Econometric Study of Credit Cycles and Sectoral Risk with NLP-based Credit Risk Index Construction

A Thesis

Submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

Submitted by:

© Lubdhak Mondal

Indian Institute of Science Education and Research, Pune

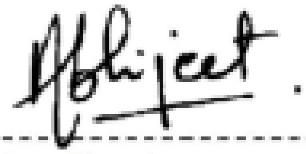
Under the supervision of:

Dr. Abhijeet Chandra (IIT Kharagpur)

TAC Member: Dr. Anindya Goswami (IISER Pune)

Certificate

This is to certify that this dissertation entitled **Credit Cycle and Sectoral Risk** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Lubdhak Mondal of Indian Institute of Science Education and Research, Pune under the supervision of Dr. Abhijeet Chandra, Assistant Professor, Vinod Gupta School of Management, Indian Institute of Technology, Kharagpur, during the academic year 2022-2023.



Abhijeet Chandra
Vinod Gupta School of
Management
Indian Institute of Technology,
Kharagpur

Committee:

Supervisor: Dr. Abhijeet Chandra (IIT Kharagpur)
TAC Member: Dr. Anindya Goswami (IISER Pune)

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

IISER PUNE

Declaration

I hereby declare that the matter embodied in the report entitled **Credit Cycle and Sectoral Risk**, are the results of the work carried out by me at the Asian Institute of Digital Finance, NUS Singapore and VGSOM, IIT Kharagpur under the supervision of Dr. Abhijeet Chandra (IIT Kharagpur), and the same has not been submitted elsewhere for any other degree.

Lubdhak Mondal

Lubdhak Mondal

This Thesis is Dedicated to

Anubhab Chatarjee
&
Aniruddha Gupta

Those who build me now, shape the person that I am!

Academic Acknowledgements

Lockdown allowed us to have much too much free time, and I'm sure I'm not alone in remembering that year vividly—2020. Lockdown was challenging, but it helped mold me. I wasn't the brightest student at college, but I always managed to have a smile on my face. I get a chuckle out of remembering the version of myself from five years ago when I didn't have my own computer, didn't understand what "coding" was, and just linked computers with playing video games.

I'm optimistic since lockdown allowed me to learn about finance, programming, and even cinematography! My interest in behavioral finance peaked in my third year of college, I reached out to Prof. Chandra, one of the country's foremost authorities in the topic. He assigned me research on the connecting crypto market and NLP. I didn't know much about (NLP) at the time, but because to his enthusiasm, we were able to complete and document a very unusual project. As a result of this job, I now have a deep appreciation for both NLP and finance.

The worth of time and how to make the most of opportunities became clear to me during this period of introspection. I worked on many projects at once so that I could learn as much as possible about the financial world. I've never been one to learn well in a classroom setting and have always preferred independent study. Lockdown provided this wonderful chance of zero attendance, which was really a gift for me. I could spend a lot of time learning about economics and enhancing my programming abilities. I found myself really enjoying the research itself. It was challenging to juggle numerous projects at once, but I was continuously pushing myself to do better.

I approached Prof. Chandra in December to serve as my master's thesis adviser. Despite my lack of expertise, he once again assigned me research on credit risk. As he briefed me and handed me the necessary paperwork, I immediately got to work on the job. Even though I had no idea how to construct a study subject on Credit Risk and NLP, I was adamant about include some NLP elements. Prof. Chandra's courage in taking a chance and maybe failing is quite inspiring, he never stopped me. To top it all off, he presented me with an incredible chance to spend six months as a visiting intern at the National University of Singapore Business School, working with a group of experts in the fields of credit risk and NLP. I owe him a great debt of gratitude for giving me this

opportunity, which was formative to my professional development. Because I wasn't going for PhD, I felt the need to educate myself as much as possible about the business world. I relocated to Singapore for the internship, put in my hours from 9 to 6:30, and put in my time at night and on the weekends to complete writing my thesis. I wouldn't encourage someone take on such a demanding job while doing a MS Thesis at the same time. This experience has taught me a great deal; I now understand how research is conducted in business settings, and most significantly, I gained my financial independence for next 2-3yrs .

All three of the incredible folks I met there, my thesis was successful thanks to the efforts of Prof. Jin Chuan Duan, my team leader Xuan Yao, and my advisor Hao Zhuang.

Working with Professor Goswami has been a pleasure as well. He guided me through a semester-long math project and two crucial courses, and he was always there when I needed assistance with my Thesis.

If Ramasish and Kapil weren't here, this article wouldn't be complete. I was mentoring Ramasish, a younger me at IIT, on a project he was developing. I have no clue how he juggled IIT with scraping the BSE stock market history, but he was a superb coder. And Kapil of course, with some similarities our theses, we're having a conversation, working together, and exchanging ideas. He was the key to finishing my thesis. My late-night pleas for assistance or explanations were never turned down by him.

We're in same boat , let's keep working together, please!

Personal Acknowledgements

As I sit here, overlooking the tranquil expanse of Marina Bay Sands, I can't help but reminisce about the journey that brought me here. The Fifth Year marks the end of one chapter and the start of another, but it would be an injustice not to acknowledge the late nights spent studying, the countless cups of tea that fueled me, and the sense of triumph that came with each milestone.

I still remember the trepidation I felt when I first arrived at IISER, surrounded by brilliant minds from all over the country. As a small-town kid, how could I ever hope to compete with such luminaries? But in time, I discovered that IISER wasn't just a roof over my head, but a place to call home. It was 3rd semester, I still recall my mother's voice, tinged with concern, as she asked me when I was planning to return "home". And I replied, with a sense of quiet conviction, "*Mom, I am already home*". It was here that I found solace and comfort amidst the chaos of academia, and where I forged lifelong friendships that would endure long after we parted ways.

My IISER experience wouldn't have been complete without Prantik and Ishita(Dipayan), my roommates and partners in mischief. Their impact on my life is immeasurable, and the memories we made together will always hold a special place in my heart. And then there was Saket, whose passion for film-making and social service inspired me every day. There were Sahil and Deepesh, whose eccentricity and fervour brought an unforgettable dynamism to my existence.

But it wasn't just my close friends who made my IISER journey so special. The experiences I had, from Disha to the enchanting Karavaan opening performance, from STS to Conclave, and of course, the unforgettable KALPA, all marked milestones in my personal and creative growth. Each one contributed to the vibrant and diverse community that made IISER such a unique and memorable experience.

And how could I forget Sagu Da and Ayan Da, whose guidance and mentorship helped me found the Student's Media Body? They were towering seniors whose impact on my life was profound, and I will always be grateful for their support.

Last but not least, I must pay tribute to the musical genius that Saikat and Ayan da were. Forming Advita is still a reminiscent memory.

I can't help but feel a sense of nostalgia creeping over me when I'm writing this. It seems like only yesterday that I arrived at IISER, full of wonder and anticipation. And yet, here I am, at the end of my journey. It was a long and winding road, with plenty of highs and lows, but through it all, I was cocooned by the amazing people I met at IISER even if I don't know what they gained from me and even if I did justice to them or not. There's definitely the temptation of writing a giant size letter to more people like Kunal, Purva, Goirik, Aditya, Amit and countless others whose names elude me now.

I have been blessed with the unwavering support and kindness of those around me. As I bring this letter to a close, I cannot help but wonder what the future holds. But whatever comes next, I know that I will carry the memories of my time at IISER with me always, a shining beacon of hope and light in the darkness.

Signing off

"*Let there be Kaavish, Mohineer Ghoraguli, Floyd, Murakami and ofcourse Krishnochura Full.*"

– Lubdhak Mondal

Abstract

This thesis proposes a study on the relationship between credit cycles and sectoral risks in the Indian context. The research will use a novel credit cycle index and a novel sectoral risk indicator based on firm-level data to provide more accurate and dynamic indicators of credit and market risks. Natural language processing (NLP) methods will also be applied to create a firm-specific sentiment index related to credit risk.

The primary objective of this research is to develop a novel sectoral risk indicator for eleven sectors, to investigate the relationship between credit cycle index and sectoral risks, and to develop investment strategies based on these studies. In addition to shedding light on the impact of macroprudential regulations and other credit risk mitigating variables, the study also aims to develop an NLP-based firm-specific sentiment index for credit risk.

The outcomes of this study may give investors and regulators with vital information for making educated investment choices.

Contents

1 Introduction **1**

1.1 Background of the Study 1

1.2 Problem Statement 3

1.3 Research Objectives 3

1.4 Disposition 4

2 Literature Review **5**

2.1 Understanding the Credit Cycle: Definition, Phases, and Impact on Financial Stability & Macroprudential Policies 5

2.2 Beyond Credit Quantity: The Need for a Risk-Based Credit Cycle Index 8

2.3 The VaR Controversy: Lessons Learned from the Subprime Mortgage Crisis 10

2.4 Beyond Financial Data: The Role of News Media in Credit Risk Assessment 11

3 Methodology **13**

3.1 Methodological Approach 13

3.2 Methodological Theories 14

3.2.1 Stationarity, Autoregression and Heteroskedasticity 14

3.2.2 Panel Data 17

3.2.2.1 Fixed Effect Model 19

3.2.2.2 Random Effect Model 21

3.2.2.3 Robust Covariance Estimation 23
3.2.2.4 Nickell Bias 25
3.2.2.5 Panel models with Driscoll and Kraay standard errors 26
3.2.2.6 Bias-corrected estimation of linear dynamic panel data models 28
3.2.3 Vector Autoregression 30

4 Classical Data and Alternative Data 37

4.1 Classical Data 37
4.1.1 The Credit Cycle Index 37
4.1.2 Sectoral Risk Indicator 38
4.1.2.1 Construction of the Indicator 40
4.1.3 Macroprudential Index Data 41
4.1.4 Control Variables 42
4.2 Alternate Data 43
4.2.1 Data Collection 44
4.2.2 Named Entity Recognition (NER) 44
4.2.3 Source Latent Dirichlet Allocation 44
4.2.4 Sentiment analysis by TABSA-BERT 45
4.2.5 Sentiment aggregation 46

5 Analysis and Discussion 49

5.1 Introduction 49
5.2 Panel Data 49
5.2.1 Annual Case 50
5.2.2 Monthly Case 52
5.3 VAR Results 55

5.3.1 Test Statistics 57
5.3.2 Impulse Response Function and Forecasting 57
5.4 NLP Results 72
5.4.1 Analyzing Credit Risk and Greenness Sentiment: User Input and Relevance Assessment 73
5.4.2 Real-Time Assessment of Credit Risk & Greenness Sentiment for Corporate Entities 76
5.4.3 Case Study 78
5.4.3.1 Credit Suisse 78

6 Conclusions **83**
6.1 Summary of Panel Regression Study 83
6.2 Summary of VAR Study 83
6.3 Summary of NLP Study 85
6.4 Recommendations for Further Research 86

Bibliography **88**

A Additional Reference and Selected Test Results **93**
A.1 Selected Forecast Error Variance Decomposition and Granger Causality 93
A.2 Probability Implied Rating 94

List of Figures

2.1 India CCI versus Credit-to-GDP cycle 9

4.1 Monthly CCI of Indian Banking Sector 38

4.2 Term Structure 38

4.3 Distribution of Defaults 38

4.4 No. of Companies 38

4.5 CVaR for general loss distribution 39

4.6 Sectoral Distribution of Nifty Bank 40

4.7 Sentence-pair fine-tuning with BERT 45

4.8 Full Methodology for Generating Entity-Specific Daily Credit Sentiment for Indian Case 47

5.1 Orthogonalized IRF Function for Banking Sector 58

5.2 Orthogonalized IRF Function for IT Sector 61

5.3 Orthogonalized IRF Function for Healthcare Sector 62

5.4 CCI & Sectoral Risk Forecasting for Banking, IT & Healthcare sectors 63

5.5 Orthogonalized IRF Function for Consumer Discretionary Sector . . 64

5.6 Orthogonalized IRF Function for Financial Services Sector 65

5.7 Orthogonalized IRF Function for Utility Sector 65

5.8 CCI & Sectoral Risk Forecasting for Consumer Discretionary, Financial Services & Utility sectors 66

5.9 Orthogonalized IRF Function for Energy Sector 67

5.10 Orthogonalized IRF Function for Industrial Sector 68

5.11 Orthogonalized IRF Function for Commodities Sector 69

5.12 CCI & Sectoral Risk Forecasting for Energy, Industrials & Commodity Sector sectors 70

5.13 Orthogonalized IRF Function for Reality Sector 70

5.14 Orthogonalized IRF Function for Telecommunications Sector 72

5.15 CCI & Sectoral Risk Forecasting for Reality & Telecommunications Sector sectors 73

5.16 Wordcloud for Credit Risk topic in English 74

5.17 Wordcloud for Credit Risk topic in Chinese 74

5.18 Wordcloud for Credit Risk topic in Bahasa 74

5.19 Wordcloud for Credit Risk topic in Vietnamese 74

5.20 Wordcloud for Greenness topic in English 74

5.21 Wordcloud for Greenness topic in Chinese 74

5.22 Wordcloud for Greenness topic in Bahasa 74

5.23 Wordcloud for Greenness topic in Vietnamese 74

5.24 Input of a user input article regarding Evergrande Group 74

5.25 Entity Specific Sentiment: 1 75

5.26 Entity Specific Sentiment: 2 75

5.27 Overview of the NS Monitor 76

5.28 Entity Specific Sentiment 77

5.29 Entity Specific Articles 77

5.30 Entity Specific Sentiment Index 78

5.31 Events Chronology at Credit Suisse *(Source: Morningstar)* 79

5.32 PD of Credit Suisse 80

5.33 Credit Suisse Sentiment Index 81

6.1 Impact of independent variables on sector-level Credit Risk Containment 84

A.1 Commodity Sector 93

A.2 Consumer Discretionary Sector 93

A.3 Energy Sector 93

A.4 Mapping 10-day moving average 1-year CRI PD to the S&P experience 94

List of Tables

1.1 Short and Long Term Credit requirements in different Indian Sectors 2

2.1 India's Significant Credit Events and CCI's Performance During Them 9

5.1 Summary Statistics for Annual Case. Time-period: 2000-2017 50

5.2 The lag structure of Sectoral risk wrt to Credit Cycle 51

5.3 Summary Statistics for Monthly case. Time-period: 2006-2020 53

5.4 Sectoral Risk and Credit Cycle Interplay Including the Crisis Period . 53

5.5 Summary Statistics for after crisis Monthly case. Time-period: 2009-2020 54

5.6 Sectoral Risk and Credit Cycle Interplay excluding the Crisis Period . 55

5.7 Granger Causality test for Baking Sector 59

5.8 Cholesky forecast-error variance decomposition for Banking Sector . . 60

6.1 Credit Channel Theory & Financial Instability Theory explanation of Credit requirements in different Indian Sectors 84

6.2 Credit Risk and Sectoral Risk Forecast Tab 85

1

Introduction

This section offers a quick overview of the study subject. The writer also addresses the issue, the objective of the research, and the topics covered. In order to provide the reader with a summary, we finish the chapter by discussing the disposition of the study.

1.1 Background of the Study

Credit cycle is expressed as credit growth rate or credit-to-GDP ratio in an economy. It measures total volume/quantity of credit available, predominantly seen in terms of short-term capital flows, in the economy. Policymakers use credit cycles to address macroeconomic situations such as loosening (or tightening) of capital flow. If such flows are not well managed, they may compound unwarranted currency appreciation and asset price increases, which forces policymakers to make choices to reverse capital flows, resulting in abrupt currency depreciation and production reduction owing to capital limitations.

As an early warning system, credit cycle metrics play a crucial role in anticipating financial crises. The aftermath of the global financial crisis of 2008 refocused attention on the credit view argument of Minsky, Kindleberger, and other academics who contend that financial crises are the result of a failed credit boom. During expansionary times, banks get too enthusiastic about the creditworthiness of borrowers and aggressively offer additional loans. This paves the way for an increase in credit risk, which may induce systemic risk and culminate in a financial crisis in the wake of an economic shock. A growing corpus of research has proven a link between credit expansion and the incidence of financial crises. For instance, Aikman et al. [1] and Schularick and Taylor [2] discovered a *two-year lag* between rapid credit growth and financial crises. Based on the "reckless lending" thesis and actual evidence, policymakers think that restricting credit expansion may prevent crises. As a result, regulatory programs intended to avert crises often use credit cycle quantity metrics as the intermediate goal variable to evaluate their efficacy.

However, it might have unique repercussions for economic agents operating in different sectors. For example, a cyclical credit flow shall affect the firms in banking, financial services, and insurance (BFSI) sector differently than the ones in automobiles or IT services sector, for each player might have different credit requirements and strategies. The table 1.1 provides an overview of the short and long-term credit requirements of India's various industries.

Table 1.1: Short and Long Term Credit requirements in different Indian Sectors

Sector	**Short-Term Credit Requirement**	**Long-Term Credit Requirement**
Energy	Low to Medium	High
Materials	Medium	High
Industrials	Medium	High
Consumer Discretionary	Medium	High
Health Care	Low to Medium	High
Financials	High	Medium to High
Information Technology	Low to Medium	High
Communication Services	Low to Medium	High
Utilities	Low to Medium	High
Real Estate	Medium	Medium to High

Looking at another vertical, in recent years, there has been a rising interest in using alternate data sources to get a deeper comprehension of financial markets and credit risk. The business press, which is often regarded as a barometer for market sentiment, is one such data source. The purpose of using business press text to generate alternative data is to capture the collective opinion and sentiment of market players about certain businesses, industries, or economic situations. This is especially pertinent in the context of credit risk, since sentiment research may give information into how creditworthy borrowers are regarded to be.

A wealth of information is a fundamental benefit of employing business press text as an alternate data source. There is an abundance of business press sources, such as newspapers, journals, blogs, and social media, from which sentiment indices may be derived. Furthermore, these sources give quick and current information, which is essential in the context of credit risk, since shifts in sentiment may have instantaneous effects on credit risk. The impacts of media sentiment on financial analysis for such as stock return, volatility and credit risk have been studied but are still in the early phase of a growing literature; for example, Grob Klubmann et al. [3], Qui et al. [4], Sun et al. [5], Dunham et al. [6], Gavaldon [7].

1.2 Problem Statement

The proposed study aims to assess the relationship between a unique credit cycle indicator and sectoral risks in 11 BSE-listed Indian industries. The project aims to provide a more precise and dynamic indicator of credit cycles than macroeconomic indicators by developing a credit cycle index using firm-specific data, such as probability of default(PD) generated by Duan et al. [8]. The whole project has a vertical structure, with the one of the vertical focusing on different Indian sectors classified by Bloomberg classification and exploring the relationship between the interrelationships between the novel credit cycle and Market risk at the sector level individually. In addition, the project intends to develop a dynamic panel model to examine the impact of macroprudential policies and other credit risk mitigating variables in the Indian context across all Indian sectors. By highlighting the interrelationships between the credit cycle index and sectoral risks, we also hope to contribute through the findings of our research to develop certain investment strategies based on the predictability of the credit cycle risks across varying time horizons. And for our last vertical structure aimed at the firm level, we aimed to produce a firm-specific sentiment index connected to credit risk, which would eventually assist investors and regulators in making more informed choices.

1.3 Research Objectives

We can divide the whole study thematically in two parts one being deploying the econometrics tools to study the problem for the first four verticals and for the last vertical we'll be using various state of the art tools from Natural Language Processing (NLP) to address the issue.

Following are the functional research objectives of the study:

1. Constructing the novel credit cycle index for 11 different sectors from probability of default values obtained from Credit Research Initiative, NUS in contrary to frequently used aggregate economy-based measures like Credit to GDP ratio.
2. Generating a sectoral risk indicator utilising a firm-level risk measure and aggregated for all the firms within the sample sector. Since market risk is additive, such an aggregate (or weighted average) measure of risk will denote the required Sectoral Risk Indicator.
3. To empirically examine the relationship between the novel credit cycle index and sectoral risks separately, to highlight the credit risk movement(s) in the various Indian sectors, and to obtain specific investment strategies based on the predictability of the credit cycle risks across varying time horizons.

4. Examine the influence of macro-prudential policies, Repo Rate, GDP growth and other Credit Risk factors on minimizing India's sectoral credit risk across all sectors by using a dynamic panel model.

5. Provide a framework for generating firm level credit risk sentiment index from publicly accessible business news and corporate filings using cutting-edge NLP techniques.

1.4 Disposition

Chapter one of the thesis gives an introduction and context of our study. It concludes by articulating the study's goals and the required procedures to achieve those objectives. In addition, the introduction includes a summary of the study's academic and industrial value.

Chapter two comprises the literature review. Throughout the literature review, the study's underlying econometric models are presented and defined. We utilize the literature review as a foundation for identifying theories, definitions, and key ideas, as well as past research conducted on the subject of the study. We also briefly highlighted the new research direction of using NLP to analyze Credit Risk, as well as some preliminary efforts in this subject.

Chapter three describes the technique used in our investigation and explains the reasons behind the selected strategy. This section concentrates mostly on the conventional econometric models and approaches we used for the first two verticals of our study. In addition, it presents some critical discussions on the methods used for the research.

Chapter four gives crucial insights into the sometimes ignored methods of data collection and construction. It also has a section titled *Alternative Data* that describes how we gathered and analysed newspaper data using cutting edge NLP techniques for the final research vertical. In addition, it briefly outlines the important NLP ideas used in our research.

Chapter five offers the study's results as well as our analysis and interpretation. The primary purpose of this chapter is to ultimately give insight into the research challenges described in the introduction.

Chapter six concludes the study. We provide an overview of the study's key results. In addition, the chapter gives recommendations and proposals for practical and scholarly ramifications.

2

Literature Review

In the literature review, we discuss the design of the Credit Cycle Index and the Sectoral Risk Indicator, as well as previous research on pertinent topics and the advantages and disadvantages of the methodology used in this study. In addition, the writers describe the ideas, models, and terminology used in the research. Lastly, we presented a short summary of a few seminal works on the use of NLP techniques to financial studies, along with an explanation of how they inspired and informed our research.

2.1 Understanding the Credit Cycle: Definition, Phases, and Impact on Financial Stability & Macroprudential Policies

The credit cycle has become an area of renewed interest and attention for both policymakers and academic researchers following the 2008 financial crisis. Unlike the business cycle, the credit cycle has no unified definition or measure. In contemporary market economies, the credit cycle is the cyclical pattern of credit expansion followed by credit contraction and economic stagnation, and is typically measured using credit quantity-related parameters such as the credit to GDP ratio or credit growth rate. The credit cycle consists of four phases:

1. Downturn
2. Repair
3. Recovery
4. Expansion

During the expansion phase, lenient lending practices and relaxed financial health checks may lead to reckless lending, increasing credit risk and the likelihood of borrower default. This, in turn, leads to the downturn phase, during which lenders become cautious, access to credit contracts, and credit growth slows down. The repair and recovery phases see companies improving their financial health, lending gradually increasing, and economic growth.

The emphasis of research on the credit quantity cycle has been on detecting and defining the cycle, analyzing its causes, and investigating its interaction with the business cycle and other macroeconomic variables. Drehmann et al. [9], Mouatt [10], Kiyotaki and Moore [11].

One key area of research in the study of the credit cycle is the identification and measurement of the cycle itself. Some researchers have argued that the credit cycle is driven by financial frictions that lead to fluctuations in the availability of credit, while others have suggested that changes in lending standards or regulatory policy may be responsible for the cycle. One influential paper in this area is the work of Drehmann and Juselius [9], who use a structural vector autoregression (SVAR) framework to identify the credit cycle and analyze its interactions with the business cycle. They find that the credit cycle has an independent effect on output, and that credit growth tends to lead output growth by several quarters.

The connection between the credit cycle and financial stability is an additional significant topic of study. Aikman et al.[1] have evaluated the association between credit growth and the probability of banking crises and discovered that rapid credit expansion is highly connected with an elevated risk of crisis two years lag. Fast credit growth is highly correlated with a greater likelihood of financial crises, and the "*Minsky moment*" hypothesis [12] suggests that rapid credit growth can create a self-reinforcing cycle of asset price increases, speculative behavior, and eventual collapse. Several studies have also explored the relationship between the credit cycle and financial stability, like the work of Schularick and Taylor [2], who contend that credit booms may lead to an increase in systemic risk and the probability of financial catastrophes.

A macroprudential policy is a method for financial regulation that seeks to reduce the frequency and severity of financial crises. Typically, the effectiveness of this program is determined by its potential to achieve its ultimate goal. Using cross-country panel regressions, scholars such as Fendoglu [13] and Cerutti et al. [14] have produced indices of the macroprudential policy stance for widely used instruments and assessed their efficiency in restricting credit quantity/growth. Their results indicate that macroprudential policy exerts substantial moderating effects on lending volume. While it is often believed that macroprudential policy is effective at controlling loan amount and, thus, has a moderating effect on an economy's credit risk cycle, this transitivity theory presents some issues.

According to Barrell et al. [15], not all credit-to-GDP amplifications are inevitably "credit booms gone wrong" caused by irresponsible lending. In certain instances, financial intermediation distributes loans for productivity increases as opposed to risky lending. In such situations, macroprudential devices may reduce market efficiency and heighten the risk level. In addition to the quantity channel, macroprudential instruments may impact credit risk through various transmission mechanisms. For example, with severe credit limits, banks may be pushed to explore riskier initiatives to maintain profitability and competitiveness. These channels may work in the same or opposite direction as the quantity channel, resulting in increased or reduced policy effects. Hence, not only may credit quantity impact future risk, but risk perception may also encourage self-adjustment in credit quantity through market hazards.

In the aftermath of the global financial crisis of 2008, macroprudential policies have been progressively implemented to enhance financial stability by addressing credit risk, a significant contributor to financial instability. These strategies aim to address both the cross-sectional and temporal elements of credit risk. This is strongly connected to the credit cycle, which refers to the gradual buildup of risk in the financial system. The objective of macroprudential policies is to lessen the pro-cyclicality of the financial system by amassing buffers during times of economic prosperity that may be used during times of economic stress.

It has been proposed that the credit-to-GDP ratio and credit growth are reliable predictors of financial crises. The Basel Committee on Banking Supervision has approved the credit-to-GDP ratio as a guidance for national regulators on countercyclical capital cushions. Yet, a number of studies have shown that not all financial crises are related with excessive growth and that high credit expansion does not inevitably result in financial instability. Thus, it is possible that these metrics may not reliably foresee financial crises and are insufficient to promote financial stability.

Traditional credit risk proxies, such as default rates, firm failure rates, and non-performing loans, are ex-post result measurements that are incapable of serving as early warning indicators for coming crises. In addition, they may not be best for random occurrences, such as corporate defaults, if the sample size is small since they may be quite noisy.

Yet, these assessments may exclude crucial data on the allocation and riskiness of the credit system, which are essential for financial stability. In addition, it is impossible to determine in real time where the economy is in its cycle and when to intervene to prevent a financial disaster. As a consequence, it is preferable to focus on direct credit risk indicators while implementing macroprudential tools and assessing their efficacy.

Unfortunately, much studies have not been done in these contexts, though Duan et al. [8] sheds light on the aforementioned difficulties by establishing a unique credit cycle indicator that reflects the movement of credit risk. The credit cycle typically has a standard effect on the overall economy, changes in the availability of credit can influence the performance of firms in different sectors in different ways. For instance, firms in the automobile sector may react differently than firms in the construction industry. Finally, the Indian context remains understudied, and research is needed to explore the impact of market risks on the credit cycle at sector levels.

The proposed new credit cycle indicator derives on the probabilities of default (PDs) measure created by the Credit Research Initiative (CRI) at National University of Singapore. It has been proven that these indices are autoregressive and indicative of well-known events that impact the outlook for credit risk. In addition, the credit risk cycle has a separate cyclical pattern from the quantity cycle, as measured by the ratio of credit to GDP. Cross-Country Panel research by Cerutti et al. [14] and Duan et al. [8] revealed that whereas macroprudential policy has considerable moderating impacts on credit growth and the growth is definitely linked with greater risks with a two-year lag, the policy has not itself mitigated credit risk. This shows that people may have neglected alternative transmission routes that relate the use of macroprudential devices to an increase in credit risk. The entire analysis reveals that the credit risk for 52 nations increased between 2000 and 2013. In our study, we collected sector-level data for eleven distinct industries in the Indian setting and analyzed their credit cycle. We also found some similar trends and results to Duan et al. [8] with few other narratives to ponder upon. At the country level, Duan contends that even if macroprudential policy's influence on credit quantity has been considerably reduced, its ineffectiveness on credit risk and the correlation between credit quantity and risk continue.

2.2 Beyond Credit Quantity: The Need for a Risk-Based Credit Cycle Index

Widespread implementation of macroprudential measures has reduced the incidence and severity of financial crises. Often, these policies are evaluated based on their capacity to attain this final objective. Concerns exist, however, about the efficacy of macroprudential interventions in managing credit risk cycles. Although it has been established that such measures may have a substantial moderating influence on loan volume, the transitivity argument relating credit volume and credit risk cycles is questionable. Not all increases in the credit-to-GDP ratio are connected with "*failed credit booms*" and "*reckless lending.*" [2] In certain instances, financial intermediation distributes loans for productivity enhancements as opposed to hazardous lending. Taxing such situations using macroprudential instruments may reduce market efficiency and increase risk.

In addition, macroprudential policies may influence credit risk via pathways outside credit quantity. To protect profitability and competitiveness, strict lending constraints may incentivize banks to undertake more risky projects, while other channels may operate in a similar or reverse direction as the quantity channel, resulting in policy impacts that are either strengthened or reduced. Hence, judging the efficacy of macroprudential policies based purely on credit quantity metrics may not be adequate, and the development of a risk-based credit cycle index is required.

Year	Credit Event	Description
1997-98	Asian Financial Crisis	The financial crisis affecting many Asian countries, leading to a sharp depreciation of the Indian rupee. Both CCI and CCI Bank remained very high slowly coming down around 2000
2001-02	Dot-com Bubble and Economic Slowdown	The bursting of the dot-com bubble and economic slowdown led to a decline in credit growth and rising NPAs. This was originated from the USA and India affected a little later. But we can clearly see rising trend in both CCI and CCI Bank from mid-2000.
2008	Global Financial Crisis/ Sub Prime Crisis	Global financial crisis triggered by the collapse of Lehman Brothers, leading to a slowdown in credit growth. This is similar to dotcom bubble originated in USA and later affects India. This case has been discussed earlier.
2013	Indian Rupee Depreciation and Current Account Deficit	Depreciation of the Indian rupee and high current account deficit led to concerns about the country's finances. This can be seen with a sudden but slight increase in CCI Bank from end of 2012.
2015	Kingfisher Airlines Loan Default	Kingfisher Airlines defaulted on loans, leading to a rise in NPAs and concerns about the banking sector. A clear upward trend can be seen in CCI from the end of 2014.
2016	Asset Quality Review of Banks by RBI	RBI conducted an asset quality review of banks, leading to increased NPAs and concerns about bank stability. This too fueled the massive increase in CCI from the end of 2014
2020-2021	COVID-19 Pandemic and Lockdowns	The pandemic and lockdowns led to a decline in economic activity, affecting credit growth and NPAs. The CCI Bank shows a clearly upward trend from the beginning of 2020, whereas the Credit-to-GDP accounted for in late 2020 and 2021.

Table 2.1: India's Significant Credit Events and CCI's Performance During Them

Duan et al. [8] has developed a novel credit cycle index, which has been proven reliable in US markets. Credit events, such as the 2008 subprime crisis, have been re-evaluated in light of these newly developed indicators. Credit Cycle Index data was received from CRI-NUS, Credit and GDP quarterly data for India was got from Fred, and a Credit-to-GDP index was developed. Based only on the Credit-to-GDP index, it was determined that credit amount follows a cycle with a period of around 12 to 13 years, which is consistent with the financial/credit cycle described in prior research [9], [1]. The Credit Cycle Index (abbreviated CCI) represents the credit risk cycle.

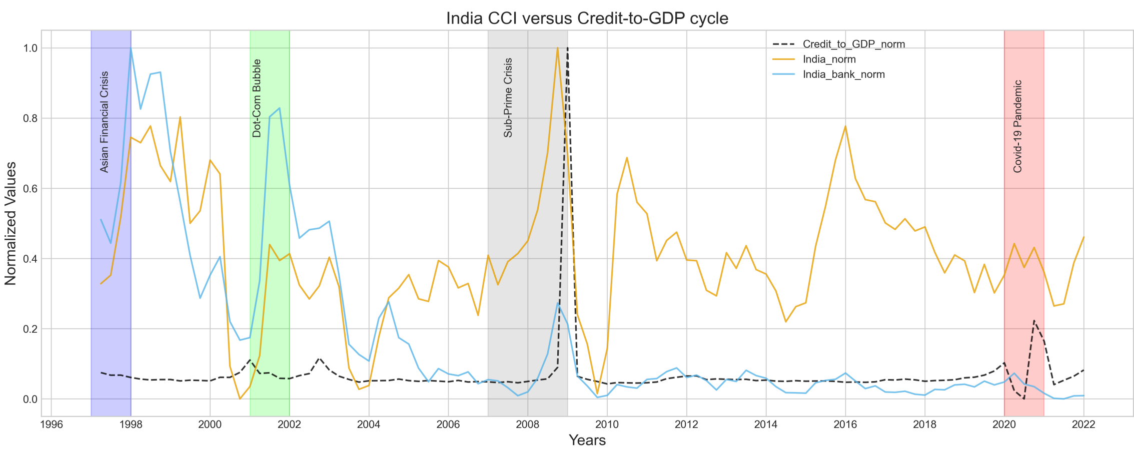


Figure 2.1: India CCI versus Credit-to-GDP cycle

The innovative Credit Cycle Index functions as an early warning signal. From picture 2.1 we can observe, at the beginning of 2007, both CCI India and CCI Indian Bank have risen slowly, indicating a major credit event. This is made much evident by examining the credit-to-GDP ratio and CCI Indian bank during the subprime crisis, when the crisis occurred when the credit quantity was at its lowest point, yet CCI and CCI bank provided early warnings. The Credit Cycle Index is a valid instrument for assessing the credit risk cycle because it takes into account variables other than credit amount that might have a major influence on credit risk. Table 2.1 illuminates India's significant credit events and the performance of CCI and CCI Bank during those time in brief. With the exception of the dot-com bubble, which started mostly in the United States, and the shock of the U.S. economic downturn, which reached India in a later phase, the CCI has signaled all significant credit events in advance.

2.3 The VaR Controversy: Lessons Learned from the Subprime Mortgage Crisis

Market risk is the risk of incurring losses as a result of fluctuations in market factors such as stock price or volatility. As market risk is evaluated using stock market data, our research examines how the CCI values/credit cycle effects the market parameters of enterprises belonging to various industries.

Of all market risk measures, Value at Risk(VaR) and Conditional Value at Risk (CVaR) are of greatest relevance to us. VaR is a statistical method used to estimate, with a specified degree of confidence, the potential losses that an investment portfolio or trading position may incur within a certain time period. Conditional Value at Risk, also known as Expected Shortfall (ES), is a risk metric that computes the anticipated loss in the distribution's tail beyond the Value at Risk (VaR). CVaR is a companion measure to VaR, which estimates the greatest loss that may occur with a given probability. CVaR, on the other hand, calculates the average loss that may occur over the VaR threshold when a portfolio or trading position is subjected to extreme market circumstances.

From its introduction, VaR in particular has been heavily criticized since it only analyzes the frequency of significant losses (β near to 1) and not their severity. ES handles the key "what-if" question more well. Nevertheless, VaR is not a subadditive risk measure in general. that is, given two places C_1 and C_2, it is feasible that,

$$\mathrm{VaR}_\beta\left(C_1+C_2\right)>\mathrm{VaR}_\beta\left(C_1\right)+\mathrm{VaR}_\beta\left(C_2\right) \text{ for some } 0<\beta<1.$$

This complicates the case for risk aggregation and diversification. In contrast, CVaR is always sub-additive as specified in Definition 9 in Chapter 4. These factors contributed to the subprime mortgage crisis of 2007–2009 (for pertinent debates and for in-depth study please refer Donnelly Embrechts [16], McNeil et al. [17]. Even if industry in the mid-1990s did not heed early warnings about the VaR problem in abnormal markets, the negative concerns underpinning VaR-based regulation are now clearly obvious.

2.4 Beyond Financial Data: The Role of News Media in Credit Risk Assessment

Natural Language Processing (NLP) techniques offer a means of analyzing unstructured data sources, such as social media posts, news articles, and customer reviews etc. Topic modeling can assist in identifying patterns and themes in news articles, which in turn can inform Credit Risk assessment. NLP techniques can also enhance Credit Risk modeling accuracy and efficiency. For instance, text classification can categorize loan applications into different risk categories, while named entity recognition can identify essential entities, including borrowers' addresses and names.

In the business world, textual information from sources such as formal filings and analyst reports has played a crucial role in decision-making for decades. With the digitalization of text-based material, these sources are now suitable for machine-based analysis. However, text streams are unstructured and present greater complexity than numerical data. Research into the effects of media sentiment on variables such as stock return, volatility, and credit risk are still nascent, but the potential for utilizing these text streams for decision-making is vast. This study emphasizes the extraction of topic-focused and entity-specific media sentiments at the article level, as opposed to generic sentiments about companies.

Aziz et al. [18] present a unique method for assessing credit risk utilizing topic modeling approaches. They believe that conventional credit risk models, which depend primarily on financial and accounting data, may not be enough in today's information-rich world, when a large amount of information is accessible in the form of unstructured data, such as news articles and social media postings. To enhance credit risk assessment, they suggest a hybrid technique that blends standard financial data with unstructured data sources. In particular, they employ Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF) to a corpus of financial news articles and stock prices data in order to extract latent themes from the text data, and then utilize these topics as features to forecast stock price changes. The authors show that topic modeling may be used to discover pertinent aspects and give insight into the credit risk drivers, investment strategies and decision-making in finance.

In order to represent the credit-centric nature of the research, Duan et al. [19] designed a method for determining an article's relevance to credit risk and then use it to weight the views stated about all businesses featured in the article. Naturally, the sentiment stated about businesses are conveyed at the sentence level, but the it has aggregated at the article level. As a result, entity-specific sentiment ratings might change based on the sentiments expressed about various companies in the same article. Business press has been chosen over social media because the perspectives stated by professional journalists are arguably more perceptive and impartial, and the editorial control aids in ensuring quality and consistency. This research relies on the daily business press rather than other forms of texts since they provide as systematic and timely information sources.

In this study, a three-class categorical outcome variable was developed to predict default risk, with values of 0, 1, or 2 representing survival, default, or other exit, respectively. However, establishing the degree of an article's relevance to credit risk is challenging due to the broad nature of credit risk as a topic. To tackle this problem, a current sentence-based alternative called targeted aspect-based sentiment analysis (TABSA-BERT) was employed, which involves grouping the target (corporate name) and aspect (topic) to create an auxiliary sentence to supplement the original sentence and link it to the expressed sentence-based sentiment. However, as credit risk cannot be conveyed through a single word or sentence, it should be evaluated holistically, similar to how people do. They used Source-LDA of Wood [20] to extract article-level topics from a corpus of around 1.7 million English articles. We choose Source-LDA over LDA of Blei et al. [21] due to the necessity to assist the extraction of credit risk this incremental transformation really helped the algorithm to sensibly extract topics related to credit risk.

I got the opportunity to work with the NLP team of AIDF NUS. Using Duan et al. [19] paper as a foundation, I contributed to the extraction of other themes pertaining to greenness, gold, and commodities, among others. In addition to English and Chinese, we built language models for Vietnamese, Bahasa, Indonesian, and Malay, which allowed us detect news at a more regional level. And last, we created a website that accesses 30 or more major RSS feeds from across the world and generates hourly firm-specific credit risk and greenness sentiments from news stories. I discussed more about this in the end of the chapter 5.

3

Methodology

In this chapter, the We explain how the research was conducted. The fundamental principles and ideas behind the approach adopted for the investigation are outlined. Explanation and presentation of the data material and its organization utilized in the research. The chapter concludes with a review of the study's reliability and validity, followed by criticism of the methodology used and the study's shortcomings.

3.1 Methodological Approach

The choice of methodological approach for the research relies on the description of the issue and the objective of the investigation. Our objective is to use existing theories to produce results rather than create new ones, which necessitates the use of numerous econometrics methods such as Panel Models, Vector Auto-regressive Methods (VAR), Impulse Response Functions, Forecasting and the Granger Causality Test, among others. In this whole chapter, we'll examine several underling theories in terms of the methodologies they're acquainted with, and then we'll get into data collecting and compilation in next chapter, a tedious but frequently disregarded procedure.

Given that the investigated phenomena are portrayed statistically and not in narrative form, we have opted to use a quantitative methodology. The quantitative research is comprised of monthly and annual measurements of eleven industries collected between 2000 and 2020. We utilised both monthly and annual data depending on the necessity. By utilizing monthly data, it is possible to see greater variety in the time series. Due to the fact that some financial indicators (such as GDP Rate and Policy Rate) are only accessible quarterly (or when RBI modifies them), missing values are occasionally created using linear extrapolation. In case of long time series data, both eras of severe economic contraction and expansion are included. Initially, we followed the benchmark regression by [14] with the inclusion of the market risk indicator (CVaR) as a predictor variable; after doing further study on the issue statement, we occasionally incorporated few other predictor variables also.

During the research, we paid special attention to small-cap stocks because the market risk indicator construction demonstrates that corporate actions or defaults occur more frequently in small-cap companies and these stocks are inherently more volatile. Small-cap equities expose investors to a higher credit risk than other-cap stocks.

India is one of the fastest-growing emerging countries and has the fifth-largest GDP, yet the Bombay Stock Exchange has not been studied much for credit risk, particularly at the sector level. The US financial crisis that extended to India is a good example of how markets are interconnected through globalization. The Indian market is quite dynamic and fascinating to analyze. In addition, we see a substantial amount of volatility in GDP growth and the RBI Repo Rate over the study period.

3.2 Methodological Theories

3.2.1 Stationarity, Autoregression and Heteroskedasticity

Before starting with this section this is to be noted that these all four concepts can be used in both time-series and panel settings. Because of the page constraints we'll be elaborating on using these techniques on panel data settings mainly as those are little more mathematically involved, once one has a proper understanding of these concepts in panel settings intuitively it is quite easy to understand applying these techniques in time-series settings. The name of the tests can be different but one can intuitively figure out that panel data is combination of cross section and time series data so time component is inherently there.

Definition 1 (Stationarity)**.** *A time series X_t is considered weak stationary if and only if the following requirements are satisfied,*

- ***Mean-stationarity****: The mean of the series remains constant with respect to time, i.e.,$E(X_t) = \mu$ for all t*
- ***Covariance-stationarity****: The covariance of the series is constant over time, i.e., $\mathbf{Cov}(X_t, X_{t+k}) = \sigma^2$ for all t and k, where σ^2 is the variance of X_t.*

Panel data approaches, including fixed effects and random effects models, require stationary variables. Panel data sets containing non-stationary variables may yield skewed and inconsistent estimates, making variable relationships difficult to determine. Panel data unit root tests establish if a panel data collection contains unit roots. Non-stationary time series that return to their mean after oscillations have unit roots. Panel data with unit roots can complicate econometric analysis by making it harder to model variables and draw conclusions. Panel data unit root tests look for unit roots in each time series and their relationships. The LevinLin–Chu [22], Harris–Tzavalis, Breitung, Im–Pesaran–Shin, and Fisher-type test are some well-known tests for stationarity Choi test. Due to the asymptotic condition of the LLC test, we will focus mostly on the most well-known test, the LevinLin–Chu test.The assumption that N is constant as $T \to \infty$ is standard in macroeconomic studies like mine. In order to pass the LLC test, T must rise faster than N, so that $N/T \to 0$.

Definition 2 (Levin–Lin–Chu Test)**.** *LLC test is used to determine whether a panel data set contains unit roots or is stationary,*

- H_0*: All of the individual time series in the panel data set have a unit root, i.e., the series are non-stationary.*
- H_a*: Atleast one of the individual time series in the panel data set is stationary, i.e., the series is stationary.*

Let us begin with a simple panel-data model with a first-order autoregressive term,

$$x_{it} = \alpha_i x_{i,t-1} + \mathbf{y}'_{it}\delta_i + \epsilon_{it} \tag{3.1}$$

where $i = 1, \ldots, N$ indexes panels $t = 1, \ldots, T_i$ indexes time; x_{it} is the variable being tested; and ϵ_{it} is a stationary error term. Testing circumstances might specify that x_{it} is the means across all panels, means across all panels plus a temporal trend, or nothing at all. By default, $\mathrm{y}_{it} = 1$, so that the term $\mathrm{y}'_{it}\delta_i$ represents panel-specific means (fixed effects). Panel unit-root tests are used to test the null hypothesis $H_0 :$ $\alpha_i = 1$ for all i versus the alternative $H_a : \alpha_i < 1$. Depending on the test, H_a may hold, for one i, a fraction of all i or all i; the output of the respective test precisely states the alternative hypothesis. Equation (3.1) is often written as

$$\Delta x_{it} = \beta_i x_{i,t-1} + \mathrm{y}'_{it}\delta_i + \epsilon_{it}$$

so that the null hypothesis is then $H_0 : \beta_i = 0$ for all i versus the alternative $H_a :$ $\beta_i < 0$. While conducting an LLC test, it is assumed for simplicity's sake that all panels have the same autoregressive value so that $\alpha_i = \alpha$ for all i.

Definition 3 (Heteroskedasticity)**.** *Let A be the dependent variable in a linear regression model and B the vector of independent variables. Model representation is,*

$$A = \alpha_0 + \alpha_1 B_1 + \alpha_2 B_2 + ... + \alpha_k B_k + \epsilon$$

where $\alpha_0, \alpha_1, \alpha_2, ..., \alpha_k$ are the coefficients of the independent variables and ϵ is the error term. Heteroskedasticity arises when the variance of the error term ϵ, indicated by $Var(\epsilon)$, varies across data. Heteroskedasticity may be expressed mathematically as:

$$Var(\epsilon) = \sigma_i^2 = \sigma^2(B)$$

where σ_i^2 is the conditional variance of the error term for the ith observation, and $\sigma^2(B)$ is a function of the independent variables B.

Heteroskedasticity influences the characteristics of the ordinary least squares (OLS) estimators and the standard errors of the coefficients, potentially making the OLS estimators inefficient and the standard errors inaccurate. For this reason, we've to rely on several robust standard error panel data approaches when we found out heteroskedasticity is indeed there in our data. There is a special methodology to assess heteroskedasticity in stata, which uses the General Least Square method (GLS) in panel settings and the Likelihood-ratio (LR) test after estimation, using the null hypothesis that the parameter vector of a statistical model meets a smooth constraint. Since iterative GLS with just heteroskedasticity yields maximum-likelihood parameter values, an LR test is straightforward to do. The steps are as follows,

1. Fit the model with heteroskedasticity at the panel level and preserve the likelihood.
2. Conform the data to the model without heteroskedasticity.
3. When fitting nested models, *lrtest* normally infers the number of constraints based on the number of estimated parameters. However, with GLS, panel-level variances are calculated as nuisance parameters, and their count is omitted from the estimated parameters. Therefore, we will need to inform *lrtest* of the number of constraints we have suggested.
4. Then we may do the *lrtest*, with assumptions nested inside heteroscedastic relationships. If the p-value is less than 0.05, we may thus reject the null hypothesis and infer that heteroskedasticity is actually present in the panel.

Definition 4 (Autoregression)**.** *The mathematical definition of autoregression is a regression model in which the dependent variable is described as a linear combination of its own previous values plus an error component. This is the mathematical equation for an autoregressive model of order m,*

$$A_t = \mu + \alpha_1 A_{t-1} + \alpha_2 A_{t-2} + ... + \alpha_m A_{t-m} + \epsilon_t$$

where,

- A_t *is the value of the time series at time* t.
- μ *is a constant term.*
- $\alpha_1, \alpha_2, ..., \alpha_m$ *are coefficients of the lagged values.*
- $A_{t-1}, A_{t-2}, ..., A_{t-m}$ *are the lagged values of the time series.*
- ϵ_t *encapsulates the unexpected and random elements of a series*
- m *represents the order of the autoregression model, which indicates the number of lagged data utilized by the model.*

In this model, each value of A_t is reliant on its previous values, with the coefficients $\alpha_1, \alpha_2, ..., \alpha_m$ capturing the intensity and direction of this dependency. Typically, it is assumed that the error term ϵ_t is i.i.d. with mean 0 and constant variance. Duan [8] have previously discovered that the developed Credit Cycle Index has a *robust autoregressive characteristic*. Therefore, we must use **dynamic panel models** for our research. Duan [8] also verified that the CCI is stationarity.

As shown in the next section, our panel data are definitely heteroskedastic, and there is a significant autoregressive relationship in our dependent variable. We investigated many panel models, but ultimately settled on dynamic panel models in which the error structure is assumed to be heteroskedastic, corrects the coefficients for standard errors and is applicable for Small N and Large T or Small N and Small T circumstances, also addressing the heteroskedasticity issue.

3.2.2 Panel Data

Panel data or longitudinal data are repeated measurements throughout time on the same particular unit, such as a company, state, industry, or nation. Then, regressions may capture both variation across units and variation over time, analogous to regression on cross-sectional data. Cross-sectional data approaches are more sophisticated than panel data methods. The standard errors of estimators of panel data must be updated since each subsequent data period is not independent of preceding ones. Panel data demands far more complex models and estimating techniques. Panel data, like time series data, are a set of sequential observations gathered at predetermined intervals and in chronological sequence. Panel data, like cross-sectional data, are collected observations from a large sample of people.

There are several benefits to panel data,

1. Individual heterogeneity may be explicitly accounted for in panel data (where "individual" refers to micro-units).
2. By combining data in two dimensions, panel data give larger data variance, reduced collinearity, and more degrees of freedom.
3. Panel data are superior than cross-sectional data for examining the dynamics of change. For example, it is useful for comprehending transitional behaviour, such as firm bankruptcy or credit risk events.
4. Panel data enables the analysis of more complex behavioral models, such as the effects of technological progress and economic cycles.
5. Panel data may lessen the effects of aggregation bias when large groups of firms are pooled.

Drawing from Zellner's seemingly unrelated regression models [23], panel data models denote the dependent and explanatory variables with two (or more) subscripts that indicate the variable's characteristics typically individual and time, but possibly location, group, etc. Fixed effects models and random effects models are particular instances of the conventional linear regression model and the extended regression model, respectively. If all cross-sectional units have the same number of time series observations, the panel is balanced; otherwise, it is imbalanced. A simplest example of balanced panel will be,

$$y_{it} = \mathrm{x}'_{it}\boldsymbol{\gamma} + \phi_i + \varepsilon_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{3.2}$$

where dependent variable y_{it} satisfies a linear model with an intercept that is specific to individual i. When using matrix notation for the data set, the Kronecker product notation is helpful. In contrast, clustered or grouped data models often presume the number of observations per group i might vary between groups, which in this notation would substitute the common number of "time periods" T with a groups-specific number T_i of individuals. Typically, for panel data models, observations are stacked in the reverse order of subscripts, i.e., first collecting the observations over time for each individual as,

$$\underset{(T\times 1)}{\mathbf{y}_i} = \underset{(T\times K)}{\mathbf{X}_i}\,\boldsymbol{\gamma} + \phi_i \eta_T + \varepsilon_i \tag{3.3}$$

for $i = 1, \ldots, N$, where $\mathbf{y}_j$ and ε_j are T-vectors and $\mathbf{X}_i$ is a $T \times K$ matrix,

$$\underset{(T\times 1)}{\mathbf{y}_i} = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \cdots \\ y_{iT} \end{pmatrix}, \quad \underset{(T\times 1)}{\varepsilon_i} = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \cdots \\ \varepsilon_{iT} \end{pmatrix}, \quad \underset{(T\times K)}{\mathbf{X}_i} = \begin{pmatrix} \mathbf{x}'_{i1} \\ \mathbf{x}'_{i2} \\ \cdots \\ \mathbf{x}'_{iT} \end{pmatrix},$$

and η_T is a T-dimensional column vector of ones. Then, stacking the entire data set by individuals,

$$\underset{(NT\times 1)}{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \cdots \\ \mathbf{y}_N \end{pmatrix}, \quad \underset{(NT\times 1)}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_N \end{pmatrix}, \quad \underset{(NT\times K)}{\mathbf{x}} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \cdots \\ \mathbf{x}_N \end{pmatrix}$$

and defining

$$\underset{(N\times 1)}{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \cdots \\ \phi_N \end{pmatrix}$$

the data may be represented by the single (very simple) equation,

$$\mathbf{y} = \mathbf{X}\gamma + \mathbf{A}\phi + \varepsilon$$

when,

$$\underset{(NT\times N)}{\mathbf{A}} \equiv \mathbf{I}_N \otimes \eta_T$$

In panel data models, the "individual intercept" ϕ is intended to adjust for the impact of unobservable regressors that are particular to individual i so as to account for the influence of unobservable regressor so that,

$$\phi_i = \mathrm{w}_i'\lambda$$

Different theories about the connection between both the observable regressors x_{it} and the intercept term ϕ_i and consequently the non-observable regressors w_i, provide versions of the conventional and GR models. With this nomenclature, it is evident that the matrix X lacks a single column vector of ones, since a linear combination of the matrix $\mathbf{A}$ of individual-specific "dummy variables" would result in such a vector,

$$\begin{aligned}\mathbf{A}\cdot\eta_N &= (\mathbf{I}_N \otimes \eta_T)(\eta_N \otimes 1)\\ &= (\mathbf{I}_N \cdot \eta_N \otimes \eta_T \cdot 1)\\ &= \eta_{NT},\end{aligned}$$

such that the combined regressor matrix $[\mathbf{X}, \mathbf{A}]$ does not have full column rank.

3.2.2.1 Fixed Effect Model

Definition 5 (Fixed Effect)**.** *If the individual intercepts, represented as ϕ_i, are viewed as fixed constants that may be randomly correlated with the regression vectors x_{it}, the resulting model is known as the fixed-effects model. This model is a particular instance of the linear model that is predicated on certain assumptions, such as $\mathbf{X}$ being non-arbitrary, $[\mathbf{X}, \mathbf{A}]$ having complete column rank, and $E(\varepsilon) = \mathbf{0}$, and $\mathbf{V}(\varepsilon) = \sigma^2\mathbf{I}_{NT}$.*

In conventional OLS "residual regression" computations, the standard least-squares estimator for the sub-vector γ of regression coefficients is obtained through a specific approach,

$$\hat{\gamma}_{FE} = \left(\overline{\mathrm{X}}'\overline{\mathrm{X}}\right)^{-1}\overline{\mathrm{X}}'\overline{\mathrm{y}}$$

where $\overline{\mathbf{y}} \equiv \left(\mathbf{I}_{NK} - \mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'\right)\mathbf{y}$ are the residuals of a regression. Because of the distinctive structure of the matrix $\mathbf{A}$, it is simple to show that the sub vector of $\overline{\mathbf{y}}$ corresponding to each i may indeed be expressed as,

$$\begin{aligned}\overline{\mathbf{y}}_i &= \left(\mathbf{I}_T - \eta_T\left(\eta_T'\eta_T\right)^{-1}\eta_T'\right)\mathbf{y}_i\\ &= \mathbf{y}_i - y_{i.}\eta_T\end{aligned}$$

where $y_i.$ is the average value of y_{it} over t

$$y_{i\cdot} \equiv \frac{\eta_T' \mathbf{y}_i}{\eta_T' \eta_T} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$$

from the original structural equation,

$$y_{it} = x_{it}'\gamma + \phi_i + \varepsilon_{it} \tag{3.4}$$

hence,

$$y_{i.} = x_i'\gamma + \phi_i + \varepsilon_{i.} \tag{3.5}$$

and subtracting 3.4 from 3.5 gives,

$$\begin{aligned} y_{it} - y_{i\cdot} &= (\mathbf{x}_{it} - \mathbf{x}_i)' \boldsymbol{\gamma} + (\phi_i - \phi_i) + (\varepsilon_{it} - \varepsilon_{i.}) \\ &= (x_{it} - x_i)' \boldsymbol{\gamma} + (\varepsilon_{it} - \varepsilon_{i\cdot}) \end{aligned}$$

By subtracting the respective time-averages of y_{it} and x_{it} from their individual values, the constant effect ϕ_i is eliminated from the structural equation. This is similar to the way that intercept factor is removed from a conventional regression model by considering deviations from means. Another approach to estimate the fixed effect, $\hat{\boldsymbol{\gamma}}_{FE}$, involves calculating the first-differences across time instead of variances from time-averages in order to remove the fixed effect ϕ_i:

$$\begin{aligned} \Delta y_{it} &\equiv y_{it} - y_{i,t-1} \\ &= \Delta \mathrm{x}_{it}'\gamma + \Delta\phi_i + \Delta\varepsilon_{it} \\ &= \Delta \mathbf{x}_{it}'\gamma + \Delta\varepsilon_{it}, \end{aligned}$$

For $i = 1, \ldots, N$ and $t = 2, \ldots, T$, the outputs of the least-squares regression of Δy_{it} on $\Delta \mathrm{x}_{it}$ for balanced panels are equivalent to those of $y_{it} - y_{i\cdot}$ on $\mathrm{x}_{it} - \mathbf{x}_i$. Regardless of the case, it is apparent that regression coefficients on elements of $\mathbf{x}_{it}$ that remain constant over time ($\mathrm{x}_{it} \equiv \mathrm{x}_{is}$) are unidentifiable since that component of either $\Delta \mathrm{x}_{it}$ or $\mathrm{x}_{it} - \mathrm{x}_{i\cdot}$ will be zero. Only variations in the regressors over time for a given individual allow identification of the appropriate coefficient with respect to the individual-specific, time-invariant intercept ϕ_i. If the error terms ε_{it} are independently and identically distributed (i.i.d) and normally distributed with a mean of zero and variance σ^2, then the maximum likelihood (ML) estimate of $\hat{\boldsymbol{\gamma}}_{FE}$ is also the ML estimator of $\boldsymbol{\gamma}$, and the ML estimator of σ^2 is determined,

$$\hat{\sigma}^2_{ML} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - y_{i.}\left(\mathrm{x}_{it} - \mathrm{x}_{i\cdot}\right)' \hat{\boldsymbol{\gamma}}_{FE}\right)^2.$$

This estimator is biased, and if $N \to \infty$ for a given T the ML estimator of σ^2 is also inconsistent, with

$$\hat{\sigma}^2_{ML} \xrightarrow{p} \frac{(T-1)}{T}\sigma^2$$

This discrepancy of the ML estimator of σ^2 is a typical illustration of Neyman and Scott's [24] pernicious "*incidental parameters issue*." Since as N rises, the number N of intercept terms ϕ_i increases to infinity, as does the related ML estimator,

$$\begin{aligned}\hat{\phi}_i \equiv & y_{i\cdot} - \mathbf{x}_{i\cdot}' \hat{\gamma}_{FE} \\ & \xrightarrow{p} \phi_i + \varepsilon_i\end{aligned}$$

is inconsistent for ϕ_i, and this inconsistency leads to the ML estimator $\hat{\sigma}^2_{ML}$ being unreliable. Thankfully, this does not result inconsistency of $\hat{\beta}_{FE}$, which is an accurate and consistent estimator.

$$s^2_{ML} = \frac{1}{N(T-1)-K} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - y_{i\cdot} - (\mathbf{x}_{it} - \mathbf{x}_{i\cdot})' \hat{\beta}_{FE} \right)^2$$

of σ^2 is easily accessible. It should be noted that the term "fixed effect" does not imply that the regressors x_{it} or intercept terms ϕ_i must be considered as non-random or fixed in a quantitative sense. Instead, they can be seen as random and jointly distributed as long as the criteria on ε_{it} are satisfied based on the realizations of X and ε. A "fixed effect" model is distinguished by the lack of imposed structure on the relationship between ϕ_i and x_{it}.

Regressors in fixed-effects models may be linked with individual-level effects, making it necessary to eliminate or account for the fixed effects in order to reliably estimate regression parameters. For our study we'll looking at the fixed effect models specifically as with large T and small N in case of monthly there is likely to be little difference, so FE is preferable as it is easier to compute which will be clear after the definition of both of these.

3.2.2.2 Random Effect Model

One drawback of the fixed-effect model is that it cannot identify the γ coefficients that relate to the regressors that remain constant for an individual over time. To obtain such coefficients, more stringent conditions must be imposed on the connection between the individual-specific intercept ϕ_i and the regressors x_{it}.

Definition 6 (Random Effect)**.** *This model relies on a strong assumption that the intercept ϕ_i is an unrelated r.v to x_{it} and ε_{it}, in order to simplify the relationship between them,*

$$E(\phi_i) = \phi, \operatorname{Var}(\phi_i) = \sigma^2_\phi, \text{ and } \operatorname{Cov}(\phi_i, \varepsilon_{it}) = 0$$

all assumed independent of x_{it}. Relabeling the variance of s_{it} as $\operatorname{Var}(s_{it}) \equiv \sigma^2_e$, the original panel data model can be rewritten as,

$$y_{it} = x_{it}'\gamma + \phi + v_{it} \tag{3.6}$$

where

$$v_{it} \equiv (\phi_i - \phi) + \varepsilon_{it}$$

has

$$\begin{aligned} E\left(v_{it}\right) &= 0, \\ \operatorname{Var}\left(v_{it}\right) &= \sigma_\phi^2 + \sigma_\varepsilon^2, \\ \operatorname{Cov}\left(v_{it}, u_{in}\right) &= \sigma_\phi^2, \ \ \textit{and} \\ \operatorname{Cov}\left(u_{it}, u_{js}\right) &= 0 \qquad \textit{if } i \neq j. \end{aligned}$$

Hence, the model may be expressed in matrix form as,

$$\mathbf{y} = \mathbf{X}\gamma + \phi\eta_{NT} + \mathbf{v} \tag{3.7}$$

where

$$\begin{aligned} E(\mathbf{v}) &= \mathbf{0} \\ \mathbf{V}(\mathbf{v}) &= \sigma_\phi^2 \mathbf{A}\mathbf{A}' + \sigma_\varepsilon^2 \mathbf{I}_{NT} \\ &\equiv \sigma_\varepsilon^2 \mathbf{\Omega} \end{aligned}$$

with

$$\begin{aligned} \Omega &\equiv \mathbf{I}_{NT} + \frac{\sigma_\phi^2}{\sigma_\varepsilon^2}\mathbf{A}\mathbf{A}' \\ &= \mathbf{I}_{NT} + \frac{\sigma_\phi^2}{\sigma_\varepsilon^2}\left(\mathbf{I}_N \otimes \eta_T \eta_T'\right) \end{aligned}$$

In the unlikely event that the ratio $\pi \equiv \sigma_\phi^2/\sigma_\varepsilon^2$ were known - hence, Ω did not include any unknown parameters; Aitken's GLS estimator of γ and ϕ and would have the conventional form of,

$$\begin{pmatrix} \hat{\gamma}_{GLS} \\ \hat{\phi}_{GLS} \end{pmatrix} = \left(\mathbf{Z}'\Omega^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\Omega^{-1}\mathbf{y}$$

where $\mathbf{Z} \equiv [\mathbf{X}, \eta_{NT}]$. Similar to the fixed effects estimator, the GLS estimator $\hat{\beta}_{GLS}$ of the slope coefficients has many algebraically equivalent forms. y_{it}^* on x_{it}^* may be interpreted using the coefficients of a Least Squares regression where,

$$\begin{aligned} y_{it}^* &\equiv y_{it} - y_{i\cdot} + \omega \cdot \left(y_{i\cdot} - y_{\cdot\cdot}\right) \\ \mathbf{x}_{it}^* &\equiv \mathbf{x}_{it} - \mathbf{x}_{i\cdot} + \omega \cdot \left(\mathbf{x}_{i\cdot} - \mathbf{x}_{\cdot\cdot}\right) \end{aligned}$$

for

$$\begin{aligned} y_{\cdot\cdot} &\equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it} \\ \mathbf{x}_{\cdot\cdot} &\equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbf{x}_{it} \end{aligned}$$

and

$$\begin{aligned} \omega &\equiv \sqrt{\frac{\sigma_\varepsilon^2}{T\sigma_\varepsilon^2 + \sigma_\phi^2}} \\ &\equiv \sqrt{\frac{1}{T\pi + 1}} \end{aligned}$$

where as above $\pi \equiv \sigma_\phi^2/\sigma_\varepsilon^2$. As the variability of the random effect σ_ϕ^2 declines to zero for σ_ε^2 fixed, $\omega \to 1$, and the GLS estimator reduces to the usual LS regression of the deviations $y_{it} - y.$ on the associated deviations in the regressor variable from its mean value $\mathrm{x}_{it} - \mathrm{x}..$ in regressors. Applying this conclusion, the GLS estimator may be interpreted as a matrix-weighted average,

$$\hat{\boldsymbol{\gamma}}_{GLS} = \mathbf{D}\hat{\boldsymbol{\gamma}}_{FE} + (\mathbf{I} - \mathbf{D})\hat{\boldsymbol{\gamma}}_B$$

of the fixed effect estimator $\hat{\boldsymbol{\gamma}}_{FE}$ defined above and the "between estimator" (The form $\mathbf{D} \equiv \mathbf{D}(\theta)$ is taken from [25])

$$\hat{\boldsymbol{\gamma}}_B \equiv \left[\sum_{i=t}^{N} (\mathbf{x}_{i.} - \mathbf{x}_{..}) (\mathbf{x}_{i}, -\mathbf{x}_{..})'\right]^{-1} \sum_{i=t}^{N} (\mathbf{x}_{i}, -\mathbf{x}_{..}) (y_{i.} - y_{..})$$

of γ can be obtained from the LS regression of the time-averages $y_i.$ on x_i, and a constant. When the variances σ_γ^2 and σ_ε^2 are unknown, an estimator of $\theta \equiv \sigma_\gamma^2/\sigma_\varepsilon^2$ or $\omega \equiv (1 + T\pi)^{-1/2}$ is required to build a Feasible GLS estimator. The unbiased estimator s_{FE}^2 of σ_ε^2 based on the fixed-effect estimator $\hat{\gamma}_{FE}$ will indeed be consistent; for constant T, the variance estimate for the "between" estimator will also be accurate.

$$s_B^2 \equiv \frac{1}{N-K-1} \sum_{i=1}^{N} \left((y_{i.} - y_{..}) - (\mathrm{x}_{i.} - \mathrm{x}_{..})' \hat{\boldsymbol{\gamma}}_B\right)^2$$

will be unbiased and consistent for $\mathrm{Var}(\varepsilon_i) = \sigma_\phi^2 + (\sigma_\varepsilon^2/T)$. Hence

$$\hat{\omega} \equiv \frac{s_{FE}^2}{T s_B^2} \xrightarrow{p} \omega,$$

and may be substituted for ω when constructing a Feasible GLS estimator. Observe that the matching estimator of σ_ϕ^2

$$s_\phi^2 \equiv s_B^2 - \left(s_{FE}^2/T\right)$$

cannot be certain to be positive.

3.2.2.3 Robust Covariance Estimation

Errors showing heteroskedasticity or serial correlation are particularly problematic in the statistical analysis of panel data, as they are in other applications of generalized least squares (GLS) methods. Model

$$y_{it} = \mathbf{w}_{it}'\mu + v_{it} \tag{3.8}$$

where $\mathbf{w}_{it}'$ contains the row vector of regressors x_{it}' and the appropriate row of the matrices $\mathbf{A}$ and $\mathbf{R}$ of dummy variables for the fixed effect, we may assume that the errors v_{it} are uncorrelated across individuals i and have arbitrary variance-covariance patterns across time t for each individual i. Now say,

$$\begin{aligned} E\left(v_{it}\right) &= 0 \\ \operatorname{Cov}\left(v_{it}, v_{is}\right) &= \sigma_{i,ts}, \\ \operatorname{Cov}\left(v_{it}, v_{js}\right) &= 0 \quad \text{ if } i \neq j. \end{aligned}$$

arranging the observations in the conventional matrix format,

$$\mathrm{y} = \mathrm{W}\mu + \mathrm{v} \tag{3.9}$$

which readily follows,

$$\mathbf{V}(\mathbf{v}) \equiv \beta = \left[\begin{array}{cccc} \beta_1 & 0 & \ldots & 0 \\ 0 & \beta_2 & \ldots & \ldots \\ \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & \mathbf{0} & \beta_N \end{array}\right]$$

for

$$\begin{gathered} \beta_j \equiv \left[\sigma_{j,ts}\right] \\ (T \times T) \end{gathered}$$

It might be fair to employ the traditional LS estimator in the lack of a parametric form for β_j,

$$\hat{\mu}_{LS} = \left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}'\mathbf{y}$$

of the μ coefficients (i.e., the slope coefficients γ and any individual or time fixed effects), should be stable and follow a normal distribution under certain conditions on the $\mathbf{v}$, allowing for reliable statistical inference. The goal of this GLS application is the same as that of previous ones: to get a reliable estimate for,

$$\operatorname{plim} \frac{1}{N}\mathbf{V}\left(\hat{\boldsymbol{\mu}}_{LS}\right) = \operatorname{plim}\left(\frac{1}{N}\mathbf{W}'\mathbf{W}\right)^{-1}\left(\frac{1}{N}\mathbf{W}'\beta\mathbf{W}\right)\left(\frac{1}{N}\mathbf{W}'\mathbf{W}\right)^{-1},$$

the asymptotic covariance matrix of the LS estimator. The middle matrix is the most difficult because,

$$\frac{1}{N}\mathbf{W}'\beta\mathbf{W} \equiv \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{n=1}^{T}\sigma_{i,tn}\mathbf{w}_{it}\mathbf{w}_{in}',$$

using the same logic that lead to the Huber-Eicker-White robust covariance matrix estimator gives,

$$\frac{1}{N}\mathbf{W}'\hat{\beta}\mathbf{W} \equiv \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{n=1}^{T}\hat{v}_{it}\hat{v}_{in}\mathbf{w}_{it}\mathbf{w}_{is}'$$

as a consistent estimator of the middle matrix of $plim\frac{1}{N}\mathbf{V}\left(\hat{\boldsymbol{\mu}}_{LS}\right)$, where $\hat{v}_{it}$ are the LS residuals

$$\hat{v}_{it} \equiv y_{it} - \mathbf{w}_{it}'\hat{\mu}_{LS}$$

Applying the same constraints as the Huber-Eicker-White which is resistant to heteroskedasticity covariance estimate, this estimator will also be reliable as $N \to \infty$

. This resilient covariance matrix estimator conveniently extends to clustered data problems where the number of “time periods” or “group members” T_i is dependent on the group index i the previous formulations are generalized by swapping T for T_i across.

3.2.2.4 Nickell Bias

Individual-specific effects are treated as fixed parameters in a fixed effects model. This method may correct for unobserved time-invariant heterogeneity that changes among individuals, however employing a lagged dependent variable as a regressor might result in bias. The bias caused by employing fixed effects with a lagged dependent variable is referred to as the Nickell bias [26], named after the economist who originally found it, Stephen Nickell. The Nickell bias occurs when the lagged dependent variable is linked with the individual-specific effects, which may occur when the lagged dependent variable is endogenous or serially correlated. And because we have previously demonstrated that the CCI has an autoregressive character, we are compelled to use the dynamic panel models.

In a dynamic panel data model, the time demeaning operation of fixed effects,

$$y_{it} = \alpha_i + \beta y_{it-1} + \epsilon_{it}$$

which takes us to a modified regression model

$$y_{it} - y_{i.} = \beta \left(y_{it-1} - y_{i\cdot-1} \right) + \left(\epsilon_{it} - \epsilon_{i.} \right)$$

where dots indicate time averages. Here, error terms $(\epsilon_{it} - \epsilon_i .)$ and regressors $(y_{it-1} - y_{i\cdot-1})$ are correlated even as $N \to \infty$, where N is the number of units in the panel. This can be shown formally, but essentially follows from the observation that $y_i.$ contains future y_{it} which are generated by past y_{it} which, in turn, are generated by past ϵ_{it} which are contained in $\epsilon_{i.}$ Hence, even as $N \to \infty$, the FE estimator will not consistently estimate β. Nickell Shows the inconsistency to be approximately equal to,

$$-\tfrac{1+\beta}{T-1} \tag{3.10}$$

Now for $T \to \infty$ this bias tends to zero and because our monthly scenario has T is quite large, we may use any traditional FE estimators with a lagged dependent variable as regressors. In this instance, we used the DK [27] estimator, which permits robust estimations that meet our needs.

For annual cases in which T is quite small and things become borderline. Here we can expect the Nickell bias is likely to persist. To address these situations, we employed an estimator that explicitly corrects for the bias introduced by Kripfganz et al. [28]. In the next two parts, we provided a quick overview of the concepts behind the two estimators we would use in our investigation.

3.2.2.5 Panel models with Driscoll and Kraay standard errors

This estimator was used particularly for monthly cases for which we have data from 2006 to 2020 and which are, in essence, several time series. I performed the study using the Stata program -xtscc, which computes Driscoll-Kraay standard errors and permits any correlation across cross-section and general serial correlation throughout time, as well as robust inference, which we describe in the previous part. Additionally, when T is big and N is small, we may incorporate the lag dependent variable directly in the panel configuration, since the Nickell bias introduced by utilizing fixed effects with a lag dependent variable is negligible when T is large as shown in the last section. As described in section 3.2.2.4, it is necessary to use a dynamic panel model as our dependent variable CCI having autoretrogressive structure. Duan et al.[8] used traditional Arellano-Bond GMM estimator which is consistent with when $N > T$ but in our case $T > N$ in both cases where in monthly case we have a large T and that's why as shown in the last chapter the Nickell Bias for this case will tends to zero and in case of annual data, we have small T we need to use some bias corrected dynamic models to account for the nickell bias.

Our discussion will be limited to linear regression, however Driscoll and Kraay's [29] covariance matrix estimator may be used with any kind of model.

Driscoll and Kraay standard errors for pooled OLS estimation

Following Driscoll and Kraay [29] and Daniel Hoechle [27] we focus on the below linear model,

$$y_{it} = \mathbf{x}_{it}'\mu + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T \tag{3.11}$$

here the dependent variable y_{it} is a scalar, $\mathbf{x}_{it}$ is a $(K+1) \times 1$ vector of independent variables which starts with 1, and μ is a $(K+1) \times 1$ vector whose coefficients is not known. Here i represents cross-sectional individuals and t represents time. Commonly, observations are arranged as follows:

$$\mathbf{y} = [\ y_{1t_{11}} \quad \dots \quad y_{1T_1} \quad y_{2t_{21}} \quad \dots \quad y_{NT_N}\ ]'$$

and

$$\mathbf{X} = [\ \mathbf{x}_{1t_{11}} \quad \dots \quad \mathbf{x}_{1T_1} \quad \mathbf{x}_{2t_{21}} \quad \dots \quad \mathbf{x}_{NT_N}\ ]'$$

This approach enables an unbalanced panel, since only a subset for each individual i is picked and $t_{i1}, \dots, T_i$ with $1 \leq t_{i1} \leq T_i \leq T$ of all T observations are potentially accessible. For all values of s, t, it is considered that now the regressors $\mathbf{x}_{it}$ have no association with the scalar disturbance component ε_{is} (suggesting robust exogeneity, which is necessary for our situation). However, it is permissible for the disturbances to be autocorrelated, heteroscedastic, and cross-sectionally dependent. According to these assumptions, μ may be reliably predicted via OLS regression, which produces,

$$\widehat{\mu} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y}$$

As the square roots of the diagonal components of the asymptotic (robust) covariance matrix, the DK standard errors for the coefficient estimations are subsequently computed as,

$$V(\widehat{\mu}) = (\mathbf{X}'\mathbf{X})^{-1}\,\widehat{\mathbf{S}}_T\,(\mathbf{X}'\mathbf{X})^{-1}$$

where $\widehat{\mathbf{S}}_T$ is from Newey and West [30]:

$$\widehat{\mathbf{S}}_T = \widehat{\mathbf{\Omega}}_0 + \textstyle\sum_{j=1}^{m(T)} w(j,m)\left[\widehat{\mathbf{\Omega}}_j + \widehat{\mathbf{\Omega}}'_j\right] \tag{3.12}$$

In the (3.12), $m(T)$ denotes the lag length up to which the residuals may be autocorrelated and the modified Bartlett weights,

$$w(j,m) = 1 - j/\{m(T)+1\}$$

We need to make sure semi-definiteness of $\widehat{\mathbf{S}}_T$ is positive and smoothen the sample autocovariance function such that higher order lags get less weight. The $(K+1)\times(K+1)$ matrix $\widehat{\Omega}_j$ is defined as,

$$\widehat{\Omega}_j = \textstyle\sum_{t=j+1}^{T} \mathbf{h}_t(\widehat{\mu})\mathbf{h}_{t-j}(\widehat{\mu})' \quad \text{with} \quad \mathbf{h}_t(\widehat{\mu}) = \sum_{i=1}^{N(t)} \mathbf{h}_{it}(\widehat{\mu}) \tag{3.13}$$

Note that in (3.13), the sum of the time t moment conditions $\mathbf{h}_{it}(\widehat{\mu})$ ranges from 1 to $N(t)$, where N is permitted to vary with t. This little modification to the original estimator developed by Driscoll and Kraay [29] is sufficient to make their estimate suitable for use with imbalanced panels. Individual orthogonality criteria $\mathbf{h}_{it}(\widehat{\mu})$ in (3.13) are the $(K+1)\times 1$ dimensional moment conditions of the linear regression model in the case of pooled OLS estimation, i.e.,

$$\mathbf{h}_{it}(\widehat{\mu}) = \mathbf{x}_{it}\widehat{\varepsilon}_{it} = \mathbf{x}_{it}\left(y_{it} - \mathbf{x}'_{it}\widehat{\mu}\right)$$

The covariance matrix estimator of DK is equivalent to the heteroscedasticity and autocorrelation consistent covariance matrix estimator of Newey and West applied to the time series of cross-sectional averages of the $\mathbf{h}_{it}(\widehat{\gamma})$. Using cross-sectional averages, this method estimates standard errors that are constant regardless of the cross-sectional size of the panel N. Driscoll and Kraay [27] demonstrate that this consistency finding remains true even when $N \to \infty$. In addition, using this method to estimate the covariance matrix gives standard errors that are resilient to extremely broad kinds of cross-sectional and temporal dependence.

Fixed-effects regression with Driscoll and Kraay standard errors

For our study, we are primarily interested in estimating fixed-effects (within) regression models with Driscoll and Kraay standard errors; nonetheless, knowing OLS with

DK standard errors is essential for laying the groundwork for understanding Fixed effect regression with DK standard error. In two phases, the relevant fixed-effects estimator is built. In the first stage, all model variables $z_{it} \in \{y_{it}, \mathbf{x}_{it}\}$ are within transformed as follows:

$$\widetilde{z}_{it} = z_{it} - \bar{z}_i + \bar{\bar{z}}, \quad \text{where} \quad \bar{z}_i = T_i^{-1} \sum_{t=t_{i1}}^{T_i} z_{it} \quad \text{and} \quad \bar{\bar{z}} = \left(\sum T_i\right)^{-1} \sum_i \sum_t z_{it}$$

Knowing that the within-estimator relates to the OLS estimator of,

$$\widetilde{y}_{it} = \widetilde{\mathbf{x}}_{it}' \mu + \tilde{\varepsilon}_{it} \tag{3.14}$$

next phase involves estimating the modified regression model in (3.14), as described in the preceding section, using pooled OLS estimation with Driscoll and Kraay standard errors.

3.2.2.6 Bias-corrected estimation of linear dynamic panel data models

As noted in section 3.2.2.4, our design of the problem statement for annual data does not meet the $N > T$ (and T is fairly tiny in this instance) criterion for Arellano and Bond estimator to operate consistently hence we can't follow the methodology in [8] of using Arellano and Bond GMM [31] estimator. As we discussed already classical FE or RE estimator with a presence of lagged dependent term will be biased in the presence of unobserved group-specific heterogeneity [26]. The only way to reduce this bias when Time horizon is quite large as like our monthly case. Thus, these estimators may not be particularly trustworthy. In any event, you should not use the two-step version of these estimators, since finding the best weighting matrix for the second step with such a tiny N is futile. Given extremely persistent macro data, one can anticipate that the Nickell bias will persist. If all factors other than the lagged dependent variable are absolutely exogenous, an estimate that explicitly corrects for bias may be preferable to a GMM estimator.

Kripfganz et al. [28] introduced a novel bias-corrected (BC) estimator with desirable finite-sample features and developed a programme *xtdpdbc* in Stata. It has certain desirable qualities, which we will attempt to describe shortly. Firstly, as the analytical form of the bias is known, the BC estimator may immediately rectify it at the source by modifying the momentary circumstances. The slight variance between FE and RE estimators is maintained. We can accommodate higher order autoregressive models using this. This BC estimator follows a known asymptotic distribution, and it is an estimate based on the technique of moments. This allows for the determination of standard errors. There is a way to make standard errors resistant to cross-sectional dependency. This BC estimator's FE counterpart, the adjusted profile likelihood estimator of Dhaene and Jochmans [32], is mathematically equivalent. Since there is just one lag of the predictor variables, it is also equivalent to the iterative BC estimator proposed by Bun and Carree [33]. Unlike Kiviet's [34] bias estimate, which need a prior consistent estimator, this BC estimator may be used without one.

$$Y_{it} = \sum_{k=1}^{m} \alpha_k Y_{i,t-k} + \mathbf{X}_{it}'\theta + \underbrace{\phi_i + v_{it}}_{=e_{it}}, \quad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T \tag{3.15}$$

Some basic model assumptions are as follows,

1. Higher-order autoregressive model with p lags of the dependent variable and minimum initial observations regularity criteria.
2. Purely exogenous regressors X_{it} with regard to the idiosyncratic error term: $E\left[\mathrm{X}_{it} v_{is}\right] = \mathbf{0}$ for all t and s.
3. Unobserved group-specific "fixed effects", $E\left[\mathrm{X}_{it}\phi_i\right] \neq 0$, or "random effects", $E\left[\mathrm{X}_{it}\phi_i\right] = 0$.
4. There will be no serially uncorrelated idiosyncratic errors as $E\left[v_{it} v_{is}\right] = 0$ for all t and s, but it could be heteroskedastic, $E\left[v_{it}^2\right] = \sigma_i^2$.

Let us now focus on the Bias-corrected method of moments estimator. For demonstration purpose let consider $m = 1$ and define $\Omega = (\alpha_1, \theta')'$. We'll be focusing FE estimator for our case, This FE-BC estimator solves for,

$$\hat{\mathbf{\Omega}} = \arg\min_{\mathbf{\Omega}} \left(\sum_{i=1}^{N} \mathbf{M}_i(\mathbf{\Omega})\right)' \left(\sum_{i=1}^{N} \mathbf{M}_i(\mathbf{\Omega})\right)$$

with moment functions

$$\mathbf{M}_i(\mathbf{\Omega}) = \frac{1}{T}\sum_{t=1}^{T}\left[\begin{pmatrix} Y_{i,t-1} - \bar{Y}_{-1,i} \\ \mathbf{X}_{it} - \bar{X}_i \end{pmatrix} - \begin{pmatrix} \frac{T}{T-1} b\left(\alpha_1\right)\left(e_{it} - \bar{e}_i\right) \\ \mathbf{0} \end{pmatrix}\right] e_{it}$$

where $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{X}_{it}$ and $b\left(\alpha_1\right) = -\frac{1}{T^2}\sum_{t=0}^{T-2}\sum_{s=0}^{t} \alpha_1^s$,

such that $E\left[\mathbf{M}_i(\mathbf{\Omega})\right] = \mathbf{0}$.

Though we'll be applying FE-BC estimator for our case, but let us briefly familiarise with the other side of the coin of FE-BC estimator which is RE-BC estimator and let's just glance over a bit.

The RE-BC estimator solves for,

$$\hat{\mathbf{\Omega}}^{(k)} = \arg\min_{\mathbf{\Omega}} \left(\sum_{i=1}^{N} \mathbf{M}_i(\mathbf{\Omega})\right)' \mathbf{F} \left(\sum_{i=1}^{N} \mathbf{M}_i(\mathbf{\Omega})\right)$$

with moment functions

$$\mathbf{M}_i(\mathbf{\Omega}) = \frac{1}{T}\sum_{t=1}^{T}[\underbrace{\begin{pmatrix} Y_{i,t-1} - \bar{Y}_{-1,i} \\ \mathbf{X}_{it} - \bar{X}_i \\ \mathrm{X}_{it} \end{pmatrix}}_{=\mathrm{z}_{it}} - \begin{pmatrix} \frac{T}{T-1} b\left(\alpha_1\right)\left(e_{it} - \bar{e}_i\right) \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}]e_{it}$$

and weighting matrix $\mathbf{F} = \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{z}_{it}\mathbf{z}_{it}'\right)^{-1}$ for the 1-step estimator $\hat{\mathbf{\Omega}}^{(1)}$, or $\mathbf{F} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{M}_i\left(\hat{\mathbf{\Omega}}^{(1)}\right)\mathbf{M}_i\left(\hat{\mathbf{\Omega}}^{(1)}\right)'\right)^{-1}$ for the 2-step estimator $\hat{\mathbf{\Omega}}^{(2)}$.

This section was intended to provide an understanding of the concept behind the use of BC estimators using Stata's *-xtdpdbc* command. For further specifics, readers might turn to Kripfganz et al. [28].

3.2.3 Vector Autoregression

A stationary time-series variable a_t may typically be treated as dependent on its own lagged values in a univariate autoregression,

$$a_t = \mu_0 + \mu_1 a_{t-1} + \mu_2 a_{t-2} + \cdots + \mu_k a_{t-k} + \varepsilon_t$$

The vector autoregression (VAR) model is a natural extension of the autoregressive model for the analysis of multiple time series; in this model, each variable in the vector is depicted as depending on its own lag and on the lags of all other variables in the vector. A two-variable VAR with one lag appears as follows,

$$\begin{aligned} a_t &= \mu_0 + \mu_1 a_{t-1} + \mu_2 b_{t-1} + \varepsilon_{1t} \\ b_t &= \rho_0 + \rho_1 a_{t-1} + \rho_2 b_{t-1} + \varepsilon_{2t} \end{aligned}$$

Models of this kind are used by applied macroeconomists to characterize macroeconomic data, draw causal inference, and give policy recommendations. There are two primary model-selection choices that must be taken while developing a VAR model. The research topic and theoretical framework inform these choices. Those are as follows,

1. Choosing which variables to include into the VAR model is the first major choice to be made. Researchers often choose variables that are both theoretically grounded and relevant to their research subject.

2. The selection of the lag duration is the second option to be made. It's possible to use a heuristic model, such include a year of delays, or more rigorous lag-length selection criteria to make this choice. The number of previous periods utilized to make predictions about the present period is determined by the lag duration. The VAR model's accuracy and its ability to predict the future are very sensitive to the lag duration that is chosen.

After settling on the variables and lag time, the next step is to estimate the VAR model's parameters. Methods of estimation might vary according on the nature of the study's inquiry and the information at hand. Post-estimation checks on the model fit may be conducted with the help of the model's parameters.

Estimation and Stability

Let's begin with a model we'll be using in our case with annual data of Credit Cycle, Sector Risk (CVaR) and Repo Rate. The order of the variables in the VAR model is quite important which we've discussed later in this section.

$$\begin{bmatrix} \text{Reporate}_t \\ \Delta\text{CCI}_t \\ \boldsymbol{\Delta}\textbf{CVaR}_t \end{bmatrix} = \mathbf{b_0} + \mathbf{B_1} \begin{bmatrix} \text{Reporate}_{t-1} \\ \Delta\text{CCI}_{t-1} \\ \boldsymbol{\Delta}\textbf{CVaR}_{t-1} \end{bmatrix} + \cdots + \mathbf{B_k} \begin{bmatrix} \text{Reporate}_{t-k} \\ \Delta\text{CCI}_{t-k} \\ \boldsymbol{\Delta}\textbf{CVaR}_{t-k} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

$b_0$is a vector of intercept terms and each of B_1 to B_k is a 3×3 matrix of coefficients. The next step is to determine an appropriate lag duration. For all estimates and tests, we will use *Stata Software* and its **"var"** package. After establishing the model, it is necessary to determine whether or not the variables are stationary. It may be examined using any standard unit root test, such as the *Augmented Dickey–Fuller* or *Phillips–Perron tests*. In case any variables found non-stationary we can use standard first difference techniques or Log transform techniques. Then, diagnostics for proper lag-order selection must be tested. The *"varsoc"* command displays the outcomes of many lag-order selection experiments. The likelihood ratio test and Akaike's information requirements will be discussed, as well as the proposed lag order. Two objects must be estimated when variables and lag length are available: the coefficient matrices and the error term covariance matrix. Using least squares, coefficients may be calculated equation by equation. With a sample covariance matrix of residuals, one may derive the covariance matrix of errors. var command does both jobs and presents the coefficients by default; an estimate of the error terms' covariance may be determined from the output $e(\sigma)$. The results of var are organized by equation, where each "equation" corresponds to its dependent variable; hence, there are Repo Rate equations, CCI equations, and CVaR equations. $e(\sigma)$ represents the covariance matrix of the calculated VAR residuals, which can be accessed easily from Stata's output. Note that residuals are connected across equations. Not including the constant terms, a VAR with n variables and k lags will have kn^2 coefficients; so it might get tricky to interpret all the results from the output. So, it is common practice to exclude the var output table and instead provide more meaningful post-estimation information. Granger causality tests and impulse–response functions are two prominent post-estimation statistics that are used to evaluate VAR output.

Before proceeding with post-estimation statistics and theories, we must ensure that our VAR models are stable. Stata's *"varstable"* command will execute the required stability assessment for the var model with the optimum lag duration. And last, we must verify that there is no auto-correlation at the given lag length using the Lagrange multiplier test and the *"varlmar"* function in Stata. Before moving on to the next part, we will briefly present the definitions and hypotheses of three test statistics: the Augmented Dickey Fuller Test, the Stability Check for the VAR model, and the Lagrange-multiplier test for the VAR model.

Definition 7 (Augmented Dickey–Fuller (ADF) test). *A non-stationary time series is characterized by its unit root. The ADF test is included in the unit root test. Technically speaking, a unit root in a time series is declared to exist when μ equals 1.*

$$X_t = \mu X_{t-1} + \theta A_e + \epsilon$$

when X_t is value of the time series at time t and A_e is an exogenous variable. A non-stationary time series is indicated by the existence of a unit root.

The Dickey-Fuller test is a unit root test that examines the null hypothesis that μ equals to 1 for the following model equation. The coefficient of the first lag on x is denoted by μ.

$$x_t = c + \theta t + \mu x_{t-1} + \phi \Delta X_{t-1} + \varepsilon_t$$

with the hypothesis,

$$H_0 : \mu equals\ to 1.$$

The aforementioned equation is utilized in the Augmented Dickey Fuller Test, which is a popular Unit Root test. The ADF test incorporates high-order regressive processes into the original Dickey Fuller test equation.

$$x_t = c + \theta t + \mu x_{t-1} + \phi_1 \Delta X_{t-1} + \phi_2 \Delta X_{t-2}.. + \phi_m \Delta X_{t-m} + \varepsilon_t$$

The Dickey Fuller test does not affect the status quo, however.

Because $\mu = 1$ under the null hypothesis, the p-value must be less than the significance threshold (say, 0.05 or 0.01) before the null hypothesis may be rejected. The series must be stationary because of this.

Definition 8 (VAR Stability). *It is essential to have covariance stationary variables for inference after var. In x_t, variables with a constant covariance are ones whose first two moments do not change with time. The condition under which a variable x_t is considered covariance stationary is as follows:*

1. $\mathbf{E}[x_t]$ *is finite and independent of t.*
2. $Var[x_t]$ *is finite and independent of t.*
3. $Cov[x_t, x_m]$ *is a finite function of $|t - m|$ but not of t or m alone.*

Yet, a more stringent stability requirement is required for making sense of VAR models. If a VAR is stable, it may be inverted and its representation as a vector moving average has infinite order. The meanings of interpretations of impulse–response *functions and forecast-error variance decomposition are known if the VAR is stable. The calculated VAR is stable if the modulus of each eigenvalue of the matrix* $\mathbf{B} < 1$*, as shown by [35]. The companion matrix B is represented as below,*

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \dots & \mathbf{B}_{m-1} & \mathbf{B}_m \\ \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} \end{pmatrix}$$

and gets its eigenvalues using the eigenvalues of the matrix. $\sqrt{a^2+b^2}$ *is the modulus of the complex eigenvalue* $a+bi$*. As shown by [35], the VAR is stable if the modulus of each eigenvalue of* $\mathbf{B} < 1$.

Definition 9 (Lagrange Multiplier (LM) test for VAR)**.** *Most VAR post-estimation studies assume that disturbances are not autocorrelated. The LM test is conducted at lags* $i = 1, ..., p$*. The null hypothesis of the test is that there is no autocorrelation at lag* i *for each individual* i*. The formula for the LM test statistic at lag* i *is,*

$$\mathrm{LM}_m = (M - a - 0.5)\ln\left(\frac{|\widehat{\mathbf{\Delta}}|}{\left|\widetilde{\mathbf{\Delta}}_m\right|}\right)$$

where M *is the number of observations in the VAR;* Δ *is the maximum likelihood estimate of* Δ*, the variance–covariance matrix of the disturbances from the VAR; and* Δ_m *is the maximum likelihood estimate of* Δ *from the following augmented VAR.*

Evaluating the output of a VAR: Granger causality tests

In the econometric study of time series, causal relationships are often founded on the idea of predictability and are demonstrated by testing for Granger causality [36], [37]. Granger causality is popular because it is detected using reduced-form VAR models that may be applied to a group of potentially jointly driven variables. This is to be noted that same concept we can apply in panel settings too concept introduced by Lopez et al. [38]. Granger causality is a prediction-based statistical notion of causation. Each system variable is described as a linear mixture of its own past values and the past values of all other system variables in a VAR model. The Granger causality test is used to examine if the past values of one system variable aid to predict the future values of another system variable.

Definition 10 (Granger causality)**.** *Granger causality states that if a signal A "Granger-causes" a signal B, then previous values of A should include information that helps forecast B in addition to the information contained in past values of B alone.*

$$H_0 : A \text{ does not Granger Cause } B$$

$$H_1 : A \text{ Granger Causes } B$$

Rule of decision, if p value is,

- $< 0.05 = A$ Granger causes B at the 5% significance level.
- $> 0.05 = A$ does not Granger cause B at the 5% significance level.

The vargranger post-estimation command does several Granger causality tests.

Evaluating the output of a VAR: Impulse responses

The second set of data often used to assess a VAR consists of modeling system shocks and tracking their effects on endogenous variables. When the error components are connected across equations and the shocks are linked across equations. One method to approach this problem is to suppose that there are underlying structural shocks v_t which are not correlated while these stresses are linked to the reduced-form shocks according to the following relationship:

$$\begin{aligned} \epsilon_t &= \mathbf{B}\mathbf{v}_t \\ E\left(\mathbf{v_t}\mathbf{v_t'}\right) &= \mathbf{I} \end{aligned}$$

If we denote the covariance matrix of the error terms by $\mathbf{\Delta}$, then the $\mathbf{B}$ matrix is linked to $\mathbf{\Delta}$ via,

$$\begin{aligned} \mathbf{\Delta} &= E\left(\epsilon_t\epsilon_t'\right) \\ &= E\left(\mathbf{B}\mathbf{v_t}\mathbf{v_t'}\mathbf{B}'\right) \\ &= \mathbf{BE}\left(\mathbf{v_t}\mathbf{v_t'}\right)\mathbf{B}' \\ &= \mathbf{BB}' \end{aligned}$$

Because we have estimated $\hat{\mathbf{\Delta}}$, the problem is to construct $\hat{\mathbf{B}}$ from,

$$\hat{\mathbf{\Delta}} = \hat{\mathbf{B}}\hat{\mathbf{B}}' \tag{3.16}$$

Many B matrices fulfill (3.16). Assuming that $\mathbf{B}$ is lower-triangular is one technique to reduce the number of potential choices; $\mathbf{B}$ may then be uniquely identified using a Cholesky decomposition of $\mathbf{\Delta}$. The imposition of a lower-triangular structure on matrix $\mathbf{B}$ in a VAR model implies a certain sequence in the variables under consideration, and different arrangements of variables can lead to distinct $\mathbf{B}$ matrices. The economic implications of this structure are that a shock to a given equation will instantaneously affect variables in the lower sequences, while variables in higher sequences will remain unaffected. In this analysis, we will adhere to the same sequence as that in which the VAR equations are presented.

Impulse response function (IRF) analysis is affected by the ordering of the variables in a VAR model. This is due to the fact that the IRF demonstrates how a change in one variable influences the other variables in the system, and this is dependent on the ordering of the variables. There is no universally "best" ordering for the variables in a VAR model, since the optimum ordering is dependent on the research topic and the economic theory driving the connections among the variables. According to one alternative explanation, the Repo rate influences both the credit cycle index and sectoral risk, but the credit cycle index has a more immediate and direct impact on sectoral risk. In this situation, it may be reasonable to arrange the variables as follows: Repo rate, credit cycle index, sectoral risk (CVaR). *The Credit Channel Theory* [39] and *Minsky's Financial Instability theory* [12], which were briefly mentioned in Chapter 6, are the two major theories we utilized to organize our variables in this manner. These topics are discussed in further detail in Chapter 5.

With **B**, we may produce shocks which are not correlated between equations and determine their effect on the VAR variables. We may generate impulse–response functions using the irf create command and then graph the output using the irf graph command. On the impulse–response graph, each row represents an impulse and each column represents a response variable. The horizontal axis of each graph corresponds to the temporal units in which our VAR is computed, in this instance month/year. In this scenario, all variables are measured in percentage points, hence all vertical units are % changes. The horizontal axis represents the passage of time period. All findings are contingent on the **B** matrix or the parameter ordering in the VAR. Diverse orderings will yield various **B** matrices, resulting in diverse impulse responses. When we suspect that the variables in our VAR model may show contemporaneous effect, we should often apply Orthogonalized IRFs. Nevertheless, OIRFs may not be appropriate if we are only concerned about the short-term effects of shocks, as they typically emphasize the model's longer-term behavior as we are looking into 5years of impact we'll be adopting OIRF.

4

Classical Data and Alternative Data

In this chapter, we discuss how different data for the study were gathered and analysed, and built. This chapter also introduces a new study path in which the power of Natural Learning Processing will be used to explore Credit Risk.

4.1 Classical Data

In this section we'll be focusing on the how country and sector specific different econometric data has been collected, computed and the rationale behind it.

4.1.1 The Credit Cycle Index

The CRI PD is the likelihood that a debtor will be unable to fulfill its financial obligations. It is derived from the forward intensity model by Duan et al. [40] of the CRI, which includes 17 common and firm-specific risk variables. Prediction horizons, or "term structures", range from one month to sixty months. The PD and AS are accessible for over 800,000 publicly traded companies worldwide. The current actively traded companies, including a few from India, have their valuations updated daily. The Credit Cycle Index will be constructed from several sectors, with industry categorization based on the Bloomberg Industry Classification Standard (BICS) 2020. Using CRI PD, Aggregate PD and Actuarial Spread(AS) were calculated. The aggregate PD and AS indicate the creditworthiness of an area, economy, or industry. They are derived as the median value of each firm's PDs and ASes within the selected group. This will be used to construct our Sectoral Credit Cycle Index.

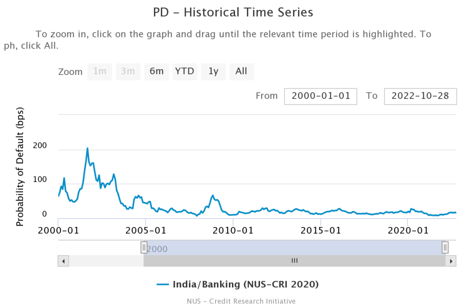


Figure 4.1: Monthly CCI of Indian Banking Sector

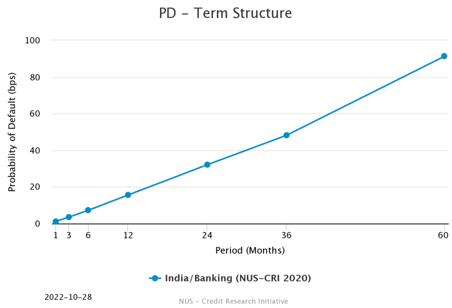


Figure 4.2: Term Structure

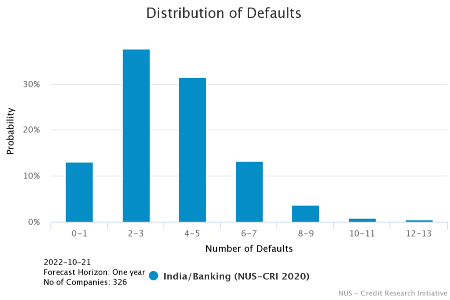


Figure 4.3: Distribution of Defaults

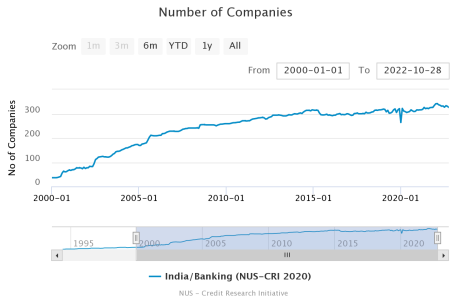


Figure 4.4: No. of Companies

Duan and Zhu showed [8] credit cycle index, denoted by $CCI^h_{i,t}$ reflects the aggregate credit risk of an entity i for a horizon of h months at time t. The entity might be a country, a region, or a particular sector. Hence, this new credit cycle index is more properly a family of indices constructed in the same manner as widely-referenced stock market indices, where component businesses and weightings are picked to fulfill a predefined objective. This index is unique in the world of credit risk management since it may concentrate on a specific or a range of prediction horizons. As an illustration, sector i's CCI is calculated as the median PD of all firms that domiciled in sector i. That is,

$$CCI^h_{i,t} := \underset{1\leq j\leq n, j\in \text{ domiciled in } i}{\text{median}} \left(PD^h_{j,t}\right) \tag{4.1}$$

where $PD^h_{j,t}$ is the h-month probability of default for firm j at time t. In case of picture 4.1 the graph denotes the CCI of Indian Banking sector with forecast horizon(h) is 12 months.

For our investigation, we set h to 12 months for all intents and purposes.

4.1.2 Sectoral Risk Indicator

Conditional Value at Risk (CVaR) was used to generate our Sectoral Risk component. CVaR is also known as Expected Shortfall in many literatures. In order to comprehend the notion of CVaR, we must first define **Value-at-Risk (VaR)**, which is closely connected. VaR is a common market risk metric that is even acknowledged by the *Basel Market Risk Framework*.

Following [41] we get the first element is a portfolio $\mathcal{H}$ of financial holdings each with a well-defined value at given moment t. At time t the value of the portfolio is v_t . The value of the portfolio one time period into the future, viewed from the present, is a random variable denoted by V_{t+1}. In risk management, one is interested in the (positive) tail of the so-called profit-and-loss (PnL) distribution function (*df*) of the random variable,

$$L_{t+1} = -\left(V_{t+1} - v_t\right) \tag{4.2}$$

Definition 11 (Value at Risk)**.** *Suppose L as given above and let* $0 < \beta < 1$*, then,*

$$\mathrm{VaR}_\beta(L) = \inf\left\{x \in \mathbb{R} : F_L(x) \geq \beta\right\} \tag{4.3}$$

[i.e., $\mathrm{VaR}_\beta(L)$ *is the* $100\beta\%$ *quantile of* F_L *].*

Definition 12 (Conditional Value at Risk)**.** *The conditional value at risk or expected shortfall of L at confidence level β is given by,*

$$\mathrm{CVaR}_\beta(L) = \frac{1}{1-\beta}\int_\beta^1 \mathrm{VaR}_u(L)\mathrm{d}u. \tag{4.4}$$

For F_L *continuous, the definition of ES is equivalent to,*

$$\mathrm{CVaR}_\beta(L) = \mathbb{E}\left[L \mid L > \mathrm{VaR}_\beta(L)\right] \tag{4.5}$$

VaR is not a **coherent measure** of risk and specifically doesn't follow the most desired property of **sub-additivity** for the purpose of our study as discussed in chapter 2. In field of Financial Economics a coherent risk measure is a function that satisfies properties of monotonicity, **sub-additivity**, and homogeneity.

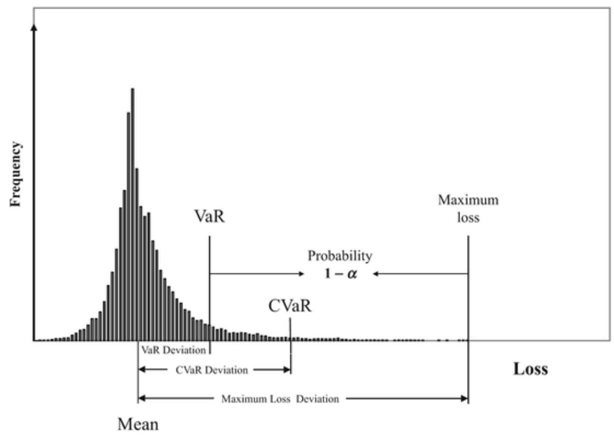


Figure 4.5: CVaR for general loss distribution

Before jumping to step by step processes of construction of the indicator let us familiarise with the three process of computation of CVaR.

1. Parametric Approach
2. Simulation Approach
3. Historical Approach

The simulation method (sometimes called the Monte Carlo simulation method) creates a random walk by drawing random samples from a particular distribution in order to anticipate future rates and revenues but combining all stocks from a sector, let alone all sectors, was physically impossible and would have been computationally costly. The parametric technique (also known as the variance-covariance approach) assumes normalcy which is not always true in real market scenario The CVaR was calculated using the Historical Method under this supposition.

4.1.2.1 Construction of the Indicator

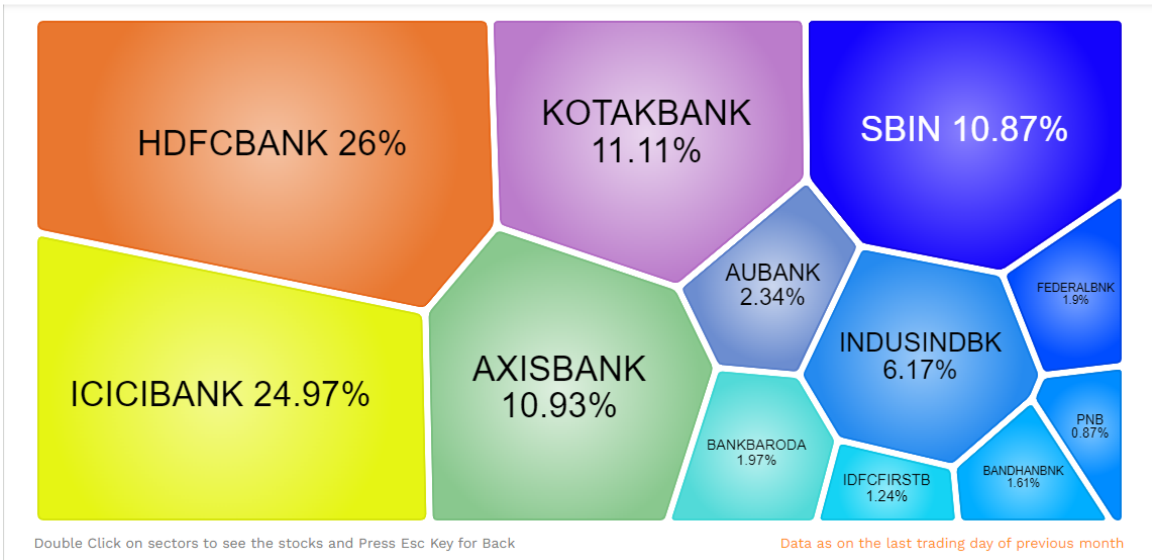


Figure 4.6: Sectoral Distribution of Nifty Bank

1. To begin, we acquired all Index Constituent data for a certain sector from the (see figure 4.6) BSE Sectoral Indices Website.
2. Weights of each stock that have been utilized to indicate how much % they contribute in each sector (as seen in figure 4.6), that data will also be used to calculate CVaR. The weight percentage of each stock may be determined by dividing the stock's free float market capitalization by the total free float market capitalization of all companies in the sector.
3. It is not expected that all the stocks are listed in the market at same time, therefore we used a dynamic method to account for the weight of each stock.

Let's understand this using an example, in the figure 4.6 we can see there's a stock name AU Small Finance bank which has been listed in the market on 14th July 2017. Hence, it was not included in our computation prior to its listing, and after it has been added, its market capitalization will be accounted for, along with the rest of the stocks.

4. Following that, daily closing price data for these stocks were obtained from the BSE between the years 2000 and 2020. Then, for each stock, the daily percent change in closing price or the daily return was determined. The formula used to create a daily price change portfolio is,

$$Portfolio_{dr} = \sum_{i=1}^{n} w_i v_i$$

where, w_i = weight of ith stock and v_i = daily return of ith stock and n = number of stocks in the sector .If we assemble every single day's weighted averaged percent change of close price data, we will get the daily return of closing price data for that sector.

5. Finally, we can aggregate the sector's daily percent change in closing price data on a monthly or annual basis and use the CVaR function to get the sector's monthly or yearly CVaR. For our purposes, we'll look at monthly/annual CVaR depending on cases.

6. In this manner, we may also develop a country-specific risk indicator for specified time horizon. For example, using the same reasoning as previously, we can collect the daily closing price data of the SP & BSE 500 index, which consists of 500 stocks and is a very excellent reflection of the Indian market, and then calculate monthly/annual CVaR using the same processes.

4.1.3 Macroprudential Index Data

We've mainly used two macroprudential index data one is monthly and another is yearly and each of those data has been compiled in little different format and we'll be introducing those in detail manner. First, we'll look into the MPI index for monthly frequency and then we'll be looking towards the yearly frequency.

For the monthly instance, we used the *IMF's iMaPP* database (International Monetary Fund). The IMF's iMaPP database is an Excel file including (1) dummy-type indications of tightening and easing actions of various macroprudential policy instruments, (2) a unique numerical indicator of regulatory constraints on the loan-to-value ratio, and (3) a description of each policy action. Alam et al. [42] compiled data from five existing databases and the IMF's Macroprudential Policy Survey to create the iMaPP database. National sources, IMF official publications, and the websites of the Bank for International Settlements (BIS) and the Financial Stability Board also contribute to the iMaPP database (FSB). This database has a great amount of information, however we are mostly interested in **MaPP indicators**. This is a policy activity indicator for each instrument. Each instance of tightening is categorized as a plus one, each instance of loosening as a minus one, and neutral or no action as a zero. **SUM 17** is the sum of the policy action indicator values for all 17 instruments. Comprehensive discussion of each policy instrument examined for the MPI Index is beyond the scope of this thesis; nevertheless, one may consult Alam et al. [42]. Limitations on loan-to-value ratios (LTV), counter-cyclical capital buffer (CCB), capital conservation buffer, Tax, Liquidity, Loan-to-Deposit, etc. are a few prominent devices. This database contains monthly MaPP statistics for several nations, including India. We collected these data from the iMaPP database and filtered them for India before using them in our research.

For annual cases, we derive the index from Cerutti et al. [14], whose data comes from the Global Macroprudential Policy Instruments (GMPI) research conducted by the IMF's Monetary and Capital Department. The survey covers the years 2000 to 2017 and a total of 18 separate policy instruments, of which 12 are included in the MPI. Readers may consult Cerutti et al. [14] for further information on the survey questionnaire (2017). The total MPI is calculated by summing the 12 binary readings for each instrument.

For both cases we choose lag of the MPI as the regressor variable since such precautionary strategies often need further time to take effect at least a year so in case of monthly it was taken around 12 lags and for yearly one/two lags. This also mitigates the notion of reverse causality, when policymakers decide to enact such policies after seeing credit growth.

4.1.4 Control Variables

We've used few control variables for panel data regression and we'll be discussing about those in this section. First, there is the GDP growth rate, which indicates the demand drive. For the yearly regression example, we used Bloomberg's data on the annual GDP growth rate. Normally, India discloses the quarterly GDP growth rate; hence, in the case of monthly data gaps, we employ interpolation to assure heterogeneity.

Our next variable is systemic risk (srisk) for India, developed by Brownlees et al. [43] from the Volatility and Risk Institute (V-Lab) at NYU Stern School of Business and this data is obtained from V-Lab's official website. Srisk measures the possible effect of a distressed financial institution on the whole financial system. It calculates the amount of capital a financial institution would need to raise in order to retain its present level of risk and avoid distress, taking into account the possible losses it may cause for the financial system. Srisk is based on a network model of the financial system that accounts for the interconnection of financial institutions and the spread of distress across the system. It is derived using publicly accessible data on the balance sheets and market positions of financial firms. The same network model of the financial system is used to determine country-specific srisk as is used to calculate global srisk. The model focuses on financial institutions and their relationships inside a particular country, as opposed to the global economic system as a whole. The intricate details of computation are out of the scope of this thesis one can refer to [43].

While macroprudential devices are typically employed in tandem with monetary tools, the policy rate is a crucial variable to regulate. In the instance of yearly regression, we used annual repo rate data from the RBI website. Similar to the GDP Growth Rate, RBI does not modify the repo rate every month (sometime even yearly too); hence, we interpolated the missing value for heterogeneity in the data.

The CPI is the Consumer Price Index. It represents the average change in prices paid by consumers for a basket of goods and services across time. The CPI is used to quantify inflation in an economy, and central banks and governments often use it to direct monetary policy and adjust economic indicators for inflation. The CPI is determined by calculating the weighted average of the prices of goods and services in a representative basket of commodities and services routinely bought by families, such as food, housing, transportation, and medical care. This data we only able to obtain for annual frequency from FRED.

Finally, Market Crash is a binary indicator that captures the presence of a market crash over a specified time frame. A market crash is a significant decline in India's two major stock indices, NSE and BSE. These crash dates are readily accessible from the Bloomberg, and we create the indicator so that it is either one when crash occurs else zero.

Similar to the MPI index, for both situations of regression we choose the ideal lag of all the control variables. The rationale for this is that it will take time for the impacts of altering each variable to manifest in CCI or market risk.

4.2 Alternate Data

As introduced the key motivation of this kind of study in chapter 2, I'll introduce the key steps we follow to retrieve this alternative data. Here we'll presenting a novel method for extracting media sentiments that are precisely focused on themes and entities at the article level, as opposed to conveying generic sentiments about firms following Xuan and Duan [19] developed by NLP-Team AIDF, NUS where I got an opportunity of interning for last 6month. Given the credit-centric nature of our research, [19] have established a novel way to identify an article's relevance to credit risk by weighing the views stated on all companies featured in the article. Despite the fact that the views stated about businesses are sentence based, we aggregate them at the article level.

4.2.1 Data Collection

We primarily web-crawled over seventy or more business journals from 1998 to the present, as well as in multiple languages. Even though the majority of collected news is in English and we concentrated primarily on collecting business-related news, the majority of collected news is in English. The frequency of news articles per year in the early years is significantly lower than it is currently.

Three NLP components are crucial to the extraction of sentiments from the business media corpus. After collecting sentiments, it is required to aggregate them across articles, business publications, and a changing time range. The goal is to provide alternative data time series of credit-focused, entity-specific media sentiments at the article level.

4.2.2 Named Entity Recognition (NER)

Identifying corporate and other names in the business media corpus is essential for this project. To build a general credit risk topic distribution that can be applied to all companies, we needed to exclude such names, and we did so. Afterwards, we match corporate names inserted back into articles with those in the NUS-CRI database, enabling us to aggregate media sentiments for publicly traded companies. We detect corporate and other names using the Named Entity Recognition (NER) approach of natural language processing. Coreference Resolution of Stanza by Qi et al.[44] is used to combine several forms of the same corporation (*HDFC* and *Housing Development Finance Corporation Limited*) and associate pronouns with the appropriate company. We construct similarity scores to map such ambiguous elements in articles to companies in the NUS-CRI database, and then use numerous rule-based pre-processing procedures to increase precision and efficiency.

4.2.3 Source Latent Dirichlet Allocation

Source-LDA, a variant of LDA, is used to extract the word distribution of the credit risk topic from a business media corpus. LDA is a bag-of-words approach introduced by Blei et al. [21] which represents each article as a combination of topics weighted by article-specific probabilities, where each topic is defined by a common word distribution across all articles. LDA finds a best representation of articles in a corpus by MCMC utilizing a collapsed Gibbs sampler. To generate the credit risk topic, a strong prior word distribution is applied. For our purpose, we only set two topics one **credit-risk** and another one is **greenness**. In upcoming future AIDF-NLP team is focusing on extracting more credit and market centric topics. The probability of an article on the credit risk topic is determined by the weight attributed by Source-LDA. To implement Source-LDA, company and other names are removed by NER, and only nouns and verbs are kept by POS tagging. Additionally, bigrams forming a specific meaning in discussions of credit risk are identified, and stopwords and sentiment words are removed. According to Wood [20], the LDA model converges after 1,000 iterations.

4.2.4 Sentiment analysis by TABSA-BERT

In recent years, the Devlin et al. [45] BERT (Bidirectional Encoder Representations from Transformers) model has gained prominence due to its performance in sentiment analysis. BERT is a pretrained language model that use a transformer architecture to facilitate text-based supervised learning tasks. Its adaptable architecture facilitates the execution of downstream NLP tasks, like as classification, by fine-tuning the parameters of a pre-trained model and an additional layer of a supervised learning neural network.

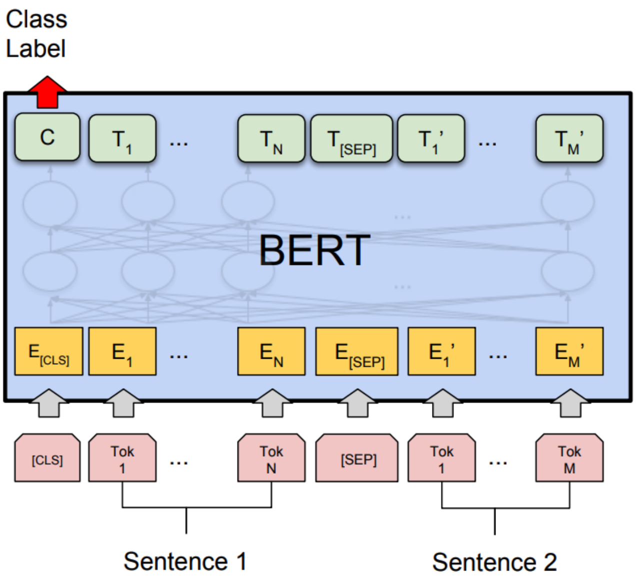


Figure 4.7: Sentence-pair fine-tuning with BERT

BERT accepts either a single sentence or a pair of sentences as input text. BERT-base and BERT-large encode tokens with vector lengths of 768 and 1024 bytes, respectively. We applied RoBERTa-large, an enhanced version of Liu et al. [46]. RoBERTa-encoding large's represents word, segment, and position embedding, with various encoding for the same word to reflect its individual meaning inside a phrase. Two distinct tags, [CLS] and [SEP], serve as sentence embedding and sentence separator, respectively.

A sentence-pair strategy has been employed for sentiment analysis, in which a pseudo-sentence is constructed for a target to couple with the real phrase. The method generates a pair of sentences for each target-specific sentiment. In our example, the targets are corporate names, but their job does not specify a particular element, since relevance to credit risk should be evaluated holistically at the article level. Figure 4.7 depicts the whole architecture and process of BERT for sentence-pair refinement.

The sentiment labels have an ordinal five-point scale from -2 to 2, they perform ordinal classification instead of categorical classification. RoBERTa-large turns a phrase pair into 1024 features, with "C" serving as the input vector for the add-on layer of the classification neural network. The ordinal classification network includes one hidden layer with four nodes, and after averaging the node values, the logistic function has been inverted and four cut-off values have been used in order to get the five probabilities corresponding to the five ordinal classes.

A sentence extracted from a newspaper:
Credit rating agency Fitch Ratings has compared two major Indian telecom companies, Bharti Airtel and Vodafone Idea, and concluded that while both face intense competition and regulatory hurdles, Vodafone Idea's high debt levels make it more vulnerable to financial stress than Bharti Airtel.

The Processed Sentence:
Credit rating agency Fitch Ratings has compared two major Indian telecom companies, Target1 and Target2, and concluded that while both face intense competition and regulatory hurdles, Target2 high debt levels make it more vulnerable to financial stress than Target1

Sentence-pair 1 tagged by a label of +1 (moderately positive)
1.1: The processed sentence
1.2: Target1

Sentence-pair 1 tagged by a label of -2 (strongly negative)
2.1: The processed sentence
2.2: Target2

An example of TABSA-BERT task

4.2.5 Sentiment aggregation

First article level sentiment has been aggregated to entity level sentiments. Each article's credit risk subject weight may be added to each article's sentiment to generate credit-focused, entity-specific sentiment at the article level. The credit-focused, entity-specific sentiment at the article level is calculated by multiplying the entity-specific sentiment obtained from TABSA-BERT at the article level by the article's credit risk subject weight obtained from Source-LDA. On any given day, a business publication may publish many articles on the same firm. We thus average the credit-focused, entity-specific sentiment ratings at the article level from the same media source. A further average with similar weighting is given to the commercial presses. There may not be news coverage on successive days, yet news may have lasting effects. Thus, we establish a seven-day moving window and calculate the seven-day moving average for credit-focused, entity-specific sentiments to provide daily time series for each media-covered organization. The absence of sentiment scores will not be replaced. In the seven-day time-frame, the moving average will represent the number of non-missing sentiment ratings.

Figure 4.8 below summarizes all the procedures from news collection to generating an entity-specific daily sentiment for Indian Case.

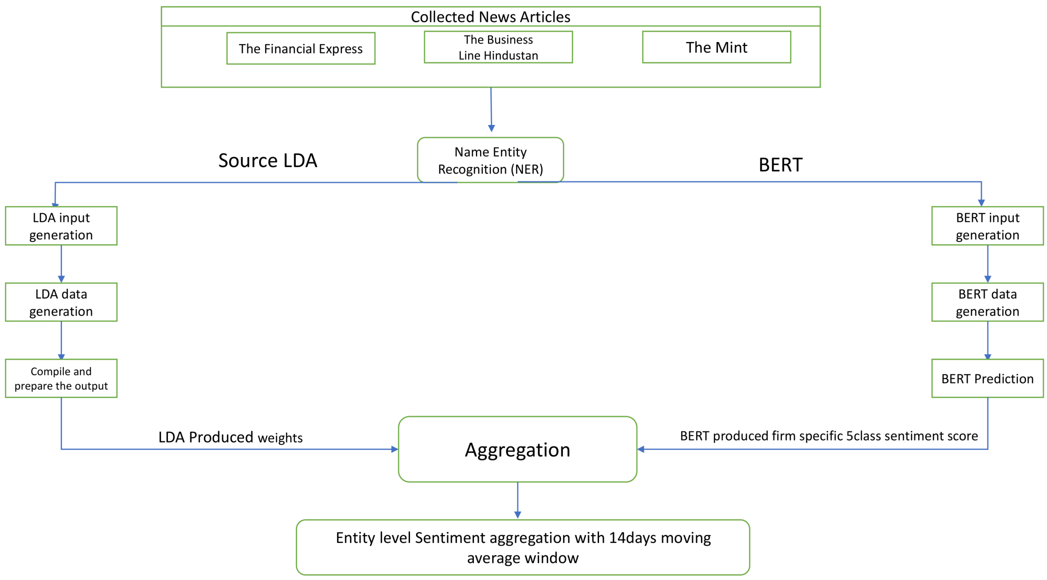


Figure 4.8: Full Methodology for Generating Entity-Specific Daily Credit Sentiment for Indian Case

5

Analysis and Discussion

This chapter describes the panel models and VAR models that were used, as well as the outcomes and conclusions. Using a handful of case examples, we also discuss the outcomes of our investigation into Credit Risk utilizing NLP tools. The discussion culminates in a consensus regarding the answer to the research objectives of the study.

5.1 Introduction

In this chapter, using the methodology described in chapter 3, we set up the panel models and VAR models and analyze the dataset described in chapter 4 for our first study vertical. A handful of case studies were also used to briefly convey the NLP study outcomes. The findings are presented, examined, and debated.

5.2 Panel Data

In this part, we will concentrate on the fourth research aim indicated in section [1.3] of the first chapter, which is the establishment of two distinct panel models for two distinct data frequencies (annual and monthly data, respectively) for eleven distinct Indian sectors. There are few variables that can be collected for yearly data, such as the CPI, however it was difficult to get for monthly instance. In monthly instance, we may anticipate better granularity and more comprehensive outcomes. In general, though, we discovered a comparable structure for both cases.

5.2.1 Annual Case

VARIABLES	**Observations**	**mean**	**std**	**min**	**max**
Dependent variable					
ΔCCI	187	-0.0001499	0.0021204	-0.011409	0.009705
Independent variables					
$\Delta CVaR$	187	0.0028954	0.0185447	-0.0607887	0.0545179
MPI	187	2.235294	1.115981	1	4
Repo Rate	187	6.936765	.948121	5.125	8.75
GDP Growth	187	6.767318	1.630778	3.0867	8.4976
Market Crash	187	0.1117647	0.4934742	0	1
Systematic Risk	187	34.02941	25.67347	2	75
CPI	187	100.2779	29.87907	62.01716	151.9678

Table 5.1: Summary Statistics for Annual Case. Time-period: 2000-2017

The descriptive statistics for all of our variables from 2000 to 2017 are shown in Table 5.1 The average change in the CCI over the last 12 months is -1.49 bps, indicating that the credit risks of the eleven distinct sectors have reduced throughout the study period. We see that the average sectoral risk is positive, indicating that the sectoral tail risk has grown throughout this time period. The macroprudential policy index, which spans from 1 to 4 with a mean of 2.23 and a standard deviation of 1.12, does not display significant variance. Throughout the years, the RBI has maintained a modest degree of monetary policy tightening, with considerable variance based on economic circumstances and policy goals. On an average 11% times in the sample period, we observed a market crash. Consumer Price Index (CPI) indicates that the CPI in the nation has had various variations throughout the years, with periods of considerable inflation as shown by the maximum value of 150, but has, on average, stayed reasonably steady. The systematic risk indicates that the country's financial system has been exposed to a moderate level of systemic risk over the years, with a moderate probability of widespread financial distress, such as a market crash or a credit crunch, caused by the failure of one or more financial institutions, and significant variation in the risk level over time. The baseline regression has been adopted from Cerutti et al. [14] which is,

$$Y_{i,t} = Y_{i,t-1}\alpha + MPI'_{i,t-1}\beta + GDP'_{i,t-1}\gamma + BankCrisis'_{i,t-1}\delta + Policy'_{i,t-1}\theta + \mu_i + \epsilon_{i,t} \tag{5.1}$$

Where $Y_{i,t}$ is the dependent variable here which is ΔCCI and rest are independent variables described in detail in the Chapter 4.
Based on the baseline regression we added few more explanatory variables such as Systematic Risk(SRISK), CPI, Market Crash etc (For detailed description see Chapter 4)

$$\begin{aligned} Y_{i,t} = Y_{i,t-1}\alpha + CVaR'_{i,t-1}\lambda + MPI'_{i,t-1}\beta + GDP'_{i,t-1}\gamma + Reporate'_{i,t-1}\theta \\ + SRISK'_{i,t-1}\delta + MarketCrash'_{i,t-1}\phi + CPI'_{i,t-1}\tau + \mu_i + \epsilon_{i,t} \end{aligned} \tag{5.2}$$

For such a dynamic panel model with fixed effects, degrading the equation or choosing the first difference would always cause an endogeneity problem, resulting in skewed estimates from widely used estimators. In addition, we have an instance where T narrowly surpasses N. To handle such a scenario, we used the BC estimator [28]. We consider CVaR and all control variables MPI, GDP growth, Market Crisis dummy variable, Repo Rate, and SRISK to be endogenous. Table 5.2 is a compilation of our panel regression results for the yearly case.

Table 5.2: The lag structure of Sectoral risk wrt to Credit Cycle

VARIABLES	(1) ΔCCI	(2) ΔCCI	(3) ΔCCI	(4) ΔCCI
L1.ΔCCI	-0.286***	-0.289***	-0.232**	-0.189***
	(0.0649)	(0.0574)	(0.0921)	(0.0551)
ΔCVaR	0.0100**			
	(0.00489)			
L1.MPI	1.32e-3***	1.16e-3***	1.03e-3***	0.00122***
	(4.01e-4)	(2.81e-4)	(2.48e-4)	(2.85e-4)
L1.GDPRate	0.000117**	6.71e-05	1.90e-05	-6.05e-05
	(5.24e-05)	(4.50e-05)	(7.89e-05)	(4.36e-05)
L1.RepoRate	5.80e-4***	5.69e-4*	1.45e-4	2.62e-4**
	(2.17e-4)	(3.36e-4)	(1.81e-4)	(1.31e-4)
L1.MarketCrash	-9.06e-4***	-7.05e-4	-4.8e-4**	-7.85e-4**
	(3.38e-4)	(4.66e-4)	(2.31e-4)	(3.09e-4)
L1.CPI	-1.01e-4	-7.85e-05	-1.23e-4	-0.1.73e-4*
	(8.07e-05)	(1.33e-4)	(1.28e-4)	(9.59e-05)
L1.SystematicRisk	-2.81e-05*	-2.29e-05*	-6.62e-06	-1.77e-05
	(1.70e-05)	(1.28e-05)	(1.48e-05)	(1.11e-05)
L1.ΔCVaR		3.96e-3		
		(1.78e-2)		
L2.ΔCVaR			0.0153	
			(9.95e-3)	
L3.ΔCVaR				-0.0176***
				(5.18e-3)
Constant	-5.99e-3	-5.58e-3	-2.58e-3	-2.27e-3
	(2.04e-3)	(2.98e-3)	(2.13e-3)	(1.61e-3)
Observations	176	176	165	154
Number of SectorID	11	11	11	11

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5.2 reveals that our primary variable of interest, Sectoral Risk (ΔCVaR), is positive and highly significant in almost all lags up to the third lag, demonstrating that a rise in sectoral risk is connected with an increase in credit risk. Nevertheless, in the third lag, the sign of Sectoral Risk has reversed, indicating that the link between sectoral risk and credit risk is positive in the short term but negative in the long term. One potential economic explanation for this trend is that, in the near term, a rise in sectoral risk may result in a rise in credit risk owing to increased default probability and worse recovery rates in that industry. Long-term, however, the rise in sectoral risk may result in a tightening of credit standards and a decrease in lending, which might lead to a drop in total credit risk. This trend may also be affected by changes in macroeconomic circumstances, such as Macroprudential policy or the Repo Rate's reactions to sectoral concerns. This is very consistent with the Aikman et al. [1] and Schularick and Taylor [2] arguments that credit growth is substantially related with credit crises with a *two-year lag*.

In all other respects, the impact of MPI is comparable to the findings of Duan et al. [8]. MPI is positive in each of the four scenarios, demonstrating that a tightening of macroprudential policy is associated with an increase in credit risk. This indicates that the greater use of macroprudential policy tools has not decreased credit risk. Due to the restricted availability of loans, the credit risk may grow. Models couldn't find any link between GDP growth rate and credit risk. Though it is significant and positive in model (1) and follows the reckless lending argument, which states that during "good times" when real economic growth is strong, firms and financial institutions tend to be overly optimistic and make riskier investments, leading to a higher credit risk in an emerging country like India.

We also tested for Granger causality in a panel setting using Dumitrescu et al. [47] idea of GC Test in a panel setting. This conclusion may imply, from an economic standpoint, that sectoral risk is a leading indication of credit risk. This may be due to the fact that changes in the amount of sectoral risk may impact the capacity of borrowers in that industry to repay their loans, hence impacting the overall degree of credit risk. We find out similar relation in monthly cases too hence we'll not be adding those results exclusively.

Dumitrescu and Hurlin Granger causality test results for annual case

Optimal Number of lags (AIC): 2 (lags tested 1 to 3).
W-bar = 6.1899
Z-bar = 6.8156 (p-value = 0.0000)
Z-bar tilde = 3.7896 (p-value = 0.0002)

H_0: CVaR doesn't Granger-cause ΔCCI
H_1: CVaR does Granger-cause ΔCCI for atleast one panel (Sector).

We find out CCI doesn't granger cause Sectoral risk but Sectoral Risk does granger cause Sectoral Risk.

5.2.2 Monthly Case

The baseline regression for monthly case is,

$$\begin{aligned} Y_{i,t} = Y_{i,t-1}\alpha + CVaR'_{i,t-1}\lambda + MPI'_{i,t-1}\beta + GDP'_{i,t-1}\gamma \\ + Reporate'_{i,t-1}\theta + MarketCrash'_{i,t-1}\phi + \mu_i + \epsilon_{i,t} \end{aligned} \tag{5.3}$$

Variables	Observations	mean	std	min	max
dependent variable					
ΔCCI	1,969	6.16e-06	0.0093905	-0.1739582	0.1414304
independent variables					
CVaR	1,969	0.0279058	0.0189983	-0.2438909	0.0006619
MPI	1,969	0.1117318	0.7392244	-3	2
GDP Growth	1,969	1.710071	4.293956	-23.0582	21.39841
Repo Rate	1,969	6.629888	0 .975837	4	9
Market Crash	1,969	0.1005587	0.3008198	0	1

Table 5.3: Summary Statistics for Monthly case. Time-period: 2006-2020

Table 5.3 displays the descriptive statistics for all of our variables from 2006 to 2020. As this is a monthly instance, we may go more deeply into the specifics of numerous factors. The average change in the CCI over the last year has been 0.06 basis points, showing that the credit risks of the eleven separate industries have grown somewhat throughout the research period. Similar to the annual example, the sectoral risk stays positive, suggesting a rise in risk. The macroprudential policy index, which ranges from -3 to 2 and has a mean of 0.11 and a standard deviation of 0.73, has a large range of values. The Repo rate is comparable to what we've witnessed annually. GDP Growth Rate has a wide range from -23% to 21%, with a mean of 1.71%, indicating that India expanded well throughout this era. On an average we see 10% of times that market has been crashed.
In case of monthly data we too use a dynamic panel model with fixed effects and similarly degrading the equation or choosing the first difference would always cause an endogeneity problem, resulting in skewed estimates from widely used estimators. But here, we have an instance where T is quite large than N. In such case we can directly use the lagged dependent variable as an independent variable as when $T \rightarrow \infty$ the Nickell bias doesn't affect and hence we'll use the Driscoll Krayy estimator [29]. This has been already discussed in Chapter 3.

Table 5.4 demonstrated a striking match with our major Yearly Case data. Our primary variable of interest, Sectoral Risk (ΔCVaR), is positive and statistically significant at the 1% level. This shows that increased sectoral risk correlates with increased credit risk. This makes economic sense, since sectoral risk may directly affect the creditworthiness of borrowers in these areas. For instance, if a sector undergoes a recession or crisis, the creditworthiness of borrowers in that area may decline, hence increasing the credit risk for lenders.
As we discovered in the yearly example, MPI's one-year lag demonstrates that it has failed to reduce Credit risk.
The statistical significance of market collapse implies that market crashes are connected with greater credit risk. This is due to the fact that market collapses may directly affect the value of collateral, the availability of financing, and the creditworthiness of borrowers, hence increasing the credit risk for lenders. In addition, market collapses might be an indication of a larger economic downturn, which can raise credit risk.

Table 5.4: Sectoral Risk and Credit Cycle Interplay Including the Crisis Period

VARIABLES	(1) ΔCCI
L1.ΔCCI	-0.4687**
	(0.0964)
CVaR	0.1519**
	(3.9e-3)
L12.MPI	6.63e-04***
	(6.62e-05)
L14.GDPRate	-2.26e-04
	(6.69e-05)
L11.RepoRate	-5.37e-04**
	(7.38e-05)
L11.MarketCrash	2.35e-03***
	(1.86e-4)
Constant	8.00e-4
	(5.26e-4)
Observations	1,815
Number of groups	11

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Post Crisis

The descriptive statistics for all of our variables for the post-crisis period from 2009 to 2020 are shown in Table 5.4. We see an increase in credit and sectoral risk over the last year, as measured by the average change in the CCI and Sectoral Risk, respectively. The macroprudential policy index, which varies from -2 to 2 and has a mean of 0.06, decreased from the previous example, but the huge variance raises the issue of efficacy. The Repo rate and GDP growth rate stay almost the same, but the maximum repo rate decreases from 9% to 8%, indicating RBI's gradual reduction of the Repo rate.

Variables	Observations	mean	std	min	max
dependent variables					
ΔCCI	1,584	0.0000221	0.0074648	-0.1739582	0.1414304
independent variables					
CVaR	1,584	0.0252867	0.0158324	-0.16947	-0.0024512
MPI	1,584	0.0694444	0.6838823	-2	2
GDP Growth	1,584	1.683036	4.760279	-23.0582	21.39841
Repo Rate	1,584	6.434028	0.948885	4	8
Market Crash	1,584	0.0486111	0.2151216	0	1

Table 5.5: Summary Statistics for after crisis Monthly case. Time-period: 2009-2020

Table 5.6 is a compilation of our panel regression results for the monthly case excluding the Subprime Crisis era.

Table 5.6: Sectoral Risk and Credit Cycle Interplay excluding the Crisis Period

	(1)
VARIABLES	ΔCCI
L.ΔCCI	-0.395***
	(0.0933)
CVaR	0.111***
	(0.0330)
L18.MPI	2.52e-4
	(1.66e-4)
L16.GDPRate	2.37e-4
	(1.27e-4)
L9.PolicyRate	2.38e-4**
	(1.92e-4)
L13.MarketCrash	-9.80e-4**
	(3.92e-4)
Constant	1.51e-3
	(0.00130)
Observations	1,386
Number of groups	11

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

MPI and Sectoral Risk continue to be positive in the post-crisis period, but the impact of Sectoral Risk on Credit Risk has decreased by around 26%, leading us to infer that MPI and policy tools are effective in reducing Credit Risk, although more improvement is necessary. An intriguing observation in the post-crisis age is that the Market Crash has no beneficial effect on Credit Risk.

5.3 VAR Results

As briefly introduced in chapter 3 we'll diving deep into the details of our VAR model. Below we'll be introducing our 3-lag VAR model for annual data.
Variables: Repo Rate (RR), Credit Cycle Index (CCI), Sectoral Risk (CVaR). Lags(3)

$$\Delta RR_t = a_1 + \sum_{k=1}^{k=3} b_{11}\Delta RR_{t-k} + \sum_{m=1}^{m=3} b_{12}\Delta CCI_{t-m} + \sum_{n=1}^{n=3} b_{13}\Delta CVaR_{t-n} + u_t$$
$$\Delta CCI_t = a_2 + \sum_{k=1}^{k=3} b_{21}\Delta RR_{t-k} + \sum_{m=1}^{m=3} b_{22}\Delta CCI_{t-m} + \sum_{n=1}^{n=3} b_{23}\Delta CVaR_{t-n} + v_t$$
$$\Delta CVaR_t = a_3 + \sum_{k=1}^{k=3} b_{31}\Delta RR_{t-k} + \sum_{m=1}^{m=3} b_{32}\Delta CCI_{t-m} + \sum_{n=1}^{n=3} b_{33}\Delta CVaR_{t-n} + w_t$$

Matrix Representation,

$$\begin{bmatrix} \Delta RR_t \\ \Delta CCI_t \\ \Delta CVaR_t \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} \Delta RR_{t-1} \\ \Delta CCI_{t-1} \\ \Delta CVaR_{t-1} \end{bmatrix} + \ldots + \begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix}$$

First, we'll dive into some details of identifications tests to verify the stability of the VAR Models mentioned in chapter 3. Then the above-mentioned Matrix representation has previously been seen, and we will now go further into Cholesky Decomposition and Identification techniques for our specific situation. Afterwards, we will discuss our results. This section's presentation will be as follows:

1. First, we will provide the impulse response function of three sectors at a time, followed by the 5-year projection of CCI and Sectoral Risk. Before mentioning the IRF we'll provide some test statistics to ensure the different VAR models for different sectors are stable and doesn't suffer for autocorrelation. We can observe all the test statistics follows a similar pattern and don't suffer from any stability issues.

2. In addition, we will give the findings of the Granger causality test between the CCI, Sectoral Risk, and Repo Rate, as well as the results of the Cholesky prediction error variance decomposition for chosen sectors. Refer to the appendix for results from few more other sectors, since include all findings for each particular sector would make the chapter & Thesis very lengthy and illegible. Therefore, we focused on the most important and intriguing results.

Let's begin by analyzing the Cholesky Decomposition for our situation.

$$\begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} \Delta RR_{t-1} \\ \Delta CCI_{t-1} \\ \Delta CVaR_{t-1} \end{bmatrix}$$

As shown by the above matrix, this structure will be applied to all sectors. According to the ordering, we limit the short and long-term effects of certain factors; in this situation, the ordering is crucial. Our assumptions for this specific ordering are as follows:

1. CCI and Sectoral Risk have no short-term influence on Repo Rate, and it will take Repo Rate to react to changes in CCI and Sectoral Risk. Repo Rate will respond to changes in CCI and CVaR in a timely manner not in short run.

2. Sectoral Risk has no influence on the CCI in the near term, while Repo Rate does.

3. Moreover, both the CCI and the Repo Rate have a short-term impact on sectoral risk.

This link between Repo rate, credit cycle index, and sectoral risk is supported by the **Credit Channel Theory** [39]. Changes in monetary policy, such as the Repo rate, may alter the availability of credit and the cost of borrowing for enterprises

and consumers, according to this idea. As the Repo rate is decreased, the cost of borrowing decreases and credit becomes more accessible, which may lead to an increase in lending and investment activity. This may result in a rise in the credit cycle index and a reduction in sectoral risk. In contrast, when the Repo rate is increased, borrowing costs rise and credit becomes less accessible, which may reduce lending and investment activity and increase sector risk.
The **Minsky financial instability** hypothesis [12] is a second economic theory that supports this link. According to Minsky, the credit cycle is comprised of periods of stability followed by times of instability, with the instability being the result of the accumulation of financial fragility in the economy. Many variables, including the availability of credit and fluctuations in the Repo rate, might contribute to the emergence of financial instability. As the Repo rate is dropped, borrowing costs fall and credit becomes more accessible, which may contribute to a rise in financial fragility and sector risk. This might then result in a rise in the credit cycle index. In contrast, when the Repo rate is increased, borrowing costs rise and credit becomes less accessible, which may lead to a reduction in financial fragility and sectoral risk, which may result in a reduction in the credit cycle index. The ordering of the variables as Repo rate, credit cycle index, and sectoral risk (CVaR) is based on the economic theory that changes in the Repo rate may impact both the credit cycle index and sectoral risk, but the credit cycle index has a more immediate and direct effect on sectoral risk. Two economic theories that support this link between the variables are the credit channel theory and Minsky's financial instability hypothesis.

5.3.1 Test Statistics

In this part, we will examine three distinct test statistics to validate the integrity of our VAR models. The following are,

1. First, we will use the augmented dickey fuller test to check that all variables are stationary. If this is not the case, we must use error-corrected VECM or ARDL models. For all variables and sectors, the estimated MacKinnon p-value for Z(t) = 0.0000 was derived using the Dickey Fuller test statistics; hence, we can reject the null hypothesis and conclude that the variables are really stationary.

2. Next, we must examine var's stability state using the "varstable" stata command. All eigenvalues of a stable VAR will fall inside the unit circle, and the VAR will satisfy the stability requirement. Our VAR model is inconsistent without this. We absolutely verified this situation.

3. Finally, we must do the Lagrange multiplier test to ensure that there is no autocorrelation at the specific lag lengths that have been selected for various situations. All of our p-values are more than 0.05, thus we cannot reject the null hypothesis that there is no autocorrelation in lag order.

5.3.2 Impulse Response Function and Forecasting

An IRF assesses the impact of a shock to an endogenous variable on itself or on another endogenous variable [35] and [48]. Becketti [49] gives a straightforward and

moderate introduction to IRF analysis. For our study, we will be using Stata to generate orthogonalized IRFs.
Let's start by saying that all shocks have a standard deviation of one. All the grey region represents the standard error, the X-axis represents the future periods (years), and we constructed five future periods for the OIRF function. The percentage variation is shown on the Y-axis. All variables are analyzed in yearly frequencies, and their starting dates are either January 2000, January 2002, or January 2005. And all concludes in February 2022. In 2022, we only have data for two months.
FEVD estimates the fraction of an endogenous variable's forecast-error variance that may be attributable to orthogonalized shocks to itself or another endogenous variable [35] and [48]. Cholesky FEVD will be used among the various FEVD varieties.
The Variance decomposition reveals the proportion of inaccuracy in predicting a variable over time attributable to a particular shock.

1. In other words, the proportion of the variability in the dependent variable that can be attributed to "*its own shocks*" as opposed to "*shocks in the other variables in the system*".

2. As in the IRF, the Cholesky Decomposition is used for identification reasons in the variance decomposition.

Sector: Bank, IT and Healthcare

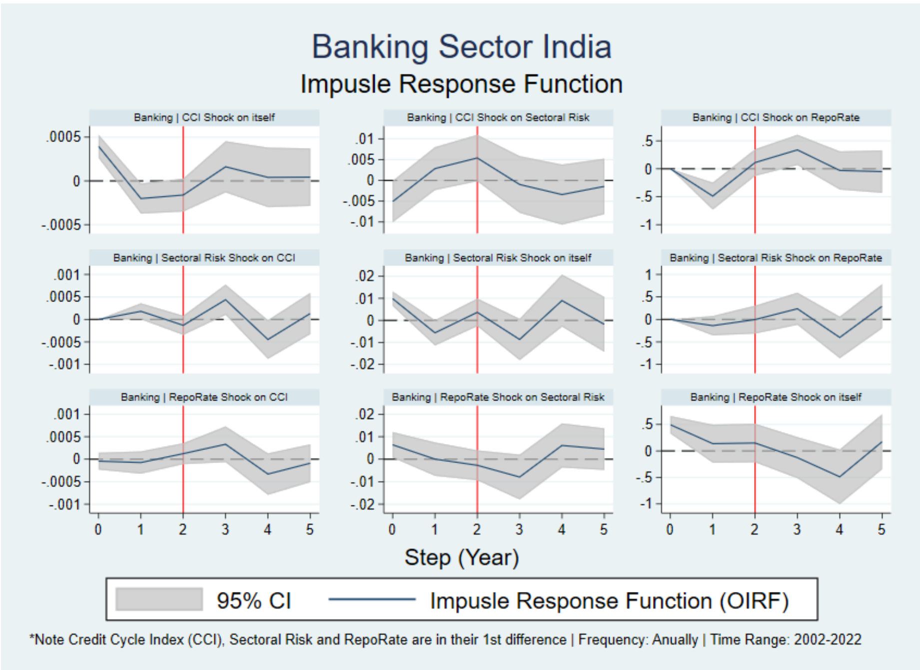


Figure 5.1: Orthogonalized IRF Function for Banking Sector

The image 5.1 depicts the OIRF of three distinct variables for the Banking industry. We won't focus mostly on diagonal pictures, which essentially describe the impulse on its own. All of these variations are explained by "**Base Effects**". We are mainly concerned with the interactions between cross-diagonal variables.

Equation	Excluded	Prob > chi2
RepoRate	CCI	0.000
RepoRate	CVaR	0.069
RepoRate	ALL	0.000
CCI	RepoRate	0.020
CCI	ALL	0.000
CCI	CVaR	0.000
CVaR	CCI	0.0907
CVaR	RepoRate	0.120
CVaR	ALL	0.149

Table 5.7: Granger Causality test for Baking Sector

From the table 5.7 we see,

- There's a bidirectional relationship between CCI and Repo Rate both helps each other to predict at 5% significance level.
- CCI also helps to predict CVaR at 5% significance level.

The CCI shock on the repo rate shows that the first reduction in the repo rate may be attributable to the fact that banks are likely to require additional liquidity from the central bank in the near run to cover their elevated credit risk in short run. As a consequence, the central bank may react by decreasing the repo rate in order to make more liquidity accessible to banks. Moreover, as we can see, the time frame we studied is from January 2002 to February 2022, after covid the central bank was already lowering the repo rate. But, over time, the central bank may grow worried about the possible inflationary effects of low interest rates and their effects on the economy as a whole. In order to preserve price stability and prevent inflation from reaching unmanageable levels, the central bank may begin gradually increasing the repo rate over the following two years. Considering we are presently in 2023, we can see that this is precisely the behaviour we witness.

When we examine the CCI shocks on Sectoral Risk, we see that a rise in credit risk may result in an increase in the default risk of borrowers in the banking sector. Thus, investors may become more risk averse and demand more compensation for keeping riskier assets in the sector, causing an initial rise in the sectoral risk index. Nevertheless, over time, the central bank may take measures to alleviate the effect of the credit risk shock, such as providing liquidity assistance to the banking sector, which may assist in stabilizing the sector risk index. We observe the sharpest decline in sectoral risk in this industry, particularly beginning in the third year. Though an increase in the repo rate may have a negative impact on investments in the stock market, we must also consider that a rise in the repo rate improves the bank's health and gives them the opportunity to receive higher interest rates from customers. As a result, people will be more inclined to invest in the banking sector in this scenario.

We may also observe that the Repo Rate Shock has a negative impact on Sectoral Risk, since the early rise in Repo Rate will encourage more individuals to invest in safer choices, such as FDs, which will now offer greater returns.

Steps(Year)	Response Shocks	RR				CCI				CVaR		
		RR	CCI	CVaR		RR	CCI	CVaR		RR	CCI	CVaR
1		1	0	0		.011982	.988018	0		.249604	.157993	.592402
2		.505532	.458073	.036396		.031179	.831672	.137149		.20077	.166598	.632632
3		.514351	.451517	.034133		.076724	.755842	.167434		.190173	.248509	.561318
4		.40669	.491542	.101768		.212725	.399533	.387742		.282844	.163055	.554101
5		.473013	.320501	.206486		.25902	.268764	.472216		.283985	.1448	.571215

Table 5.8: Cholesky forecast-error variance decomposition for Banking Sector

Table 5.8 displays the Cholesky variance decomposition for prediction error for the Banking Sector. In the beginning of this section, explain how FEVD enables us to see the bigger picture. Due to page limitations, we will only cover the FEVD of the Banking sector in this section; all other FEVDs may be found in the Appendix.
Many of the table's observations are as follows:

1. According to Choleskey decomposition, we established short and long-run criteria for the response variables stated before. Hence, the short run influence of CCI and CVaR on RepoRate is zero, and similarly, the short run effect of CVaR on CCI is likewise zero. This will be true for all industries.

2. The CCI shock is pushing an increase in the Repo Rate, which will continue until the fourth year. By the end of 5 years, 47% and 32% of the change in the Repo Rate may be explained by shocks caused by itself and the CCI, respectively.

3. We may also note that shock of CVaR increases Credit Risk (CCI) in the Sector over the long term. By the end of 5 years, shocks in Sector Risk and Repo Rate may explain 47% and 25% of the change in CCI, respectively.

4. We observe that the Repo Rate Shock affects Sectoral Risk and is capable of reducing Sectoral Risk until the third year. Similarly, we see that Sector Risk responds to the CCI shock by growing until the third year and then decreasing somewhat, leading us to conclude that by the end of the third year, other economic forces have kicked in and the sector has adapted to its new credit environment. By the end of the fifth year, the Repo Rate and CCI shocks explain 28% and 14% of the change in Sector Risk, respectively.

The image 5.2 depicts the OIRF of three distinct variables for the Information Technology industry. Similar to last case We won't focus mostly on diagonal pictures, which essentially describe the impulse on its own. We are mainly concerned with the interactions between cross-diagonal variables.

The CCI shock on the repo rate shows the initial increase in the repo rate as a result of the shock to the credit cycle index is consistent with the traditional theory of monetary policy transmission. A rise in credit risk would increase the cost of borrowing, compelling the central bank to boost its policy interest rate to maintain price stability. The subsequent decline in the repo rate in the second year may be

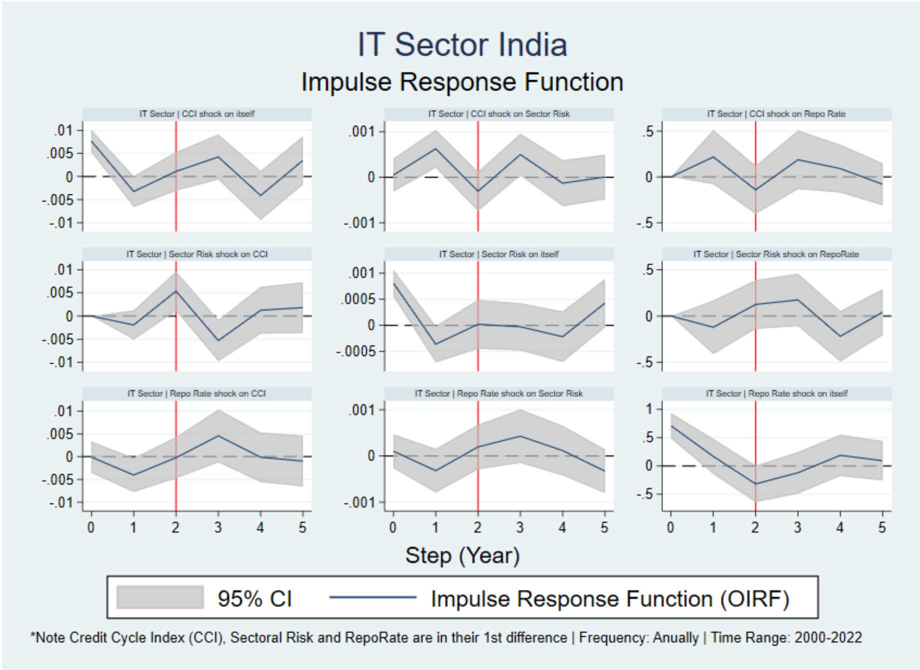


Figure 5.2: Orthogonalized IRF Function for IT Sector

related to the IT industry's low need for short-term funding. This suggests that the impact of the credit cycle shock on the IT sector's borrowing costs may have been temporary, leading in a drop in the repo rate the following year. The modest increase in the repo rate in the third year could be due to the reversal of the temporary effect from the previous year. In spite of this, the repo rate falls steadily over the next two years, maybe because the IT sector's credit risk has steadied and the central bank has lowered the policy interest rate to stimulate economic growth.
From Credit Cycle Theory we get to know that IT sector Short-Term Credit Requirement is quite low than any other industries hence the behavior of the IRF indicates that the IT sector's credit demand is relatively insensitive to changes in credit risk and that the impact of monetary policy on the sector's borrowing costs is temporary. CCI shocks on Sectoral Risk reveals that a shock to the credit cycle index results in an initial increase in the sectoral risk index, indicating an increase in credit risk and a possible drop in profitability. Successive declines and increases show sector adaptation and a shifting economic climate. However, based on the x-axis of the graph of CCI shock on Sectoral Risk, the change in CCI shock on Sectoral Risk is rather minimal, yet this provides a comprehensive picture. Similarly we can also see the Repo Rate shock on Sector Risk is also negligible.

The image 5.3 depicts the OIRF of three distinct variables for the Healthcare industry. The CCI shock on repo rate indicates an increase in credit risk, which may increase borrower default rates. After the Covid Crisis, healthcare financing is plentiful worldwide. As lenders grow risk-averse and demand higher-quality collateral, the repo rate may fall temporarily. Yet, as the repo rate rises, risk stabilizes and its effects slowly diminishes while coming to 5th year. According to Minsky's financial

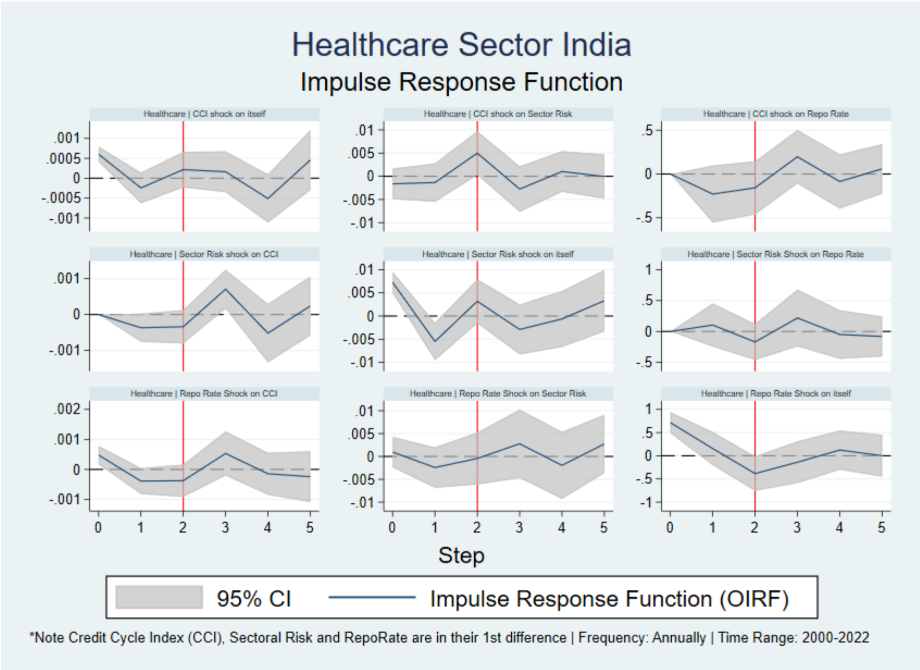


Figure 5.3: Orthogonalized IRF Function for Healthcare Sector

instability hypothesis, healthcare is less sensitive because to consistent demand and smaller leverage over the long term. CCI's shock didn't affect repo rate as much as other sectors.
As predicted, the repo rate shock on CCI decreases as repo rate climbs, and hence, credit risk decreases. CCI shock on Sectoral Risk demonstrates that the sectoral risk rises for the first two years and subsequently falls, indicating that the Repo Rate impact has already set in and reduced the risk somewhat by that point.
Figure 5.4 depicts CCI and Sectoral Risk projections for the aforementioned industries for the next five to six years. The fact that the predicted upper and lower bounds spread with time explains why Time series forecasting methods perform poorly with time.

For **Banking** sector CCI momentarily decreases and then begins to rise. It is corroborated by the fact that banks performed well in 2022, but as we have seen banks take chances in stable times and as the Repo Rate was lower, they handed out more number of loans(more number of loans more possibility of being band loans), which began the process of credit risk accumulation. Looking at the Adani Group financial problem in 2023 shows that our VAR model accurately forecasted this increase in CCI. Maybe as a result of Repo Rate, sectoral risk diminishes. Increasing the repo rate benefits the banking industry, hence reducing sectoral risk.

For **Healthcare** sector, Both CCI and Sectoral Risk exhibited a cyclical up-and-down trend around zero. But, if we look carefully, we can observe that sectoral risk almost replicates the pattern of CCI with a one-year lag. Thus, we can conclude the market need time to respond to such credit risk news.

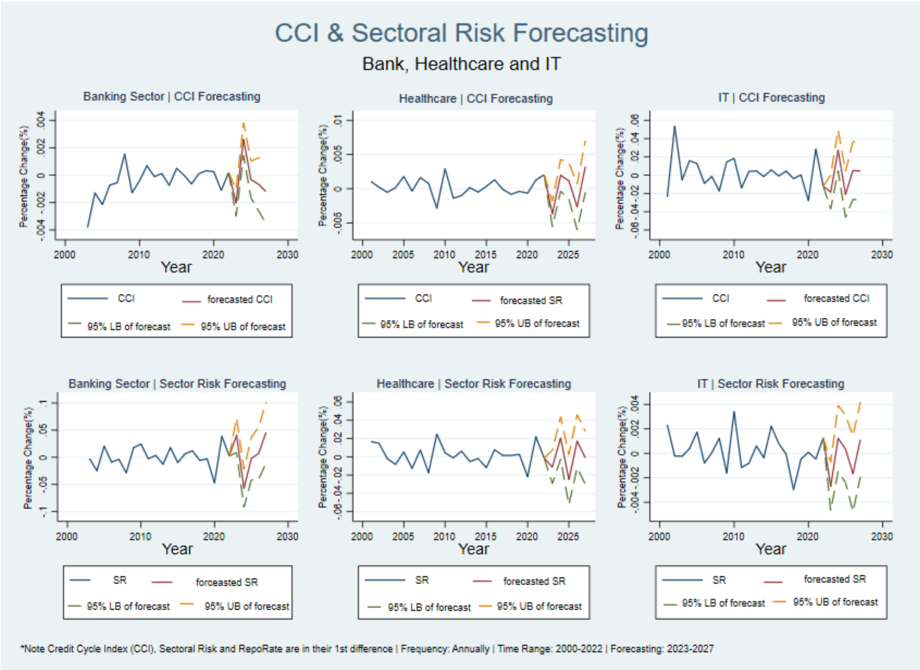


Figure 5.4: CCI & Sectoral Risk Forecasting for Banking, IT & Healthcare sectors

In the **Information Technology** industry, CCI and sectoral risk decrease in the short term, but increase in the long run. We may infer that individuals are more acclimated to working from home after a Covid-19 crisis. This increased the health of the IT sector, but as time passes, individuals will return to their usual schedules, resulting in losses for the sector's participants and a rise in Credit and Sector Risk in the long term.

Sector: Consumer Discretionary, Financial Services and Utilities

The image 5.5 depicts the OIRF of three distinct variables for the Consumer Discretionary industry. Technology industry. The CCI shock on Sector Risk shows an increase in credit risk can reduce credit availability, which reduces investment and growth in the Consumer Discretionary sector. Hence, market risk may rise, raising the sectoral risk index. However, the sector may be able to adapt to the higher credit risk environment by reducing its reliance on external credit and seeking alternative financing, stabilizing the sectoral risk index. Which implies that credit risk and sectoral risk for this case are driven by the same underlying factors, such as macroeconomic conditions, and are therefore co-moving in the short term. Yet, the influence of these common variables may diminish with time, resulting in a stability of the sectoral risk index.

Repo Rate Shocks on Sectoral Risk indicates that the relationship between the repo rate and sectoral risk is temporary and negative. This can be explained from Credit Channel theory which indicates for this sector is sensitive to interest rate changes due to high capital costs. A hike in the repo rate could restrict the availability of

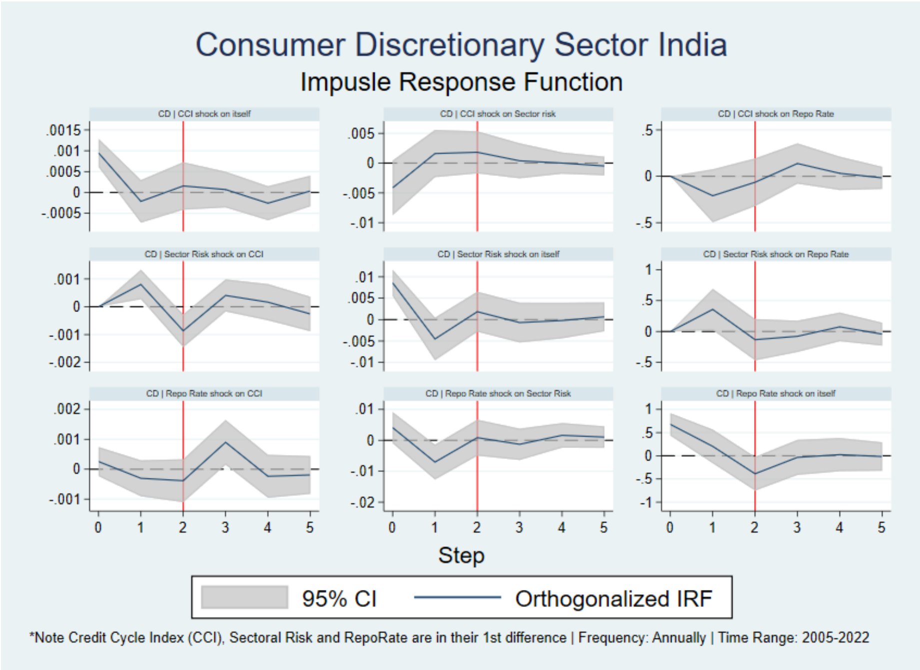


Figure 5.5: Orthogonalized IRF Function for Consumer Discretionary Sector

credit, which could in turn reduce investment and slow the growth of the Consumer Discretionary sector. Thus, the market risk associated with the sector may decrease, resulting in a short-term decrease in the sectoral risk index. Over time, however, the sector may be able to adapt to the higher repo rate environment by exploring alternative funding sources or by reducing its reliance on external credit, resulting in an increase in the sectoral risk index in the medium term. Another possible explanation is that the repo rate shock may have an indirect effect on sectoral risk through changes in other macroeconomic factors, such as inflation or currency rates, which may affect the industry's demand and supply conditions.
Image 5.6 illustrates the OIRF for three unique variables within the Financial Services sector. This segment gave equally comparable outcomes to the banking industry due to its similar nature. One may thus go to the banking sector's description of IRF from previously to grasp the idea.
Image 5.7 illustrates the OIRF for three unique variables within the Utility sector. The CCI shock on Sector Risk shows rise in CCI implies increase in Credit Risk which raises borrowing prices. This affects utilities companies' finances, increasing sectoral risk. However, over time, the sectoral risk starts to decrease. Utility companies may have adjusted to the shifting credit climate by reducing debt or taking other steps to manage financial risk. Business strength reduces sectoral risk. The overall economic climate may have improved. When the economy recovers, sectoral risk may decrease if the credit cycle index is rising due to a transitory shock like a recession or crash/default risk of any big firm of the sector(Adani Situation in Indian Market(2023)).

Figure 5.8 depicts CCI and Sectoral Risk projections for the aforementioned industries

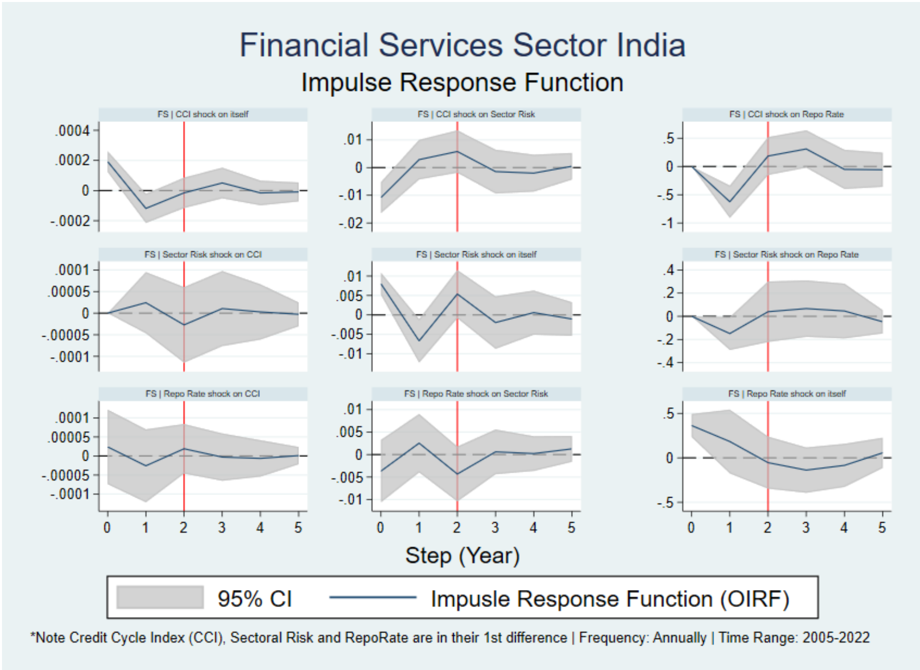


Figure 5.6: Orthogonalized IRF Function for Financial Services Sector

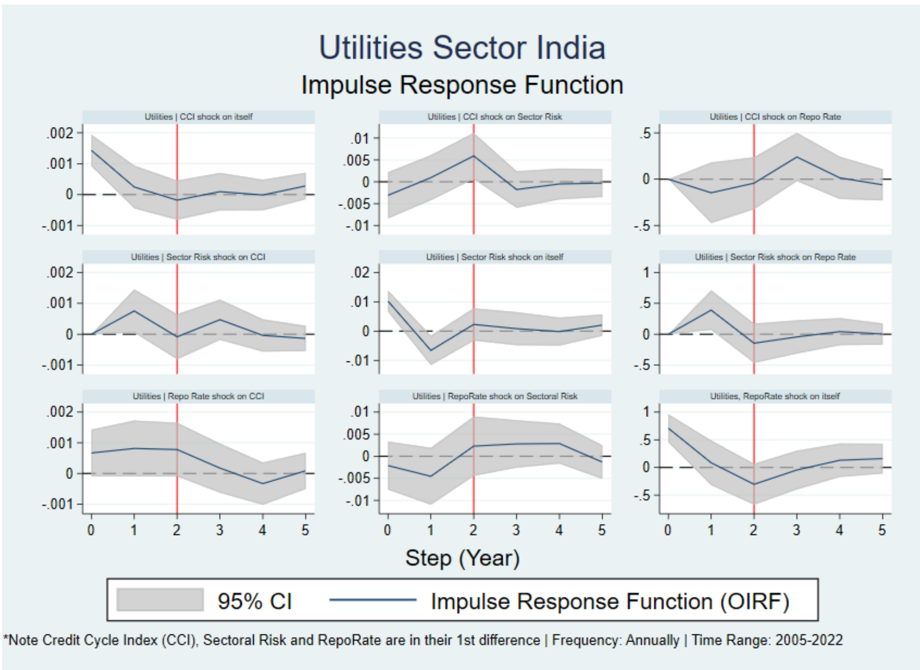


Figure 5.7: Orthogonalized IRF Function for Utility Sector

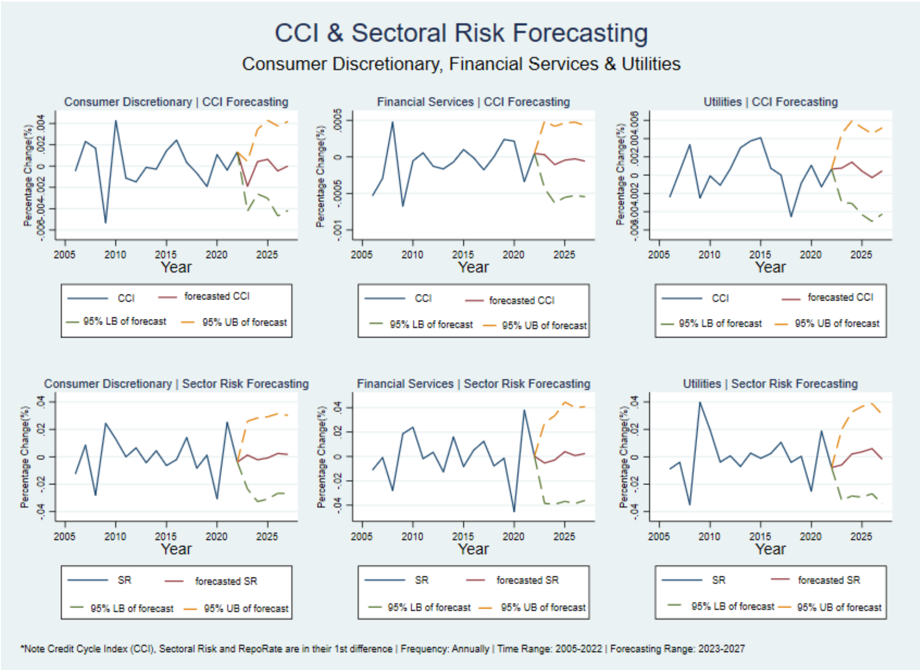


Figure 5.8: CCI & Sectoral Risk Forecasting for Consumer Discretionary, Financial Services & Utility sectors

for the next five to six years. We see the spreads are much wider in these cases, as in these cases we have data only from 2005 it affects the model estimation. Any time series model performs well if we feed more amount of past data upto a threshold.
For **Consumer Discretionary** CCI forecast shows decrease in short term and slight increase in long term. But we see the Sectoral Risk forecast shows gradual increase in both short and long term.
For **Financial Services** sector, Both CCI and Sectoral Risk pretty stable around zero with CCI having an initial dip and Sectoral Risk having a short term increase. These observations (and later trend of the graph) shows that sectoral risk almost replicates the pattern of CCI with a one-year lag. Thus, we can conclude the market need time to respond to such credit risk news.
In the **Utility** industry, CCI forecast shows short term increase in CCI and this behaviour can be explained by gradual increase in energy crisis in post covid world but in long term we see slight decrease in CCI. Similarly we see there is a gradual increase in both short and long term in the forecast of Sectoral Risk.

Sector: Energy, Commodities and Industrials

Image 5.9 illustrates the OIRF for three unique variables within the Energy sector. We observe a similar behaviour of the shocks in this sector like Utilities sector. CCI shock on Sectoral Risk shows in the short term, a higher credit cycle index may lower credit supply to the energy industry, which might reduce investment and economic activity. Investors may perceive increased risk, raising the Sectoral Risk Index. Over time, the industry may react to greater credit risk and obtain other funding or become more efficient at controlling credit risk. The Sectoral Risk Index may gradually decrease and stabilize. Post covid the energy crisis actually risen in India

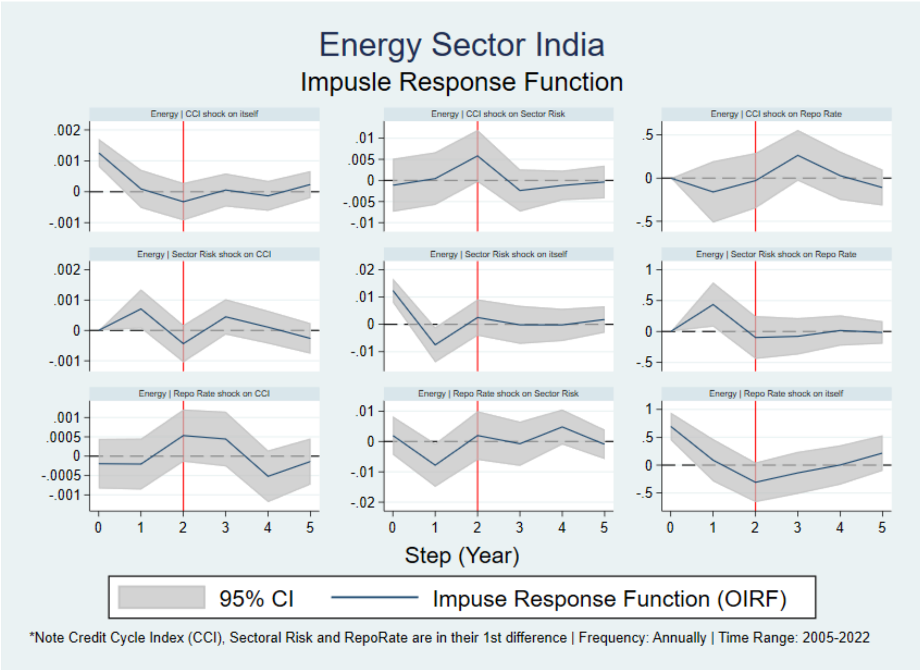


Figure 5.9: Orthogonalized IRF Function for Energy Sector

which is indicated by the higher value of CCI in the period 2021-2022. Investors may demand greater risk premiums because credit risk increases energy sector default risk. Restructuring debt or improving operations may lower the Sectoral Risk Index as enterprises react to the new risk environment.

Then Repo Rate shock on Sectoral Risk depicts a very interesting story. Energy corporations may pay more to borrow if the Repo Rate rises. This can reduce sector investment and economic activity, raising investor risk perception and the Sectoral Risk Index. Higher borrowing rates raise company default risk and the Sectoral Risk Index. The Repo Rate increase may decrease consumer demand for energy products and services, lowering energy sector revenues and profits (Current push towards EV and alternate energy is big example of this) . The Sectoral Risk Index may rise if sector valuations fall.

Image 5.10 illustrates the OIRF for three unique variables within the Industrial sector. Credit risk transmission through the economy explains our CCI shock on Sectoral Risk. Lenders become more cautious when the CCI rises because it indicates a higher default risk. This decreases loans to industrial enterprises, lowering investment and production.

Short-term sectoral risk index grows when companies become more exposed to external shocks due to lower investment and production faced by covid-19 crisis. Over time, enterprises respond to the changing economic conditions and strengthen their risk management techniques, lowering the sectoral risk index.

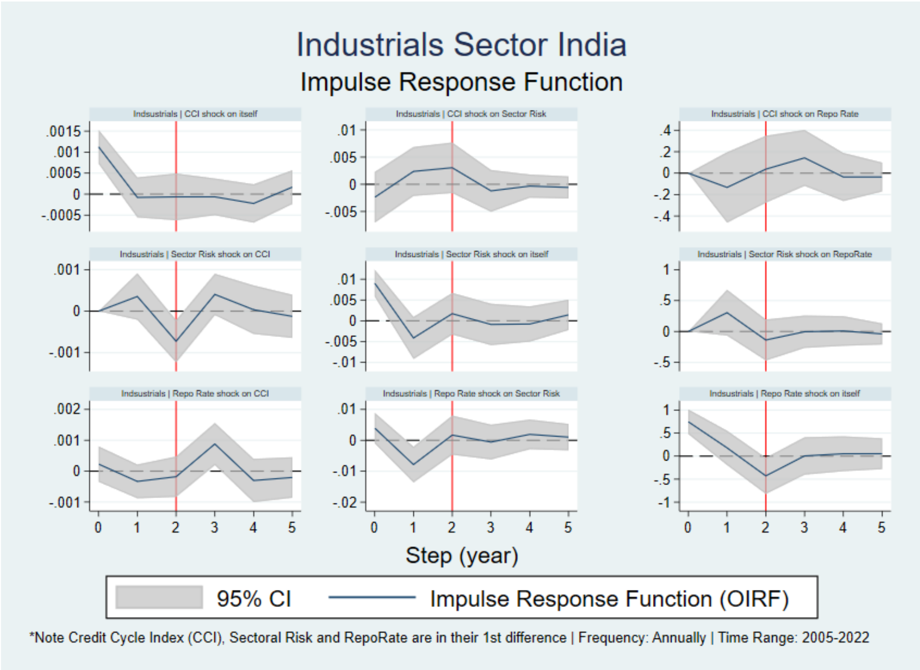


Figure 5.10: Orthogonalized IRF Function for Industrial Sector

Regulators and central banks may stabilize the credit market and restore investor confidence, lowering the sectoral risk index in the medium run. Liquidity injections, interest rate modifications, and bank capital requirements are examples.

After a credit shock, enterprises adjusted to new economic conditions and regulatory bodies and central banks stabilized the credit market, lowering the sectoral risk index. The repo rate rises as the central bank tightens monetary policy, reducing lending to the economy, notably the industrial sector in this case. The sectoral risk index drops in the short term as industrial businesses cut investment and production. In the medium term, industrial enterprises may enhance operational efficiency and minimize loan reliance. Cost-cutting, asset utilization, and equity or asset-backed financing may be used.

Industrial investment and output declines can also lower demand for energy, raw materials, and labor. This can lower input prices, which can boost construction and transportation enterprises' profits. Credit cycle index may rise which can be seen from Repo Rate shock on CCI graph. So, the sectoral risk index may rise in the medium run as the credit cycle index rises and industrial enterprises improve their operational efficiency. Yet, as industrial businesses adjust to the new economic conditions and the economy adjusts to the new monetary policy regime, the sectoral risk index may settle down.

Figure 5.11 depicts the OIRF for three distinct variables in the Commodities industry. This sector's IRF performance closely resembles that of the Industrials sector, which we reviewed before.

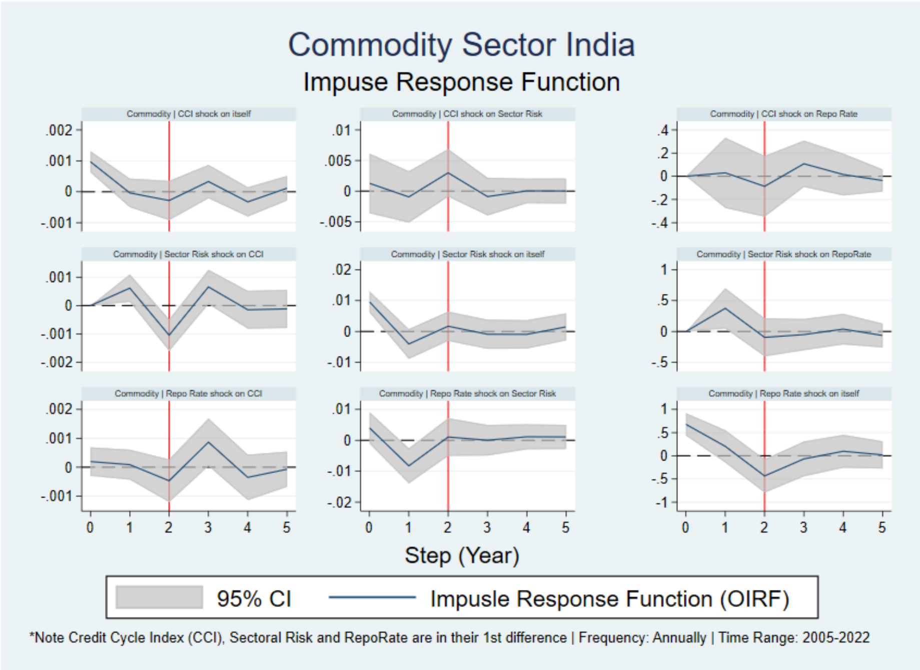


Figure 5.11: Orthogonalized IRF Function for Commodities Sector

Figure 5.12 depicts CCI and Sectoral Risk projections for the aforementioned industries for the next five to six years. We see the spreads are much wider in these cases, as in these cases we have data only from 2005 it affects the model estimation as we discussed earlier.

The CCI projection for the **Energy** sector indicates a high growth in the near term and a minor reduction after three years, but an overall gain from the base level. Similarly, Sectoral Risk likewise grows over time, but in the short run it drops and then surges steeply from which we can conclude the market need time to respond to such credit risk news.

For both **Industrials** sector and **Commodity** sector CCI decreases in the short run but rises substantially in the long run. And overall, sectoral risk rises gradually for both scenarios.
These projections indicate that these three industries will experience short- and long-term stress.

Sector: Reality and Telecommunications

Figure 5.13 depicts the OIRF for three distinct variables in the Reality sector. The sectoral risk index drops in the first year following the CCI shock on Sectoral Risk because higher credit risk might tighten lending conditions and diminish real estate investment. When market exposure decreases with lesser investment, the sector's risk may diminish. Also, coming from covid-19 crash reality sector in India was

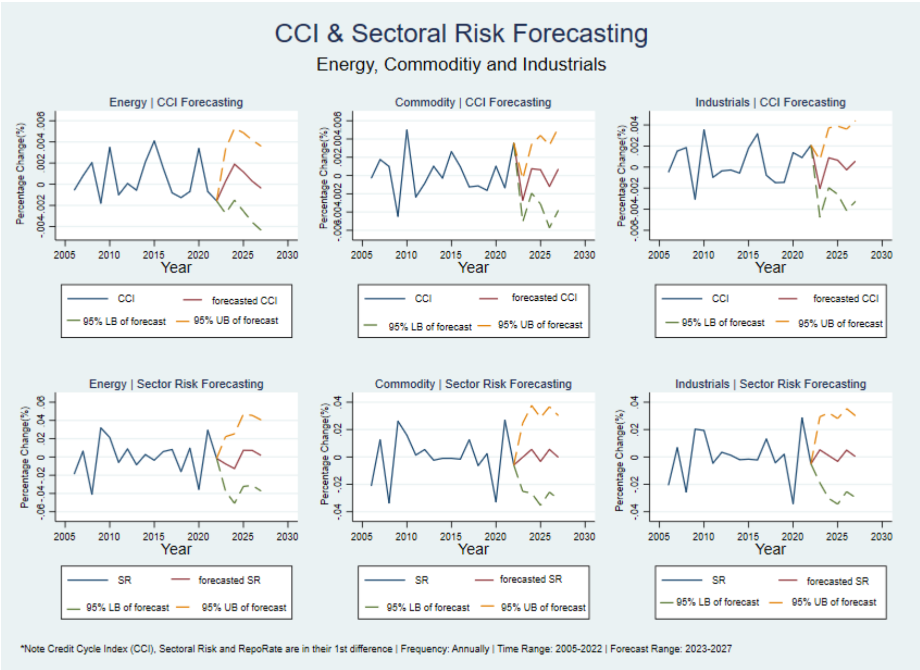


Figure 5.12: CCI & Sectoral Risk Forecasting for Energy, Industrials & Commodity Sector sectors

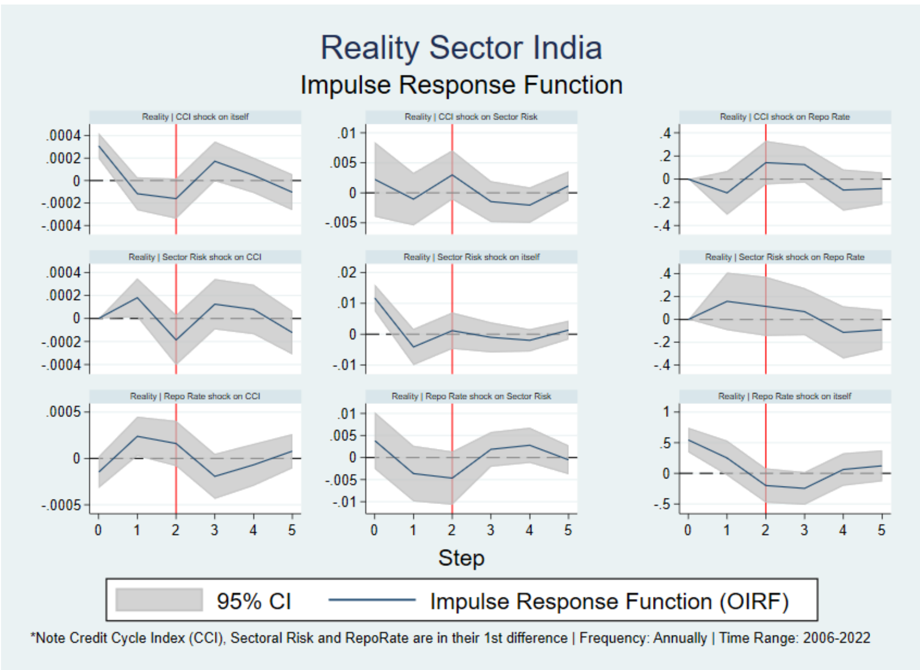


Figure 5.13: Orthogonalized IRF Function for Reality Sector

performing well for various reasons, like: Low Interest Rates, Government massive infrastructure push, bullish sentiment amongst people sentiment to own house, growth of IT/ITeS/ E-Commerce Sector and increase in the demand for office spaces warehouse etc.

The sectoral risk index increased in the second year, suggesting that the real estate industry is becoming more susceptible to credit cycle concerns. This may imply that the industry is more subject to economic changes or market sentiment movements and more likely to undergo significant drops in a downturn. In the third year, the sectoral risk index decreased, which may suggest market players' attempts to minimize credit cycle risks or a change in real estate attitude. Lastly, the sectoral risk index stabilization in the following years implies that market players have adjusted to the new credit cycle normal and that the sector has recovered some stability.

The Repo Rate shock on Sectoral Risk shows the sectoral risk index initially decreased in the first two years following the repo rate shock because a higher repo rate might raise real estate sector borrowing costs and tighten lending. This may reduce real estate demand and investment, lowering risk.

In the next three years, the sectoral risk index gradually increased, suggesting that the real estate industry is becoming increasingly susceptible to high repo rate concerns. This suggests that the sector is more susceptible to economic changes and market mood movements and may fall more sharply in a downturn.
Credit Channel Theory [39] suggests this sector is sensitive to interest rate changes due to high capital costs. The real estate business is largely dependent on finance, therefore changes in credit costs or availability can significantly affect its performance. As real estate investments are long-term and need large upfront money, the industry may be more vulnerable to interest rate movements than other industries. The sectoral risk index's response to a repo rate shock shows the various economic dynamics that affect India's real estate industry.

Figure 5.14 depicts the OIRF for three distinct variables in the Telecommunication sector. An increase in the credit cycle index first raises the telecommunication sector's credit risk. Due to tighter credit, higher default rates or other factors that increase the likelihood of loan defaults. For this sector Indian Market always been dominated 3-4 players, and one of them Vodaphone default risk rose quite significantly in last few years. Hence, the short-term telecommunication sectors risk index rises. Over time, sector businesses may change their business plans, diversify their funding sources, or take other measures to reduce credit risk. Even if the credit cycle index stays high, these efforts might progressively lower the sectoral risk index and stabilize it. Market dynamics, regulatory laws, and government schemes may also affect the sectoral risk index.

In 2021, the Indian government approved a rescue package for the debt-ridden telecom firms in the country and reserved rights to convert interest on them into equity at a

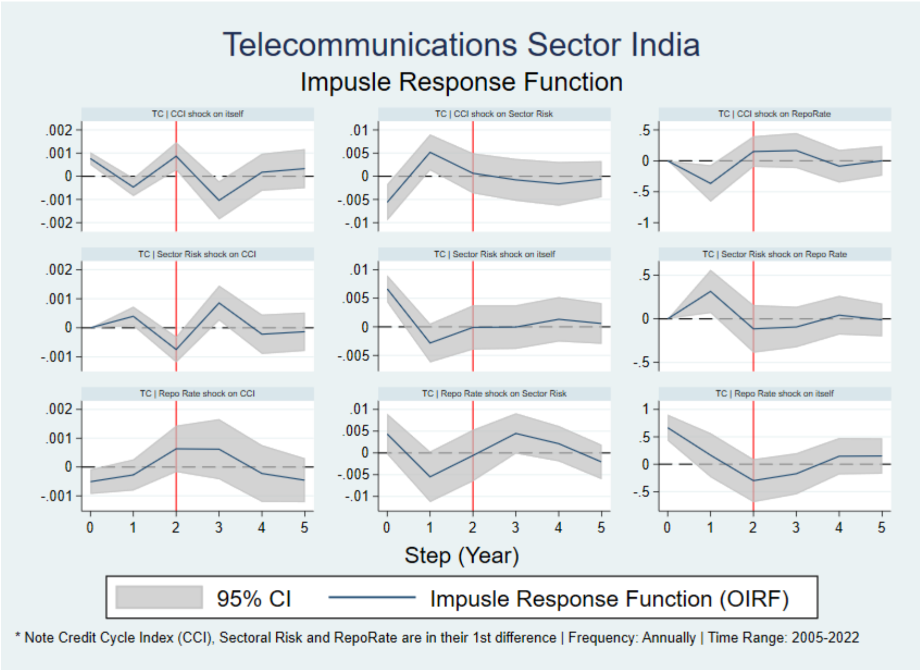


Figure 5.14: Orthogonalized IRF Function for Telecommunications Sector

later date.
Figure 5.15 depicts CCI and Sectoral Risk projections for the aforementioned industries for the next five to six years. We see the spreads are much wider in these cases, as in these cases we have data only from 2005 it affects the model estimation as we discussed earlier.

The CCI projection for the **Reality** sector shows that CCI doesn't increase much from the baseline but Sectoral Risk forecast shows an upward trend which implies the Sectoral Risk might rise in long term for this sector.

For **Telecommunications** sector both CCI and Sectoral Risk projections show an increasing trend, indicating that this sector will be subject to credit-related pressures, which will translate into an increase in Sectoral Risk. In the short term, however, the CCI demonstrates a slight decline, which indicates government intervention and the approval of a rescue plan for the insolvent telecom companies.
These projections indicate that these three industries will experience short- and long-term stress.
Reader might difficulties to go through such detailed prediction results, hence we compile the all forecasting results in a single chart in Chapter 6.

5.4 NLP Results

In this part, we will examine a few of the outcomes of our suggested NLP research attempt. The whole research was completed when I was an intern at AIDF NUS Business School, and a substantial portion of my work, such as codes and finished

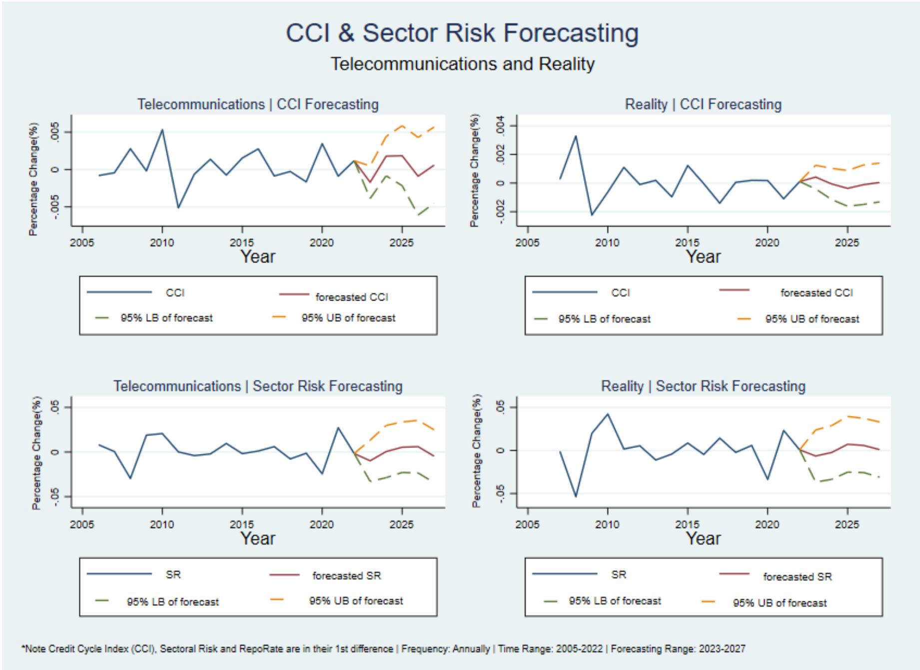


Figure 5.15: CCI & Sectoral Risk Forecasting for Reality & Telecommunications Sector sectors

webpages, must be kept private since it was developed in a corporate setup. Nonetheless, we will demonstrate some key results, detailed steps on how to extract credit risk sentiment from user-input articles, a couple of case studies to familiarise ourselves with how to produce daily entity-specific credit risk sentiments, and we will conclude the chapter by discussing the AIDF NLP Team's flagship product, **NS Monitor**, which is capable of producing hourly firm-specific sentiments regarding credit risk and greenness. During my internship at AIDF, I was mostly responsible for setting up the whole backend of this flagship project. This was my most significant contribution.

In addition, during this time period, we included a new subject, **Greenness**, along with **Credit Risk**, to contribute to the Green Finance research initiative at AIDF, and we covered practically all **ASEAN** languages in addition to English and Traditional Chinese.

5.4.1 Analyzing Credit Risk and Greenness Sentiment: User Input and Relevance Assessment

Let us first view wordcoulds we got relating to both of the topics in 4 distinct languages employing Source LDA.

Word clouds assist visualize a text or subject model's most common and essential terms. A topic modelling word cloud may determine the most important words related to a subject or cluster of topics.

Topic modelling uses word frequency and co-occurrence in a corpus to identify subjects. For each subject or cluster of topics, a word cloud may instantly determine the most important and relevant terms. Word clouds may quickly and intuitively reveal

Figure 5.16: Word-cloud for Credit Risk topic in English

Figure 5.17: Word-cloud for Credit Risk topic in Chinese

Figure 5.18: Word-cloud for Credit Risk topic in Bahasa

Figure 5.19: Word-cloud for Credit Risk topic in Vietnamese

a corpus's content and organization by displaying the topic's most common terms.

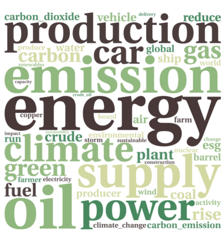


Figure 5.20: Word-cloud for Greenness topic in English

Figure 5.21: Word-cloud for Greenness topic in Chinese

Figure 5.22: Word-cloud for Greenness topic in Bahasa

Figure 5.23: Word-cloud for Greenness topic in Vietnamese

All the above images wordcloud for topic Credit Risk & Greenness in 4 different languages.

Using the concepts discussed in chapter 4, let's examine a few examples of how a user can input their own article and receive the degree of relevance associated with the input article in terms of Credit Risk and Greenness, as well as the sentiment score associated with each identified corporate entity within the input article.

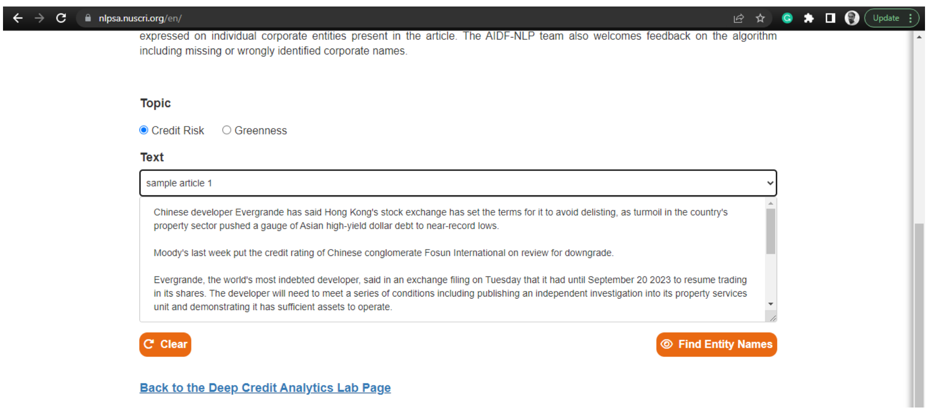


Figure 5.24: Input of a user input article regarding Evergrande Group

Figure 5.24 demonstrates how we may copy and paste articles from any source directly into the user input field. Next, based on the relevance of the article, we have the choice of selecting between two alternatives, Credit Risk or Greenness.

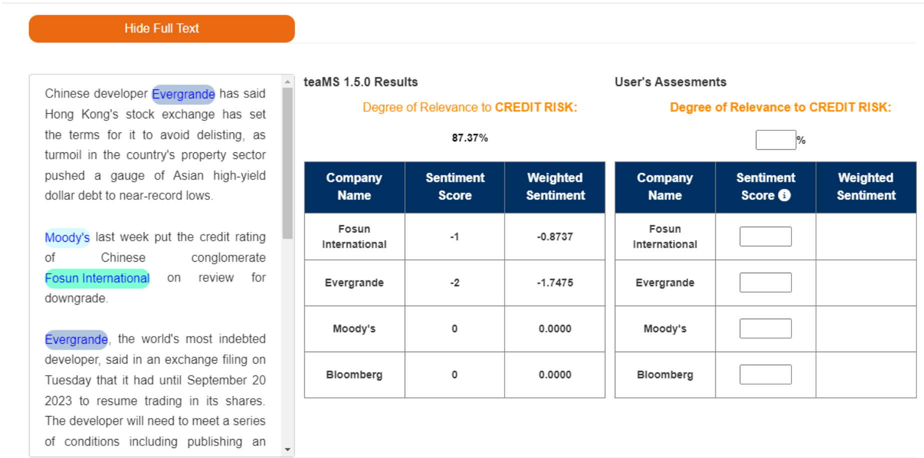


Figure 5.25: Entity Specific Sentiment: 1

In figure 5.25 on the left column, we can see the user-entered text as well as the entities that the NER algorithm was able to recognize. The article's relevance to Credit Risk is then shown in the upper-middle portion of the table. The next column displays the sentiment scores and weighted sentiment for each of the discovered entities. As stated in Chapter 4, the Sentiment Score is essentially a five-class classification (-2,-1,0,1,2) utilizing the TABSA-BERT. And sentiment is aggregated using weights derived from Source LDA and the BERT classification score. For comprehensive instructions, please refer to Section 4.2.4 and 4.2.5 of Chapter 4.

Figure 5.26: Entity Specific Sentiment: 2

We provide a second illustration in figure 5.26 using an article from *The Financial Express* discussing the deteriorating performance of *State Bank of India* in the Asia-Pacific region.

5.4.2 Real-Time Assessment of Credit Risk & Greenness Sentiment for Corporate Entities

Currently, based on this methodology, we have built AIDF's flagship product, **NS Monitor**, which is capable of hourly firm-specific Credit Risk and Greenness sentiment monitoring.
A concise summary of the project is as follows:

1. We automate the crawling of over 40 significant business press RSS feeds, with each source being crawled every 5 minutes.

2. We process this extracted data in real time using an AIDF-developed algorithm. Hence, the sentiment ratings for any publicly traded company may be updated in real time.

3. We also used daily sentiment to develop a sentiment index similar to prominent stock indices for the vast majority of listed firms.

This project is currently under development, and only developers and paying clients have access to the website. We will share screenshots of the final demo website and discuss them as we go along.

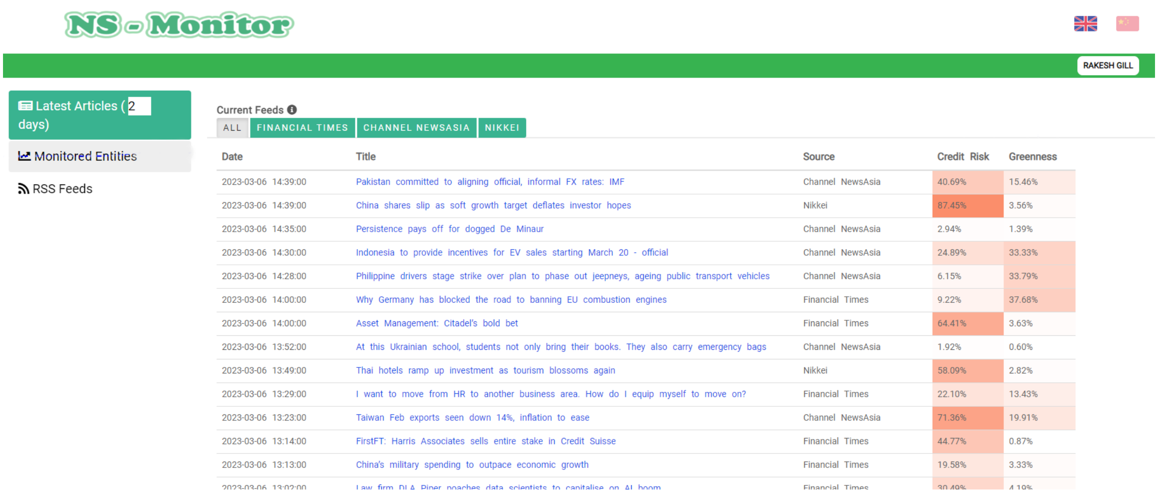


Figure 5.27: Overview of the NS Monitor

Figure 5.27 demonstrates that user will be forwarded to this page after logging in. Here on the left one can see three tabs labelled recent articles, monitored entities, and RSS Feed. First, let's examine the tab for the most recent article, which has five columns labelled Date, Title, Source, Credit Risk, and Greenness. The articles are ordered in chronological order, with the most recent items appearing first. One may see the Title and Source of each article, as well as the Credit Risk and Greenness columns, which indicate the article's importance to the respective themes. More it is relevant to the subject, the more prominent the colour will become.

One may pick the option monitored entities to precisely select a publicly traded company in the search field and get results similar to figure 5.28 for numerous stocks simultaneously. Figure 5.28 shows current(when the snap was taken) Credit Risk and Greenness Sentiment Score of Apple Inc. and Tencent Holdings Ltd.

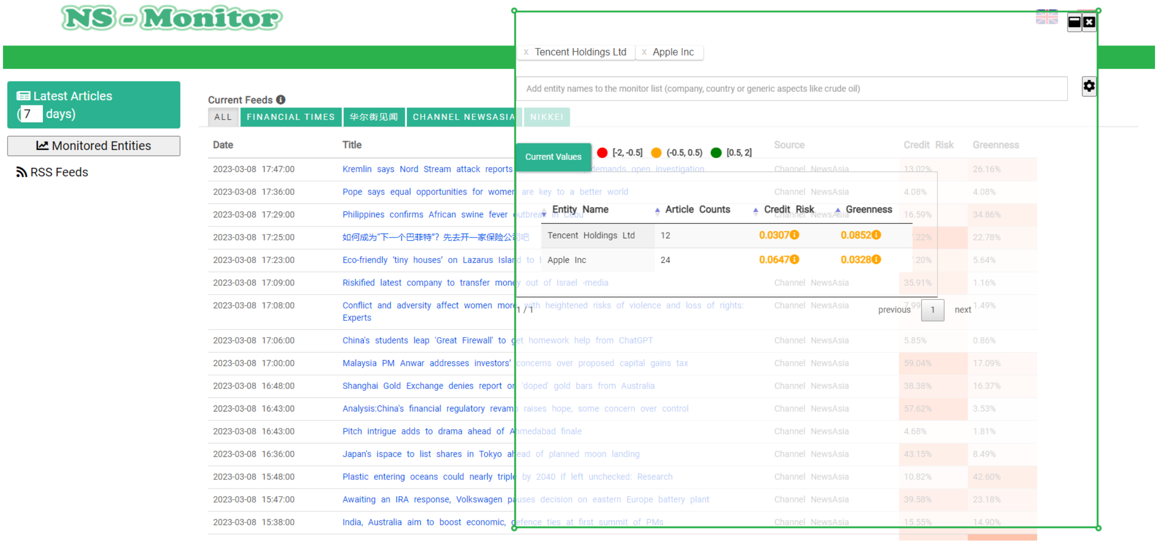
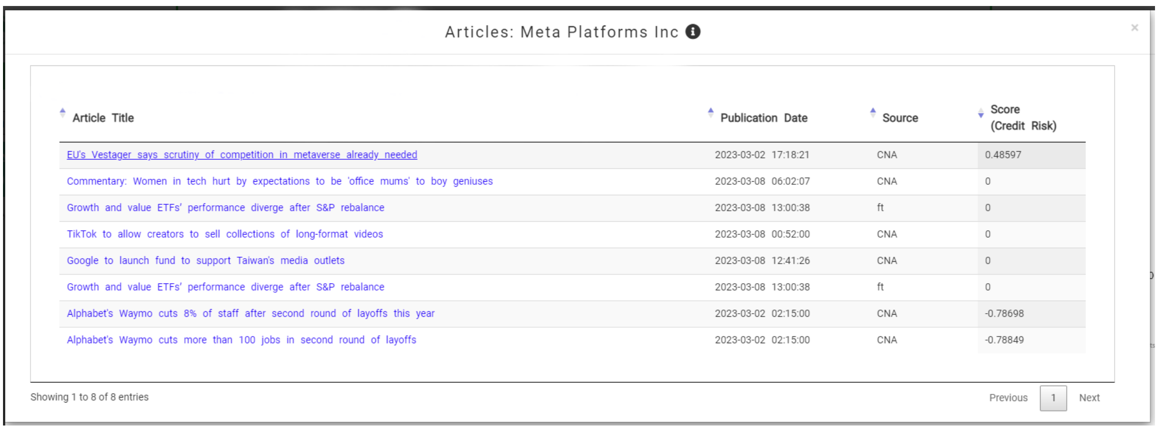


Figure 5.28: Entity Specific Sentiment

One may be interested in retrieving all articles for a certain company or a particular topic. We also have that feature. Figure 5.29 displays all of the Meta Inc. articles presently (at the time of the snapshot) stored in NS Monitor's database.



Articles: Meta Platforms Inc

Article Title	Publication Date	Source	Score (Credit Risk)
EU's Vestager says scrutiny of competition in metaverse already needed	2023-03-02 17:18:21	CNA	0.48597
Commentary: Women in tech hurt by expectations to be 'office mums' to boy geniuses	2023-03-08 06:02:07	CNA	0
Growth and value ETFs' performance diverge after S&P rebalance	2023-03-08 13:00:38	ft	0
TikTok to allow creators to sell collections of long-format videos	2023-03-08 00:52:00	CNA	0
Google to launch fund to support Taiwan's media outlets	2023-03-08 12:41:26	CNA	0
Growth and value ETFs' performance diverge after S&P rebalance	2023-03-08 13:00:38	ft	0
Alphabet's Waymo cuts 8% of staff after second round of layoffs this year	2023-03-02 02:15:00	CNA	-0.78698
Alphabet's Waymo cuts more than 100 jobs in second round of layoffs	2023-03-02 02:15:00	CNA	-0.78849

Showing 1 to 8 of 8 entries

Previous 1 Next

Figure 5.29: Entity Specific Articles

And lastly, if a user is interested in checking the historical moving average sentiment score for a certain entity for a particular topic, we also provide that option. Figure 5.30 depicts how the historical moving average sentiment score for the subject credit risk for the company Meta Inc. may be seen. As previously said, the structure of an entity-specific index is remarkably similar to that of stock indexes. And we offer similar viewing choices for the past 7 days, 1 month, 3 months, and one year. The current sentiment score is shown by the yellow dot. The left axis represents the range of sentiment scores, from -2 to +2. The X-axis denotes indicate the dates, while green bars represent the daily article counts for the given company.

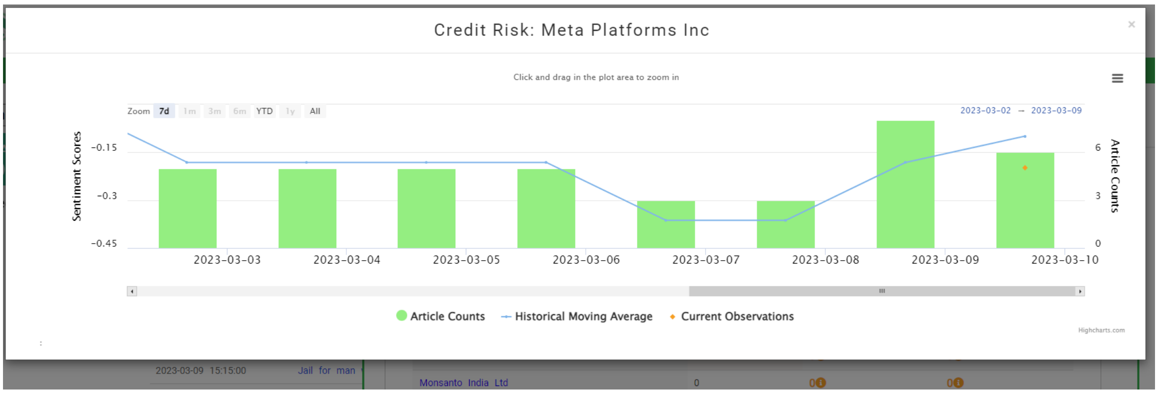


Figure 5.30: Entity Specific Sentiment Index

Several viewing choices, such as Topic-Specific and Source-Specific, are also available to meet the needs of the viewer. Now, we are in the process of adding multiple articles from different sources to improve on the primitive work we shown here. We really believe that our study will benefit numerous bond investors, rating agencies, and scholars.

5.4.3 Case Study

In this segment, we will introduce one case study regarding a very recent infamous instances of credit stress event. This case study will demonstrate briefly how the Entity specific credit sentiment index developed by the NLP Team of AIDF-NUS helps to signal these events and how it enhances the model used by CRI-NUS to calculate the probability of defaults for various entities.

5.4.3.1 Credit Suisse

Credit Suisse, one of Switzerland's top banks, has suffered controversies and losses since 2020. The bank's CEO, Tidjane Thiam, resigned in February 2020 after a spying controversy. The bank engaged private detectives to follow its former wealth management boss, Iqbal Khan, who had joined UBS just blocks away. Credit Suisse planned and executed most of seven espionage activities between 2016 and 2019, misleading Switzerland's banking regulator about the scale. Credit Suisse altered a surveillance invoice to hide organizational issues detected by the regulator.

In March 2021, British financial business Greensill Capital collapsed, forcing Credit Suisse to terminate four funds with $10 billion invested. The Swiss Financial Market Supervisory Authority (FINMA) said the bank "*gravely disregarded its supervisory responsibilities*" and mandated increased reporting and protections. Four former Credit Suisse executives faced enforcement procedures. Three weeks after Greensill, U.S. family office Archegos Capital Management collapsed, costing Credit Suisse $5.5 billion. Paul Weiss, Rifkind, Wharton Garrisson found a "basic

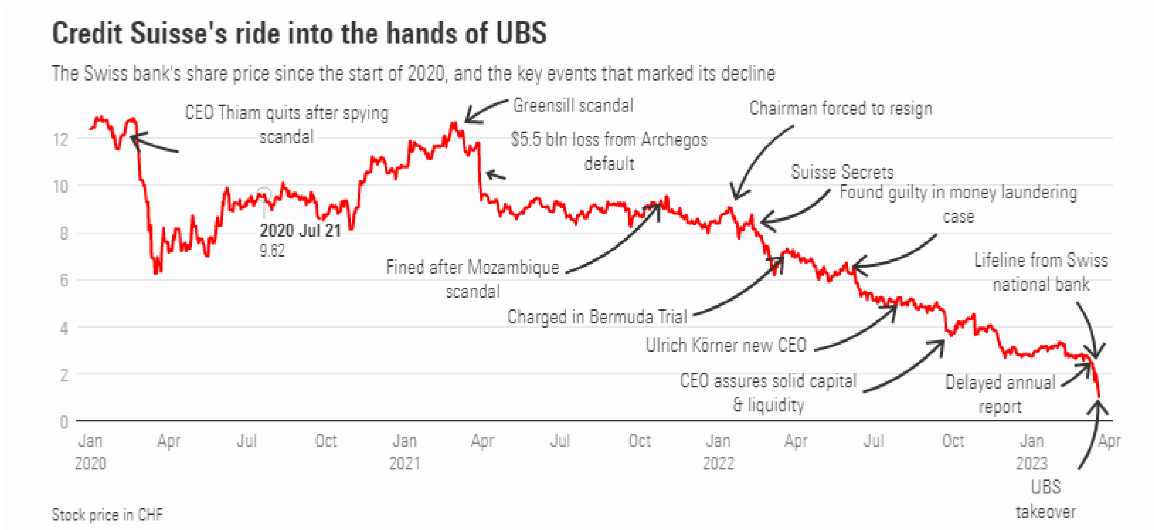


Figure 5.31: Events Chronology at Credit Suisse *(Source: Morningstar)*

failure of management and controls" in the investment bank and prime services business, which focused on short-term earnings and failed to rein in Archegos's risk-taking.

In October 2021, American and British authorities fined Credit Suisse $475 million for bribery regarding loans to Mozambican state-owned firms. The bank funded the People's Republic of Mozambique $1.3 billion for maritime surveillance, fisheries, and shipyard projects between 2012 and 2016, however bribes were taken. When the IMF discontinued financing, Mozambique's hidden Credit Suisse debt caused an economic disaster. Credit Suisse's governance, risk management, and compliance operations are vulnerable after these scandals and losses, requiring major reforms and cultural changes. After the Archegos and Greensill collapses in 2022, Lloyds Banking Group's former CEO Antonio Horta-Osorio became chairman to fix the bank. After nine months, Horta-Osorio resigned over Covid violations.

In February 2022, Credit Suisse's Suisse Secrets data breach revealed drug trafficking and money laundering on 18,000 bank accounts worth over $100 billion. Credit Suisse denies this. A former Credit Suisse senior banker's protracted deception cost Credit Suisse's local life insurance business $553 million in March. The bank was fined CHF 2 million ($2.1 million) in June for failing to stop Bulgarian cocaine traffickers from laundering money. In last October, a new leadership team revealed intentions to cut 9,000 jobs and raise $4 billion, including 9.9% from Saudi National Bank. Investment banking would be separated from First Boston's revitalization. The bank's worst yearly loss since the financial crisis was in February 2023 when client deposits totaled over CHF 110 billion ($119 billion). Results dropped the bank's shares 15%.

Due to these events, we can observe that the CRI-NUS PD index for Credit Suisse is rising slowly but steadily commencing in 2022, crossing and downgrading to **BB**

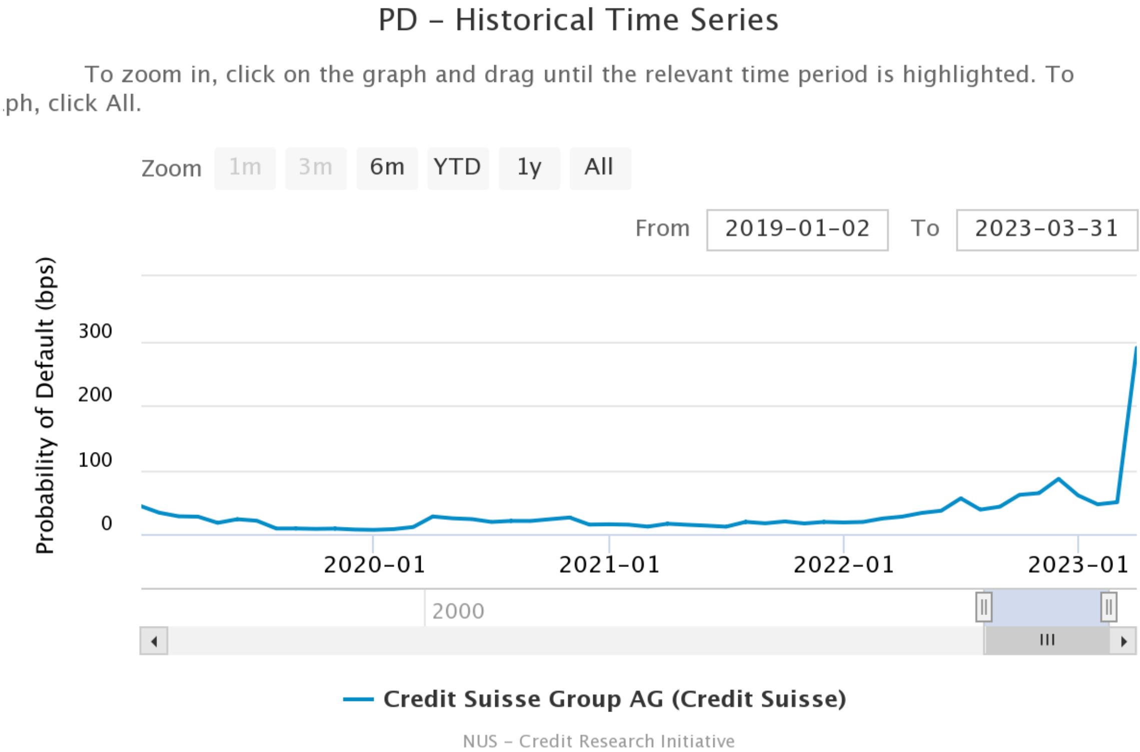


Figure 5.32: PD of Credit Suisse

credit rating equivalent to S&P rating [1] in the middle of the year. And the by the end of the year it's almost touching **B+** level which is really a critical level. By this point a lot of major credit rating agencies has downgraded it to lower levels.

In March 2023, Credit Suisse's "material weaknesses" in financial controls caused clients to leave and funding prices to rise. Clients left despite a CHF 50 billion liquidity rescue from the Swiss National Bank. UBS took over Credit Suisse for CHF 3 billion ($3.23 billion) in equity, absorbing losses of up to CHF 5 billion ($5.4 billion) and erasing low-ranking bonds worth CHF16 billion ($17 billion). Shareholders received CHF 0.76, 59% less than the previous closure. At this point, we can see that the Credit Suisse was on the verge of default; consequently, the CRI-PD jumped abruptly after the main announcement, but it had been signaling a crisis since early 2022, when the PD was gradually increasing.

Figure 5.33 shows media was constantly showing pessimistic opinion about Credit Suisse from beginning of 2021. In the case of Credit Suisse, the use of sentiment analysis appears capable of bridging the information divide and supplementing the commonly used structural financial variables to enhance early credit warning.
Credit Suisse and other cases of a similar nature (it has already shown by CRI regarding a famous case of *Wirecard*) compel us to seek a more effective method for

[1] refer appendix for The Probability of Default Implied Rating (PDiR) which supplements the CRI-PD by providing a convenient and intuitive overview of a company's credit quality by translating the CRI PD to letter grades employed by major rating agencies.

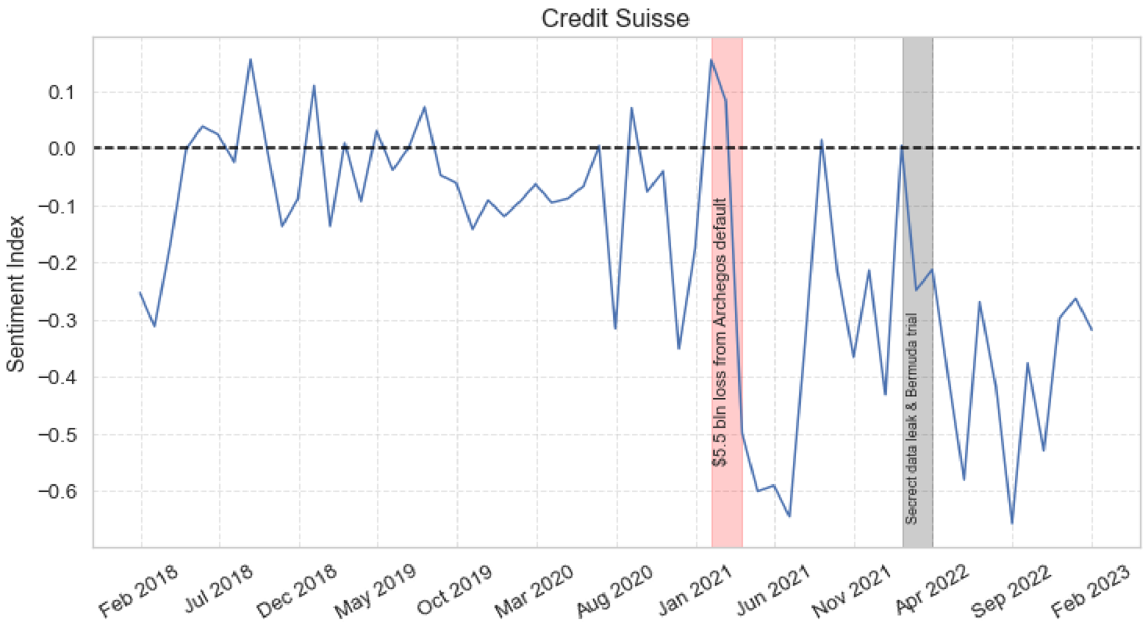


Figure 5.33: Credit Suisse Sentiment Index

detecting increased credit risks in companies that engage in fraudulent accounting and market manipulation. When analyzing cases involving fraudulent accounting practice, the current NUS CRI PD model, which relies exclusively on structural financial variables derived from financial statements and market data, is an inadequate credit assessment instrument. Financial forgeries, regardless of how deftly they are devised, may leave traces that arouse suspicion and are caught up by the media. The media sentiment score offers a glimmer of optimism that a quantitative system such as the NUS-CRI PD model can be practically enhanced to detect early credit deterioration that has been purposefully concealed by perpetrators of financial frauds.

6

Conclusions

The subsequent chapter concludes the investigation and presents the results obtained. In addition, the results are presented in a comparison table for easy perusal. The chapter concludes with a discussion of the implications of the thesis's findings and suggestions for future research.

6.1 Summary of Panel Regression Study

As previously stated, the objective was to assess the effect of sectoral risk, macro-prudential policies, market crash, and GDP growth on lowering India's sectoral credit risk across all sectors by using a panel model. We followed benchmark regression by Cerutti et al. [14] periodically adding a few more variables. 2000-2017 was studied for the Yearly case, and 2006-2020 for the Monthly case. We split the monthly example into two halves, one of which includes the whole time period specified, while the other begins in 2009, bypassing the subprime crisis phase. We also find evidence of strong one directional causality from Sectoral Risk to Credit Cycle. The objective was to examine the impact of the policy in two distinct time periods. Figure 6.1 below summarizes our primary results.

6.2 Summary of VAR Study

Credit Channel Theory [39] posits that fluctuations in the availability of credit may have a substantial effect on the economy and various sectors. According to this idea, as credit conditions tighten, credit-dependent industries, such as construction and real estate, may see a sharper fall in economic activity and an increase in credit risk.

According to the **Minsky Financial Instability** Hypothesis [12], financial crises are an inherent characteristic of a capitalist economy, and times of stability may lead to overconfidence and greater risk-taking. In the context of sectoral credit risk, this theory indicates that particular industries, including banking and technology, may be more susceptible to speculative excesses and financial instability. This may enhance the credit risk in these industries during times of economic duress.

Independent Variables	Effect on Credit Risk (CCI/Dependent Variable) Annual Case	Effect on Credit Risk (CCI) Monthly case	
		Including Subprime Crisis era	Excluding Subprime Crisis era
Sectoral Risk	Increased until two years lag then decreased in 3rd year.	Increased	Increased but 26% reduction from subprime era
Macroprudential Policies	Increased with one year lag	Increased with 12months lag	Increased with 18months lag
GDP Rate	Decreased but not significant	Decreased but not significant	Decreased with 16months lag
Policy Rate	Increased with one year lag	Increased with 11months lag	Increased but not significant
Market Crash	Decreased with one year lag	Increased with 11months lag	Decreased with a 13months lag

Figure 6.1: Impact of independent variables on sector-level Credit Risk Containment

Sector	Credit Channel Theory	Financial Instability Theory
Energy	Sensitive to interest rate changes due to high capital costs	Vulnerable to financial fragility due to high leverage and risk-taking
Commodities	Sensitive to interest rate changes due to high capital costs	Vulnerable to financial fragility due to high leverage and risk-taking
Industrials	Sensitive to interest rate changes due to high capital costs	Vulnerable to financial fragility due to high leverage and risk-taking
Consumer Discretionary	Sensitive to interest rate changes due to high capital costs	Vulnerable to financial fragility due to high leverage and risk-taking
Health Care	Sensitive to interest rate changes due to high capital costs	Less vulnerable due to stable demand and lower leverage
Banking	Sensitive to interest rate changes due to interest income	Prone to instability due to high leverage and interconnectedness
Information Technology	Less sensitive to interest rate changes due to high growth	Less vulnerable due to high profitability and lower leverage
Telecommunication	Less sensitive to interest rate changes due to stable demand	Less vulnerable due to stable demand and lower leverage
Utilities	Less sensitive to interest rate changes due to stable demand	Less vulnerable due to stable demand and lower leverage
Real Estate	Sensitive to interest rate changes due to high capital costs	Prone to instability due to high leverage and asset price fluctuations

Table 6.1: Credit Channel Theory & Financial Instability Theory explanation of Credit requirements in different Indian Sectors

Both the credit channel theory and Minsky's financial instability theory provide insights into how sectoral credit risk may vary. Thus, it is essential to study both theories in order to comprehend how sectoral credit risk might vary. The credit channel hypothesis highlights the significance of bank lending in conveying monetary policy shocks to the real economy and implies that particular industries may be more susceptible to changes in bank lending conditions. In the meanwhile, Minsky's theory of financial instability emphasizes the potential for financial fragility to arise in specific sectors during times of credit growth, which might result in an increase in credit risk for such sectors. By studying these ideas, analysts may get a deeper understanding of the dynamics of credit risk in various industries and formulate more educated risk management methods.
We use the Vector Auto Regression Model with three variables Repo Rate, Credit Cycle Index, and Sectoral Risk Index to grasp the interrelationships between these variables, especially the latter two, and to anticipate CCI and Sectoral Risk for the next 5-6 years. The findings pertaining to sectoral forecasting have been summarized

Sector	Credit Risk Forecast Short Term	Credit Risk Forecast Long Term	Sectoral Risk Forecast Short Term	Sectoral Risk Forecast Long Term
Energy	Increase	Slight Decrease from short term but increased from base level	Slight Decrease	Increase
Commodity	Decrease	Increase	Increase	Increase
Industrials	Decrease	Increase	Increase	Increase
Consumer Discretionary	Decrease	Slight Increase from short term but overall stays neutral	Decrease	Very Slight Increase
Financial Services	Neutral	Neutral	Decrease	Slight Increase
Bank	Decrease	Increase	Decrease	Increase
Information Technology	Neutral	Very Slight Increase	Decrease	Increase
Healthcare	Decrease	Increase	Neutral	Neutral
Communication Services	Decrease	Increase	Very Slight Decrease	Increase
Utilities	Increase	Increase	Increase	Increase
Real Estate	Increase	Increase	Increase	Increase

Table 6.2: Credit Risk and Sectoral Risk Forecast Tab

in Table 6.2 for ease of reading.

6.3 Summary of NLP Study

Main effort of this part was the creation of a real-time assessment instrument of publicly traded companies regarding Credit Risk and greenness sentiment. Using topic modeling and sentiment analysis techniques, the tool analyzes a corpus of data from more than 40 RSS feeds of business press. Each publicly traded company's sentiment assessment is computed in real time and displayed on a website. The sentiment index is also calculated using hourly updated sentiment scores. The sentiment score is derived from a five-class TABSA-BERT classification, and sentiment is aggregated using weights derived from Source LDA and the BERT classification score. The website for NS Monitor presents the most recent articles, monitored entities, and RSS feeds. The Credit Risk and Greenness sentiment scores are displayed for each article & each entities, and users can enter the name of a publicly traded company in the search field to obtain sentiment scores for a variety of equities. The tool also gives users the option to retrieve all articles pertaining to a specific company or topic, as well as view the entity's historical moving average sentiment score similar to the various stock indices of publicly traded companies. Overall, the instrument is a valuable resource for investors and analysts interested in real-time monitoring of Credit Risk and Greenness sentiment for publicly traded companies.

6.4 Recommendations for Further Research

By analyzing the interaction between credit risk and sectoral risk, we identified the methodological limitations of panel data construction. As our yearly scenario has bias-related difficulties, one may go further into the different estimators and provide more accurate estimates. Whilst in the monthly example we are not subject to bias-related issues, it is clear that the majority of the variables included in our experiments served as controls. To bring greater heterogeneity into a dataset, one might attempt to include more sector-specific variables.

And sectoral study is not the limit; using the concept of construction of CCI and Sectoral Risk Index, one can delve into numerous studies, such as between different market capitalizations, between government-owned firms and private firms, and between various thematic portfolios; the possibilities are limitless.

By examining the effect of many variables on credit risk and sectoral risk, more research ideas may improve upon the current body of knowledge. Although specific variables, such as legislative changes and technical advances, have been researched in isolation, there is a dearth of study on how the interaction of many factors impacts credit risk and sectoral risk in diverse industries. The proposed study will thus investigate the combined influence of legislative changes, technical improvements, ESG variables, political events, and risk management tactics on credit risk and sector risk.

The research might examine the effect of regulatory reforms, like as Basel III and Dodd-Frank, on credit risk and sectoral risk, as well as their interaction with other variables. Changes in legislation, for instance, may have unanticipated effects on credit risk and sectoral risk, which may be exacerbated by other variables such as political events or technology breakthroughs. Similarly, technology advancements like blockchain and artificial intelligence may have ramifications for credit risk and sectoral risk, which may be impacted by other variables like ESG concerns or legislative changes.

Moreover, one may investigate how other variables affect the effect of ESG issues, such as climate change, labor standards, and board diversity, on credit risk and sectoral risk. For instance, legislative changes, technical advancements, or political events may worsen the effect of climate change on credit risk and sectoral risk.

Future study might evaluate the efficacy of various risk management measures, such as diversification and hedging, in lowering credit risk and sectoral risk in various industries. By examining the efficacy of various tactics and their interactions with other elements, the research might provide useful insights into how financial institutions can manage credit risk and sector risk.

Instead of sector-specific data, we significantly depend on country-specific control variables for our VAR study (or even panel model study). Several of these data are quite difficult to get. While we believe that certain country-specific control factors, such as the Repo Rate and the GDP Rate, will play a significant role in predicting Credit Risk and Sector Risk, the addition of more sector-specific data will result in more robust findings.

The credit-focused, entity-specific sentiment at the article level for enterprises receiving media attention has much more predictive power than the conventional structured financial indicators. In addition, the emotion score may greatly improve the predictive capacity for various sorts of corporate defaults and other exits.

These NLP approaches may be relevant to various financial and business challenges, for which articles in the business press may give further information beyond what is already present in the regularly used structured variables. To meet the needs of various users, one may try to include specialized topics such as ESG, Gold, Bonds, and commodities, among others using the proposed algorithm.

References

1. Aikman, D., Haldane, A. G. & Nelson, B. D. Curbing the credit cycle. *The Economic Journal* **125,** 1072–1109 (2015).
2. Schularick, M. & Taylor, A. M. Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008. *American Economic Review* **102,** 1029–1061 (2012).
3. Groß-Klußmann, A., König, S. & Ebner, M. Buzzwords build momentum: Global financial Twitter sentiment and the aggregate stock market. *Expert Systems with Applications* **136,** 171–186 (2019).
4. Qiu, X. *et al.* Pre-trained models for natural language processing: A survey. *Science China Technological Sciences* **63,** 1872–1897 (2020).
5. Sun, C., Huang, L. & Qiu, X. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. *arXiv preprint arXiv:1903.09588* (2019).
6. Dunham, L. M. & Garcia, J. Measuring the effect of investor sentiment on liquidity. *Managerial Finance* **47,** 59–85 (2021).
7. Azqueta-Gavaldón, A. Causal inference between cryptocurrency narratives and prices: Evidence from a complex dynamic ecosystem. *Physica A: Statistical Mechanics and its Applications* **537,** 122574 (2020).
8. Duan, J.-C. & Zhu, Y. Credit risk cycle indices-properties and macroprudential policy. *National University of Singapore working paper* (2020).
9. Drehmann, M., Borio, C. E. & Tsatsaronis, K. Characterising the financial cycle: don't lose sight of the medium term! (2012).
10. Mouatt, S. Credit cycles: freewheeling, driving or driven? *International Journal of Social Economics* (2015).
11. Kiyotaki, N. & Moore, J. Credit cycles. *Journal of political economy* **105,** 211–248 (1997).
12. Minsky Ph D, H. P. The financial-instability hypothesis: capitalist processes and the behavior of the economy (1982).
13. Fendoğlu, S. Credit cycles and capital flows: Effectiveness of the macroprudential policy framework in emerging market economies. *Journal of Banking & Finance* **79,** 110–128 (2017).
14. Cerutti, E., Claessens, S. & Laeven, L. The use and effectiveness of macroprudential policies: New evidence. *Journal of Financial Stability* **28,** 203–224 (2017).

15. Barrell, R., Karim, D. & Macchiarelli, C. Towards an understanding of credit cycles: do all credit booms cause crises? *The European Journal of Finance* **26,** 978–993 (2020).
16. Donnelly, C. & Embrechts, P. The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. *ASTIN Bulletin: The Journal of the IAA* **40,** 1–33 (2010).
17. Demarta, S. & McNeil, A. J. The t copula and related copulas. *International statistical review* **73,** 111–129 (2005).
18. Aziz, S., Dowling, M., Hammami, H. & Piepenbrink, A. Machine learning in finance: A topic modeling approach. *European Financial Management* **28,** 744–770 (2022).
19. Duan, J.-C. & Yao, X. Media Sentiments for Enhanced Credit Risk Assessment.
20. Wood, J., Tan, P., Wang, W. & Arnold, C. *Source-LDA: Enhancing probabilistic topic models using prior knowledge sources* in *2017 IEEE 33rd International Conference on Data Engineering (ICDE)* (2017), 411–422.
21. Blei, D., Ng, A. & Jordan, M. Latent dirichlet allocation. *Advances in neural information processing systems* **14** (2001).
22. Bornhorst, F. & Baum, C. LEVINLIN: Stata module to perform Levin-Lin-Chu panel unit root test (2006).
23. Zellner, A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association* **57,** 348–368 (1962).
24. Neyman, J. & Scott, E. L. Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society,* 1–32 (1948).
25. Ruud, P. A. *et al.* An introduction to classical econometric theory. *OUP Catalogue* (2000).
26. Nickell, S. Biases in dynamic models with fixed effects. *Econometrica: Journal of the econometric society,* 1417–1426 (1981).
27. Hoechle, D. Robust standard errors for panel regressions with cross-sectional dependence. *The stata journal* **7,** 281–312 (2007).
28. Breitung, J., Kripfganz, S. & Hayakawa, K. Bias-corrected method of moments estimators for dynamic panel data models. *Econometrics and Statistics* **24,** 116–132 (2022).
29. Driscoll, J. C. & Kraay, A. C. Consistent covariance matrix estimation with spatially dependent panel data. *Review of economics and statistics* **80,** 549–560 (1998).
30. Newey, W. K. & West, K. D. Hypothesis testing with efficient method of moments estimation. *International Economic Review,* 777–787 (1987).
31. Arellano, M. & Bond, S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies* **58,** 277–297 (1991).
32. Dhaene, G. & Jochmans, K. Likelihood inference in an autoregression with fixed effects. *Econometric Theory* **32,** 1178–1215 (2016).

33. Bun, M. J. & Carree, M. A. Bias-corrected estimation in dynamic panel data models. *Journal of business & economic statistics* **23,** 200–210 (2005).

34. Kiviet, J. F. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of econometrics* **68,** 53–78 (1995).

35. Lütkepohl, H. *New introduction to multiple time series analysis* (Springer Science & Business Media, 2005).

36. Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society,* 424–438 (1969).

37. Granger, C. W. Some recent development in a concept of causality. *Journal of econometrics* **39,** 199–211 (1988).

38. Lopez, L. & Weber, S. Testing for Granger causality in panel data. *The Stata Journal* **17,** 972–984 (2017).

39. Bernanke, B. S. & Gertler, M. Inside the black box: the credit channel of monetary policy transmission. *Journal of Economic perspectives* **9,** 27–48 (1995).

40. Duan, J.-C., Sun, J. & Wang, T. Multiperiod corporate default prediction—A forward intensity approach. *Journal of Econometrics* **170,** 191–209 (2012).

41. Embrechts, P. & Hofert, M. Statistics and quantitative risk management for banking and insurance. *Annual Review of Statistics and Its Application* **1,** 493–514 (2014).

42. Alam, Z. *et al. Digging deeper–Evidence on the effects of macroprudential policies from a new database* (International Monetary Fund, 2019).

43. Brownlees, C. & Engle, R. F. SRISK: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies* **30,** 48–79 (2017).

44. Qi, P., Zhang, Y., Zhang, Y., Bolton, J. & Manning, C. D. Stanza: A Python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082* (2020).

45. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

46. Liu, Y. *et al.* Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).

47. Dumitrescu, E.-I. & Hurlin, C. Testing for Granger non-causality in heterogeneous panels. *Economic modelling* **29,** 1450–1460 (2012).

48. Hamilton, J. *Time series econometrics* 1994.

49. Becketti, S. *et al. Introduction to time series using Stata* (Stata Press College Station, TX, 2013).

Appendix A

Additional Reference and Selected Test Results

A.1 Selected Forecast Error Variance Decomposition and Granger Causality

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
RepoRate	LCCI	.39597	2	0.820
RepoRate	LCVaR	8.9847	2	0.011
RepoRate	ALL	11.84	4	0.019
LCCI	RepoRate	3.1066	2	0.212
LCCI	LCVaR	38.53	2	0.000
LCCI	ALL	42.56	4	0.000
LCVaR	RepoRate	11.334	2	0.003
LCVaR	LCCI	3.8272	2	0.148
LCVaR	ALL	13.055	4	0.011

(a) Granger Causality

Results from Commodities

Step	(1) fevd	(2) fevd	(3) fevd
0	0	0	0
1	.14837	.01472	.83691
2	.431829	.01287	.555301
3	.410357	.054968	.534675
4	.407279	.058345	.534376
5	.409354	.057756	.53289

(1) irfname = Commodities, impulse = RepoRate, and response = LCVaR.
(2) irfname = Commodities, impulse = LCCI, and response = LCVaR.
(3) irfname = Commodities, impulse = LCVaR, and response = LCVaR.

(b) FEVD response CVaR

Results from Commodities

Step	(1) fevd	(2) fevd	(3) fevd
0	0	0	0
1	.038978	.961022	0
2	.033645	.689095	.27726
3	.096643	.3716	.531757
4	.251681	.279662	.468657
5	.265827	.288104	.446069

(1) irfname = Commodities, impulse = RepoRate, and response = LCCI.
(2) irfname = Commodities, impulse = LCCI, and response = LCCI.
(3) irfname = Commodities, impulse = LCVaR, and response = LCCI.

(c) FEVD response CCI

Figure A.1: Commodity Sector

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
RepoRate	LCCI	1.1198	2	0.571
RepoRate	LCVaR	7.0693	2	0.029
RepoRate	ALL	11.937	4	0.018
LCCI	RepoRate	10.485	2	0.005
LCCI	LCVaR	27.446	2	0.000
LCCI	ALL	37.093	4	0.000
LCVaR	RepoRate	8.5672	2	0.014
LCVaR	LCCI	3.4076	2	0.182
LCVaR	ALL	9.6981	4	0.046

(a) Granger Causality

Results from CD

Step	(1) fevd	(2) fevd	(3) fevd
0	0	0	0
1	.157208	.15983	.682962
2	.367957	.11001	.522033
3	.357285	.123278	.519437
4	.361616	.122617	.515767
5	.370238	.120914	.508847

(1) irfname = CD, impulse = RepoRate, and response = LCVaR.
(2) irfname = CD, impulse = LCCI, and response = LCVaR.
(3) irfname = CD, impulse = LCVaR, and response = LCVaR.

(b) FEVD response CVaR

Results from CD

Step	(1) fevd	(2) fevd	(3) fevd
0	0	0	0
1	.067964	.932036	0
2	.09017	.540941	.368889
3	.114459	.363912	.521629
4	.30615	.267245	.426606
5	.30878	.27418	.41704

(1) irfname = CD, impulse = RepoRate, and response = LCCI.
(2) irfname = CD, impulse = LCCI, and response = LCCI.
(3) irfname = CD, impulse = LCVaR, and response = LCCI.

(c) FEVD response CCI

Figure A.2: Consumer Discretionary Sector

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
RepoRate	LCCI	.7079	2	0.702
RepoRate	LCVaR	7.2137	2	0.027
RepoRate	ALL	10.553	4	0.032
LCCI	RepoRate	10.088	2	0.006
LCCI	LCVaR	7.3998	2	0.025
LCCI	ALL	17.329	4	0.002
LCVaR	RepoRate	6.9103	2	0.032
LCVaR	LCCI	4.3987	2	0.111
LCVaR	ALL	13.511	4	0.009

(a) Granger Causality

Results from Energy

Step	(1) fevd	(2) fevd	(3) fevd
0	0	0	0
1	.024493	.008257	.96725
2	.235066	.00548	.759454
3	.214804	.110516	.67468
4	.212417	.125749	.661834
5	.262869	.121032	.616099

(1) irfname = Energy, impulse = RepoRate, and response = LCVaR.
(2) irfname = Energy, impulse = LCCI, and response = LCVaR.
(3) irfname = Energy, impulse = LCVaR, and response = LCVaR.

(b) FEVD response CVaR

Results from Energy

Step	(1) fevd	(2) fevd	(3) fevd
0	0	0	0
1	.023117	.976883	0
2	.036176	.733394	.23043
3	.131537	.617374	.251089
4	.177301	.540218	.282481
5	.240207	.498519	.261274

(1) irfname = Energy, impulse = RepoRate, and response = LCCI.
(2) irfname = Energy, impulse = LCCI, and response = LCCI.
(3) irfname = Energy, impulse = LCVaR, and response = LCCI.

(c) FEVD response CCI

Figure A.3: Energy Sector

A.2 Probability Implied Rating

Rating Category	Initial Assignment		Upgrade To		Downgrade To	
	lb (bps)	ub (bps)	lb (bps)	ub (bps)	lb (bps)	ub (bps)
AAA	0	0.0035	0	0.0027	-	-
AA+	0.0035	0.1044	0.0027	0.0035	0.1044	0.3060
AA	0.1044	0.3060	0.0035	0.1044	0.3060	0.4069
AA-	0.3060	0.4069	0.1044	0.3060	0.4069	1.2928
A+	0.4069	1.2928	0.3060	0.4069	1.2928	3.0646
A	1.2928	3.0646	0.4069	1.2928	3.0646	3.9506
A-	3.0646	3.9506	1.2928	3.0646	3.9506	9.9936
BBB+	3.9506	9.9936	3.0646	3.9506	9.9936	22.0796
BBB	9.9936	22.0796	3.9506	9.9936	22.0796	28.1227
BBB-	22.0796	28.1227	9.9936	22.0796	28.1227	46.2056
BB+	28.1227	46.2056	22.0796	28.1227	46.2056	82.3715
BB	46.2056	82.3715	28.1227	46.2056	82.3715	100.4544
BB-	82.3715	100.4544	46.2056	82.3715	100.4544	357.0556
B+	100.4544	357.0556	82.3715	100.4544	357.0556	870.2578
B	357.0556	870.2578	100.4544	357.0556	870.2578	1126.8589
B-	870.2578	1126.8589	357.0556	870.2578	1126.8589	1630.8764
CCC+	1126.8589	1630.8764	870.2578	1126.8589	1630.8764	2638.9113
CCC	1630.8764	2638.9113	1126.8589	1630.8764	2638.9113	3142.9287
CCC-	2638.9113	3142.9287	1630.8764	2638.9113	3142.9287	4449.8571
CC	3142.9287	8370.6423	2638.9113	7063.7139	4449.8571	8777.9817
C	8370.6423	10000	-	-	8777.9817	10000

Figure A.4: Mapping 10-day moving average 1-year CRI PD to the S&P experience