regarded as gauges of business activity. This is intuitive as a high economic growth rate measured by these predictors leads to a high growth rate at the firm-level, which reduces the likelihood of default for the firm. This is the same case for the other predictors such as UMCSENT, which

is to measure consumer sentiment to the economic environment; GPDI gauges the amount of gross private domestic Investment; USPHCI is directly used to measure economic activity.

However, there is a special scenario that these predictors can misunderstand the normal recognition of default risk. As discussed previously, higher growth rate means a more active economic environment, which reduces the default risk among the industrial sectors. Undoubtedly, higher growth rate may also mean high and increased supply in the market, which may result in deflation (Borio et al., 2015). In this context, a firm faces high competition within the market with less demand from the customers. The reduced profit and increased problem of illiquidity can lead to higher default. Therefore, those predictors fail to capture higher defaults over time when the market changes randomly and dynamically.

Figure 5-7 Dynamic change of specific EP on explaining IDPs (4)

The total sample is from April 1980 to December 2014. 5-year as the size of each internal is used to extract a set of effective predictors. I use rolling window by a month based 1980/04on the unit of 5-year interval. Then there are 357 sub-intervals from the first interval from April 1980 to March 1985 to the last interval from January 2010 to December 2014. According to the classification of NBER recession periods, I split the 357 intervals into 9 samples. The first sample covers all the intervals. The second sample covers the 16 intervals that can fully cover the 1980s recession periods. The third sample has 32 intervals which do not cover any recession periods. The fourth sample includes 52 intervals which cover the 1990s recession period. The fifth sample has no recession period which has 160 intervals. The sixth sample contains 52 intervals with covering 2001s recession. The seventh sample covers no recession period with 13 intervals. The eighth sample includes 42 intervals which considering the 2007-2009 recession period. The last sample does not contain any recession period with 7 intervals inside. Then I track each predictor in explain default risk in each interval and count the frequency of them. IPUTIL is the industrial production: electric and gas utilities. UMCSENT is the University Michigan consumer sentiment. GPDI is the gross private domestic investment. USPHCI is the coincident economic activity index.



e. Labour Market Conditions

In Table 5-4, two moderate EPs, MANEMP and CE160V, are from the category of labour market conditions. The first is used for measuring all employees in manufacturing, and the second is to measure broader sectors. From a theoretical perspective, Agrawal and Matsa (2013) study the impact of labour unemployment risk on corporate financing in the US, and they find that there is a strong relation between unemployment risk and financial policy among firms facing tight financial constraints. They suppose that reductions in labour unemployment risk are associated with increases in corporate leverage and reducing leverage decreases the probability that a firm will encounter financial distress. Meanwhile, Jaggia and Thakor (1994) find that firms can reduce the probability of distress by taking less risky projects, or reduce workers' losses in distress by redesigning job tasks to require fewer firm-specific skills. Then, this labour cost can give the sign of default risk. However, in my finding, these type of predictors also cannot have a consistent predictive power for corporate defaults. Figure 5-8 provides evidence of how both EPs explain IDPs over time. Interestingly, they have different trends, indicating that the degree of impact on IDPs for each predictor is also different. For the dynamic influence on IDPs, it may be explained as technogical improvement reducing the number of employees in the maunufacturing sector or broad sectors. Consequently, firms with fewer employees are still healthy without any defaults.

f. Cost of Capital

Spread ratio, such as SPREAD.MOODY.1 (Spread between Moody's Seasoned Baa Corporate Bond Yield and Moody's Seasoned Aaa-Corporate Bond Yield), has been widely used by previous studies for anticipating corporate defaults. Jorion and Zhang (2007) and Collin-Dufresne et al. (2003) find that there is a close relationship between default events and trivial increase of credit spread. Koopman et al. (2009) study the relation between macroeconomic fundamentals and credit cycles by employing various databases including macroeconomic data from FRED and corporate rating transition and default data from Standard and Poor's US from 1980 to 2005. They find that credit spread is an important determinant of default and rating migration intensities. However, several studies also find

that credit spread fails to predict default risk. Giesecke et al. (2011) study and predict the industry level of default probabilities in the US corporate bond market by using yearly data from 1866 to 2008. In using a regime-switching framework, they find that changes in credit spreads do not appear to be related to subsequent default rates. They explain that changes in credit and liquidity risk premia mainly drive the changes in the time variation in credit spreads and changes in the default risk can only marginally drive the changes in credit spreads, which is consistent with the results obtained by Bai et al. (2012). It is clear that the influence of credit spread on default risk is not constant by dynamic. This dynamic pheonomenon is shown in Figure 5-8. This predictor is only able to explain 37% of all the intervals and explain around 40% of the interals in the 8 subsamples except the sixth and eighth subsamples (that do not cover recession periods).

g. International Competiveness and Balance of Payments

This section is to explain three moderate EPs from the group of international competiveness and balance of payments. Specifically, the first group has TWEXMMTH which can explain 37% of all the intervals. In Table 5-4, the second group has IMPGSC96 and EXPGSC1 which can explain 40% and 36% of all the intervals, respectively. Since both groups of factors can give a picture of how international trade has impact on IDPs in the US, then this study discusses them as follows.

TWEXMMTH is used for measuring the effects of dollar appreciation and depreciation against foreign currencies on the competitiveness of the US products relative to goods. This index is calculated by 7 major currencies trading widely in currency markets outside their respective home areas. The changes in this index have influence on US exporters and importers. Suppose the US dollar appreciates rapidly. It would be beneficial for importers to reduce the cost of production and increase profits since the price of imported goods decreases. Meanwhile, appreciation of currency can attract more capital flowing from outside to inside, which brings a lot of liquidity into the domestic market. By contrast, exporters may face a big loss in foreign trade as the price goes up and the competition of price decreases. Therefore the probabilities of default for these firms increase. The two aspects are reflected by changes in IMPGSC96 and EXPGSC1 since the former is to measure the amount of imports thereby

showing whether the importers face potential financial distress, and the latter is to measure the amount of exports thereby displaying whether exporters face unwanted financial distress. Figure 5-8 shows how these EPs explain IDPs over time. Both IMPGSC96 and EXPGSC1 have stronger predictive power for IDPs compared with TWEXMMTH. In particular, imports have more impact on the period from the beginning of 1990 to the end of 2001. Exports are still able to impact IDP in the longer period that ends at the beginning of 2007. TWEXMMTH has a very stable and poor performance in the 8 groups of sub-intervals.

The 1997 Asian crisis provides a good example for explaining how these predictors cannot consistently influence IDPs. In Figure 5-1, there is a month with higher defaults in January 1998 and the black line as the IDPs reaches the peak level of 0.09. It should be noted that these high defaults only last one month and no corporate default clustering occurs in consecutive months. This result can be interpreted by the unwanted changes of exchange rates among these East Asia countries. Since the Asian economic and financial crisis has attracted attention to the trade links among the US and countries throughout Asia, trade links are reflected by the exchange rates. Pollard and Coughlin (1999) figure that "as the effects on Asian economies became more prounced and the effects of crisis spread to other countries in spring and summer 1998, changes in US trade became more visible and concerns about the possiblity of a US recession increased."[29] However, the US economy had not suffered from the influence of the rapid decrease of exports and its economy was still robust in 1998. They determine that although the commodity prices decrease after a result of declining demand in Asia, the growth of US economy obtains support from the expansion of domestic consumption, investment spending, and decreasing interest rate.[30] In addition, numerous

[29] The effects of the Asian financial crisis on U.S. economy can be explained from three perspectives. The first is that the growth of U.S. economy has been derived by the sharp increase of the export sector since the mid-1980s. For example, from 1992 to 1997, growth in exports accounted for more than 30 percent of the growth in real GDP; secondly, during that period, the Asia-Pacific area has become the key important importers, which received 31 percent of U.S. exports in comparison with only 22 percent recorded from 1970 to 1975. In contrast, the pattern of U.S. exports has been decrease importance of the European market and Europe only received 25 percent of U.S. exports in comparison with the record of 32 percent from 1970 to 1975. Finally, it should be noted that Japan is not the key contributor in the Asian-Pacific area but the rest of East Asia. Therefore, it is clear that East Asia market plays an increasing significance of role in terms of driving U.S. economy expansion (Pollard and Coughlin, 1999).

[30] This is partly caused by the flight-to-quality capital inflows and the Fedral Reserve's easing of monetary policy.

studies identify that the Asian crisis only has an impact on certain economic sectors (Noland, 1998, Werling and McCarthy, 1998). For instance, Noland (1998) states that "the effect of the decline in exports on the output of a particular industry depends on the extent to which that industry is dependent on the East Asian market to sell its output". Consequently, there is no obvious default clustering, which also reflects why the NBER's Business Cycle Dating Committee did not regard it as a recession. Therefore, it can be seen that exchange rate risk has a certain impact on default risk. Many elements are able to result in the degree of influence of exchange risk on default risk, such as trade links between US and other trading partners, industrial structure, interaction between monetary policy and exchange rate policy, hedging policy etc.[31] Additionally, exchange risk can influence both exports and imports. Therefore, three predictors are not able to predict the default risk constantly over time.

h. Firm Profitability

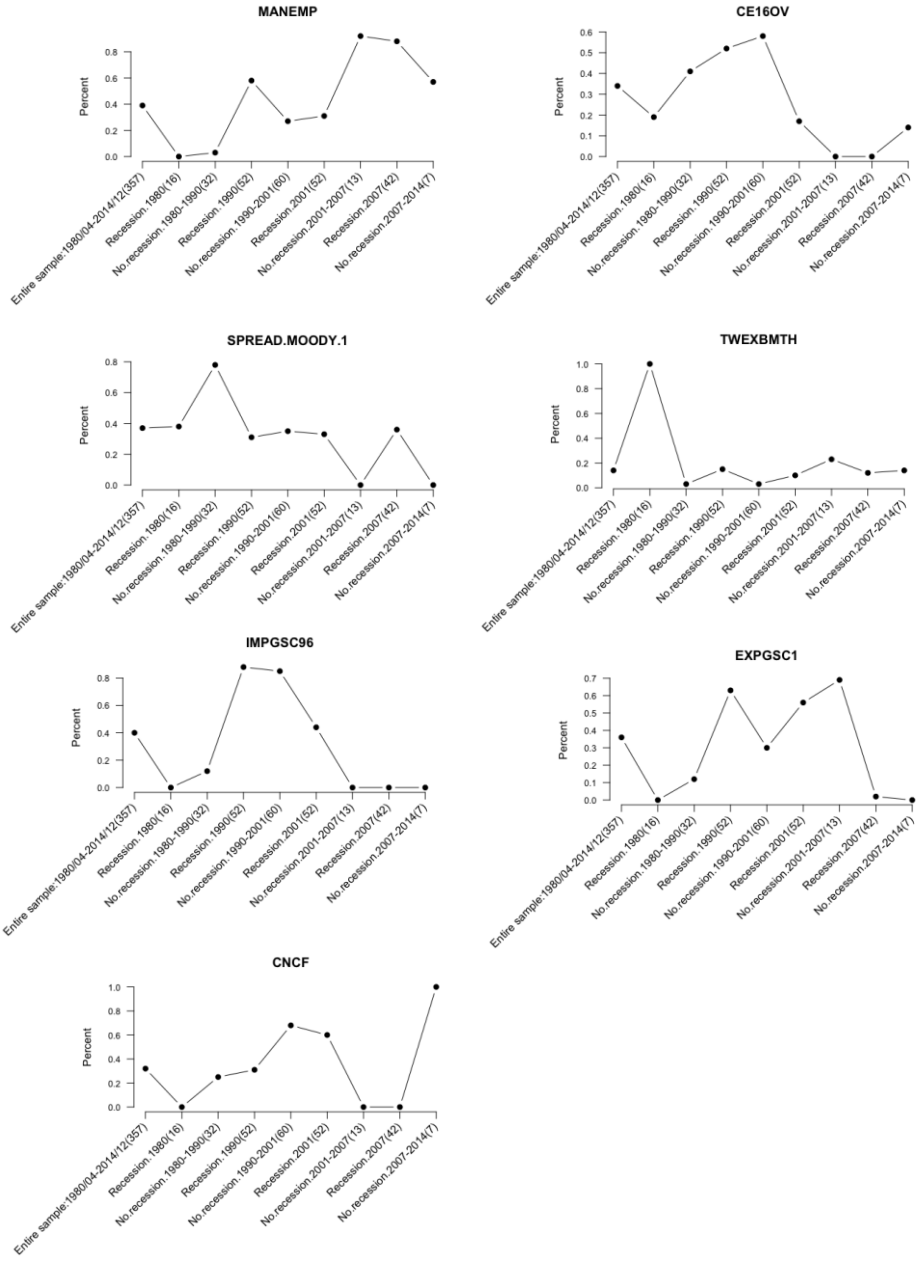
The moderate EP, CNCF, is related to firm profitability, which is short for corporate net cash flow. In the 32 percent of all the intervals in Table 5-4, IDPs can only be explained by this cash flow predictor. It is intuitive that the firms having larger cash flow in their asset and investment portfolio should be free from the default. Firms with a rich of cash holdings have a lower probability of default. For this type of puzzle finding, a significant study done by Acharya et al. (2012) provides evidence for understanding the story behind this phenomenon. They focus on the firms that issue public bonds and use various data resources, such as public bonds from the Fixed Income Securities Database (FISD), default data from Moody's DRD, financial data from Compustat, and executive compensation data from ExecuComp etc. from December 1996 to September 2010. Their study empirically explores the role of liquid assets in default-predicting models including z-score developed by Altman (1968), O-score created by Ohlson (1980), the three other reduced form models from Zmijewski (1984), Shumway (2001), and Chava and Jarrow (2004). They identify that the correlation between liquidity and defaults relies on the time horizon over which default is being measured. More

[31] Many studies such as Burnside et al. (2001) and Eichengreen and Hausmann (1999) find that using the hedging policies for reducing the exchange rate risk can lead to the lower probability of bankruptcy.

specifically, the short-term probability of default is lower when there is a large amount of liquid asset reserves. However, higher cash holdings increase the long-term probability of default based on horizons longer than one year. These empirical results are consistent with the theoretical findings after constructing a theoretical model to show this phenomenon based on the endogeneity of a levered firm's cash policy. As their paper explains, it is reflected by the precautionary motive for saving cash; that is, the riskier firms accumulate higher cash reserves thereby avoiding future defaults. Figure 5-8 displays how CNCF explain IDPs over time. Therefore, this predictor cannot always predict default risk correctly since the financial frictions are dynamic.

Figure 5-8 Dynamic change of specific EP on explaining IDPs (5)

The total sample is from April 1980 to December 2014. 5-year as the size of each internal is used to extract a set of effective predictors. I use rolling window by a month based 1980/04on the unit of 5-year interval. Then there are 357 sub-intervals from the first interval from April 1980 to March 1985 to the last interval from January 2010 to December 2014. According to the classification of NBER recession periods, I split the 357 intervals into 9 samples. The first sample covers all the intervals. The second sample covers the 16 intervals that can fully cover the 1980s recession periods. The third sample has 32 intervals which do not cover any recession periods. The fourth sample includes 52 intervals which cover the 1990s recession period. The fifth sample has no recession period which has 160 intervals. The sixth sample contains 52 intervals with covering 2001s recession. The seventh sample covers no recession period with 13 intervals. The eighth sample includes 42 intervals which considering the 2007-2009 recession period. The last sample does not contain any recession period with 7 intervals inside. Then I track each predictor in explain default risk in each interval and count the frequency of them. MANEMP is the all employees: manufacturing. CE160V is the civilian employment. SPREAD.MOODY.1 is the spread between Moody's seasoned Baa corporate bond yield and Moody's seasoned Aaa yield. TWEXBMTH is the trade weighted U.S. Dollar index: Broad. IMPGSC96 is the real imports of goods and services. EXPGSC1 is the real exports of goods & services. CNCF is the corporate net cash flow.



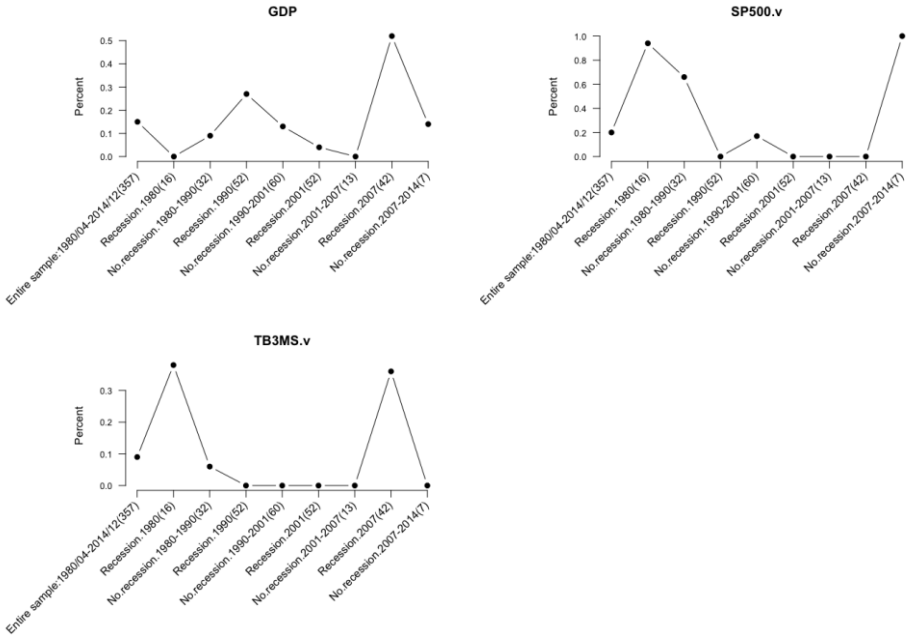
i. Other popular predictors used by prior studies

The rest of the EPs that are not listed in the group of key EPs are weaker for consistantly predicting default risk in the period from 1980 to 2014. In particular, three predictors including trailing 1-year return on the S&P500 (SP500.v), GDP growth (GDP), and 3-month Treasury rate (TB3MS.v) belong to this group. Prior work by Duffie et al. (2007), Das et al. (2007), Couderc et al. (2008), and Duffie et al. (2009) suggests that these macroeconomic factors have strong power in default prediction. For example, Das et al. (2007), Couderc et al. (2008), and Duffie et al. (2009) determine that GDP growth is a signficant exploratory variable of predicting default risk. In contrast, Duffie et al. (2007) find that GDP growth has no power to predict defaults. In this study, it is apparent that the performance of these factors can only predict default risk in a few intervals out of 357 intervals from 1980 to 2014. The implication of the current findings is that market conditions are dynamic in terms of predicting default risk. Changes of default risk can be caused by different econcomic environments. It would be cautious to select EPs for predicting the default risk since the market conditions always change over time.

Figure 5-9 illustrates how these EPs dynamically impact on IDPs. Basically, they are quite volatile and have poor performance in explaining IDPs. For several sub-samples, these predictors cannot even explain IDPs in any interval. For example, GDP has no any prediction effects on IDPs in 16 intervals that cover the 1990 recession. This is the same case in 13 intervals that consist of the period from 2001 to 2007. Likewise, both SP500 and 3 months T-bill rates are able to explain limited periods.

Figure 5-9 Dynamic change of specific EP on explaining IDPs (6)

The total sample is from April 1980 to December 2014. 5-year as the size of each internal is used to extract a set of effective predictors. I use rolling window by a month based 1980/04on the unit of 5-year interval. Then there are 357 sub-intervals from the first interval from April 1980 to March 1985 to the last interval from January 2010 to December 2014. According to the classification of NBER recession periods, I split the 357 intervals into 9 samples. The first sample covers all the intervals. The second sample covers the 16 intervals that can fully cover the 1980s recession periods. The third sample has 32 intervals which do not cover any recession periods. The fourth sample includes 52 intervals which cover the 1990s recession period. The fifth sample has no recession period which has 160 intervals. The sixth sample contains 52 intervals with covering 2001s recession. The seventh sample covers no recession period with 13 intervals. The eighth sample includes 42 intervals which considering the 2007-2009 recession period. The last sample does not contain any recession period with 7 intervals inside. Then I track each predictor in explain default risk in each interval and count the frequency of them. GDP is the growth domestic production; SP500.v is the return of SP500; TB3MS.v is the 3 months T-bill rates.



5.4 Robustness Testing

To perform robustness testing, two issues should be tested. One is related to the value of gamma (γ) for filtering time series of data; the other is associated with whether there is an influence from the method of constructing IDPs in the current results.

In this study, I use an HP filter with gamma (γ) being equal to 5. The value chosen for γ differs depending on the sampling frequency of the data and represents the trade-off between smoothness and goodness of fit in the resulting series. A smoothing constant of 100/1600/14400 for annual/quarterly/monthly data has been taken as a standard by many researchers in macroeconomics. I follow this rule to filter our data. Meanwhile, it should be noted that more γ I used, the more smoothness the data I have, which means that some

information hidden in the data may be deleted. Then I used 5, 10, 20, 30, and 50 as γ in this study. The motivation for selecting 5 is threefold. First, this study focuses on short-term fluctuation rather than medium-term or long-term fluctuation. Second, more γ means that more information is removed from the data. Some macro factors had a gentle trend of growth in the past so they are very sensitive to using γ. Third, when γ equals 5, the results of a set of effective predictors are stable even using different proportions of training sample. Thus this study uses 5 for smoothing the data.

In addition, I construct another time series of IDPs for the listed firms without concerning merging the microeconomic database. This is to test further whether the current results are influenced by adjusting the method of screening listed defaulters. Table 5-6 presents the results of key EPs. Compared with Table 5-4, they have the same EPs, even the order of EPs are almost the same. The only big difference between them is the value of frequency of each EP in terms of explaining corresponding IDPs. Therefore, the results are reasonable and credible in this study.

Table 5-6 Key EPs without merging microeconomic database as previous studies

	Predictors			Category
	Abbreviation	Frequency %	Full name	
1	BORROW	54%	Total Borrowings of Depository Institutions from the Federal Reserve	Bank lending condition
2	CPILFESL	52%	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy	Monetary policy indicator
3	NAPMII	47%	ISM Manufacturing: Inventories Index	Business cycle leading indicator
4	GPSAVE	47%	Gross Private Saving[32]	Monetary policy indicator
5	AMBSL	44%	St. Louis Adjusted Monetary Base	Monetary policy indicator
6	M12MTVUSM227NFWA	42%	Moving 12-Month Total Vehicle Miles Traveled	Business cycle leading indicator
7	IPUTIL	41%	Industrial Production: Electric and Gas Utilities	General market indicator
8	IMPGSC96	40%	Real Imports of Goods and Services	Balance of Payments
9	MANEMP	40%	All Employees: Manufacturing	Labour market conditions
10	TOTALNS	37%	Total Consumer Credit Owned and Securitized, Outstanding	Bank lending condition
11	UMCSENT	37%	Consumer Sentiment	General market indicator
12	MICH	37%	University of Michigan: Inflation Expectation	Monetary policy indicator
13	TWEXMMTH	36%	Trade Weighted U.S. Dollar Index: Major Currencies	International competiveness
14	EXPGSC1	36%	Real Exports of Goods & Services	Balance of Payments
15	REQRESNS	36%	Required Reserves of Depository Institutions	Bank lending condition
16	SPREAD.MOODY.1	36%	Interest Spread (Difference between 10-year BAA yield and10-year AAA yield)	Cost of capital
17	ULCNFB	34%	Nonfarm Business Sector: Unit Labor Cost	Cost of labour
18	CE16OV	34%	Civilian employment	Labour market conditions
19	GPDI	32%	Gross Private Domestic Investment	General market indicator

[32] Private savings are the sum of savings by people and companies. Company savings are cyclical since businesses hold liquid reserves to cushion themselves against the economic cycle and to provide funds for expansion. National saving is comprised of two subparts: public (government) and private saving. Government saving is the difference between government receipts and government expenditures. Private saving is the difference between income and outlays of the private sector. Private saving can be further disaggregated into saving by individuals (personal saving) and saving by firms (retained corporate earnings). http://www.frbsf.org/education/publications/doctor-econ/2005/august/401k-ira-national-savings-rate

20	USPHCI	32%	Coincident Economic Activity Index	General market indicator
21	CNCF	32%	Corporate net cash flow	Firm profitability
22	ALTSALES	31%	Light Weight Vehicle Sales: Autos and Light Trucks	Business cycle leading indicator
23	REALLNNSA	31%	Total real estate loans	Bank lending condition
24	W068RCQ027SBEA	30%	Government Total Expenditure	General market indicator

Note: Default database is not constructed by merging DRD with CRSP/Compustat. The total sample is from 1980/04 to 2014/12. 5-year as the size of each internal is used to extract a set of effective predictors. 5-year as the size of each internal is used to extract a set of effective predictors. I use rolling window by a month based 1980/04on the unit of 5-year interval. Then there are 357 sub-intervals from the first interval from 1980/04 to 1985/03 to the last interval from 2010/01 to 2014/12. Then there are 357 groups of effective predictors. Frequency % means that the percentage of frequency of occurrence in the 357 groups of effective predictors. If the number is more than 50%, it means that this predictor can explain more than half of sample period.

5.5 Conclusion

This chapter focuses on investigation of dynamic change in EPs for predicting corporate default probabiltities in US industrial firms. The time span is from April 1980 to December 2014 with 92 macroeconomic factors. To explore this dynamic phenomenon, this study designs a mechanism by using LASSO to extract EPs in using different sub-samples based on sliding and rolling window analysis.

The study finds that no EPs can constantly predict IDPs over time; that is, the 357 groups of EPs extracted from 357 intervals are different. The set of EPs is dynamic rather than stable. The strongest EP is the total borrowings of depository institutions from the Federal Reserve which can explain IDPs in 54% of 357 intervals. If this study is only interested in strong EPs that can predict IDPs in more than 45% of all intervals, there are five EPs selected. One is about bank lending conditions, three are related to montary policy indicators. The last one is business cycle leading indicator. Interestingly, two of them are monetary tools, which are borrowings of depository institutions and monetary base, respectively. The other strong EPs, such as core consumer price index and gross private savings, can give a sign of a monetary target from an economic policymaker. Then it can be concluded that monetary policies have strong association with historical corporate defaults, suggesting that they may result in default clustering.

It can be seen that both bank lending conditions and monetary policy indicators have more explanatory power for corporate defaults in the US in comparison with the other macroeconomic factors. The former reflects the degree of monetary liquidity; the latter shows the price level in the economy. This result mirrors the interactions between monetary policies and credit market stability; in other words, the implementation of monetary policy may result in severe defaults happening thereby leading to instability of the credit market. Defaults especially exhibit higher counter-cyclicality. For the decision makers, it is cautious to make an appropriate monetary policy shaping risk-taking to simultaneously stabilise the economy and the financial market in practice. In particular, several monetary tools related to borrowing of depositary institutions from the central bank and monetary base should be considered

prudent since they have had strong predicting power for default risk in the past. Money market fund as a supplementary tool can provide liquidity to the market for the banks and corporates, which also has significant influence on corporate defaults particulary in the 2007-2009 financial crisis. This finding provides emprical support of recent argument of whether monetary market fund impacts on the financial staibiltiy.

In addition to four out of five EPs noted above, the following significant EPs for IDP are business cycle leading indicators, such as inventories index, vehicle miles travelled, and vehicles sold. These correpond to intuition; that is, leading indicators generally move earlier than the occurrence of the business cycle. Earlier literature finds that default clustering tends to follow the trend of the business cycle. It is clear that the leading indicators may provide predictive information for explaining corporate default probabilties.

In contrast, three popular EPs used by prior empirical studies on default risk modelling have poor predictive performance in the 357 intervals. This finding figures the weakness of using fixed sets of risk predictors, even though they have contributed to developing advanced techniques, such as reduced form models to predict default probabilities by considering various risk predictors from the macroeconomic environment.

Future research about prediction of macro indicator can further test its performance of predicting default cycle among all the firms rather than only in the industrial firms. Another possible direction is to investigate dynamic EPs for predicting defaults in the other economic sectors, such as financial services, utilities, and transportation etc. These results may provide some insights into how macroeconomic conditions dynamically influence default risk in specific economic sectors in the past, which can help decision makers understand how default risk spreads with the change of macroeconomic conditions.

CHAPTER 6

6. Investigating the Relationship between the Dynamic of Various Macro Indicators and Economic Recessions

6.1 Introduction

This chapter constructs macro indicators by measuring the interactions among macroeconomic and financial variables in order to capture the critical transitions in the macroeconomic system. Based on these indicators, this chapter employs standard logit model to examine whether these proposed indicators individually have one-month ahead prediction to US economic recessions from September 1980 to December 2014. In particular, hundreds of macroeconomic and financial variables are used to build different macro indicators for predicting US recessions.

There are two main motivations. First, according to Scheffer et al. (2012), the majority of work on generating empirical indicators to detect critical transitions in the complex system has been carried out in ecology and climate science. No study has used this method for measuring changes in a macroeconomic system. The main reason is that economists do not consider cascade effects exist in the economic system. According to Helbing (2013), the equilibrium paradigm and the representative agent approach are the major pillars in economic study. Regarding the equilibrium paradigm, economic systems have a tendency to evolve towards an equilibrium state. There are no bubbles and crashes, suggesting that precautions are unnecessary in practice. Even though external shocks can cause sudden changes in the economy, for economists, they are not likely to recognise that there are interactions between

system elements that can lead to propagating cascade effects even if all components relax to their equilibrium state.

Representative agent models assume that companies act in the way a representative individual would optimally choose, and they consider the dynamic processes in the representative agent. However, Helbing (2013, p.56) explains that "such models cannot capture processes well if random events, the diversity of system components, the history of the system or correlations between variables matter a lot. It can even happen that representative agent models make predictions opposite to those of agent-based computer simulations assuming the very same interaction rules." Both theories including equilibrium and representative agent models are not essentially compatible with probabilistic cascade effects. In fact, cascade effects also exist in the economic system. These effects are able to result in a system being away from its previous (equilibrium) state, and the representative dynamics are not available since diverse possible paths of events may be very different. Helbing (2013) explains that the spread of innovations and products can cause cascade effects and then there is no exception in the current economic system.

Assuming that if there is an external shock in the economic system, several macro factors respond at the same time, this small number of factors may lead to changes in the other macro factors. The interactions among them will increase, which implies that amplification of cascade effects is ongoing. For instance, a central bank as the policymaker may increase interest rates thereby curbing an overheated economy, which increases the cost of capital. At the macro level, the number of loans decreases rapidly and the banks slash credit to the public. The firm faces a higher operation cost and cost of financing. Then employees are laid off and the salaries for other employees reduce, which can be reflected by numerous macro factors, such as unemployment ratios, working hours, and initial claims etc. Poor economic conditions spark crisis in the economy, which leads to a decrease in the amount of industrial production and even negatively impacts on future expectations of investors and consumers. Consequently, this deterioration in economic conditions further pricks the bubble in any markets. It is clear that the cascade effects are quite obvious in the economic system. It can be concluded that studying cascade effects in the economic system would be helpful to capture changes in the macroeconomic system.

Macroeconomic conditions change over time within a business cycle, which moves between expansion and contraction. When there is an expansion of the economy, economic conditions tend to be positive; when the economy is contracting, economic conditions become negative. The macroeconomic conditions are reflected by the co-movements among numerous macroeconomic and financial variables. In the business cycle forecast literature, the phenomenon of co-movements provides a sign of the current state of the macroeconomy. High co-movement implies that macroeconomic conditions tend to be worse; otherwise, the conditions tend to be better. That is, if there are higher correlations or interactions among these macro factors, then economic conditions become worse, implying that the probability of economic recession occurring goes up; otherwise, the economic conditions become better with a lower probability of occurrence of economic recession. In this context, constructing macro indicators to measure macroeconomic conditions is introduced, which can shed new light on how to improve predictive performance on recessions by using alternative recession factors instead of conventional ones, such as yield spreads, stock market return, and interest rates (Estrella and Mishkin, 1998, Nyberg, 2010, Ng, 2012).

Most literature focuses on application of probit models and its extensions, such as various dynamic probit models that have been proposed in order to handle autocorrelated errors and multiple break points across business cycles (Kauppi and Saikkonen, 2008, Nyberg, 2010). This study is to apply standard logit model instead of using these probit models to predict the recessions since this method is as good a binary response model as probit models. I believe that using a key recession factor that always provides the sign of future changes in the macroeconomic system plays a more significant role of forecasting economic recessions than using more sophisticated models.

The main findings can be summarised as follows. First, I find that both macro indicators separately constructed by procyclical factors and factors from six economic groups, except the effective predictors extracted by Lasso, have the most accurate one-month ahead prediction on US recession months in the out-of-sample analysis. This result implies that the interaction among macro factors has predictive content for US future recessions. The other indicators that are constructed by Lasso also have a very good performance, suggesting that Lasso can better capture the predictors that potentially explain economic recessions. Second,

most indicators have better out-of-sample estimation results than the in-sample estimation results, indicating that macro indicators are able to predict future recessions in the US. Third, 25% as the threshold for the logit model can generate more accurate forecasts than 50%, which is consistent with the previous studies (Ng, 2012).

This chapter is organized as follows. The next section describes the data. The method of how to construct macro indicators is also mentioned in this section. In Section 6.3, the empirical results are presented including the in-sample and out-of-sample predictions of recession periods in the US, which is followed by discussions in Section 6.4. Section 6.5 concludes.

6.2 Data

The database in this chapter is constructed from FRED, which contains 114 macro level and micro level economic factors. Based on Koopman et al. (2011), these variables are grouped into two broad categories including macro and business cycle (BC) conditions and micro level conditions. There are 85 factors in the macro and BC conditions, which have been classified into 8 economic groups: (a) bank lending conditions; (b) general macro indicators; (c) labour market conditions; (d) monetary policy indicators; (e) liquidity from non-banks; (f) balance of payments; (g) terms of trade; (h) labour cost/wages; (i) cost of capital; (j) cost of resources; and (k) equity market conditions. Micro level conditions contain 4 economic groups: (a) cost of labour/wages; (b) cost of capital; (c) cost of resources; and (d) equity market conditions. All the macro factors with their corresponding classifications are listed in Table 6-2. For the interpretation of each factor, they are all detailed in Table 4-1 in Chapter 4. I use the definition of economic recession from NBER, which is exhibited in Table 6-1.

Table 6-1 Historical recession from 1980 to 2014 in the US

Date of cycle	Length of cycle	Historical background	
January 1980 – July 1980	6 months	Contractionary monetary policy undertaken by the Federal Reserve to handle double digit inflation; Residual effects of the oil crisis.	Savings and loans crisis
July 1981 – November 1982	16 months		
July 1990– March 1991	8 months	Junk bond defaults	
March 2001– November 2001	8 months	Dot-com crisis	
December 2007– June 2009	18 months	Subprime crisis	

Source from FRED.

Table 6-2 Data description for macro factors

Summary listing	Factors										Total
	Full name	Abbreviation	Cyclical indicator	Relation with business cycle	Relation with default risk	Full name	Abbreviation	Cyclical indicator	Relation with business cycle	Relation with default risk	
Macro and BC conditions (85 factors)											
Bank lending conditions	Loans and Leases in Bank Credit, All Commercial Banks	LOANS	Lagging factor	Procyclical	Negative	Household debt/income-ratio	TDSP	Lagging factor	Countercyclical	Positive	11
	Real Estate Loans, All Commercial Banks,1943	REALLN	Lagging factor	Procyclical	Negative	Household obligations/income	FODSP	Lagging factor	Countercyclical	Positive	
	Total Consumer Credit Owned and Securitized, Outstanding	TOTALSL	Lagging factor	Procyclical	Negative	Interbank Loans, All Commercial Banks	IBLACBM027SBOG	Lagging factor	Procyclical	Negative	
	Commercial and Industrial Loans, All Commercial Banks	BUSLOANS	Lagging factor	Procyclical	Negative	Borrowings, All Commercial Banks	BOWACBM027SBOG	Lagging factor	Procyclical	Negative	
	Consumer Loans at All Commercial Banks	CONSUMER	Lagging factor	Procyclical	Negative	Required Reserves of Depository Institutions	REQRESNS	Lagging factor	Countercyclical	Positive	
	Federal Debt: Total Public Debt	GFDEBTN	Lagging factor	Countercyclical	Positive						
General macro indicators	Economic activity index	USPHCI	Coincident factor	Procyclical	Negative	Real Manufacturing and Trade Industries Sales	CMRMTSPL	Coincident factor	Procyclical	Negative	36
	Industrial Production Index	INDPRO	Coincident factor	Procyclical	Negative	Smoothed recession probabilities	RECPROUSM156N	Coincident factor	Countercyclical	Positive	
	Industrial Production: Mining: Drilling oil and gas wells	IPN213111S	Coincident factor	Procyclical	Negative	Uni Michigan consumer sentiment	UMCSENT	Leading factor	Procyclical	Negative	
	Industrial Production: Manufacturing (SIC)	IPMANSICS	Coincident factor	Procyclical	Negative	Real final sales of domestic product	A190RL1Q225SBEA	Leading factor	Procyclical	Negative	
	Industrial Production: Mining	IPMINE	Coincident factor	Procyclical	Negative	Final Sales to Domestic Purchasers	FSDP	Leading factor	Procyclical	Negative	
	Industrial Production: Electric and Gas Utilities	IPUTIL	Coincident factor	Procyclical	Negative	Expenditure durable goods	PCEDG	Leading factor	Procyclical	Negative	
	Industrial Production: Materials	IPMAT	Coincident factor	Procyclical	Negative	New One Family Houses Sold	HSN1F	Leading factor	Procyclical	Negative	
	Personal income	PI	Coincident factor	Procyclical	Negative	Capacity Utilization: Manufacturing (NAICS)	MCUMFN	Leading factor	Procyclical	Negative	
	Real disposable personal income	DSPIC96	Coincident factor	Procyclical	Negative	Capacity Utilization: Total Industry	TCU	Leading factor	Procyclical	Negative	
	Personal Consumption Expenditures	PCE	Coincident factor	Procyclical	Negative	Moving 12-Month Total Vehicle Miles Traveled	M12MTVUSM227NFWA	Leading factor	Procyclical	Negative	
	Personal Consumption Expenditures: Chain-type Price Index	PCEPI	Coincident factor	Procyclical	Negative	Light Weight Vehicle Sales: Autos & Light Trucks	ALTSALES	Leading factor	Procyclical	Negative	
	Government expenditure	W068RCQ027SBEA	Coincident factor	Countercyclical	Positive	Housing Starts	HOUST	Leading factor	Procyclical	Negative	

	GDP	GDP	Coincident factor	Procyclical	Negative	Building Permits	PERMIT	Leading factor	Procyclical	Negative	
	Gross private domestic investment	GPDI	Coincident factor	Procyclical	Negative	ISM Manufacturing: New Orders Index	NAPMNOI	Leading factor	Procyclical	Negative	
	Private Nonresidential Fixed Investment	PNFI	Coincident factor	Procyclical	Negative	ISM Manufacturing: Inventories Index	NAPMII	Leading factor	Procyclical	Negative	
	Change in private inventories	CBI	Leading factor	Procyclical	Negative	ISM Manufacturing: Supplier Deliveries Index	NAPMSDI	Leading factor	Procyclical	Negative	
	Private Residential Fixed Investment	PRFI	Coincident factor	Procyclical	Negative	ISM manufacturing index	NAPM	Leading factor	Procyclical	Negative	
	Gross National Product	A001RP1Q027SBEA	Coincident factor	Procyclical	Negative	The months' supply is the ratio of houses for sale to houses sold.	MSACSR	Leading factor	Countercyclical	Positive	
Labour market conditions	Initial Claims	ICSA	Leading factor	Countercyclical	Positive	Civilian Employment	CE16OV	Coincident factor	Procyclical	Negative	13
	Weekly Hours Worked: Manufacturing for the United States	HOHWMN02USM065S	Leading factor	Procyclical	Negative	All Employees: Total Nonfarm Payrolls	PAYEMS	Coincident factor	Procyclical	Negative	
	Employment Level: Part-Time for Economic Reasons, Slack Work or Business Conditions, Nonagricultural Industries	LNS12032198	Leading factor	Procyclical	Negative	All Employees: Manufacturing	MANEMP	Coincident factor	Procyclical	Negative	
	ISM Manufacturing: Employment Index	NAPMEI	Leading factor	Procyclical	Negative	Average (Mean) Duration of Unemployment	UEMPMEAN	Lagging factor	Countercyclical	Positive	
	Civilian Unemployment Rate	UNRATE	Coincident factor	Countercyclical	Positive	Of Total Unemployed, Percent Unemployed 27 Weeks and over	LNU03025703	Lagging factor	Countercyclical	Positive	
	Labor Market Conditions Index	FRBLMCI	Coincident factor	Procyclical	Negative	Number of Civilians Unemployed for 15 Weeks & Over	UEMP15OV	Lagging factor	Countercyclical	Positive	
	Civilian Employment-Population Ratio	EMRATIO	Coincident factor	Procyclical	Negative			Coincident factor	Procyclical	Negative	
Monetary policy indicators	Gross Saving	GSAVE	Coincident factor	Procyclical	Negative	Consumer Price Index for All Urban Consumers: Housing	CPIHOSSL	Lagging factor	Procyclical	Negative	15
	Gross Private saving	GPSAVE	Coincident factor	Procyclical	Negative	Consumer Price Index for All Urban Consumers: All Items	CPIAUCSL	Lagging factor	Procyclical	Negative	
	Personal Saving	PMSAVE	Coincident factor	Procyclical	Negative	Monetary Base	BOGMBASE	Leading factor	Procyclical	Negative	
	GDP deflator, implicit	GDPDEF	Coincident factor	Procyclical	Negative	M1 Money Stock	M1SL	Leading factor	Procyclical	Negative	
	University of Michigan Inflation Expectation	MICH	Leading factor	Procyclical	Negative	M2 Money Stock	M2SL	Leading factor	Procyclical	Negative	
	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy	CPILFESL	Lagging factor	Procyclical	Negative	M3 for the United States	MABMM301USM189S	Leading factor	Procyclical	Negative	

	Consumer Price Index for All Urban Consumers: Energy	CPIENGSL	Lagging factor	Procyclical	Negative	MZM Money Stock	MZMSL	Leading factor	Procyclical	Negative	
	Consumer Price Index for All Urban Consumers: Transportation	CPITRNSL	Lagging factor	Procyclical	Negative						
Liquidity from non-banks	Retail Money Funds	RMFSL	Leading factor	Procyclical	Negative	Institutional Money Funds	IMFSL	Leading factor	Procyclical	Negative	2
Firm profitability	Corporate Profits After Tax	CP	Lagging factor	Procyclical	Negative	Corporate net cash flow	CNCF	Lagging factor	Procyclical	Negative	3
	Net corporate dividends	B056RC1A027NBEA	Lagging factor	Procyclical	Negative						
Terms of trade	Trade Weighted U.S. Dollar Index: Broad	TWEXBMTH	Leading factor	Procyclical	Uncertain	Trade Weighted U.S. Dollar Index: Major Currencies	TWEXMMTH	Leading factor	Procyclical	Uncertain	2
Balance of payments	Net Exports of Goods and Services	NETEXP	Leading factor	Countercyclical	Uncertain	Real Imports of Goods & Services, 3 Decimal	IMPGSC96	Leading factor	Procyclical	Uncertain	3
	Real Exports of Goods & Services	EXPGSC1	Leading factor	Countercyclical	Uncertain						
Micro-level conditions (29 factors)											
Labour cost/wages	Unit labor cost: nonfarm business	ULCNFB	Coincident factor	Procyclical	Positive	Business Sector: Real Output Per Hour of All Persons	OPHPBS	Coincident factor	Procyclical	Positive	5
	Nonfarm Business Sector: Real Compensation Per Hour	COMPRNFB	Coincident factor	Procyclical	Positive	Nonfarm Business Sector: Real Output Per Hour of All Persons	OPHNFB	Coincident factor	Procyclical	Positive	
	Nonfarm Business Sector: Compensation Per Hour	COMPNFB	Coincident factor	Procyclical	Positive						
Cost of capital	Effective federal funds rate	FEDFUNDS	Lagging factor	Countercyclical	Positive	10-Year Treasury Constant Maturity Rate	GS10	Lagging factor	Countercyclical	Positive	14
	30 year mortgage rate	MORTG	Lagging factor	Countercyclical	Positive	Treasury bond yield, 10 years(Baa)	BAA10YM	Lagging factor	Countercyclical	Positive	
	AAA corporate bond yield	AAA	Lagging factor	Countercyclical	Positive	3-Month Treasury Bil	TB3MS	Lagging factor	Countercyclical	Positive	
	BAA corporate bond yield	BAA	Lagging factor	Countercyclical	Positive	10-Year Treasury Constant Maturity Minus Federal Funds Rate	T10YFF	Leading factor	Countercyclical	Positive	
	Treasury bond yield, 10 years(Aaa)	AAA10YM	Lagging factor	Countercyclical	Positive	Interest spead: Difference between 10-year Treasury constant maturity rate and 1-year Treasury constant maturity rate	SPREAD.GS	Leading factor	Countercyclical	Positive	
	Bank prime loan rate	MPRIME	Lagging factor	Countercyclical	Positive	Interest spead: Difference between 1-year BAA yield and 1-year AAA yield	SPREAD.MOODY.1	Leading factor	Countercyclical	Positive	
	1-Year Treasury Constant Maturity Rate	GS1	Lagging factor	Countercyclical	Positive	Interest spead: Difference between 10-year BAA yield 10-year and AAA yield	SPREAD.MOODY.2	Leading factor	Countercyclical	Positive	

Cost of resources/Labour cost	PPI all commodities	PPIACO	Lagging factor	Procyclical	Negative	PPI industrial commodities	PPIIDC	Lagging factor	Procyclical	Negative	6
	PPI interm. energy goods	PPIIEG	Lagging factor	Procyclical	Negative	PPI crude energy materials	PPICEM	Lagging factor	Procyclical	Negative	
	PPI finished goods	PPIFGS	Lagging factor	Procyclical	Negative	PPI intermediate materials	PPIITM	Lagging factor	Procyclical	Negative	
Equity indices and respective volatilities	SP500 index	SP500	Leading factor	Procyclical	Negative	Russell 2000 index	RU2000	Leading factor	Procyclical	Negative	4
	NASDAQ composite index	NASDAQ	Leading factor	Procyclical	Negative	Wilshire 5000 Total Market Full Cap Index	WILL5000	Leading factor	Procyclical	Negative	
					Total						114

Note: All variables except exchange rates, interest rates, and industrial material price index are seasonally adjusted and they are all transformed to growth rate except four yields yield spreads including T10YFF, SPREAD.GS, SPREAD.MOODY.1, and SPREAD.MOODY.2. The method of classifying these macroeconomic and financial time series into three groups including leading, cocindent, and lagging indicator follows the study done by Conference Board (2001) and Shiskin and Moore (1967).

6.2.1 Construction of Six Macro Indicators

(1) Method of construction of macro indicator

Covariance is a measure of how much two variables change together, which is also able to reflect how the factors interact with each other. This study uses this method to measure the interactions among numerous factors thereby constructing macro indicators for capturing critical transitions in the macroeconomic system. Specifically, the indicator is based on the calculation of covariance (v), and then we collect information by using similar approach as the indicator method above. The formula of calculating the indicator is shown as follows:

$$\text{Indicator}_p = \|v\|_p = \left(\sum_{j>k} |v_{jk}|^p \right)^{\frac{1}{p}}, \tag{65}$$

where v_{jk} $(j > k > 0)$ is actually covariance between X and Y that represent two variables, $v_{X,Y} = \text{cov}(X, Y)$. Specifically, I have one dataset $\{X_1, , \ldots, X_n\}$ containing n values and another dataset containing n values. $\text{cov}(X, Y)$ is the covariance between X and Y. Then I have $n(n-1)/2$ covariance between various variable, v_{jk}. It should be noted that these values are located in the non-diagonal part of the covariance matrix.

In this study, the covariance matrix is composed of macro factors mentioned in the previous section. It is clear that the current method can convert a specific covariance matrix to an observation. Since this study focuses on the prediction of an indicator, then sliding time windows based on 8 months is used for obtaining the time series of the indicator. If there is a time moment t, the Indicator_p is calculated in the time window $[t-7, t]$, which precedes t.

First, past economic crises normally have experienced this length of period. Then calculating covariance by using same length of interval is straightforward. NBER's Business Cycle Dating Committee establishes and maintains the chronology of US business cycles and the average recession, and this committee determines that length of recession tends to average about 9 months if the extremes are eliminated.

The motivation for using 8 months is threefold. First, it is motivated by the length of historical recessions. Post-war, the shortest recession lasted 6 months, occurring in the early 1980s. The second shortest recession lasted 8 months, in 2001 (Kliesen, 2003). 8 months is more preferable for constructing macro indicators. Second, from a statistical perspective, a short interval can rapidly reflect the changes of interactions in the system and a long interval cannot respond quickly in current economic conditions since many irrelevant interactions have been considered in long intervals. Additionally, a short interval such as 6 months may generate biased interaction results since the sample size is quite small. Meanwhile, when using 6 months, I find that the predicting performance does not improve a lot. If using a bigger window size than 8 months, then there is a significant lag effect in the indicator, which reduces the ability of capturing economic recessions. Third, 8 months is close to the average length of historical recession (9 months), implying that 8 months theoretically reflects the macroeconomic conditions. Therefore, 8 months is used in this study.

(2) Justifications of construction of six macro indicators

This chapter uses 6 different groups of macro factors to construct macro indicators, which are shown as follows.

a. All the factors

The first indicator is constructed by all the macro factors in order to reflect relatively complete macroeconomic conditions. The reason for choosing all the factors is that a future event is unpredictable; that is, it is not possible to predict which groups of elements in the macroeconomic environment can lead to the co-movement of the other factors. This is supported by theoretical studies. Specifically, a broad category can provide protection against significant changes in cyclical behaviour that may result from such factors as technological developments, changing consumer tastes, or the rapid growth or decline of single products or industries (Shiskin and Moore, 1967, Kliesen, 2003, Ng, 2012). These broad economic factors can continuously capture the potential changes in the business cycle, even though some factors that performed well in the past deteriorate in the present. In addition, economists still cannot find an effective indicator from numerous macro factors to make future predictions of macroeconomic conditions since prior studies identify that the environment is

always evolving over time and the current group of factors may only mimic dynamic change of macroeconomic conditions in the specific interval (Moore, 1961, Shiskin and Moore, 1967, Zarnowitz, 1992b). Scheffer et al. (2012) conclude that "the methods for detection of incipient transitions from time series tend to require long, high-resolution data." That is, if more factors are collected for constructing a macro indicator, it may capture any transitions caused by the increase of correlations from the factors that have not been considered as the trustworthy factors thereby raising the ability of capturing unpredictable changes in macroeconomic conditions. Due to the issue of availability of macro factors used in this study, I cannot collect and use all the macro factors from FRED. I can only capture relatively complete macroeconomic conditions, and this disertation refers to this as augmented macroeconomic conditions.

b. Factors in six economic groups

The second indicator is constructed by 106 factors in six economic groups in Table 6-2, for example, employment and unemployment; production, income, consumption, and trade; fixed capital investment; inventories and inventory investment; prices, costs, and profits; money and credit. Shiskin and Moore (1967) discuss the issue of whether grouping the factors based on economic processes is meaningful in practice, and they determine that "under the criterion of economic significance, it would seem best to evaluate groups of series representing a general type of activity, because the economic theories which purport to explain business cycle phenomenon do not ordinarily refer to particular indicators, but rather to generalized economic processes." They propose a solution of grouping macro factors thereby generalizing the strategic processes in the business cycles and all the macro factors are classified into 9 types of economic groups. Excluding the six economic groups mentioned earlier, the other three economic groups are: (a) foreign trade and payments; (b) federal government activities; and (c) economic activity in other countries. They emphasize that the first two groups are measures of aggregate economic activity, and they are used to show the broad movements of the business cycle and to determine the dates when there is an economic expansion and contraction starting or ending. They further declare that the next four from fixed capital investment to money and credit with the first two groups can also mirror the business cycle with a causal role in the cyclical process. The last three groups do not

contribute to the cyclical fluctuations in the US; however, they significantly impact their pattern, amplitude, and duration. This study then selects the majority of the macro factors from group 1 to group 6. It can be seen that these factors may show generalized macroeconomic conditions.

c. Leading factors

This indicator is constructed by using 42 leading factors in Table 6-2. Leading factors tend to move direction before the business cycle occurs and the timing comparisons between the cyclical turn in these economic forwarding factors and the business cycle turn is one month or more, such as new orders, housing starts, and consumer sentiment. Many practitioners and scholars use leading factors to predict the future perspective for a specific country or global economy. Compared with the other two cyclical indicators, lagging and coincident factors, it can be concluded that leading factors may provide more prediction information than the other two types of factors since the interaction among them may occur earlier than the interaction among coincident factors or lagging factors. I expect that the macro indicator constructed by the group of leading factors will perform better than the other macro factors built by using lagging or coincident factors, respectively. Since these factors can only deliver limited information, this indicator is used for measuring incomplete macroeconomic conditions.

d. Procyclical factors

The fourth macro indicator is constructed by procyclical factors in Table 6-2. According to the relation between business cycle and each economic factor, this study splits them into two parts. One part contains 86 procyclical indicators; another covers 28 countercyclical indicators. Since the number of countercyclical factors is quite small with poor performance, I only use the procyclical factors for constructing the indicator. When the economy is entering into an expansion, these cyclical factors increase; otherwise, they decrease. This method can keep the macro factors which only have positive relationship with the business cycle in order to filter interactions among these factors. It can be assumed that this indicator is able to pick incomplete macroeconomic conditions.

e. Effective factors extracted by LASSO

I use LASSO to extract 25 effective factors for explaining historical economic recessions, and these factors are shown in Table 6-3. It is intuitive that the set of effective factors may outperform the other factors since these factors contain highly predictive information in explaining past default events. That is, these factors can pick the historical recessions from Feb 1980 to December 2014.[33] Therefore, the indicator constructed by effective factors can show specific macroeconomic conditions for US recessions in the past 34 years.

f. Factors about macro and business cycle conditions

The sixth group is based on 85 macro factors related to macro and business cycle conditions, which refect the economy at the macro level rather than the micro level. This indicator is based on calcuation of the co-movements among these factor from macro level.

[33] All the variables for modelling recessions by Lasso are the growth rate except 4 yield spreads. I also lag each macro factor by one-month for model estimation. In order to obtain appropriate model estimation, 6 different sizes of training samples such as 65%, 70%, 75%, 80%, 85%, and 90% are used for estimating Lasso and each method leads to the corresponding results. The standard of choosing the results depends on whether all these sets of results are similar. If Lasso can give us similar results even using different training samples, then the results obtained by Lasso are reliable. Then there are 6 sets of effective factors for explaining historical recessions based on 6 different training sample sizes. Basically, the results tend to be stable and they are almost similar after using training sample sizes such as 65%, 80% and 90%. When the percent of training sample is over 80%, the outcomes become quite similar. The number of effective factors also increases with an increasing training sample size and all the effective factors obtained by using small size of training sample are included in the group of effective factors. Thus, this study uses the results based on the training sample being equal to 80% to construct effective indicators for predicting economic recessions.

Table 6-3 Effective factors for economic recession

Abbreviation	Full name	Economic groups	Timing classification	Relation with business cycle
A001RP1Q027SBEA	Gross national product	Production, income, consumption, and trade	Coincident indicator	Procyclical
A190RL1Q225SBEA	Real final sales of domestic product	Production, income, consumption, and trade	Leading indicator	Procyclical
B056RC1A027NBEA	Net corporate dividends	Prices, cost, and profits	Lagging indicator	Procyclical
CNCF	Corporate net cash flow	Prices, cost, and profits	Lagging indicator	Procyclical
COMPRNFB	Nonfarm business sector: real compensation per hour	Prices, cost, and profits	Coincident indicator	Procyclical
CP	Corporate profits after tax	Prices, cost, and profits	Lagging indicator	Procyclical
EXPGSC1	Real exports goods, services	Foreign trade and payments	Leading indicator	Countercyclical
GDPDEF	GDP deflator, implicit	Prices, cost, and profits	Coincident indicator	Procyclical
GFDEBTN	Federal debt: Total public debt	Money and credit	Lagging indicator	Countercyclical
GPDI	Gross private domestic investment	Fixed capital investment	Coincident indicator	Procyclical
GPSAVE	Gross private saving	Money and credit	Coincident indicator	Procyclical
IMFSL	Institutional money funds	Fixed capital investment	Leading indicator	Procyclical
IPMAT	Industrial production: Materials	Production, income, consumption, and trade	Coincident indicator	Procyclical
LNS12032198	employment level: Part-Time for economic reasons, slack work or business conditions, non-agricultural industries	Employment and unemployment	Leading indicator	Procyclical
MANEMP	All employees: Manufacturing	Employment and unemployment	Coincident indicator	Procyclical
MORTG	30 year mortgage rate	Money and credit	Lagging indicator	Countercyclical
MPRIME	Bank prime loan rate	Money and credit	Lagging indicator	Countercyclical
PNFI	Private non-residential fixed investment	Fixed capital investment	Coincident indicator	Procyclical
PPICEM	PPI crude energy materials	Prices, cost, and profits	Lagging indicator	Procyclical
RECPROUSM156N	Smoothed recession probabilities	NA	Coincident indicator	Countercyclical
RU2000	Russell 2000 index	Prices, cost, and profits	Leading indicator	Procyclical
TCU	Capacity utilization: Total industry	Production, income, consumption, and trade	Leading indicator	Procyclical
TWEXBMTH	Trade weighted U.S. dollar index: Broad_	Foreign trade and payments	Leading indicator	Procyclical
UEMP15OV	Number of civilians unemployed for 15 weeks & over	Employment and unemployment	Lagging indicator	Countercyclical
USPHCI	Economic activity index	Production, income, consumption, and trade	Coincident indicator	Procyclical

(3) Basic description of 6 macro indicators

Table 6-4 provides the summary statistics of 6 macro indicators mentioned in the previous part. Basically, each statistic value for both indicator.all and indicator.6.economic is similar. It implys that the shape of their distributions is quite similar. They are quite volatile in comparison with the others since their standar deviations are higher than the others. In

contrast, the indciator constructed by effective predictors is least volatile with the smallest value of standard deviation among the others. It is interesting that the mean value of each indicator is bigger than that of median value. That is, each indicator has negative skewness distribution.

Table 6-4 Summary statistics of 6 macro indicators

The table shows the basic summary statistics for 6 macro indciators. In the 1st column, from the 2nd row to the 7th row, the indicator are shown as follows: (1) Indicator.all is constructed by 114 macro factors; (2) Indicator.leading is constructed by 42 leading factors; (3) Indicator.procyclical is constructed by 86 macro factors; (4) Indicator.6.economic is constructed by 106 factors from 6 economic groups; (5) Indicator.effective is constructed by 25 effective predictors; (6) Indicator.BC is constructed by 85 macro factors related to macro and business cycle conditions. The content of each column from 1st column to 9th column shows the name of macro indicator, the number of observations, the mean value, the standard deviation, the minimum value, the value of 25%, the median value, the value of 75%, and the maximum value, respectively.

Macro Indicator	**N**	**Mean**	**St. Dev.**	**Min**	**Pctl(25)**	**Median**	**Pctl(75)**	**Max**
Indicator.all	412	9.243	9.517	1.599	3.751	5.873	10.518	57.994
Indicator.leading	412	1.962	2.148	0.305	0.665	1.08	2.326	12.733
Indicator.procyclical	412	3.813	5.534	0.420	1.113	1.945	4.173	40.207
Indicator.6.economic	412	7.242	8.578	1.049	2.540	4.375	8.034	55.808
Indicator.effective	412	0.618	0.864	0.058	0.200	0.301	0.665	7.207
Indicator.BC	412	4.983	5.601	0.669	1.952	2.973	5.807	38.536

6.3 Empirical Analysis

This section builds the indicators to measure the changes in macroeconomic conditions. As discussed in the previous section, in economic downturns, many macro factors start moving rapidly and the interactions among these factors becomes higher than in non-recession periods. Based on the method of constructing macro indicators in the previous section, the macro indicator is built to capture their co-movements over time. The method for building the macro indicator is Equation (65). This study uses 6 groups of factors for constructing the recession indicators. They are constructed by all the factors: leading factors, procyclical factors, factors in the six economic groups, effective factors, and factors related to macro and business cycle conditions.

6.3.1 Graphical Analysis

Figure 6-1 shows the dynamic change in six macro indicators and NBER recessions from 1980 to 2014. Basically, the trend of each macro indicator is different, suggesting that the variability among different groups of factors is different in the same period.

The first graph in Panel A is the dynamic in macro indicator constructed by all the factors. Before NBER defined recession in July 1981, there was a downward trend and then it increased rapidly and peaked in November 1981. This indicator decreased again until NBER defined the ending point of recession in November 1982, and then it peaked in January 1983.

This macro indicator has experienced a stable trend without any fluctuations until a new peak occurred in January 1987. However, the peak is not higher than the previous two peaks, indicating that the variability for each factor is slightly more. Interestingly, this peak was followed by Black Friday in October 1987. Then this indicator fell until 1990 and it reached a small peak in the starting point of recession defined in July 1990, which was followed by a slight decrease.

In May 1990, the macro indicator increased to around 13 before it reached a high peak in March 1991. The Federal Reserve defined the official NBER cyclical peak in July 1990 and cyclical trough in March 1991. The index then experienced quite a long stable period with a lower value for almost 10 years. Prior to the next recession occurring in March 2001, this

indicator peaked in September 2000 and decreased. When the recession ended in November 2001, this indicator still experienced an increased trend until March 2002 and it increased to around 30. After that, this indicator experienced another longer stable period. April 2007 is the turning point for this indicator since it started increasing after this month. 8 months later the Federal Reserve defined a new recession called financial crisis in December 2007. This recession lasted 18 months until June 2009. During this period, there are two peaks for this index. It should be noted that the index reached its highest peak in this period, suggesting that the variability for each factor is higher or more and more factors vary together in comparison with before. After experiencing the second peak in May 2009, this indicator decreased rapidly and stabilized at a very low level until December 2014. The cyclical trough defined by the Federal Reserve is in June 2009, which is one month later than the indicator starting decreasing.

There are several interesting findings in Figure 6-1. First, this indicator started increasing before three recessions occurring and in two recessions in 1990 and 2007, and it started decreasing when NBER recession almost ended. Second, macro factors vary greatly together in the subprime crisis in 2007 than in the other three recessions, which is followed by two recessions in the early 1980s and in July 1990. The recession in 2001 is the slightest one. Third, before Back Friday occurred, this indicator experienced a slight increase and then decrease. Fourth, this indicator seemingly failed to capture the recession in 2001.

The second graph in Panel B is the indicator constructed by leading factors. There was rapid decrease before the 1981 recession occurred, which was followed by a quick uptrend until December 1981. This indicator has had a stable downtrend for almost three years, and then it increased to the next peak (the value is nine) in December 1986, which is one month earlier than the indicator constructed by all the factors. It seems that this indicator provides an earlier signal than the indicator based on all the factors for Black Friday. This indicator decreased after this relatively small peak. It is interesting that there was no clear sign of increase in the indicator before the NBER recession in July 1990. In October 1999, the indicator increased remarkably and it finally peaked in March 1991 when this recession ended.

The macro indicator continually decreased and stabilized in a low level before the recession occurring in March 2001. During this recession period, the indicator did not peak until it

peaked recession ended. Before this indicator experienced a tremendous increase in advance of next recession in 2007, there were two small peaks occurring in 2003. In the financial crisis 2007-2009, the indicator reached highest level in June in 2009 and at the same time Federal Reserve ended the recession. After that this indicator decrease with a slight increase in January 2010 and then decrease. This indicator fluctuated around one in the rest of period.

Compared with the findings using the indicator constructed by all the factors, this indicator has different trend during the recession occurring in 1981 and there were two peaks in 2003. In addition, the highest peak occurred in the end of recession in December 2007 rather the first peak in this recession period. However, for the other findings, they are almost same.

The third graph in Panel C is the index constructed by procyclical factors. Basically, there are three major differences compared with trend of indicator constructed by all the factors. First, prior to occurrence of recession in July 1981, the decrease of this indicator is quite slight. Second, there is no clear peak in that recession. Third, there is no salient peak around 1987. The other findings can be still found in using this indicator.

The fourth graph in Panel D is about the indicator constructed by the factors from six economic groups. Generally speaking, the trend is almost same as the trend of the indicator constructed by all the factors. Then the major findings from former remain the same as from the latter.

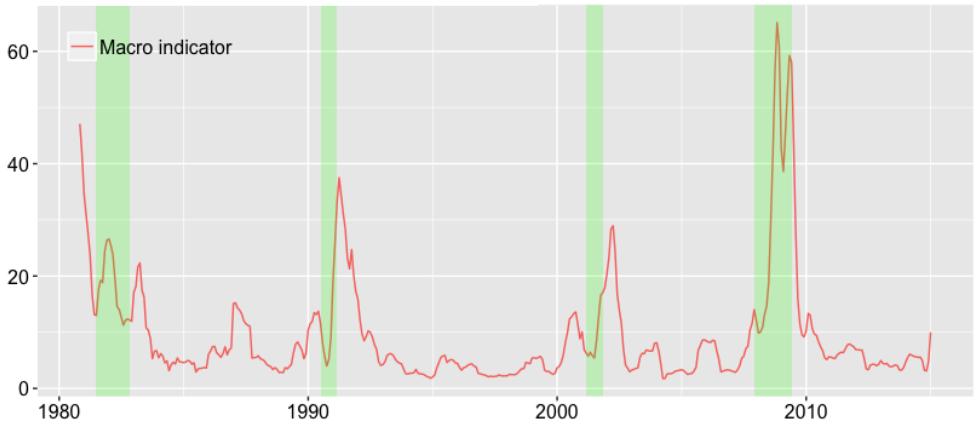
The fifth graph in Panel E provides dynamic of indicator built by 25 effective factors. The trend is quite different from the other indicators' trends. Several findings found by the other indicators are not obvious in using this indicator. Specifically, during the recession beginning in July 1981, the peak is slight lag of the other indicators. The peak is quite small even compared with the peaks in the recession occurring in March 2001. There was no any peak in the cyclical cycle from March 1991 to November 2001. In December 2014, this indicator remarkably increased from 0.96 to 4.45.

The last index in Panel F is shown in the sixth graph. This historical trend is almost same as the trend of the indicator constructed by procyclical factors. Then the findings are same as the findings based on that index. In summary, if there is rapid increase trend in these indicators, the economy tends to enter into recession. In addition, this indicator may give a

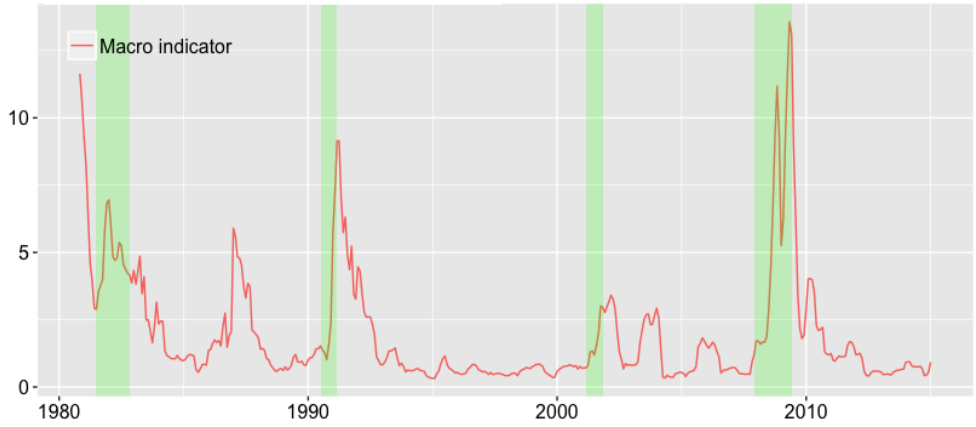
sign when the recession ends. That is, when various macro factors tend to vary together greatly, then the economy system is not stable but may have high tendency of occurrence of recession. When they do not start varying together, the economy tends to be stable since each macro factor is about to behave their own way. The recession may end at this point. If there is no any recession, each indicator tends to fluctuate along a horizontal line.

Figure 6-1 Dynamic change of recession indicators in red lines and economic recessions in green bar from 1980 to 2014

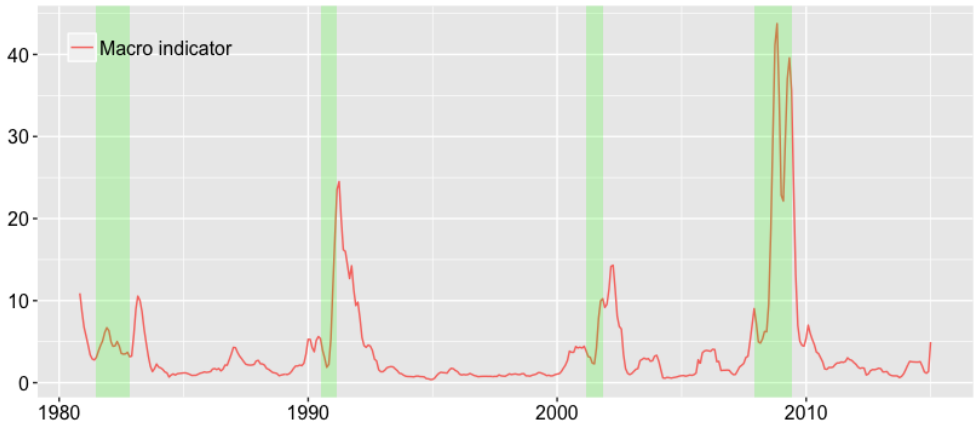
Panel A: All the factors (114)



Panel B: Leading factors (42)

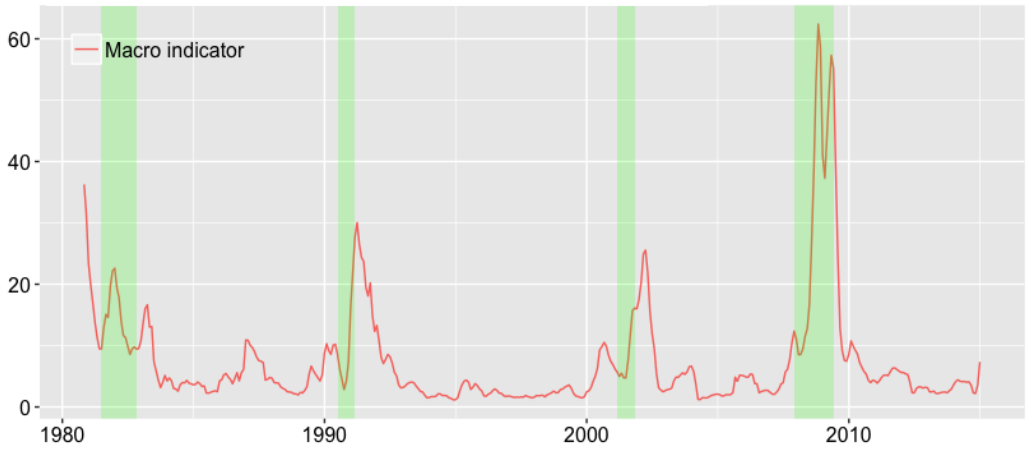


Panel C: Procyclical factors (86)

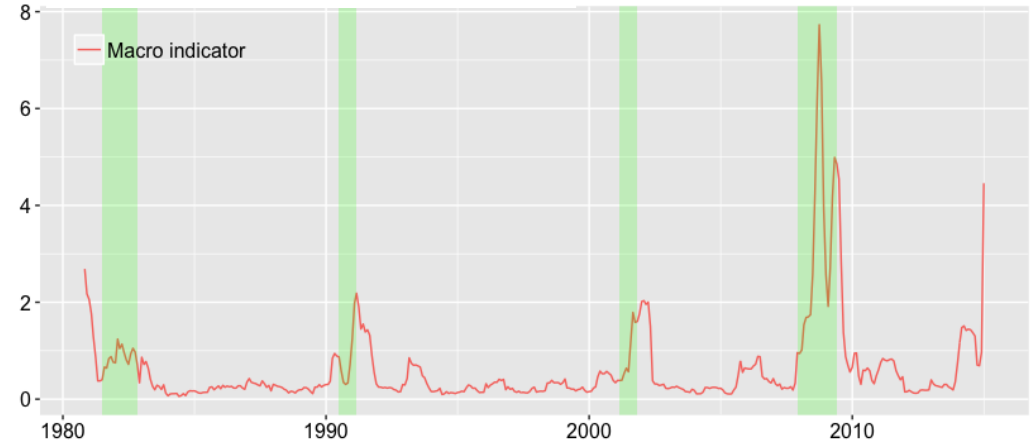


Continued

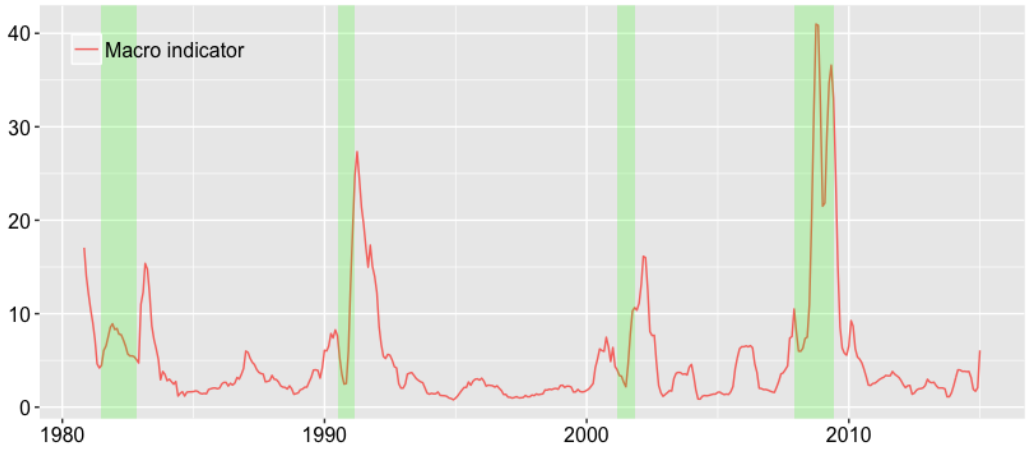
Panel D: Factors from six economic groups (106)



Panel E: Effective factors (30)



Panel F: Factors related to Macro and business cycle conditions (85)



6.3.2 Logistic Regression Analysis

This section provides empirical results after using in-sample analysis and out-of-sample analysis. In particular, the in-sample analysis is to do model estimation based on the entire sample from September 1980 to December 2014 and do model prediction in the full sample. The out-of-sample analysis is to use the first half of the full sample from September 1980 to October 1997 to do model estimation and to predict economic recessions from November 1997 to December 2014.

(1) In-sample Analysis

a. Model Estimation and Diagnostics

Table 6-5 provides logistic regression results of measuring economic recessions by six macro indicators. The coefficients reported by this table are odd ratios. Basically, all the indicators are highly significant for explaining economic recession. The effective factors have the largest coefficients, which means one unit change of this indicator results in 357% probability that recession will occur compared to that the recession will not happen. The second largest coefficient is for the indicator constructed by leading factors, and the probability of recession that will occur to the recession that will not occur is 65%, which is followed by the indicator constructed by procyclical factors, the factors related to macro and business cycle conditions, the factors in six economic groups, and all the factors.

Prior to verifying whether the prediction is accurate for each indicator, this study should test whether the models pass the diagnostics tests including deviance and Pearson goodness-of-fit statistics in Panel A and the specification test in Panel B in Table 6-6.

Since the chi-square values in Panel A are all smaller than the critical value, we cannot reject the null hypothesis of model fitting data well. In Panel B, there are only two models in column (3) and column (6) where linear predicted value squared (hastq) are not significant. Then I can conclude that model (3) and (6) fit data well in comparison with the other models. Thus, two models' estimation for the indicators including procyclical factors and factors about macro and business cycle conditions pass all the diagnostics tests; the other 4 indicators' models only pass two diagnostics tests.

b. Model Prediction

This section is to test which indicator has better predictive power for the economic recession. Two methods, Pseudo R-squared and ROC, are used in this study. For the results of Pseudo R-squared (indicator constructed by factors from six economic groups) has the largest value of 0.2659, which is followed by 0.2496 (indicator constructed by the effective factors), 0.2343 (indicator constructed by procyclical factors), 0.2337 (indicator constructed by all the factors), and 0.2251 (indicator constructed by leading factors). The smallest one is for the indicator that is built by the factors from the group of macro and business cycle conditions. This order implies the performance of each indicator from the best one to the worst one. It should be emphasized that since all the Pseudo R-squared values range from 0.2 to 0.4, then all the models also fit data well.

Figure 6-2 exhibits the plot of ROC curves and AUC values. It is clear that the largest AUC is for the indicator built by using effective factors, suggesting that this indicator has the most accurate prediction for measuring economic recession. The second largest value is the indicator constructed by factors from six economic groups, and the indicator built by procyclical factors has the third largest AUC value. The next three indicators for AUC values are constructed by leading factors, all the factors, and macro and business cycle factors, respectively. However, all the AUC are larger than 80%, which means that all the models perform well in discriminating between occurrence of economic recession and occurrence of non-recession.

Table 6-7 provides in-sample estimation results of the percentage of incorrect prediction of recession for various probit models. This table reports two different results based on two threshold levels including 50% and 25%, following prior studies (Nyberg, 2010, Ng, 2012). Basically, the error rates of correctly predicting economic recessions by using 25% classification rate are smaller than that by 50%, which is consistent with Ng (2012)'s finding. If 50% is used to classify the recession months, all the indicators can only accurately explain around 25% of recession periods. The most accurate prediction is generated by using the indicator constructed by factors from 6 economic groups, which is followed by the indicator separately constructed by procyclical indicators and factors related to macro and business cycle conditions. If the results are obtained by using 25% as the threshold, the predicting

performance has improved remarkably. The best macro indicator for predicting recession is constructed by leading factors. The following two best indicators are built by effective factors and factors from 6 economic groups.

Therefore, for the in-sample analysis, I can conclude that 25% should be used to measure recession months. The indicator constructed by leading factors and factors from 6 economic groups have the best predictive performance of measuring US historical economic recessions. However, if the model diagnostics results are concerned in this case, the indicator constructed by procyclical factors should be used for prediction. The in-sample performance of these indicators is shown in Figure 6-3.

(2) Out-of-sample Analysis

Regarding previous research, the in-sample results are seldom consistent with the out-of-sample results. The out-of-sample analysis is implemented in this study in order to further verify the performance for each indicator with one-month-ahead prediction.

a. Model Estimation and Diagnostics

I equally split the sample into two parts, following the prior studies such as Estrella and Mishkin (1998), Kauppi and Saikkonen (2008), and Liu and Moench (2016). The first half sample is for model estimation which is from September 1980 to October 1997; the rest of the sample is for out-of-sample prediction.

Table 6-8 illustrates the regression results of measuring economic recession by using 6 indicators in the sample from September 1980 to October 1997. All the indicators are highly significant for predicting economic recession. Likewise, the indicator built by effective factors has the largest coefficients, followed by indicators that are calculated by using leading factors, procyclical factors, factors related to macro and business cycle conditions, and all the factors.

For the results of measures of goodness of fit tests in Table 6-9, all the models pass the deviance and Pearson tests and do not reject the null hypothesis in Panel A, indicating that they fit the data well. Panel B shows the results of using specification test. One out of six models cannot pass the test since the variable hatsq is still significant, implying that this

model has specification errors; that is, this model does not cover all the relevant predictors and the linear combination of both indicator and economic recession is not sufficient. Thus the model for the indicator calculated by effective factors does not fit the data. I can trust the results from the other models based on the other five indicators.

b. Model Prediction

As mentioned in the in-sample analysis, both methods consisting of Pseudo R-squared and ROC can give a basic picture that the model performs well in terms of correctly predicting economic recessions. For the sample from September 1980 to October 1997, two indicators are constructed by procyclical factors and factors from six economic groups have better predictive performance concerning the results of diagnostics tests.

However, this section is interested in how each indicator predicts economic recession by using the out-of-sample approach. Table 6-10 is used in this case, which shows the two groups of prediction results from November 1997 to December 2014 after using the estimated model with the sample from September 1980 to October 1997.

As with the in-sample estimation results, the results by using 25% threshold for classifying recession months are more accurate than using 50% threshold. For the ranking of accuracy of predicting recessions based on 50% as the threshold, the indicators are calculated by effective factors, procyclical factors, factors from six economic groups, factors related to macro and business cycle conditions, all the factors, and leading factors, respectively.[34] If 25% is used as the threshold, the ranking results are coincident with the results in using 50%. In particular, the best predictor is constructed by effective factors, which corresponds to intuition. These effective factors are extracted from the entire sample, indicating that these factors have predictive content for historical recessions in the US economy. Consequently, even using out-of-sample analysis, the indicator constructed by effective factors has good performance of predicting economic recessions.

However, if I compare the out-of-sample estimation results with the in-sample estimation results, the predicting performance of an indicator constructed by lead factors only

[34] Since the model based on indicator measured by effective factors does not pass the diagnostics tests, this section does not concern this indicator

deteriorates. The other indicators have better predictive performance in the out-of-sample approach than using the in-sample approach. In addition, the predictive performance of both indicators individually constructed by procyclical factors and factors from six economic groups are comparable with each other. These two indicators and the indicator calculated by effective factors also pass diagnostics tests, and they are reliable to do forecasts in the future. Figure 6-5 illustrates the out-of-sample performance of the probit model with these macro indicators mentioned above. The forecast horizon is 1 month.

6.3.3 Robustness Testing

This study focuses on the out-of-sample estimation results since this approach can only give a real picture of how each indicator forecasts US economic recessions in practice. To further evaluate the predictive accuracy of the six probit models with different indicators, I follow the out-of-sample analysis as in Estrella and Mishkin (1996).

First, I consider a series of out-of-sample forecasts obtained from various estimation samples running from 1980:M9–1997:M12 up to 1980:M9–2014:M12. I use each sample one at a time, and then the six probit models are estimated. I do predictions for a given forecast horizon, one month in this study. I then add one more month to the estimation period, and repeat the estimation and predicting procedures. For example, if the first estimation period is 1980:M9–1997:M12 and the forecast horizon is a period of one month ahead, the forecast refers to the predicted recession probability in January 1998. The second estimation period is then 1980:M9–1998:M1 and the prediction will be calculated for February 1998. These procedures are repeated until the last prediction for December 2014 is achieved.

Table 6-11 shows the estimation results by using this iterative out-of-sample analysis. Basically, the results are similar to the results by using original out-of-sample method. The threshold 25% is more favorable for being used to classify recession months than 50%. The best three predictors are still the same as using the normal out-of-sample approach; even the error rates are the same. Therefore, the estimated results in the out-of-sample analysis are solid in this context.

Table 6-5 Logistic regression of economic recessions in the full sample from September 1980 to December 2014 - Odds ratios

Logistic regression model for explaining recession in relation to macro indcaior is shown as follows $L_t = \ln(P_t/1 - P_t) = \beta_1 + \beta_2 X_{t-1} + u_t$, where P_t and $1 - P_t$ are the probability of having recession and having no recession at date t, respectively. $P_t/(1 - P_t)$ is the odds ratio in favour of occurance of the recession - the ratio of the probabilty that a recession will occure to the probability that it will not occur. X_t is is an independent variable (representing six macro indicators) at date t. The macro indicators in the column from (1) to (6) are constructed by six groups of factors, which are all the factors, leading factors, procyclical factors, factors in the six economic groups, effective factors, and factors related to macro and business cycle conditions, respectively. Since the construction of the indicator has already concerned one-month ahead prediction, then the indicator should not be lagged by one month again. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$. Standard errors are reported in parentheses underneath the coefficients.

	Recession (=1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator.all	1.121***					
	-0.016					
Indicator.leading		1.646***				
		-0.069				
Indicator.procyclical			1.240***			
			-0.034			
Indicator.6.economic				1.158***		
				-0.021		
Indicator.effective					4.571***	
					-0.227	
Indicator.BC						1.201***
						-0.027
Constant	0.034	0.036	0.046	0.033	0.039	0.041
	-0.281	-0.277	-0.249	-0.283	-0.269	-0.261
Observations	412	412	412	412	412	412
Pseudo R-squared	0.2337	0.2261	0.2343	0.2659	0.2496	0.2101
Log Likelihood	-116.692	-117.86	-116.606	-111.785	-114.268	-120.286
Akaike Inf. Crit.	237.385	239.721	237.212	227.57	232.537	244.572

Table 6-6 Measures of goodness of fit tests for the models in Table 6-5

This table covers three types of tests for measures of goodness of fit tests for the logistic regression. Panel A shows the results of both deviance and Pearson test Panel B shows the results of specification test. The critical value of chi-square in Panel A is calculated based on degree of freedom 410 at 95% significant level. Decision means whether the null hypothesis of logit model fitting correctly is rejected or not. *p<0.1; **p<0.05; ***p<0.01. Standard errors are reported in parentheses underneath the coefficients.

Panel A Deviance and Pearson Goodness-of-Fit Statistics

Indicator	test	chiSq	df	Critical value	Decision
Indicator.all	Pearson	352.45	410	458.2108	Not reject
	Deviance	233.38	410	458.2108	Not reject
Indicator.leading	Pearson	363.42	410	458.2108	Not reject
	Deviance	235.72	410	458.2108	Not reject
Indicator.procyclical	Pearson	345.36	410	458.2108	Not reject
	Deviance	233.21	410	458.2108	Not reject
Indicator.6.economic	Pearson	334.74	410	458.2108	Not reject
	Deviance	223.57	410	458.2108	Not reject
Indicator.effective	Pearson	383.46	410	458.2108	Not reject
	Deviance	228.54	410	458.2108	Not reject
Indicator.BC	Pearson	355.29	410	458.2108	Not reject
	Deviance	240.57	410	458.2108	Not reject

Panel B Specification Test

	Recession (=1)					
	(1)	(2)	(3)	(4)	(5)	(6)
hat	0.734***	0.593***	0.894***	0.847***	0.906***	0.834***
	-0.142	-0.154	-0.138	-0.118	-0.114	-0.164
hatsq	-0.171***	-0.233***	-0.083	-0.122***	-0.107***	-0.101
	-0.065	-0.067	-0.055	-0.047	-0.028	-0.073
Constant	0.353	0.366	0.219	0.344	0.363	0.184
	-0.319	-0.29	-0.374	-0.335	-0.327	-0.34
Observations	412	412	412	412	412	412
Log Likelihood	-113.687	-112.375	-115.737	-109.191	-109.951	-119.452
Akaike Inf. Crit.	233.374	230.75	237.475	224.382	225.902	244.903

Figure 6-2 ROC curve for the logistic model as classifiers on the data from September 1980 to December 2014

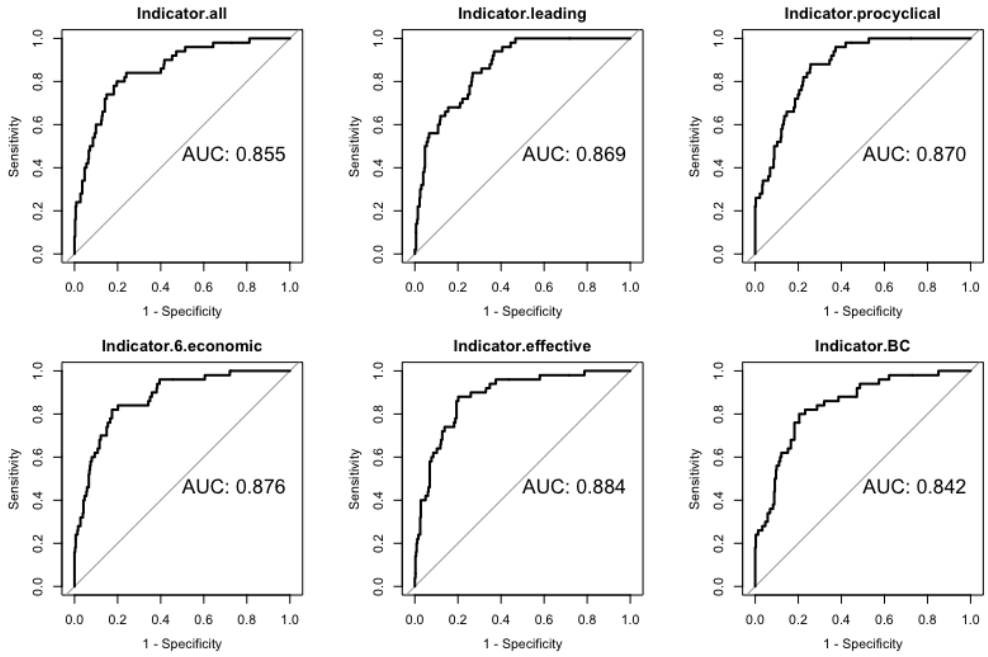


Table 6-7 In-sample estimation results for various probit models: percentage of incorrect prediction of recession

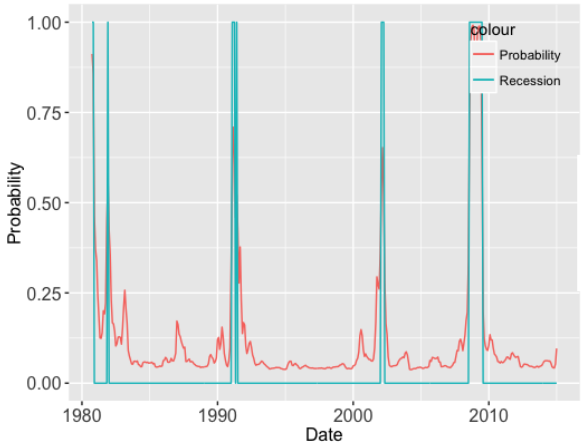
In-Sample forecasts are estimated using the full sample, from September 1980 to December 2014, and then evaluated using this sample again. Finally, I compute the predictions for the September 1980 – December 2014 period and compare them to the actual movements of the market over that time period.

	Test set error rate	
	50% Threshold of classifying recessions	25% Threshold of classifying recessions
Indicator.all	0.76	0.6
Indicator.leading	0.76	**0.5**
Indicator.procyclical	**0.74**	0.66
Indicator.6.economic	**0.72**	**0.58**
Indicator.effective	0.76	**0.56**
Indicator.BC	**0.74**	0.68

Figure 6-3 In-sample recession probabilities implied by the probit model, with selective macro indicators based on classification rate 25%

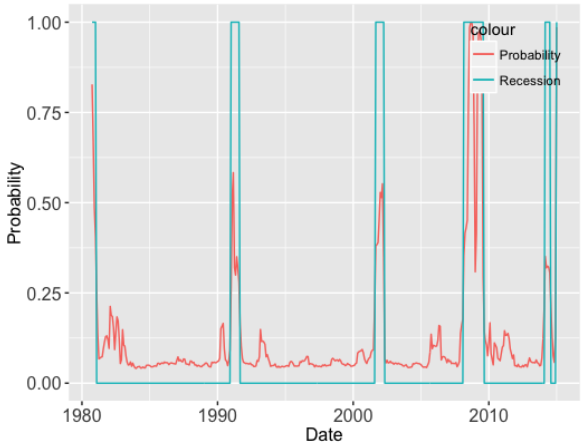
(1)

Indicator.leading



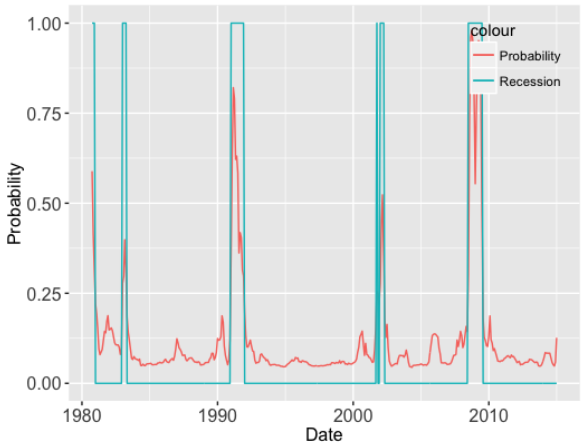
(3)

Indicator.effecti ve



(2)

Indicator.6.econo mic



(4)

Indicator.all

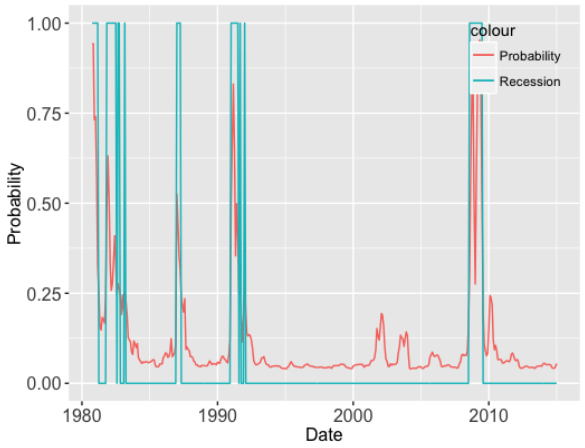


Table 6-8 Logistic regression of economic recessions in the first half of full sample from September 1980 to October 1997 - Odds ratios

Logistic regression model for explaining recession in relation to macro indcaior is shown as follows $L_t = \ln(P_t/1 - P_t) = \beta_1 + \beta_2 X_{t-1} + u_t$, where P_t and $1 - P_t$ are the probability of having recession and having no recession at date t, respectively. $P_t/(1 - P_t)$ is the odds ratio in favour of occurance of the recession - the ratio of the probabilty that a recession will occure to the probability that it will not occur. X_t is is an independent variable (representing six macro indicators) at date t. The macro indicators in the column from (1) to (6) are constructed by six groups of factors, which are all the factors, leading factors, procyclical factors, factors in the six economic groups, effective factors, and factors related to macro and business cycle conditions, respectively. Since the construction of the indicator has already concerned one-month ahead prediction, then the indicator should not be lagged by one month again. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. Standard errors are reported in parentheses underneath the coefficients.

	Recession (=1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator.all	1.162***					
	-0.03					
Indicator.leading		2.172***				
		-0.16				
Indicator.procyclical			1.342***			
			-0.062			
Indicator.6.economic				1.185***		
				-0.034		
Indicator.effective					5.484***	
					-0.322	
Indicator.BC						1.299***
						-0.053
Constant	0.023	0.027	0.026	0.025	0.022	0.023
	-0.441	-0.431	-0.436	-0.431	-0.466	-0.45
Observations	206	206	206	206	206	206
Pseudo R-squared	0.0781	0.0699	0.1383	0.1124	0.1704	0.1241
Log Likelihood	-49.282	-52.442	-46.118	-47.379	-47.239	-48.674
Akaike Inf. Crit.	102.564	108.884	96.236	98.759	98.478	101.348

Table 6-9 Measures of goodness of fit tests for the models in Table 6-8

This table covers three types of tests for measures of goodness of fit tests for the logistic regression. Panel A shows the results of both deviance and Pearson test Panel B shows the results of specification test. The critical value of chi-square in Panel A is calculated based on degree of freedom 410 at 95% significant level. Decision means whether the null hypothesis of logit model fitting correctly is rejected or not. *p<0.1; **p<0.05; ***p<0.01. Standard errors are reported in parentheses underneath the coefficients.

Panel A Deviance and Pearson Goodness-of-Fit Statistics

Indicator	test	chiSq	df	Critical value	Decision
Indicator.all	Pearson	186.2657	205	239.4034	Not reject
	Deviance	136.7007	205	239.4034	Not reject
Indicator.leading	Pearson	186.0645	205	239.4034	Not reject
	Deviance	137.9113	205	239.4034	Not reject
Indicator.procyclical	Pearson	174.3626	205	239.4034	Not reject
	Deviance	127.7774	205	239.4034	Not reject
Indicator.6.economic	Pearson	176.6156	205	239.4034	Not reject
	Deviance	131.6143	205	239.4034	Not reject
Indicator.effective	Pearson	166.7661	205	239.4034	Not reject
	Deviance	123.006	205	239.4034	Not reject
Indicator.BC	Pearson	170.2823	205	239.4034	Not reject
	Deviance	129.874	205	239.4034	Not reject

Panel B Specification Test

	Recession (=1)					
	(1)	(2)	(3)	(4)	(5)	(6)
hat	0.928***	0.920***	0.932***	0.928***	0.945***	0.943***
	-0.181	-0.191	-0.17	-0.166	-0.15	-0.2
hatsq	-0.062	-0.054	-0.06	-0.075	-0.097***	-0.039
	-0.075	-0.068	-0.049	-0.061	-0.03	-0.073
Constant	0.234	0.143	0.236	0.324	0.441	0.12
	-0.55	-0.492	-0.544	-0.557	-0.44	-0.529
Observations	206	206	206	206	206	206
Log Likelihood	-48.981	-52.187	-45.628	-46.783	-44.169	-48.55
Akaike Inf. Crit.	103.962	110.374	97.257	99.565	94.338	103.101

Figure 6-4 ROC curve for the logistic model as classifiers on the data from September 1980 to October 1997

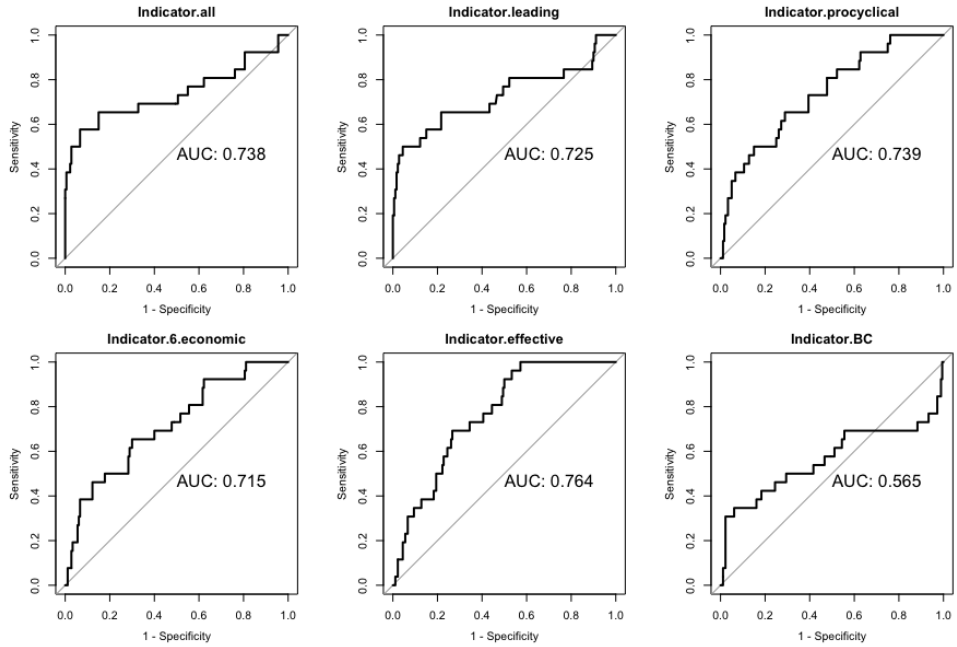


Table 6-10 Out-of-sample estimation results for various probit models: Percentage of incorrect prediction of recession

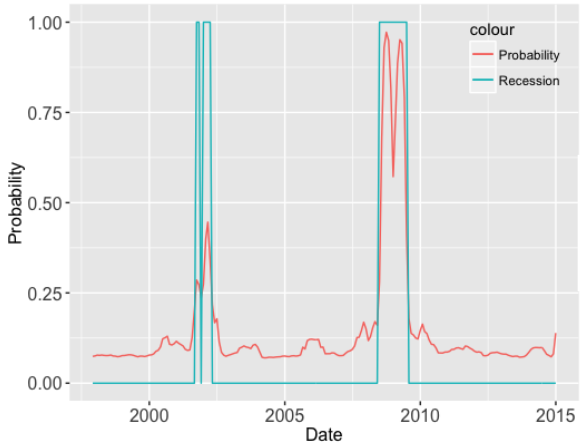
Out-of-sample forecasts are estimated using the first half of the sample, from September 1980 to October 1997, and then evaluated using the second half of the sample, from November 1997 to December 2014. Finally, we compute the predictions for the September 1980 - October 1997 period and compare them to the actual movements of the market over that time period.

	Test set error rate	
	50% Threshold of classifying recessions	25% Threshold of classifying recessions
Indicator.all	0.62	0.58
Indicator.leading	0.77	0.65
Indicator.procyclical	**0.58**	**0.46**
Indicator.6.economic	**0.58**	**0.46**
Indicator.effective	**0.46**	**0.23**
Indicator.BC	0.62	0.58

Figure 6-5 Out-of-sample recession probabilities implied by the probit model, with selective macro indicators based on classification rate 25%

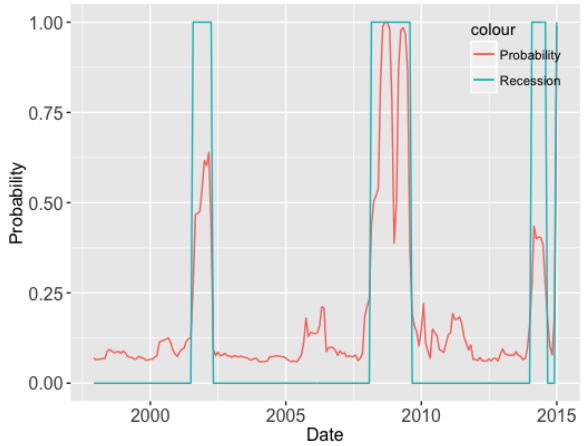
(1)

Indicator. procyclical



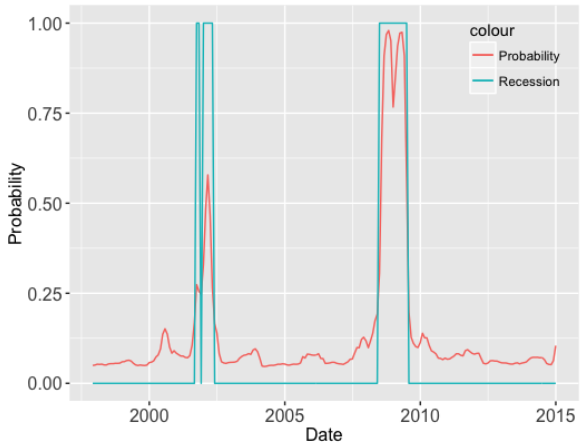
(3)

Indicator.effecti ve



(2)

Indicator.6.econo mic



(4)

Indicator.all

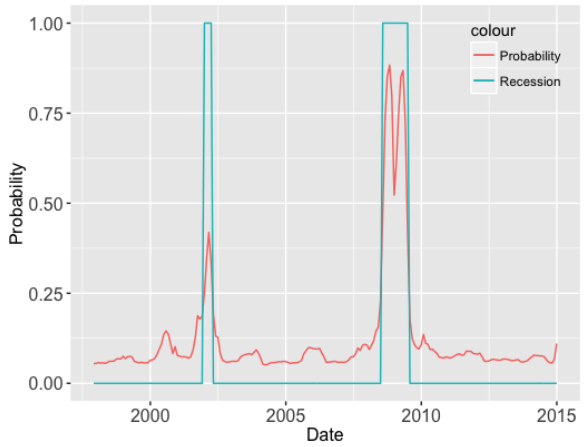


Table 6-11 Iterative out-of-sample estimation results for various probit models: percentage of incorrect prediction of recession

The iterative out-of-sample uses the sample up to December 1997. Then these estimates are used to form projections, say 1 month ahead. In this case, the projection would apply to the January 1998. After adding one more month to the estimation period, the procedure is repeated. That is, data up to the next month of January 1998 are used to make a projection for the next month of January 1998. In this way, the procedure mimics what a statistical model would have predicted with the information available at any point in the past. Data that became available subsequent to the prediction date are not used to estimate or to predict recessions.

	Test set error rate	
	50% Threshold of classifying recessions	25% Threshold of classifying recessions
Indicator.all	**0.58**	0.58
Indicator.leading	0.69	0.58
Indicator.procyclical	**0.58**	**0.46**
Indicator.6.economic	**0.58**	**0.46**
Indicator.effective[35]	**0.35**	**0.23**
Indicator.BC	**0.58**	0.54

[35] I also use an iterative approach to select the effective predictors based on the iterative out-of-sample being up to December 1997, and then construct the macro indicators. In this case, say 1 month ahead, I initially lag all the macro factors by 1 month, and then use Lasso based on different training samples to select the EPs for explaining historical recessions from February 1980 to December 1997. Using these EPs is to construct the macro indicators. The constructed indicator and recessions with period up to December 1997 is used to do estimation and prediction for the next month January 1998. The procedure is repeated. Data up to the next month of January 1998 are used to make a projection for the next month of January 1998. For example, for predicting the recession in February 1998, I add one more month to the estimation period and use Lasso to reselect the EPs for the expanded sample from February 1980 to January 1998. I reconstruct the indicator based on new set of EPs, and then I do the prediction by using this new indicator and recessions for the next month, February 1998. In practice, I use four different training samples such as 65%, 70%, 75%, and 80%. The following table shows the results for the test set error rate in terms of incorrectly predicting economic recessions. Basically, the performance of using the EPs based on 65% and 70% has the best performance of predicting economic recession in one-month ahead. In particular, even the indicator constructed by EPs that are extracted by training sample size 70% has remarkable performance in using 50% to classify recessions. The correct prediction rate is 62%, which is same as using 25% to classify recessions. Likewise, the performance of using 25% for classifying recessions outperforms that of using 50% except using training sample of 70%. Since this study mainly focuses on the macro factors that are not extracted by Lasso, then these findings are not discussed in detail. However, in the future, this will be regarded a new topic in a paper.

Sizes of training samples	Test set error rate	
	50% Threshold of classifying recessions	25% Threshold of classifying recessions
60%	0.50	0.42
65%	0.50	**0.38**
70%	**0.38**	**0.38**
75%	0.50	0.46
80%	0.58	0.46
85%	0.58	0.42

6.4 Discussions

(1) Why do several indicators have downward trend before the occurrence of the 1980s recession? Why is it that the turning points of major indicators in the recessions in July 1981 are not the cyclical trough?

From the findings in the graphical analysis, I find that all the indicators have experienced downward trend before occurrence of recession in July 1981.[36] However, several indicators have different trends during the recession period. Specifically, the indicator constructed by leading factors peaked one month earlier than the other indices. Additionally, after official

[36] Both recessions were caused for following reasons: anti-inflation policies undertaken by Federal Reserve, with employment and output displaying the largest business cycle declines in the post-World War II period. The inflation mounted before FRED defined early recession in 1980, which is triggered by two main reasons. In the 1960s and 1970s, policymakers and economists depend on the trustworthy of higher inflation that can reduce unemployment ratio as shown in the Phillips curve. These decision makers implement the "stop-go" monetary policy during the 1970s in order to fight against the high unemployment and high inflation. During the "go" periods, Federal Reserve lowered interest rates to loosen the money supply and target lower unemployment. During the "stop" periods, when inflation mounted, Federal Reserve would raise interest rates to reduce inflationary pressure. However, these policies are not stable in the trade-off between "stop" and "go" in the long-run period, which results in both higher inflation and higher unemployment ratio in the middle of 1970s. Meanwhile, the energy crisis in the middle of 1970s in the oil production also was a major contributor to rising inflation then, directly and through its negative effect on potential output.

Consequently, the economy fell into the recession in 1982 and labour market conditions deteriorated throughout the year. The unemployment rate, already high compared with historical standards in the beginning of the recession in middle of 1981, arrived at 10.8 percent at the end of 1982. More significantly, most industries cut the jobs and almost all sectors of the economy were affected adversely by the economic contraction, including the generally more stable services sector (Urquhart and Hewson, 1983).

Since inflation reflects the future price of products from market participants, then the changes of this index have large effect on the whole economy. Goodfriend and King (2005) argue that "inflation scares could be triggered by economic data, political developments or any news thought to have predictive content for negative shocks to potential output and inflation in the future. Thus, inflation scares can take on a life of their own. When this happens, inflation scares may become a possibly independent source of shocks to the economy, separate from actual shocks to potential output". Then higher inflation may slow down the whole economy. For example, Wadhwani (1986) identifies that higher inflation implies higher liquidation rates and default premia, which means that higher inflation may lead to more bankruptcies.

In this context, Federal Reserve tightened the interest rate, which further generate effects on the real economy activity including additional expectational elements, since optimizing theories of consumption and investment suggest the importance of this feature and because both consumption and investment appear to be substantially influenced by expectations in the U.S. economy (Goodfriend and King, 2005). These influences can be reflected by various macro factors.

It is clear that the index constructed by macro factors can capture the rapid changes of economic conditions. Interestingly, there are three recession indices that are more volatile in this recession in 1981, which are the indices constructed by all the factors, leading factors, and factors from six economic groups. In addition, the indices built by leading factors reach the peak in the beginning of NBER recession and then it deceases continually. It can be interpreted that the variability among leading factors falls down earlier than that among most macro factors. In fact, this corresponds to intuition.

NBER cyclical trough defined in November 1982, only this indicator continually decreases further to the low level; however, the other indicators, such as the indicator constructed by all the factors, procyclical factors, factors from six economic groups, and macro and business cycle factors, had another small peak following this recession. These findings can be interpreted as follows:

The first recession occurred in January 1980 and lasted only six months.[37] One year later, the second recession occurred in July 1981 and lasted one year and four months. All the indicators have experienced downward trend and they did not decrease to the horizontal bottom lines, suggesting that although the majority of factors did not tend to vary together, there are some factors that still fluctuated at the same time. This downward trend shows that the economy was in a recovery stage.

Westcott and Bednarzik (1981) propose that the economic recovery after the official end of the recession established in July 1980 is largely doubtful. They illustrate that "the end of the year again showed a rise in the interest rates to levels that prevailed earlier. While employment had increased steadily over the second half, it had not improved sufficiently in the housing and auto manufacturing industries-hardest hit in the downturn. Instead, fourth quarter figures for housing sales and auto purchases were relatively weak, giving little hope for a strong improvement in employment in these industries. Also, the unemployment rate had shown very little improvement from the recession high."

Additionally, Steven Rattner (1981) reports that "top officials at the Federal Reserve Board, including its chairman, Paul A. Volcker, say that their policy of reducing the expansion of money and credit will mean little or no economic growth in 1981 and continuing high interest rates". After failing to gain traction during the weak and brief recovery from the 1980 downturn, weakness in manufacturing and housing caused by rising interest rates began to have an expanded effect on related sectors beginning in the middle of 1981 (Bednarzik et al., 1982). Thus, all the indicators had upward trend until December 1981/January 1982, since macro factors varied greatly compared with before the recession. The turning point of

[37] This study does not plot this part in Figure 6-1 since the results are only from September 1980 to December 2014 after calculating indicators.

December 1981/January 1982 is due to some macro factors starting to return to normal level. The unemployment rate as a coincident factor reached 8% in October in 1981, the highest in 6 years, and then declined 0.3% in January 1982. CPI as the leading factor was 6% in November 1981, the lowest rate since 1977. In December, CPI was only 4.8% and PPI reduced to 3.6% from 15.6% in May 1981. These changes may show why there was a downward trend after December 1981/January 1982 (Goodfriend and King, 2005).

As to the reason why four indicators constructed by all the factors, procyclical factors, and macro and business cycle factors peaked again after recession ended in November 1982, Goodfriend and King (2005) explain that although Fred Reserve successfully reduced inflation from over 10 percent to around 5 percent, the deep 1981-82 recession bottomed out in November 1983, indicating that many economic factors did not return to normal level until November 1983 and then the variability in the recovery period was still high after 1982. It is interesting that all the indicators except that constructed by effective factors turn to be stable along their horizontal line in November 1983. This phenomenon can be interpreted by effective factors that have more predictive content for historical recessions.

(2) Why does the indicator constructed by the leading factors behave in a different way in the past 35 years? For example, there was a salient peak around 1997 compared with the other indicators. Likewise, this indicator also has several clear peaks occurring in the other cyclical cycles, such as from November 1982 to July 1990, from November 2001 to December 2007 and from June 2009 to December 2014.

This behaviour can be interpreted from several perspectives as regards the reason why this indicator is seemingly ahead of the other indicators. For instance, the peak occurred in December 1981 rather than January 1982; leading factors tend to move direction before the business cycle occurs and the timing comparisons between the cyclical turn in the indicators and business cycle turn is one month or more, such as new orders, housing starts, and consumer sentiment. Lagging factors tend to shift direction after the business cycle occurs or coincident series occur, such as interest rate, unit labour cost, and cost of resources. For the timing comparisons, these factors follow the business cycle turn by one month or more. Coincident factors such as employment, production, and personal income etc. are the series

that measure aggregate economic activity which are used to define the business cycle. Coincident factors are special in terms of timing comparisons since they may follow or lead turning points in the business cycle by three months or less, including that they completely coincide with the business cycle. It seems then that the indicator covering leading factors can provide information earlier than the other indicators covering a mix of factors. In contrast, these lagging or coincident factors explain why there are peaks behind the recession in July 1981 since they vary slightly together. The other indicators are a mix since they cover three types of cyclical factors including lagging factors, coincident factors, and leading factors.

It should be noted that this indicator seemingly loses the ability of measuring economic recession in advance in the following periods. Early economic literature has discussed that leading factors may become lagging factors in certain periods since the economic system is a kind of evolutionary system, and the factors are also dynamic thereby explaining the corresponding economic environment.

Meanwhile, this indicator was more volatile than the others. Zarnowitz (1992a) explains that "the leading series represent largely flow and price variables that are highly sensitive to the overall cyclical influences but also to shorter random disturbances; hence they show large cyclical rises and declines but also high volatility". Before Black Friday occurring, leading factors varied greatly together, which provided some signals of the crash in the stock market. This result corresponds to the empirical result; that is, the leading indicator has relatively poor performance compared to the other good indicators. Stock and Watson (2003) determine that the majority of leading indicators lose the ability of predicting the economic recession in March 2001.

(3) Why are all the indicators' peaks lag of the official ending point of recession starting in March 2001?

As discussed earlier, before the recession occurring in March 2001, these indicators peaked in September 2000 and decreased. When the recession ended in November 2001, all the indicators still experienced an increased trend until March 2002 and then they experienced downward trends in the following period. It can be summarized that even though recession defined by Fred Reserve ended, the majority of macro factors obviously still fluctuated and

they tended to vary together, suggesting that the economy was not stable at that moment. Until March 2002, the economy tended to recover since the fluctuation among these factors became slighter.

This has support from the economic literature. First, the recession that started in March 2001 differed from other recessions of the past thirty years. As discussed previously, the past two recessions in the early 1980s resulted from monetary policies used by the Federal Reserve. The recession of 1990 was caused by the consumption shock and this shock is related to the response of consumers in the face of the uncertainty raised by Iraq's invasion of Kuwait and the 1979 oil crisis. The cause of the 2007 recession was the housing boom and bust which in turn led to financial turmoil in the US. Taylor (2009) argues that this financial crisis was triggered by government actions and interventions through excess monetary policies. It can be assumed then that this cause of financial crisis from December 2007 to June 2009 is similar to the twin recessions of the early 1980s. Stock and Watson (2003) explain that "the recession of 2001 started neither in the shopping mall nor in the corridors of the Federal Reserve Bank, but in the boardrooms of corporate America as businesses sharply cut back on expenditures – most notably, investment associated with information technology – in turn leading to declines in manufacturing output and in the overall stock market".

In addition, recessions occur since economic developments are of insufficient magnitude to alter expenditures by households and firms so as to decrease aggregate demand, output, and employment (Kliesen, 2003). It is suggested that when there is a recession, many macro factors have the same reactions which are expressed as a change in their direction from the stable stage. This implies deterioration of economic conditions. Thus, interactions among various factors are quite high. This is the case in the early 1980s crises, the 1990 crisis, and the 2007 crisis. In contrast, the recession occurring at the onset of March 2001 does not follow the same path. Kliesen (2003) clarifies that the unique feature of this recession is its mildness, which is reflected by the decline in real GDP. Real GDP even increased 0.2 percent from the first quarter of 2001 to the fourth quarter of 2001. That is, during the recession period, real GDP still grew. Stock and Watson (2003) identify that "the economy gained substantial strength in the final quarter of 2001 and throughout 2002, and all the monthly indicators were growing by December 2001."

For example, although there was a decline in industrial production that peaked in June 2000 and manufacturing and trade sales fell during the first quarter of 2001, employment did not reach a peak until March 2001 as the starting point of the 2001 recession defined by the Federal Reserve. In addition, personal income had a small decline from a cyclical peak in November 2000 and a cyclical trough declined by 1.5 percent over the next twelve months, indicating the unusual fact that productivity growth remained strong throughout this crisis. Therefore, the higher fluctuations among these factors appear after December 2001.

(4) Which indicator should be used for measuring economic recessions?

Based on the results from out-of-sample analysis, the best two indicators are constructed by a procyclical factors and factors from six economic groups, which outperform the other indicators except the indicator constructed by effective factors. For the indicator constructed by procyclical factors, even though the indicator only uses 86 factors, its performance is still quite good in terms of correctly predicting economic recessions no matter what a kind of approach is used for testing the validation of predicted values. In particular, according to Shiskin and Moore (1967), "An idea indicator would be one that changed direction only in the event of a recession or recovery". Based on this standard, this indicator is seemingly more consistent with this ideal indicator since the time series of this indicator is smoother compared with the others in Figure 6-1.

However, regarding future events that are stochastic and economic, this is still evolving with the occurrence of technical developments, changing consumer tastes, deregulations, and even other factors such as natural disasters, the best indicator should be based on using broad factors. This suggests that the indicator constructed by 106 factors from six economic groups should be used as well.

Shiskin and Moore (1967) argue that in addition to the factors from six economic groups mentioned earlier, the other factors in three group of foreign trade and payments, federal government activities, and economic activity in other countries may significantly affect the cyclical cycle's pattern, amplitude, and duration. That is, using broad factors to construct indicators theoretically has the potential ability to capture future changes in the economic system rather than using less broadones. Scheffer et al. (2012) declare that using the methods

for detecting incipient transitions from time series has a tendency to require long, high-resolution data. In another words, if more factors with high frequency are used for constructing the indicator, then it may capture early transitions caused by the increase in interactions among these factors in the macroeconomic system. From the empirical results, the indicator constructed by 106 factors can work as well as the indicator calculated by procyclical factors. In contrast, the indicator calculated by all the factors has poorer performance than them.

Helbing (2013) provides explanations for this conflicting phenomenon. Specifically, using 'big data' may provide more information to capture the unpredictable shock for the economic system; that is, using more factors increases the ability of predicting economic recessions. However, Helbing (2013) also demonstrates that too much information may lead to incorrect decisions since they have the potential to create a more opaque picture of the economic environment. Although the indicator built by all the factors has poorer performance than the best two mentioned above, it is still to be regarded as a complementary indicator in terms of forecasting future recessions.

The indicator constructed by using only 30 effective factors has the best performance in predicting recession months. However, this indicator may not work well in the future. For example, in Figure 6-1, after the beginning of 2014, there is a significant increase in this indicator, suggesting that the probability of recession occurring increases rapidly. However, there was no recession at that time. From this result, it is implied that LASSO is able to tell us mostly what happened in the past, and even provides an almost complete picture compared to the other good indicators mentioned above in the out-of-sample results.

6.5 Conclusion

Based on the advanced theory of anticipating critical transitions in a complex system proposed by Scheffer et al. (2012), this chapter constructs several macro indicators by using hundreds of macroeconomic factors to measure the changes in the macroeconomic system thereby predicting recessions from September 1980 to December 2014 in the US.

I find that the interactions among macroeconomic and financial variables measured by covariance among these variables can provide one-month ahead prediction for US economic recessions. Specifically, this study initially classifies hundreds of macro factors into six groups which consist of all the factors, leading factors, procyclical factors, factors from six economic groups, effective factors, and factors related to macro and business cycle conditions, respectively. This study generates six macro indicators by separately collecting covariance among macro factors from six groups of macro factors mentioned above. I find that the best predictive macro indicators are constructed by employing procyclical factors and factors from six economic groups based on results from both in-sample and out-of-sample analysis. In particular, following the standard of the ideal indicator determined by Shiskin and Moore (1967), the indicator constructed by procyclical factors is preferable to that constructed by the factors from six economic groups since the former is smoother than the latter.

However, regarding prediction of economic recessions in the US, the indicator generated by using broad macro factors is also able to provide predictive power for economic recessions. This finding is consistent with prior studies as regards using more factors that can detect early transitions in the system. It also corresponds to the finding in the business cycle study that using broad macro factors may capture any transitions caused by the increase of co-movement from factors that have not been considered as trustworthy factors, thereby raising the ability of capturing unpredictable changes in the macroeconomic system. The implication of these results is to provide a new quantitative approach for central bankers and policy makers to predict economic recessions one month ahead.

The current study can be extended in several directions. In this study I have focused on whether the macro indicators used to measure interactions among macroeconomic and

financial variables can provide one-month ahead prediction of US economic recessions by using standard logistic model. It is also worthwhile considering the out-of-sample performance of these indicators from one to eight quarters ahead. Another issue of interest compares the out-of-sample performance between these indicators and traditional recession risk factors, such as interest spread, stock price, and interest rates.

According to Ng (2012), of more interest is the balance between existence of publication lags of recession reference dates and moving closer to real time recession forecasting. His study assumes that a publication lag is 3 months. Future study then should be concerned with this issue in order to exploit the predictive content in each macro indicator for US economic recessions. Also it would be interesting to examine how advanced dynamic probit models improve the out-of-sample performance with each macro indicator in forecasting US recessions.

Regarding the out-of-sample estimation results, the current macro indicator can only capture around 54% recession months correctly one-month ahead except the macro indicator constructed by effective factors (the ratio of correct prediction is 0.77). The predictive power is thus limited. Future studies, besides using effective predictors to construct macro indicators for predicting future recessions, should employ other methods to measure the interactions among macro factors in order to examine whether the predicting performance can be further improved compared with using the current macro indicators and the other popular recession factors.

Finally, as Helbing (2013) argues, the economic system also has a high connection with the other elements, such as corruption, migration, climate change etc. Therefore, another future direction is how to understand and collect the information from these elements to generate new indicators in order to predict future collapse in the economic system.

CHAPTER 7

7. Conclusion

7.1 Conclusions and Implications

7.1.1 First Empirical Chapter

Even though numerous studies have contributed to developing advanced techniques, such as reduced form models to predict DPs by considering various risk determinants from the macroeconomic environment, this study finds that the performance of the default risk determinants is not stable, which reflects the dynamic in the perception of default risk. It would be cautious to suggest that using the current popular macro factors, for instance interest spread, interest rate, and the stock return, may not predict this dynamic phenomenon in default risk.

Interestingly, both bank lending conditions and monetary policy indicators have more explanatory power for default risk in the US in comparison with the other macroeconomic factors. The former reflects the degree of monetary liquidity; the latter shows the price level in the economy. This result mirrors the interactions between monetary policies and credit market stability; that is, the implementation of monetary policy may result in severe defaults thereby leading to instability of the credit market, especially that defaults exhibit higher countercyclicality.

For decision makers, it is cautious to make an appropriate monetary policy shaping risk-taking to simultaneously stabilise the economy and the financial market in practice. In particular, several monetary tools related to borrowing of depositary institutions from the central bank, reserve requirements of depositary institutions, and the monetary base should

be considered prudent since they have strong predicting power for corporate defaults in the past.

Meanwhile, after 2001, the money supplied by non-bank groups, such as money market funds, has more predictive power than the monetary policy indicators. This impact on corporate default has experienced gradual increase for 35 years. In contrast, relevant predictors, such as the borrowing rate from central banks, has experienced the opposite trend to money market funds. Particularly in the period of financial crisis occuring in 2007 monemary fund reaches the highest level in the history in terms of explaining corporate defaults. This finding semmingly provides emprical support of strong association between monetary market fund and financial stability. Money market fund, as a supplementary tool, can provide liquidity to the market, which is able to influence the effects of monetary policies. Thus, economic policy makers need to be concerned about how to achieve policy execution objectives by allowing for the other elements, such as that a sufficiently wide set of nonbank counterparties has access to the capital market against government collateral.

7.1.2 Second Empirical Chapter

This chapter introduces an advanced method from natural sciences by constructing macro indicators for measuring changes in the macroeconomic system. By employing 114 factors, this study generates various indicators for measuring changes in the macroeconomic conditions. Based on a logit model, this study uses various macro indicators for predicting US historical recessions one-month ahead. Since this study is ex post, then the publication lags in macro factors and NBER recession dates are not considered.

I find that covariance among macro factors can provide a sign of the occurrence of recession months in the US, indicating that interactions among these factors can be used to imply the risk of unwanted collapse in the macroeconomic system. Based on out-of-sample estimation results, three macro indicators that are constructed by effective factors, procyclical factors and the factors from six economic groups have better performance than the others in predicting US recessions one month ahead.

In particular, even though the macro indicators calculated by a small number of effective predictors can correctly forecast 80% of historical recessions, this indicator may not capture future collapse in the economic system. In this context, the last two indicators mentioned are more preferable. If following the definition of standard ideal indicator proposed by Shiskin and Moore (1967), the indicator constructed by procyclical factors is preferable.

I find that this best indicator is only constructed by 86 factors and the indicators constructed by the factors from six economic groups cover 106 macro factors, which is less than the indicators constructed by all the factors. This finding implies that too much data can generate an opaque picture in the economic environment, which is consistent with Helbing (2013). However, considering the evolving economic system, this indicator built by all the factors should be regarded as a complementary indicator in terms of forecasting future recessions.

7.2 Limitations

The limitations of this thesis relate to the second empirical chapter about predicting the US economic recessions. There are three issues as follows:

First, I use past 8 months information with 8 monthly observations (such as, $t_{t-7}, \dots, t_t$) to measure the current macroeconomic conditions in the current month t. Actually, following the method described in the paper done by Scheffer et al. (2012), the current marcoeocnomic conditions t should be calculated by using the information in the period t. For example, if I calcuate the macroeocnomic conditions in the September 2016, then I need to use all the information such as daily observations, even high frequecy observations from the current month to construct a value for representing this month's macroeconomic condition. However, in reality, there is no high frequency data from macroeocnomic time series. Normally the highest frequecy for macroeocnomic variables is monthly, and for some variables, they are annually and quarterly, respectively. Thus, I could only use past 8 months information ($t_{t-7}, \dots, t_t$) to calculate the macroeconomic conditions in the date t. This is the limitation of using the method of capturing critical tranistions in the system.

Second, this research is a typical ex post study, which does not concern the problem of announcement with a delay in recession and macro factors in reality. That is, I regard that all

the data can be known in the standard date and no delay happens in each variable. This is a limitation since in reality this assumption cannot be achieved. Ng (2012) argues that financial variables and macroeconomic variables are normally available with 1- or 2-months delay before the figures are officially published. This is the same in the case of the recession indicator from the Federal Reserve. Ng (2012) mentions that the average announcement delay from 1980 to 2008 is around 9 months. Therefore, this ex post study cannot provide a full picture of how these macro indicators are capable of forecasting economic recessions.

Third, even though the macro indicators, by measuring the covariance among macro factors, can make one-month ahead prediction, the correct classification rate for classifying the recessions in the out-of-sample is only around 65%, which is 15% larger than 50% that is the probability of correctly predicting recessions without using any techniques or methods. Thus, the predictive ability for these macro indicators is limited in practice.

7.3 Future Research

For future study, there are two main directions, based on my two empirical chapters.

7.3.1 First Direction

For the first empirical chapter, it would be interesting to investigate the dynamic in EPs by using different forecast horizons in independent variables that are hundreds of macro factors. The current results imply that bank lending conditions and monetary policy indicators have relatively consistent effect on corporate defaults after lagging all the macro factors by 3 months. It can be assumed that if 6-month and 12-month are used for lagging the factors, the results may be different. The policy makers may then better understand how the changes in macroeconomic environment influence the credit market over time.

Additionally, of more interest to central banks is that I can further use causality methods to examine whether there is a causal relationship between corporate defaults and monetary policy variables, such as consumer index, monetary base, and borrowing rate from central banks. This research may give a complete story of how monetary policies impact on corporate defaults.

7.3.2 Second Direction

There are several sub-directions that can be taken based on the second empirical chapter. Future work should further investigate the predictive power for each macro indicator by using different forecast horizons, such as 3-month, 6-month, 9-month, and 12-month etc., following prior studies. In addition, it would concern the announcement of macroeconomic variables and recession date with a delay. Following the suggestions given by Ng (2012), 3 months are used for belancing the publication lags in both recession dates and macroeconomic variables. That is, all explanatory variables are known 3 months ahead of the recession dummy variable.

In order to improve forecasting performance, future study should be concerned with two issues. The first issue is to use more powerful indicators by measuring interactions among macroeconomic and financial variables. As Scheffer et al. (2012) suggested, variance and correlation can also be used for gauging the interactions among system components. It is worth using these new generated indicators to examine the predictive contents for the US recessions. The second issue is to employ more sophisticated models which can handle auto correlated errors and multiple break points across business cycles. For example, I can use dynamic probit models and then compare the forecasting performance among different models with proposed recession factors in order to determine whether the interactions can provide earlier warning signs and more accurate out-of-sample estimation for predicting economic recessions than the current popular approaches.

Regarding the research finding from Helbing (2013), the other elements, such as corruption, migration, climate change etc. have high correlation with the economic system. Thus, another possible concern is to generate new indicators that can allow for the influence from these sources to lead to economic system collapse in the future.

8. References

Acharya, V., Davydenko, S. A. & Strebulaev, I. A. 2012. Cash holdings and credit risk. *Review of Financial Studies,* 25**,** 3572-3609.

Acharya, V. V. & Carpenter, J. N. 2002. Corporate bond valuation and hedging with stochastic interest rates and endogenous bankruptcy. *Review of Financial Studies,* 15**,** 1355-1383.

Agrawal, A. K. & Matsa, D. A. 2013. Labor unemployment risk and corporate financing decisions. *Journal of Financial Economics,* 108**,** 449-470.

Aikman, D., Galesic, M., Gigerenzer, G., Kapadia, S., Katsikopoulos, K. V., Kothiyal, A., Murphy, E. & Neumann, T. 2014. Taking uncertainty seriously: Simplicity versus complexity in financial regulation. *Bank of England Financial Stability Paper*.

Allen, S. L. 2012. *Financial Risk Management: A Practitioner's Guide to Managing Market and Credit Risk*, Wiley.

Altman, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance,* 23**,** 589-609.

Anson, M. J. P., Fabozzi, F. J., Choudhry, M. & Chen, R. R. 2004. *Credit Derivatives: Instruments, Applications, and Pricing*, Wiley.

Azizpour, S. & Giesecke, K. 2008. Self-exciting corporate defaults: contagion vs. frailty. *Manuscript, Stanford University*.

Azizpour, S., Giesecke, K. & Schwenkler, G. 2013. Exploring the sources of default clustering. Stanford University working paper series.

Azizpour, S., Giesecke, K. & Schwenkler, G. 2014. Exploring the sources of default clustering. Stanford University working paper series.

Bai, J., Collin-Dufresne, P., Goldstein, R. & Helwege, J. 2012. Is Credit Event Risk Priced? *Columbia University DP*.

Barro, R. J. 1989. Economic growth in a cross section of countries. National Bureau of Economic Research.

Bcbs 2010. Basel III: A global regulatory framework for more resilient banks and banking systems Basel Committee on Banking Supervision.

Beaver, W. H. 1968. The information content of annual earnings announcements. *Journal of Accounting Research***,** 67-92.

Bednarzik, R. W., Hewson, M. A. & Urquhart, M. A. 1982. The employment situation in 1981: new recession takes its toll. *Monthly Labor Review***,** 3-14.

Bennin, R. 1980. Error rates in CRSP and COMPUSTAT: A second look. *Journal of Finance,* 35**,** 1267-1271.

Bernanke, B., Gertler, M. & Gilchrist, S. 1994. The financial accelerator and the flight to quality. National Bureau of Economic Research.

Bernanke, B. S. 1993. Credit in the macroeconomy. *Quarterly Review-Federal Reserve Bank of New York,* 18**,** 50-50.

Bernanke, B. S. & Gertler, M. 1995. Inside the black box: the credit channel of monetary policy transmission. National bureau of economic research.

Bhamra, H. S., Fisher, A. J. & Kuehn, L.-A. 2011. Monetary policy and corporate default. *Journal of Monetary Economics,* 58**,** 480-494.

Bharath, S. T. & Shumway, T. 2008. Forecasting default with the merton Distance to Default Model. *Review of Financial Studies,* 21**,** 1339-1369.

Black, F. & Cox, J. C. 1976. VALUING CORPORATE SECURITIES: SOME EFFECTS OF BOND INDENTURE PROVISIONS. *Journal of Finance,* 31**,** 351-367.

Black, F. & Scholes, M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy,* 81**,** 637-654.

Blinder, A. S. & Maccini, L. J. 1991. The resurgence of inventory research: what have we learned? *Journal of Economic Surveys,* 5**,** 291-328.

Blöchlinger, A. 2012. Validation of Default Probabilities. *Journal of Financial and Quantitative Analysis,* 47**,** 1089-1123.

Borio, C. E., Erdem, M., Filardo, A. J. & Hofmann, B. 2015. The costs of deflations: a historical perspective. *BIS Quarterly Review March*.

Burnside, C., Eichenbaum, M. & Rebelo, S. 2001. Hedging and financial fragility in fixed exchange rate regimes. *European Economic Review,* 45**,** 1151-1193.

Campbell, J. Y., Hilscher, J. & Szilagyi, J. 2008. In search of distress risk. *Journal of Finance,* 63**,** 2899-2939.

Charitou, A., Dionysiou, D., Lambertides, N. & Trigeorgis, L. 2013. Alternative bankruptcy prediction models using option-pricing theory. *Journal of Banking and Finance,* 37**,** 2329-2341.

Chauvet, M. & Potter, S. 2005. Forecasting recessions using the yield curve. *Journal of Forecasting,* 24**,** 77-103.

Chava, S. & Jarrow, R. A. 2004. Bankruptcy prediction with industry effects. *Review of Finance,* 8**,** 537-569.

Christiansen, C., Eriksen, J. N. & Møller, S. V. 2014. Forecasting US recessions: The role of sentiment. *Journal of Banking and Finance,* 49**,** 459-468.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, L. Erlbaum Associates.

Collin-Dufresne, P., Goldstein, R. & Helwege, J. 2003. Are jumps in corporate bond yields priced? Modeling contagion via the updating of beliefs. *University of California Berkeley Working Paper*.

Collin-Dufresne, P., Goldstein, R. S. & Helwege, J. 2010. Is Credit Event Risk Priced? Modeling Contagion via the Updating of Beliefs. National Bureau of Economic Research.

Collin‐Dufresne, P., Goldstein, R. S. & Martin, J. S. 2001. The determinants of credit spread changes. *Journal of Finance,* 56**,** 2177-2207.

Couderc, F., Renault, O. & Scaillet, O. 2008. Business and financial indicators: What are the determinants of default probability changes? : National Centre of Competence in Research Financial Valuation and Risk Management.

Crosbie, P. & Bohn, J. 2002. Modelling Default Risk. KMV working paper.

Curdia, V. & Woodford, M. 2010. Credit spreads and monetary policy. *Journal of Money, Credit and Banking,* 42**,** 3-35.

Das, S. R., Duffie, D., Kapadia, N. & Saita, L. 2007. Common failings: How corporate defaults are correlated. *Journal of Finance,* 62**,** 93-117.

Davydenko, S. A. When do firms default? A study of the default boundary. A Study of the Default Boundary (November 2012). EFA Moscow Meetings Paper, 2012.

Davydenko, S. A., Strebulaev, I. A. & Zhao, X. 2012. A Market-based study of the cost of default. *Review of Financial Studies,* 25**,** 2959-2999.

De Servigny, A. & Renault, O. 2002. Default correlation: empirical evidence. *Standard and Poor's*.

Delianedis, G. & Geske, R. 2001. The component of corporate credit spreads: Default. *Recovery, Tax, Jumps, Liquidity, and Market Factors", eScholarship Repository, University of California*.

Dewatripont, M., Rochet, J. C., Tirole, J. & Tribe, K. 2010. *Balancing the Banks: Global Lessons from the Financial Crisis*, Princeton University Press.

Dichev, I. D. 1998. Is the risk of bankruptcy a systematic risk? *Journal of Finance,* 53**,** 1131-1147.

Domencich, T. A. & Mcfadden, D. 1975. *Urban travel demand: a behavioral analysis*, North-Holland.

Dueker, M. J. 1997. Strengthening the Case for the Yield Curve as a Predictor of US Recessions. *Review,* 79**,** 41-51.

Duffie, D. 2005. Credit risk modeling with affine processes. *Journal of Banking and Finance,* 29**,** 2751-2802.

Duffie, D. 2011. Measuring Corporate Default Risk. *Clarendon Lectures in Finance***,** 128.

Duffie, D., Eckner, A., Horel, G. & Saita, L. 2009. Frailty correlated default. *Journal of Finance,* 64**,** 2089-2123.

Duffie, D. & Lando, D. 2001. Term structures of credit spreads with incomplete accounting information. *Econometrica,* 69**,** 633-664.

Duffie, D., Saita, L. & Wang, K. 2007. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics,* 83**,** 635-665.

Duffie, D. & Singleton, K. J. 2012. *Credit Risk: Pricing, Measurement, and Management*, Princeton University Press.

Eichengreen, B. & Hausmann, R. 1999. Exchange rates and financial fragility. National bureau of economic research.

Elizalde, A. 2006. Credit risk models II: Structural models. *Documentos de Trabajo (CEMFI)***,** 1.

Eom, Y. H., Helwege, J. & Huang, J.-Z. 2004. Structural Models of Corporate Bond Pricing: An Empirical Analysis. *Review of Financial Studies,* 17**,** 499-544.

Erlenmaier, U. & Gersbach, H. 2013. Default Correlations in the Merton Model. *Review of Finance,* 18**,** 1775–1809.

Estrella, A. & Mishkin, F. S. 1996. The yield curve as a predictor of US recessions. *Current Issues in Economics and Finance,* 2**,** 1-6.

Estrella, A. & Mishkin, F. S. 1998. Predicting US recessions: Financial variables as leading indicators. *Review of Economics and Statistics,* 80**,** 45-61.

Figlewski, S., Frydman, H. & Liang, W. 2012. Modeling the Effect of Macroeconomic Factors on Corporate Default and Credit Rating Transitions. *International Review of Economics and Finance,* 21**,** 87-105.

Finger, C. 1999. Conditional approaches for CreditMetrics portfolio distributions. *CreditMetrics Monitor,* 1**,** 14-33.

Fischer, E. O., Heinkel, R. & Zechner, J. 1989. Dynamic capital structure choice: Theory and tests. *Journal of Finance,* 44**,** 19-40.

Fisher, I. 1933. The debt-deflation theory of great depressions. *Econometrica: Journal of the Econometric Society***,** 337-357.

Fsoc 2012. Proposed Recommendations Regarding Money Market Mutual Fund Reform Washington: Financial Stability Oversight Council.

Gansecki, M. 2010. Statistical Analysis of Groundwater Data at RCRA Facilities—Unified Guidance. *Groundwater Monitoring & Remediation,* 30**,** 32-35.

Gertler, M. & Gilchrist, S. 1991. Monetary policy, business cycles and the behavior of small manufacturing firms. *The Quarterly Journal of Economics,* 109**,** 309-340.

Geske, R. 1977. The valuation of corporate liabilities as compound options. *Journal of Financial and Quantitative Analysis,* 12**,** 541-552.

Giesecke, K. 2004. Correlated default with incomplete information. *Journal of Banking and Finance,* 28**,** 1521-1545.

Giesecke, K., Longstaff, F. A., Schaefer, S. & Strebulaev, I. 2011. Corporate bond default risk: A 150-year perspective. *Journal of Financial Economics,* 102**,** 233-250.

Gilchrist, S. & Zakrajšek, E. 2011. Credit spreads and business cycle fluctuations. National Bureau of Economic Research.

Goldstein, R., Ju, N. & Leland, H. 2001. An EBIT ‐ Based model of dynamic capital structure. *Journal of Business,* 74**,** 483-512.

Goodfriend, M. & King, R. G. 2005. The incredible Volcker disinflation. *Journal of Monetary Economics,* 52**,** 981-1015.

Gorban, A. N., Smirnova, E. V. & Tyukina, T. A. 2010. Correlations, risk and crisis: From physiology to finance. *Physica A: Statistical Mechanics and its Applications,* 389**,** 3193-3217.

Gregoriou, G. N. & Hoppe, C. 2009. *The Handbook of Credit Portfolio Management*, McGraw-Hill.

Griffin, J. M. & Lemmon, M. L. 2002. Book–to–market equity, distress risk, and stock returns. *Journal of Finance,* 57**,** 2317-2336.

Gujarati, D. N. 2003. Basic Econometrics. 4th. New York: McGraw-Hill.

Gurný, P. & Gurný, M. 2013. Comparison of Credit Scoring Models on Probability of Default Estimation for Us Banks. *Prague Economic Papers,* 2013**,** 163-181.

Haldane, A. G. & May, R. M. 2011. Systemic risk in banking ecosystems. *Nature,* 469**,** 351-355.

Helbing, D. 2013. Globally networked risks and how to respond. *Nature,* 497**,** 51-59.

Hillegeist, S. A., Keating, E. K., Cram, D. P. & Lundstedt, K. G. 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies,* 9**,** 5-34.

Hodrick, R. J. & Prescott, E. C. 1997. Postwar US business cycles: an empirical investigation. *Journal of Money, Credit, and Banking,* 29**,** 1-16.

Hull, J., Predescu, M. & White, A. 2005. The valuation of correlation-dependent credit derivatives using a structural model. *Available at SSRN 686481*.

Idre. 2016. *Logistic Regression with Stata* [Online]. Los Angeles: UCLA Institute for Digital Research and Education.

Jaggia, P. B. & Thakor, A. V. 1994. Firm-specific human capital and optimal capital structure. *International Economic Review***,** 283-308.

James, G., Witten, D., Hastie, T. & Tibshirani, R. 2013. *An Introduction to Statistical Learning: with Applications in R*, Springer New York.

Janosi, T., Jarrow, R. & Yildirim, Y. 2002. Estimating expected losses and liquidity discounts implicit in debt prices. *Journal of Risk,* 5**,** 1-38.

Jarrow, R. & Protter, P. 2004. Structural versus reduced form models: a new information based perspective. *Journal of Investment Management,* 2**,** 1-10.

Jarrow, R. & Turnbull, S. 1992. Credit risk: Drawing the analogy. *Risk Magazine,* 5**,** 63-70.

Jones, E. P., Mason, S. P. & Rosenfeld, E. 1984. Contingent claims analysis of corporate capital structures: An empirical investigation. *Journal of Finance,* 39**,** 611-625.

Jorion, P. & Zhang, G. 2007. Good and bad credit contagion: Evidence from credit default swaps. *Journal of Financial Economics,* 84**,** 860-883.

Kang, J. & Pflueger, C. E. 2015. Inflation risk in corporate bonds. *Journal of Finance,* 70**,** 115-162.

Kashyap, A. K., Lamont, O. A. & Stein, J. C. 1992. Credit Conditions and the Cyclical Behavior of Inventories: A Case Studyof the 1981-82 Recession. National Bureau of Economic Research.

Kauppi, H. & Saikkonen, P. 2008. Predicting US recessions with dynamic binary response models. *Review of Economics and Statistics,* 90**,** 777-791.

Kim, I. J., Ramaswamy, K. & Sundaresan, S. 1993. Does default risk in coupons affect the valuation of corporate bonds?: A contingent claims model. *Financial Management,* 22**,** 117-131.

Kliesen, K. L. 2003. The 2001 recession: How was it different and what developments may have caused it? *Review Federal Reserve Bank of Saint Louis,* 85**,** 23-38.

Koopman, S. J., Kräussl, R., Lucas, A. & Monteiro, A. B. 2009. Credit cycles and macro fundamentals. *Journal of Empirical Finance,* 16**,** 42-54.

Koopman, S. J. & Lucas, A. 2005. Business and default cycles for credit risk. *Journal of Applied Econometrics,* 20**,** 311-323.

Koopman, S. J., Lucas, A. & Schwaab, B. 2011. Modeling frailty-correlated defaults using many macroeconomic covariates. *Journal of Econometrics,* 162**,** 312-325.

Lahiri, K. 2010. *Transportation Indicators and Business Cycles*, Emerald Group Publishing.

Lando, D. & Nielsen, M. S. 2010. Correlation in corporate defaults: Contagion or conditional independence? *Journal of Financial Intermediation,* 19**,** 355-372.

Lane, W. R., Looney, S. W. & Wansley, J. W. 1986. An application of the Cox proportional hazards model to bank failure. *Journal of Banking and Finance,* 10**,** 511-531.

Lang, W. W. & Nakamura, L. I. 1995. 'Flight to quality'in banking and economic activity. *Journal of Monetary Economics,* 36**,** 145-164.

Lee, S. H. & Urrutia, J. L. 1996. Analysis and prediction of insolvency in the property-liability insurance industry: a comparison of logit and hazard models. *Journal of Risk and Insurance***,** 121-130.

Leland, H. E. 1994. Corporate debt value, bond covenants, and optimal capital structure. *Journal of Finance,* 49**,** 1213-1252.

Leland, H. E. 2004. Predictions of default probabilities in structural models of debt. *Journal of Investment Management,* 2**,** 5-20.

Leland, H. E. & Toft, K. B. 1996. Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *Journal of Finance,* 51**,** 987-1019.

Lennox, C. 1999. Identifying failing companies: a re-evaluation of the logit, probit and DA approaches. *Journal of Economics and Business,* 51**,** 347-364.

Liu, S., Qi, H., Shi, J. & Xie, Y. A. 2013. Inferring default correlation from equity return correlation. *European Financial Management,* 9999**,** 1-27.

Liu, W. & Moench, E. 2016. What predicts US recessions? *International Journal of Forecasting,* 32**,** 1138-1150.

Lo, A. W. 1986. Logit versus discriminant analysis: A specification test and application to corporate bankruptcies. *Journal of Econometrics,* 31**,** 151-178.

Lo, A. W. 2004. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management,* 30**,** 15-29.

Löffler, G. 2004. Ratings versus market-based measures of default risk in portfolio governance. *Journal of Banking and Finance,* 28**,** 2715-2746.

Longstaff, F. A. & Schwartz, E. S. 1995. A simple approach to valuing risky fixed and floating rate debt. *Journal of Finance,* 50**,** 789-819.

May, R. M., Levin, S. A. & Sugihara, G. 2008. Complex systems: Ecology for bankers. *Nature,* 451**,** 893-895.

McDonald, C. G. & Van De Gucht, L. M. 1999. High-yield bond default and call risks. *Review of Economics and Statistics,* 81**,** 409-419.

Mcfadden, D. 1973. Conditional logit analysis of qualitative choice behavior.

McNeil, A. J., Frey, R. & Embrechts, P. 2010. *Quantitative Risk Management: Concepts, Techniques, and Tools*, Princeton university press.

Merton, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance,* 29**,** 449-470.

Mitchell, W. C. & Burns, A. F. 1938. Statistical indicators of cyclical revivals. *Statistical Indicators of Cyclical Revivals.* NBER.

Moody, S. 2015. Moody's Rating Symbols & Definitions. *Moody's Investors Service, Report*, 79004**,** 46.

Moore, G. H. 1961. Statistical indicators of cyclical revivals and recessions. *Business Cycle Indicators, Volume 1.* Princeton University Press.

Ng, E. C. 2012. Forecasting US recessions with various risk factors and dynamic probit models. *Journal of Macroeconomics,* 34**,** 112-125.

Nielsen, M. S. 2011. *Essays on Correlation Modelling*, Copenhagen Business School Copenhagen Business School, Institut for Finansiering Department of Finance.

Noland, M. 1998. *Global economic effects of the Asian currency devaluations*, Peterson Institute.

Nyberg, H. 2010. Dynamic probit models and financial variables in recession forecasting. *Journal of Forecasting,* 29**,** 215-230.

Ohlson, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research,* 18**,** 109-131.

Pollard, P. S. & Coughlin, C. C. 1999. Going down: The Asian crisis and US exports. *Federal Reserve Bank of St. Louis Review***,** 33-46.

Pozdena, R. J. 2009. Driving the economy: Automotive travel, economic growth, and the risks of global warming regulations. *Portland, Ore.: Cascade Policy Institute*.

Ramey, V. How important is the credit channel in the transmission of monetary policy? Carnegie-Rochester Conference Series on Public Policy, 1993. Elsevier, 1-45.

Richardson, F. M., Kane, G. D. & Lobingier, P. 1998. The impact of recession on the prediction of corporate failure. *Journal of Business Finance and Accounting,* 25**,** 167-186.

Romer, C. D., Romer, D. H., Goldfeld, S. M. & Friedman, B. M. 1990. New evidence on the monetary transmission mechanism. *Brookings Papers on Economic Activity***,** 149-213.

Rosengren, E. S. 2012. Money Market Mutual Funds and Financial Stability. *Remarks at the Federal Reserve Bank of Atlanta's 2012 Financial Markets Conference* Stone Mountain, Georgia: Federal Reserve Bank of Boston

Sarel, M. 1996. Nonlinear effects of inflation on economic growth. *Staff Papers,* 43**,** 199-215.

Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., Van De Koppel, J., Van De Leemput, I. A., Levin, S. A. & Van Nes, E. H. 2012. Anticipating critical transitions. *Science,* 338**,** 344-348.

Schwaab, B., Koopman, S. J. & Lucas, A. 2015. Global credit risk: World, country and industry factors. Tinbergen Institute Discussion Paper.

Sellon, G. H. & Weiner, S. E. 1996. Monetary policy without reserve requirements: analytical issues. *Economic Review-Federal Reserve Bank of Kansas City,* 81**,** 5-24.

Serven, L. & Solimano, A. 1992. Private investment and macroeconomic adjustment: A survey. *The World Bank Research Observer,* 7**,** 95-114.

Shiskin, J. & Moore, G. 1967. Indicators of Business Expansions and Contractions. *Occasional Paper,* 103.

Shumway, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business,* 74**,** 101-124.

Steven Rattner. 1981. *Fedral reserve sees little growth in '81 with contitinued high rates* [Online]. New York: New York Times.

Stock, J. H. & Watson, M. W. 1989. New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual 1989, Volume 4.* MIT press.

Stock, J. H. & Watson, M. W. 1993. A procedure for predicting recessions with leading indicators: econometric issues and recent experience. *Business cycles, indicators and forecasting.* University of Chicago Press.

Stock, J. H. & Watson, M. W. 2003. How did leading indicator forecasts perform during the 2001 recession? *FRB Richmond Economic Quarterly,* 89**,** 71-90.

Supervision, B. C. O. B. 2005. Studies on the validation of internal rating systems. *Working Paper No. 14.* Bank for International Settlements.

Surulescu, N. A. 2010. Default risk in stochastic volatility models. ETH Zurich: CER-ETH-Center of Economic Research.

Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. *Science,* 240**,** 1285-1293.

Taylor, J. B. 2009. The financial crisis and the policy responses: An empirical analysis of what went wrong. National Bureau of Economic Research.

The Conference Board. 2001. *Business Cycle Indicators Handbook,* New York.

Tian, S., Yu, Y. & 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)***,** 267-288.

Turnbull, S. M. 1979. Debt Capacity. *Journal of Finance,* 34**,** 931-940.

Turner, A. 2009. *The Turner Review: A regulatory response to the global banking crisis*, Financial Services Authority London.

Ulbricht, N. & Weiner, C. 2005. Worldscope meets Compustat: a comparison of financial databases. SFB 649 discussion paper.

Urquhart, M. A. & Hewson, M. A. 1983. Unemployment continued to rise in 1982 as recession deepened. *Monthly Labor Review,* 106**,** 3-12.

Valletta, R. & Kuang, K. 2012. Why is unemployment duration so long. *FRBSF Economic Letters,* 2012-03.

Wadhwani, S. B. 1986. Inflation, bankruptcy, default premia and the stock market. *The Economic Journal***,** 120-138.

Werling, J. F. & Mccarthy, M. 1998. The Asian Crisis and U.S. Industry: An Industry Perspective. Manufacturers Alliance.

Westcott, D. N. & Bednarzik, R. W. 1981. Employment and unemployment: a report on 1980. *Monthly Lab. Rev.,* 104**,** 4.

Wilson, T. 1997a. Portfolio Credit Risk I. 10(9), September**,** 111-117.

Wilson, T. 1997b. Portfolio Credit Risk II. 10(10), October**,** 56-61.

Zarnowitz, V. 1992a. Cyclical Indicators: Structure, Significance, and Uses. *Business Cycles: Theory, History, Indicators, and Forecasting.* University of Chicago Press.

Zarnowitz, V. 1992b. Major Macroeconomic Variables and Leading Indexes. *Business Cycles: Theory, History, Indicators, and Forecasting.* University of Chicago Press.

Zeng, B. & Zhang, J. 2002. Measuring credit correlations: equity correlations are not enough. *KMV, San Francisco*.

Zhou, C. 2001. An analysis of default correlations and multiple defaults. *Review of Financial Studies,* 14**,** 555-576.

Zmijewski, M. E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research,* 22**,** 59-82.

9. Appendix

9.1 Supplemented Document for Chapter 5

This part is to provide how to get the results in Table 5-4 and Table 5-5. The detail description is shown in Table 9-1.

Table 9-1 EPs in different subsamples from April 1980 to December 2014

1980/04-1985/03	**1980/05-1985/04**	**1980/06-1985/07**	**…**	**2009/11-2014/10**	**2009/12-2014/11**	**2010/01-2014/12**
REALLNNSA	REALLNNSA	REALLNNSA	…	IPMANSICS	LOANS	LOANS
GFDEBTN	GFDEBTN	IPMINE	…	IPMAT	TOTALNS	TDSP
IPMINE	IPMINE	NAPMII	…	USPHCI	TDSP	IPMANSICS
NAPMII	NAPMII	PI	…	GPDI	IPMAT	IPMAT
A001RP1Q027SBEA	PCEPI	PCEPI	…	UNRATE	M12MTVUSM227NFWA	NAPMII
W068RCQ027SBEA	A001RP1Q027SBEA	A001RP1Q027SBEA	…	HOHWMN02USM065S	PCEDG	M12MTVUSM227NFWA
HSN1F	W068RCQ027SBEA	W068RCQ027SBEA	…	TCU	USPHCI	PCEDG
ALTSALES	HOUST	UMCSENT	…	A190RL1Q225SBEA	A001RP1Q027SBEA	USPHCI
HOUST	FSDP	M2SL	…	PSAVERT	W068RCQ027SBEA	A001RP1Q027SBEA
FSDP	REQRESNS	MICH	…	GDPDEF	UMCSENT	W068RCQ027SBEA
MICH	**CPILFESL**	**CPILFESL**	…	BAA.v	GPDI	UMCSENT
CPILFESL	GPSAVE	CPIENGSL	…	SPREAD.MOODY.1	UNRATE	GPDI
GPSAVE	TWEXBMTH	GPSAVE	…	SPREAD.MOODY.2	HOUST	UNRATE
GDPDEF	IMFSL	TWEXBMTH	…	SP500.v	FSDP	ALTSALES
TWEXBMTH	BOPBM	IMFSL	…		PSAVERT	HOUST
IMFSL	MPRIME	SPREAD.GS	…		GPSAVE	PSAVERT
BOPBM	SPREAD.GS	SP500.v	…		GDPDEF	GPSAVE
MPRIME	SP500.v		…		FEDFUNDS	FEDFUNDS
SP500.v			…		BAA.v	BAA.v
			…		MPRIME	MPRIME
			…		PPICEM	PPICEM
			…		ULCNFB	SP500.v
			…		SP500.v	

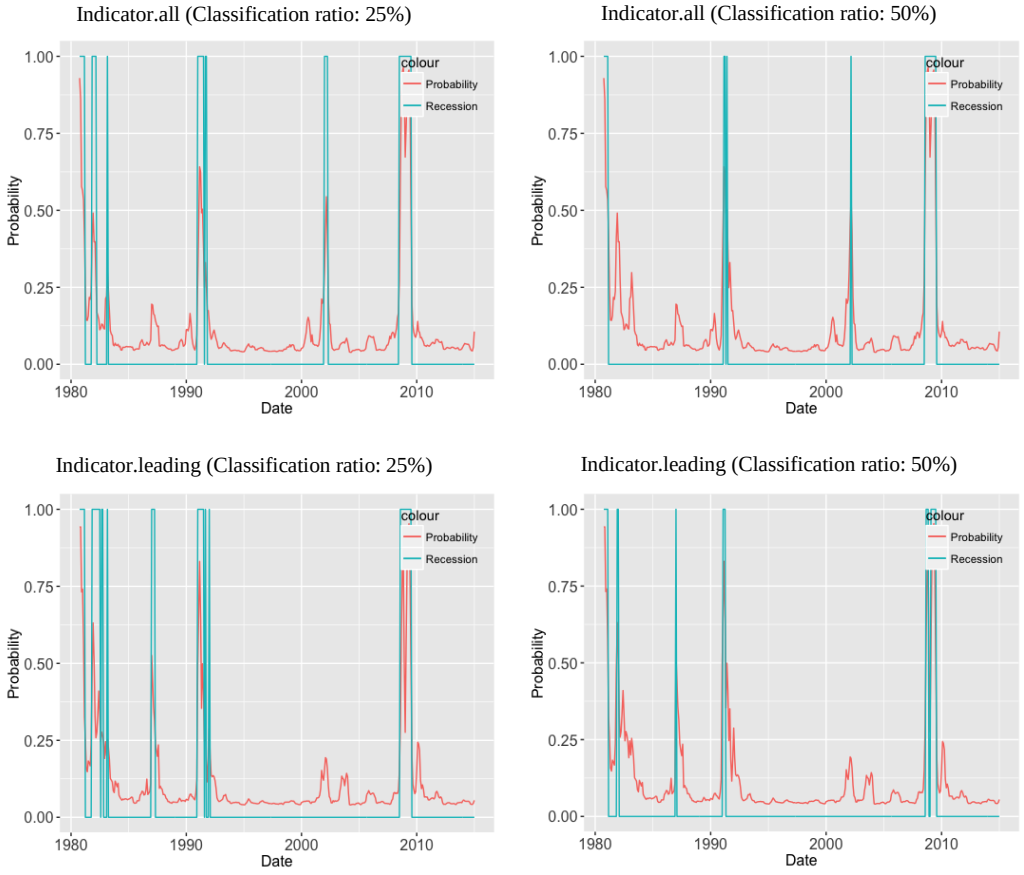
Note: The total sample is from April 1980 to December 2014, which contatins 417 observations. I split them into 357 intervals by using rolling window. There is 5-year data in the window with 60 observations, and LASSO is used to extract a set of EPs in the window. I move the first window from April 1980 to March 1985 by a month. The second interval from May 1980 to April 1985 is created. This procedure is iterated until the window ends in January 2010. The final window is then from January 2010 to December 2014. Based on the 357 intervals, I eventually get 357 groups of EPs by using LASSO. Frequency % means that the percentage of frequency of occurrence in the 357 groups of EPs. If the number is more than 50%, it means that this predictor can explain more than half of sample period. For example, an EP entitled CPILFESL occurs in 189 out of 357 groups of EPs, and the frequency for this EP is 189. In the first three subsamples, CPIFESL always appears. Likewise, I need to count how many intervals this variable can explain IDP. This number is the value of frequency, which is 189 in this case. Since I have 357 subsamples, then the percentage of frequency is 189 which is divided by 357, which is 53%.

9.2 Supplemented Document for Second Empirical Chapter

This part is to provide all the in-sample and out-of-sample prediction results by using the model with various macro indicators. All the figures plot recession probabilities implied by the logit model. Red line is the time series of recession probabilities. Green line represents the occurrence of recessions implied by the recession probabilities. I use two ratios to classify the recessions, which are 25% and 50%, respectively.

9.2.1 In-sample Forecasts for Economic Recessions

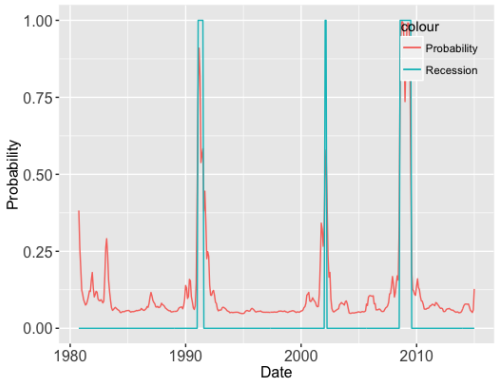
Figure 9-1 In-sample forecasts by using various macro indicators



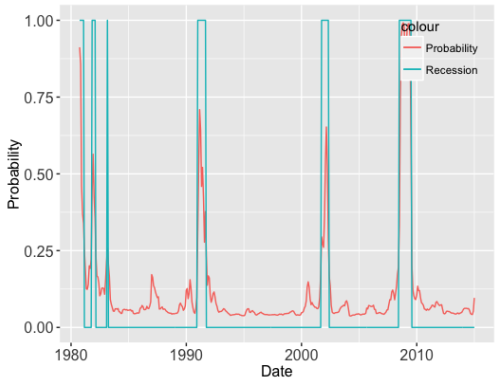
Indicator.procyclical (Classification ratio: 25%)



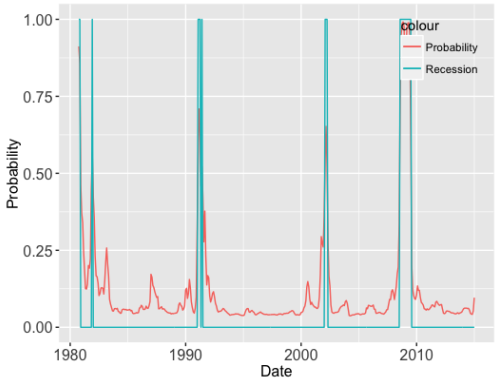
Indicator. procyclical (Classification ratio: 50%)



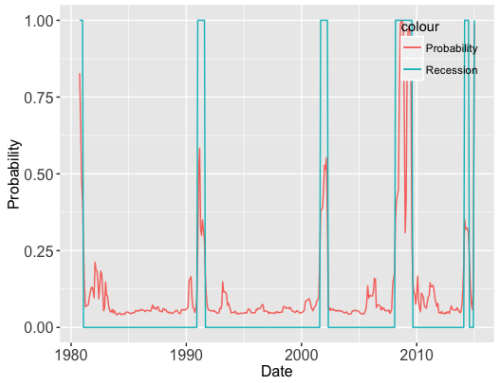
Indicator.6.economic (Classification ratio: 25%)



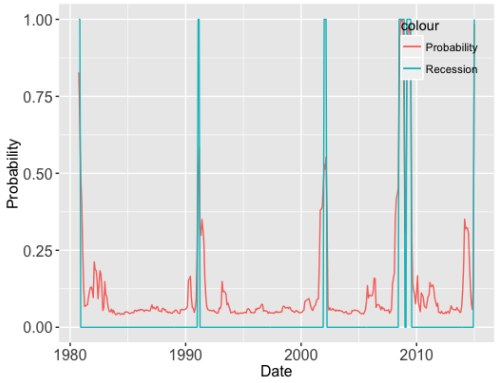
Indicator.6.economic (Classification ratio: 50%)



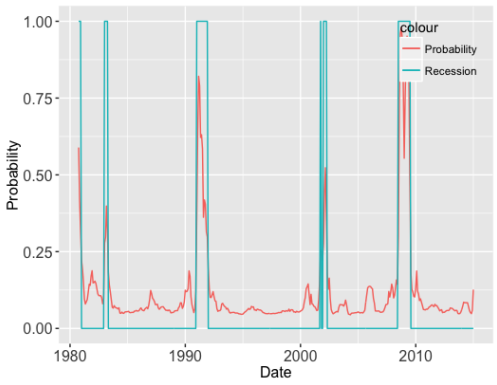
Indicator.effective (Classification ratio: 25%)



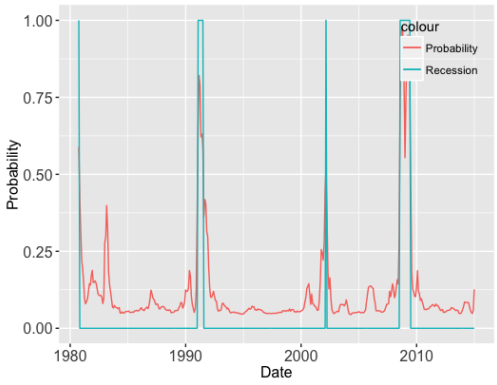
Indicator.effective (Classification ratio: 50%)



Indicator.BC (Classification ratio: 25%)



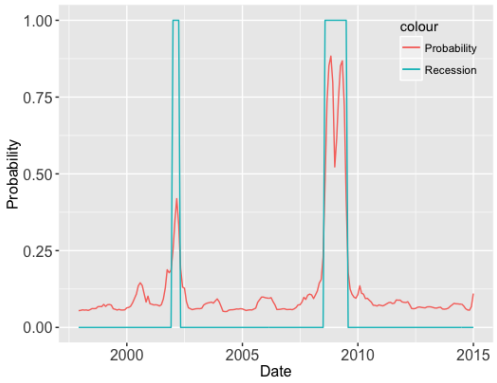
Indicator.BC (Classification ratio: 50%)



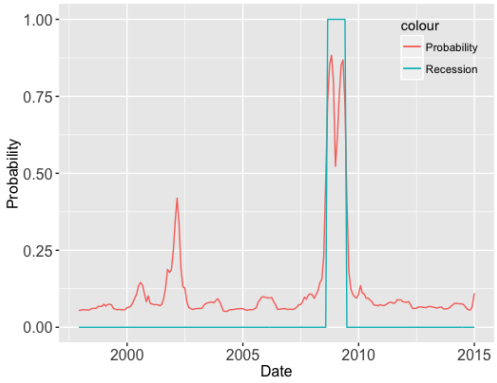
9.2.2 Out-of-sample Forecasts for Economic Recessions

Figure 9-2 Out-of-sample forecasts by using various macro indicators. The forecast starts from January 1998 to December 2014

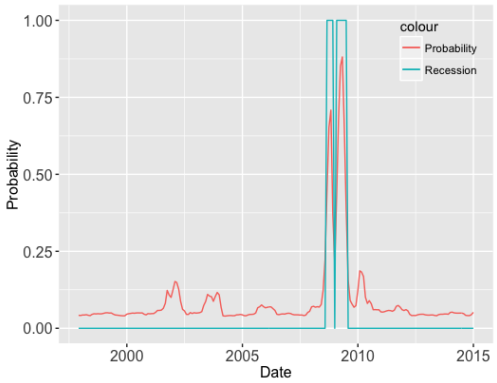
Indicator.all (Classification ratio: 25%)



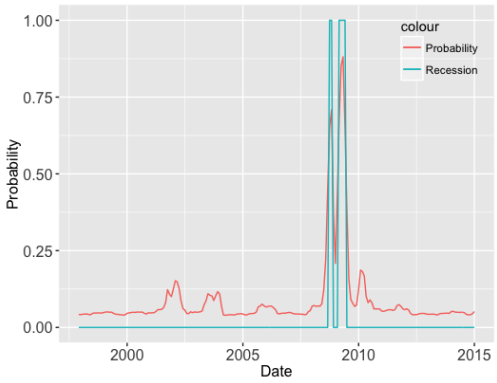
Indicator.all (Classification ratio: 50%)



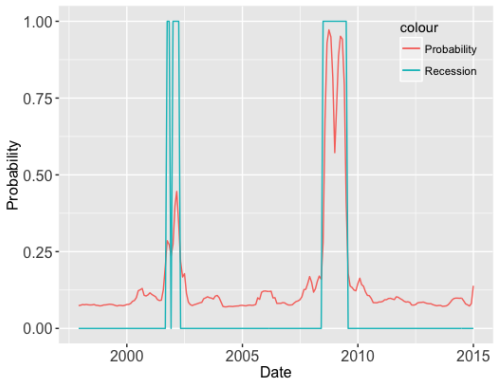
Indicator.leading (Classification ratio: 25%)



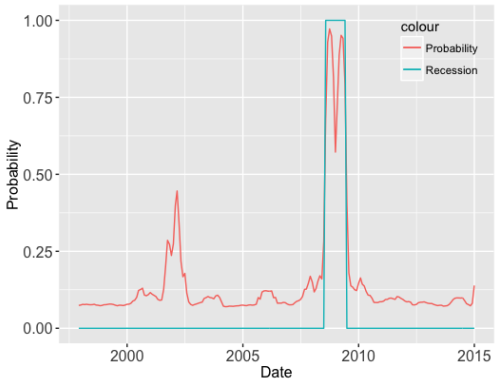
Indicator.leading (Classification ratio: 50%)



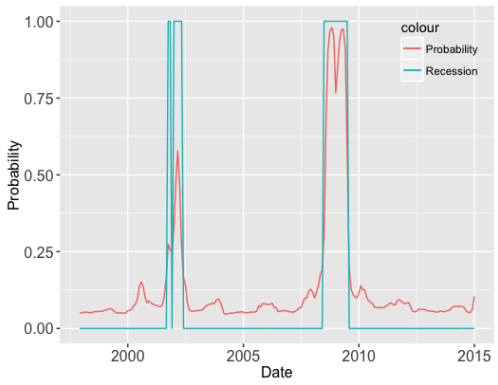
Indicator.procyclical (Classification ratio: 25%)



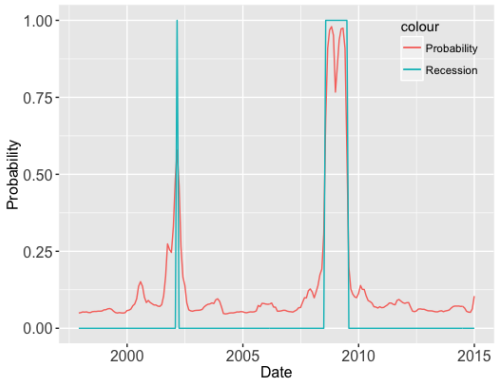
Indicator. procyclical (Classification ratio: 50%)



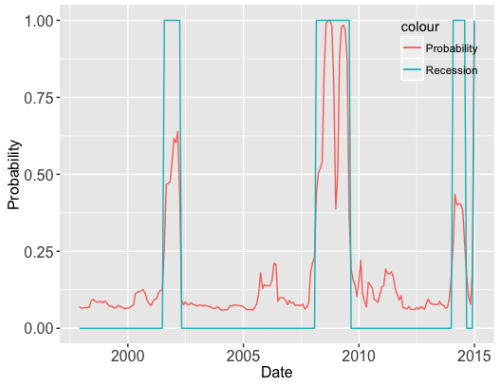
Indicator.6.economic (Classification ratio: 25%)



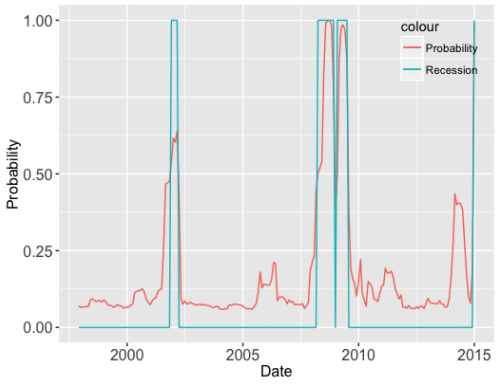
Indicator.6.economic (Classification ratio: 50%)



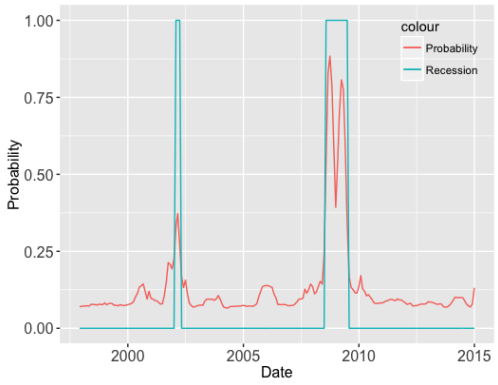
Indicator.effective (Classification ratio: 25%)



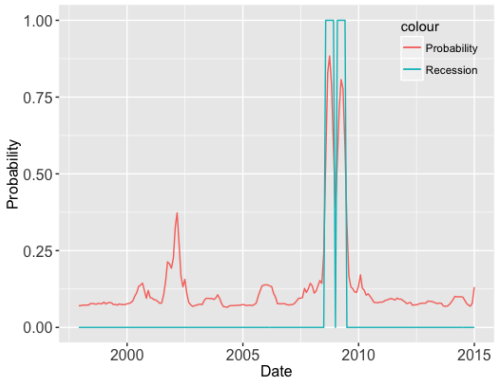
Indicator.effective (Classification ratio: 50%)



Indicator.BC (Classification ratio: 25%)



Indicator.BC (Classification ratio: 50%)



9.3 User's Manual of R Package "Kai.R.Tools"

Package "Kai.R.Tools"

June 29, 2016

Type Package for **R**

Title Tools for extracting default information and rated information from Moody's DRD, constructing two types of risk drivers including IDPs and VDPs, respectively, and statistically analysing the data and modelling IDPs/VDPs.

First Version, 2014-11-30
Second Version, 2014-12-03
Third Version, 2014-12-06
Fourth Version, 2014-12-08
Fifth Version, 2014-12-19[38]
Sixth Version, 2014-12-21[39]
Seventh Version, 2014-12-25[40]
Eighth Version, 2015-01-02[41]
Ninth Version, 2015-01-07[42]
Tenth Version, 2015-07-15[43]

Author Kai Xing PhD Candidate in Finance and Risk <lixkx@nottingham.ac.uk>

Description

Kai-DRD is a set of tools that solve the calculation of issuer-weighted default probabilities (IDPs) and value-weighted default probabilities (VDPs) and issue-weighted default probabilities from Moody's Default Recovery Database (DRD). These tools also provide the solution about how to merge the other databases including CRSP and Compustat/CCM and combine them together. It also has the plotting tools for showing the trend of various types of default probabilities.

[38] This new version is to add new functions for merging three databases: Moody's, Compustat/CCM, and CRSP.

[39] A new set of functions in the third part for construction of internal risk drivers are added into.

[40] This version is to optimize several functions such as using package "data.table" to replace "plyr" in order to work out quickly. Correcting some codes existing in several functions, for example, in the Module 6 and 7, the functions used for counting the number of defaults and firms (bonds) are corrected finally.

[41] Debugging the codes of functionalities of construction of IDPs and VDPs and adding new functionalities for constructing internal risk drivers.

[42] This version optimizes the current functions and now it can also produce the default databases for the research only on external risk drivers (it means merging Compustat/CCM is not concerned in this case). In addition, they can also concern two cases based on whether the firms are listed in US market.

[43] This version covers a set of new functions such as construction of DTD (Duan and Wang, 2012) and Sigma (Shummy, 2001), implementation of network analysis, 2 variable selection methods and 3 correlation methods etc.

In addition, this package also provides a set of tools for extracting risk drivers from CRSP (daily and monthly), Compustat (quarterly and annually), and Federal Reserve Economic Data (FRED). In particular, for the internal risk drivers, it covers more than 50 firm specific risk drivers constructed from CRSP and Compustat/CCM such as accounting ratios, market-driven variables, distance to default (DTD)(yearly/monthly type- based on KMV approach and market-proxy approach), stock return for each firm, and the idiosyncratic standard deviation of stock returns..

This package also covers a set of tools for statistical analysis used in my project such as network analysis, correlation analysis, time series analysis, variable selection analysis etc. and it offers functions for modelling default probabilities. Therefore, it will be continually updated based on the demand of my research project in the future.

Date 2016-06-29 19:00:00

R topics documented

9.2.1 Functionalities of Construction of Issuer-weighted Default Probabilities (IDP) 247

9.2.2 Algorithm of IDP and Data File Description 248

Module – ZERO: Merging two databases: Compustat (or CCM) and CRSP 249
Module – 1: Preliminary clean including country, period, economic sector 253
Module – 2: Cleaning family structures 254
Module – 3: Retaining the first default event 255
Module – 4: Cleaning defaults in terms of considering effective default data (that can be used for identifying accounting information) 256
Module – 5: Summarizing the results of defaulted firms 257
Module – 6: Preliminary clean including country, period, economic sector 258
Module – 7: Identifying effective rating firms after considering CRSP 259
Module – 8: Extracting the number of ratted firms 260
Module – 9: Identification of IDP 261
Practice – Identification of IDP in the US industrials from 1980/01/01 to 2013/12/31 263

9.2.3 Functionalities of Construction of Value-weighted Default Probabilities (VDP) 264

Module – ZERO: Merging two databases: Compustat and CRSP 267
Module – 1: Preliminary clean including country, market, period, economic sector, bond type 267
Module – 2: Cleaning family structures 268
Module – 3: retaining the first default event 270
Module – 4: Cleaning defaults in terms of considering effective default data (that can be used for identifying accounting information) 271
Module – 5: Summarizing the results of defaulted bonds 271
Module – 6: Preliminary clean including country, period, economic sector 273
Module – 7: Identifying effective rating bonds after considering CRSP/Compustat 274
Module – 8: Extracting the number of total value of ratted bonds and issues 275
Module – 9: Identification of VDP and issues-weighted DP 276
Module – 10: Summary of IDPs and VDPs 277
Practice – Identification of IDPs in the US industrials from 1980/01/01 to 2014/12/31 279

9.2.4 Functionalities of Construction of Internal Risk Drivers 280

Overview-Description of internal risk drivers 281
Module – 1: Construction of DTD.MONTHLY, SIGMA and RETURN.ANNUAL 282
Module – 2: Construction of DTD.ANNUALLY 284
Module – 3: Construction of accounting ratios 287
Module – 4: Construction of industry-level of internal risk drivers 288

9.2.5 Other key functions for empirical analysis 289

9.3.1 Functionalities of Construction of Issuer-weighted Default Probabilities (IDP)

Structure

1. Algorithm of IDP and Data File Description

2. From Module – 1st to 5th Module are about identification of denominators of Default probabilities

Module – ZERO: Merging two databases: Compustat and CRSP

Module – 1 Preliminary clean including country, period, economic sector

Module – 2 Cleaning family structures

Module – 3 Retaining the first default event

Module – 4 Cleaning defaults in terms of considering effective default data by merging CRSP

Module – 5 Summarizing the results of defaulted firms

3. From Module – 6th to 8th Module are about identification of denominators of Default probabilities

Module – 6 Preliminary clean including country, period, economic sector

Module – 7 Identifying effective rating firms after considering CRSP

Module – 8 Extracting the number of ratted firms

4. Module – 9 Identification of IDP

5. Practice - identification of IDP in the US industrials from 1980/01/01 to 2013/12/31

9.3.2 Algorithm of IDP and Data File Description

(1) Method of calculation of IDPs

IDPs are fractions in which the numerator represents the number of issuers that defaulted in a particular time period and the denominator represents the number of issuers that could have defaulted in that time period. The formula is shown below:

$$IDPs = \frac{d_t^{IDP}}{n_t^{IDP}} \tag{66}$$

Under this method, this research defines the outstanding number of credits (n_t) recursively as:

$$n_t^{IDP} = n_{t-1}^{IDP} - d_{t-1}^{IDP} - w_t^{IDP}. \tag{67}$$

The numeratorsd_t^{IDP} and d_{t-1}^{IDP} are the numbers of issuers defaulting at period t and $t-1$. The denominators n_t^{IDP} and n_{t-1}^{IDP} are the numbers of issuers that potentially could have defaulted at date t and $t-1$. w_t^{IDP} denotes the number of credits which withdraw between periods t and $t+1$ (Acharya and Carpenter, 2002, Moody, 2015).

For example, if 25 issuers defaulted in 2013-12, and that static pool consisted of 2,000 issuer ratings, the resulting annual default rate for all ratings in December would be 1.3% (25/2000). While the calculation of static pool allows for the credits that withdraw and default, it is the numbers of issuers in 2013-11 (2020) minus 10 issuers defaulted in 2013-11 minus 10 issuers withdraw between 2014-01 and 2013-12.

(2) Description of Relevant Data Files

There are 9 files used for extracting the number of defaulters at date and the number of rated firms at period t and $t-1$. In Table 9-2, there are 5 files that are used for calculation of IDPs in DRD. All the historical defaults can be found in the file called mast_dflt.txt and rated firms in each date can be found in senr_ratg_standard.txt. Mast_issr.txt shows the detailed information about each firm such as country, sector, name etc. family_structure.txt is used to determine the firms' parent firms in the history. Moodys_cusip.txt is used to merge CCM database/CRSP and DRD. These files should be uploaded into R environment.

Table 9-2 Relevant files for calculating IDP in DRD

Names	Descriptions
mast_dflt.txt	Stores details of each issuer default.
mast_issr.txt	Stores descriptive details of each issuer in the data set.
issr_ids.txt	Links Moody's Issuer Number to one or more identifiers external to this database (such as Moodys.com identifiers, tickers, and Base CUSIPs), where available.
family_structure.txt	Stores default histories in the database; defaults can have more than one "default event" listing in this table.

CRSP_daily.csv	Stores daily stock market data on major stock exchanges data on major stock exchanges (NYSE, AMEX, NASDAQ) from CRSP.
CRSP_monthly.csv	Stores monthly stock market data on major stock exchanges data on major stock exchanges (NYSE, AMEX, NASDAQ) from CRSP.
CCM_quarterly.csv	Stores quarterly accounting information from CRSP/Compustat merged database (CCM).
CCM_annual.csv	Stores annually accounting information from CRSP/Compustat merged database (CCM).
Moodys_cusip.csv	Stores all the firms that can match the firms recorded in Compustat. We extract all the firms' CUSIP values from the file called issr_ids.txt (in Moody's DRD) and then upload the 6 digits CUSIP to Wharton Research Data Services to obtain these firms' PERMNO codes in Compustat. We save the file including several columns such as "COMNAN", "NCUSIP", "TICKER", and "PERMNO" as Moodys_cusip. In order to identify the listed firms from rated firms in Moody's database, we need this file Moodys_cusip. If the function has the logical value to control whether we need listed firms or not, this file should be used.

Source: The first five files are downloaded Moody's DRD database (https://www.moodys.com/Pages/Default-and-Recovery-Analytics.aspx). For the detail description of each file, please refer to *DRD Technical Specifications* (https://www.moodys.com/sites/products/ProductAttachments/MA%20DRD%20Tech%20Specs.pdf. Both CRSP and CCM are downloaded from Wharton Research Data Services (http://wrds-web.wharton.upenn.edu/wrds/).

References:

ACHARYA, V. V. & CARPENTER, J. N. 2002. Corporate bond valuation and hedging with stochastic interest rates and endogenous bankruptcy. *Review of Financial Studies,* 15**,** 1355-1383.

MOODY'S 2015. Annual Default Study: Corporate default and recovery rates, 1920-2014. *Moody's Global Credit Policy Special Comment, Annual update.*

Module – ZERO: Merging two databases: Compustat (or CCM) and CRSP

Merging_Compustat.CRSP_1/ Merging_Compustat.CRSP_2/ Merging_Compustat.CRSP_3

Description

Merges two databases including Compustat (or CCM) and CRSP

Usage

```
Merging_Compustat.CRSP_1 (date_1, date_2, cleaning_time_1)
Merging_Compustat.CRSP_2 (x, var_quarter, var_annual)
Merging_Compustat.CRSP_3 (x, cleaning_time_2)
```

Arguments

date_1	a string value, specifying start date for extracting accounting data that is lag time of date for extracting defaults.
date_2	a string value, specifying end date for extracting accounting data that is lag time of date for extracting defaults.
cleaning_time_1	a numeric value, selecting the firms that have more than this number of monthly stock price from CRSP database. If it is 6, this means that we choose the firms that have more than 6 months' stock information.
x (in Merging_Compustat.CRSP_2)	a data frame that is constructed in the function **Merging_Compustat.CRSP_2**.
var_quarter	a string value, specifying a accounting variable in monthly Compustat and the input is abbreviation of accounting variable in Compustat.
var_annual	a string value, specifying a accounting variable in annual Compustat and the input is abbreviation of accounting variable in Compustat.
x (in Merging_Compustat.CRSP_3)	a data frame that is constructed in the function **Merging_Compustat.CRSP_3**.
cleaning_time_2	a numeric value, selecting the firms that have more than this number of quarterly stock price from quarterly Compustat database. If it is 2, this means that we choose the firms that have more than 2 months' stock information. This value should be consistent with the value of "cleaning_time_1".

Details

The function Merging_Compustat.CRSP_1 is to merge Compustat (or CCM) and CRSP in order to obtain common firm level data such as equity information, accounting information in both databases. The main standard of screening the data is to keep the data that have at least certain number of months' equity price, which is controlled by the input titled "cleaning_time_1". In particular, two databases are used in this function. One is the CRSP/Compustat Merged Database **which should be named as CCM_quarter** in Table 9-2. Another is the monthly CRSP stock database **which should be named as CRSP_monthly** in Table 9-2.

The results obtained by Merging_Compustat.CRSP_1 will be further screened by the following function named as Merging_Compustat.CRSP_2 which is to fill missing values in Compustat monthly database. The methods used for filling data is from Duffie et al. (2007). For cases with missing debt data, if the missing value corresponds to year Y, quarter Q, then

we complete the data whenever possible as follows. First, we only consider there to be missing data for firm F at year Y, quarter Q if there is debt data for firm F before and after Y , Q. Given that, if there are no missing quarterly debt data for F in year Y, we take the closest observation to quarter Q, before quarter Q, to complete the data set. If all quarterly debt observations for F in year Y are missing, we look in the yearly debt file for firm F , year Y . If this observation is not missing we take it to be the debt level for year Y, quarter Q, firm F. In all other cases of missing observations we take the data to be missing.

The third function named as Merging_Compustat.CRSP_3 is to further screen firms that have less than the certain number of stock information after filling the missing values in the monthly data by using quarterly data.

date_1 is the start date and date_2 is the date that after the end of date. Since the lag accounting period will be considered according to previous research, then setting the date should be careful. For example, if we want to predict default on 1980/01/01, then the date for extracting accounting information should start on 1979/10/01 based on 3 months lag accounting information.

cleaning_time_1 and cleaning_time_2 are different. Cleaning_time_1 is based on the number of months. If we just want to keep 6 months data, then cleaning_time_1 should be 6. Cleaning_time_2 is based on the number of quarterly data, for example, 6 months data will be reserved and then here it should be 2 rather than 6 as 2 represents two quarterly data for the stock.

References:

DUFFIE, D., SAITA, L. & WANG, K. 2007. Multi-period corporate default prediction with stochastic covariates. Journal of Financial Economics, 83, 635-665.

Examples:

```
DATA.Duffie <- Merging_Compustat.CRSP_1 ('1959-09-01', '2015-01-30', 6)
DATA <- DATA.Duffie

# 1 - cheq: Cash and Short-Term Investments

temp_DATA <- Merging_Compustat.CRSP_2(DATA, 'cheq', 'che')
temp_DATA <- Merging_Compustat.CRSP_3(temp_DATA, 2)

temp_DATA$cheq <- temp_DATA$var_quarter
temp_DATA$var_quarter <- NULL

"cheq" %in% names(temp_DATA)

# 2 - dlcq: Debt in Current Liabilities

temp_DATA <- Merging_Compustat.CRSP_2(temp_DATA, 'dlcq', 'dlc')
temp_DATA <- Merging_Compustat.CRSP_3(temp_DATA, 2)

temp_DATA$dlcq <- temp_DATA$var_quarter
temp_DATA$var_quarter <- NULL
```

```
"dlcq" %in% names(temp_DATA)

# 3 - lctq: Current Liabilities - Total
# 4 - dlttq: Long-Term Debt - Total
# 5 - atq: Assets - Total
# 6 - ltq: Liabilities - Total
# 7 - niq: Net Income
# 8 - req: Retained Earnings
# 9 - wcapq: Working Capital
# 10 - cshoq: Common Shares Outstanding
# 11 - intanq: Intangible Assets - Total
# 12 - actq: Current assets

# 13 - prccq:

temp_DATA <- Merging_Compustat.CRSP_2(temp_DATA, 'prccq', 'prcc_c')
temp_DATA <- Merging_Compustat.CRSP_3(temp_DATA, 2)

temp_DATA$prccq <- temp_DATA$var_quarter
temp_DATA$var_quarter <- NULL
"prccq" %in% names(temp_DATA)

# 14 - ebit

temp_DATA <- Merging_Compustat.CRSP_unique('ebit', temp_DATA, 2)

temp_DATA$ebit <- temp_DATA$var_quarter
"ebit" %in% names(temp_DATA)

# 15 - saleq

temp_DATA <- Merging_Compustat.CRSP_2(temp_DATA, 'saleq', 'sale')
temp_DATA <- Merging_Compustat.CRSP_3(temp_DATA, 2)

temp_DATA$saleq <- temp_DATA$var_quarter
temp_DATA$var_quarter <- NULL
"saleq"%in% names(temp_DATA) # 15 - revq

temp_DATA <- Merging_Compustat.CRSP_2(temp_DATA, 'revtq', 'revt')
temp_DATA <- Merging_Compustat.CRSP_3(temp_DATA, 2)

temp_DATA$revtq <- temp_DATA$var_quarter
temp_DATA$var_quarter <- NULL
"revtq" %in% names(temp_DATA)

DATA.Duffie <- temp_DATA
DATA_conversion <- DATA.Duffie (Note: DATA_conversion is used to screen listed firms in
the function titled NUM_defaults_5_internal.drivers)
```

Module – 1: Preliminary clean including country, period, economic sector

NUM_defaults_filter_1

Description

Extracts the basic information, such as,
(1) issuer's information
(2) Filter condition <1> - Country
(3) Filter condition <2> - Time period
(4) Filter condition <3> - Economic Sectors

Usage

```
NUM_defaults_filter_1 (country, economic_sector, date_1, date_2, all.firms = T, listed.firms = T , Non.listed.firms = T)
```

Arguments

country	a numeric value, specifying country's code, US country in Moody's is 145.
economic_esctor	a string value, specifying which economic sector that we need to choose. For Moody's database, 'INDUSTRIAL' is used for representing all of firms in industrials.
date_1	a string value, specifying start date for extracting default data.
date_2	a string value, specifying end date for extracting default data.
all.firms = T	a logical value, controlling whether all firms are considered or not.
listed.firms = T	a logical value, controlling whether listed firms are considered or not.
Non.listed.firms = T	a logical value, controlling whether Non-listed firms are considered or not.

Details

The function NUM_defaults_filter_1 is to generate a specified data frame in terms of country, time period, and economic sectors. This function uses three databases including mast_dflt.txt, mast_issr.txt from Moody's DRD database and Moodys_cusip.csv in Table 9-2. The first database covers the historical defaults information, and the second one covers all the issuers' information.

date_1 is the start date and date_2 is the end date, such as, if we end on Dec 31, 2013, and then date_2 is January 1, 2014. It should be noted that economic_sector, date_1, and date_2 are all string values and we need to add '' into the codes.

For the logical value, if non-listed firms are considered in this research, listed.firms should be FALSE. Then the Module (or function) used in the next procedures are Module 2 (for cleaning family structure), Module 3 (for cleaning default event within XXX days) and Module 6 (for obtaining denominators of default probabilities). Module 4 (for cleaning defaults in terms of considering effective default data (that can be used for identifying accounting information) and Model 7 (for identifying effective rating firms after considering CRSP) will be useless for this research (Non-listed firms have no accounting information). If all of firms are concerned, two logical value should be FALSE.

Examples:

```
nonlisted.idp_1 <- NUM_defaults_filter_1 (145, 'INDUSTRIAL', '1980-01-01', '2013-12-31', FALSE, FALSE, TRUE)

listed.idp_1 <- NUM_defaults_filter_1 (145, 'INDUSTRIAL', '1980-01-01', '2013-12-31', FALSE, TRUE, FALSE)

all.idp_1 <- NUM_defaults_filter_1 (145, 'INDUSTRIAL', '1980-01-01', '2013-12-31', TRUE, FALSE, FALSE)
```

Module – 2: Cleaning family structures

NUM_defaults_2/ NUM_defaults_3

Description

Combines defaults by firms related through parent-subsidiary relationships. There are two functions in this Module in order to handle two issues in cleaning family structure. They are:

1) Some firms experienced restructure events because of default events or expansion (acquistion/merger), then these firms have different parent firms in the past. In this context, we need to screen these firms;
2) Next issue is to identify the parent firm and their subsidiaries in the same default date and then we need to choose parent firms and delete their subsidiaries.

Usage

```
NUM_defaults_2 (x, all.firms = T,  listed.firms = T , Non.listed.firms = T)

NUM_defaults_3 (x, all.firms = T,  listed.firms = T , Non.listed.firms = T)
```

Arguments

x (in NUM_defaults_2)	a data frame that is constructed in the previous function NUM_defaults_filter_1.
x (in NUM_defaults_3)	a data frame that is constructed in the previous function NUM_defaults_2.
all.firms = T	a logical value, controlling whether all firms are considered or not.
listed.firms = T	a logical value, controlling whether listed firms are considered or not.
Non.listed.firms = T	a logical value, controlling whether Non-listed firms are considered or not.

Details

In both functions, the first function is used to determine the firms that have different parent firms in the past; the second function is to select parent firms rather than their subsidiaries. Keeping the order of using two functions is necessary.

Examples

```
all.idp_2 <- NUM_defaults_2 (all.idp_1, TRUE, FALSE, FALSE)

all.idp_3 <- NUM_defaults_3 (all.idp_2, TRUE, FALSE, FALSE)
```

Module – 3: Retaining the first default event

NUM_defaults_4

Description

Retain the first default event within a certain span of time.

Usage

```
NUM_defaults_4 (x, days)
```

Arguments

x	a data frame that is constructed in the previous function NUM_defaults_3.
days	a numeric value, specifying the number of days used for screening default events within a certain period.

Details

This function is to retain the first default event if the firm has more than two defaults within two years.

days is 740 which means 365 days * 2 years <- 730 days. If we want to consider one year, then the value of days will be 365.

Examples

```
all.idp_4 <- NUM_defaults_4 (all.idp_3, 730, TRUE, FALSE, FALSE)
```

Module – 4: Cleaning defaults in terms of considering effective default data (that can be used for identifying accounting information)

NUM_defaults_5_internal.drivers

Description

Clean defaults by merging **DATA_conversion** produced in Model-ZERO and we then obtain effective data from CRSP, and Compustat. In particular, there are two issues that should be concerned in this function, which are shown below:

1) Obtaining CUSIP value in Moody's and connecting Moody's database with **DATA_conversion**;
2) Extracting defaults (with defaults) that can successfully match with the firms existing in **DATA_conversion**.

Usage

```
NUM_defaults_5_internal.drivers (x, Internal.risk.drivers = T)
```

Arguments

x	a data frame that is constructed in the function NUM_defaults_4.
Internal.risk.drivers = T	a logical value, controlling whether the internal risk drivers are concerned in the study.

Detail

It should be noted that the firms with less than 6 months stock price information should be deleted in this codes.

Examples

```
all.idp_5 <- NUM_defaults_5_internal.drivers (all.idp_4, TRUE)
```

Module – 5: Summarizing the results of defaulted firms

output_defaults/ output_defaults_types

Description

Includes two functions in this part which are used for:

1) Summarizing defaults in each month and each year;
2) Counting the number of types of default events by default types.

Usage

```
output_defaults (x, date_1, length_time)

output_defaults_types (x)
```

Arguments

x	a data frame that is constructed in the function NUM_defaults_5_internal.drivers or NUM_defaults_4.

date_1	a string value, specifying start date for extracting defaults.
length_time	a numeric value, specifying the number of months from the start date to end date. For example, there are 408 observations [(2014-1980)*12 = 408] from Jan 1, 1980 to Dec 31, 2013.

Details

Both functions are to summarize the defaults. One is to summarize the number of defaults each month and the number of rated firms each month; another is to show the types of defaults and the number of each type of defaults. The input for x can come from different database produced by previous functions. For example, if we do not concern listed firms, then we can directly use the results from function NUM_defaults_4.

Examples

```
all.idp_output_defaults <- output_defaults(all.idp_5, '1980/1/1', 408)

defaults_types_all  <- output_defaults_types (all.idp_5)
```

Module – 6: Preliminary clean including country, period, economic sector

DEN_rated_filter_1

Description

Extracts basic information about rated firms

1) Issuer's information
2) Filter condition <1> - Country
3) Filter condition <2> - Economic Sectors

Usage

```
DEN_rated_filter_1 (country, economic_sector, date_1, date_2, x, all.firms = T, listed.firms = T, Non.listed.firms = T)
```

Arguments

country	a numeric value, specifying country's code, US country in Moody's is 145.
economic_esctor	a string value, specifying which economic sector that we need to choose. For Moody's database, 'INDUSTRIAL' is used for representing all of firms in industrials.
date_1	a string value, specifying start date for extracting accounting data that is lag time of date for extracting defaults.
date_2	a string value, specifying end date for extracting accounting data that is lag time of date for extracting defaults.
x	a data frame that is produced by Function NUM_defaults_4.
all.firms = T	a logical value, controlling whether all firms are considered or not.
listed.firms = T	a logical value, controlling whether listed firms are considered or not.
Non.listed.firms = T	a logical value, controlling whether Non-listed firms are considered or not.

Details

date_1 is the start date and date_2 is the end date, such as, if we end on Dec 31, 2013, and then date_2 is January 1, 2014. Input (x) can be the results from the other functions such as NUM_defaults_4 and NUM_defaults_3. This result of rated firms' information depends on whether we need to concern two screening conditions including parent-subsidiary relationship, and multiple defaults.

Examples

```
all.idp_DATA_1 <- DEN_rated_filter_1 (145, 'INDUSTRIAL', '1980-01-01', '2014-12-31', all.idp_4, TRUE, FALSE, FALSE)
```

Module – 7: Identifying effective rating firms after considering CRSP

DEN_rated_2

Description

This function is to merge Moody's DRD with **DATA_conversion** from Model-ZERO.

Usage

DEN_rated_2 (x, date_1, date_2, y, Internal.risk.drivers = T)

Arguments

x	a data frame that is constructed in the previous function DEN_rated_filter_1.
date_1	a string value, specifying start date for extracting accounting data that is lag time of date for extracting defaults.
date_2	a string value, specifying end date for extracting accounting data that is lag time of date for extracting defaults.
y	a data frame that is produced by the previous Function NUM_defaults_5.
Internal.risk.drivers = T	a logical value, controlling whether the internal risk drivers are concerned in this study.

Details

date_1 is the start date and date_2 is the end date.

Examples

all.idp_DATA_2 <- DEN_rated_2 (idp_DATA_1, '1980-01-01', '2013-12-31', idp_data_5_listed.firm, TRUE)

Module – 8: Extracting the number of ratted firms

DEN_rated_3

Description

This function is to extract monthly rated firms.

Usage

DEN_rated_3 (x, date_1, date_3, length_time)

Arguments

date_1	a string value, specifying start date for extracting defaults.
date_3	a string value, being used for identifying the end of each month. In particular, the monthly date should be in Feb of start year.
length_time	a numeric value, specifying the number of months from the start date to end date. For example, there are 408 observations [(2014-1980)*12 = 408] from Jan 1, 1980 to Dec 31, 2013.

Details

date_3 is used to identify the end of each month. This value is to determine the withdrawn in calculation of denominators of default probabilities.

Examples

all.idp_output_rated_firms <- DEN_rated_3 (all.idp_DATA_2, '1980-01-01', '1980-02-01', 408)

Module – 9: Identification of IDP

default_pro

Description

These functions are to calculate the issuer-weighted default probabilities and do some graphical analysis based on algorithm of IDP.

Usage

default_pro (x, y)

plot_line (x, frequency, var)

plot_bar (x, frequency, var)

plot_black_outline (x, frequency, var)

Arguments

x	a data frame that is constructed in the pervious function. output_defaults
y	a data frame that is constructed in the pervious function. output_rated_firms
frequency	a string value that is the variable in the data frame. It can be either monthly or yearly, such as 'date_YM', and 'year'.
var	A string value that is the variable in the data frame. Such as, number of defaults, number of rated firms, percent. They are named as 'defaults_volumn_numerator' and 'defaults_volumn_numerator', respectively.

Details

default_pro is to combine the number of rated firms from output_rated_firms with the number of defaulters from output_rated_firms and it also generate a new column for the default probabilities. The following three functions are used to plot time series of defaulters, rated firms, and IDP in different style, line or bar.

Examples

```
all.idp_output_IDP <- default_pro (all.idp_output_defaults, all.idp_output_rated_firmss)

plot_line(all.idp_output_IDP, 'year', 'defaults_numerator')

plot_bar(all.idp_output_IDP, 'date_YMD', 'defaults_numerator')
```

Practice – Identification of IDP in the US industrials from 1980/01/01 to 2013/12/31

The following Algorithm 1 shows how to use these functions to identify the numerator and denominator of IDP.

Algorithm 1 Extracting defaulters and rated firms in the U.S. industrial sector from 1980 to 2013

Main Inputs

'1959-09-01' (start date for collecting accounting and stock information), '2015-01-30' (end date for collecting accounting and stock information), 2/6 (Number of quarterly/monthly stock information for selecting a firm), cheq/che (quarterly/annually Cash and Short-Term Investments), dlcq/dlc (quarterly/annually Debt in Current Liabilities), 145 (country code in Moody's), 'INDUSTRIAL' (Economic Sector), '1980-01-01' (Starting Period), '2013-12-31' (Ending Period), 730 (keeping the first default event within two years), FALSE (the last logical inputs decide listed or non-listed firms)

Functions:

```
DATA <- Merging_Compustat.CRSP_1 ('1959-09-01', '2015-01-30', 6)

# 1 - cheq: Cash and Short-Term Investments

temp_DATA <- Merging_Compustat.CRSP_2 (DATA, 'cheq', 'che')
temp_DATA <- Merging_Compustat.CRSP_3 (temp_DATA, 2)
temp_DATA$cheq <- temp_DATA$var_quarter
temp_DATA$var_quarter <- NULL
"cheq" %in% names(temp_DATA)
# 2 - dlcq: Debt in Current Liabilities
temp_DATA <- Merging_Compustat.CRSP_2 (temp_DATA, 'dlcq', 'dlc')
temp_DATA <- Merging_Compustat.CRSP_3 (temp_DATA, 2)
temp_DATA$dlcq <- temp_DATA$var_quarter
temp_DATA$var_quarter <- NULL
"dlcq" %in% names(temp_DATA)
DATA_conversion <- temp_DATA

all.idp_1 <- NUM_defaults_filter_1 (145, 'INDUSTRIAL', '1980-01-01', '2013-12-31', TRUE, FALSE, FALSE)
all.idp_2 <- NUM_defaults_2 (all.idp_1, TRUE, FALSE, FALSE)
all.idp_3 <- NUM_defaults_3 (all.idp_2, TRUE, FALSE, FALSE)
all.idp_4 <- NUM_defaults_4 (all.idp_3, 730, TRUE, FALSE, FALSE)
all.idp_5 <- NUM_defaults_5_internal.drivers (all.idp_4, TRUE)
all.idp_output_defaults <- output_defaults(all.idp_5, '1980/1/1', 420)

all.idp_DATA_1 <- DEN_rated_filter_1 (145, 'INDUSTRIAL', '1980-01-01', '2013-12-31', all.idp_4, TRUE,
FALSE, FALSE)
all.idp_DATA_2 <- DEN_rated_2 (all.idp_DATA_1, '1980-01-01', '2013-12-31', all.idp_5, TRUE)
all.idp_output_rated_firms <- DEN_rated_3 (all.idp_DATA_2, '1980-01-01', '1980-02-01', 420)
all.idp_output_IDP <- default_pro (all.idp_output_defaults, all.idp_output_rated_firms)
```

Output

The number of firms and rated firms and IDP are recorded in the **all.idp_output_IDP.**

9.3.3 Functionalities of Construction of Value-weighted Default Probabilities (VDP)

Structure

1. Algorithm of VDP and Data File Description

2. From Module – 1st to 5th Module are about identification of denominators of Default probabilities

Module – ZERO: Merging two databases: Compustat and CRSP

Module – 1 Preliminary clean including country, market, period, economic sector, bond type

Module – 2 Cleaning family structures

Module – 3 Retaining the first default event

Module – 4 Cleaning defaults in terms of considering effective default data by merging CRSP

Module – 5 Summarizing the results of defaulted bonds

3. From Module – 6th to 8th they are about identification of denominators of Default probabilities

Module – 6 Preliminary clean including country, period, economic sector

Module – 7 Identifying effective rating firms after considering CRSP

Module – 8 Extracting the number of total value of ratted bonds and issues

4. From Module – 9th to 10th they are about the summary of DPs

Module – 9 Identification of VDP and issues-weighted DP

Module – 10 Summary of IDP and VDP

5. Practice - identification of VDPs in the US industrials from 1980/01/01 to 2013/12/31

Algorithm of VDP and Data File Description

(1) Method of calculation of VDP

VDP is value-weighted default probabilities. The method is calculated by defaulters and the formula is shown below:

$$VDP = \frac{d_t^{VDP}}{n_t^{VDP}} \tag{68}$$

Under this method, this research defines the outstanding number of credits (n_t^{VDP}) recursively as:

$$n_t^{VDP} = n_{t-1}^{VDP} - d_{t-1}^{VDP} - w_t^{VDP}. \tag{69}$$

The numerators d_t^{VDP} and d_{t-1}^{VDP} are the value of defaulted straight bonds at period t and $t-1$. The denominators n_{t} and $\mathrm{n}_{\mathrm{t-1}}$ are the value of rated straight bonds that potentially could have defaulted at date t and $t-1$. w_t^{VDP} denotes the number of credits which withdraw between periods t and $t+1$ (Moody, 2015).

(2) Description of Relevant Data Files

There are 11 files used for extracting the value of defaulted straight bonds at date and the number of rated firms at period t and $t-1$. In Table 9-3 there are 7 files that are used for calculation of VDP in DRD. All the historical defaults for the debt instruments can be found in the file called dflt_issu.txt and rated debts in each date can be found in senr_ratg_bond_only.txt. mast_issr.txt shows the detailed information about each firm such as country, sector, name etc. mast_issu.txt shows descriptive debt details of each issue in the data set including coupon rate, currency, maturity, etc. look_up.txt is used for identifying the code (REG) for the straight bond in DRD. family_structure.txt is used to determine the firms' parent firms in the history. Moodys_cusip.txt is used to merge CCM database/CRSP and DRD. These files should be uploaded into R environment.

Table 9-3 Relevant files for calculating VDP in DRD

Names	Descriptions
dflt_issu.txt	For each defaulted issuer, this table has information on the debt issues outstanding at the time of default (including non-defaulted issues).
senr_ratg_bond_only.txt	This table stores the issuer-level estimated senior unsecured rating & history based only on the bonds that an issuer has outstanding.
mast_issu.txt	Stores descriptive debt details of each issue in the data set including coupon rate, currency, maturity, etc.

mast_issr.txt	Stores descriptive details of each issuer in the data set.
issr_ids.txt	Links Moody's Issuer Number to one or more identifiers external to this database (such as Moodys.com identifiers, tickers, and Base CUSIPs), where available.
family_structure.txt	Stores default histories in the database; defaults can have more than one "default event" listing in this table.
Look_up.txt	For fields in other tables using abbreviated codes (except for MDY_DOMN_NUM), use this table to look up the text descriptions.
CRSP_daily.csv	Stores daily stock market data on major stock exchanges data on major stock exchanges (NYSE, AMEX, NASDAQ) from CRSP.
CRSP_monthly.csv	Stores monthly stock market data on major stock exchanges data on major stock exchanges (NYSE, AMEX, NASDAQ) from CRSP.
CCM_quarterly.csv	Stores quarterly accounting information from CRSP/Compustat merged database (CCM).
CCM_annual.csv	Stores annually accounting information from CRSP/Compustat merged database (CCM).
Moodys_cusip.csv	Stores all the firms that can match the firms recorded in Compustat. We extract all the firms' CUSIP values from the file called issr_ids.txt (in Moody's DRD) and then upload the 6 digits CUSIP to Wharton Research Data Services to obtain these firms' PERMNO codes in Compustat. We save the file including several columns such as "COMNAN", "NCUSIP", "TICKER", and "PERMNO" as Moodys_cusip. In order to identify the listed firms from rated firms in Moody's database, we need this file Moodys_cusip. If the function has the logical value to control whether we need listed firms or not, this file should be used.

Source: The first seven files are from Moody's DRD database (https://www.moodys.com/Pages/Default-and-Recovery-Analytics.aspx). For the detail description of each file, please refer to *DRD Technical Specifications* (https://www.moodys.com/sites/products/ProductAttachments/MA%20DRD%20Tech%20Specs.pdf. Both CRSP and CCM are downloaded from Wharton Research Data Services (http://wrds-web.wharton.upenn.edu/wrds/).

Module – ZERO: Merging two databases: Compustat and CRSP

Same as Module – ZERO in the Functionalities of Construction of Issuer-weighted Default Probabilities (IDPs)

Module – 1: Preliminary clean including country, market, period, economic sector, bond type

VDP_NUM_defaults_filter_1/ VDP_NUM_defaults_filter_2

Description

Extracts the basic information, such as,

(1) issues' information
(2) Filter condition <1> - Country
(3) Filter condition <2> - US bond Market
(4) Filter condition <3> - Economic Sectors
(5) Filter condition <2> - Time period

Usage

```
VDP_NUM_defaults_filter_1(country, market, economic_sector, market_used = T)
```

```
VDP_NUM_defaults_filter_2 (x, bond, date_1, date_2, all.firms = T, listed.firms = T, Non.listed.firms = T)
```

Arguments

country	a numeric value, specifying country's code, US country in Moody's is 145.
market	a string value, identifying debts issued by US firms in the US market and in the US and Non-US market in the economic sector
economic_esctor	a string value, specifying which economic sector that we need to choose. For Moody's database, 'INDUSTRIAL' is used for representing all of firms in industrials.
market_used = T	a logical value, controlling whether the input "market" is used.
x	a data frame that is constructed in the pervious function.

bond	a string value, specifying bond type (Straight bond is 'REG' in Moody's DRD).
date_1	a string value, specifying start date for extracting default data.
date_2	a string value, specifying end date for extracting default data.
listed.firms = T	a logical value, controlling whether listed firms are considered or not.
Non.listed.firms = T	a logical value, controlling whether Non-listed firms are considered or not.

Details

Both functions use following files, mast_issr.txt, mast_issu.txt, look_up.txt, deflt_issu.txt, issr_ids, and Moodys_cusip.csv. They are used to screen the country, economic sector, bond type, traded market, and time period.

VDP_NUM_defaults_filter_1 can identify debts issued by specified country's firms in that country's market, for example, if here is 'USD', it means we consider all the bonds issued in the US. This value is controlled by the logical value "market_used = T".

In fact, bond has only one value that we can select, as only straight bonds ('REG') can be studied in our current research.

For the logical value of Non.listed.firms, if non-listed firms are considered in this research, the Module (or functions) used in the next procedures are Module 2 (for cleaning family structure), Module 3 (for cleaning default event within XXX days) and Module 6 (for obtaining denominators of default probabilities). Then Module 4 (for cleaning defaults in terms of considering effective default data (that can be used for identifying accounting information) and Module 7 (for identifying effective rating firms after considering CRSP) will be useless for the study (Non-listed firms have no accounting information).

Note: the bond record is from 1890 to 2014 and actually we only choose data from 1980 to 2014 since there is no any defaults about straight bonds prior to 1980 in Moody's DRD.

Examples

```
all.vdp_1 <- VDP_NUM_defaults_filter_1 (145, 'USD', 'INDUSTRIAL', TRUE)

all.vdp_2 <- VDP_NUM_defaults_filter_2 (all.vdp_1, 'REG', '1980-01-01', '2013-12-31',
TRUE, FALSE, FALSE)
```

Module – 2: Cleaning family structures

VDP_NUM_defaults_3

Description

Combines defaults by firms related through parent-subsidiary relationships. There are two functions in this Module in order to handle two issues in cleaning family structure. They are:

1) Some firms experienced restructure events because of default events or expansion (acquistion/merger), then these firms have different parent firms in the past. In this context, we need to screen these firms;
2) Next issue is to identify the parent firm and their subsidiaries in the same default date and then we need to choose parent firms and delete their subsidiaries.

Usage

```
VDP_NUM_defaults_3(x, y, all.firms = T, listed.firms = T, Non.listed.firms = T)
VDP_NUM_defaults_4(x, all.firms = T, listed.firms = T, Non.listed.firms = T)
```

Arguments

x (in VDP_NUM_defaults_3)	a data frame that is constructed in the previous function VDP_NUM_defaults_filter_1.
x (in VDP_NUM_defaults_4)	a data frame that is constructed in the previous function VDP_NUM_defaults_filter_3.
y	a data frame that is constructed in the previous function VDP_NUM_defaults_2.
all.firms = T	a logical value, controlling whether all firms are considered or not.
listed.firms = T	a logical value, controlling whether listed firms are considered or not.
Non.listed.firms = T	a logical value, controlling whether Non-listed firms are considered or not.

Details

The first function is used for dealing with the first issue of firms having different parent firms in the past; the second function is to select parent firms rather than its subsidiaries. Keeping the order of using two functions is necessary.

Examples

```
all.vdp_3 <- VDP_NUM_defaults_3 (all.vdp_1, all.vdp_2, TRUE, FALSE, FALSE)
all.vdp_4 <- VDP_NUM_defaults_4 (all.vdp_3, TRUE, FALSE, FALSE)
```

Module – 3: retaining the first default event

VDP_NUM_defaults_5

Description

Retain the first default event within a certain span of time.

Usage

```
VDP_NUM_defaults_5 (x, days, all.firms = T, listed.firms = T, Non.listed.firms = T)
```

Arguments

x	a data frame that is constructed in the pervious function.
days	a numeric value, specifying the number of days used for screening default events within a certain period.
all.firms = T	a logical value, controlling whether all firms are considered or not.
listed.firms = T	a logical value, controlling whether listed firms are considered or not.
Non.listed.firms = T	a logical value, controlling whether Non-listed firms are considered or not.

Details

This function is to retain the first default event if the firm has more than two defaults within two years. days is 740 which means 365 days * 2 years = 730 days. If we want to consider one year, then the value of days will be 365.

Examples

```
all.vdp_5 <- VDP_NUM_defaults_5 (all.vdp_4, 730, TRUE, FALSE, FALSE)
```

Module – 4: Cleaning defaults in terms of considering effective default data (that can be used for identifying accounting information)

VDP_NUM_defaults_6_internal.drivers

Description

Clean defaults after merging two databases, Moody's DRD and CRSP and we only choose the firms that have stock price information. In particular, there are two issues that should be concerned in this function, which are shown below:

1) Obtaining CUSIP value in Moody's and connecting Moody's database with CRSP;
2) Extracting defaults (with defaults) that can successfully match with the firms existing in CRSP.

Usage

```
VDP_NUM_defaults_6_internal.drivers (x, Internal.risk.drivers = T)
```

Arguments

x	a data frame that is constructed in the pervious function in Module 5.
Internal.risk.drivers = T	a logical value, controlling whether the internal risk drivers are concerned in practice.

Detail

The firms with less than 6 months stock price information should be deleted in this codes.

Examples

```
vdp_data_6 <- VDP_NUM_defaults_6_internal.drivers (all.vdp_5, TRUE)
```

Module – 5: Summarizing the results of defaulted bonds

VDP_output_defaults_7

Description

Includes two functions in this part which are used for:

1) Summarizing defaults based on volume of bonds in each month and each year;
2) Summarizing defaults based on number of issues in each month and each year.

Usage

<1> Sum the number of total value of defaulted bonds in each month

```
VDP_output_defaults_7 (x, date_3, length_time)
```

<2> Sum the number of issuers that issued bonds going to default in each month

```
VDP_output_defaults_8 (x, date_3, length_time)
```

```
Output_debt_types (x)
```

Arguments

x	a data frame that is constructed in the function in Module 4.
date_3	a string value, specifying start date for extracting defaults; however, it should be the beginning of February of that year.
length_time	a numeric value, specifying the number of months from the start date to end date. For example, there are 408 observations [((2013+1)-1980)*12 = 408] from Jan 1, 1980 to Dec 31, 2013.

Details

All the functions are to summarize the defaults. One is to summarize the total value of defaulted bonds each month; another is to show the number of defaults based on issues. The last one is to provide the number of different types of defaulted bonds, such as senior unsecured debt, subordinated debt, and preferred stock etc.

date_3 in function VDP_output_defaults_8 is used to fill missing dates that have no defaults.

Examples

```
all.vdp_output_defaults <- VDP_output_defaults_7 (all.vdp_6, “1980-02-01”, 420)
```

```
debt_types_all <- output_debt_types (all.vdp_6)
```

Module – 6: Preliminary clean including country, period, economic sector

VDP_DEN_rated_filter_1

Description

Extracts basic information about rated firms

1) Issuer's information
2) Filter condition <1> - Country
3) Filter condition <2> - Economic Sectors
4) Filter condition <3> - Bond type

Usage

```
VDP_DEN_rated_filter_1 (country, bond, market, economic_sector, date_1, date_2, x, market_used = T, all.firms = T, listed.firms = T, Non.listed.firms = T)
```

Arguments

country	a numeric value, specifying country's code, US country in Moody's is 145.
bond	a string value, specifying bond type (Straight bond is 'REG' in Moody's DRD).
market	a string value, identifying debts issued by US firms in the US market and in the US and Non-US market in the economic sector
economic_sector	a string value, specifying the specifc economic sector. For Moody's database, 'INDUSTRIAL' is used for representing all of firms in industrials.
date_1	a string value, specifying start date for extracting accounting data that is lag time of date for extracting defaults.
date_2	a string value, specifying end date for extracting accounting data that is lag time of date for extracting defaults.
x	a data frame that is produced by Function VDP_NUM_defaults_5.
market_used = TRUE	a logical value, controlling whether previous input "market" is used.
Non.listed.firms = TRUE	a logical value, controlling whether Non-listed firms are considered or not. If it is TRUE, it means Non-listed firms are concerned; otherwise, listed firms are considered.

Details

This function is to extract country, economic sector, market, and bond.

It can also identify debts issued by specified country's firms in that country's market, for example, if here is 'USD', it means we consider all the bonds issued in the US. This is controlled by the logical value "market_used = TRUE".

Input (x) can be the results from the other functions, such as VDP_NUM_defaults_5 and VDP_NUM_defaults_4. This result of rated firms' information depends on whether we need to concern two screening conditions including parent-subsidiary relationship, and multiple defaults.

.

Examples

all.vdp_DATA_1 <- VDP_DEN_rated_filter_1 (145, 'REG', 'USD', 'INDUSTRIAL', '1980-01-01', '2013-12-31', vdp_data_5, TRUE, FALSE)

Module – 7: Identifying effective rating bonds after considering CRSP/Compustat

VDP_DEN_rated_2

Description

This function is to merge Moody's DRD with DATA_conversion from Model-ZERO.

Usage

VDP_DEN_rated_2 (x, date_1, date_2, y, Internal.risk.drivers = T)

Arguments

x	a data frame that is constructed in the previous function.
date_1	a string value, specifying start date for extracting accounting data that is lag time of date for extracting defaults.
date_2	a string value, specifying end date for extracting accounting data that is lag time of date for extracting defaults.
y	a data frame that is produced by Function - **VDP_NUM_defaults_6**.

Internal.risk.drivers = T	a logical value, controlling whether the internal risk drivers are concerned in practice.

Details

Both date inputs including date_1 and date_2 and y are used for calculating the number of bonds (firms) rated and non-defaulting bonds (firms) during this period **after merging Compustat/CRSP with Moody's**.

Examples

```
all.vdp_DATA_2 <- VDP_DEN_rated_2 (all.vdp_DATA_1, '1980-01-01', '2013-12-31', all.vdp_6, TRUE)
```

Module – 8: Extracting the number of total value of ratted bonds and issues

VDP_output_rated_3/ VDP_output_rated_4

Description

Extracts monthly rated bonds and issues.

Usage

<1> Summary of total value of bonds rated

```
VDP_output_rated_3 (x, date_1, date_3, length_time)
```

<2> Summary of total number of issues

```
VDP_output_rated_4 (x, date_1, date_3, length_time)
```

Arguments

date_1	a string value, specifying start date for extracting defaults.
date_3	a string value, identifying the end of each month. In particular, the format of date is in Feb of start year.

length_time	a numeric value, specifying the number of months from the start date to end date. For example, there are 408 observations [(2014-1980)*12 = 408] from Jan 1, 1980 to Dec 31, 2013.

Details

date_3 is used for identify the end of each month. This value is to identify the withdrawn in calculation of denominators of default probabilities.

Examples

vdp_output_rated_bonds <- VDP_output_rated_3 (vdp_DATA_2, '1980-01-01', '1980-02-01', 408)

vdp_output_rated_issues <- VDP_output_rated_4 (DATA_2, '1980-01-01', '1980-02-01', 408)

Module – 9: Identification of VDP and issues-weighted DP

VDP_default_pro

Description

This function is to combine two databases (vdp_output_defaults, vdp_output_defaults_issues, vdp_output_rated_bonds, vdp_output_rated_issues) together and then calculate value-weighted DP and issues-weighted DP.

Usage

VDP_default_pro (x, y)

plot_line (x, frequency, var)

plot_bar (x, frequency, var)

plot_black_outline (x, frequency, var)

Arguments

x (in VDP_default_pro)	a data frame that is constructed in the function vdp_output_defaults.
y	a data frame that is constructed in the function vdp_output_rated_bonds.
x (in plot_line, plot_bar, and plot_black_outline)	a data frame that is constructed in the function VDP_default_pro.
frequency	a string value that is the variable in the data frame. It can be either monthly or yearly, such as, 'date_YM', and 'year'.
var	a string value that is the variable in the data frame generated by VDP_default_pro, such as, 'defaults_volume_numerator' (the number of defaults), 'volume_issues' (the number of rated bonds), and 'Volume_percent' (their percent).

Details

VDP_default_pro is to combine the number of rated bonds from vdp_output_rated_bonds (the number of rated issues from vdp_output_rated_issues) with the number of defaulters from vdp_output_defaults (the number of defaulted issues from vdp_output_defaults_issues) and it also generate a new column for both value-weighted and issues-weighted default probabilities. The following three functions are used to plot time series of defaulters, rated bonds/issues, and VDP in different style, line or bar.

Examples

```
vdp_output_VDP <- VDP_default_pro (vdp_output_defaults, vdp_output_rated_bonds)

plot_line(vdp_output_VDP, 'date_YM', 'defaults_volumn_numerator')

plot_bar(vdp_output_VDP, 'year', 'defaults_volumn_numerator')

plot_black_outline (vdp_output_VDP, 'date_YM', 'defaults_volumn_numerator')
```

Module – 10: Summary of IDPs and VDPs

Summary_IDP_VDP

Description

This function is to combine two constructed databases including VDP and IDP together.

Usage

```
Summary_IDP_VDP (x, y)

plot_line_multiple (x, frequency, var1, var2)

plot_line (x, frequency, var)

plot_bar (x, frequency, var)
```

Arguments

x	a data frame that is constructed in the function VDP_default_pro.
y	a data frame that is constructed in the function IDP_default_pro.
frequency	a string value that is the variable in the data frame. It can be either monthly or yearly.
var1	A string value that is the variable (IDP).
var2	A string value that is the variable (VDP).
var	A string value that can be 'defaults_numerator' (No.defaulted firms), 'Firms' (No.rated firms), 'defaults_volumn_numerator' (Total value of defaulted bonds), 'volume_issues' (Total value of rated bonds), 'defaults_issuers_numerator' (No.defaulted issues), and 'Issuers' (No rated issues)

Examples

```
Summary_all_DPs_Duffie <- Summary_IDP_VDP (vdp_output_VDP, idp_output_IDP)

plot_line_multiple (Summary_all_DPs_Duffie, 'year', 'IDP', 'VDP')

plot_bar(Summary_all_DPs_Duffie, 'year', 'Firms')

plot_bar(Summary_all_DPs_Duffie, 'year', 'defaults_volumn_numerator')
```

Practice – Identification of IDPs in the US industrials from 1980/01/01 to 2014/12/31

The following Algorithm 2 shows how to use these functions to identify the numerator and denominator of VDP and merge IDP and VDP databases together.

Algorithm 2 Extracting defaulters and rated bonds in the U.S. industrial sector from 1980 to 2013

Main Inputs

'1959-09-01' (start date for collecting accounting and stock information), '2015-01-30' (end date for collecting accounting and stock information), 2/6 (Number of quarterly/monthly stock information for selecting a firm), cheq/che (quarterly/annually Cash and Short-Term Investments), dlcq/dlc (quarterly/annually Debt in Current Liabilities), 145 (country code in Moody's), 'USD' (issued by the U.S. firms), 'INDUSTRIAL' (Economic Sector), 'REG' (straight bonds), '1980-01-01' (Starting Period), '2014-12-31' (Ending Period), 730 (keeping the first default event within 2 years)

Functions:

```
DATA <- Merging_Compustat.CRSP_1 ('1959-09-01', '2015-01-30', 6)

# 1 - cheq: Cash and Short-Term Investments
temp_DATA <- Merging_Compustat.CRSP_2 (DATA, 'cheq', 'che')
temp_DATA <- Merging_Compustat.CRSP_3 (temp_DATA, 2)
temp_DATA$cheq <- temp_DATA$var_quarter
temp_DATA$var_quarter <- NULL
"cheq" %in% names(temp_DATA)

all.vdp_1 <- VDP_NUM_defaults_filter_1 (145, 'USD', 'INDUSTRIAL', TRUE)
all.vdp_2 <- VDP_NUM_defaults_filter_2 (all.vdp_1, 'REG', '1980-01-01', '2014-12-31',
TRUE, FALSE, FALSE)
all.vdp_3 <- VDP_NUM_defaults_3 (all.vdp_1, all.vdp_2, TRUE, FALSE, FALSE)
all.vdp_4 <- VDP_NUM_defaults_4 (all.vdp_3, TRUE, FALSE, FALSE)
all.vdp_5 <- VDP_NUM_defaults_5 (all.vdp_4, 730, TRUE, FALSE, FALSE)
all.vdp_6 <- VDP_NUM_defaults_6_internal.drivers (all.vdp_5, TRUE)
all.vdp_output_defaults <- VDP_output_defaults_7 (all.vdp_6, '1980-02-01', 420)
debt_types_all <- output_debt_types (all.vdp_6)

all.vdp_DATA_1 <- VDP_DEN_rated_filter_1 (145, 'REG', 'USD', 'INDUSTRIAL', '1980-01-
01', '2014-12-31', all.vdp_5, TRUE, TRUE, FALSE, FALSE)
all.vdp_DATA_2 <- VDP_DEN_rated_2 (all.vdp_DATA_1, '1980-01-01', '2014-12-31',
all.vdp_6, TRUE)
all.vdp_output_rated_bonds <- VDP_output_rated_3 (all.vdp_DATA_2, '1980-01-01',
'1980-02-01', 420)
all.vdp_output_VDP <- VDP_default_pro (all.vdp_output_defaults,
all.vdp_output_rated_bonds)
summary_DPs_all <- Summary_IDP_VDP (all.vdp_output_VDP, all.idp_output_IDP)
```

Output

The time series of IDP and VDP are recorded in the summary_DPs_all.

9.3.4 Functionalities of Construction of Internal Risk Drivers

Structure

Description of internal risk drivers

Module – 1: Construction of DTD.MONTHLY, SIGMA and RETURN.ANNUAL

Module – 2: Construction of DTD.ANNUALLY

Module – 3: Construction of accounting ratios

Module – 3: Construction of industry-level of internal risk drivers

9 Appendix

Overview-Description of internal risk drivers

Based on the finalised database about rated firms in each month, such as all.idp_DATA_2, the following functions are able to construct 35 internal risk drivers for these rated firms. All these variables are constructed using the CRSP and COMPUSTAT data items. The main methods of constructing accounting ratios are from Shumway (2001) and Campbell et al. (2008). In order to fill in the missing values in each accounting variable, we used daily, monthly, annually CRSP and quarterly and annually Compustat. For the methods of how to handle missing values, we used same approach as Duffie et al. (2007). Winsorizing each variable is used to limit the influence of outliers. Specifically, we replace any observation below the 5th percentile with the 5th percentile, and any observation above the 95th percentile with the 95th percentile. Below we provide details for 35 risk drivers in Table 9-4. In order to convert panel data to time series data for each micro factor, we average each firm's driver in each month.

Table 9-4 Data description for 35 internal risk drivers

Main	Predictors	Description	Total
(a) Profitability ratios	NITA	Net income over total assets	8
	NIMTA	Net income over market value of total assets	
	EBITTA	Earnings before interest and taxes over total assets	
	EBITMTA	Earnings before interest and taxes over market value of total	
	RETA	Retained earnings over total assets	
	REMTA	Retained earnings over market value of total assets	
	NISALE	Net income over sales	
	EBITSALE	Earnings before interest and taxes over sales	
(b) Liquidity ratios	CACL	Current asset over current liabilities	6
	METL	Market equity to total liabilities	
	WCTA	Working capital over total assets	
	WCMTA	Working capital over market value of total assets	
	CSHTA	Cash and short-term investments over total assets	
	CSHMTA	Cash and short-term investments over market value of total	
(c) Leverage ratios	TLTA	Total liabilities over total assets	6
	TLMTA	Total liabilities over market value of total assets	
	CLTA	Current liabilities over total assets	
	CLMTA	Current liabilities over market value of total assets	
	RECL	Retained earnings over current liabilities	
	MB	Market-valued total assets over book value of firm[44]	
(d) Activity ratios	REVWC	Revenue over working capital	4
	CLSALE	Current liabilities over sales	
	SALETA	Sales over total assets	
	SALEMTA	Sales over market value of total assets	
(e) Other variables	S.DEBT	"short term" debt over total debt	11
	BA	log(book asset value)	
	SIZE	Log(market capitalization)	
	SA	Log(sale)	
	PRICE	Log (equity price)	
	EXRET	log of gross excess return over value-weighted S&P 500 return	
	RETURN.MONTHLLY	The firm's trailing one-month stock return	
	RETURN.ANNUAL	The firm's trailing one-year stock return	
	SIGMA.DAILY	Standard deviation of each firm's daily stock return over the	
	DTD.ANNUALLY	Merton distance to default (MDTD) (Annually)*	
	DTD.MONTHLY	Market value proxy distance to default (MVPDTD)	
		Total	35

Note: * The calculation of both types of DTD is shown as follows.

[44] Book value of firm is calculated by Total assets- Intangible Assets- Total liabilities)

[45] The formula for calculating $EXRET_{i,t}$is $EXRET_{i,t} = \log(1 + R_{i,t}) - \log(1 + R_{S\&P500,t})$

Module – 1: Construction of DTD.MONTHLY, SIGMA and RETURN.ANNUAL

Constructing.drivers_1

Description

Constructs four interval risk drivers in Table 9-4, including:

1) DTD.MONTHLY: distance to default (DTD) based on market value proxy approach;
2) SIGMA: standard deviation of each firm's daily stock return over the past 3 months;
3) RETURN.ANNUAL: annual stock return.
4) RETURN.MONTHLLY: monthly stock return.

Usage

```
Constructing.drivers_1 (x, y, date_1, date_2)
```

Arguments

x	a data frame that should be DATA_conversion.
y	a data frame that is constructed by Module 7 in both functionalities of constructions of IDP and VDP.
date_1	a string value, specifying start date.
date_2	a string value, specifying end date.

Details

This function is to construct three internal risk drivers. The first risk driver is distance to default calculated by market value proxy method (Duan and Wang, 2012). The algorithm of measuring distance to default based on market value proxy approach is shown below. The definition of DTD is that the number of standard deviations the market value of assets is away from default for a given firm, which has been widely used by academics in financial study and extensively employed by financial practitioners in the industrial world. There are two approaches to calculate DTD. The first one is written

$$DTD = \frac{ln(V_t/L_T) + (\mu_A - \frac{\sigma_A}{2})(T-t)}{\sigma_A\sqrt{T-t}}, \quad (70)$$

where V_t is the market value of assets at time t, L_T is the firms liabilities also named as default point at time T, σ_A is the volatility of market value of assets and μ_A is the mean rate of asset growth. The second one is illustrated below:

$$DTD_t = \frac{V_t - L_T}{V_t \sigma_A}, \tag{71}$$

where V_t is the market value of assets at time t, L_t is the firms liabilities also named as default point at time T, σ_A is the volatility of market value of assets. This approach avoids to use μ_A.

From both formulas, it can be seen that if a firm' market value of assets is larger than its liabilities, then this firm is not going to default since the firm has significant buffer to absorb losses in its asset value. Another case is that if the asset volatility a firm is smaller, then the likelihood of firm default is small since this firm is stable without significant influence from external shocks. Duan and Wang (2012) classify current four popular measures to calculate DTD, which are the market value proxy method (Eom et al., 2004), the volatility restriction method (Ronn and Verma, 1986), the KMV iterative method (Crosbie and Bohn, 2002), and the transformed-data maximum likelihood (Duan, 2000). In particular, for the KMV iterative method, many following studies have created similar iterative approach in terms of how to iteratively calculate and use the volatility of market value of firm asset, such as Duffie et al. (2007), Bharath and Shumway (2008) and Vassalou and Xing (2004). Duan and Wang (2012) find that the results of using these approaches are not quite big different in using non-financial firms. The main difference is for financial firms and KMV iterative approach performs worse than the other approach since the large proportion of liabilities of financial firms cannot be accounted by this approach. Consequently, this package only concerns two approaches including market value proxy approach called MVPDTD and iterative approach named MDTD, respectively. The main difference between two approaches is that whether the market value of firm asset and its volatility are estimated in practice. This part is to illustrate MVPDTD.

The MVPDTD model is DTD based on market-proxy model, which is mainly different from MDTD in terms of no using estimated market value of firm assets and its volatility. This approach is directly estimate the values for μ_A and σ_A from daily asset values by summing daily updated market capitalization with quarterly updated book values of liabilities. Since this method is to calculate monthly DTD, then forecasting horizon is a month. That is, both μ_A and σ_A are obtained based on daily logarithmic asset returns in each month. The risk free rate is the compounded return for a month, which is shown

$$(1 + \text{return of } V\ )^{30} + 1. \tag{72}$$

Then we obtain all the variables in Equation (71) and MVPDTD can be calculated.

The second interval risk driver is to measure standard deviation of each firm's daily stock return over the past 3 months named as SIGMA, which is introduced by Shumway (2001). Simply, we regress the each stock's monthly returns in year $t-1$ on the value-weighted NYSE/AMEX index return for the same year, which is regarded as each firm's sigma for year t. Sigma is the standard deviation of the residual of this regression. If there are less than twelve months of returns, then the corresponding value should be removed in the regression. To avoid outliers, sigma are all truncated at the 99th and 1st percentile values.

Reference

BHARATH, S. T. & SHUMWAY, T. 2008. Forecasting Default with the Merton Distance to Default Model. Review of Financial Studies, 21, 1339-1369.

CROSBIE, P. & BOHN, J. 2002. Modelling Default Risk. KMV working paper.

DUAN, J.-C. & WANG, T. 2012. Measuring distance-to-default for financial and non-financial firms. Global Credit Review, 2, 95-108.

DUAN, J. C. 2000. Correction: Maximum Likelihood Estimation Using Price Data of the Derivative Contract (Mathematical Finance 1994, 4/2, 155–167). Mathematical Finance, 10, 461-462.

DUFFIE, D., SAITA, L. & WANG, K. 2007. Multi-period corporate default prediction with stochastic covariates. Journal of Financial Economics, 83, 635-665.

EOM, Y. H., HELWEGE, J. & HUANG, J.-Z. 2004. Structural Models of Corporate Bond Pricing: An Empirical Analysis. Review of Financial Studies, 17, 499-544.

RONN, E. I. & VERMA, A. K. 1986. Pricing risk-adjusted deposit insurance: An option-based model. The Journal of Finance, 41, 871-896.

SHUMWAY, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. Journal of Business, 74, 101-124.

VASSALOU, M. & XING, Y. 2004. Default risk in equity returns. Journal of Finance, 59, 831-868.

Examples

```
idp.data.1 <- Constructing.drivers_1 (DATA_conversion, all.idp_DATA_2, '1980-01-01', '2014-12-31')
```

Module – 2: Construction of DTD.ANNUALLY

Constructing.drivers_2.DTD

Description

Constructs a risk driver DTD.ANNUALLY (DTD based on Merton approach) in Table 9-4.

Usage

```
Constructing.drivers_2.DTD (x, date_1, date_2)
```

Arguments

x	a data frame that is constructed in previous function idp.data.1.
date_1	a string value, specifying start date.
date_2	a string value, specifying end date.

Details

The MDTD model is based on Merton model and this study follows a recipe similar to those of Duffie et al. (2007), Bharath and Shumway (2008), Vassalou and Xing (2004), and Crosbie and Bohn (2002). Specifically, the Merton (1974) model assumes that firms are financed by equity, with its value at time t denoted by E_t, and one single pure discount bond (denoted by D_t) with maturity date T and principal F. The asset value V_t follows a geometric Brownian motion:

$$dV_t = \mu_t V_t d_t - \sigma V_t dW_t, \tag{73}$$

where V_t is the market value of assets at time t, L_t is the firms liabilities at time t, μ_t is the expected continuously compounded return on V_t at time t, W_t is a standard Brownian motion. This formula can be interpreted as the firm that has issued just one discount bond maturing in T periods. This model further assumes that the equity of the firm is valued as a European call option on the value of assets. Specifically, the firm's equity is the call option on the underlying value of the firm, a strike price is equal to the face value of the firm's liabilities and a time-to-maturity is T. Then, the value of equity considered as a function of the total value of the firm is able to be defined by the Black-Scholes-Merton formula. The equity value at maturity T is $E_t = max(V_t - F, 0)$. Therefore, the equity value at time $t \leq T$ by the Black-Scholes option pricing formula becomes

$$E(V_t, \sigma) = V_t N(d_1) - F e^{-r(T-t)} N(d_2)\,, \tag{74}$$

where r is the instantaneous risk-free rate, $N(.)$ is the standard normal cumulative distribution function, specifically,

$$d_1 = \frac{\ln(V_t/F) + \left(r + \frac{1}{2}\sigma_A^2\right)(T-t)}{\sigma_A \sqrt{T-t}}, \tag{75}$$

and d_2 is written as follows

$$d_2 = d_1 - \sigma\sqrt{T-t}. \tag{76}$$

In order to calculate the estimated firm value and its volatility, MDTD should use two significant equations. One is based on the Black-Scholes-Merton Equation (74), showing the value of equity of a firm as a function of the value of the firm; another is associated with the volatility of its equity. Following assumptions of the Merton model, the volatility of its equity follows from Ito's lemma, which is written

$$\sigma_E = \left(\frac{V}{E}\right)\frac{\partial E}{\partial V}\sigma_A. \tag{77}$$

In the Black-Scholes-Merton model, Equation (75) can be shown as follows since $\frac{\partial E}{\partial V} = N(d_1)$

$$\sigma_E = \left(\frac{V}{E}\right) N(d_1)\sigma_A, \tag{78}$$

where d_1 is defined in Equation (75).

Both nonlinear equations (75) and (78) are used to calculate the value and volatility of a firm's equity.

According to the Merton (1974) model, the option' value is observed as the total value of the firm's equity, while the value of the underlying asset (the total value of the firm) is not directly observable. Thus, while V must be inferred, E can be easily calculated in the marketplace by multiplying the firm's shares outstanding by its current stock price. Similarly, in the Merton DTD model, the volatility of equity, σ_E , can be estimated but the volatility of the underlying firm, σ_V, must be inferred. This study uses the following step to calculated MDTD:

1) The first procedure is to estimate σ_E ,the volatility of equity, from historical stock returns from the CRSP database. I use daily data from the past 12 months to obtain this value;

2) Next is to select a forecasting horizon and a measure of face value of firm's liabilities. In this case, 1 year is used to forecast MDTD and the method of calculating firms' debt is the sum of short-term debt and one half of its long-term debt from Compustat, which is same as Duffie et al. (2007).

3) The third step is to collect values of the risk-free rate named one-year T-bill rate and the market equity of the firm.

4) After finishing these three steps, all the values except for V and σ_V , the total value of the firm and the volatility of firm value, respectively, are included in Equations (75) and (78). Then we returns the sum of squared errors for the two equations in order to calculate V and σ_V. In this case, the algorithm developed by Byrd et al. (1995) is used to solve this nonlinear optimization problem for identifying market value of the firm and the volatility of firm value.

5) The final step is to employ Equation (71) to calculate MDTD.

In particular, for simplicity, we avoid the calculation μ_A, the mean rate of asset growth used by prior studies mentioned early.

Reference

BHARATH, S. T. & SHUMWAY, T. 2008. Forecasting Default with the Merton Distance to Default Model. Review of Financial Studies, 21, 1339-1369.

BYRD, R. H., LU, P., NOCEDAL, J. & ZHU, C. 1995. A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing, 16, 1190-1208.

CAMPBELL, J. Y., HILSCHER, J. & SZILAGYI, J. 2008. In search of distress risk. Journal of Finance, 63, 2899-2939.

CROSBIE, P. & BOHN, J. 2002. Modelling Default Risk. KMV working paper.

DUAN, J.-C. & WANG, T. 2012. Measuring distance-to-default for financial and non-financial firms. Global Credit Review, 2, 95-108.

DUAN, J. C. 2000. Correction: Maximum Likelihood Estimation Using Price Data of the Derivative Contract (Mathematical Finance 1994, 4/2, 155–167). Mathematical Finance, 10, 461-462.

DUFFIE, D., SAITA, L. & WANG, K. 2007. Multi-period corporate default prediction with stochastic covariates. Journal of Financial Economics, 83, 635-665.

EOM, Y. H., HELWEGE, J. & HUANG, J.-Z. 2004. Structural Models of Corporate Bond Pricing: An Empirical Analysis. Review of Financial Studies, 17, 499-544.

MERTON, R. C. 1974. ON THE PRICING OF CORPORATE DEBT: THE RISK STRUCTURE OF INTEREST RATES*. Journal of Finance, 29, 449-470.

RONN, E. I. & VERMA, A. K. 1986. Pricing risk‐adjusted deposit insurance: An option‐based model. The Journal of Finance, 41, 871-896.

SHUMWAY, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. Journal of Business, 74, 101-124.

VASSALOU, M. & XING, Y. 2004. Default risk in equity returns. Journal of Finance, 59, 831-868.

Examples

idp.data.2 <- Constructing.drivers_2.DTD (idp.data.1, “1980-02-01”, “2015-01-01”)

Module – 3: Construction of accounting ratios

Constructing.drivers_3.ratios

Description

Constructs the other interval risk drivers in Table 9-4 except five risk drivers mentioned early.

Usage

Constructing.drivers_3.ratios (x, lag_time)

Arguments

x	a data frame that is constructed by Constructing.drivers_2.DTD or Constructing.drivers_1.
lag_time	a numeric value that is used to reset date for predicting IDPs/VDPs and it is based on monthly data.

Details

This function is to construct the other 30 internal risk drivers for predicting IDPs/VDPs in Table 9-4. lag_time is equal to 3 and it means to lag 35 variables by 3 months in Table 9-4.

Examples

```
idp.data.3 <- Constructing.drivers_3.ratios (idp.data.2, 3)
```

Module – 4: Construction of industry-level of internal risk drivers

Constructing.drivers_4.aggregation

Description

Constructs the other interval risk drivers in Table 9-4 except five risk drivers mentioned early.

Usage

```
Constructing.drivers_4.aggregation (x, date_1, date_2, y, lag_time, VDP = T)
```

Arguments

x	a data frame that is constructed by the function Constructing.drivers_3.ratios.
y	a data frame that is generated by the function Summary_IDP_VDP.
date_1	a string value, specifying start date.
date_2	a string value, specifying end date.
lag_time	a numeric value that is used to reset date for predicting IDPs/VDPs and it is based on monthly data.
VDP = T	a logical value, controlling whether the interval risk drivers are extracted for issuers or bonds' issuers (in IDP or VDP) or not.

Details

This function is used to convert firm level data (panel data of 35 firm specific variables) to industrial level data (time series of data for each variable). lag_time is the number of months which is used to lag each variable. It should be noted that although the database (x) is lagged, here this function still need same value for lagging variable. The reason of using repeated command is that in the future we may directly use the firm level data for the study. Then we still keep this command in this function.

Examples

idp.data.4 <- Constructing.drivers_4.aggregation (idp.data.3, '1980-01-01', '2014-12-31', summary_DPs_all, 3, F)

9.3.5 Other key functions for empirical analysis

data.factors

Description

Constructs various indicators.

Usage

data.indicator (data.factors, date.start, date.end, window.size, correlation_type, boundary.upper, penalised.norm, types.indicator = c ("correlation", "edge", "covariance", "variance"), calibrating.indicator = c ("sum", "average.factor", "average.pair"))

Arguments

data.factors	a data frame that each column represents various variable, and the row represents time series.
date.start	a string value, specifying start date.
date.end	a string value, specifying end date.
window.size	a numeric value that is used to give the window size for calculating the correlation among variables in the correlation matrix.
correlation_type	a string value, controlling the correlation methods, such as "pearson", "kendall", and "spearman".
boundary.upper	a numeric value that is used to control the minimum value of correlation coefficient that are considered in the calculation of indicator.
penalised.norm	a numeric value that is a penalty form in the indicator.

types.indicator	a string value, controlling which method is used for constructing the indicator. There are four methods including correlation, edge, covariance, and variance.
calibrating.indicator	a string value, controlling the method used to calibrate the indicator.

Details

The method of construction of the indictor is to capture the features in the observations, such as correlation, edges, covariance, and variance. This function applies the methods used by Gorban et al. (2010) for collecting correlation information. Specifically, the indictor is based on the Pearson correlation coefficient (r), and then we collect the these correlation information by using an approach called L_p norm

$$\|r\|_p = \left(\sum_{j>k} |r_{jk}|^p \right)^{\frac{1}{p}}, \tag{79}$$

where r_{jk} $(j > k)$ is actually the Pearson's r beween two variables. p is the penalty parameter. There are $n(n-1)/2$ correlation coefficients between various variable, r_{jk}. It should be noted that these values are loacted non-diagnoal part of the correlation matrix. If the strong correlations among variables is the major concern, then we can delete terms with values below threshold, $\alpha < 0.5$, from the sum $G_{p,\alpha}$

$$G_{p,\alpha} = \left(\sum_{j>k, |r_{jk}|>\alpha} |r_{jk}|^p \right)^{\frac{1}{p}}. \tag{80}$$

The quantity $G_{p,\alpha}$ is a p-weight of the $G - correlation\ graph$. The vertices of this graph correspond to variables, and these vertices are connected by the edges, if the absolute value of the correspondent sample correaltion coefficient exceeds α: $|r_{jk}| > \alpha$. In order to concern the dynamic change of components of matrix, we calibrate the Equation (79) and then the indcator is as follows:

$$\text{Indicator}_{correlation} = \frac{G_{p,\alpha}}{N}, \tag{81}$$

where N is the number of factors in the correlaiton matrix in Equation (80). For example, the correlation matrix is composed of macro factors from FRED. It is clear that the current method can convert specific correlation matrix to an observation. Since this study focuses on the prediction of indicator, then sliding time windows based on 9 months is used for obtaining the time series of indicator. If there is a time moment t, the indicator $G_{p,\alpha}$ is calculated in the time window $[t-9, t-1]$, which precedes t.

If the indicator is constructed by edges, then the indicator is written as follows:

$$\text{Indicator}_{edge} = \frac{\left(\sum_{j>k, |r_{jk}|>\alpha} |e_{jk}|^p \right)^{\frac{1}{p}}}{N}, \tag{82}$$

where e_{jk} is the edge between two variables. p is the penalty parameter . There are $n(n-1)/2$ edges between various variable, e_{jk}. It should be noted that these values are loacted non-diagnoal part of the correlation matrix. If the strong correlations among variables is the major concern, then we can delete terms with values below threshold, $\alpha < 0.5$. The rest of them decides how many edges among variables. N is the number of factors in the correlaiton matrix.

If the indicator is constructed by covariance, then the indicator is as follows:

$$\text{Indicator}_{covariance} = \|cov\|_p = \frac{\left(\sum_{j>k}\left|cov_{jk}\right|^p\right)^{\frac{1}{p}}}{N}, \tag{83}$$

where cov_{jk} $(j > k > 0)$ is actually covarnace between X and Y represent two variables, $cov_{X,Y} = \text{cov}(X, Y)$. p is the penalty parameter. Specifically, we have one dataset $\{X_1, , \ldots, X_n\}$ containing n values and another dataset containing n values. $\text{cov}(X, Y)$ is the covariance between X and Y . Then we have $n(n-1)/2$ covariance between various variable, cov_{jk}. It should be noted that these values are loacted non-diagnoal part of the covariance matrix. N is the number of factors in the cavariance matrix.

If the indicator is constructed by variance, then the indicator is as follows:

$$\text{Indicator}_{variance} = \|v\|_p = \frac{\left(\sum|v|^p\right)^{\frac{1}{p}}}{N}, \tag{84}$$

where v is actually variance of each variable. p is the penalty parameter. N is the number of factors.

The data frame is specific. The first five columns are not used to calculate the indicator. Then it should be cautious to construct the data frame. correlation_type depends on the method of calculation correlation coefficient, r_{jk}. boundary.upper is controlled by α in Equation (80) and (82). penalised.norm is controlled by p in Equation (80), (82), (83), and (84). Normally, the best value in calculating the indicator is one. calibrating.indicator is the value of N in the four equations mentioned above. There are three methods to calibrate the indicator. First, I sum up all the values. Second, I average each indicator value by the number of variables in the data frame. Third, I only concern the number of pairs after removing the non-favorable relationship by controlling α. This is only used in the Equation of (80) and (82).

References:

GORBAN, A. N., SMIRNOVA, E. V. & TYUKINA, T. A. 2010. Correlations, risk and crisis: From physiology to finance. *Physica A: Statistical Mechanics and its Applications*, 389**,** 3193-3217.

Examples

df <- data.indicator (dataframe, “1980-01-31”, “2014-12-31”, 9, “pearson”, 0.5, 1, types.indicator = c (“correlation”), calibrating.indicator = c (“sum”))

plot_network/plot_network.star

Description

Executing two types of network analysis.

Usage

plot_network (x, correlation_type, z, boundary.upper, boundary.down)

plot_network.star (x, correlation_type, z, boundary.upper, boundary.down)

Arguments

x	a data frame that each column represents various variable, and the row represents time series.
correlation_type	a string value, controlling the correlation methods, such as "pearson", "kendall", and "spearman".
z	a data frame that used to mark the variables that in the data frame x.
boundary.upper	a numeric value that is used to control the value of correlation coefficient that are considered in the calculation of indicator.
boundary.down	a numeric value that is used to control the value of correlation coefficient that are considered in the calculation of indicator.

Details

The function named as plot_network is used to do network analysis based on correlation analysis. Both boundary.upper and boundary.down are to control the boundaries of correlation coefficients. The former one is to control the larger boundary; and the latter one is to control the lower boundary. For example, if we choose 80% as the larger boundary in boundary.upper, and the correlation between two variables that have larger correlation coefficient (more than 80%) is represented by blue line. If we choose the 50% as the lower boundary in boundary.down, the correlation between two variables (that their correlation coefficients are more than 50% and smaller than 80%) is represented by red line.

plot_network.star is different from plot_network in terms of how to draw correlation graph among these variables in the data frame. This function is to show how a variable correlates with the other variables. The variable in the first column is the central variable, and the other variables are around along the circle.

Examples

plot_network (dataframe, "pearson", dataframe, 0.7, 0.5)

plot_network.star (dataframe, "pearson", dataframe, 0.7, 0.5)

9.4 User's Manual of Ph.D. Project Web Applications: KAI.APP.PhD/KAI.APP.EIA

Ph.D. Project Web Applications KAI.APP.PhD/KAI.APP.EIA

USER'S MANUAL

Kai PhD Candidate in Finance and Risk, Developer

Nottingham University Business School

July 01, 2016

Revision Sheet

Release No.	Date	Revision Description
1.0	01/03/2016	Initial Revision (Only covering KAI.APP.PhD)
2.0	15/03/2016	Second Revision (Adding KAI.APP.EIA)
3.0	12/04/2016	Third Revision (Debugging)
4.0	09/05/2016	Fourth Revision (Adding causality analysis).

USER'S MANUAL

TABLE OF CONTENTS

9.4.1 1. GENERAL INFORMATION 297

1.1 Apps Overview 297
1.2 Authorized Use Permission 298
1.3 Contact 298

9.4.2 2. APPLICATIONS SUMMARY 300

2.1 Understanding of Web Apps 300
2.2 Structure of the Inputs 300

9.4.3 3. GETTING STARTED 303

3.1 KAI.APP.PhD 303
3.1.1 Structure of KAI.APP.PhD 303
3.1.2 Overview the KAI.APP.PhD 304
3.1.3 Tab 1 and Tab 2 306
3.1.4 Tab 3.1 – 3.6 310
3.1.5 Tab 4.1 – 4.3 324
3.2 KAI.APP.EIA 325
3.2.1 Structure of KAI.APP.EIA 325
3.2.2 Overview the KAI.APP.EIA 326
3.2.3 Tab 3.0 – 3.6 328
3.2.4 Tab 4.1 – 4.3 328

1. GENERAL INFORMATION

9.4.1 1. GENERAL INFORMATION

1.1 Apps Overview

In order to do interactive exploration, I designed and developed two web applications including KAI.APP.PhD and KAI.APP.EIA by Shiny based on R programming languages. Both apps are intended to do basic empirical analysis in support of my PhD project. Basically, they have different roles in my PhD project which can achieve their specific purposes.

Specifically, KAI.APP.PhD has three main purposes. The first purpose is to find whether there is relationship between the correlation among specific group of factors (known as risk drivers) and economic recession and issuer-weighted default probabilities (IDP) by using network analysis. The second one is to construct various indicators based on different algorithms[46] and analyse the relationship between each indicator and IDP by using various statistical methods, such as correlation analysis, causality analysis, and statistical inference analysis. In particular, a set of relevant model diagnostics tests can be done in this app in order to verify whether the data fits the statistical models well. The third purpose is to investigate whether these indicators can explain US economic recession by using logistic regression. Meanwhile, by using LASSO (the least absolute shrinkage and selection operator), this app can do variable selection for dummy variable representing economic recession. A group of effective factors are extracted which can strongly explain historical economic recession. Based on this group of factors, this app can construct the indicator for further determining whether this indicator can explain economic recession.

"EIA" in the name of the application KAI.APP.EIA represents "Effective Indicator Analysis". Compared with KAI.APP.PhD, this app has some common features. For example, KAI.APP.EIA uses similar empirical techniques used in KAI.APP.PhD. In contrast, there are two major differences in the research purposes. First, KAI.APP.EIA focuses on the selected group of effective factors that can explain historical IDP by using LASSO. Second, this app is able to investigate dynamic change in the group of effective factors for explaining IDP. KAI.APP.EIA can provide empirical results in the chapter in my thesis which is related to investigation of dynamic change in the effective factors. This is why this application is named as "Effective Indicator Analysis" since this app focuses on the effective factors rather than all the macro factors in the different groups in KAI.APP.PhD.

It should be noted that KAI.APP.PhD provides basic empirical results in my first two papers called *Macroeconomic Conditions, Corporate Defaults, and the Adaptive Market Hypothesis:*

[46] The detailed description of how to construct various indicators is shown in the manual called "Kai.R.Tools".

Principal Indicator and Early Warning Signal and *A New Approach to Measuring Economic Recessions*, respectively.

Both applications are not restricted to my current databases. If I am interest in the other countries' data, I am still able to use two apps in practice. Then there is a flexibility of using them.

1.2 Authorized Use Permission

Both KAI.APP.PhD and KAI.APP.EIA are open to the public. Since both applications are wholly owned by Nottingham University and Kai Xing, then they cannot be used or referenced without their express consent. Clicking the following links is to open the applications:

a) http://shiny.maths.nottingham.ac.uk:3838/shiny/KAI.APP.PhD/

b) http://shiny.maths.nottingham.ac.uk:3838/shiny/KAI.APP.EIA/

1.3 Contact

For additional information, user can contact with the developer Kai Xing PhD candidate in Nottingham University Business School (lixkx@nottingham.ac.uk).

2. APPLICATION SUMMARY

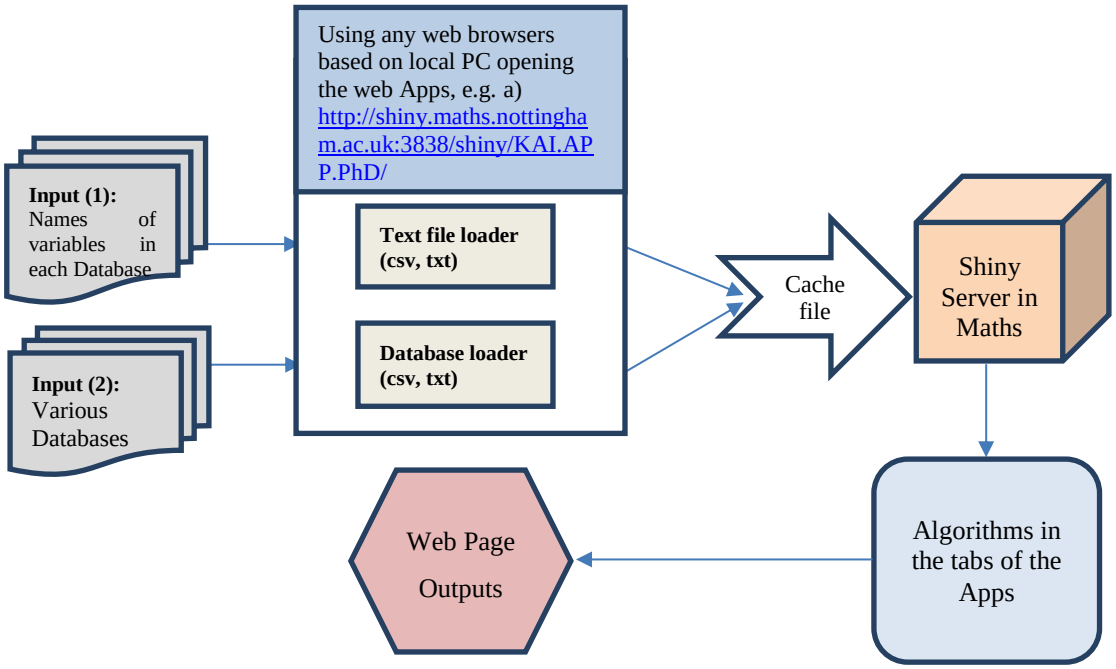
9.4.2 2. APPLICATIONS SUMMARY

2.1 Understanding of Web Apps

The mechanism of the web apps is simple, which is shown in Figure 9-3. Basically, the apps have three parts. First, two types of data are uploaded via text file loader and database loader. Second, these data are sent to Shiny Server in the School of Mathematical Sciences in Nottingham University for analysis in using relevant algorithms. The apps contains different algorithms in order to do relevant analysis. In particular, these algorithms are combined in the specific tabs in the web apps. That is, different tabs can achieve several specific targets. Third, the outputs are shown in the web pages under the corresponding tab.

These web apps can be opened by any web browsers, such as Google Chrome, IE, and Firefox etc. There are two data inputs for using the apps, which are Input (1) and (2). Input (2) is the database containing various time series variables. Input (1) shows the information of these variables in Input (2), for example, it consists of the detailed explanation for each variable. The detailed explanation is illustrated in the next part.

Figure 9-3 Flow chat of web apps



2.2 Structure of the Inputs

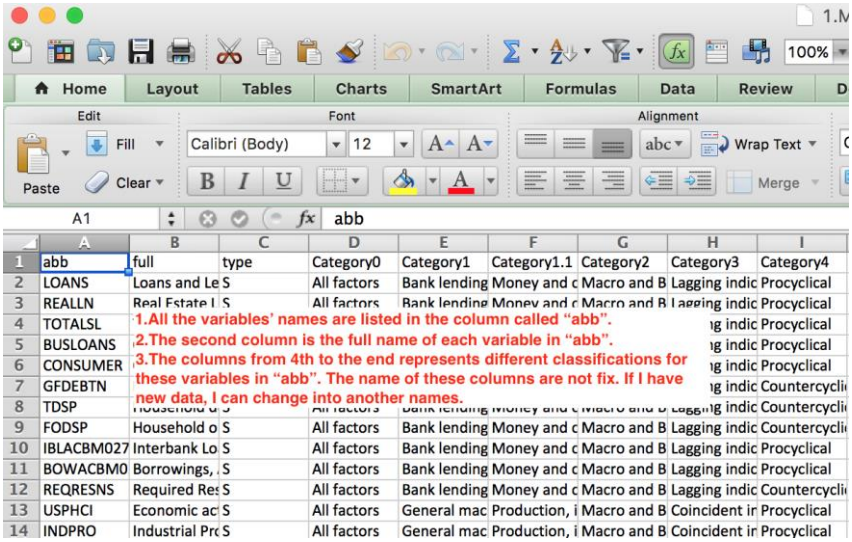
The data for both web apps are the same. In order to enable apps work correctly, there are several requirements for creating the data inputs including Input (1) and (2) in Figure 9-4

shows the screen captures for the two data inputs. In each graph, there are three requirements in red colour, which guide to create the databases.

In my case, Input (1) contains information about hundreds of macroeconomic factors from Federal Economic Database (FRED). Specifically, the abbreviation for each macro factor is recorded in the column named as "abb". Their full names are in the second column called "full". The third one records the type of each variable various classifications.

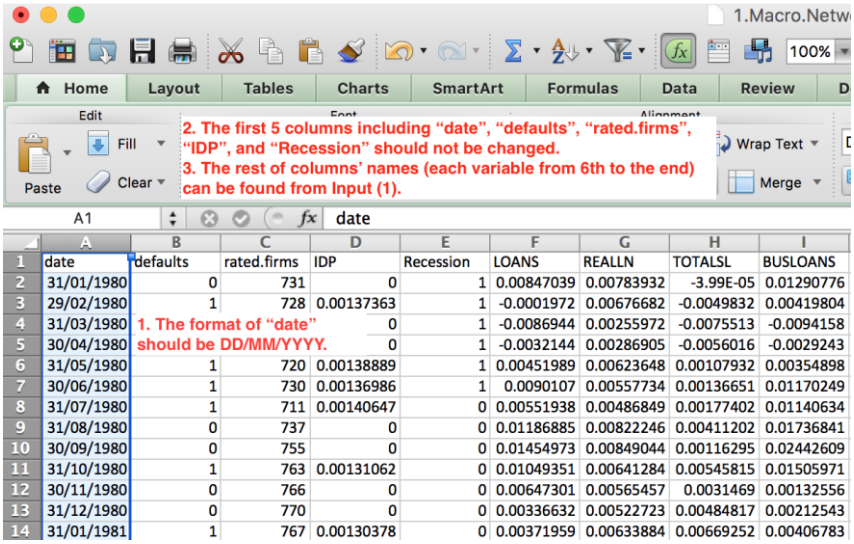
Figure 9-4 Structure of Inputs for the apps

Input (1)



	A	B	C	D	E	F	G	H	I
1	abb	full	type	Category0	Category1	Category1.1	Category2	Category3	Category4
2	LOANS	Loans and Le	S	All factors	Bank lending	Money and c	Macro and B	Lagging indic	Procyclical
3	REALLN	Real Estate L	S	All factors	Bank lending	Money and c	Macro and B	Lagging indic	Procyclical
4	TOTALSL							ıg indic	Procyclical
5	BUSLOANS							ıg indic	Procyclical
6	CONSUMER							ıg indic	Procyclical
7	GFDEBTN							ıg indic	Countercyclic
8	TDSP							ıg indic	Countercyclic
9	FODSP	Household o	S	All factors	Bank lending	Money and c	Macro and B	Lagging indic	Countercyclic
10	IBLACBM027	Interbank Lo	S	All factors	Bank lending	Money and c	Macro and B	Lagging indic	Procyclical
11	BOWACBM0	Borrowings,	S	All factors	Bank lending	Money and c	Macro and B	Lagging indic	Procyclical
12	REQRESNS	Required Res	S	All factors	Bank lending	Money and c	Macro and B	Lagging indic	Countercyclic
13	USPHCI	Economic act	S	All factors	General mac	Production, i	Macro and B	Coincident ir	Procyclical
14	INDPRO	Industrial Pro	S	All factors	General mac	Production, i	Macro and B	Coincident ir	Procyclical

Input (2)



	A	B	C	D	E	F	G	H	I
1	date	defaults	rated.firms	IDP	Recession	LOANS	REALLN	TOTALSL	BUSLOANS
2	31/01/1980	0	731	0	1	0.00847039	0.00783932	-3.99E-05	0.01290776
3	29/02/1980	1	728	0.00137363	1	-0.0001972	0.00676682	-0.0049832	0.00419804
4	31/03/1980				0	1	-0.0086944	0.00255972	-0.0075513	-0.0094158
5	30/04/1980				0	1	-0.0032144	0.00286905	-0.0056016	-0.0029243
6	31/05/1980	1	720	0.00138889	1	0.00451989	0.00623648	0.00107932	0.00354898
7	30/06/1980	1	730	0.00136986	1	0.0090107	0.00557734	0.00136651	0.01170249
8	31/07/1980	1	711	0.00140647	0	0.00551938	0.00486849	0.00177402	0.01140634
9	31/08/1980	0	737	0	0	0.01186885	0.00822246	0.00411202	0.01736841
10	30/09/1980	0	755	0	0	0.01454973	0.00849044	0.00116295	0.02442609
11	31/10/1980	1	763	0.00131062	0	0.01049351	0.00641284	0.00545815	0.01505971
12	30/11/1980	0	766	0	0	0.00647301	0.00565457	0.0031469	0.00132556
13	31/12/1980	0	770	0	0	0.00336632	0.00522723	0.00484817	0.00212543
14	31/01/1981	1	767	0.00130378	0	0.00371959	0.00633884	0.00669252	0.00406783

3. GETTING STARTED

9.4.3 3. GETTING STARTED

3.1 KAI.APP.PhD

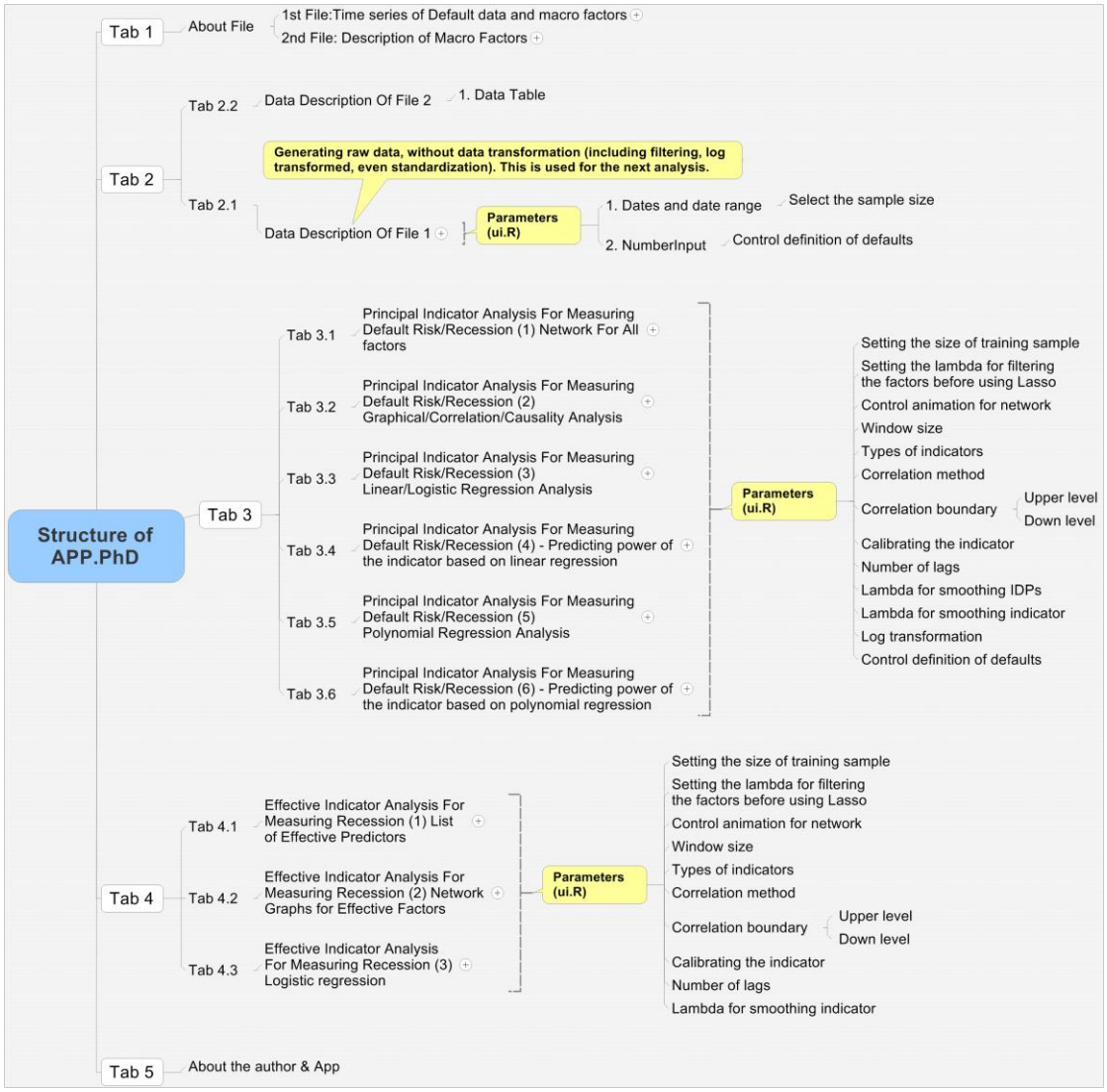
3.1.1 Structure of KAI.APP.PhD

Figure 9-5 shows the structure of the KAI.APP.PhD. There are five tabs which have specific tasks in this app. Tab 1 is to show the basic information of uploaded data, such as file name, size, and type. Tab 2 is to illustrate the data description for each uploaded data. Based on the data from Tab 2, this app do all the empirical analysis in Tab 3 or Tab 4. There is no connection between Tab 3 and Tab 4 since they focuses on different topics.

The yellow box showing “Parameters (ui.R)” indicates the relevant settings in the app. These settings correspond to the widgets that have specific instructions to do their jobs. That is, the widgets are used to control and adjust the empirical analysis in the corresponding tab. Specifically, if you change the starting and ending date in the app, then the sample size used for analysis is changed. Also the number of minimum defaults can be controlled in order to removing certain months with small default events. In the following part, this manual will detail how to use this app under different tabs.

Principal indicator analysis is based on the factors without using LASSO. In contrast, effective indicator analysis is based on the effective factors extracted by LASSO.

Figure 9-5 Structure of the KAI.APP.PhD



3.1.2 Overview the KAI.APP.PhD

Figure 9-6 shows the appearance of KAI.APP.PhD. There are three red boxes, which represent different areas. Each area has specific purpose which are described in black font.

Figure 9-6 Appearance of the KAI.APP.PhD

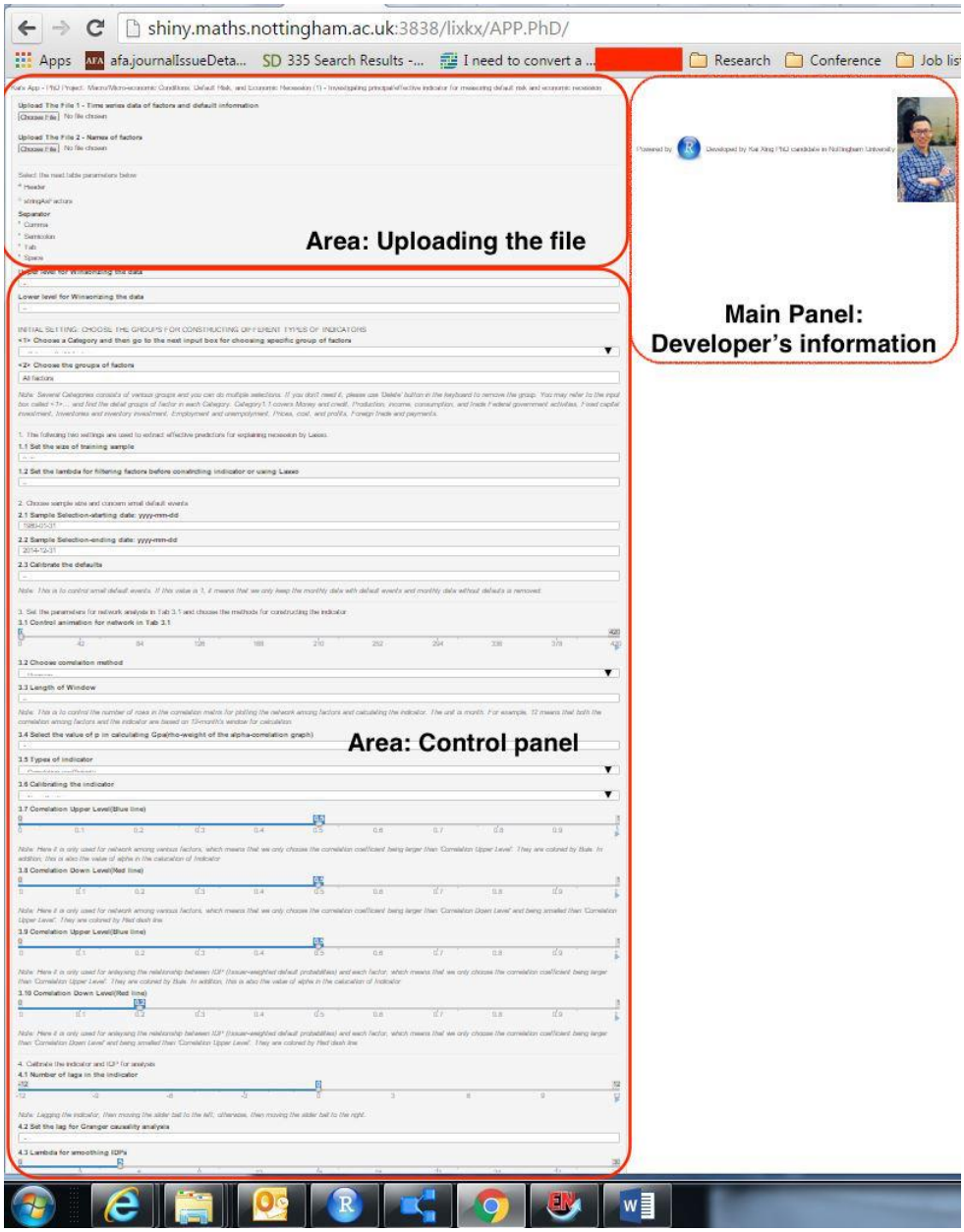


Figure 9-7 describes the part called "Area: Uploading the file" in Figure 9-6.

After uploading the two inputs including Input (1) called MarcoName.csv and Input (2) called 1.Macro.Network Analysis.industrial.csv, you can see the inside of the app in Figure 9-8.

Figure 9-7 Detail description of "Area: Uploading the file" in Figure 9-6

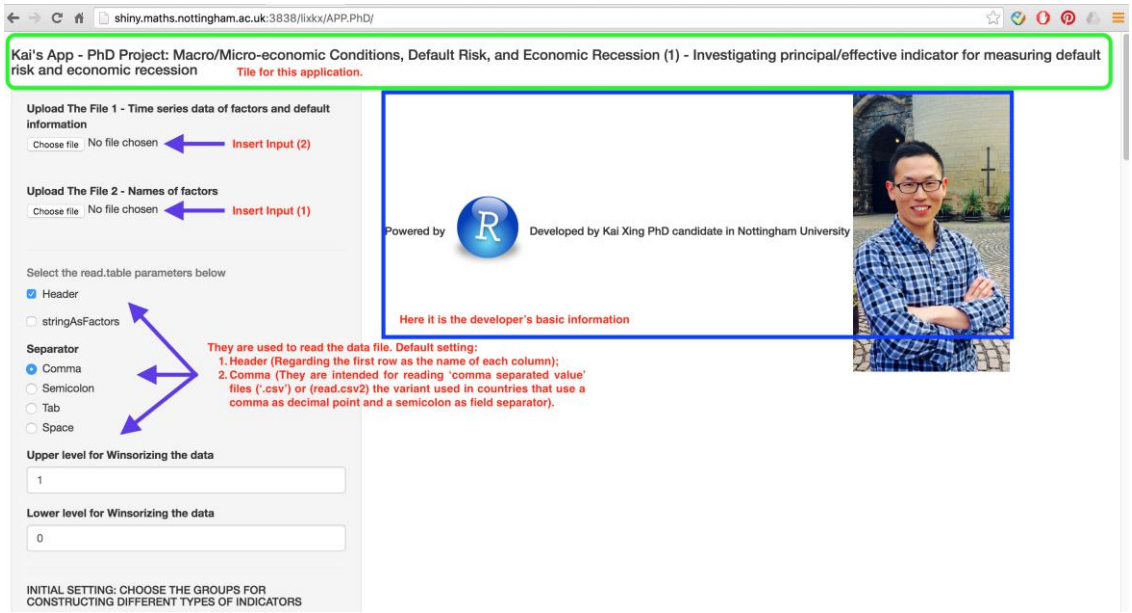
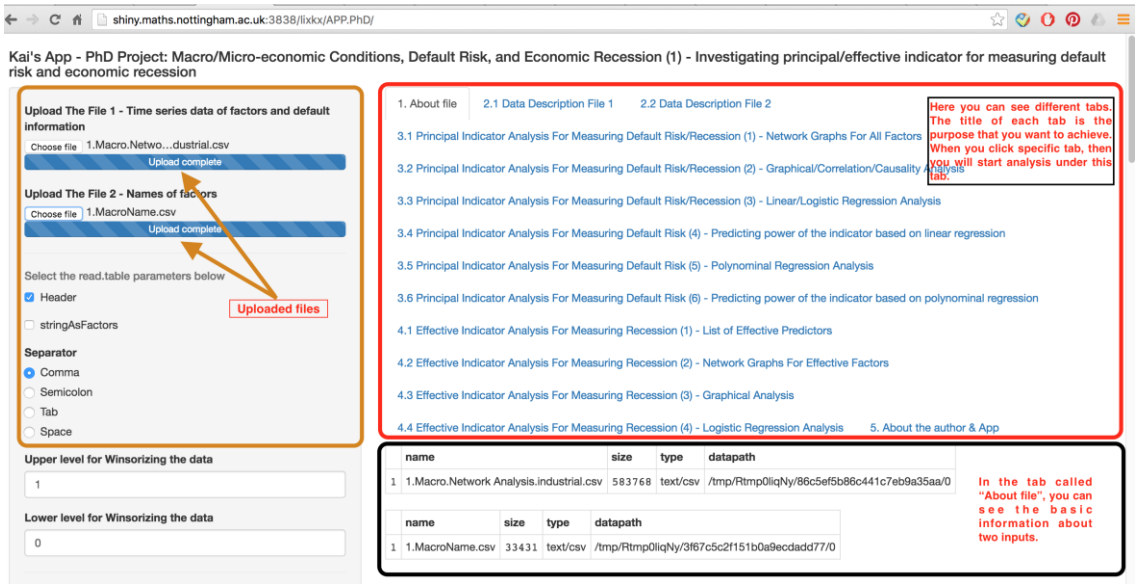


Figure 9-8 Uploading the file and reading the file



3.1.3 Tab 1 and Tab 2

Figure 9-9 illustrates the structure of Tab 1 and Tab 2. After selecting the groups of factors and setting the starting date and ending date, you can see the data table, summary of each variable in the data, plot defaults/exposures/IDP, and the breakpoints in the time series of IDP.

Figure 9-10 provides an example after setting the starting and ending dates “1980-01-31” and “2014-12-31”, group of factors “Category0:All factors” and “All factors”, and number of minimum defaults “0”. The approach of choosing the factors is shown in Figure 9-11.

Notes: If you want to identify the breakpoints, you should use “4.3 Lambda for smoothing IDPs” to reduce the zero in the time series of IDP; otherwise, you cannot get the breakpoints in IDP.

Figure 9-9 Structure of Tab 1 and Tab 2

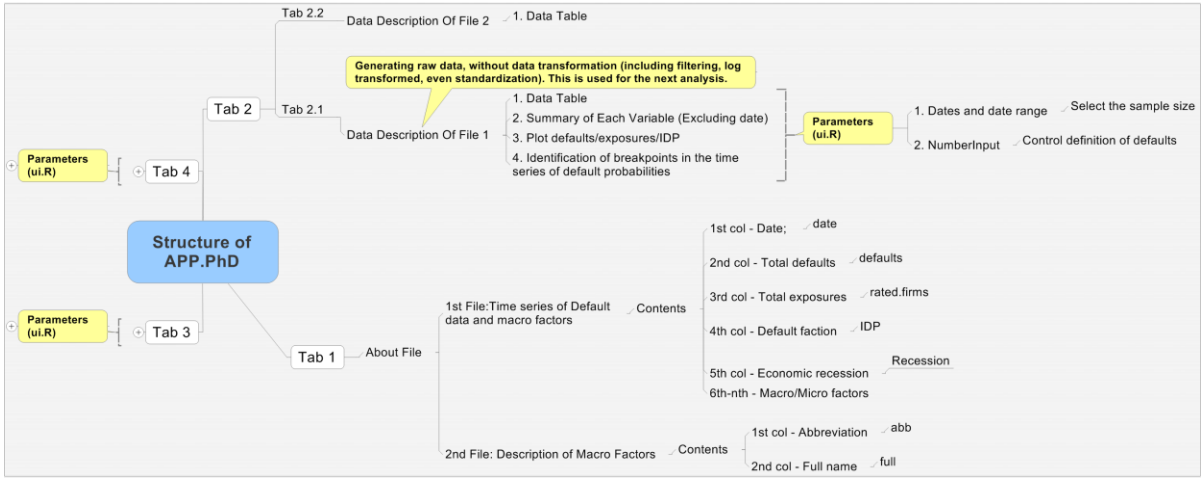


Figure 9-10 Appearance and outputs of Tab 2.1

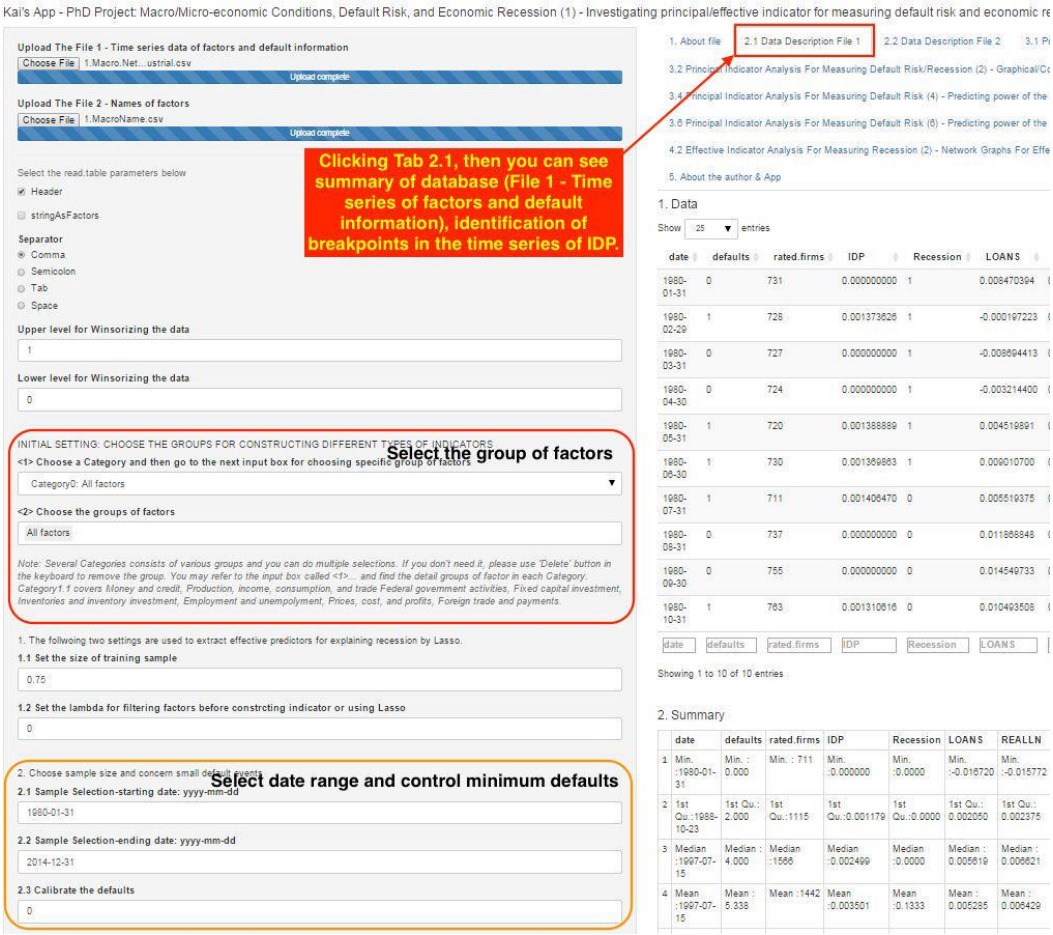
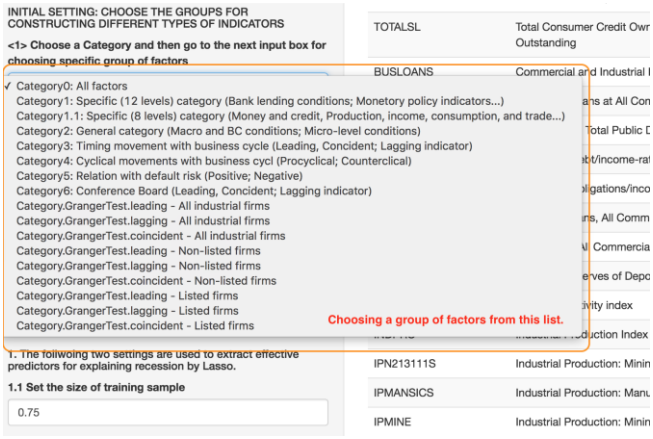
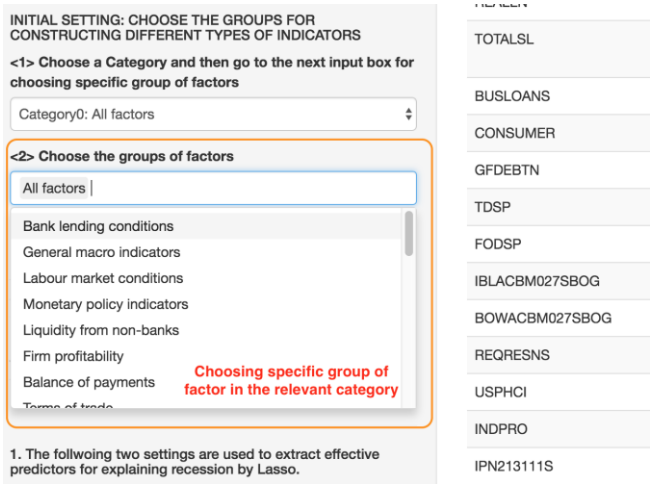


Figure 9-11 Selection of the group of factors

First selection of category



Second selection of factors in the category



3.1.4 Tab 3.1 – 3.6

Figure 9-12 presents the structure of Tab 3 in this app. Tab 3 covers 6 sections and different sections have different purposes by employing relevant techniques. The main issue of doing empirical analysis in Tab 3 is to know how to use the app to implement network analysis and investigate the relationship between indicators and IDP or relationship between indicators and economic recessions.

The implementation of network analysis is in Tab 3.1. Figure 9-13 shows how to use app to do network analysis. In Panel A, there are orange, red, and green boxes. Each box shows the description of how to use the slider bar, selective input, and numerical input in order to execute the network analysis based on group of macro factors. Panel B shows the outputs based on the settings in Panel A.

For the rest of tabs from Tab 3.2 to Tab 3.6, we need to use two sets of control widgets to send the messages to Shiny in order to achieve our purposes. They are shown in Figure 9-14. Panel A exhibits six widgets which are used to construct various indicators. Panel B indicates how to calibrate the indicator and IDP. By using different settings in these widgets, we can get different indicators and different statistical inference test results in the six tabs. When clicking the specific tab, the algorithms hidden in this tab will work based on the information sent from the relevant widgets. Then you can see the outputs under the tab.

Note: there are some notes below several widgets. Please read them carefully before changing the settings!

Figure 9-12 Structure of Tab 3

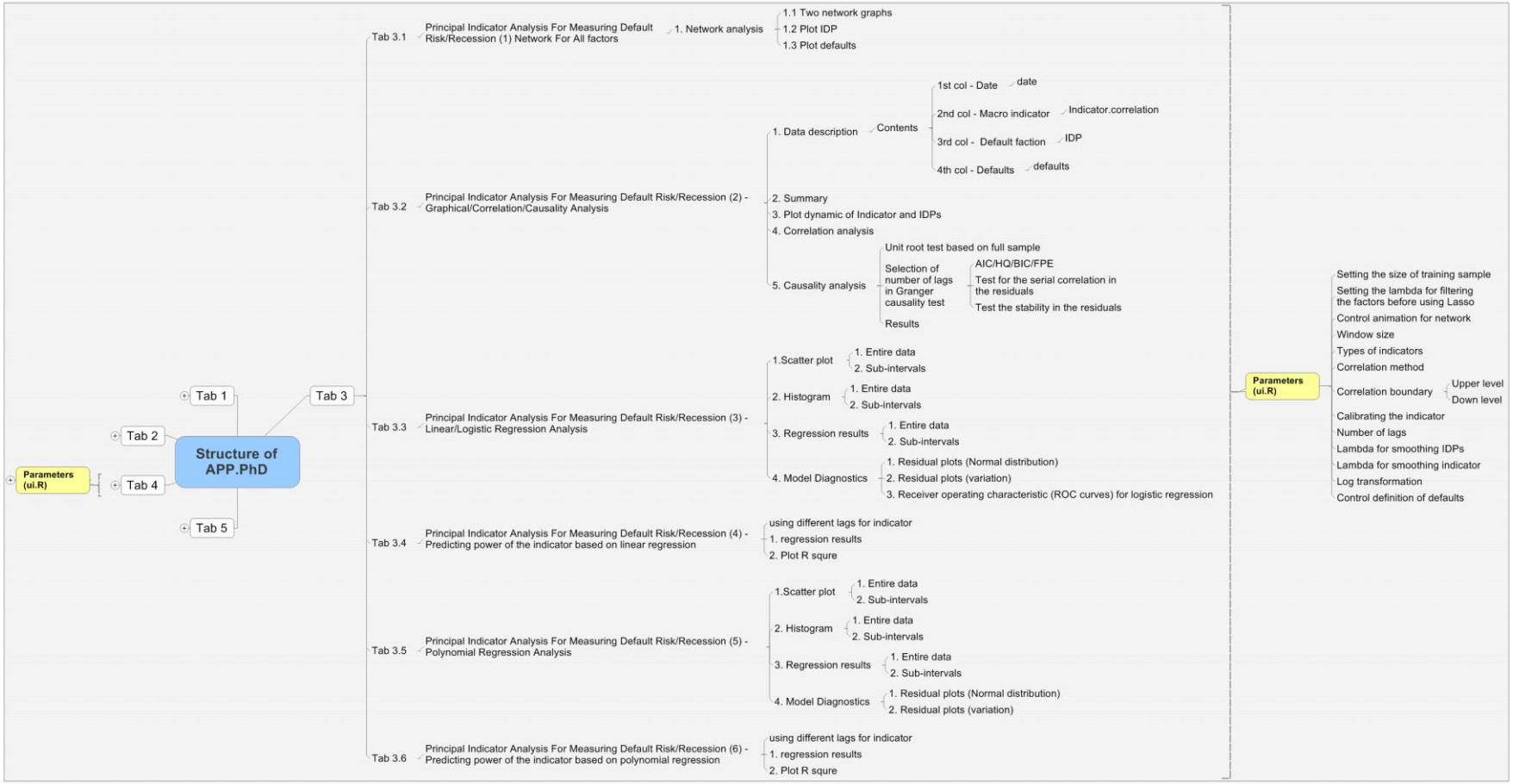
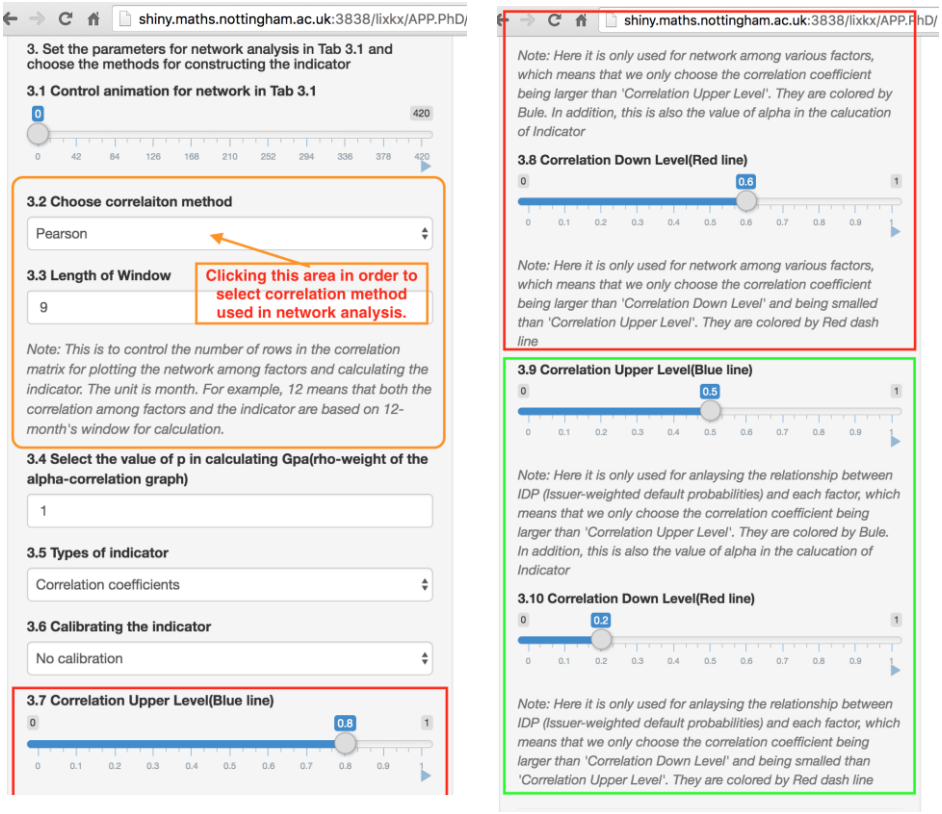


Figure 9-13 Appearance and outputs of Tab 3.1

Panel A



Panel B

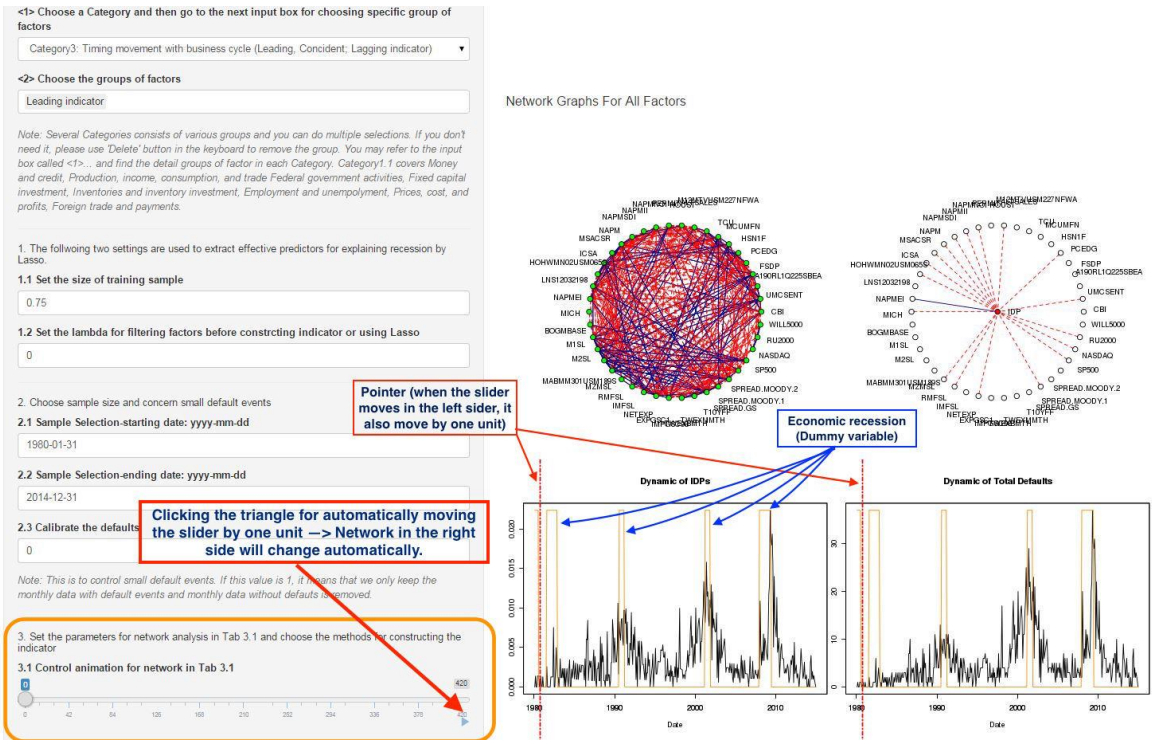
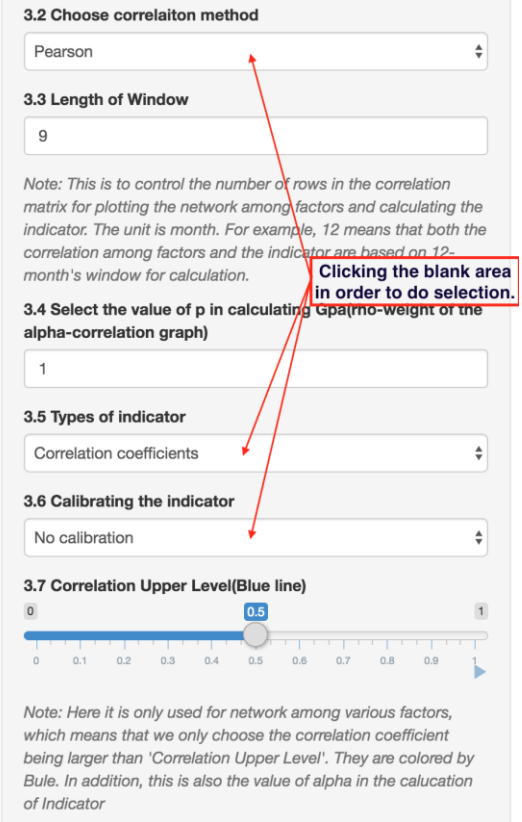
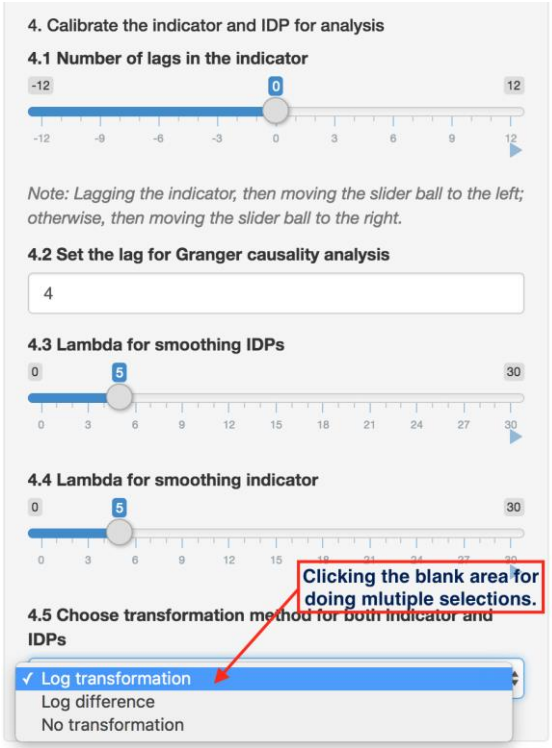


Figure 9-14 Construction of the indicator and adjustment of both indicator and IDP

Panel A

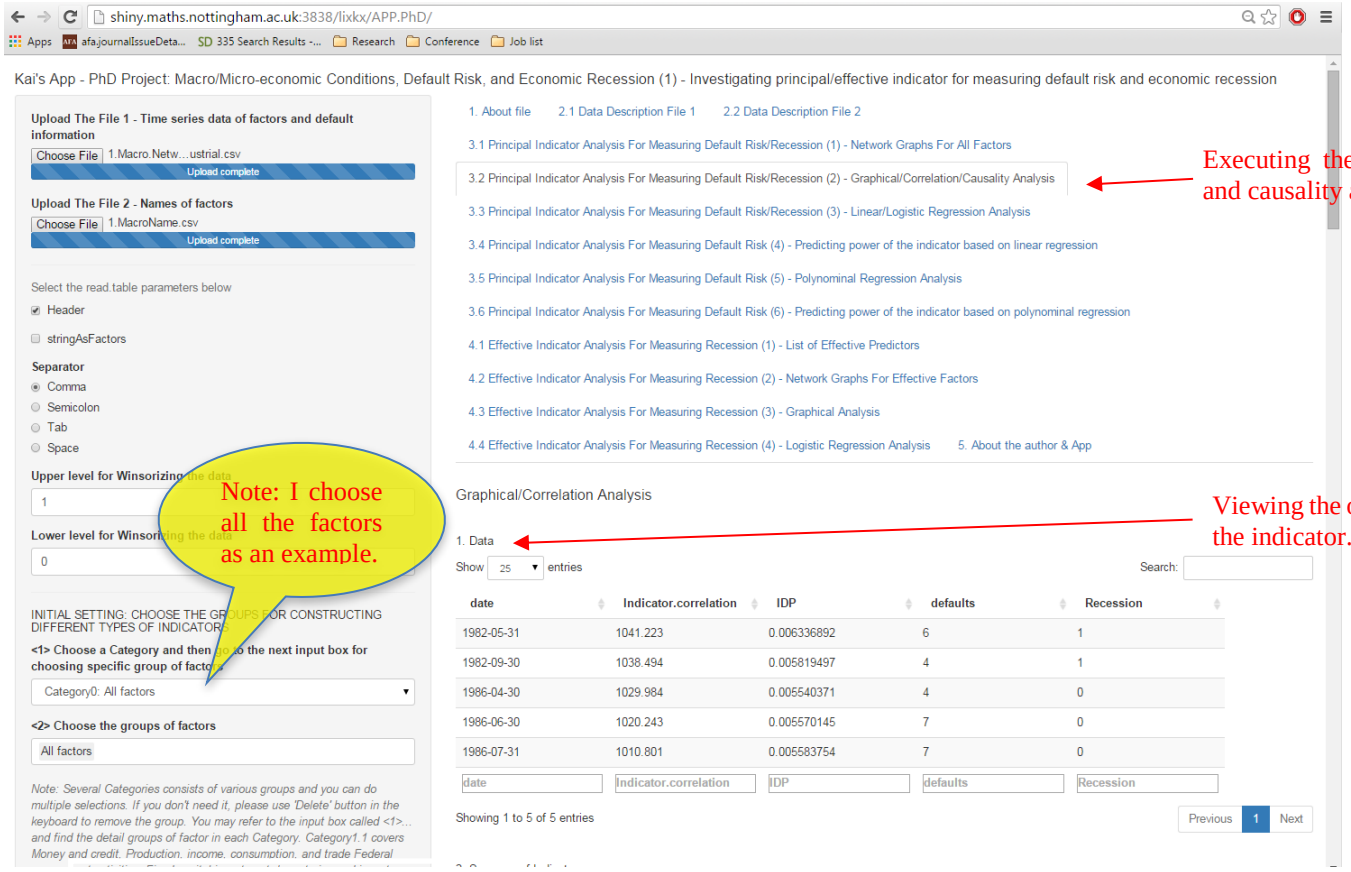


Panel B



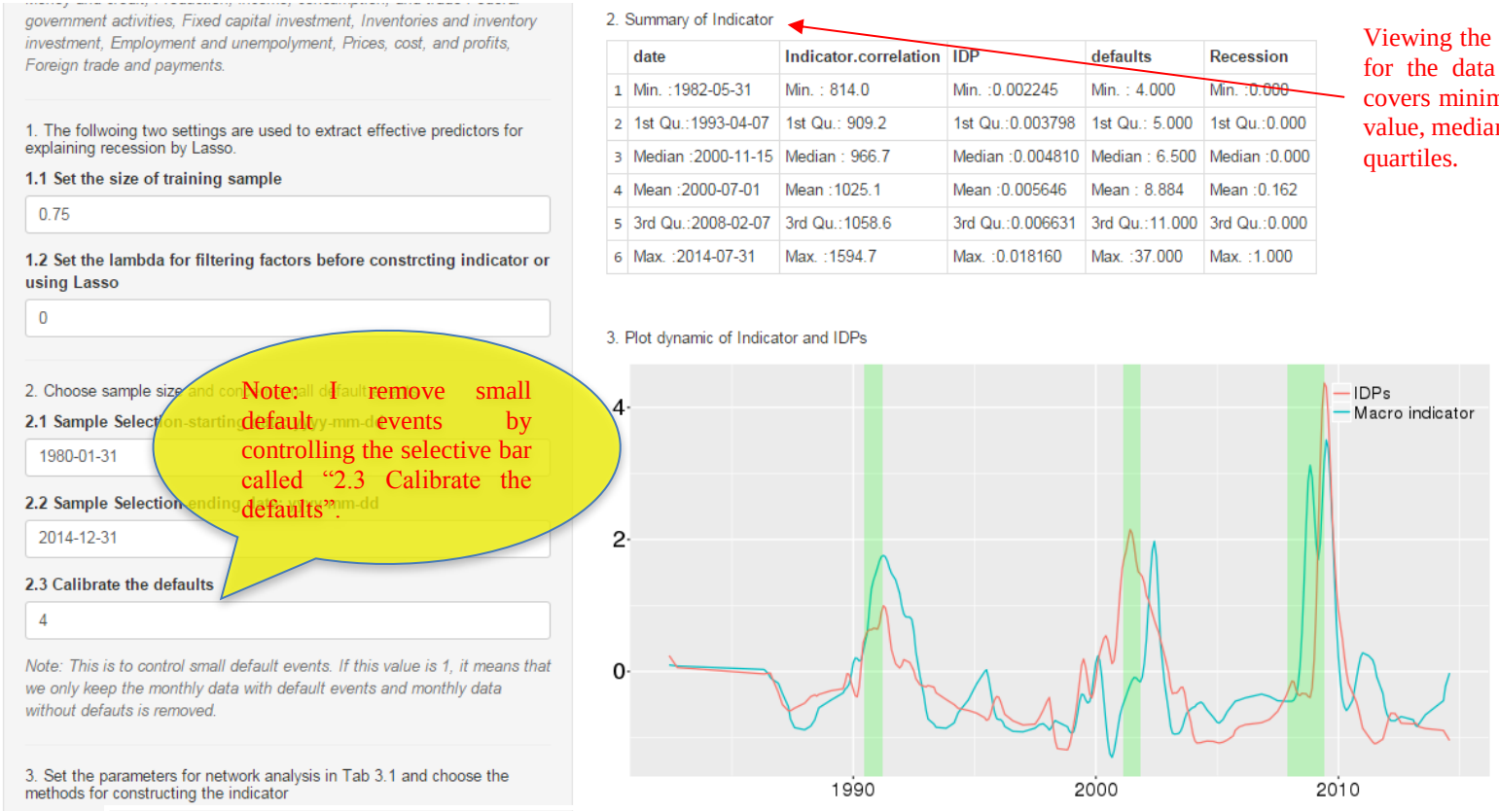
After completing the relevant settings for investigation between IDP and indicator in Figure 9-14, this manual is to show how I work with this app in order to analyse the data. The detail procedure is shown in the following figures.

Figure 9-15 Work with Tab 3.2 (1)



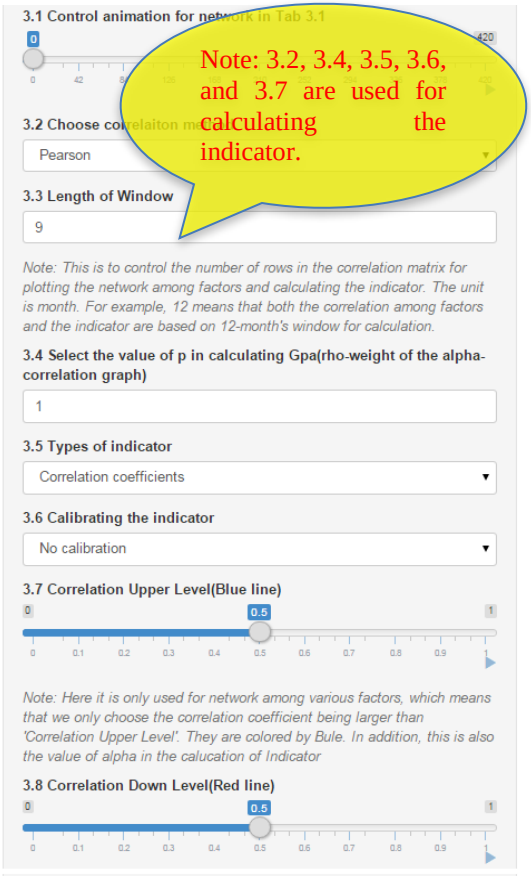
Continued

Figure 9-16 Work with Tab 3.2 (2)



Continued

Figure 9-17 Work with Tab 3.2 (3)



Note: 3.2, 3.4, 3.5, 3.6, and 3.7 are used for calculating the indicator.

4. Correlation between Indicator and IDP in each interval.

4.1 Split the intervals based on breakpoints

	Interval	Obs	Pearson	Spearman
b0	1982.05-2014.07	216	0.558***	0.499***
b1	1982.05-1982.09	2	N.A.	N.A.
b2	1986.04-1993.05	54	0.828***	0.829***
b3	1993.10-1998.11	29	0.299	0.328
b4	1998.12-2004.02	60	0.396**	0.491***
b5	2004.04-2014.07	71	0.744***	0.598***

4.2 Split the intervals based on 10 years' interval

	Interval	Obs	Pearson	Spearman
b0	Indicator.1986.04-2014.07	214	0.569***	0.513***
b1	Indicator.1986.04-1994.09	58	0.892***	0.881***
b2	Indicator.1995.01-2004.12	90	0.435***	0.432***
b3	Indicator.2005.01-2014.07	66	0.719***	0.527***

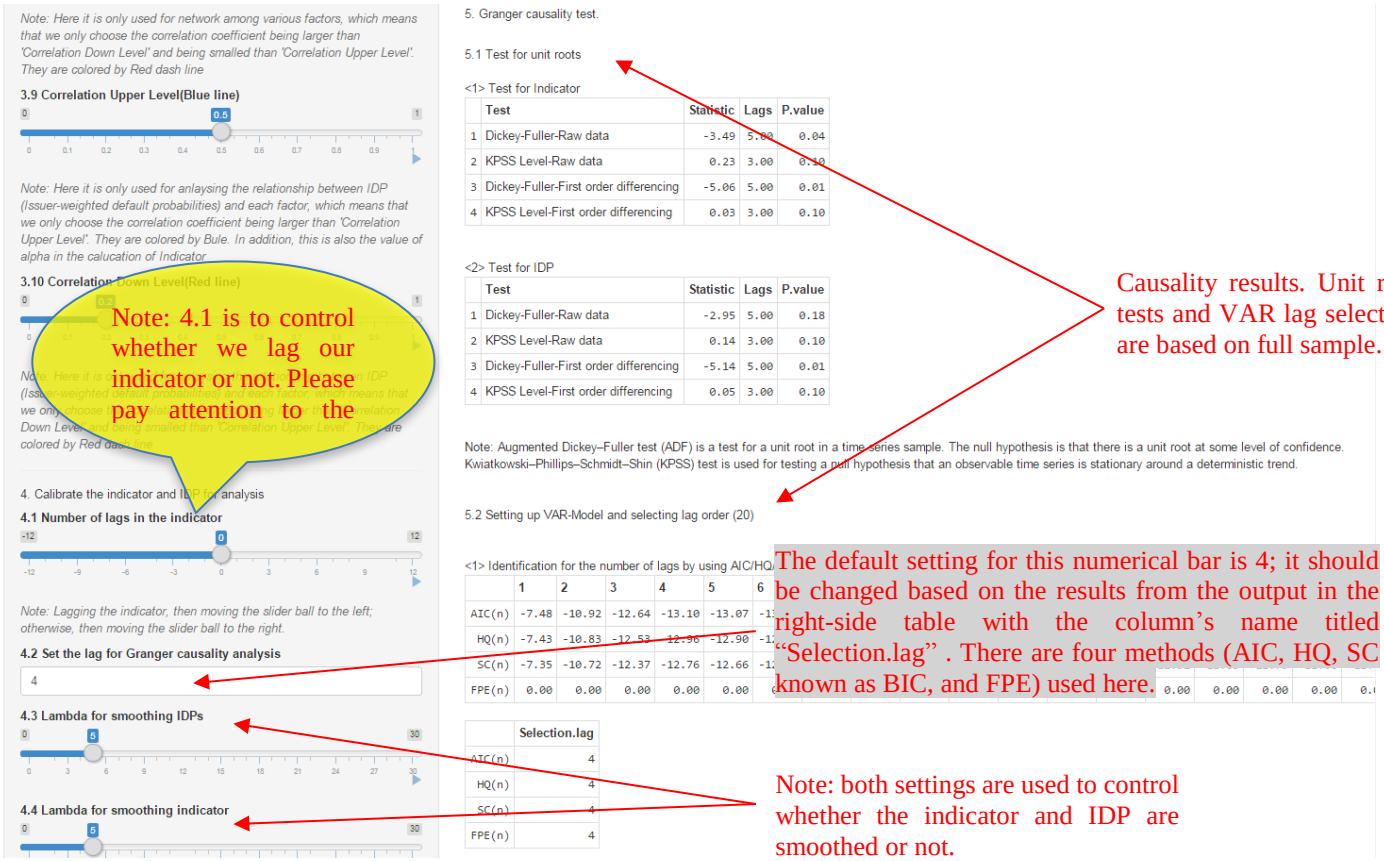
4.3 Split the intervals based on whether the interval covers the recession

	Interval	Obs	Pearson	Spearman
b0	Indicator.1982.05-2014.07	216	0.569***	0.517***
b1	Indicator.1982.05-1982.09	2	N.A.	N.A.
b2	Indicator.1986.04-1990.05	26	0.673***	0.616***
b3	Indicator.1990.07-1991.03	9	0.657	0.917***
b4	Indicator.1991.04-2001.02	75	0.398***	0.391***
b5	Indicator.2001.03-2001.11	9	-0.357	-0.383
b6	Indicator.2001.12-2007.10	40	0.681***	0.416**
b7	Indicator.2008.01-2009.06	18	0.483*	0.253
b8	Indicator.2009.07-2014.07	37	0.725***	0.243

Correlation results based on different samples. For the first one (4.1), the breakpoints are identified in the 4th part called "4. Identification of breakpoints in the time series of default probabilities" in the Tab 2.1.

Continued

Figure 9-18 Work with Tab 3.2 (4)



Continued

Figure 9-19 Work with Tab 3.2 (5)

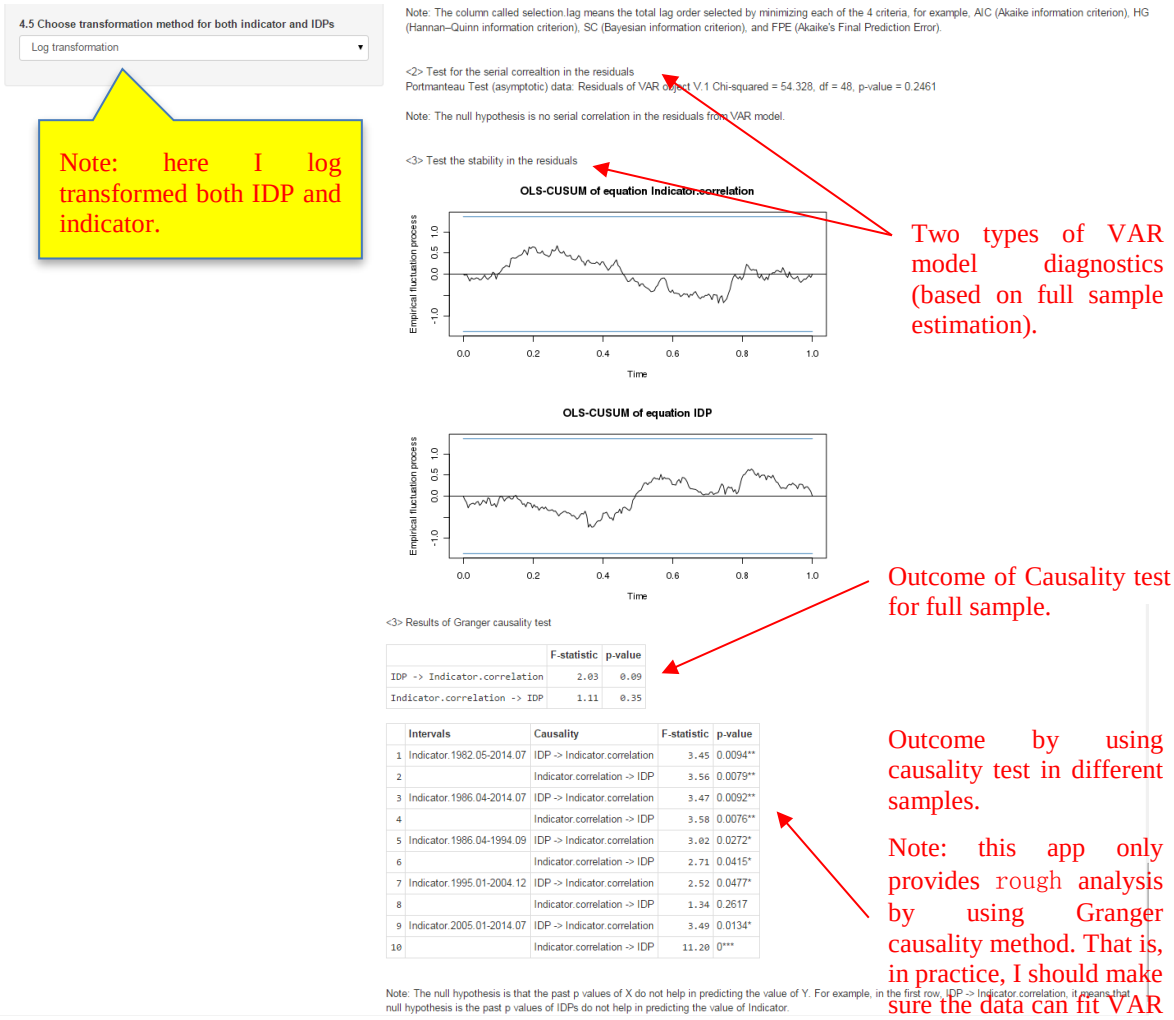
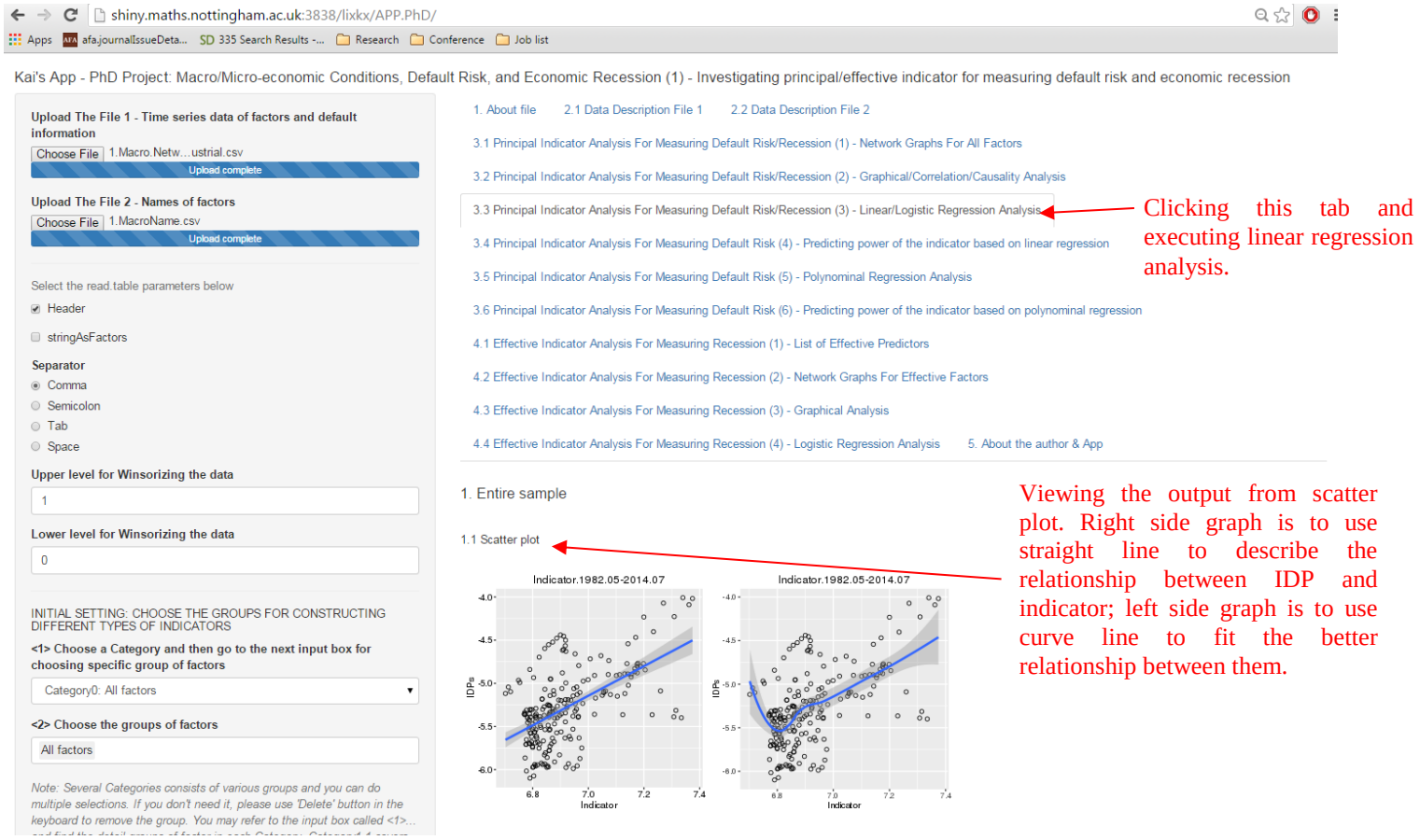
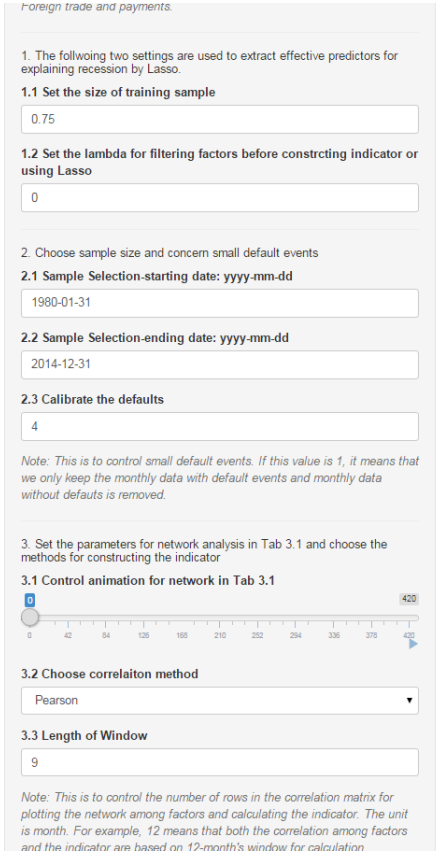


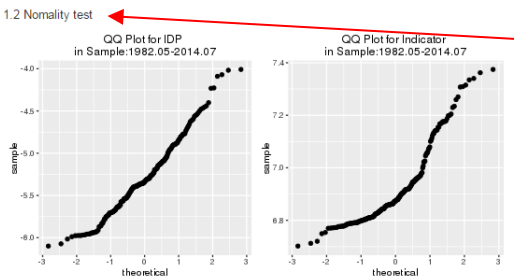
Figure 9-20 Work with Tab 3.3 (1)



Continued

Figure 9-21 Work with Tab 3.3 (2)





Normality tests (QQ-plot) for the indicator and IDP.

Viewing the output of three linear regression results.

Viewing the output of logistic regression results (Dependent variable: Dummy variable, recession; Independent variable, indicator).

1.3 Linear/Logistic regression results

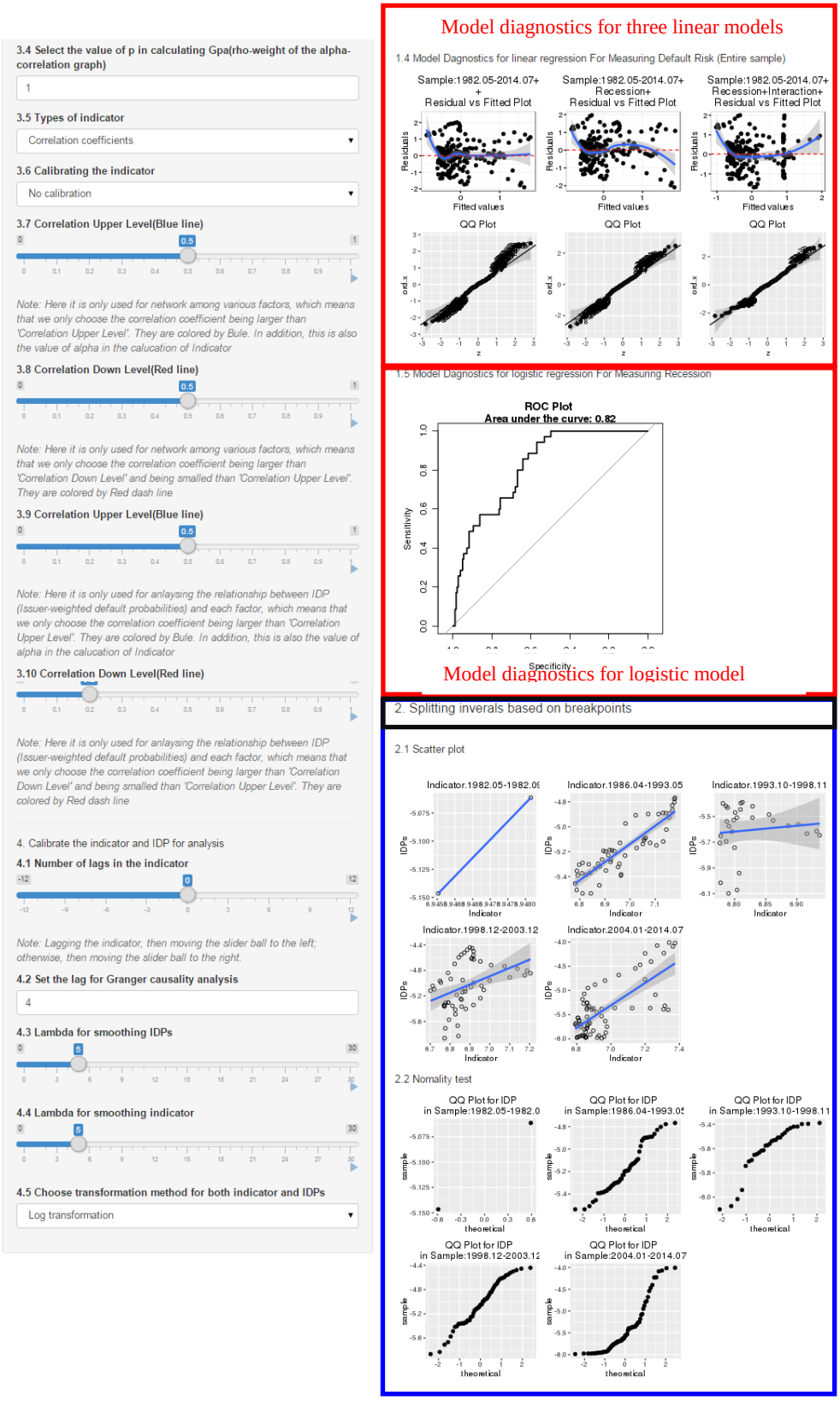
Regression Results

	Dependent variable:			
	IDPs			Recession (=1)
	OLS			*logistic*
	(1)	(2)	(3)	(4)
Indicator.1982.05-2014.07	0.569***	0.486***	0.638***	
	(0.056)	(0.061)	(0.067)	
Recession		0.507***	0.952***	
		(0.166)	(0.186)	
Indicator.1982.05-2014.07*Recession			-0.627***	
			(0.137)	
Indicator.1982.05-2014.07				7.190***
				(1.288)
Constant	0.000	-0.082	-0.053	-51.743***
	(0.056)	(0.061)	(0.059)	(9.025)
Observations	216	216	216	216
R^2	0.323	0.352	0.410	
Adjusted R^2	0.320	0.346	0.401	
Log Likelihood				-77.641
Akaike Inf. Crit.				159.282
Residual Std. Error	0.825 (df = 214)	0.809 (df = 213)	0.774 (df = 212)	
F Statistic	02.264*** (df = 1; 214)	57.775*** (df = 2; 213)	49.074*** (df = 3; 21	
Note:	p<0.1; p<0.05; p<0.01			

Standard errors are reported in parentheses underneath the coefficients.

Continued

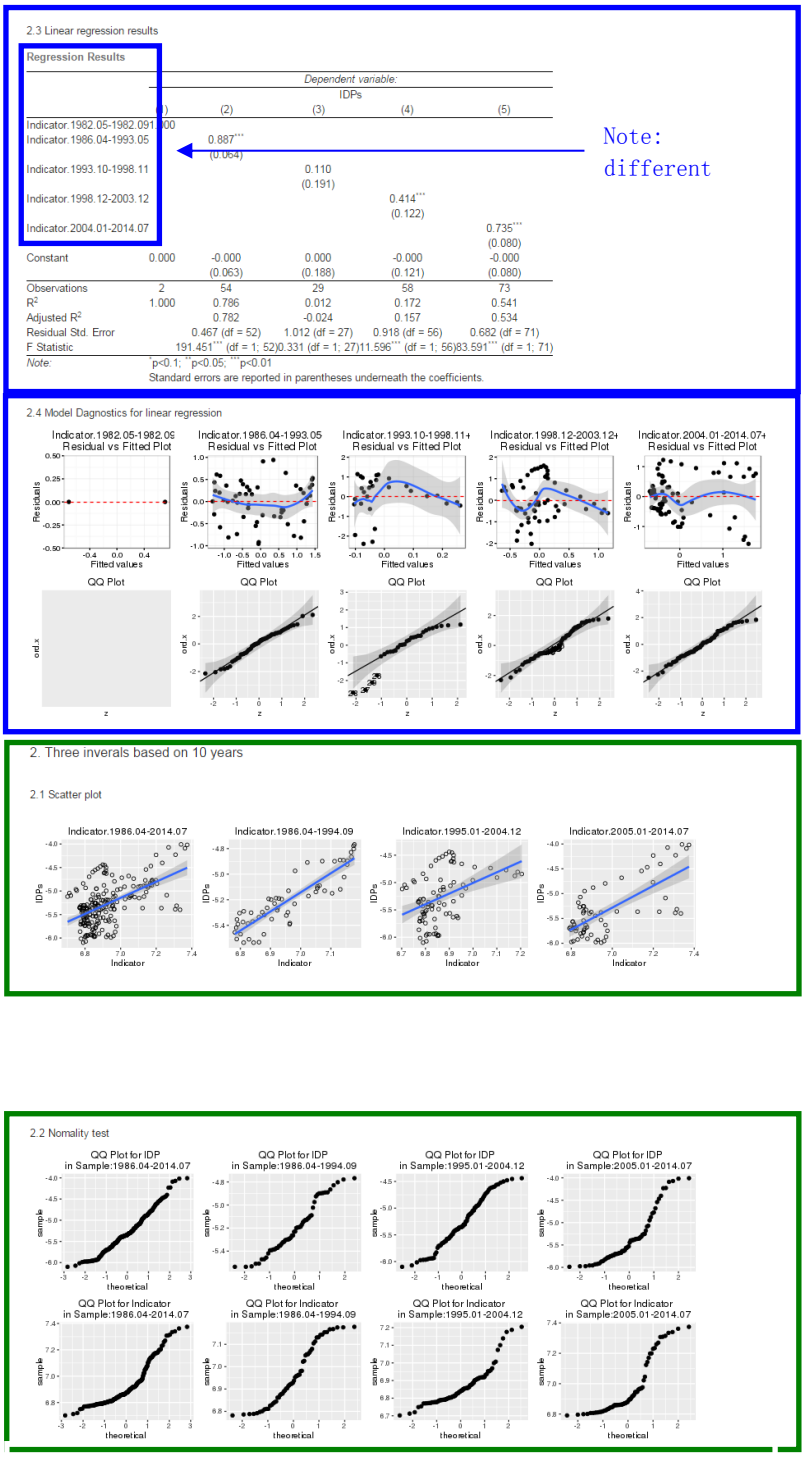
Figure 9-22 Work with Tab 3.3 (3)



Continued

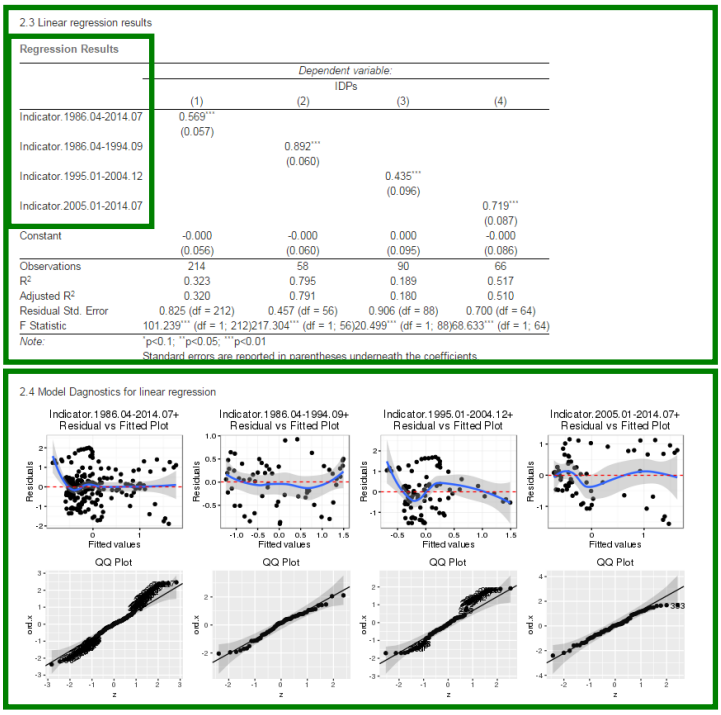
Note: A set of empirical analysis tools except logistic regression in red box in Figure 9-21 and Figure 9-22 are also used in blue box and green box in Figure 9-22, Figure 9-23, and Figure 9-24.

Figure 9-23 Work with Tab 3.3 (4)



Continued

Figure 9-24 Work with Tab 3.3 (5)



2.3 Linear regression results

Regression Results

	Dependent variable:			
	IDPs			
	(1)	(2)	(3)	(4)
Indicator.1986.04-2014.07	0.569***			
	(0.057)			
Indicator.1986.04-1994.09		0.892***		
		(0.060)		
Indicator.1995.01-2004.12			0.435***	
			(0.096)	
Indicator.2005.01-2014.07				0.719***
				(0.087)
Constant	-0.000	-0.000	0.000	-0.000
	(0.056)	(0.060)	(0.095)	(0.086)
Observations	214	58	90	66
R^2	0.323	0.795	0.189	0.517
Adjusted R^2	0.320	0.791	0.180	0.510
Residual Std. Error	0.825 (df = 212)	0.457 (df = 56)	0.906 (df = 88)	0.700 (df = 64)
F Statistic	101.239*** (df = 1; 212)	217.304*** (df = 1; 56)	20.499*** (df = 1; 88)	68.633*** (df = 1; 64)
Note:	*p<0.1; **p<0.05; ***p<0.01			
	Standard errors are reported in parentheses underneath the coefficients.			

2.4 Model Dagnostics for linear regression

3.1.5 Tab 4.1 – 4.3

Figure 9-25 provides the structure of Tab 4 in this app. As mentioned early, this tab is to extract the effective factors and then use them to construct the indicators for analysing whether the indicator can measure the economic recession. This tab need two more widgets for executing LASSO in the app, which are shown in Figure 9-26.

Figure 9-25 Structure of Tab 4

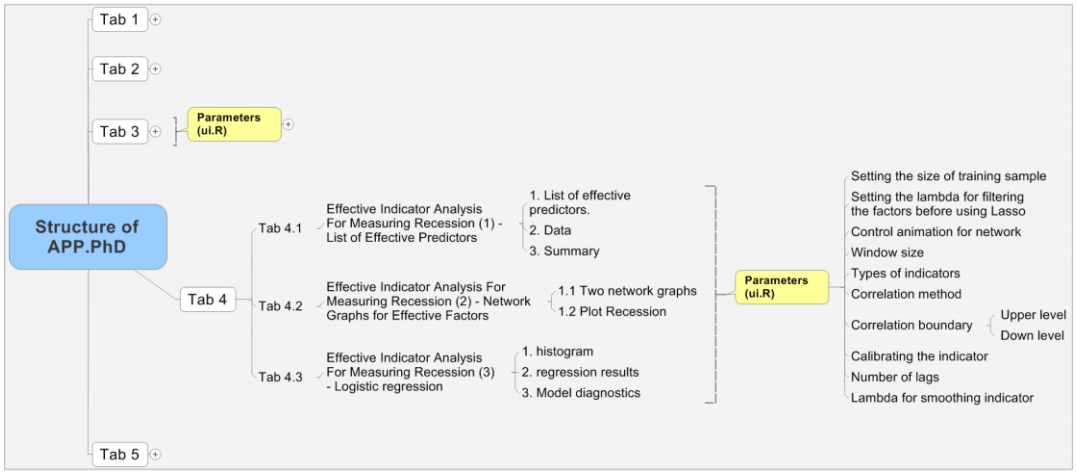
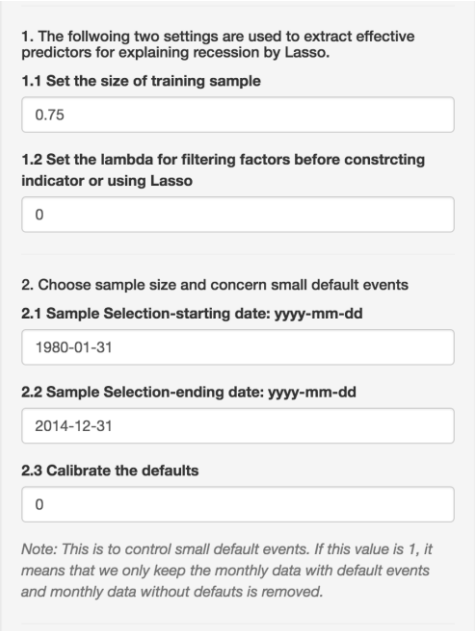


Figure 9-26 Implementation of LASSO

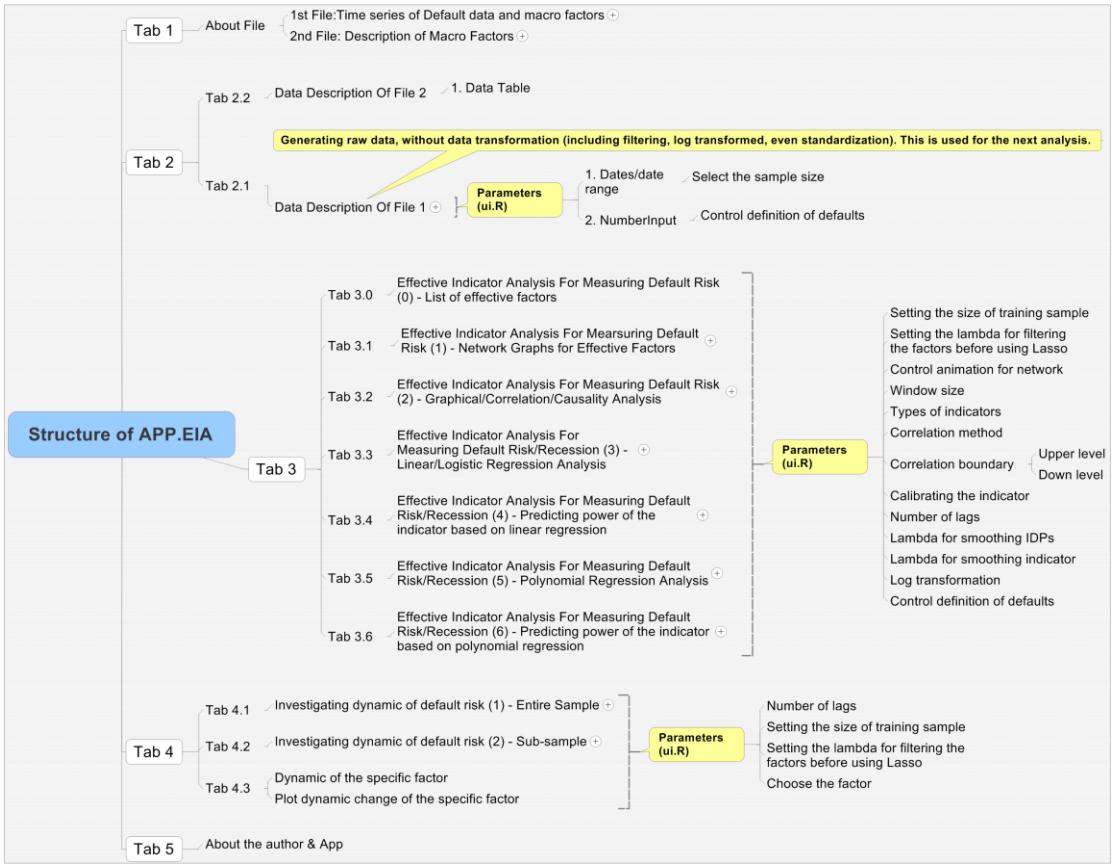


3.2 KAI.APP.EIA

3.2.1 Structure of KAI.APP.EIA

The application KAI.APP.EIA also has four tabs and each tab is to achieve specific purposes in Figure 9-27. Tab 1 and Tab 2 are used to show the basic information of uploaded databases, summary statistics of the variables, and IDP, which is the same as the functions of Tab 1 and Tab 2 in the KAI.APP.PhD. Likewise, the major functions in Tab 3 in KAI.APP.EIA are also similar as that in KAI.APP.PhD. The only difference between them is that KAI.APP.EIA has an extra procedure of extracting effective factors from macro factors database before doing same empirical analysis in Tab 3 in KAI.APP.PhD. The information followed by the yellow box with "Parameters" insider indicates which widgets in the sider panel in the app can control and adjust data analysis. There is no connection between Tab 3 and Tab 4. Both tabs use the data from Tab 2. Tab 4 is the new section compared with KAI.APP.PhD. There are 3 sub-parts in Tab 4, which are used to investigate the dynamic change in the effective predictors for IDP. Tab 5 provides the basic information for the this app and the developer' contact.

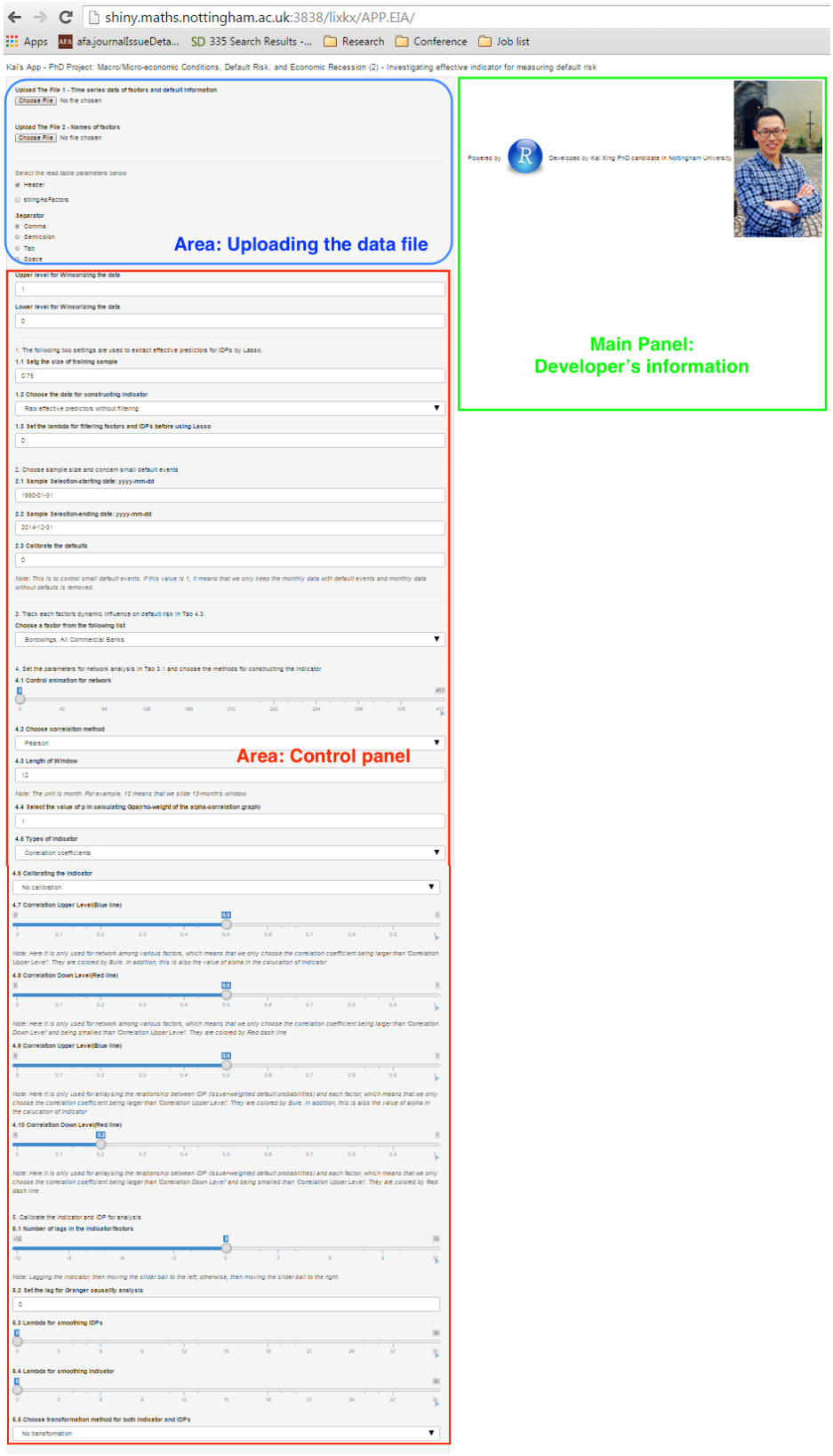
Figure 9-27 Structure of the KAI.APP.EIA



3.2.2 Overview the KAI.APP.EIA

Figure 9-28 shows the appearance of KAI.APP.EIA. There are three boxes with different colours. Each area has specific purpose, which is described by using the corresponding colour font. Basically, the layout of KAI.APP.EIA is similar as that of KAI.APP.PhD.

Figure 9-28 Appearance of the KAI.APP.EIA



3.2.3 Tab 3.0 – 3.6

Since both KAI.APP.PhD and KAI.APP.EIA have same features in terms of using network analysis, correlation analysis, causality analysis, and statistical inference analysis, then the widgets in both apps are almost same for doing these analysis. Compared with KAI.APP.PhD, KAI.APP.EIA needs to use extra three widgets for doing variable selection by LASSO. Figure 9-29 shows an example of how to use these widgets get effective factors. The effective factors are shown in the area called "List of Effective Factors".

Note: In the widget marked as "1.2 Choose the data for constructing indicator", there are two choices including "Filtered effective predictors" and "Raw effective predictors without filtering". The former one means that this app uses Hodrick-Prescott (HP) filtering method to remove the cyclical component from the time series of each variable and then it uses LASSO to get effective factors. The latter one means that this app directly uses the raw data to select the effective factors without using filtering method. The numerical input bar called "Set the lambda for filtering factors and IDPs before using Lasso" can only work when "Filtered effective predictors" is selected.

After determining the effective factors, this app uses the same empirical techniques in Tab 3 in KAI.APP.PhD to investigate two main issues:

1) Whether there is relationship between the correlation among the effective factors and economic recession/IDP;
2) Whether there is relationship between macro indicators constructed by effective factors and IDP;
3) How these macro indicators explain IDP.

Since the Tab 3.1 – 3.6 KAI.APP.EIA follows same structure, design, and mechanism as that in KAI.APP.PhD. Then this manual does not provide details in this case.

3.2.4 Tab 4.1 – 4.3

These three tabs are new features in comparison with KAI.APP.PhD. They are used to investigate the dynamic of effective factors that can explain IDP. The structure of Tab 4.1 – 4.3 is shown in Figure 9-30. It can be seen that the analysis is separated into three parts:

1) Investigating the dynamic change in effective factors in the full sample in Figure 9-31, Figure 9-32, and Figure 9-33. Two approaches are used to investigate the dynamic of effective factors, which is shown in Figure 9-32 and Figure 9-33.
2) Studying the dynamic change in effective factors in 8 different samples. The method of splitting full sample into 8 parts is based on whether the sample covers recession period or non-recession period. Figure 9-30 provides the descriptions of 8 samples in Tab 4.2 in Tab 4. Figure 9-34 provides an example.
3) Exploring the dynamic in the specific factor or finding how the specific factor explains IDP in the full and 8 sub-samples. Figure 9-35 shows an example.

Note: The time span for extracting effective factors is based on 5 years which covers 60 monthly observations. The reason of using 5 years is that the minimum number of observations is 60. Empirically, I find that more sample sizes used for selection of effective factors do not impact the results. Theoretically, the better way to explore how the effective factors dynamically change over time is to use smaller sample size rather than bigger one.

Figure 9-29 Using LASSO to extract effective factors

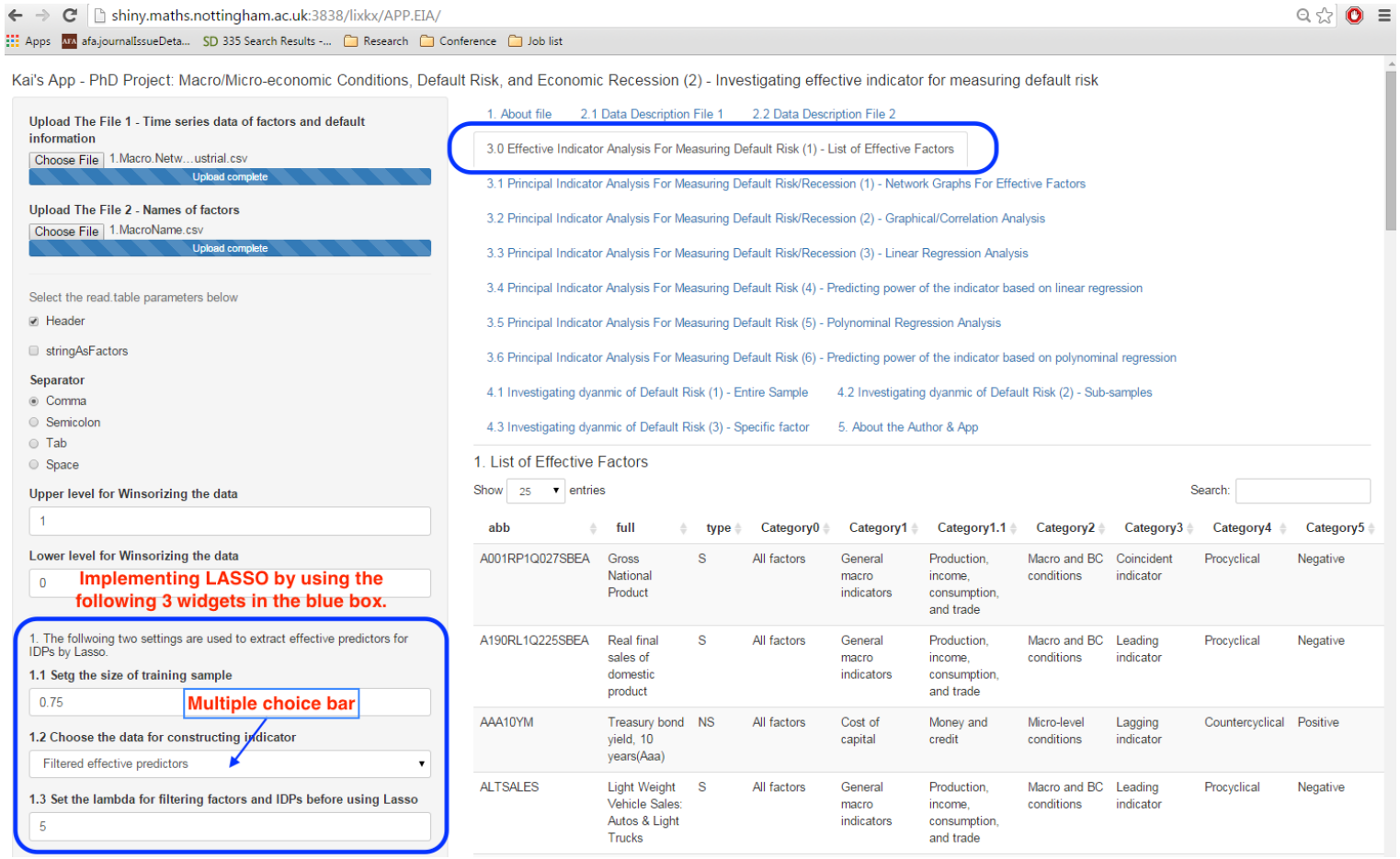


Figure 9-30 Structure of Tab 4.1 – 4.3

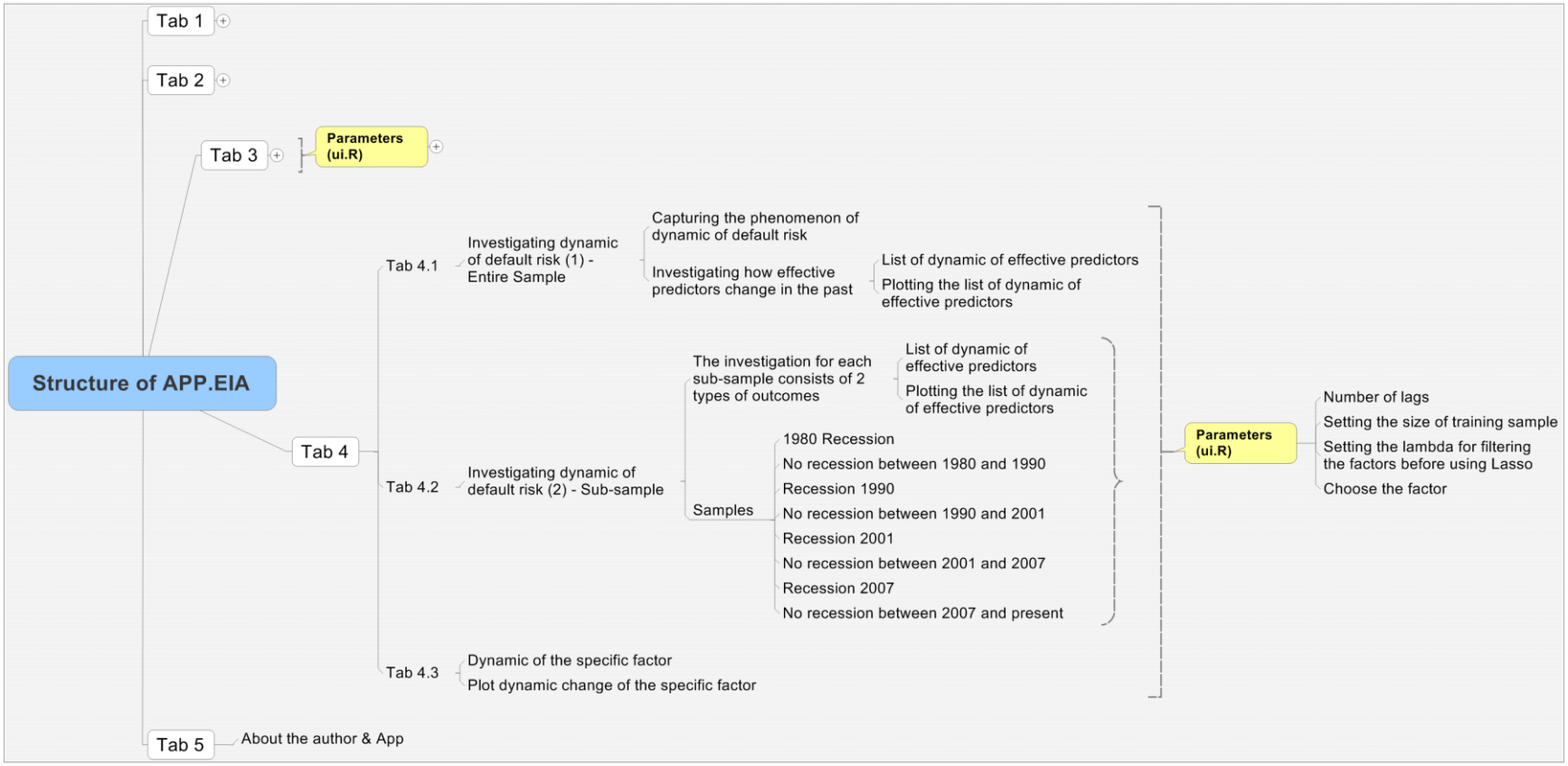
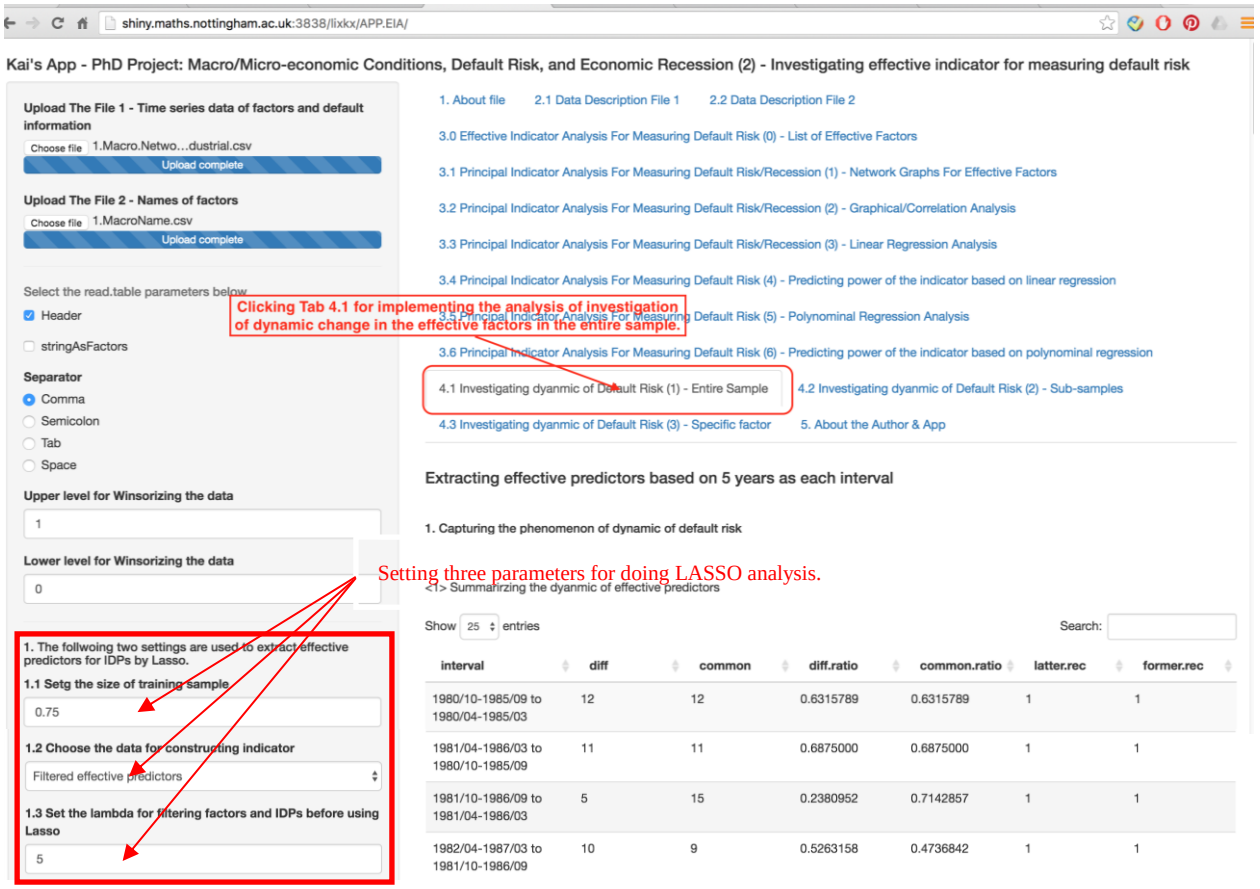
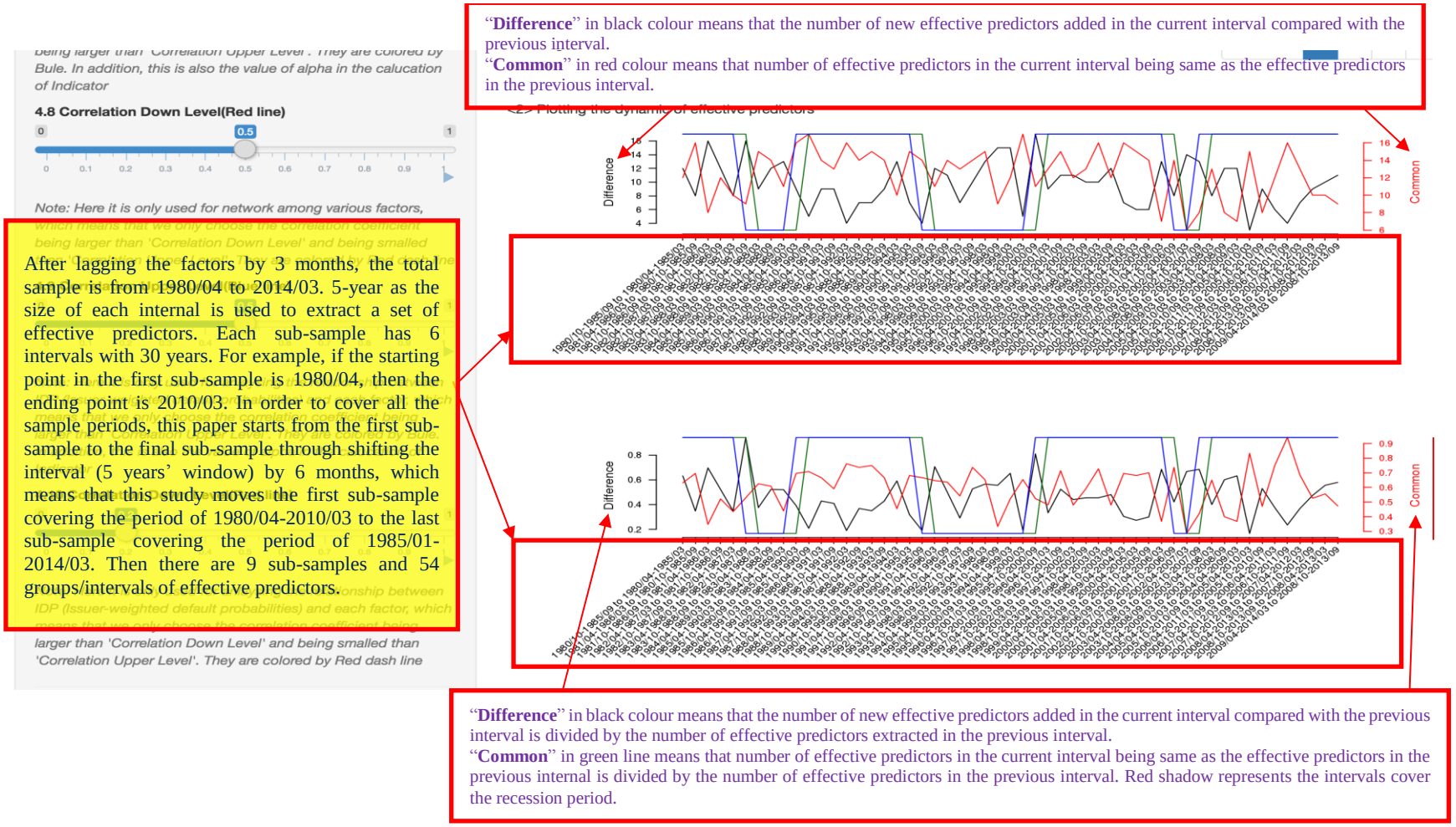


Figure 9-31 Investigation of dynamic in the effective factors in the full sample by Tab 4.1 (1)



Continued

Figure 9-32 Investigation of dynamic in the effective factors in the full sample by Tab 4.1 (2)



Continued

Figure 9-33 Investigation of dynamic in the effective factors in the full sample by Tab 4.1 (3)

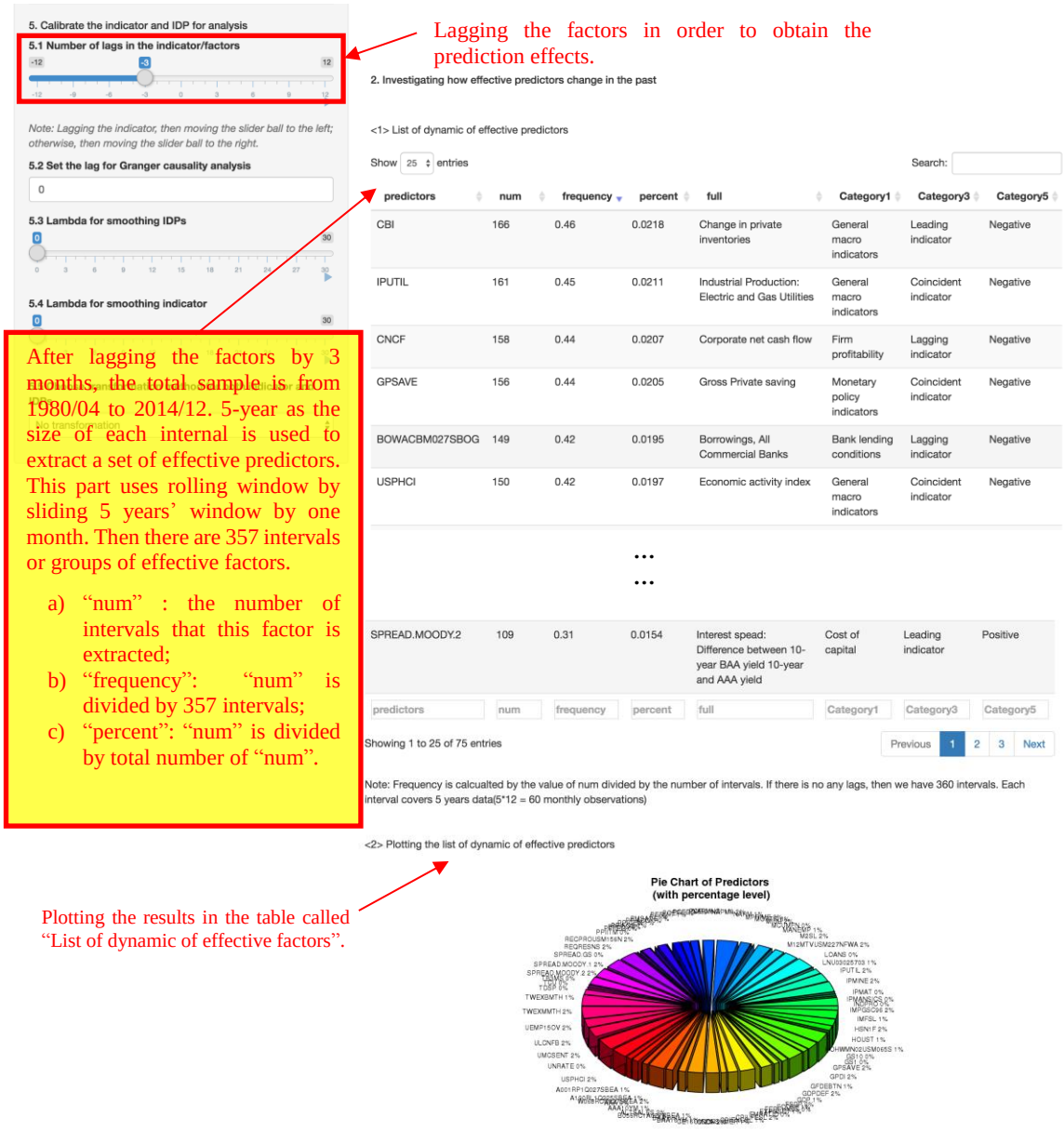


Figure 9-34 Investigation of dynamic in the effective factors in the 8 sub-samples by Tab 4.2

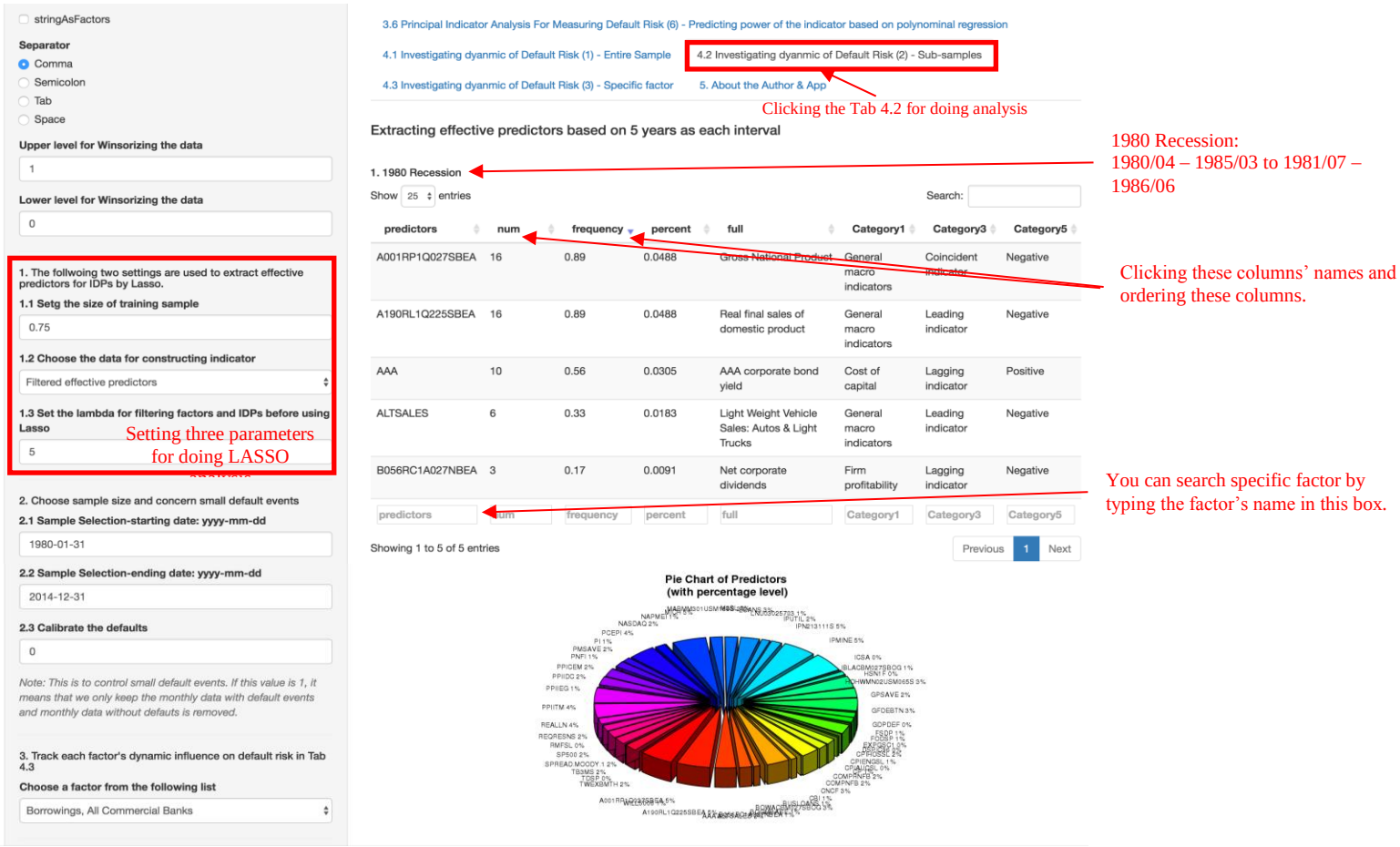


Figure 9-35 Investigation of dynamic in the spcific factor in the 8 sub-samples by Tab 4.3

