# ESSAYS ON BUSINESS ANALYTICS AND GAME THEORY

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ECONOMICS

MAY 2023

By
Minh Nguyen

Dissertation Committee:

Ruben Juarez, Chairperson
Peter Fuleky
Miroslav Gabrovski
Yuanzhang Xiao
Shirley Daniel, University Representative

© Copyright 2023
by
Minh Nguyen
All Rights Reserved

To my families, teachers, and friends.

# Acknowledgements

First and foremost I would like to deeply thank every member in my dissertation committee. I would especially like to thank my advisor, Dr. Ruben Juarez, for his inspiration, guidance and support during my PhD life. I would also like to acknowledge other members in the committee, including Dr. Peter Fuleky, Dr. Miroslav Gabrovski, Dr. Yuanzhang Xiao, and Dr. Shirley Daniel for guiding and supporting for years, especially during the dissertation stage. All members in the committee play crucial roles during my PhD journey, especially they guide me to write, encourage me and give me constructive feedback so that I can finish this dissertation. Some of them are my co-authors of one chapter in the dissertation. I am grateful to Dr. Shirley Daniel, Dr. Yuanzhang Xiao, Dr. Ting-Tsen Yeh, and Dr. Barbara DeBaryshe for giving me the opportunity to work as their RA during my last three years at UH Manoa. I thank Dr. Yuanzhang Xiao, Dr. Jonghyun Lee, and Dr. Susanne Still for being my first teachers in machine learning and deep learning at UH Manoa. I am deeply grateful to Dr. Ruben Juarez, Dr. Peter Fuleky, Dr. Shirley Daniel, Dr. Yuanzhang Xiao, Dr. Farzana Nasrin, Dr. Nori Tarui, and Dr. Barbara DeBaryshe for supporting me when I was on the 2022-23 job market.

I feel lucky and thankful to many people who support me a lot so that I can go to UH Manoa to do my PhD. I will never forget the help from Professor Cuong Le Van (PSE) and Professor Nguyen Thi Canh (VNU-HCM) who wrote letters of recommendations for me to directly help me being admitted to the UH Economics Department as a PhD student. I also want to thank other people who helped me during my PhD applications, especially Professor Pham Hoang Van (Baylor), Professor André Grimaud (TSE), and Dr. Pham Khanh Nam (UEH). Moreover, I would like to thank the Vietnam Center of Research in Economics, Management and Environment (VCREME), my VCREME's friends and instructors who helped me prepare the knowledge of mathematics and economics in several pre-PhD courses. I thank my teachers and my cohorts at VNU-HCM (UEL & HCMUS) and the Vietnam-Netherlands Program at UEH. I thank colleagues and students at UFM and HCMUSSH.

I would like to thank the Department of Economics and Department of Mathematics, UH Manoa, for providing Graduate Assistantship to me for my first three years at UH Manoa and very nice offices for me to study and work for years. I am thankful to the current Grad Chair (Dr. Nori Tarui) and former Grad Chair (Dr. Ruben Juarez) for supporting me a lot in the last six years. I am thankful to many professors who taught me at the UH Econ and Math Department in the last six years. I thank seminar organizers, including Ruben Juarez, Michael Roberts, Timothy Halliday, Lester Lusher, Teresa Molina, and Evan Gawlik. I thank Dr. Ruben Juarez, Dr. Ekaterina Sherstyuk, and Dr. Miroslav Gabrovski for supporting my third year paper. I especially feel thankful to Mindy Scordilis, the secretary of the UH Econ Department for her constant support of me on countless things.

I am grateful to many entities for providing and allowing me to join summer/winter schools and short courses including VIASM, VNU-HCM, VAST, AMS, and BFS during my PhD. I am thankful to instructors and speakers in these schools and short courses. I also want to thank instructors at Coursera and Datacamp who I never met in-person but who teach me invaluable knowledge about machine learning, data science, and financial engineering.

I am appreciative of many friends, colleagues, and teachers at UH Manoa for supporting me academically and non-academically especially Chi Diep, Hung Vo, Thuy Doan, Quynh Tran, Binh Le, Ngan Pham, Quang-Vu Bach, Ben Chu, Evan Gawlik, Malik Younsi, Vincent Chung, Hien Ha, Irvin Chang, Ikenna Nometa, Adib Rahman, Anna Jeong, Sanghum Cho, Ziyi Chen, Wilson Lau, Po-Cheng Chen, and Yu-Chieh Wu. I also would like to thanks many friends and colleagues from other institutions including Minh-Ly Lieu, Bang Nguyen, Nguyen Le Hoang Anh, Anh Tran, Dung Le, Duc Le, Yen Trinh, Diu Tran, Imelda, Lining Han, Tuan Ho Quoc, Ting-Sen Yeh, Trang Phan, Quynh Vo, Thuc Cao, Duc Thanh Son Vu, Van Thanh Nguyen, Thanh Ngo, Hoa Vu, Toan Nguyen, many friends at the Vietnamese Student Association in Hawaii, peers at East-West Center (EWC), and my wife and I's friends (especially chi Uyen, co Thu, anh Matt, Don, and David). These people helped me a lot along my PhD journey. I also want to thank my classmates (Pratistha, Kang, Rangrang, Sisi, Maya, Rachel, Dongkyu, Younghan, Cheol-Kon, Chuong, and Leila) and peers at the UH Econ Department for going together on this journey.

I am thankful to all faculty and staff at the UH Center on the Family (COF), especially Barbara and JoAnn, for supporting and helping me during my last year at UH Manoa. I thank my COF colleagues (Barbara, Javzan, Ivette, and Qi Chen) for collaborating and caring. I am thankful to EWC, EWC Housing, and EWC's staff for supporting me for five years when I was an EWC Student Affiliate. I thank International Baptist Ministries Hawaii for supporting food and the Holy Spirit. I also thank Food Vault Hawaii for supporting food. I thank Newman Center Hawaii for providing a peaceful place to come.

Finally, I want to thank my families for their unconditional support for a very long time. Especially, I feel so lucky to have my wife, Nga Pham, who loves, supports and shares countless experiences and memories with me in Vietnam and the US. And the appearance of our boy, Leomin Nguyen, during my PhD life is a precious gift to my wife and I.

# Abstract

This collection of essays explores several research topics in business analytics and game theory. The first essay, Chapter 2, applies text mining and machine learning to quantify the impact of natural disaster risk on firm performance. The second essay, Chapter 3, uses machine learning and the Merton model to predict corporate financial distress in a transition economy. The third essay, Chapter 4, investigates optimal investment strategies in finite and mean field games, considering the presence of risk-seeking agents and using a hyperbolic absolute risk aversion (HARA) utility function. Each chapter focuses on a specific type of shock, including climate shocks (Chapter 2), economic shocks (Chapter 3), and stochastic shocks (Chapter 4).

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1
# Introduction

This dissertation comprises three essays, each of which explores a research topic in business analytics or game theory with focusing on a specific type of shock.

## 1.1 Motivations

Shocks can exist in many forms. Each chapter in this dissertation explores a type of shock.

*Climate shocks.* Chapter 2 investigates natural disasters such as climate shocks and their effects on U.S. public firm performance. In this chapter, climate shocks are proxy for environmental shocks or environmental uncertainty. There are many studies in the literature examining (objective) environmental uncertainty/shocks (Downey et al. 1975; Swamidass and Newell 1987; Faucheux and Froger 1995; Kreiser and Marino 2002; López-Gamero et al. 2011; Yu et al. 2018) and perceived environmental uncertainty/shocks (Duncan 1972; Lewis and Harvey 2001; Freel 2005; Yu et al. 2018). This chapter contributes to the understanding of both objective and perceived environmental uncertainty in terms of climate shocks and its effects on the US firm performance. In this chapter, perceived natural disaster risk is a proxy for perceived environmental uncertainty.

*Economic shocks.* Chapter 3 predicts corporate financial distress of firms in a transition economy, Vietnam, under economic shocks (e.g. 2008-9 financial crisis, macroeconomic fluctuations, and the COVID-19 pandemic). There exist very few studies about financial distress prediction under shocks in Vietnam up to this point. Vo et al. (2019) predict financial distress prediction at the industry level during the global financial crisis 2007-2009 and the post-global financial crisis between 2010 and 2017. In contrast, we will focus on the firm-level data in this chapter.

*Stochastic shocks.* Brownian motions are considered as stochastic shocks. Chapter 4 will examine Brownian motions as stochastic shocks (or stochastic uncertainty) in the framework of a $n$-agent games and mean field games on the financial markets. The key new aspects in our setting are that there are presence of risk-seeking agents and relative performance motivation in the stochastic environment. Brownian motions are used extensively in the literature of optimal investment and portfolio selection to model the dynamics of wealth (Merton 1969; Liu 2000; Espinosa and Touzi 2015; Lacker and Zariphopoulou 2019).

## 1.2 General research question

The general research question of this dissertation is: *How do shocks affect the performance and behavior of economic agents (e.g. firms and individual investors)?*

## 1.3 Dissertation contributions

This dissertation contributes to the understanding of how shocks affect the performance and behavior of economic agents, including firms and individual investors. The three essays in this dissertation present empirical and theoretical findings on the relationships between climate shocks and firm performance (Chapter 2), economic shocks and corporate financial distress (Chapter 3), and stochastic shocks and investment behavior (Chapter 4).

Chapter 2 contributes to the empirical understanding of the relationship between climate shocks and firm performance. It also contributes to the understanding of the comparison among models performance using different machine learning techniques in predicting firm performance. Particularly, we propose a new way to measure the perceived natural disaster risk by using text mining. We also propose a new dictionary of words related to natural disasters and natural hazards. To the best of my knowledge, this chapter is the first study that measures the perceived natural disaster risk using Form 10-Ks. Moreover, this chapter contributes to the further understanding of the relationships between perceived natural disaster risk, government-reported damages of natural disasters and hazards, and firm performance. Furthermore, this chapter contributes to the understanding of the relative performance among several machine learning models (classification and regression trees or CART, neural works, and linear regression) in predicting firm performance under natural disaster risks.

Chapter 3 makes significant contributions to the limited literature on predicting corporate financial distress in a transition economy, Vietnam, during economic shocks. Specifically, this study provides empirical insights into predicting financial distress in a transition economy like Vietnam, where market conditions and data may not be as available or reliable as in many developed economies. It is the first study to employ three methods, including accounting-based, market-based, and machine-learning models, to investigate financial distress among public firms in Vietnam. Additionally, this study is the first to use the synthetic minority oversampling technique (SMOTE) to address biased results caused by imbalanced financial distress data in Vietnam. Another contribution is the provision of new and explicit data for the Merton model, which can be valuable for future research.

Chapter 4 contributes to the theoretical understanding of the behavior of risk-averse and risk-seeking agents on the financial markets. For the case of strictly concave utility function, We prove that there exists the Nash equilibrium for the $n$-agent games and the mean-field equilibrium for the mean-field games (MFG) in both exponential and power utility functions. Under some mild conditions, the equilibrium is unique. For the case of strictly convex utility function, we prove that there exists a unique corner solution in both $n$-agent games and MFG for both exponential and power cases. We also quantify the qualitative effects of personal and market parameters on the optimal investment strategies in both $n$-agent and mean field games.

## 1.4 The outline of the dissertation

The organization of this dissertation, aside from the introduction (Chapter 1) and conclusion (Chapter 5), consists of two main blocks: business analytics and game theory. The relationship between these two blocks and their related chapters is illustrated in a tree in the following figure.

Figure 1.1: The organizational diagram of two blocks of the dissertation

$$
\begin{tikzcd}
 & & & \text{Essays on business analytics and game theory} \arrow[dll] \arrow[dr] & \\
 & \begin{array}{c}\text{Business analytics}\\ \text{(Chapter 2, 3)}\end{array} \arrow[dl] \arrow[dr] & & & \begin{array}{c}\text{Game theory}\\ \text{(Chapter 4)}\end{array} \arrow[d] \\
\begin{array}{c}\text{Modeling}\\ \text{firm performance}\end{array} & & \begin{array}{c}\text{Corporate financial}\\ \text{distress prediction}\end{array} & & \begin{array}{c}\text{Finite and}\\ \text{mean field games}\end{array}
\end{tikzcd}
$$

This dissertation examines the performance and behavior of various economic agents, including both firms and individual investors.

Figure 1.2: The organizational diagram of agents in the dissertation

$$
\begin{tikzcd}
 & \text{Agents in the dissertation} \arrow[d] & \\
 & \text{Microeconomic agents} \arrow[dl] \arrow[dr] & \\
\begin{array}{c}\text{Firms}\\ \text{(Chapter 2, 3)}\end{array} & & \begin{array}{c}\text{Individual investors}\\ \text{(Chapter 4)}\end{array}
\end{tikzcd}
$$

# Chapter 2
# Natural disaster risk and firm performance: Text mining and machine learning approach

## Abstract

We develop a perceived measure of firms' disaster exposure and/or preparedness equal to the number of words related to natural disaster events in the firms' Form 10-Ks. We then link this measure to contemporary and future firm decision-making and performance. We find that this perceived natural disaster risk and the government-reported damages of natural hazards this year are negatively associated with firm profitability next year. However, the perceived natural disaster risk is not associated with sales growth and Tobin's Q ratio. Specifically, the perceived natural disaster risk negatively affects firm profitability in the services sector but not in the manufacturing sector. The firm profitability in the services sector is also negatively affected by the billion-dollar natural disasters in the same year. Finally, we find that advanced machine learning models robustly outperform linear regression in predicting firm performance under natural disaster risks. The main implication from this study is that we can employ textual data in financial reports to measure the perceived natural disaster risk and predict its effects on firm performance.

**Keywords**: Natural disaster risk, firm performance, Form 10-Ks, text mining, machine learning.

**JEL Codes**: C45, C53, L25, M21, Q51, Q54.

---

This is a joint work with Ting-Tsen Yeh, Yuanzhang Xiao, and Shirley Daniel. This paper is now under review at *International Review of Economics and Finance*. We would like to thank Ruben Juarez, Peter Fuleky, Miroslav Gabrovski, Michael Roberts, Max Biggs, Toan Phan, Marc Gronwald, Imelda Wang, Jonghyun Lee, Teresa Molina, Lester Lusher, Justin Tyndall, Ekaterina Sherstyuk, Chak Tan, Anh Tran, Sanghum Cho, Anna Jeong, Minh-Ly Lieu, Quy Ta, Hoa Vu, and participants at the Workshop on Energy and Environmental Research, Applied Miro Seminar, and Graduate Mathematical Student Seminar at the University of Hawaii at Manoa, the New Advances in Statistics and Data Sciences Conference, the 5th International Conference on Econometrics and Statistics, the 4th Advances in Decision Analysis Conference, the 2022 Asia Meeting of the Econometric Society, the 2022 Asian Meeting of the Econometric Society in East and South-East Asia, the 24th International Conference on Computational Statistics, the SIAM Conference on Mathematics of Data Science, the 2022 INFORMS Annual Meeting, the Hawaii Economic Association Graduate Student Showcase, University of Utah, and Washington State Employment Security Department for helpful comments and discussions. All errors are our own.

## 2.1 Introduction

Natural disasters and natural hazards are common phenomena that all public authorities, firms, and academics are concerned about. Academics might measure the effects (direction and magnitude) of the natural disasters and natural hazards on many outcomes such as damages, economic growth, or employment. Public authorities might try to find good strategies and guidance to help vulnerable communities and economic sectors to adapt and respond to natural disasters and hazards. Firms might want to identify factors and mechanisms how natural disasters and natural hazards affect their performance so that they can survive and thrive especially during and after each major weather and climate event. Therefore, a study on environmental uncertainty (natural disasters and natural hazards) is necessary for public authorities, firms, and academics.

This study aims to address several research gaps in the literature. The first research gap is that despite the increasing of number of the billion-dollar natural disaster events as well as its damages as seen in Figure 2.1 and Figure 2.2 respectively and the growing of number studies of the effects of natural disasters on supply chain (Ye and Abe 2012; Abe and Ye 2013), the literature of the relationship between natural disasters and firm performance is relatively limited and fragmented. Many studies just focus on one specific aspect of firms' activities such as firm investment (Hosono et al. 2016) or operating performance (Hsu et al. 2018). At the moment since the evidence of the relationship between natural hazards and disasters and firm performance is relatively limited, this study aims at providing further understanding about this relationship.

The second research gap is that even though the perception of natural hazards and disasters has long been studied in the literature, for example Burton and Kates (1963), the way to measure it mostly bases on the Likert-type scale questions. This type of scale might be subjective and might not be comparable.

Another research gap in the literature of firm performance is that most current empirical studies rely heavily on the linear regression method. This method, however, is not always the best tool for predicting firm performance. This study aims at going beyond linear regression by investigating systematically several common machine learning techniques to predict firm performance. Specifically, this paper will model firm performance under environmental uncertainty by using linear regression (baseline model) and several other advanced machine learning techniques (classification and regression trees or CART, neural networks, and linear regression). The reason to use CART and neural networks is that they are promising tools to capture the nonlinear interactions between natural disasters and firm performance and among other variables in this study. It is necessary to examine more machine learning techniques to have a comparative perspective of the performances among these techniques, which might help to select the suitable models in predicting firm performance.

This study aims to address these above gaps by answering the following three questions: (i) How to construct an alternative measure for the perceived natural disaster risk? (ii) How do this perceived risk of natural disasters and the government-reported of natural hazards and disasters affect firm performance? (iii) How are differences in performance among several machine learning models (linear regression, CART, and neural networks) in predicting firm performance under natural disaster risks?

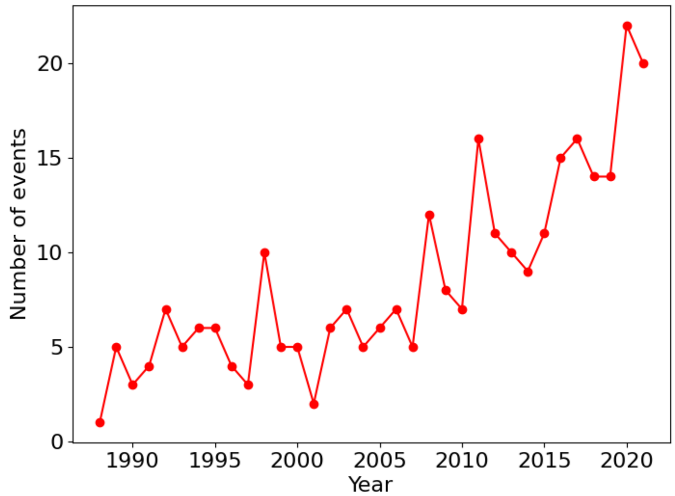
**Contributions**. This study makes the following primary contributions to the existing literature on natural disasters and firm performance. First, we propose a new way to measure the perceived risk of natural hazards and disasters (or perceived natural disaster risk for short) by using text mining. Second, in order to create a measure of perceived risk of natural disasters, we propose a new dictionary of words related to natural disasters disasters and hazards (see Table 2.11). Third, to the best of my knowledge, this is the first study measures the perceived natural disaster risk using Form 10-Ks. Four, this study contributes to the further understanding of the relationships between perceived natural disaster risk, government-reported damages of natural disasters and hazards, and firm performance. Five, this study contributes to the understanding of the relative performance among several machine learning models (CART, neural works, and linear regression) in predicting firm performance under natural disaster risks.

### 2.1.1 Motivations

#### An increase in number of billion-dollar natural disasters

The number of billion-dollar weather and climate disaster events in the U.S. between 1988 and 2021 is shown in Figure 2.1. In general, there is an upward trend of the number of billion-dollar natural disasters in the given period with the trough being in 1988 (1 disaster) and the peak is in 2020 (22 disasters). The average number of billion-dollar natural disasters is approximately 8.44 per year. Note that this upward trend is not affected by inflation since the damages are already adjusted to CPI in 2021. In this study, natural disasters are a proxy for environmental uncertainty. In the sense of frequency, we can consider the billion-dollar disasters as black swan events since they are rare and it is hard to predict when a disaster happening in a particular area.

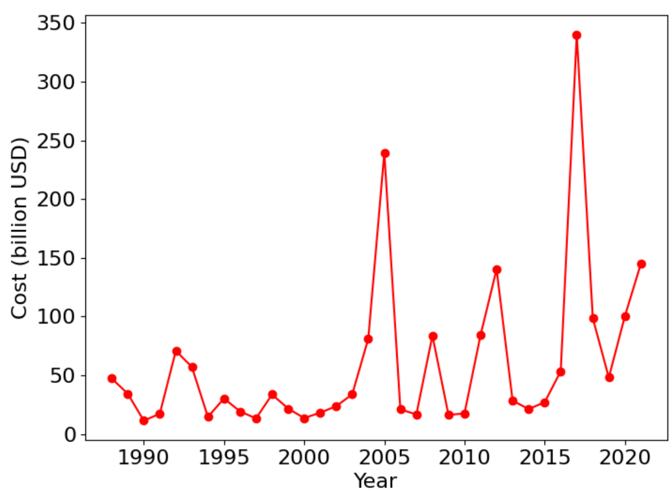
Figure 2.1: Number of billion-dollar natural disasters in the U.S. between 1988 and 2021




*Source*: The National Oceanic and Atmospheric Administration (NOAA)

Figure 2.2 illustrates the damages of the billion-dollar natural disasters adjusted to CPI in 2021 in the U.S. in the 1988-2021 period. We can observe that the second highest peak and the highest peak are in 2005 with the record-breaking hurricane Katrina[1] and in 2017 with hurricane Harvey, Maria, Irma, and the largest wildfire season ever in California. On average, the damage of billion-dollar disasters is $59.43 billion per year in the U.S. over the 1988-2021 period.

Figure 2.2: Damages of billion-dollar natural disasters in the U.S. between 1988 and 2021

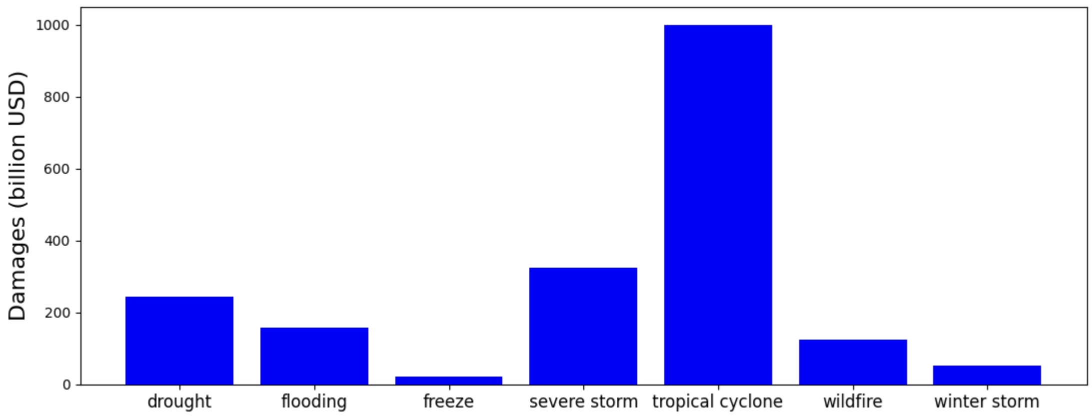



*Source*: NOAA

[1]Hurricane Katrina caused 1,833 deaths and $178.8 billion in damages. For more details, please see https://www.ncdc.noaa.gov/billions/events.

Note that natural disasters in this study include seven types of natural disasters which are tropical cyclones, severe storm, drought, flooding, wildfire, winter storm, and freeze. It can be seen that tropical cyclones are the main billion-dollar natural disasters in the United States.

Figure 2.3: The damages of seven types of natural disasters in the U.S between 1988-2021




*Source*: NOAA

Table 2.3 shows the damages of the U.S. natural hazards and U.S. billion-dollar natural disasters between 1988 and 2021. The differences between these damages are discussed more details in Section 2.3.1. Note that these damages are already adjusted to CPI in 2021. It can be seen that Texas, Louisiana, Florida, and California are the four most affected states by the damages of both natural hazards and billion-dollar natural disasters in this period.

Table 2.1: Damages of the U.S. natural hazards and billion-dollar natural disasters by state between 1996 and 2021 (unit: $ billion)

| State | Damage (hazards) | Damage (disasters) | State | Damage (hazards) | Damage (disasters) |
|---|---|---|---|---|---|
| AK | 0.5306 | 1.7800 | MT | 0.4548 | 9.0050 |
| AL | 14.7559 | 43.1300 | NC | 19.4234 | 79.3075 |
| AR | 6.7640 | 13.7600 | ND | 8.9996 | 13.4275 |
| AZ | 5.3592 | 5.6500 | NE | 7.8058 | 20.0050 |
| CA | 42.0547 | 110.0250 | NH | 0.4074 | 1.4925 |
| CO | 9.8094 | 30.6775 | NJ | 32.8633 | 66.1300 |
| CT | 0.3078 | 6.6125 | NM | 3.4049 | 5.5775 |
| DE | 0.2454 | 2.1675 | NV | 1.5264 | 2.4725 |
| FL | 77.1236 | 172.8375 | NY | 6.7471 | 63.1825 |
| GA | 10.8816 | 30.5525 | OH | 8.9132 | 22.7075 |
| HI | 0.3525 | 0.0000 | OK | 11.6543 | 33.4800 |
| IA | 15.1654 | 39.1650 | OR | 5.5070 | 11.4050 |
| ID | 1.2038 | 4.9775 | PA | 5.6403 | 23.3850 |
| IL | 7.3788 | 34.8275 | RI | 0.1615 | 1.5200 |
| IN | 3.9683 | 19.0575 | SC | 2.2795 | 18.4800 |
| KS | 4.9935 | 23.9325 | SD | 2.1579 | 13.3050 |
| KY | 5.0951 | 15.7875 | TN | 9.8923 | 29.9775 |
| LA | 119.7575 | 258.3025 | TX | 147.8666 | 342.3525 |
| MA | 1.1323 | 5.0600 | UT | 1.2495 | 1.8175 |
| MD | 1.9651 | 13.2100 | VA | 4.0211 | 15.6600 |
| ME | 0.8270 | 1.5100 | VT | 1.9174 | 1.7625 |
| MI | 6.9556 | 8.2100 | WA | 4.0305 | 6.1600 |
| MN | 7.3728 | 19.9275 | WI | 5.8471 | 14.1550 |
| MO | 10.9232 | 33.2300 | WV | 1.7502 | 3.0300 |
| MS | 44.3708 | 51.9075 | WY | 0.2957 | 3.2925 |

*Notes*: For the U.S. billion-dollar natural disasters (disasters for short), NOAA only provides the estimation of the interval of damages for each state every year. I take the average damages of each interval in each year, then summarize all of these averages to get the total damages for each state in this table. For the U.S. natural hazards, the damages are not adjusted to CPI. Therefore, they are adjusted to CPI in 2021 to be consistent with that of the U.S. billion-dollar natural disasters.

### The revolution of textual data

At the firm level, textual data is increasingly widely used. Recently, there are many studies using textual data to examine many aspects of business such as corporate disclosures (Li 2011), firm performance prediction (Qiu et al. 2014), bankruptcy prediction (Mai et al. 2019), corporate innovation (Bellstam et al. 2021), product sales (Li et al. 2019), and corporate culture (Li et al. 2021).

The textual data obtained from Form 10-Ks seems a good channel which reflects firm performance. Cohen et al. (2020) find that "changes to the language and construction of financial reports have strong implications for firms' future returns and operations".

[1]See more details at https://www.businessinsider.com/world-cup-favorites-2014-4

Specifically, they show that the changes to Form 10-Ks predict the firm's future earnings and profitability. One of the first textual-data studies using Form 10-Ks is the seminar paper (Loughran and McDonald 2011). Specifically, they develop a word list that reflects the tone in financial text and link it to 10-Ks filings. There are several other studies using Form 10-Ks such as (Kang et al. 2018; Ladas 2021).

There are a number of previous studies using the qualitative Management Discussion and Analysis (MD&A) section, which is one of the richest information sections in Form 10-Ks, to explore bankruptcy prediction (Mai et al. 2019), company's earnings (Feldman et al. 2010; Bochkay and Levin 2019), firm investment (Cho and Muslu 2021), and corporate credit rating prediction (Choi et al. 2020). Our study take into account the MD&A Section as well as other sections in the Form 10-Ks.

### 2.1.2 Related literature

**Objective and perceived environmental uncertainty**

Environmental uncertainty is an object of many fields including economics of disasters, environmental economics, environmental management, strategic management, organization theory, microeconomics, and business economics. We will review two different kinds of environmental uncertainty which are actual environmental uncertainty and perceived environmental uncertainty.

Objective environmental uncertainty (OEU), in the context of organizational environment, has long been examined in the management literature (e.g. Downey et al. 1975; Swamidass and Newell 1987; Faucheux and Froger 1995; Kreiser and Marino 2002; López-Gamero et al. 2011; Yu et al. 2018). However, the literature of environmental uncertainty in the context of the natural environment is fairly limited. Many previous studies do not have explicit ways to measure environmental uncertainty either in organizational or environmental context. This study empirically uses the government-reported damages of natural disasters as the proxy of objective environmental uncertainty.

Regarding perceived environmental uncertainty (PEU), there are two types of PEU. The first type of PEU is in the sense of the *organizational* (or *business*) environment and it has long been studied in the fields of strategic management and organization theory (e.g. Duncan 1972; Downey et al. 1975; Miles et al. 1978; Miller 1993; Freel 2005). Meanwhile, the second type of PEU is in the sense of the *natural* environment, which is studied much less and much later compared to the first type of PEU. Indeed, the first research investigating PEU in the natural environment is Lewis and Harvey (2001), which is thirty-four years later compared to the first research about PEU in the organizational environment by Lawrence and Lorsch (1967). This study, therefore, focuses on the second type of PEU to fulfill the gap. Let's call the second type of PEU *perceived natural environmental uncertainty* (or PNEU). Note that in contrast to many previous studies in the organization theory literature about PEU using self-report questionnaire (e.g. Lawrence and Lorsch 1967; Tosi et al. 1973; Tung 1979; Miller 1993; Lewis and Harvey 2001), this study employs the text mining technique.

There are several studies examining both OEU and PEU in the context of organizational environment (e.g Downey et al. 1975; Milliken 1987; Yu et al. 2018). It seems that most studies explore both OEU and PEU in the context of organizational environment. This study, however, focuses on objective and perceived environmental uncertainty in the context of *natural* environment. In this paper, the government-reported damages of natural hazards and disasters are proxy for the objective natural environmental uncertainty (ONEU) and the perceived natural hazard risk is proxy for the perceived natural environmental uncertainty (PNEU).

**The effects of natural disasters on economy and firms**

There are two ways that natural disasters affect the economy through two channels: observed damages and perceived measure. For the first channel, the literature shows limited and mixed results about the effects of observed damages of natural disasters on the economy at the macro and micro level. It seems that the effects of natural disasters on both the economy (at the macro level) and firms (at the micro level) are insignificant.

At the macro level, the neoclassical growth model anticipates that the growth effect is negative in the very short run and the growth is temporarily higher than the balanced growth after that (Felbermayr and Gröschl 2014). The previous macroeconomic empirical studies seem to support this result. Indeed, natural disasters generally have negative effects on the direct costs such as lost lives or physical destruction and an insignificant impact on the indirect costs (consequences physical destruction) (Botzen et al. 2019; Lazzaroni and Bergeijk 2014). Also, since the proportion of indirect costs increases in larger disasters and might constitute a larger fraction of total costs in large disasters than in small disasters (National Research Council 1999), the total impact of natural disasters might be insignificant. For example, Noy and Vu (2010) find evidence in a developing country, Vietnam, that more deadly disasters result in lower output growth, destroy more property and capital but these negative effects appear in turn to boost the economy in the short-run. They argue that these outcomes can be supported by the so-called 'investment-producing destruction' hypothesis.

At the firm level, natural disasters might have negative effects (e.g. Hsu et al. 2018; Huang et al. 2018; Pankratz et al. 2023), positive effects (e.g. Noth and Rehbein 2019) or mixed effects on firms (e.g. Zhou and Botzen 2021; Leiter et al. 2009). On the one hand, Hsu et al. (2018) find that firms with factories located in states highly affected by natural disasters are much less profitable than states less affected. Also, firms located in countries with more severe weather are likelier to hold more cash and less likely to distribute cash dividends

(Huang et al. 2018). Moreover, Pankratz et al. (2023) find that increasing exposure to extremely high temperatures reduces firms' revenues and operating income. On the other hand, Noth and Rehbein (2019) examine firms' outcomes after a major flood in Germany in 2013 and find that firms located in the disaster areas have significantly higher turnover, lower leverage, and higher cash in the period after 2013. Furthermore, Zhou and Botzen (2021) investigate the effects of natural disasters on the firm level in Vietnam. They find that flooding increases labor and capital growth but reduces sales growth notably up to three years after flooding. Also, Leiter et al. (2009) find that, in the short run, floods lead to an increase in total assets and employment growth of firms in regions affected by flooding more than firms in regions unaffected by flooding. However, the productivity of the former firms is negatively affected by flooding. One explanation is that the increase in investments in assets and employment offset the damaged production capabilities. This research will provide further evidence on whether natural disasters have a significant effect on firm performance or not and whether the direction of effect is positive or negative.

Regarding the second channel, perceived measure, the natural disasters might affect the behavior and decisions of the firm's executives and managers and firm's related agents (e.g investors). Several previous studies examine the reactions of executives, managers, and investors to natural disasters (McKnight and Linnenluecke 2017; Alok et al. 2020; Huynh and Xia 2022; Huang et al. 2022).

### 2.1.3 Comparison among machine learning techniques

One of the main goals of this study is to compare the modeling performances under uncertainty using different machine learning methods. Some comparisons in the literature are as follows.

**Linear regression vs. CART**. There are relatively few papers comparing the performances of linear regression and CART models in comparison to other techniques like ANN with linear regression. In health science, Razi and Athappilly (2005) use linear regression, neural networks (NNs), and CART to forecast the number of days in bed due to illness. They find that NNs and CART provide better prediction accuracy than linear regression model[2]. In finance, Zhu et al. (2011) propose a combination of CART and logistic regression for stock ranking. The work of Barth et al. (2023) implies that CART models are a good tool to investigate the annual relations between share price and accounting amounts (amount of accounting information).

**Linear regression vs. Neural networks**. As opposed to the class of linear regressions, neural networks can be classified as the multivariate nonlinear nonparametric models (White 1989; Zhang et al. 1998). Neural networks have some advantages compared to traditional

[2]Even though the authors mentioned that the regression model in their paper is nonlinear. They are in fact linear in parameters.

econometric methods. One of the advantages of neural networks over model-based methods like linear regression is that neural networks are data-driven methods which need very few assumptions about the model itself[3]. Since neural networks are not restricted by many assumptions they are more flexible then the model-based methods like linear regression. As a result, in many complicated problems in which we do not have good models, neural networks are feasible and promising ways for modeling as long as we have enough reliable data. Another advantage is that neural networks (and also many other machine learning techniques) have strong power of prediction. This is because, with a neural networks model, one usually splits data into two parts: train data and test data[4]. One uses train data to build the training machine learning models (e.g neural networks, CART, and linear regression), then validating this model with unseen test data. By the nature that neural networks always work with unseen data, they are attractive and promising for forecasting tasks (Sharda and Patil 1992; Swanson and White 1995; Adya and Collopy 1998; Huang et al. 2007) and management science (Sharda 1994). Further, neural networks perform well with high noise data (Marquez et al. 1991).

The prediction ability of neural networks has improved recently relative to traditional methods like linear regression. Early forecasting studies observe that traditional methods including linear regression or vector autoregressions perform better (lower mean squared errors) than neural networks (e.g. Swanson and White 1995, 1997) or equivalent to and sometimes better than neural networks (e.g. Hill et al. 1994). Moreover, these studies also argue that neural networks appear to be promising models for forecasting even though further refinement of neural networks is needed. However, more recent studies in the literature find that neural networks outperform linear regression models in prediction (e.g. Hill et al. 1996; Desai and Bharati 1998; Fadlalla and Lin 2001; Anyaeche and Ighravwe 2013; Pombeiro et al. 2017). Li and Ma (2010) discover that neural networks are valuable tools for forecasting in financial economics due to the learning, generalization, and nonlinear behavior properties. Anyaeche and Ighravwe (2013), in a study of profitability forecasting, find that neural networks have MSE lower than multiple-linear regression model. Further, Pombeiro et al. (2017) find that neural networks models (together with fuzzy model) outperform linear regression in predicting electricity consumption. For a systematic review of neural networks in business in the last two decades, please read Tkac and Verner (2016).

A noticeable point of neural networks is that simply increasing the number of hidden layers may not improve the performance of the neural networks much when working with the high-frequency financial data (Chen et al. 2017).

[3] One typical assumption of neural networks is that the neural networks are *fully* connected. Of course, one can relax this assumption if necessary.

[4] In many cases, one splits data into three parts: train data, validation data, and test data. Also note that in the old days, economists did not call train data and test data. Rather, they call it in-sample and out-of-sample data (e.g. Swanson and White 1995, 1997; Altay and Satman 2005).

**CART vs. Neural networks**. Yildiz et al. (2017) employ several machine learning models including support vector regression (SVR), regression trees (RT or CART), artificial neural networks (ANNs), nonlinear autoregressive network with exogenous inputs (NARX), and Multivariate regression model (MLR) for commercial building electricity load forecasting. They find that ANNs and NARX relatively perform better than other techniques.

### 2.1.4 Lagged dependent variables in machine learning models

Lagged dependent variables have long been studied extensively in econometrics of regression in the forms of autocorrelation or serial correlation (e.g. Taylor and Wilson 1964; Durbin 1970; Godfrey 1978; Inder 1984; Arellano and Bond 1991). They play important roles in regression in economics and other social sciences. For example, Wilkins (2018) argues that including the additional lagged dependent variables leads to more accurate parameter estimates in regression. Moreover, Keele and Kelly (2006) find that adding lagged dependent variables remains an appropriate model used to estimate dynamic phenomena. In the case of adding different orders of lagged variables into regression models we need to pay attention to the multicollinearity issues among lagged variables.

Lagged dependent variables are not only examined in regression but also in the CART models. Gocheva-Ilieva et al. (2019) employ both lagged dependent and independent variables in the CART models. They find that the first-order of lagged dependent variables are the most relatively important variables in two towns in Bulgaria. Ou et al. (2017) use many lagged features in CART as well as other techniques to forecast traffic flow in urban roads in Kunshan City, China.

Lagged dependent (and independent) variables are also studied widely in many neural network models. Doganis et al. (2006) find that ANN models with lagged variables of sales are better than traditional methods in forecasting such as linear regression or moving average model. Swanson and White (1995) use both lagged dependent and lagged independent variables in linear regression as well as ANN to predict future spot rates. Tsoumakas (2019) surveys a number of previous studies in many machine learning models (moving average, radial basis function network, ensemble approach, long short-term memory, and deep neural networks) for food sales prediction in which lagged variables play very important roles.

### 2.1.5 Firm performance indicators

This study investigates four common measures of firm performance in the business literature which are return on assets (ROA), earnings before interest and taxes on assets (EBITAT)[5], Tobin's Q ratio, and sales growth. Here, ROA and EBITAT represent firm

[5] Note that in this study I use EBITAT for in-text sentences. The corresponding lower word, `ebitat`, is used as a variable. Similarly for ROA and `roa`.

profitability, Tobin's Q represents the investment opportunity, and sales growth is an indicator representing the efficiency of firm operations.

**Return on assets (ROA)**. ROA is a measure that represents the profitability (or efficiency) of a firm in relation to its total assets. Specifically, return on assets is calculated by dividing a firm's net income by its total assets as follows

$$\text{ROA} = \frac{\text{Net income}}{\text{Total assets}}, \tag{2.1}$$

where

$$\text{Net income} = \text{Total revenue} - \text{Total expenses}.$$

Therefore, we can say that ROA measures the profitability or efficiency of a firm. The higher ROA, the more efficient assets are and vice versa. Note that net income and total assets are normally measured in the same period of time.

**Earnings before interest and taxes on assets (EBITAT)**. EBITAT is another indicator of a firm's profitability. It is the ratio of earnings before interest and taxes on assets (EBIT) to total assets[6]. To make it comparable with ROA, EBIT is divided by total assets to obtain the earnings before interest and taxes on total assets.

$$\text{EBITAT} = \frac{\text{Earnings before interest and taxes}}{\text{Total assets}}. \tag{2.2}$$

Similarly to (2.1), in (2.2) EBIT and total assets are measured in the same period of time.

**Tobin's** $Q$. Tobin's Q (or just $Q$ for short) is an indicator as a proxy for investment opportunities. It was first studied by two macroeconomists: Nicholas Kaldor and James Tobin. Kaldor (1966) introduces the evaluation ratio $v$ (or Kaldor's $v$) of the stock market in the following formula

$$\text{Kaldor's } v = \frac{\text{Firm's market value}}{\text{Replacement cost}},$$

where the replacement cost means the capital employed by the firms, which is normally approximately by the book value of total assets. It is named after the $q$ theory of investment in Tobin (1969).

More recently, Lindenberg and Ross (1981) develop a procedure for calculating $q$ ratio in which the numerator is not the market value of shares like in the Kaldor's $v$ formula. Rather, the numerator should be the firm's total market value including the market value of

[6]The ratio of EBIT/total assets is a common measure of profitability in many previous studies (e.g. Margaritis and Psillaki 2010; Cornett et al. 2008; Yu et al. 2009).

a firm's debt, the market value of common stock, and the market value of preferred stock. The denominator, which measures the replacement cost, is more complicated to measure. Loosely speaking, the denominator can be approximated by the book value of the firm's total assets. Therefore, Tobin's $Q$ is defined as the ratio of the market value of a firm's equity and liabilities to its corresponding book values, i.e.

$$\begin{aligned}\text{Tobin's } Q &= \frac{\text{Equity market value} + \text{Liabilities market value}}{\text{Equity book value} + \text{Liabilities book value}}\\ &= \frac{\text{Equity market value} + \text{Liabilities market value}}{\text{Total assets}},\end{aligned}$$

where Total assets = Shareholders'equity + Liabilities = Equity market value + Liabilities book value in which the first equality is the basic relation in accounting.

In this study, I denote $Q$ as Tobin's $Q$ and estimate it by the similar formula as that in Smirlock et al. (1984) given by

$$Q = \frac{\text{Equity market value} + \text{Total preferred stock} + \text{Total long-term debt} + \text{Total debt in current liabilities}}{\text{Total assets}}. \tag{2.3}$$

**Sales growth**. Sales growth (or revenue growth) is one of the most common measures of firm performance. There are many previous studies that use revenue growth (i.e. sale growth) as an indicator of firm performance (e.g. Thornhill 2006; Mithas et al. 2012; Mironov 2013).

$$\text{Sales growth}(t) = \frac{\text{Sales}(t+1) - \text{Sales}(t)}{\text{Sales}(t)}. \tag{2.4}$$

The remainder of this chapter is organized as follows. Section 2.2 presents the methods that we use in this chapter. Section 2.3 describes the data and variables. The results of different models are presented in Section 2.4. We discuss several noticeable points in Section 2.5 before concluding chapter in Section 2.6.

## 2.2 The methods

As famous statistician Leo Breiman suggested in his 2001 paper that "we need to move away from exclusive dependence on data models and adopt a more diverse set of tools" (Breiman 2001), this study employ three different types of machine learning techniques representing for three different statistical methods including classical parametric method (linear regression), non-parametric method (CART), and flexible nonlinear models (neural networks).

### 2.2.1 Linear regression

The specification of the first linear regression model is given by

$$y_{ijkt} = \beta_0 + \beta_l \mathbf{X}_{lit} + \beta_e \mathbf{X}_{eit} + \beta_m \mathbf{X}_t + \beta_1 \text{disaster}_{jt} + \beta_2 \text{disaster}_{j,t+1} + \beta_3 \text{disaster10K}_{it} + \nu_j + \varphi_k + \xi_t + \epsilon_{ijkt},$$

where

- $y_{ijkt}$ is a specific firm performance indicator for firm $i$ in state $j$ in industry $k$ in year $t$,
- $\mathbf{X}_{lit}$ is the list of $m$ state controls of firm $i$ at year $t$, where $2 \leq l \leq m$,
- $\mathbf{X}_{eit}$ is the list of $n - m$ action controls of firm $i$ at year $t$, where $m + 1 \leq e \leq n$,
- $\mathbf{X}_t$ are the Herfindahl-Hirschman index and macroeconomic controls in year $t$,
- $\text{disaster}_{jt}$ and $\text{disaster}_{j,t+1}$ are the damages of natural disasters/hazards in state $j$ at year $t$ and year $t + 1$,
- $\text{disaster10K}_{it}$ is the firm $i$'s reporting natural disasters in Form 10K in year $t$,
- $\nu_j$ is the state fixed effects, where $j = 1, ..., 49$ represents for forty nine states,
- $\varphi_k$ is the industry fixed effects, where $k = 1, 2, ..., 61$ (historical SIC classification),
- $\xi_t$ is the year fixed effects, where $t = 1994, ..., 2021$,
- $\epsilon_{ijt}$ is the error term.

Note that the above year fixed effects $\xi_t$ excluded one base year (year_2003) to avoid the perfect multicollinearity issue. Similarly, the industry fixed effects $\varphi_k$ and state fixed effects $\nu_j$ already excluded one industry (sich2d_59) and one state (WY) to avoid perfect multicollinearity problem. Note also that the year fixed effects, industry fixed effects, and state fixed effects allow to eliminate bias from unobservables that change over states and industries but constant over time, bias from unobservables that change over time and state but constant over industries, and bias from unobservables that change over time and industries but constant over states.

The second linear regression specification is given by

$$y_{ijkt} = \beta_0 + \beta_l \mathbf{X}_{lit} + \beta_e \mathbf{X}_{eit} + \beta_m \mathbf{X}_t + \beta_1 \text{damage}_{jt} + \beta_2 \text{damage}_{j,t+1} + \beta_3 \text{disaster10K}_{it} + \nu_j + \varphi_k + \xi_t + \epsilon_{ijkt},$$

where $\xi_t$ is the year fixed effects, where $t = 1997, ..., 2021$. Note that for natural hazards we only have data from 1996 to 2021 instead of from 1993 to 2021 as that of the billion-dollar natural disasters.

### 2.2.2 Classification and regression trees

Classification and regression trees (also known as decision trees or CART) are a non-parametric supervised machine learning method for classification and regression tasks. It was first introduced by Breiman et al. (1984). Since the predicted outcomes in this study are continuous, we will focus on regression trees. Potential applications of CART in economics are discussed intensively in Athey and Imbens (2019).

It is necessary to give a quick overview of the empirical CART models[7]. Given a sample of training data $(X^t_{i1}, ..., X^t_{iK}, Y_i)$, for $i = 1, ..., N$ and $t = 1, ..., T$, where $K$ is the number of features and $t$ denotes time at the period $t$. Here, at one time, only one predicted outcome $Y_i$ is considered. The idea of regression trees is to split sequentially the sample of training data into subsamples, then using regression method to estimate the average predicted outcome in each subsample. Each split is only based on a threshold criterion of one feature.

Before the split, the empirical sum of squared errors of the training data is given by

$$Q = \sum_{i=1}^{N} (Y^t_i - \bar{Y}^t)^2, \qquad \text{where} \quad \bar{Y}^t = \frac{1}{N}\sum_{i=1}^{N} Y^t_i.$$

Starting with the full training data sample, at each internal node, split the sample into two subsamples based on one single feature $X_{ik}$ and a threshold criterion $c_j$, where $j = 1, ..., J$. The goal of the split is to minimize the empirical sum of squared errors of two training subsamples after splitting

$$\min \quad Q(k, c_j) = \sum_{i:X^t_{ik} \leq c_j} (Y^t_i - \bar{Y}^t_{k,c_j,l})^2 + \sum_{i:X^t_{ik} > c_j} (Y^t_i - \bar{Y}^t_{k,c_j,r})^2$$

where

$$\bar{Y}^t_{k,c_j,l} = \frac{\sum_{i:X^t_{ik} \leq c_j} Y^t_i}{\sum_{i:X^t_{ik} \leq c_j} 1} \qquad \text{and} \qquad \bar{Y}^t_{k,c_j,r} = \frac{\sum_{i:X^t_{ik} > c_j} Y^t_i}{\sum_{i:X^t_{ik} > c_j} 1}$$

are the average predicted outcomes in the two subsamples and $l$ and $r$ denote left and right, respectively. Note that the split point is the internal node such that the above sum of squared is minimum among all possible sums of squared errors. Repeat until the stopping criterion is reached. The common stopping criterion is the minimum number of training

[7] A similar framework can be found in Athey and Imbens (2019).

observations (or *sample* as a common word used in the machine learning language) at each leaf node.

The key feature importance metric for CART is the weighted impurity reduction (or mean decrease in impurity or variance reduction). Note that in the tree, a node will be split if this split induces a decrease of impurity greater than or equal to this impurity value. The weighted impurity decrease (or variance decrease) formula is given as follows

$$\text{weighted impurity decrease} = \frac{N_t}{N}(\text{impurity} - \frac{N_{tR}}{N_t} * \text{right\_impurity} - \frac{N_{tL}}{N_t} * \text{left\_impurity}),$$

where $N$ is the total number of samples, $N_t$ is the number of samples at the current node, $N_{tL}$ and $N_{tR}$ are the number of samples in the left and right child, respectively[8]. Note that the CART model in this study contains *only one* decision tree. This is in contrast to other several ensemble learning methods like XGBoost (Extreme Gradient Boosting), AdaBoost (Adaptive Boosting), or Random Forest which use several different trees.

### 2.2.3 Neural networks

This study employs neural networks with a single hidden layer called artificial neural network (ANN) and neural networks with two or more hidden layers called deep neural networks (DNN)[9]. Specifically, an ANN is a network with one input layer, one hidden layer, and one output layer[10]. As we will see, the only difference between ANN and DNNs is the number of hidden layers. In this study, the word "deep" in deep neural networks stands for many (at least two) hidden layers.

Deep neural networks (DNNs) are a class of neural networks that compose two or more hidden layers. Inspired by the notations used in Amasyali and El-Gohary (2021), in this study, DNN[2], DNN[3], and DNN[4] are defined as the DNN model with two, three, and four hidden layers, respectively. Assume that all DNN models are fully connected neural networks. In all DNN models, Rectified Linear Unit (ReLU) is used as the activation function. Note that by the Universal Approximation Theorem, any complex continuous functions can be approximated by neural networks.

The goal of learning in DNNs is to adjust the set of weights so that the performances of the networks are as good as possible (e.g. as small mean squared error as possible). In order to do that we need to adjust a list of parameters such as number of hidden layers,

[8]For more details, please see https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html.

[9]Note that this classification is just with the naming purpose. In reality, ANN and DNN might be used interchangeably.

[10]This definition is similar to that in Amasyali and El-Gohary (2021). Meanwhile, other studies might define that ANNs are neural networks with one or two hidden layers and DNNs are neural networks with more than two hidden layers (e.g. Merkel et al. 2018).

number of neurons in each layer, optimizer, batch size, number of epochs, initialization (using random seed), and dropout rate. They are normally called hyperparameters, which are the parameters of weights (or coefficients).

Note that in contrast with the linear regression models, CART and neural networks are nonlinear models. This implies that CART and neural networks can capture the nonlinear relationships and heterogeneous effects among features which are usually ignored in the linear models. However, some costs of using neural networks are computational intensity, losing the level of interpretability, curse of dimensionality, and overfitting problems.

### 2.2.4 Comparison among models

Meanwhile there are many studies that compare linear regression, CART, and neural networks in other fields (Razi and Athappilly 2005; Kim 2008; Wang et al. 2016), there are much less studies comparing models using these techniques in the literature of business, management, and finance. In the business, management, and finance literature, Bensic et al. (2005) studies small business credit scoring using neural networks, decision trees, and logistic regression rather than linear regression. Also, Schumacher et al. (2010) predict success of actuarial students using neural networks, classification trees, and logistic regression rather than regression trees and linear regression. It seems there are no studies in the literature of business, management, and finance using linear regression, CART (decision trees), and neural networks at the same time. Further, even though there might exist some studies using these three comparative techniques, it is unlikely that these studies focus on firm performance prediction. This is a big gap in the literature of business, management, accounting, and finance and this study aims to partially fulfill this gap.

## 2.3 The data and variables

### 2.3.1 Data sources

The datasets and corresponding sources used in this research are shown in Table 2.2. All data are annual and they are collected from 1993 and 2021 except for the U.S. natural hazard data that is collected between 1996 and 2021. The firm's characteristics are collected from the Compustat-Capital IQ (or just Compustat for short) database from the Wharton Research Data Services (WRDS). Macroeconomic variables are collected from the World Bank's website. The costs of natural disasters are collected from the National Oceanic and Atmospheric Administration (NOAA)'s billion-dollar weather and climate disasters database. Form-10Ks are collected from the Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) from the U.S. Securities and Exchange Commission (SEC).

Table 2.2: The data and corresponding sources

| Data | Frequency | Period | Source |
|---|---|---|---|
| Firm's characteristics | Annually | 1993-2021 | Compustat |
| Macroeconomic variables | Annually | 1993-2021 | The World Bank |
| Natural hazards | Annually | 1996-2021 | NOAA |
| Billion-dollar natural disasters | Annually | 1993-2021 | NOAA |
| Form 10-Ks (reporting) | Annually | 1993-2021 | EDGAR filings |

Regarding the U.S. billion-dollar natural disasters, there are 287 billion-dollar natural disaster events (e.g. floods, tropical cyclones, droughts) that occurred between 1993 and 2021 in the United States. This is equivalent to, on average, 8.44 billion-dollar natural disaster events per year. The number of events at the state level is clearly less than 8.44 since not all events occur in all states, which makes billion-dollar disasters as rare events in the U.S. at the state level.

**U.S. natural hazards vs. U.S. billion-dollar natural disasters**. The U.S natural hazards data (officially the U.S. Natural Hazard Statistics[11]) provides the statistical information on damages caused by weather related hazards. There are some important differences between the U.S. natural hazards data and the U.S. billion-dollar natural disasters. First, the natural hazard data include all hazardous weather events either big or small weather phenomena. In contrast, the U.S. billion-dollar natural disasters only covers the major weather and climate disasters in the U.S. Second, the U.S. natural hazardous damages are the *real* values obtained or estimated from each event while the U.S. billion-dollar disasters are the *interval* estimates. Those interval estimates are averaged to get a unique value of damage of each natural disaster. Third, it seems that the damages of the U.S. natural hazardous events are the estimated *direct* damages while that of the U.S. billion-dollar natural disasters are the *total* (direct and relevant indirect) damages of natural disasters. Moreover, the average damages of the U.S. natural hazard is approximately \$1.008 billion per year while that of the U.S. billion-dollar natural disasters is around \$2.8538 billion per year. Table 2.3 summarizes the comparison between the U.S. natural hazards and the U.S. billion-dollar natural disasters.

[11] See `https://www.weather.gov/hazstat/`

Table 2.3: Comparison between U.S. natural hazards and billion-dollar natural disasters

| | Natural hazards | Natural disasters |
|---|---|---|
| Size of disasters | Small & big | Big (billion-dollar events) |
| Types of costs | Direct | Direct & Indirect |
| Values of estimates | Unique real number | Interval |
| CPI adjustment to 2021 | Yes | Yes |
| Average damages per year | $1.008 billion | $2.8538 billion |

This study takes inflation (i.e Consumer Price Index or CPI) into account. The U.S. billion-dollar natural disasters are adjusted to CPI in 2021. However, the damages of the U.S. natural hazards are measured every year. To make sure these two damages are comparable, we adjust the damages of the U.S. natural hazards to CPI in 2021. That is, both damages of the U.S. natural hazards and billion-dollar natural disasters are adjusted to CPI in 2021. By taking CPI into account, we reduce the possibility of the endogeneity issue regarding the changes of the values of damages of natural hazards and disasters over time.

### 2.3.2 The variables

#### Dependent variables

The dependent variables and the corresponding definitions are shown in Table 2.4. Note that in this table $t$ stands for the current year and $t + 1$ stands for the next year.

Table 2.4: The list of dependent variables

| Variable | Definition |
|---|---|
| `roa` | Net income(t+1)/Total assets(t+1) |
| `ebitat` | Income before interest and taxes(t+1)/Total assets(t+1) |
| `Q` | (Equity market value(t+1) + Total preferred stock(t+1) + Total long-term debt(t+1) + Total debt in current liabilities (t+1))/Total assets(t+1) |
| `salegrowth` | (Sales(t+1) - Sales(t)) / Sales(t) |

Since all of the firm-level state independent variables are at time $t$ and all dependent variables are at time $t + 1$, this is a prediction problem. Specifically, this is a short-term prediction problem since data this year is used to predict the next year firm performance.

#### Independent variables

In order to investigate the relationships between uncertainty variables and firm performance, various control variables, including firm-wide state and action variables, industry-wide variable, and macroeconomic variables, are utilized. These controls are generated based on the literature on bankruptcy, firm performance, and accounting, specifically many variables are borrowed from Mai et al. (2019). When predicting firm performance, it is natural to consider variables at the firm-wide, industry-wide, and the economy-wide variables. The firm-level variables are split into two groups: firm-wide state independent variables and firm-wide action independent variables. Here the word "action" implies that firms can adjust these variables to obtain the desired firm performance indicators next year based on the current performance and the effects of natural disasters this year. The definitions of these controls are described in detail in the following tables.

**Firm-wide state independent variables**. Table 2.5 shows the list of firm-wide state independent variables (features). Most of variables in this list are the ratio variables with the denominator is total assets or sales. One important advantage of the ratio variables with total assets as denominator is that it helps to deal with the so-called *asset inflation*. As a result, the ratio variables might help dealing with the *endogeneity* problem because of the increase of the values of (total) assets over time. The similar argument might make sense when the denominator is sales.

Table 2.5: The list of firm-wide state independent variables

| Variable | Definition |
|---|---|
| `roalag1` | Prior year ROA |
| `ebitatlag1` | Prior year EBITAT |
| `Qlag1` | Prior year Q |
| `salegrowthlag1` | Prior year sales growth |
| `lctat` | Current liabilities(t)/Total assets(t) |
| `chlct` | Cash(t)/Current liabilities(t) |
| `lctlt` | Current liabilities(t)/Total liabilities(t) |
| `relct` | Retained earnings(t)/Current liabilities(t) |
| `atemp` | Total assets(t)/Employees(t) |
| `logat` | Ln(Total assets(t)) |
| `cheat` | Cash and short-term investment(t)/Total assets(t-1) |
| `saleat` | Sales(t)/Total assets(t-1) |
| `reat` | Return earnings(t)/Total assets(t-1) |
| `dtat` | Total debt including current(t)/Total assets(t-1) |
| `seqat` | Total stockholers' equity(t)/Total asset(t-1) |
| `invtsale` | Inventories(t)/Sales(t-1) |
| `xintsale` | Total interest and related expense(t)/Sales(t-1) |
| `invtgrowth` | (Inventories(t) – Inventories(t-1))/Inventories(t-1) |

*Notes*: This table includes four blocks of variables: lagged, ratio, and normalized variables. First block contains four lagged dependent variables. Second block contains the next five ratio variables. Third block only contains the logarithm of total assets. The last block contains nine state variables, which are normalized either by last year total assets or last year sales. Normalization by last year total assets or sales is necessary to avoid the endogeneity problem.

**Firm-wide action independent variables**. A list of firm-wide action independent variables used in this study is shown in Table 2.6.

Table 2.6: The list of firm-wide action independent variables

| Variable | Definition |
|---|---|
| `xintgrowth` | (Total Interest and Related Expense(t+1) - Total Interest and Related Expense(t)) /Total Interest and Related Expense(t) |
| `capxgrowth` | (Capital expense(t+1) - Capital expense(t))/Capital expense(t) |
| `empgrowth` | (Employees(t+1) - Employees(t))/Employees(t) |
| `dvcib` | Dividends common, ordinary(t+1)/Income before extraordinary items(t+1) |

*Note*: These variables are called *action* variables since firms can observe the effects of natural disasters on their performance to adjust these actions (or controls) to achieve their goals.

**Herfindahl-Hirschman Index and macroeconomic variables**. The Herfindahl-Hirschman Index (or Herfindahl index or industrial concentration index) is considered as a measure for industrial concentration. This index is calculated by the sum of square of market shares of all firms within a particular 3-digit SIC industry in a particular year[12]. Macroeconomic variables, including GDP growth, inflation, and real interest rate, can be found from the World Bank's database[13] and the Bureau of Economic Analysis (BEA) database[14], and the Federal Reserve Economic Data (FRED) database[15]. The Herfindahl index and three macroeconomic variables used in this study are shown in Table 2.7.

Table 2.7: The list of exogenous variables

| Variable | Definition | Source |
|---|---|---|
| `hhi` | Industrial concentration(t) | Calculated by author |
| `gdpgrowth` | GDP growth, annual %(t) | World Bank, BEA |
| `inflation` | Inflation, annual %(t) | World Bank |
| `interestrate` | Real interest rate, annual %(t) | World Bank, FRED |

*Notes*: The Herfindahl index is calculated based on the data of sales and 3-digit historical Standard Industrial Classification (or SICH) from the Compustat database with four steps. Step 1: Calculate the "total sales" of all firms within a 3-digit SICH industry in every year; Step 2: Calculate the market share for each firm within this 3-digit SICH industry in every year; Step 3: Square every market share; Step 4: Sum of all of the squares by 3-digit SICH industry in every year to obtain the Herfindahl index. Moreover, at the time we accessed, 06/06/2022, the World Bank database, the U.S. GDP growth in 2021 is not available. Therefore, we used data of U.S GDP growth in 2021 from the Bureau of Economic Analysis. Similarly, since U.S. real interest rate in 2021 is not available on 06/06/2022 in the World Bank database, we used the average of all monthly U.S. real interest rates from the Federal Reserve Economic Data (FRED) database as the U.S. real interest rate in 2021.

[12] For more details about 3-digit SIC industry classification, see `https://siccode.com/`.

[13] See `https://data.worldbank.org/country/united-states?view=chart`

[14] See `https://www.bea.gov/news/2022/gross-domestic-product-fourth-quarter-and-year-2021-advance-estimate`

[15] See `https://fred.stlouisfed.org/series/REAINTRATREARAT1YE`

**Variables of interest and sources**. Table 2.8 shows the list of all variables of interest, the data type, definition, and the data sources. Note that in this table, the damages of natural disasters in the last year (or year $t$) are denoted as `disaster` and the damages of natural disasters in the current year (or year $t+1$) are denoted as `disaster0`. Similarly for `damage` and `damage0` for the case of natural hazards. There is only one perceived natural disaster risk for both natural hazards and natural disasters, which is denoted as `disaster10K`.

Table 2.8: The list of variables of interest

| Variable | Type | Definition | Source |
|---|---|---|---|
| `disaster` | Government-reported | Damages of natural disasters(t) | NOAA database |
| `disaster0` | Government-reported | Damages of natural disasters(t+1) | NOAA database |
| `damage` | Government-reported | Damages of natural hazards (t) | NOAA database |
| `damage0` | Government-reported | Damages of natural hazards(t+1) | NOAA database |
| `disaster10K` | Firms' self-reported | Reporting disaster(t) | Form 10-Ks |

*Notes*: Note that there are two variables related to the damages of natural billion-dollar disasters and two corresponding for natural hazards. However, for each linear regression specification, we only consider these two kinds of damages in two separate cases. For more details, see Subsection 2.2.1.

### 2.3.3 Data collection procedure

Table 2.9 describes the steps to collect data (the number of firms and firm-year observations). At the beginning (Step 1), there are 29,908 firms and 315,051 firm-year observations. Note that at Step 1 not all observations are firm-based observations. The reason is that many stock ticker symbols are bonds or exchange-traded funds (ETFs). The following steps will remove these non-firm stocks.

The raw data from Step 1 contains so many rows with missing values in most of the columns. These rows should be deleted. Step 2 removes all rows with missing values in more than 75% of the columns. Many of them might be bonds and ETFs since they are not firm-based stocks, then there are not common financial statement variables such as number of employees, sales, total assets, and so on. After Step 2, there are 23,950 firms and 247,477 firm-year observations. At Step 3, the data getting from Step 2 is merged with the actual damages of natural disasters and reporting disasters data.

Step 4 and Step 5-8 help to exclude the non-firm stocks. In total, our final sample size includes 10,575 firms and 97,761 firm-year observations with no missing values. Table 2.9 shows the procedure to get the final sample size as well as all number of firms and firm-year observations at each step. Note that the sample size excludes the firms in the utility, finance, insurance, and real estate sector (Step 5) since these firms typically have the capital structures and other characteristics different compared to manufacturing and other sectors.

Table 2.9: The data collection procedure

| Step | Observations | Number of firms |
|---|---|---|
| 1. Total U.S. firms from Compustat database | 315,051 | 29,908 |
| 2. Firms with data in at least 75% columns | 247,477 | 23,950 |
| 3. U.S. firms with state and disasters data | 141,229 | 16,601 |
| 4. U.S. firms with SICH, CIK, and GVKEY number | 139,968 | 16,452 |
| 5. Exclude U.S. utilities, financial, insurance, real estate firms | 113,009 | 13,689 |
| 6. U.S. firms with total assets at least $0.1 million | 111,361 | 13,569 |
| 7. U.S. firms with positive sales | 107,019 | 13,074 |
| 8. U.S. firms with at least 1 employee | 103,401 | 12,776 |
| 9. Sample after handling missing values | 97,761 | 10,575 |
| Final sample size | 97,761 | 10,575 |

The approximate number of firms by sector and the corresponding proportions are shown in Table 2.10. It is apparent that most of the firms in our sample are manufacturing (45.72%) and services (27.02%). These proportions seem representative for all public firms. Note that the number of firms in Table 2.10 and that in Table 2.9 are close to each other with 11,373 and 97,761, respectively, but not the same. The reason is because a number of firms belong to two or more sectors. For example, some firms are classified to the whole trade sector but they are also classified to the retail trade sector because they do both wholesale and retail trading.

Table 2.10: Number of firms and proportions by sector

| Sector | Number of firms | Proportion (%) |
|---|---|---|
| Agriculture, Forestry, Fishing | 21 | 0.18 |
| Mining | 586 | 5.15 |
| Construction | 172 | 1.51 |
| Manufacturing | 5,200 | 45.72 |
| Transportation | 951 | 8.36 |
| Wholesale Trade | 545 | 4.79 |
| Retail Trade | 827 | 7.27 |
| Services | 3,071 | 27.02 |
| Total | 11,373 | 100 |

### 2.3.4 Constructing a natural disaster dictionary

In order to capture the intensity of the natural disasters to accurately reflect the perceived natural disaster risk we need the following assumption.

**Assumption 2.3.1.** *People (i.e. firms) are likely to discuss the big natural disasters more often than the relatively smaller ones in their Form 10-Ks.*

This assumption is necessary to examine the different intensities of different natural disasters. Without this assumption, the perceived natural disaster risk might be biased. For example, a Form 10-K of a company mentions both Hurricane Katrina in 2005 (one of the most powerful hurricane in the U.S. in the history) and Hurricane Iniki (the most powerful hurricanes striking Hawaii in recorded history) in 1992. Obviously, these two hurricanes have different intensities. Separately, one word related to Hurricane Katrina is not equivalent to one word related to Hurricane Iniki (in the sense that these hurricanes were damaged differently). By Assumption 2.3.1, on average, firms tend to talk about Hurricane Katrina more than Hurricane Iniki. This obviously makes sense in the real world. Note that this assumption can be applied to not only the same types of natural disasters but also the different ones (e.g hurricanes in Louisiana vs. wildfires in California).

Table 2.11 shows the natural disaster dictionary including 133 words or phrases related to natural disasters in alphabetical order. This dictionary is generated based on the words related to natural disasters from the National Oceanic and Atmospheric Administration (NOAA) website[16], the Emergency Events Database (EM-DAT) glossary on natural disasters[17], Oxford's English dictionaries from Oxford Languages[18], and the well-known Merriam-Webster dictionary[19].

[16] https://www.ncei.noaa.gov/access/billions/events
[17] https://www.emdat.be/Glossary
[18] https://languages.oup.com/google-dictionary-en/
[19] https://www.merriam-webster.com/

Table 2.11: The natural disaster dictionary

| | | | |
|---|---|---|---|
| adaptive capacity | droughts | heat wave | seiches |
| air burst | earthquake | heat waves | severe winter condition |
| airburst | earthquakes | heavy rainfall | severe winter conditions |
| airbursts | El Nino | heavy snow | snow |
| apocalypse | extreme heat | high wind | snowstorm |
| apocalypses | extreme rain | high winds | snowstorms |
| ash fall | extreme rainfall | hurricane | storm |
| ashfall | extreme rains | hurricanes | storms |
| avalanche | extreme temperature | hydrological hazard | thunderstorm |
| avalanches | extreme temperatures | hydrological hazards | thunderstorms |
| blizzard | extreme weather | La Nina | tornado |
| blizzards | firestorm | landfall | tornadoes |
| calamity | firestorms | landfalls | tremor |
| camilities | flood | landslide | tsunami |
| cataclysm | flooded | landslides | tsunamis |
| cataclysms | flooding | lava flow | twister |
| climate change | floodings | lava flows | typhoon |
| climate warms | floods | microburst | typhoons |
| coastal erosion | fog | microbursts | volcanic |
| cold wave | fogs | mudslide | volcano |
| cold waves | forest fire | mudslides | volcanos |
| coping capacity | forest fires | natural hazard | whirlpool |
| cyclone | freeze | natural hazards | whirlpools |
| cyclones | freezes | rainfall | wildfire |
| debacle | freezing | rainfalls | wildfires |
| debacles | gale | rainstorm | wind |
| derecho | gales | rainstorms | winds |
| derechos | geophysical hazard | rock fall | windstorm |
| disaster | geophysical hazards | rock-fall | windstorms |
| disaster risk | global warming | rockfall | winterstorm |
| disaster risk management | hail | rockfalls | winterstorms |
| disaster risks | hailstorm | rogue wave | |
| disasters | hailstorms | rogue waves | |
| drought | hazard mitigation | seiche | |

*Notes*: The words list related to natural disasters in this dictionary bases on several reliable sources including NOAA, EM-DAT glossary, Oxford Languages (Oxford's English dictionaries), and Merriam-Webster dictionary.

### 2.3.5 Using the natural disaster dictionary to explore Form 10-Ks

The natural disaster dictionary in Table 2.11 is used to count the number of words related to natural disasters in Form 10-Ks from 1993 to 2021. Note that our raw data includes 195,229 firm-year Form 10-Ks and 35,640 unique firms. That is, on average, each firm has approximately 5.48 Form 10-Ks.

Table 2.12 shows the most frequent words, number of firms mentioned that word, and number of word count. Two most frequent words related to natural disasters between 1993 and 2021 are *flood* (from 1996 to 2004) and *disasters* (in 2007 and from 2012 to 2021). The word *hurricanes* is the most frequent in 2005 and 2006, which might be caused by Hurricane Katrina in late August 2005. The word *earthquake* and *wind* are the most frequent words related natural disasters in several years.

Table 2.12: The most frequent words related to natural disasters in Form 10-Ks

| Year | Most frequent word | Number of firms mentioned | Word count |
|---|---|---|---|
| 1993 | earthquake | 114 | 373 |
| 1994 | earthquake | 110 | 359 |
| 1995 | earthquake | 262 | 798 |
| 1996 | flood | 620 | 1,432 |
| 1997 | flood | 688 | 1,796 |
| 1998 | flood | 689 | 1,623 |
| 1999 | flood | 650 | 1,436 |
| 2000 | flood | 695 | 1,560 |
| 2001 | flood | 636 | 1,465 |
| 2002 | flood | 931 | 2,220 |
| 2003 | flood | 967 | 2,144 |
| 2004 | flood | 1,177 | 2,390 |
| 2005 | hurricanes | 1,290 | 4,677 |
| 2006 | hurricanes | 1,341 | 4,470 |
| 2007 | disasters | 2,411 | 4,479 |
| 2008 | wind | 905 | 5,494 |
| 2009 | wind | 1,036 | 7,713 |
| 2010 | wind | 1,093 | 8,947 |
| 2011 | wind | 1,067 | 7,283 |
| 2012 | disasters | 3,208 | 7,298 |
| 2013 | disasters | 3,357 | 7,680 |
| 2014 | disasters | 3,481 | 8,178 |
| 2015 | disasters | 3,522 | 8,363 |
| 2016 | disasters | 3,523 | 8,484 |
| 2017 | disasters | 3,592 | 9,113 |
| 2018 | disasters | 3,677 | 9,533 |
| 2019 | disasters | 3,893 | 10,804 |
| 2020 | disasters | 4,193 | 12,209 |
| 2021 | disasters | 808 | 2,427 |

Table 2.13 describes the top ten topics related to natural disasters in Form 10-Ks. We can observe that the words or phrases belong to the topic of *natural disasters*, *wind*, and *flooding* are the most common in Form 10-Ks with the corresponding number of word count are 204,007 terms, 116,670 terms, and 112,338 terms. For the full list of the frequencies of all words/phrases related natural disasters in Forms 10-Ks, please see Appendix 5.

Table 2.13: Top 10 topics related to natural disasters in Form 10-Ks

| Topic | Word/Phrase | Word count |
|---|---|---|
| natural disasters | disaster, disasters, natural hazard, natural hazards | 204,007 |
| wind | wind, winds, windstorm, windstorms, high wind, high winds | 116,670 |
| flooding | flood, floods, flooding, floodings, heavy rainfall, rainfalls, extreme rains, extreme rainfall, rainstorm, rainstorms | 112,338 |
| earthquake | earthquake, earthquakes | 93,743 |
| hurricanes | hurricane, hurricanes | 81,314 |
| storm | storm, storms, winterstorm | 73,710 |
| climate change | climate change, global warming, climate warms | 67,725 |
| freeze | freeze, freezes, freezing, snow, heavy snow, snowstorm, snowstorms | 51,873 |
| drought | drought, droughts | 19,194 |
| tornado | tornado, tornadoes | 16,365 |

### 2.3.6 Data processing

In order to obtain reliable input data, it is necessary to do the following steps. First of all, all variable (except fixed effects) are winsorized at the 1st and 99th percentiles[20]. Next, the correlation matrix of all independent variables is generated. Thus, one of the two variables in pairs with correlations greater than 0.8 will be dropped. This step is repeated until obtaining all correlations less than 0.8. For the models with the billion-dollar natural disasters, the following table shows the top 20 correlations of the independent variables[21].

[20] These percentiles are common to handle outliers, e.g Dessaint and Matray (2017), Mai et al. (2019).

[21] The top 20 correlations of the models corresponding with the damages of the natural *hazards* are exactly the same that of natural *disasters* so we do not show it here.

Table 2.14: Top 20 absolute correlations of Model 1 and Model 2

| Model 1 (ROA) | | | Model 2 (EBITAT) | | |
|---|---|---|---|---|---|
| Variable 1 | Variable 2 | Correlation | Variable 1 | Variable 2 | Correlation |
| logoil | interestrate | 0.732368 | logoil | interestrate | 0.732368 |
| lctat | reat | 0.687432 | lctat | reat | 0.687432 |
| lctat | seqat | 0.635743 | reat | ebitatlag1 | 0.642279 |
| reat | roalag1 | 0.621879 | lctat | seqat | 0.635743 |
| lctat | roalag1 | 0.542714 | cheat | chlct | 0.541779 |
| cheat | chlct | 0.541779 | reat | relct | 0.534726 |
| reat | relct | 0.534726 | lctat | ebitatlag1 | 0.529939 |
| lctlt | logat | 0.515386 | lctlt | logat | 0.515386 |
| dtat | seqat | 0.467097 | logat | ebitatlag1 | 0.490036 |
| reat | seqat | 0.463247 | dtat | seqat | 0.467097 |
| logat | roalag1 | 0.456176 | reat | seqat | 0.463247 |
| logat | reat | 0.427153 | relct | ebitatlag1 | 0.428460 |
| seqat | roalag1 | 0.414960 | logat | reat | 0.427153 |
| gdpgrowth | inflation | 0.395303 | gdpgrowth | inflation | 0.395303 |
| logoil | gdpgrowth | 0.393173 | logoil | gdpgrowth | 0.393173 |
| lctlt | dtat | 0.389953 | lctlt | dtat | 0.389953 |
| xintsale | roalag1 | 0.388670 | xintsale | ebitatlag1 | 0.385681 |
| relct | roalag1 | 0.383257 | seqat | ebitatlag1 | 0.384161 |
| reat | xintsale | 0.381331 | reat | xintsale | 0.381331 |
| logat | relct | 0.377250 | logat | relct | 0.377250 |

Table 2.15: Top 20 absolute correlations of Model 3 and Model 4

| Model 3 (Tobin's Q) | | | Model 4 (Sales growth) | | |
|---|---|---|---|---|---|
| Variable 1 | Variable 2 | Correlation | Variable 1 | Variable 2 | Correlation |
| logoil | interestrate | 0.732368 | logoil | interestrate | 0.732368 |
| lctat | reat | 0.687432 | lctat | reat | 0.687432 |
| lctat | seqat | 0.635743 | lctat | seqat | 0.635743 |
| cheat | chlct | 0.541779 | cheat | chlct | 0.541779 |
| reat | relct | 0.534726 | reat | relct | 0.534726 |
| lctlt | logat | 0.515386 | lctlt | logat | 0.515386 |
| dtat | seqat | 0.467097 | dtat | seqat | 0.467097 |
| reat | seqat | 0.463247 | reat | seqat | 0.463247 |
| logat | reat | 0.427153 | logat | reat | 0.427153 |
| reat | Qlag1 | 0.425811 | gdpgrowth | inflation | 0.395303 |
| gdpgrowth | inflation | 0.395303 | logoil | gdpgrowth | 0.393173 |
| logoil | gdpgrowth | 0.393173 | lctlt | dtat | 0.389953 |
| lctlt | dtat | 0.389953 | reat | xintsale | 0.381331 |
| reat | xintsale | 0.381331 | logat | relct | 0.377250 |
| logat | relct | 0.377250 | lctat | saleat | 0.377027 |
| lctat | saleat | 0.377027 | xintsale | salegrowthlag1 | 0.376152 |
| cheat | relct | 0.370620 | cheat | relct | 0.370620 |
| lctat | logat | 0.368511 | lctat | logat | 0.368511 |
| lctat | xintsale | 0.349713 | lctat | xintsale | 0.349713 |
| gdpgrowth | interestrate | 0.340178 | gdpgrowth | interestrate | 0.340178 |

*VIF test*. To eliminate severe multicollinearity problem, besides using correlation coefficient matrix, we do the Variance Inflation Factor (VIF) test by calculating $\text{VIF}_j = \frac{1}{1-R_j^2}$, where $R_j^2$ is the goodness of fit value obtained by regressing the $j$th variable on the remaining independent variables. Empirically, variables with VIF's values less than 8 are kept as shown in Table 2.51[22].

[22]The top VIF's values of the linear regression for the case of natural hazards are shown in Appendix 4.

Table 2.16: Top 20 VIF's values for linear regression models

| Model 1 | VIF | Model 2 | VIF | Model 3 | VIF | Model 4 | VIF |
|---|---|---|---|---|---|---|---|
| inflation | 7.017203 | inflation | 7.013184 | lctlt | 7.334174 | inflation | 6.998687 |
| lctlt | 6.916450 | lctlt | 6.922942 | inflation | 7.009825 | lctlt | 6.955896 |
| logat | 6.449909 | logat | 6.549338 | logat | 6.347809 | logat | 6.380665 |
| lctat | 6.235576 | lctat | 6.233981 | lctat | 6.120241 | lctat | 6.187584 |
| interestrate | 4.376896 | interestrate | 4.373795 | seqat | 4.406900 | interestrate | 4.379291 |
| seqat | 4.281882 | seqat | 4.277134 | interestrate | 4.374459 | seqat | 4.300491 |
| reat | 3.871019 | reat | 3.900545 | gdpgrowth | 3.834821 | gdpgrowth | 3.832852 |
| gdpgrowth | 3.833718 | gdpgrowth | 3.832756 | reat | 3.829549 | reat | 3.760325 |
| saleat | 3.639975 | saleat | 3.697453 | saleat | 3.607631 | saleat | 3.655896 |
| cheat | 3.240497 | cheat | 3.250151 | cheat | 3.242551 | cheat | 3.248636 |
| hhi | 2.661929 | hhi | 2.663623 | hhi | 2.661298 | hhi | 2.661279 |
| relct | 2.282047 | relct | 2.295641 | dtat | 2.258163 | dtat | 2.258033 |
| dtat | 2.257078 | dtat | 2.256507 | relct | 2.243393 | relct | 2.241613 |
| chlct | 2.186465 | ebitatlag1 | 2.239404 | chlct | 2.197794 | chlct | 2.178986 |
| roalag1 | 2.136983 | chlct | 2.187009 | Qlag1 | 2.112606 | disaster10K | 1.685062 |
| disaster10K | 1.688482 | disaster10K | 1.691070 | disaster10K | 1.686413 | invtsale | 1.627244 |
| invtsale | 1.582029 | invtsale | 1.585414 | invtsale | 1.579723 | atemp | 1.560506 |
| atemp | 1.558577 | atemp | 1.558415 | atemp | 1.558741 | xintsale | 1.553171 |
| xintsale | 1.424234 | xintsale | 1.418182 | xintsale | 1.407173 | salegrowthlag1 | 1.403611 |
| capxgrowth | 1.228371 | capxgrowth | 1.228950 | capxgrowth | 1.224616 | capxgrowth | 1.224616 |

### 2.3.7 Data transformation

Note that in this study we use two datasets: original data and transformed data. For the comparison between linear regression and CART, we use the original data. For comparison among linear regression, CART, and neural networks, we use the transformed data.

To compare among all models (linear regression, CART, and neural networks), the entire data (both train and test data) is scaled by min-max scaler. Consider a feature or target $x$ with values $x_i, i = 1, 2, ..., n$. The original values of $x$ will be transformed as follows

$$x_{\text{scaled}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}, \tag{2.5}$$

where $x_{\text{min}}$ and $x_{\text{max}}$ is the minimum and the maximum of feature (or variable) $x$, respectively.

By using min-max scaler, all targets and features of both train and test data will be transformed into the range $[0, 1]$. This means that the minimum and maximum value of a feature or target is 0 and 1, respectively. In Python, one can use the `MinMaxScaler()` built-in function.

### 2.3.8 Exploratory data analysis

Exploratory data analysis (EDA) is a crucial step of performing initial investigations on data. Three common tools for EDA are summary statistics, histogram, and heat map.

Table 2.17 illustrates the summary statistics of all dependent and lagged dependent variables. Note that even though the mean ROA and EBITAT are negative, the median ROA and EBITAT are about 1.9% and 5.1%, respectively, which seem rational.

Table 2.17: Summary statistics of the dependent and lagged dependent variables

| | roa | ebitat | Q | salegrowth | roalag1 | ebitatlag1 | Qlag1 | salegrowthlag1 |
|---|---|---|---|---|---|---|---|---|
| count | 97761 | 97761 | 97761 | 97761 | 97761 | 97761 | 97761 | 97761 |
| mean | -0.1597 | -0.0885 | 2.3457 | 0.0487 | -0.1664 | -0.0927 | 2.2709 | 0.0494 |
| std | 0.6107 | 0.4922 | 3.3787 | 0.7767 | 0.6381 | 0.5093 | 3.2011 | 0.7812 |
| min | -4.2418 | -3.2768 | 0.3079 | -0.8855 | -4.4444 | -3.4259 | 0.2866 | -0.8855 |
| 25% | -0.1127 | -0.0677 | 0.8721 | -0.1840 | -0.1153 | -0.0694 | 0.8552 | -0.1838 |
| 50% | 0.0186 | 0.0515 | 1.3185 | -0.0637 | 0.0193 | 0.0509 | 1.2901 | -0.0637 |
| 75% | 0.0652 | 0.1061 | 2.3400 | 0.0505 | 0.0666 | 0.1063 | 2.2770 | 0.0505 |
| max | 0.3100 | 0.3346 | 25.0567 | 5.8950 | 0.3232 | 0.3411 | 23.4122 | 5.9450 |

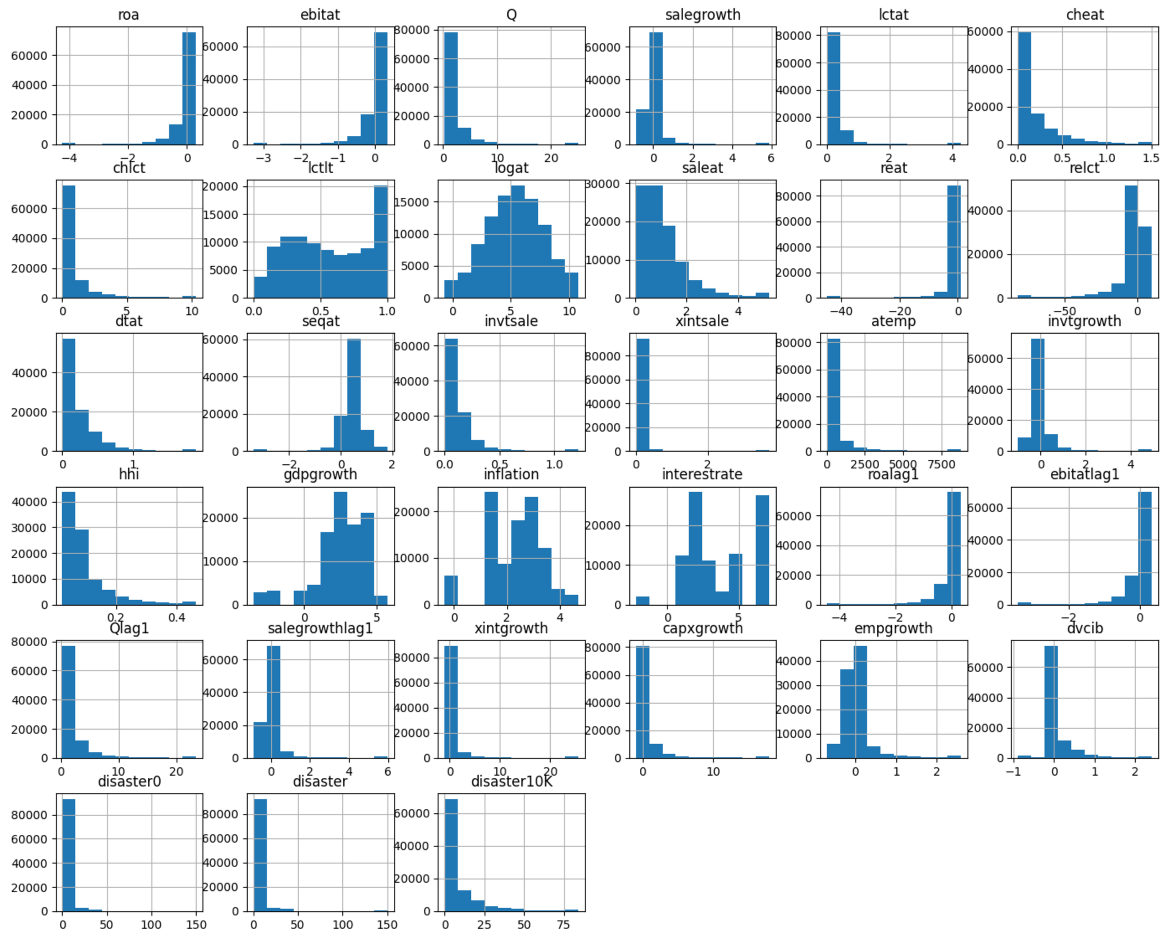
Table 2.18 shows the summary statistics of action independent variables and variables of interest (natural disasters).

Table 2.18: Summary statistics of action independent variables and variables of interest

| | xintgrowth | capxgrowth | empgrowth | dvcib | disaster0 | disaster | disaster10K |
|---|---|---|---|---|---|---|---|
| count | 97761 | 97761 | 97761 | 97761 | 97761 | 97761 | 97761 |
| mean | 0.6231 | 0.6050 | 0.0257 | 0.1154 | 2.8538 | 2.7275 | 8.9035 |
| std | 3.1686 | 2.4821 | 0.4038 | 0.3731 | 11.1612 | 10.7287 | 14.5812 |
| min | -1.0000 | -0.9939 | -0.7000 | -0.8834 | 0.0000 | 0.0000 | 0.0000 |
| 25% | -0.2339 | -0.3252 | -0.1214 | 0.0000 | 0.0525 | 0.0525 | 0.0000 |
| 50% | 0.0000 | -0.0142 | -0.0207 | 0.0000 | 0.3750 | 0.3750 | 3.0000 |
| 75% | 0.2782 | 0.4742 | 0.0654 | 0.0177 | 1.5000 | 1.5000 | 11.0000 |
| max | 25.6000 | 17.9381 | 2.5854 | 2.4079 | 150.0000 | 150.0000 | 84.0000 |

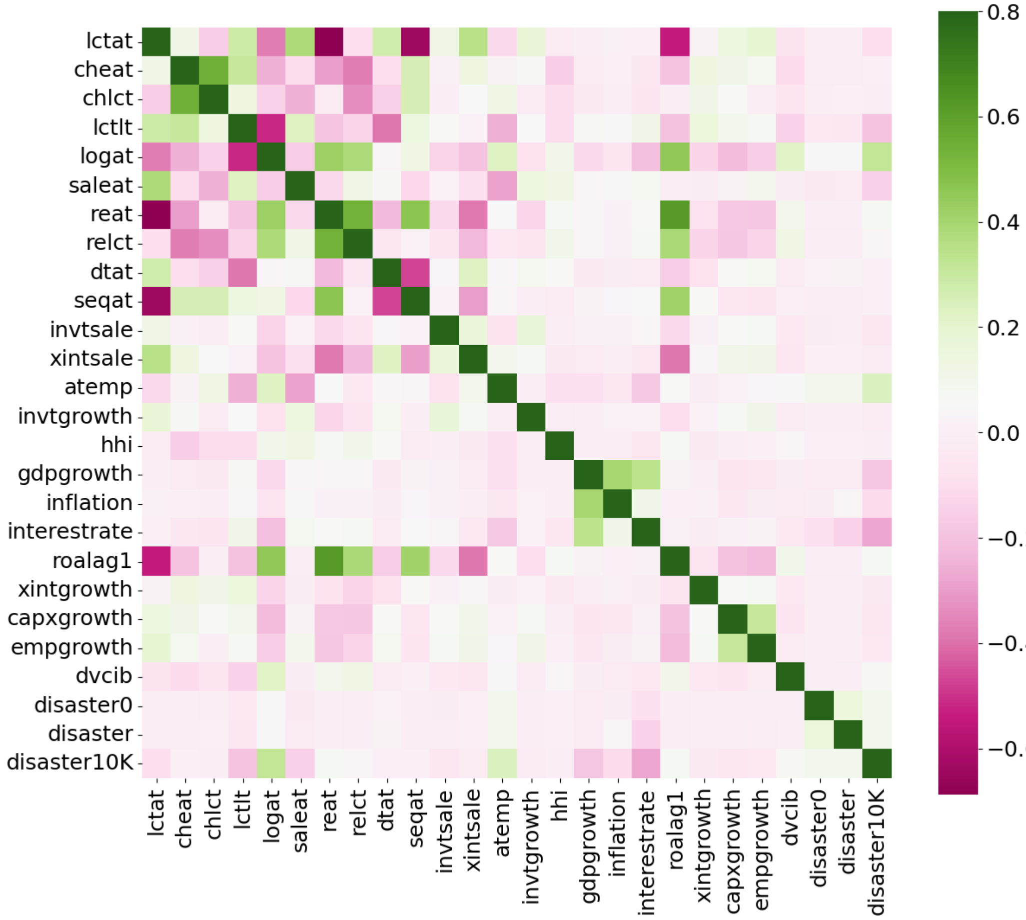
The histograms of all variables of the original data are shown in Figure 2.4.

Figure 2.4: The histograms of the original data




*Heat map of original features.* The heat map of all of the original features of the entire data (both train and test data) are shown in Figure 2.5. Heat map gives us a general image of the relationships among variables in terms of direction and magnitude. The green implies the positive relationship while the pink implies the negative relationship between a pair of two variables. Note that the bolder green cells represent the more positive correlation and the bolder pink cells imply the more negative correlation.

Figure 2.5: Heat map of the original features of Model 1 (ROA)




*Notes*: The figure shows that the correlation coefficient between any pair of variables is less than 0.8. This implies that there are no severe correlations among independent variables in Model 1. Note that, for the illustration purpose, only the heat map of Model 1 is shown in Figure 2.5. Other heat maps are shown in Appendix 3.

### 2.3.9 Train and test data

Data is split into train data (or in-sample data) and test data (or out-of-sample data). Train data is used to build and train models and unseen test data will be used to validate the models.

In order to quantify the effects of natural disasters on firm performance in the linear regression and CART model for the original data, we use the 50/50 split ratio in this study for two reasons. The reason is because, on average, the billion-dollar disasters are rare events at the state level and also at the firm level. Therefore, for the linear regression and CART models, we randomly split data into 50% of training data and 50% of test data to avoid the bias of splitting due to the nature of rare events (billion-dollar natural disasters).

In order to compare the performances among neural networks, CART, and the benchmark linear regression models using the scaled data, we use the most common split ratios in machine learning including 80/20, 90/10, and 50/50[23]. Note that the 80/20 split ratio means that the entire data is split into 80% train data and 20% test data (hold-out sample), and similarly for other split ratios. In turn, the train data is split into 80% of training data and 20% of validation data for the 80/20 split, which yields 64% ($0.8 * 80\%$) training data, 16%($0.2 * 80\%$) validation data, and 20% unseen test data. Similarly for other splits. Table 2.19 shows the proportions of training data, validation data, and test data for neural networks models.

Table 2.19: Training, validation, and test data for different splits

| Split ratio | Training data | Validation data | Test data |
|---|---|---|---|
| 80/20 | 64% | 16% | 20% |
| 90/10 | 81% | 9% | 10% |
| 50/50 | 25% | 25% | 50% |

## 2.4 Results

The results in this paper are obtained from running data using Python version 3.7, Numpy version 1.19.5, and Tensorflow version 2.7.0.

### 2.4.1 The linear regression results

**The effects of natural hazards and disasters on firm performance**. We can observe from Table 2.20 that the observed damages of natural disasters are not associated with any firm performance criteria. However, the perceived natural disaster risk (i.e. `disaster10K`) are negatively associated with the firm's profitability (ROA and EBITAT) at the 99% and 90% confidence level, respectively. Moreover, perceived natural disaster risk is not associated with Tobin's Q and sales growth.

Let's dive deeper into the meaning of the coefficients of the variables of interest. The coefficient of `disaster10K` for ROA equals to $-0.0004$, which means that if Form 10-Ks include one more additional word related to natural disasters then ROA tends to decrease by 0.0004%. Linearly, we can say that if Form 10-Ks include 10 more additional words (which approximately equals the mean of `disaster10K`) related to natural disasters then ROA is likely to decrease by 0.004%. Also, if Form 10-Ks include 15 more additional words (which is approximately 1 standard deviation) related to natural disasters then ROA is likely to decrease by 0.006%. These results imply that the effects of perceived natural disasters on

[23] Some previous studies also utilize these three split ratios of 50/50, 80/20, and 90/10, for example Wilson and Sharda (1994).

ROA is insignificant. Moreover, the effects of perceived natural disasters on EBITAT is two times smaller than that on ROA.

Table 2.20: Linear regression results

| | roa | ebitat | Q | salegrowth |
|---|---|---|---|---|
| const | -0.2226*** | -0.1496*** | 0.7068*** | 0.0278 |
| lctat | -0.0929*** | -0.0555*** | 0.4708*** | 0.1444*** |
| cheat | -0.0533*** | -0.0738*** | 0.7328*** | 0.0485*** |
| chlct | -0.0024 | -0.0028*** | 0.0295*** | 0.0129*** |
| lctlt | -0.0198** | -0.0087 | 0.0124 | -0.0133 |
| saleat | 0.0537*** | 0.0373*** | -0.1842*** | -0.0732*** |
| logat | 0.0307*** | 0.0233*** | -0.0765*** | -0.0111*** |
| reat | 0.0097*** | 0.0048*** | -0.0234*** | 0.0066*** |
| relct | 0.0036*** | 0.0028*** | -0.0016 | -0.0047*** |
| dtat | 0.0121 | 0.0244*** | 0.0156 | -0.0131 |
| seqat | 0.0805*** | 0.0596*** | 0.0140 | 0.0136 |
| invtsale | -0.0116 | -0.0108 | 0.1272* | 0.1112*** |
| xintsale | -0.0917*** | -0.0600*** | 0.1430*** | 0.1845*** |
| atemp | 0.0000 | -0.0000* | 0.0000** | 0.0000 |
| invtgrowth | 0.0065** | 0.0032 | -0.0298** | 0.0033 |
| hhi | 0.0233 | 0.0205 | -0.2553 | 0.0892 |
| gdpgrowth | -0.0030 | -0.0013 | 0.0051 | -0.0059* |
| inflation | -0.0089 | -0.0111 | 0.1078* | -0.0168 |
| interestrate | 0.0005 | 0.0004 | 0.0065 | 0.0038* |
| roalag1 | 0.4251*** | | | |
| ebitatlag1 | | 0.5423*** | | |
| Qlag1 | | | 0.6879*** | |
| salegrowthlag1 | | | | 0.0794*** |
| xintgrowth | -0.0021*** | 0.0006 | -0.0017 | 0.0131*** |
| capxgrowth | -0.0100*** | -0.0062*** | -0.0355*** | 0.0176*** |
| empgrowth | 0.0478*** | 0.0452*** | -0.5972*** | 0.6607*** |
| dvcib | 0.0147*** | 0.0108*** | 0.0512* | 0.0110 |
| disaster | -0.0002 | -0.0002 | -0.0015 | -0.0003 |
| disaster0 | -0.0002 | -0.0002 | 0.0007 | -0.0002 |
| disaster10K | -0.0004*** | -0.0002* | -0.0003 | 0.0002 |
| Year FE | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes |
| N | 48881 | 48881 | 48881 | 48881 |
| $R^2$ | 0.5587 | 0.6438 | 0.5716 | 0.2386 |

*Notes*: In this table, each column shows the coefficients and its levels of significance in the corresponding model. Here, $*$, $**$, and $***$ stand for $p$-value are less than 0.1, 0.05, and 0.01, respectively.

**Natural hazards vs. Natural disasters**. Table 2.21 compares the effects of the U.S. natural hazards and U.S. (billion-dollar) natural disasters together with the perceived natural disaster risk on firm performance. It can be seen that the effects of perceived natural disaster risk on firm profitability are the same in both cases. However, the effects of government-reported damages of natural hazards on firm profitability are different with that of the billion-dollar natural disasters. Specifically, the damages of natural hazards are negatively associated with firm profitability while that of the billion-dollar natural disasters have no effects on firm profitability. Regarding Tobin's Q and sales growth, both the damages of natural hazards and that of the billion-dollar natural disasters have no effects on these criteria of firm performance.

Table 2.21: The effects of natural hazards vs. natural disasters on firm performance

| | Natural hazards | | | | Billion-dollar natural disasters | | | |
|---|---|---|---|---|---|---|---|---|
| | `roa` | `ebitat` | `Q` | `salegrowth` | `roa` | `ebitat` | `Q` | `salegrowth` |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| `damage` | -0.0012*** | -0.0009*** | 0.0004 | -0.0006 | | | | |
| `damage0` | -0.0001 | -0.0004 | -0.0023 | 0.0004 | | | | |
| `disaster` | | | | | -0.0002 | -0.0002 | -0.0015 | -0.0003 |
| `disaster0` | | | | | -0.0002 | -0.0002 | 0.0007 | -0.0002 |
| `disaster10K` | -0.0004** | -0.0002* | -0.0007 | 0.0002 | -0.0004*** | -0.0002* | -0.0003 | 0.0002 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 46376 | 46376 | 46376 | 46376 | 48881 | 48881 | 48881 | 48881 |
| $R^2$ | 0.5792 | 0.6567 | 0.5839 | 0.2347 | 0.5587 | 0.6438 | 0.5716 | 0.2386 |

**Natural disasters vs. Perceived risk of natural disasters**. We now run the models with only natural disasters and only perceived risk of natural disasters to examine the sensitivity of the coefficients of variables of interest. Table 2.22 compares the results of the variables of interest in the case of billion-dollar natural disasters only and perceived risk of natural disasters only. Comparing the coefficients of variables of interest in the model of natural disasters only, perceived risk of natural disasters only, and both, we observe that the signs, magnitudes, and levels of statistical significance of the coefficients in the first two models are exactly the same as that in the third model. This implies that the regression coefficients are robust either we take into account natural disaster and perceived risk of natural disasters separately or at the same time.

Table 2.22: The effects of the natural disasters vs. perceived risk of natural disasters

| | Natural disasters | | | | Perceived risk of natural disasters | | | | Both | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | `roa` | `ebitat` | `Q` | `saleg` | `roa` | `ebitat` | `Q` | `saleg` | `roa` | `ebitat` | `Q` | `saleg` |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| `disaster` | -0.0003 | -0.0002 | -0.0015 | -0.0003 | | | | | -0.0002 | -0.0002 | -0.0015 | -0.0003 |
| `disaster0` | -0.0002 | -0.0002 | 0.0007 | -0.0002 | | | | | -0.0002 | -0.0002 | 0.0007 | -0.0002 |
| `disaster10K` | | | | | -0.0004*** | -0.0002* | -0.0003 | 0.0002 | -0.0004*** | -0.0002* | -0.0003 | 0.0002 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 48881 | 48881 | 48881 | 48881 | 48881 | 48881 | 48881 | 48881 | 48881 | 48881 | 48881 | 48881 |
| $R^2$ | 0.5586 | 0.6438 | 0.5716 | 0.2386 | 0.5587 | 0.6438 | 0.5715 | 0.2386 | 0.5587 | 0.6438 | 0.5716 | 0.2386 |

*Note*: `saleg` stands for sales growth and $*p < .1, **p < .05, ***p < .01$.

The results and implications in the case of natural hazards are similar compared to that of natural disasters. Note that the perceived risk of natural disasters are the same in both Table 2.22 and Table 2.23.

Table 2.23: Natural hazards vs. perceived risk of natural disasters

| | Natural hazards | | | | Perceived risk of natural disasters | | | | Both | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | roa | ebitat | Q | saleg | roa | ebitat | Q | saleg | roa | ebitat | Q | saleg |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| damage | -0.0012*** | -0.0009*** | 0.0004 | -0.0006 | | | | | -0.0012*** | -0.0009*** | 0.0004 | -0.0006 |
| damage0 | -0.0001 | -0.0005 | -0.0023 | 0.0004 | | | | | -0.0001 | -0.0004 | -0.0023 | 0.0004 |
| disaster10K | | | | | -0.0004*** | -0.0002* | -0.0007 | 0.0002 | -0.0004** | -0.0002* | -0.0007 | 0.0002 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 46376 | 46376 | 46376 | 46376 | 46376 | 46376 | 46376 | 46376 | 46376 | 46376 | 46376 | 46376 |
| $R^2$ | 0.5792 | 0.6567 | 0.5839 | 0.2347 | 0.5791 | 0.6566 | 0.5839 | 0.2346 | 0.5792 | 0.6567 | 0.5839 | 0.2347 |

*Note*: saleg stands for sales growth and $*p < .1, **p < .05, ***p < .01$.

**Manufacturing vs. Services sector**. Since manufacturing and services are two dominant sectors in the U.S. economy (see Table 2.10), it is worth to examine the effects of natural disasters on these sectors separately. Specifically, the effects of natural disasters and natural hazards on firm performance of manufacturing sector, services sector, and the whole economy are shown in Table 2.24 and Table 2.25, respectively. We can observe in both tables that, for the manufacturing sector, perceived natural uncertainty is negatively associated with Tobin's Q. Meanwhile, for the services sector actual damages in year t+1 and perceived environmental uncertainty in year t are negatively associated with firm profitability in year t+1.

Table 2.24: The effects of natural disasters on the manufacturing vs. services sector

| | Manufacturing sector | | | | Services sector | | | | Whole economy | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | roa | ebitat | Q | saleg | roa | ebitat | Q | saleg | roa | ebitat | Q | saleg |
| controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| disaster | -0.0002 | -0.0002 | -0.0007 | -0.0002 | -0.0001 | -0.0003 | -0.0002 | -0.0007 | -0.0002 | -0.0002 | -0.0015 | -0.0003 |
| disaster0 | 0.0001 | -0.0001 | -0.0016 | 0.0002 | -0.0014** | -0.0009** | 0.0052 | -0.0010 | -0.0002 | -0.0002 | 0.0007 | -0.0002 |
| disaster10K | 0.0003 | 0.0002 | -0.0032*** | 0.0006 | -0.0011* | -0.0009** | 0.0056 | -0.0004 | -0.0004*** | -0.0002* | -0.0003 | 0.0002 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 24269 | 24269 | 24269 | 24269 | 10713 | 10713 | 10713 | 10713 | 48881 | 48881 | 48881 | 48881 |
| $R^2$ | 0.6209 | 0.7058 | 0.6029 | 0.1917 | 0.5459 | 0.6143 | 0.5475 | 0.2996 | 0.5587 | 0.6438 | 0.5716 | 0.2386 |

*Note*: saleg stands for sales growth and $*p < .1, **p < .05, ***p < .01$.

Table 2.25: The effects of natural hazards on the manufacturing vs. services sector

| | Manufacturing sector | | | | Services sector | | | | Whole economy | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | roa | ebitat | Q | saleg | roa | ebitat | Q | saleg | roa | ebitat | Q | saleg |
| controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| damage | -0.0005 | -0.0003 | 0.0004 | 0.0021 | -0.0023 | -0.0015 | 0.0079 | -0.0016 | -0.0012*** | -0.0009*** | 0.0004 | -0.0006 |
| damage0 | 0.0003 | -0.0001 | 0.0033 | -0.0003 | -0.0020 | -0.0022** | 0.0029 | -0.0013 | -0.0002 | -0.0002 | 0.0007 | -0.0002 |
| disaster10K | -0.0001 | 0.0000 | -0.0023* | 0.0012*** | -0.0014** | -0.0010** | 0.0021 | -0.0005 | -0.0004*** | -0.0002* | -0.0003 | 0.0002 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 22909 | 22909 | 22909 | 22909 | 10377 | 10377 | 10377 | 10377 | 48881 | 48881 | 48881 | 48881 |
| $R^2$ | 0.6083 | 0.6894 | 0.6022 | 0.1913 | 0.5372 | 0.6171 | 0.5556 | 0.3195 | 0.5587 | 0.6438 | 0.5716 | 0.2386 |

*Note*: saleg stands for sales growth and $*p < .1, **p < .05, ***p < .01$.

**Small vs. Large firms**. We check the regression results according to the firm's size by splitting the data into two groups: Small firms and big firms. Small firms are those have total assets less than the median of total assets of all firms (237.119 million USD) while big firms are those have total assets greater than or equal the median of total assets of all firms. The regression results of small and big firms corresponding to the billion-dollar natural disasters and natural hazards are shown in Table 2.26 and Table 2.27, respectively. The effects of natural disasters on firm profitability is consistently negatively in all three cases (small, large, or all firms).

Table 2.26: The effects of natural disasters on small vs. large firms

| | Small firms | | | | Large firms | | | | All firms | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | roa | ebitat | Q | saleg | roa | ebitat | Q | saleg | roa | ebitat | Q | saleg |
| controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| disaster | 0.0001 | -0.0003 | -0.0007 | -0.0021 | -0.0002*** | -0.0002*** | -0.0001 | -0.0003*** | -0.0002 | -0.0002 | -0.0015 | -0.0003 |
| disaster0 | -0.0009 | -0.0015*** | 0.0061* | 0.0005 | 0.0000 | -0.0000 | -0.0003 | -0.0000 | -0.0002 | -0.0002 | 0.0007 | -0.0002 |
| disaster10K | -0.0002 | -0.0001 | -0.0059* | 0.0001 | -0.0001*** | -0.0001** | 0.0002 | 0.0002*** | -0.0004*** | -0.0002* | -0.0003 | 0.0002 |
| N | 24679 | 24679 | 24679 | 24679 | 23766 | 23766 | 23766 | 23766 | 48881 | 48881 | 48881 | 48881 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.5230 | 0.5955 | 0.5313 | 0.1815 | 0.4469 | 0.6612 | 0.7321 | 0.3873 | 0.5587 | 0.6438 | 0.5716 | 0.2386 |

*Note*: saleg stands for sales growth and $*p < .1, **p < .05, ***p < .01$.

Table 2.27 shows a clear and consistent results that natural hazards are consistently negatively associated with firm profitability. Moreover, for large firms, natural hazards in year $t$ are negatively associated with sales growth in year $t+1$.

Table 2.27: The effects of natural hazards on small vs. large firms

| | Small firms | | | | Large firms | | | | All firms | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | roa | ebitat | Q | saleg | roa | ebitat | Q | saleg | roa | ebitat | Q | saleg |
| controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| damage | 0.0003 | 0.0000 | -0.0094 | 0.0032 | -0.0003*** | -0.0002** | 0.0010 | -0.0005** | -0.0002 | -0.0002 | -0.0015 | -0.0003 |
| damage0 | -0.0007 | -0.0025*** | 0.0062 | -0.0018 | 0.0001 | -0.0001 | -0.0004 | -0.0002 | -0.0002 | -0.0002 | 0.0007 | -0.0002 |
| disaster10K | -0.0013** | -0.0010** | -0.0052 | 0.0014 | -0.0000 | 0.0000 | 0.0004 | 0.0001 | -0.0004*** | -0.0002* | -0.0003 | 0.0002 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 23430 | 23430 | 23430 | 23430 | 22529 | 22529 | 22529 | 22529 | 48881 | 48881 | 48881 | 48881 |
| $R^2$ | 0.5188 | 0.5814 | 0.5106 | 0.1807 | 0.4429 | 0.6669 | 0.7302 | 0.4076 | 0.5587 | 0.6438 | 0.5716 | 0.2386 |

*Note*: saleg stands for sales growth and $*p < .1, **p < .05, ***p < .01$.

### 2.4.2 Goodness of fit

Recall that the goodness of fit (or $R^2$) is a measure that explains the variation of the dependent (or target) variable due to the variation of the independent variables (or features). Formally, the goodness of fit metric is defined as follows

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2},$$

where $n$ is the size of the test data, $y_i, \hat{y}_i$, and $\bar{y}$ are the ground truth, estimated, and mean of $y_i$ values, respectively.

Note that even though some previous studies criticize $R^2$ and prefer error measures (e.g. Razi and Athappilly 2005), $R^2$ still brings some useful insights about the models and used as a criterion for evaluating forecasting models (e.g. Swanson and White 1995). In order to avoid the bias in evaluating the models, we decided to use both $R^2$ and error measures to compare the performances among these models, including linear regression, CART, and DNNs. However, we need to keep in mind that error measures seem more reliable criteria to evaluate the models in relation to $R^2$. In this subsection, we focus on evaluation models using $R^2$.

**The architecture of the CART models**. In order to find a good architecture we use two steps: trial and error method, and grid search. we first use the trial and error method to derive the lists of hyperparameters in Table 2.28. The trial and error step is very important since by doing this we have an insight of what are the suitable ranges of hyperparameters that we should choose for hyperparameter optimization.

Table 2.28: The list of hyperparameters for the CART models

| Hyperparameter | Range |
|---|---|
| `list_max_depth` | [None, 5, 6, 7, 8, 9, 10, 11, 12] |
| `list_max_features` | [None, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00] |
| `list_min_samples_split` | [2, 3] |
| `list_min_samples_leaf` | [10, 20, 30, 40, 50, 60, 70, 80, 90, 100] |

*Notes*: For the CART models, here we only consider four important and common hyperparameters including `max_depth`, `max_features`, `min_samples_split`, and `min_samples_leaf`. However, the CART models indeed have several more hyperparameters such as `max_leaf_nodes`, `min_weight_fraction_leaf`, `max_leaf_nodes`, `min_impurity_decrease`, and `ccp_alpha`.

We then use grid search (multiple for loops) based on the lists of hyperparameters from the trial and error method to tune the hyperparameters to derive the optimal hyperparameters for each target variable.

Table 2.29: Two methods using for hyperparameter optimization

| Step | Method | Output |
|---|---|---|
| Step 1 | Trial and error | "Best" of lists of hyperparameters |
| Step 2 | Grid search | Optimal combination of hyperparameters |

*Notes*: Step 1 is necessary because without this step, Step 2 (grid search) would take too much time to be implemented. If the ranges of hyperparameters are not restricted at Step 1, it might take many days to run just a model at the Step 2 stage. For Step 1 itself, to obtain the "best" lists of hyperparameters, hundreds of lists have been run to select the best ones. That is why Step 1 is called *trial and error method*.

The optimal hyperparameters for the models for predicting ROA, EBITAT, Tobin's Q, and sales growth are shown respectively in Table 2.30. Note that, for simplicity, in this table we use a dependent variable to represent for each corresponding model.

Table 2.30: Hyperparameters optimization

| | roa | ebitat | Q | salegrowth |
|---|---|---|---|---|
| max_depth | 9 | 12 | 10 | 11 |
| max_features | 0.85 | 0.7 | 0.65 | 0.9 |
| min_samples_split | 2 | 2 | 2 | 2 |
| min_samples_leaf | 90 | 60 | 100 | 80 |

The comparison of goodness of fit $R^2$ for the original data between linear regression and CART is shown in Table 2.31[24]. It can be seen that the $R^2$ of CART and that of the linear regression are not much different. The goodness of fit of CART is slightly higher than that of linear regression model for the models of ROA and sales growth while $R^2$ of the linear regression models are slightly higher than that of CART for the models of EBITAT and Tobin's Q[25].

Table 2.31: Comparison goodness of fit among models with test data

| | Linear regression | CART |
|---|---|---|
| roa | 0.5587 | 0.5618 |
| ebitat | 0.6438 | 0.6399 |
| Q | 0.5716 | 0.5668 |
| salegrowth | 0.2386 | 0.2496 |

*Notes*: This table shows that the goodness of fit ($R^2$) of the linear regression and CART models are not much different in all models. This implies $R^2$ might not be the best predictive metrics to evaluate the models. In the next subsections, we will investigate two important metrics that are feature importance and prediction errors.

### 2.4.3 Feature importance

Feature importance is a scoring metric used to measure the relative importance of each input feature in relation to other features of a predictive model. Specifically, feature importance

[24] Not that since $R^2$ is not a common criterion of neural networks, we only measure $R^2$ of CART and linear regression model.

[25] Note that the list of variables in this study is restricted by some restrictions of the linear regression model. For example, in order to avoid the severe multicollinearity problem and highly correlated between variables, we removed several independent variables out of the initial list. Since multicollinearity problem is an issue for linear regression, but not for the CART model, adding some more variables (which we cannot add to the linear regression model due to the multicollinearity issue) to the CART model might help increase $R^2$ of this model.

is a technique help to select the most important variables in the sense that how much each of these variables help to reduce the level of predicted error (i.e. impurity or uncertainty) of the target variables[26]. In the CART model, impurity can be understood as the level of uncertainty present at every node. The goal of the CART model is to minimize impurity (or minimize uncertainty) of the target variables at each of the leaf neurons.

There are two main roles of the feature importance metric. Firstly, the feature importance metric provides insights for the feature selection process. A natural way is to keep those features that help reduce the level of impurity the most and remove the variables that contribute nothing or very little to decrease the level of impurity. Note that even if we do not want to remove any features, we still know which variables contribute to the most and which variables contribute the least. Secondly, as a consequence of the first role, by removing some irrelevant (non-informative) or redundant features, the feature importance metric is a good tool for the dimensionality reduction purpose and decreasing the computational cost of processing data and training models. Thirdly, reducing the number of inputs by removing irrelevant or redundant features might help to improve the performances of our models in terms of error measures.

We can observe one noticeable result from Table 2.32 that the uncertainty variables (`disaster0`, `disaster`, and `disaster10K`) are much relatively less important, in the sense of helping to decrease the impurity level, compared to lagged dependent variables (e.g. `roalag1`, `ebitatlag1`, `Qlag1`, `salegrowthlag1`) and financial statement state variables (e.g. `empgrowth`, `reat`, `logat`, `saleat`, `seqat`).

[26] A good reference for data-driven feature selection in the CART model is Gocheva-Ilieva et al. (2019).

Table 2.32: Feature importance of the CART models

| Model 1 (ROA) | | Model 2 (EBITAT) | | Model 3 (Tobin's Q) | | Model 4 (sales growth) | |
|---|---|---|---|---|---|---|---|
| variable | impurity decrease | variable | impurity decrease | variable | impurity decrease | variable | impurity decrease |
| roalag1 | 0.731264 | ebitatlag1 | 0.894507 | Qlag1 | 0.910080 | empgrowth | 0.554372 |
| reat | 0.142301 | logat | 0.024963 | logat | 0.024171 | saleat | 0.215514 |
| logat | 0.046613 | seqat | 0.021086 | empgrowth | 0.012403 | salegrowthlag1 | 0.110306 |
| seqat | 0.023266 | reat | 0.018697 | capxgrowth | 0.008595 | reat | 0.037373 |
| saleat | 0.014844 | saleat | 0.009083 | reat | 0.007689 | xintgrowth | 0.028039 |
| xintsale | 0.013228 | xintsale | 0.008432 | relct | 0.005536 | capxgrowth | 0.020958 |
| lctat | 0.007838 | relct | 0.007126 | inflation | 0.005440 | logat | 0.005717 |
| relct | 0.004745 | empgrowth | 0.005834 | gdpgrowth | 0.005412 | lctat | 0.005061 |
| capxgrowth | 0.003734 | lctat | 0.002278 | cheat | 0.004037 | disaster10K | 0.004982 |
| empgrowth | 0.003187 | interestrate | 0.001717 | seqat | 0.003520 | invtsale | 0.004648 |
| interestrate | 0.002582 | capxgrowth | 0.001518 | interestrate | 0.002652 | cheat | 0.003900 |
| cheat | 0.002469 | chlct | 0.001170 | lctlt | 0.002152 | inflation | 0.002572 |
| dvcib | 0.001411 | lctlt | 0.001019 | xintsale | 0.001980 | gdpgrowth | 0.001174 |
| invtsale | 0.001158 | dtat | 0.000797 | xintgrowth | 0.001278 | chlct | 0.001124 |
| invtgrowth | 0.000432 | dvcib | 0.000701 | disaster10K | 0.001219 | seqat | 0.000819 |
| xintgrowth | 0.000313 | xintgrowth | 0.000417 | chlct | 0.000827 | hhi | 0.000779 |
| inflation | 0.000281 | atemp | 0.000171 | saleat | 0.000756 | interestrate | 0.000751 |
| gdpgrowth | 0.000130 | cheat | 0.000147 | atemp | 0.000637 | lctlt | 0.000727 |
| chlct | 0.000074 | invtgrowth | 0.000145 | invtgrowth | 0.000503 | atemp | 0.000639 |
| dtat | 0.000044 | invtsale | 0.000073 | lctat | 0.000433 | invtgrowth | 0.000395 |
| disaster10K | 0.000039 | gdpgrowth | 0.000045 | invtsale | 0.000372 | relct | 0.000151 |
| lctlt | 0.000037 | inflation | 0.000032 | hhi | 0.000200 | dtat | 0.000000 |
| atemp | 0.000008 | disaster10K | 0.000025 | dtat | 0.000070 | xintsale | 0.000000 |
| disaster0 | 0.000001 | hhi | 0.000016 | disaster0 | 0.000038 | dvcib | 0.000000 |
| hhi | 0.000000 | disaster0 | 0.000000 | dvcib | 0.000000 | disaster0 | 0.000000 |
| disaster0 | 0.000000 | disaster | 0.000000 | disaster | 0.000000 | disaster | 0.000000 |

A node will be split if this split induces a decrease of the impurity (mean square error) greater than or equal to this current value[27]. For example, we illustrate the trees with max_depth=2. Intuitively, the trees with mean squared error (squared_error), number of samples (samples), and the average predicted value of ROA, EBITAT, Tobin's Q, and sales growth are respectively shown in Figure 2.6, Figure 2.7, Figure 2.8, and Figure 2.9.

[27] https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html

Figure 2.6: Decision tree of the CART model for predicting ROA

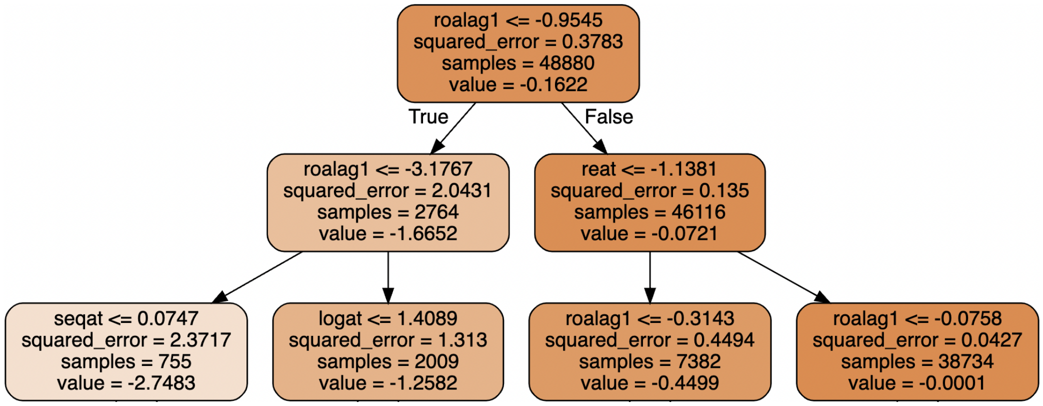



Figure 2.7: Decision tree of the CART model for predicting EBITAT

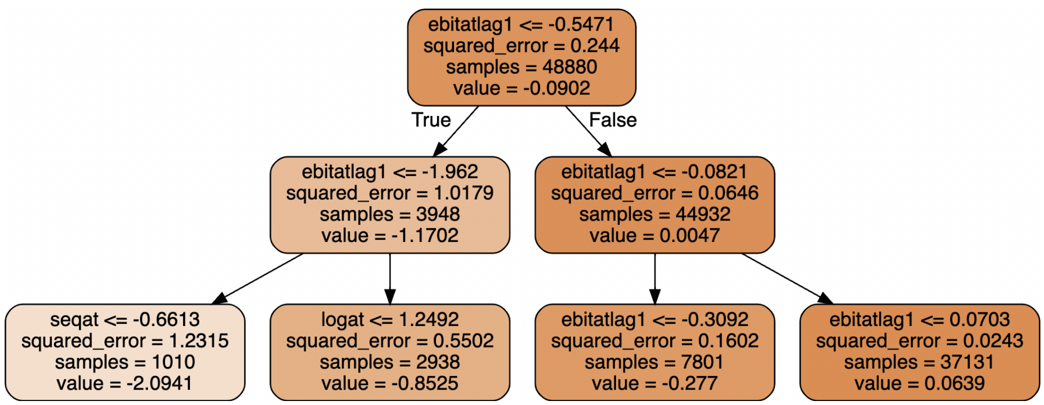



Figure 2.8: Decision tree of the CART model for predicting Tobin's Q

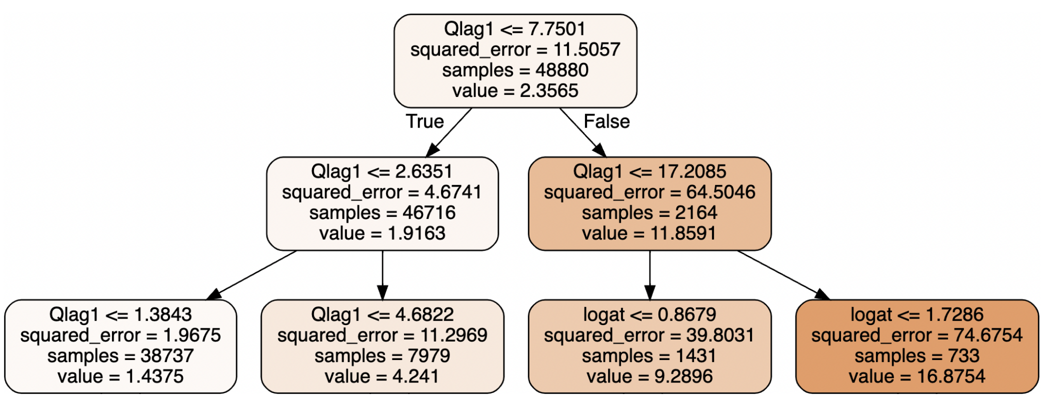
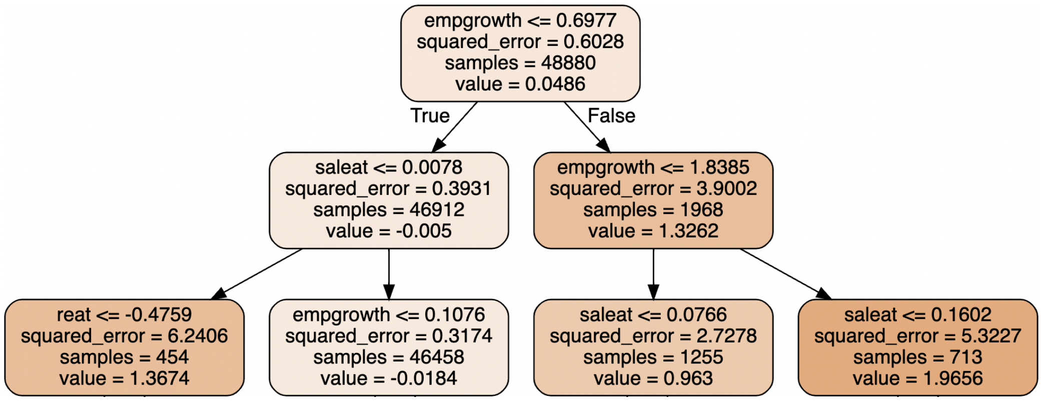


Figure 2.9: Decision tree of the CART model for predicting sales growth




### 2.4.4 Prediction errors analysis

Since we are working on the prediction problems, a natural way to measure the quality of our predictive models is how close from predicted values to the actual ground truth values. That is, how large the error measures are. Error measures are popular indicators used to compare the performances between linear regression model and DNNs (e.g. Doganis et al. 2006; Anyaeche and Ighravwe 2013; Pombeiro et al. 2017). Formally, the mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) are defined as follows.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|,$$

where $n$ is the size of the test set, $y_i$ and $\hat{y}_i$ are the actual and predicted values, respectively.

Note that there is no absolute criterion for a "good" value of MSE, RMSE, or MAE. The best way we can do this is to compare these criteria among models. The smaller error measures the better. Since the definition and the behavior of each dependent variable is different compared to the others, we are going to investigate four different models corresponding to four different firm performance criteria including ROA, EBITAT, Tobin's Q, and sales growth.

### Model 1: Predicting ROA

**Neural networks**. Neural networks with one hidden layer, two hidden layers, three hidden layers, and four hidden layers are denoted as ANN, DNN[2], DNN[3], and DNN[4], respectively. Google Colab with the Numpy version 1.19.5 and the Tensorflow version 2.7.0 are used to run the neural network models. Choosing the optimal number of neurons is the first important step to build the architecture of neural networks.

*Hyperparameter tuning the number of neurons.* Finding the optimal architectures for neural networks depends critically on determining a good set of hyperparameters. Since there are many hyperparameters, there are many ways that we can tune the hyperparameters. First, we need to find the optimal number of neurons (or units) in each layer. In order to tune hyperparameters we use the Keras tuner library. This library contains four types of algorithms, including Hyperband, random search, sklearn, and Bayesian optimization. We choose Hyperband for tuning the number of inputs for the input and hidden layers[28]. Note that, in our architecture of neural networks, the number of neurons in the input layer equals the number of independent variables (features), which is 27, and the number of output is only one (e.g ROA). The following table shows the optimal number of neurons for the hidden layer with `epochs=500`.

Table 2.33: The optimal number of neurons of neural networks for predicting ROA

| | min_value | max_value | Step | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|---|---|---|
| Input layer | 27 | 27 | - | 27 | 27 | 27 | 27 |
| Hidden layer [1] | 8 | 512 | 4 | 464 | 272 | 456 | 360 |
| Hidden layer [2] | 8 | 512 | 4 | - | 488 | 288 | 368 |
| Hidden layer [3] | 8 | 512 | 4 | - | - | 272 | 288 |
| Hidden layer [4] | 8 | 512 | 4 | - | - | - | 212 |
| Output layer | 1 | 1 | - | 1 | 1 | 1 | 1 |

Now, we are going to identify the types of activation function for each layer. For the output layer, since our problem is a prediction problem with continuous target variables it is classified as *regression* problem. In practice, the common activation function for regression problems is linear. For the input and hidden layer, in practice, there are three common activation functions for input and hidden layers including Rectified Linear Unit (ReLU), logistic (sigmoid), and hyperbolic tangent (tanh). In modern neural networks, the default recommendation is to use the ReLU activation function (Goodfellow et al. 2016). Therefore,

[28]For more details about the Hyperband algorithm, see Li et al. (2018) or the republised version at **?**. Li et al. (2018) compare the Hyperband algorithm and the popular Bayesian optimization methods and find that the Hyperband algorithm “can provide over an order-of-magnitude speedup over our competitor set on a variety of deep-learning and kernel-based learning problems”.

we use the ReLU activation function in this model. Recall that a ReLU activation function has the form $g(z) = \max\{0, z\}$.

The architectures of all neural networks models are shown in Table 2.34. From the number of inputs in Table 2.33, we obtain the number of parameters in Table 2.34. The number of parameters is calculated as follows. Suppose $n$ is the number of neurons in the previous layer and $m$ is the number of neurons in the current layer. By adding the bias to the previous layer, we have $n + 1$ neurons. The number of parameters between the previous layer and the current layer is equal to $(n + 1)m$. Specifically, the number of trainable parameters of neural networks for predicting ROA is shown as follows[29].

Table 2.34: The number of trainable parameters of neural networks for predicting ROA

| | Activation function | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|---|
| Input layer | ReLU | - | - | - | - |
| Hidden layer [1] | ReLU | 12992 | 7616 | 12768 | 10080 |
| Hidden layer [2] | ReLU | - | 133224 | 131616 | 132848 |
| Hidden layer [3] | ReLU | - | - | 78608 | 106272 |
| Hidden layer [4] | ReLU | - | - | - | 61268 |
| Output layer | linear | 465 | 489 | 273 | 213 |
| Total parameters | - | 13,457 | 141,329 | 223,265 | 310,681 |

*Hyperparameter tuning number of epochs.* In order to find the optimal number of epochs, we first run the Keras sequential `model.fit()` function for ROA with `epochs=500`, `batch_size=32`, `validation_split=0.5`[30]. For example, for the 80/20 split data, we run the `index(min(val_mse_per_epoch))` function and get the optimal number of epochs for ANN, DNN[2], DNN[3], and DNN[4], which are 282, 157, 157, and 103, respectively. We then rerun the `model.fit()` function with these optimal epochs, i.e. epochs=282 for ANN, epochs=157 for DNN[2], epochs=157 for DNN[3], and epochs=103 for DNN[4]. Finally, we evaluate the model on the test data by using the `model.evaluate()` function to obtain the error measures for unseen test data.

[29] For example, for the ANN model, the number of parameters between the input layer and the hidden layer equals to $(27 + 1) \times 464 = 12,992$. Also, the number of parameters between the hidden layer and the output layer equals $(464 + 1) \times 1 = 465$. Therefore, the total parameters equals $12,992 + 465 = 13,457$. The numbers of parameters in other neural network models are calculated in a similar manner.

[30] Note that the validation_split parameter shows the fraction of training data to be used as validation data. Since about 66.66% of our data is training set, the validation_split=0.5 implies that the fraction of validation data is approximately 33.33% of the entire data. So the fraction of each train set, validation set, and test set is equally one third of the data or approximately 33.33%.

Table 2.35: Optimal epochs of neural networks for predicting ROA

| Model | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|
| Epochs range | [1, 500] | [1, 500] | [1, 500] | [1, 500] |
| Optimal epoch (80/20 split) | 282 | 157 | 157 | 112 |
| Optimal epoch (90/10 split) | 405 | 197 | 158 | 119 |
| Optimal epoch (50/50 split) | 384 | 204 | 160 | 132 |

*Visualization of MSE and MAE of training and validation data.* The variations of the mean squared error (MSE) and mean absolute error (MAE) of training and validation data with the 80/20 split ratio of the ANN, DNN[2], DNN[3], and DNN[4] model are shown in Figure 2.10-2.13. We can observe that both MSE and MAE drop dramatically at the very few initial epochs then decrease very slowly after that before going up eventually.

Figure 2.10: The MSE and MAE loss function of the ANN model for predicting ROA

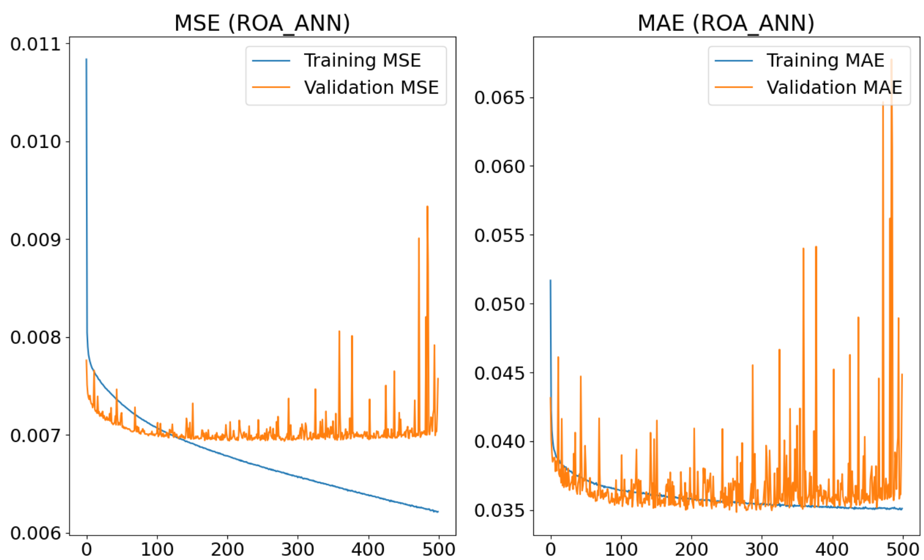


Figure 2.11: The MSE and MAE loss function of the DNN[2] model for predicting ROA

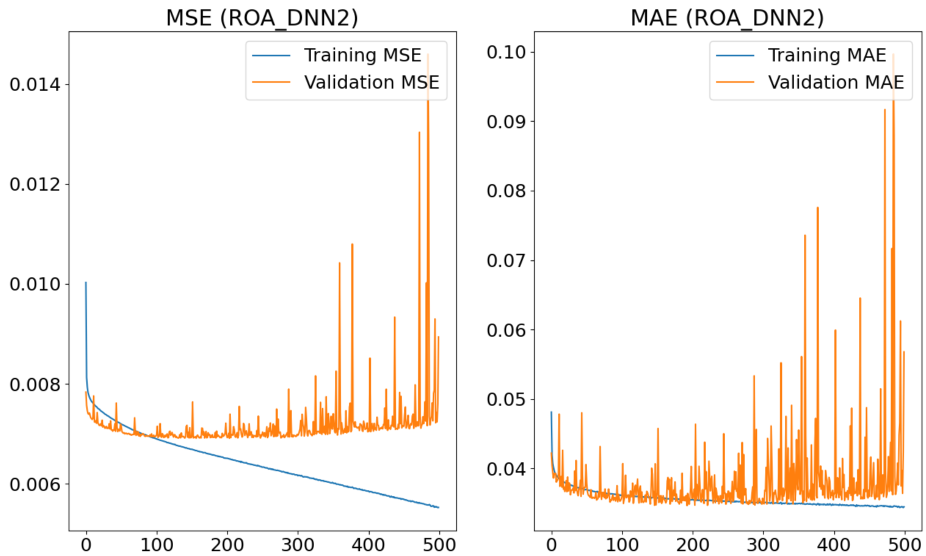



Figure 2.12: The MSE and MAE loss function of the DNN[3] model for predicting ROA

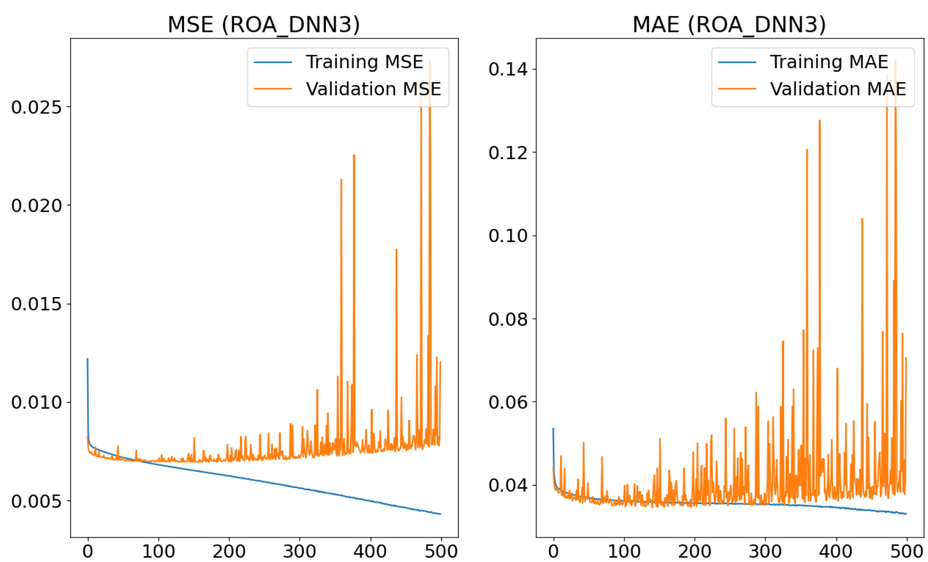
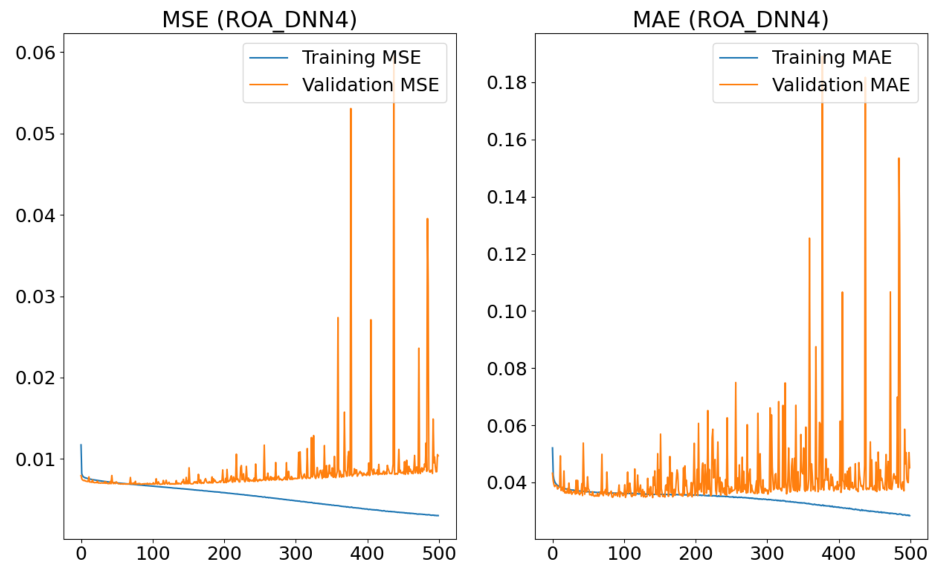


Figure 2.13: The MSE and MAE loss function of the DNN[4] model for predicting ROA




**CART model**. For the CART model, first the trial and error method is used to derive a good list of values for each hyperparameter. Four common hyperparameters and their lists of values are shown in Table 2.28. Then, four *for loops* are used to examine all combinations of the lists of given values of four hyperparameters. The optimal combinations of the hyperparameters for the CART models for testing data is shown in Table 2.30.

*Performance comparison.* Table 2.36 compares the performances on the test data (or holdout sample) between CART, several neural network models, and the benchmark linear regression model with different split ratios (80/20, 90/10, and 50/50 train and test set respectively) and in terms of MSE, RMSE, and MAE. It can be seen that CART and neural network models outperform linear regression in all given criteria and with any given split ratios in predicting ROA. Comparing CART and neural networks, it is clear that CART is better than NNs if the split ratio is 50/50, and if the split ratio is 80/20 and in terms of MAE. Meanwhile, neural networks are better than CART if the split ratio is 90/10, and if the split ratio is 80/20 and in terms of MSE or RMSE. Notice that simply increasing the number of hidden layers does not consistently improve the accuracy of predictions. This finding is similar to the result in Chen et al. (2017).

Table 2.36: Comparison of models performance on the test data for predicting ROA

| | 80/20 split | | | 90/10 split | | | 50/50 split | | |
|---|---|---|---|---|---|---|---|---|---|
| | MSE | RMSE | MAE | MSE | RMSE | MAE | MSE | RMSE | MAE |
| Regression | 0.0078 | 0.0882 | 0.0393 | 0.0075 | 0.0868 | 0.0391 | 0.0079 | 0.0888 | 0.0393 |
| CART | 0.0073 | 0.0854 | 0.0349 | 0.0072 | 0.0851 | 0.0353 | 0.0071 | 0.0845 | 0.0344 |
| *Change* | -6.41% | -3.17% | **-11.19%** | 4.00% | 1.96% | -9.72% | **-10.13%** | **-4.84%** | **-12.47%** |
| ANN | 0.0072 | 0.0849 | 0.0355 | 0.0069 | 0.0831 | 0.0351 | 0.0074 | 0.0860 | 0.0362 |
| *Change* | **-7.69%** | **-3.74%** | -9.67% | **-8.00%** | **-4.26%** | -10.23% | -6.33% | -3.15% | -7.89% |
| DNN[2] | 0.0072 | 0.0849 | 0.0357 | 0.0069 | 0.0831 | 0.0360 | 0.0074 | 0.0860 | 0.0363 |
| *Change* | **-7.69%** | **-3.74%** | -9.16% | **-8.00%** | **-4.26%** | -7.93% | -6.33% | -3.15% | -7.63% |
| DNN[3] | 0.0072 | 0.0849 | 0.0353 | 0.0069 | 0.0831 | 0.0352 | 0.0074 | 0.0860 | 0.0366 |
| *Change* | **-7.69%** | **-3.74%** | -10.18% | **-8.00%** | **-4.26%** | -9.97% | -6.33% | -3.15% | -6.87% |
| DNN[4] | 0.0072 | 0.0849 | 0.0360 | 0.0070 | 0.837 | 0.0348 | 0.0074 | 0.0860 | 0.0359 |
| *Change* | **-7.69%** | **-3.74%** | -8.39% | -6.67% | -3.57% | **-10.99%** | -6.33% | -3.15% | -8.65% |

### Model 2: Predicting EBITAT

*Hyperparameter tuning number of neurons.* Similarly for models for predicting ROA, the Hyperband method is used to tune the number of neurons in each layer in the neural network models for predicting EBITAT. The following table shows the optimal number of neurons for the input and hidden layers with epochs=500.

Table 2.37: The optimal number of neurons of neural networks for predicting EBITAT

| | min_value | max_value | Step | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|---|---|---|
| Input layer | 27 | 27 | - | 27 | 27 | 27 | 27 |
| Hidden layer [1] | 8 | 512 | 4 | 276 | 128 | 476 | 424 |
| Hidden layer [2] | 8 | 512 | 4 | - | 504 | 448 | 464 |
| Hidden layer [3] | 8 | 512 | 4 | - | - | 472 | 252 |
| Hidden layer [4] | 8 | 512 | 4 | - | - | - | 44 |
| Output layer | 1 | 1 | - | 1 | 1 | 1 | 1 |

The architectures of the neural network models for EBITAT are shown in Table 2.38.

Table 2.38: The number of trainable parameters of neural networks for predicting EBITAT

| | Activation function | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|---|
| Input layer | ReLU | - | - | - | - |
| Hidden layer [1] | ReLU | 7728 | 3584 | 13328 | 11872 |
| Hidden layer [2] | ReLU | - | 65016 | 213696 | 197200 |
| Hidden layer [3] | ReLU | - | - | 211928 | 117180 |
| Hidden layer [4] | ReLU | - | - | - | 11132 |
| Output layer | linear | 277 | 505 | 473 | 45 |
| Total parameters | - | 8,005 | 69,105 | 439,425 | 337,429 |

*Hyperparameter tuning number of epochs.* For each model, the range from 1 to 500 epochs is run to obtain the optimal epoch. Holding other things constant, the "optimal" epoch for each model is the one that leads to the minimum value of the MSE of that model.

Table 2.39: Optimal epochs of neural networks for predicting EBITAT

| Model | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|
| Epochs range | [1, 500] | [1, 500] | [1, 500] | [1, 500] |
| Optimal epoch (80/20 split) | 276 | 169 | 157 | 132 |
| Optimal epoch (90/10 split) | 358 | 185 | 210 | 122 |
| Optimal epoch (50/50 split) | 310 | 188 | 172 | 149 |

*Visualization of MSE and MAE of training and validation data.* The variations of MSE and MAE of training and validation data with the 80/20 split ratio of the ANN, DNN[2], DNN[3], and DNN[4] model for predicting EBITAT are shown in Figure 2.14-2.17.

Figure 2.14: The MSE and MAE loss function of the ANN model for predicting EBITAT

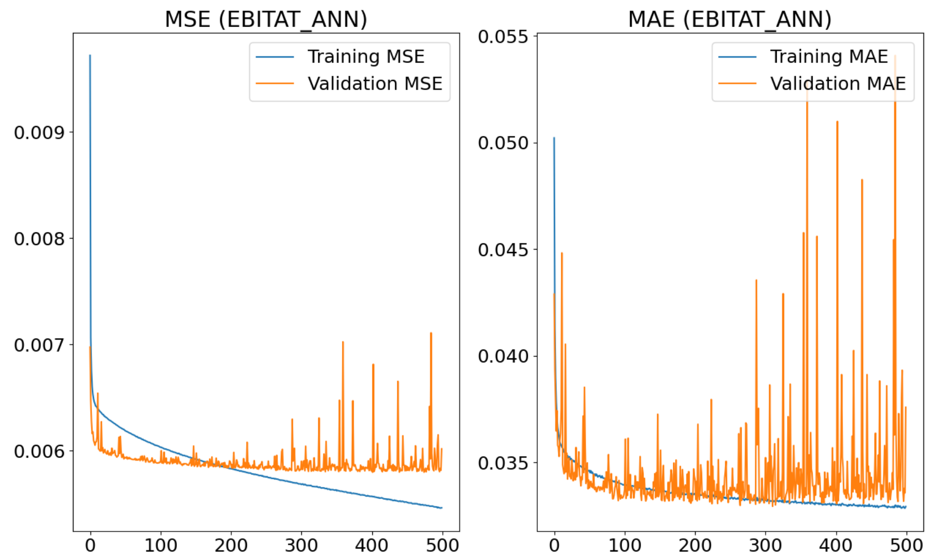



Figure 2.15: The MSE and MAE loss function of the DNN[2] model for predicting EBITAT

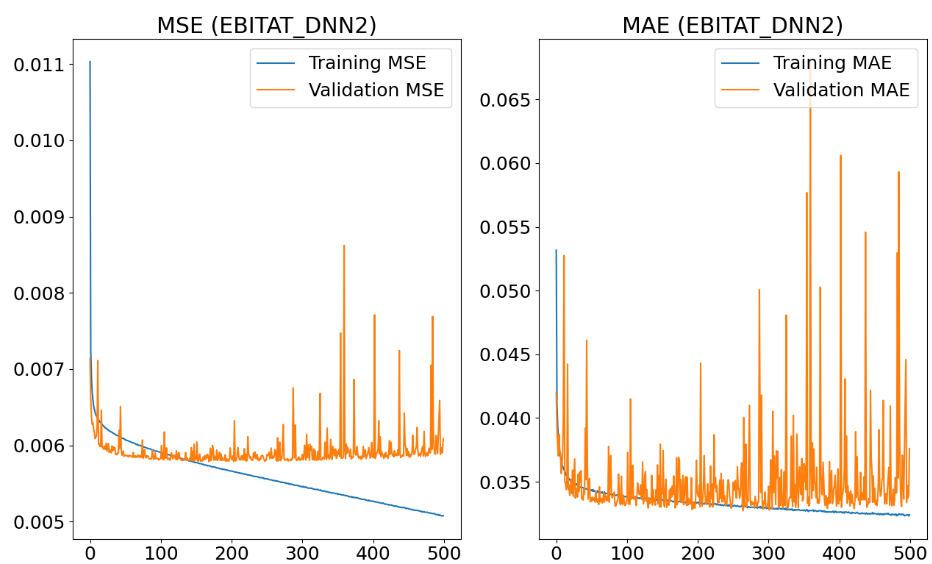


Figure 2.16: The MSE and MAE loss function of the DNN[3] model for predicting EBITAT

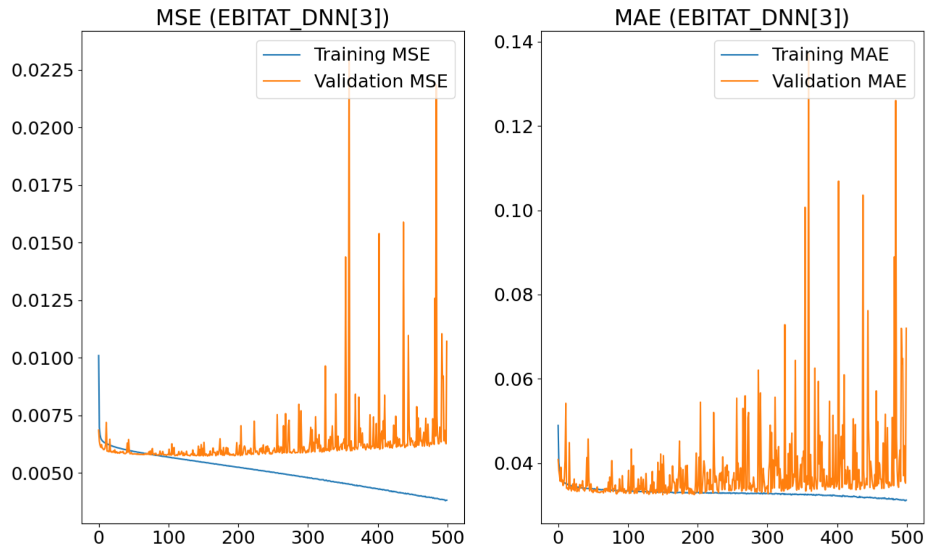
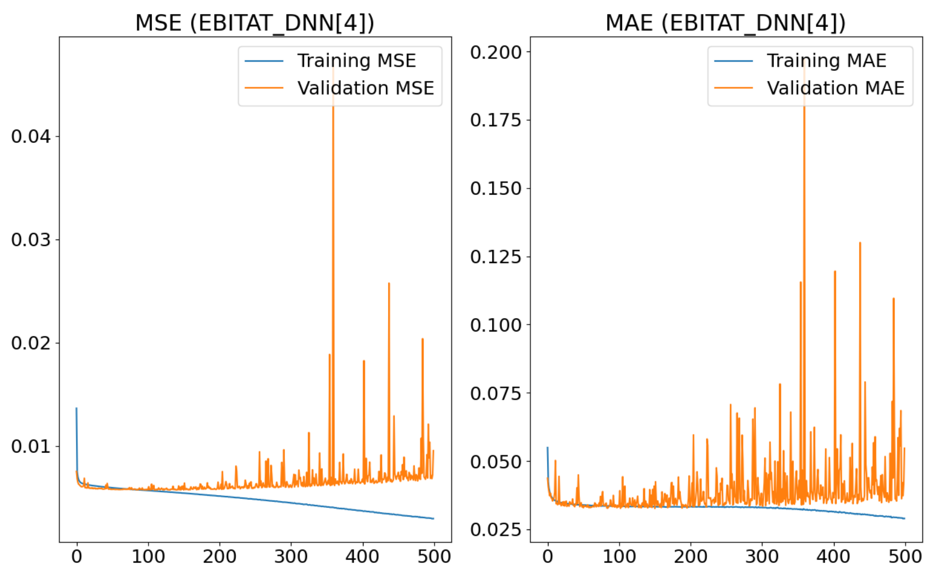



Figure 2.17: The MSE and MAE loss function of the DNN[4] model for predicting EBITAT




*Performance comparison.* Table 2.40 compares the performances on the test data among linear regression, CART, and neural networks, it is obvious that CART and neural networks are better than linear regression models in predicting EBITAT with any given split ratios and in terms of any given error measures. Moreover, in general, neural networks are better than CART if the split ratio is 80/20 or 90/10 meanwhile CART is better than neural networks if the split ratio is 50/50.

Table 2.40: Comparison of models performance on the test data for predicting EBITAT

| | 80/20 split | | | 90/10 split | | | 50/50 split | | |
|---|---|---|---|---|---|---|---|---|---|
| | MSE | RMSE | MAE | MSE | RMSE | MAE | MSE | RMSE | MAE |
| Regression | 0.0066 | 0.0813 | 0.0358 | 0.0063 | 0.0791 | 0.0353 | 0.0066 | 0.0813 | 0.0357 |
| CART | 0.0063 | 0.0792 | 0.0329 | 0.0060 | 0.0775 | 0.0329 | 0.0061 | 0.0778 | 0.0323 |
| *Change* | -4.55% | -2.58% | **-8.10%** | -4.76% | -2.02% | -6.79% | **-7.58%** | **-4.31%** | **-9.52%** |
| ANN | 0.0062 | 0.0787 | 0.0333 | 0.0057 | 0.0755 | 0.0322 | 0.0062 | 0.0787 | 0.0337 |
| *Change* | -6.06% | -3.19% | -6.98% | **-9.52%** | **-4.55%** | **-8.78%** | -6.06% | -3.19% | -5.60% |
| DNN [2] | 0.0062 | 0.0787 | 0.0330 | 0.0057 | 0.0755 | 0.0325 | 0.0062 | 0.0787 | 0.0338 |
| *Change* | -6.06% | -3.19% | -7.82% | **-9.52%** | **-4.55%** | -7.93% | -6.06% | -3.19% | -5.32% |
| DNN [3] | 0.0061 | 0.0781 | 0.0330 | 0.0058 | 0.0761 | 0.0324 | 0.0063 | 0.0794 | 0.0338 |
| *Change* | **-7.58%** | **-3.94%** | -7.82% | -7.94% | -3.79% | -8.21% | -4.55% | -2.34% | -5.32% |
| DNN [4] | 0.0061 | 0.0781 | 0.0333 | 0.0057 | 0.0755 | 0.0324 | 0.0062 | 0.0787 | 0.0336 |
| *Change* | **-7.58%** | **-3.94%** | -6.98% | **-9.52%** | **-4.55%** | -8.21% | -6.06% | -3.19% | -5.88% |

### Model 3: Predicting Tobin's Q

*Hyperparameter tuning number of neurons.* Similarly to Model 1 and Model 2, the Hyperband method is used for tuning the number of neurons in Model 3. Table 2.43 shows the optimal number of neurons for the input and hidden layers of different neural networks with epochs=500.

Table 2.41: The optimal number of neurons of neural networks for predicting Tobin's Q

| | min_value | max_value | Step | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|---|---|---|
| Input layer | 27 | 27 | - | 27 | 27 | 27 | 27 |
| Hidden layer [1] | 8 | 512 | 4 | 432 | 312 | 240 | 216 |
| Hidden layer [2] | 8 | 512 | 4 | - | 480 | 424 | 428 |
| Hidden layer [3] | 8 | 512 | 4 | - | - | 416 | 416 |
| Hidden layer [4] | 8 | 512 | 4 | - | - | - | 120 |
| Output layer | 1 | 1 | - | 1 | 1 | 1 | 1 |

The architectures of the neural network models for Tobin's Q are shown in the following table.

Table 2.42: The number of parameters of neural networks for predicting Tobin's Q

| | Activation function | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|---|
| Input layer | ReLU | - | - | - | - |
| Hidden layer [1] | ReLU | 12096 | 8736 | 6720 | 6048 |
| Hidden layer [2] | ReLU | - | 150240 | 102184 | 92876 |
| Hidden layer [3] | ReLU | - | - | 176800 | 178464 |
| Hidden layer [4] | ReLU | - | - | - | 50040 |
| Output layer | linear | 433 | 481 | 417 | 121 |
| Total parameters | - | 12,529 | 159,457 | 286,121 | 327,549 |

*Hyperparameter tuning number of epochs.* The range from 1 to 500 epochs is run to obtain the optimal epoch for each model.

Table 2.43: Optimal epochs of neural networks for predicting Tobin's Q

| Model | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|
| Epochs range | [1, 500] | [1, 500] | [1, 500] | [1, 500] |
| Optimal epochs (80/20 split) | 294 | 256 | 139 | 143 |
| Optimal epochs (90/10 split) | 323 | 216 | 149 | 141 |
| Optimal epochs (80/20 split) | 291 | 253 | 148 | 148 |

*Visualization of MSE and MAE of training and validation data.* The variations of MSE and MAE of training and validation data with the 80/20 split ratio of the ANN, DNN[2], DNN[3], and DNN[4] model for predicting Tobin's Q are shown in Figure 2.18-2.21.

Figure 2.18: The MSE and MAE loss function of the ANN model for predicting Tobin's Q

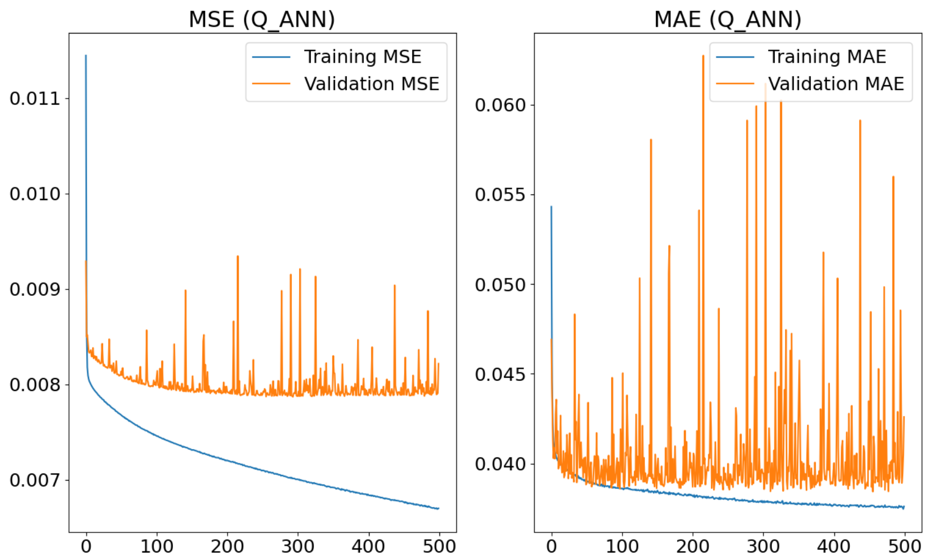



Figure 2.19: The MSE and MAE loss function of the DNN[2] for predicting Tobin's Q

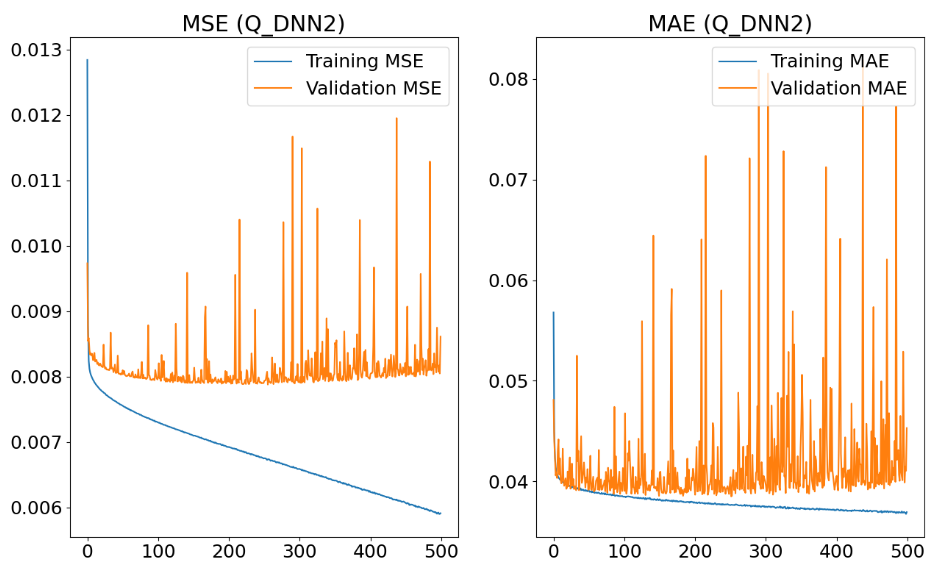


Figure 2.20: The MSE and MAE loss function of the DNN[3] for predicting Tobin's Q

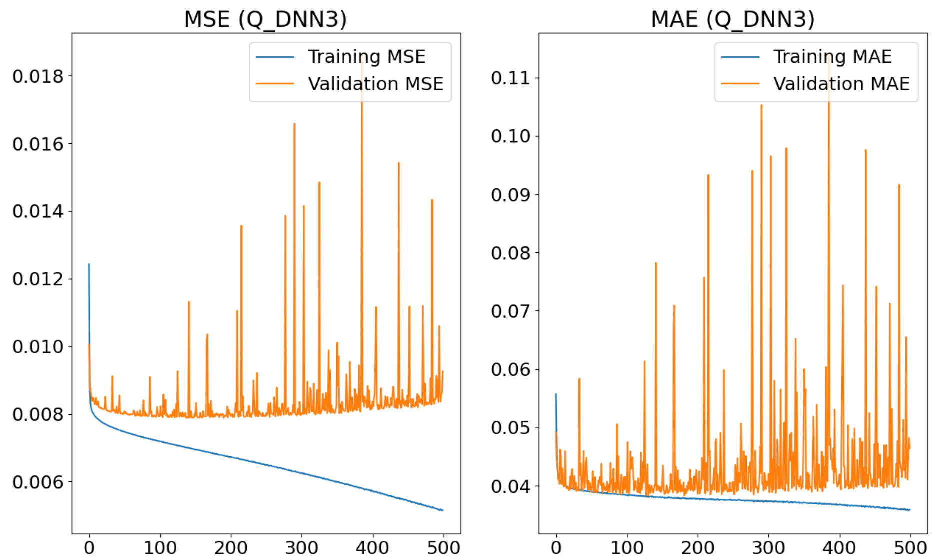
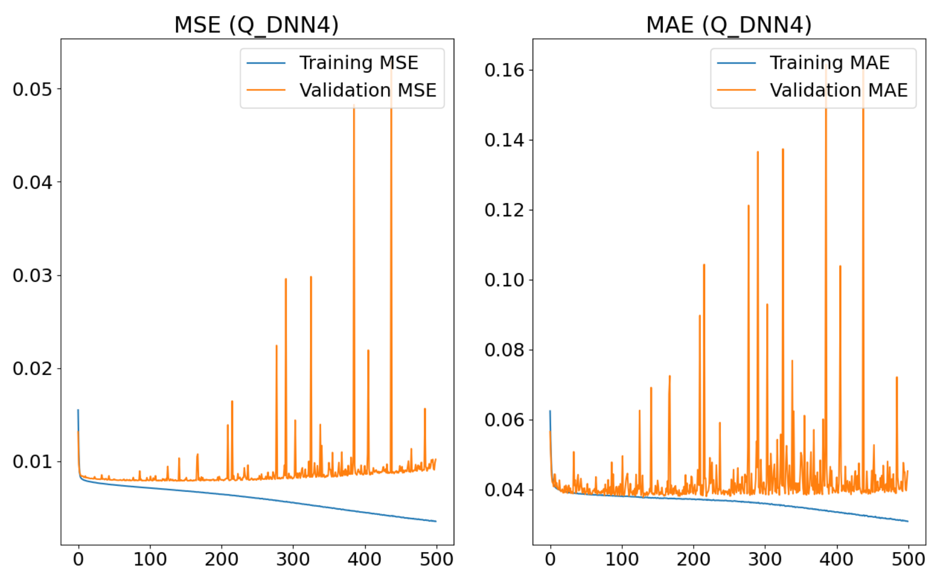



Figure 2.21: The MSE and MAE loss function of the DNN[4] for predicting Tobin's Q




*Performance comparison.* Table 2.44 compares the performances of the linear regression, CART, and neural network models on the test data in terms of MSE, RMSE, and MAE with the 80/20, 90/10, and 50/50 split ratios. Overall, CART and neural networks outperform linear regression models in predicting Tobin's Q with any given split ratios and in terms of any given error measures. Comparison between neural networks and CART, in general, neural networks are better predictive models when the split ratio is 80/20 and 90/10. Meanwhile, CART models are better than neural networks when the split ratio is 50/50.

Table 2.44: Comparison of models performance on the test data for predicting Tobin's Q

| | 80/20 split | | | 90/10 split | | | 50/50 split | | |
|---|---|---|---|---|---|---|---|---|---|
| | MSE | RMSE | MAE | MSE | RMSE | MAE | MSE | RMSE | MAE |
| Regression | 0.0081 | 0.0898 | 0.0399 | 0.0078 | 0.0883 | 0.0397 | 0.0081 | 0.0901 | 0.0399 |
| CART | 0.0076 | 0.0869 | 0.0374 | 0.0076 | 0.0875 | 0.0376 | 0.0074 | 0.0859 | 0.0369 |
| *Change* | -6.17% | -3.23% | **-6.27%** | -2.56% | -0.91% | -5.29% | **-8.64%** | **-4.66%** | **-7.52%** |
| ANN | 0.0076 | 0.0869 | 0.0382 | 0.0072 | 0.0849 | 0.0376 | 0.0078 | 0.0883 | 0.0393 |
| *Change* | -6.17% | -3.23% | -4.26% | -7.69% | -3.85% | -5.29% | -3.70% | -1.99% | -1.50% |
| DNN [2] | 0.0076 | 0.0869 | 0.0382 | 0.0072 | 0.0849 | 0.0373 | 0.0078 | 0.0883 | 0.0387 |
| *Change* | -6.17% | -3.23% | -4.26% | -7.69% | -3.85% | **-6.05%** | -3.70% | -1.99% | -3.01% |
| DNN [3] | 0.0075 | 0.0866 | 0.0384 | 0.0071 | 0.0843 | 0.0375 | 0.0078 | 0.0883 | 0.0391 |
| *Change* | **-7.41%** | **-3.56%** | -3.76% | **-8.97%** | **-4.53%** | -5.54% | -3.70% | -1.99% | -2.01% |
| DNN [4] | 0.0075 | 0.0866 | 0.0375 | 0.0072 | 0.0849 | 0.0384 | 0.0078 | 0.0883 | 0.0389 |
| *Change* | **-7.41%** | **-3.56%** | -6.02% | -7.69% | -3.85% | -3.27% | -3.70% | -1.99% | -2.51% |

### Model 4: Predicting sales growth

*Hyperparameter tuning number of neurons.* Similarly to previous models, the Hyperband method is used for tuning the number of neurons in Model 4. The following table shows the optimal number of neurons for the input and hidden layers of several neural network models with epochs=500.

Table 2.45: The optimal number of neurons of neural networks for predicting sales growth

| | min_value | max_value | Step | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|---|---|---|
| Input layer | 27 | 27 | - | 27 | 27 | 27 | 27 |
| Hidden layer [1] | 8 | 512 | 4 | 404 | 256 | 328 | 496 |
| Hidden layer [2] | 8 | 512 | 4 | - | 376 | 212 | 196 |
| Hidden layer [3] | 8 | 512 | 4 | - | - | 440 | 332 |
| Hidden layer [4] | 8 | 512 | 4 | - | - | - | 412 |
| Output layer | 1 | 1 | - | 1 | 1 | 1 | 1 |

The architectures of the neural network models for sales growth are shown in Table 2.46.

Table 2.46: The number of parameters of neural networks for predicting sales growth

| | Activation function | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|---|
| Input layer | ReLU | - | - | - | - |
| Hidden layer [1] | ReLU | 11312 | 7168 | 9184 | 13888 |
| Hidden layer [2] | ReLU | - | 96632 | 69748 | 97412 |
| Hidden layer [3] | ReLU | - | - | 93720 | 65404 |
| Hidden layer [4] | ReLU | - | - | - | 137196 |
| Output layer | linear | 405 | 377 | 441 | 413 |
| Total parameters | - | 11,717 | 104,177 | 173,093 | 314,313 |

*Hyperparameter tuning number of epochs.* The range from 1 to 500 epochs is run to obtain the optimal epoch (i.e. smallest MSE) for each model.

Table 2.47: Optimal epochs of neural networks for predicting sales growth

| Model | ANN | DNN[2] | DNN[3] | DNN[4] |
|---|---|---|---|---|
| Epochs range | [1, 500] | [1, 500] | [1, 500] | [1, 500] |
| Optimal epochs (80/20 split) | 458 | 288 | 166 | 128 |
| Optimal epochs (90/10 split) | 479 | 207 | 207 | 204 |
| Optimal epochs (50/50 split) | 199 | 149 | 165 | 82 |

*Visualization of MSE and MAE of training and validation data.* The MSE and MAE loss functions of training and validation data with the 80/20 split ratio of the neural networks for predicting sales growth are shown in Figure 2.22-2.25.

Figure 2.22: The MSE and MAE loss function of the ANN model for predicting sales growth

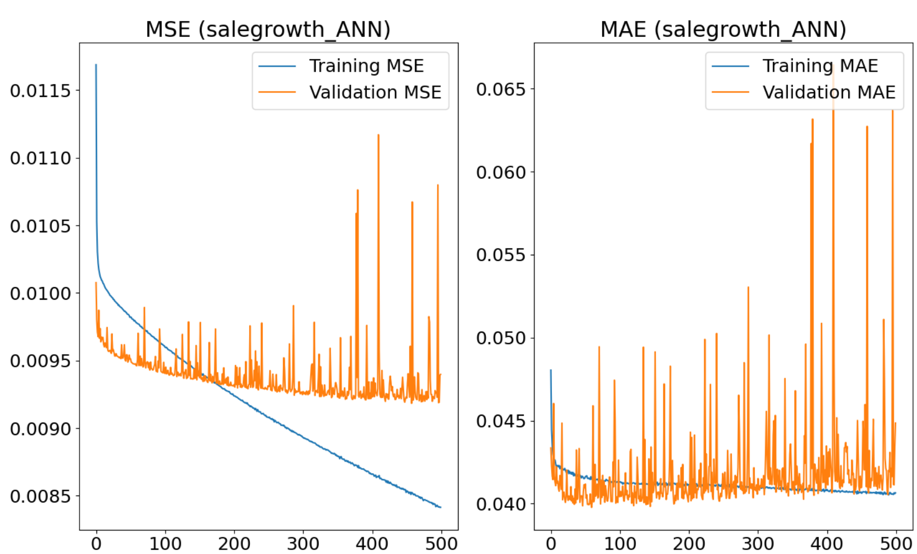


Figure 2.23: The MSE and MAE loss function of the DNN[2] for predicting sales growth

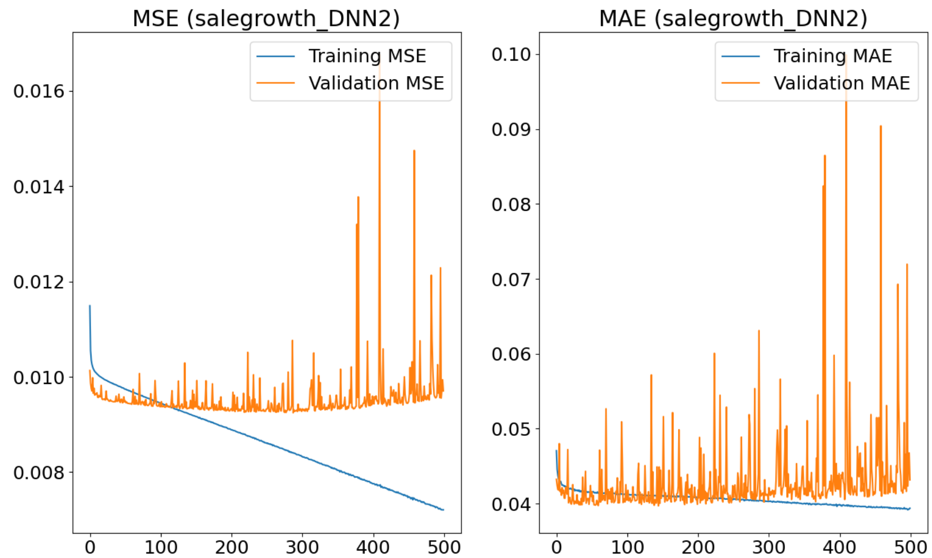



Figure 2.24: The MSE and MAE loss function of the DNN[3] for predicting sales growth

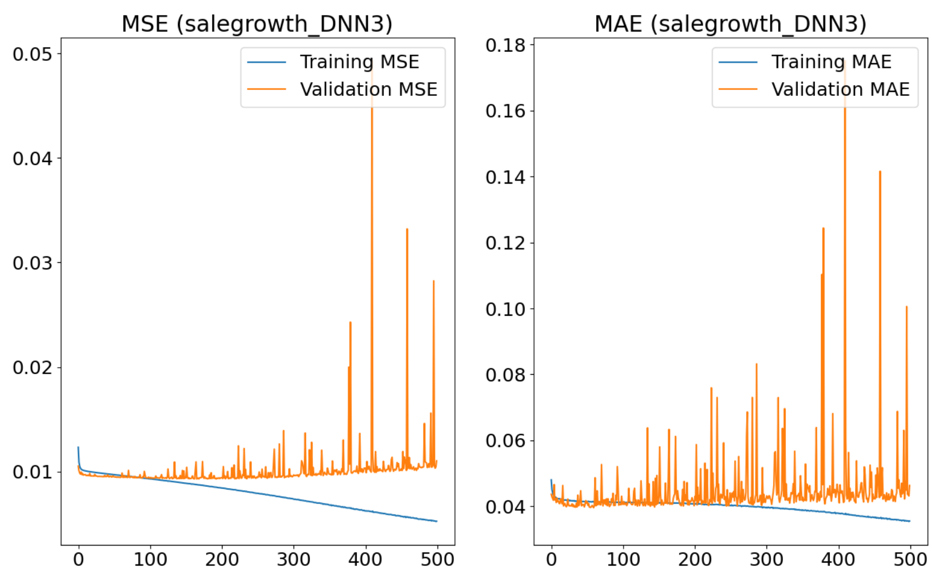
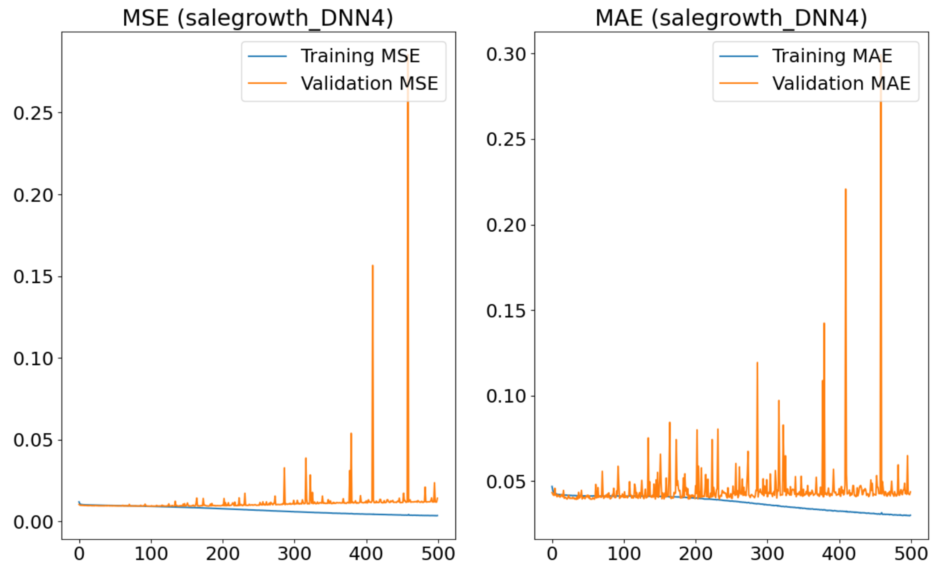


Figure 2.25: The MSE and MAE loss function of the DNN[4] for predicting sales growth




*Performance comparison.* The MSE, RMSE, MAE of linear regression, neural networks, and CART models for predicting sales growth are shown in Table 2.48. It is obvious that CART and neural networks outperform linear regression models in predicting sales growth with any given split ratios and in terms of any given error measures. Comparison between CART and neural networks, it seems that CART have the best predictive models for 80/20 and 50/50 split ratios, and for 90/10 split ratio in terms of MAE. Meanwhile, neural networks are only better than CART for 90/10 split ratio in terms of MSE or RMSE. In general, CART seem to be the best predictive models in predicting sales growth in comparison to linear regression and neural networks.

Table 2.48: Comparison of models performance on the test data for predicting sales growth

| | 80/20 split | | | 90/10 split | | | 50/50 split | | |
|---|---|---|---|---|---|---|---|---|---|
| | MSE | RMSE | MAE | MSE | RMSE | MAE | MSE | RMSE | MAE |
| Regression | 0.0103 | 0.1016 | 0.0420 | 0.0104 | 0.1019 | 0.0420 | 0.0102 | 0.1008 | 0.0421 |
| CART | 0.0098 | 0.0991 | 0.0407 | 0.0102 | 0.1010 | 0.0417 | 0.0094 | 0.0967 | 0.0400 |
| *Change* | **-4.85%** | **-2.46%** | **-3.09%** | **-1.92%** | **-0.88%** | -0.71% | **-7.84%** | **-4.07%** | **-4.99%** |
| ANN | 0.0100 | 0.1000 | 0.0414 | 0.0103 | 0.1015 | 0.0417 | 0.0099 | 0.0995 | 0.0409 |
| *Change* | -2.91% | -1.57% | -1.43% | -0.96% | -0.39% | -0.71% | -2.94% | -1.29% | -2.85% |
| DNN[2] | 0.0100 | 0.1000 | 0.0417 | 0.0103 | 0.1015 | 0.0415 | 0.0099 | 0.0995 | 0.0412 |
| *Change* | -2.91% | -1.57% | -0.71% | -0.96% | -0.39% | -1.19% | -2.94% | -1.29% | -2.14% |
| DNN[3] | 0.0100 | 0.1000 | 0.0412 | 0.0103 | 0.1015 | 0.0416 | 0.0100 | 0.1000 | 0.0416 |
| *Change* | -2.91% | -1.57% | -1.90% | -0.96% | -0.39% | -0.95% | -1.96% | -0.79% | -1.19% |
| DNN[4] | 0.0100 | 0.1000 | 0.0412 | 0.0104 | 0.1019 | 0.0407 | 0.0100 | 0.1000 | 0.0410 |
| *Change* | -2.91% | -1.57% | -1.90% | 0.00% | 0.00% | **-3.09%** | -1.96% | -0.79% | -2.61% |

In summary, comparing Table 2.36, Table 2.40, Table 2.44, and Table 2.48 yields three noticeable results. First, it is obvious and robust that neural networks and CART models outperform linear regression models in predicting any given firm performance criteria with any given split ratios and in terms of any given error measures. This finding is consistent with many previous studies (Hill et al. 1996; Desai and Bharati 1998; Fadlalla and Lin 2001; Weatherford et al. 2003; Razi and Athappilly 2005; Anyaeche and Ighravwe 2013; Pombeiro et al. 2017). Second, in general, neural networks are better than CART when the split ratio is 80/20 or 90/10 in predicting ROA, EBITAT, and Tobin's Q while CART are better than neural networks in predicting ROA, EBITAT, and Tobin's Q when the split ratio is 50/50. Third, for predicting sales growth, it seems that CART are better than neural networks when the split ratio is 80/20 or 50/50. Meanwhile neural networks seem to be better predictive models when the split ratio is 90/10.

## 2.5 Discussion

### 2.5.1 The effects of natural disasters on firm performance revisited

In the introduction section, we know that, at the firm level, the effect of natural disasters on firm performance is mixed or insignificant but these findings might be misleading. One of the reasons is that many previous studies only examine one particular type of disaster or one specific area or in a short period of time; therefore the findings might not be general enough. For example, Noth and Rehbein (2019) only study a major flood in Germany in only one year in 2013. Therefore, their results might not be general enough for the whole picture of general natural disasters. Similarly, Zhou and Botzen (2021) and Leiter et al. (2009) only focus on flooding, but not other types of disasters. Our study, however, considers disasters in its general sense which include all main types of natural disasters such as drought, flooding, freeze, severe storm, tropical cyclone, wildfire, and winter storm.

It is necessary to address that even though the relationships between disasters and firm performance are in the sense of correlations, they indeed have a cause and effect relationship. It is impossible that firms' ROA causes natural disasters since natural disasters are exogenous shocks. Firm profitability cannot cause disasters. Hence, the conclusion is that both actual and reporting disasters negatively impact the firm's ROA in the next year.

### 2.5.2 Objective versus perceived natural environmental uncertainty revisited

The results in this study shows the importance of both *numerical* and *textual* data in accessing the effects of the natural environmental uncertainty on firm performance. Most of previous studies only examine either numerical or textual data in predicting firm performance. The implication is that we should pay more attention on textual data, specifically Form 10-Ks and their roles in firm performance analysis as shown, for example, in Feldman et al. (2010), Loughran and McDonald (2011), Bochkay and Levin (2019), Choi et al. (2020), Cohen et al. (2020), and Cho and Muslu (2021).

More importantly, this is probably the first study investigating the *jointly* effects of the actual *and* perceived natural environmental uncertainty. This implies there is a big gap in the literature since, to the best of my knowledge, there are no previous studies in the literature that consider these two natural environmental uncertainties at the same time. This study partially helps to fill this gap. Our results actually indicate that both of the natural environmental uncertainties simultaneously impact on firm performance.

### 2.5.3 CART and neural networks vs. linear regression

Generally, in most models, CART and neural network models outperform linear regression in predicting any given firm performance criteria. This finding is consistent with the current literature on comparison between neural networks and linear regression (e.g. (Hill et al. 1996; Fadlalla and Lin 2001; Doganis et al. 2006; Anyaeche and Ighravwe 2013; Pombeiro et al. 2017)). It also shows that neural networks are suitable tools for forecasting in finance and economics as discussed by Li and Ma (2010).

### 2.5.4 The importance of the holistic approach

One of the essential contributions of this study is employing the holistic approach for firm performance modeling under actual and perceived environmental uncertainty. In this study, we examine four different criteria of firm performance including ROA, EBITAT, Tobin's Q, and sales growth. The first two criteria (ROA and EBITAT) represent the *profitability* of a firm. Tobin's Q represents the *investment opportunity* of a firm. Meanwhile, sales growth reflects the *operational efficiency* of a firm. These all criteria are essential to evaluate how well a firm performs. If only one in these four criteria is used to evaluate the performance of a firm, the comparison among models might be biased or not general enough.

Another aspect of the holistic approach in this research is that employing different models, including linear regression, CART, and neural networks, allows us to choose the suitable model(s) for each firm performance criterion. Particularly, as shown in the results section, sales growth performs best with the CART model while Tobin's Q performs best with NNs. Moreover, ROA and EBITAT are predicted best by either neural networks or CART depending on the error measure metrics. Using multiple models together with multiple criteria in evaluating firm performance prediction provides the robust and unbiased (or probably less biased) results.

## 2.6 Conclusion

This study proposes a new dictionary of words related to natural hazards and disasters. Thus, the text mining technique is used to count the number of words in Form 10-Ks which also appear in the dictionary. The constructed indicator is called the perceived risk of natural disasters. This textual measure of perceived risk of natural disasters is combined with the government-reported damages of natural hazards and disasters, and a number of control variables to predict the effects of natural hazards and disasters on the U.S firm performance in the 1993-2021 period.

We find that this self-reported perceived risk of natural hazards and disasters and government-reported of the damages of natural hazards in the current year are negatively associated with firm profitability next year. However, the perceived risk indicator is not associated with sales growth and Tobin's Q ratio. In particular, the perceived risk of natural hazards and disasters negatively affect firm profitability in the services sector but not in the manufacturing sector. The firm profitability in the services sector is also negatively affected by the billion-dollar natural disasters in the same year but not in the previous year. This implies that there is a lag effect of perceived risk of natural disasters on firm profitability in the services sector while there is no lag effect of the government-reported damages of the billion-dollar natural disasters on firm profitability in this sector. Finally, we find that CART and neural networks robustly outperform linear regression in predicting firm performance under natural disaster risks. The main suggestion from this study is that we can use textual data in financial reports (e.g Form 10-K filings) to measure the perceived risk of natural disasters and predict its effects on general firm performance and firm profitability specifically.

As would be expected, this study contains several limitations that pave several ways for potential future works. Future works might use the factory-level data set rather than the firm-level one to capture the situation that a firm has the headquarter in one location but have factories in several locations. One also explores the causal relationship between natural hazards and disasters and firm performance by using several techniques like instrumental variables or causal machine learning.

# Appendices

## Appendix 1.1 An illustration of Form 10-K (Alphabet Inc. and Google Inc.)

**UNITED STATES**
**SECURITIES AND EXCHANGE COMMISSION**
**Washington, D.C. 20549**

**FORM 10-K**

**ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934**

For the fiscal year ended December 31, 2021

**OR**

**TRANSITION REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934**

For the transition period from ______ to ______.

Commission file number: **001-37580**

**Alphabet Inc.**

**(Exact name of registrant as specified in its charter)**

| | |
|---|---|
| **Delaware** | **61-1767919** |
| (State or other jurisdiction of incorporation or organization) | (I.R.S. Employer Identification No.) |

**1600 Amphitheatre Parkway**
**Mountain View, CA 94043**
(Address of principal executive offices, including zip code)
**(650) 253-0000**
(Registrant's telephone number, including area code)

**Securities registered pursuant to Section 12(b) of the Act:**

*Source*: SEC

## Appendix 1.2 An illustration of the content in a Form 10-K.

Figure 2.26: The content of a Form 10-K

**Alphabet Inc.**
**Form 10-K**
**For the Fiscal Year Ended December 31, 2021**

**TABLE OF CONTENTS**

Note About Forward-Looking Statements

**PART I**

| | |
|---|---|
| Item 1. | Business |
| Item 1A. | Risk Factors |
| Item 1B. | Unresolved Staff Comments |
| Item 2. | Properties |
| Item 3. | Legal Proceedings |
| Item 4. | Mine Safety Disclosures |

**PART II**

| | |
|---|---|
| Item 5. | Market for Registrant's Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities |
| Item 6. | [Reserved] |
| Item 7. | Management's Discussion and Analysis of Financial Condition and Results of Operations |
| Item 7A. | Quantitative and Qualitative Disclosures About Market Risk |
| Item 8. | Financial Statements and Supplementary Data |
| Item 9. | Changes in and Disagreements With Accountants on Accounting and Financial Disclosure |
| Item 9A. | Controls and Procedures |
| Item 9B. | Other Information |
| Item 9C. | Disclosure Regarding Foreign Jurisdictions that Prevent Inspections |

**PART III**

| | |
|---|---|
| Item 10. | Directors, Executive Officers and Corporate Governance |
| Item 11. | Executive Compensation |
| Item 12. | Security Ownership of Certain Beneficial Owners and Management and Related Stockholder Matters |
| Item 13. | Certain Relationships and Related Transactions, and Director Independence |
| Item 14. | Principal Accountant Fees and Services |

**PART IV**

| | |
|---|---|
| Item 15. | Exhibits, Financial Statement Schedules |
| Item 16. | Form 10-K Summary |

*Source*: SEC

## Appendix 2. Two-digit SIC codes[31]

Table 2.49: Two-digit SICH codes and corresponding industries

| Two-digit SIC code | Industry |
|---|---|
| 01 | Agricultural Production - Crops |
| 02 | Agricultural Production - Livestock |
| 07 | Agricultural Services |
| 08 | Forestry |
| 09 | Fishing, Hunting and Trapping |
| 10 | Metal Mining |
| 12 | Bituminous Coal and Lignite Mining |
| 13 | Oil and Gas Extraction |
| 14 | Mining and Quarrying of Nonmetallic Minerals, except Fuels |
| 15 | Building Construction General Contractors and Operative Builders |
| 16 | Heavy Construction other than Building Construction Contractors |
| 17 | Construction Special Trade Contractors |
| 20 | Food and Kindred Products |
| 21 | Tobacco Products |
| 22 | Textile Mill Products |
| 23 | Apparel and other Finished Products Made from Fabrics and Similar Materials |
| 24 | Lumber and Wood Products, except Furniture |
| 25 | Furniture and Fixtures |
| 26 | Paper and Allied Products |
| 27 | Printing, Publishing, and Allied Industries |
| 28 | Chemicals and Allied Products |
| 29 | Petroleum Refining and Related Industries |
| 30 | Rubber and Miscellaneous Plastics Products |
| 31 | Leather and Leather Products |
| 32 | Stone, Clay, Glass, and Concrete Products |
| 33 | Primary Metal Industries |
| 34 | Fabricated Metal Products, except Machinery and Transportation Equipment |
| 35 | Industrial and Commercial Machinery and Computer Equipment |
| 36 | Electronic and other Electrical Equipment and Components, except Computer Equipment |
| 37 | Transportation Equipment |
| 38 | Measuring, Analyzing, and Controlling Instruments; Photographic, Medical and Optical Goods; Watches and Clocks |
| 39 | Miscellaneous Manufacturing Industries |
| 40 | Railroad Transportation |
| 41 | Local and Suburban Transit and Interurban Highway Passenger Transportation |
| 42 | Motor Freight Transportation and Warehousing |

[31] See https://siccode.com/

## Appendix 2. Two-digit SIC codes (cont.)

Table 2.50: Two-digit SICH codes and corresponding industries (cont.)

| Two-digit SIC code | Industry |
|---|---|
| 43 | United States Postal Service |
| 44 | Water Transportation |
| 45 | Transportation by Air |
| 46 | Pipelines, except Natural Gas |
| 47 | Transportation Services |
| 48 | Communications |
| 49 | Electric, Gas and Sanitary Services |
| 50 | Wholesale Trade-Durable Goods |
| 51 | Wholesale Trade-Nondurable Goods |
| 52 | Building Materials, Hardware, Garden Supply, and Mobile Home Dealers |
| 53 | General Merchandise Stores |
| 54 | Food Stores |
| 55 | Automotive Dealers and Gasoline Service Stations |
| 56 | Apparel and Accessory Stores |
| 57 | Home Furniture, Furnishings, and Equipment Stores |
| 58 | Eating and Drinking Places |
| 59 | Miscellaneous Retail |
| 70 | Hotels, Rooming Houses, Camps, and other Lodging Places |
| 72 | Personal Services |
| 73 | Business Services |
| 75 | Automotive Repair, Services, and Parking |
| 76 | Miscellaneous Repair Services |
| 78 | Motion Pictures |
| 79 | Amusement and Recreation Services |
| 80 | Health Services |
| 81 | Legal Services |
| 82 | Educational Services |
| 83 | Social Services |
| 84 | Museums, Art Galleries, and Botanical and Zoological Gardens |
| 86 | Membership Organizations |
| 87 | Engineering, Accounting, Research, Management, and Related Services |
| 88 | Private Households |
| 89 | Miscellaneous Services |

*Notes*: These SICH codes exclude the Finance, Insurance, Real Estate sector (two-digit SICH code from 60 to 67) and Public Administration sector (two-digit SICH code from 91 to 99). The reason is that companies in the Finance, Insurance, Real Estate sector usually have dissimilar patterns of capital structures compared to companies from other sectors. For the companies in the Public Administration sector, their goals might be not profit or not only profit (profit might be less important than other goals for example social welfare).

## Appendix 3. Histograms of independent variables of Model 2-4

Figure 2.27: Heat map of the original features of Model 2 (EBITAT)

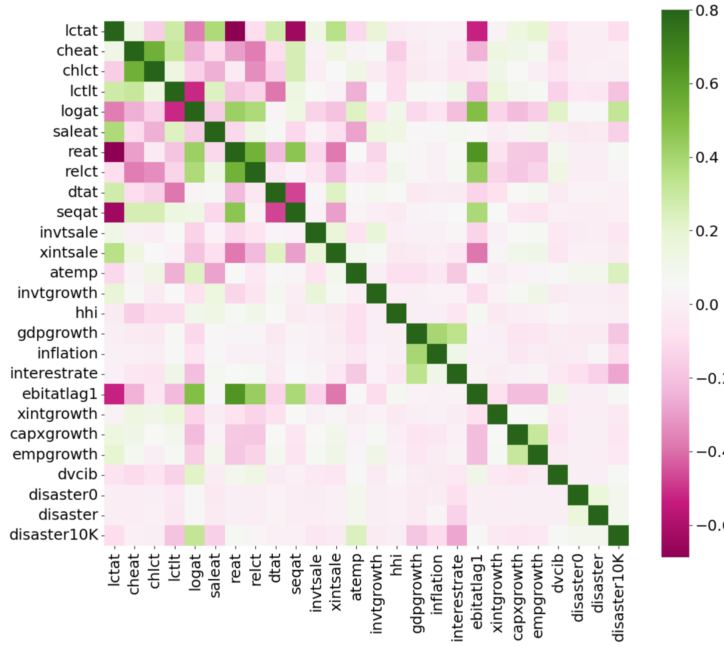



Figure 2.28: Heat map of the original features of Model 3 (Tobin's Q)

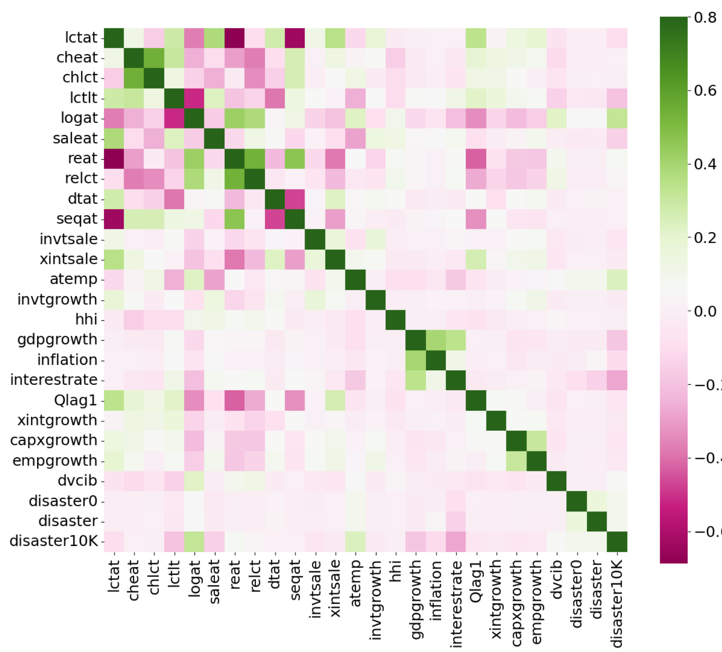
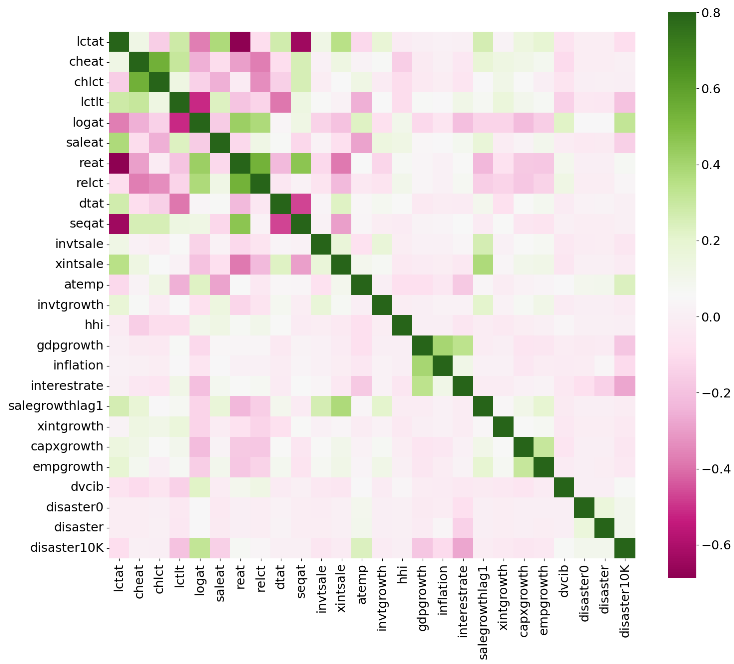


Figure 2.29: Heat map of the original features of Model 4 (sales growth)




## Appendix 4: Top 20 VIFs for the case of damages of natural hazards.

Table 2.51: Top 20 VIF's values for linear regression models for the case of natural hazards

| Model 5 | VIF | Model 6 | VIF | Model 7 | VIF | Model 8 | VIF |
|---|---|---|---|---|---|---|---|
| lctlt | 6.881487 | lctlt | 6.890518 | lctlt | 7.278854 | lctlt | 6.910421 |
| inflation | 6.691684 | inflation | 6.688276 | inflation | 6.682604 | inflation | 6.670935 |
| logat | 6.391205 | logat | 6.488870 | logat | 6.287555 | logat | 6.320445 |
| lctat | 6.205631 | lctat | 6.203778 | lctat | 6.097546 | lctat | 6.164566 |
| seqat | 4.287896 | seqat | 4.283971 | seqat | 4.410722 | seqat | 4.305235 |
| interestrate | 4.169784 | interestrate | 4.166773 | interestrate | 4.165903 | interestrate | 4.170176 |
| reat | 3.866809 | reat | 3.895254 | reat | 3.823880 | reat | 3.755915 |
| gdpgrowth | 3.699480 | gdpgrowth | 3.698220 | gdpgrowth | 3.699915 | gdpgrowth | 3.698156 |
| saleat | 3.579374 | saleat | 3.636816 | saleat | 3.548044 | saleat | 3.594935 |
| cheat | 3.229502 | cheat | 3.239009 | cheat | 3.231313 | cheat | 3.238179 |
| hhi | 2.683497 | hhi | 2.685514 | hhi | 2.682847 | hhi | 2.682481 |
| dtat | 2.319157 | dtat | 2.318606 | dtat | 2.320167 | dtat | 2.319864 |
| relct | 2.268091 | relct | 2.282542 | relct | 2.232440 | relct | 2.230530 |
| chlct | 2.176366 | ebitatlag1 | 2.239430 | chlct | 2.186312 | chlct | 2.168748 |
| roalag1 | 2.133635 | chlct | 2.177203 | Qlag1 | 2.084342 | disaster10K | 1.695737 |
| disaster10K | 1.699093 | disaster10K | 1.701731 | disaster10K | 1.697062 | invtsale | 1.609227 |
| invtsale | 1.566088 | invtsale | 1.569528 | invtsale | 1.563703 | atemp | 1.551872 |
| atemp | 1.549942 | atemp | 1.549855 | atemp | 1.550098 | xintsale | 1.543207 |
| xintsale | 1.416533 | xintsale | 1.410274 | xintsale | 1.399127 | salegrowthlag1 | 1.397037 |
| capxgrowth | 1.227708 | capxgrowth | 1.228246 | capxgrowth | 1.224133 | capxgrowth | 1.224133 |

## Appendix 5: The frequency of natural disaster words/phrases

| Word/Phrase | Word count | Word/Phrase | Word count | Word/Phrase | Word count |
|---|---|---|---|---|---|
| disasters | 126,207 | extreme temperatures | 1228 | rainfalls | 96 |
| wind | 100,011 | landfall | 1,193 | lava flow | 86 |
| disaster | 76,652 | high winds | 1,105 | avalanches | 80 |
| climate change | 61,559 | thunderstorms | 1,102 | gale | 78 |
| hurricanes | 56,363 | blizzards | 1,053 | debacle | 64 |
| flood | 55,596 | tremor | 1,008 | rainstorm | 57 |
| earthquake | 47,391 | cyclone | 988 | disaster risk | 55 |
| earthquakes | 46,352 | typhoon | 946 | extreme rain | 47 |
| storm | 45,501 | whirlpool | 922 | gales | 47 |
| floods | 37,389 | natural hazards | 878 | landfalls | 47 |
| storms | 28,197 | fog | 761 | firestorms | 46 |
| hurricane | 24,951 | mudslides | 761 | hazard mitigation | 46 |
| freeze | 20,829 | forest fires | 640 | rock fall | 32 |
| flooding | 18,280 | cyclones | 616 | twister | 31 |
| snow | 17,748 | snowstorms | 577 | microbursts | 26 |
| drought | 13,426 | extreme heat | 548 | La Nina | 25 |
| tornadoes | 11,609 | heavy rainfall | 544 | ash fall | 21 |
| extreme weather | 11,159 | heavy snow | 523 | cold waves | 18 |
| freezing | 8,739 | volcano | 457 | extreme rainfall | 16 |
| volcanic | 7,724 | avalanche | 405 | fogs | 16 |
| windstorm | 7,046 | derechos | 401 | winterstorm | 12 |
| rainfall | 6,835 | extreme temperature | 395 | air burst | 9 |
| wildfires | 6,645 | landslide | 389 | airburst | 9 |
| hail | 6,582 | heat waves | 364 | floodings | 9 |
| global warming | 6,159 | forest fire | 348 | adaptive capacity | 8 |
| droughts | 5,768 | rainstorms | 346 | apocalypse | 8 |
| windstorms | 4,882 | snowstorm | 343 | volcanos | 8 |
| tornado | 4,756 | hailstorm | 330 | climate warms | 7 |
| tsunamis | 3,420 | El Nino | 310 | cold wave | 6 |
| winds | 3,412 | natural hazard | 270 | extreme rains | 5 |
| tsunami | 3,184 | whirlpools | 241 | firestorm | 5 |
| freezes | 3,114 | mudslide | 226 | cataclysms | 3 |
| typhoons | 3,062 | high wind | 214 | debacles | 3 |
| wildfire | 2,537 | blizzard | 213 | cataclysm | 2 |
| derecho | 2,225 | thunderstorm | 204 | seiche | 2 |
| landslides | 2,055 | lava flows | 172 | ashfall | 1 |
| calamity | 2,008 | disaster risks | 155 | microburst | 1 |
| hailstorms | 1,477 | coastal erosion | 146 | rockfall | 1 |
| flooded | 1,309 | heat wave | 144 | rockfalls | 1 |
| | | | | rogue waves | 1 |

Table 2.52: The frequency of words/phrases related to natural disasters in Form 10-Ks

# Chapter 3
# Corporate financial distress prediction in a transition economy


## Abstract

Forecasting the financial distress of corporations is a difficult task in economies undergoing transition, as data is scarce and highly imbalanced. This research tackles these difficulties by gathering reliable financial distress data in the context of a transition economy and employing the synthetic minority oversampling technique (SMOTE). The study employs five different models, including linear discriminant analysis (LDA), logistic regression, support vector machines (SVM), neural networks, and the Merton model, to predict financial distress of public firms in Vietnam between 2011 and 2021. The first four models use accounting-based variables, while the Merton model utilizes market-based variables. The findings indicate that while all models perform fairly well in predicting results for non-delisted firms, they perform somewhat poorly in predicting results for delisted firms in terms of various accuracy measures such as balanced accuracy, precision, recall, and $F_1$ score. The study shows that the models that incorporate both the Altman's and Ohlson's variables consistently outperform those that only use the Altman's or Ohlson's variables in terms of balanced accuracy. Additionally, the study finds that neural networks are consistently the most effective models in terms of both balanced accuracy and Matthews correlation coefficient (MCC). The most important variable in Altman's variables as well as the combination of the Altman's and Ohlson's variables is `reat` (retained earnings over total assets), whereas `ltat` (total liabilities over total assets) and `wcapat` (working capital over total assets) are the most important variables in Ohlson's variables. The study also reveals that in most cases, the models perform better in predicting results for big firms than for small firms, and for good years than for bad years in terms of MCC.



**Keywords**: financial distress, transition economy, Altman's variables, Ohlson's variables, Merton model, machine learning.

**JEL Codes**: C45, C53, G33, M21.


---


This is a joint work with Bang Nguyen and Minh-Ly Lieu and now is under review at *Journal of Forecasting*. We thank Thanh Ngo for supporting us in collecting the data for this study. We thank Ruben Juarez, Peter Fuleky, Miroslav Gabrovski, Shirley Daniel, Yuanzhang Xiao, and Ting-Tsen Yeh for their helpful comments and discussions. All errors are our own.

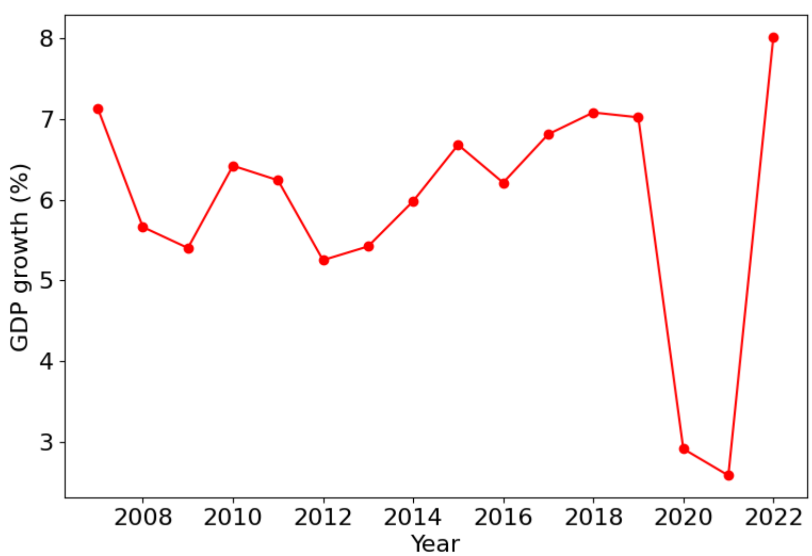
## 3.1 Introduction

### 3.1.1 Motivations

Corporate financial distress prediction is very critical in corporate risk management. It is even more critical in transition economies where data is limited and might not be reliable[1]. This task is more challenging in the economic distress periods.

In the last fifteen years, Vietnam experienced some economic distress periods including the 2008-2009 financial crisis, macroeconomic economic fluctuation in 2012-2014, and the COVID-19 pandemic in 2020-2021. Table 3.1 shows the GDP growth of Vietnam's economy in the 2007-2022 period. This dynamic market is also a challenge for all entities in that market. The need to predict the stable performance of companies is essential in many activities such as when making investment decisions, financing decisions, and lending decisions. The main goal of this study is to explore models to predict the corporate financial distress of the public firms in Vietnam during the period with several economic distress. Since firms tend to face financial distress situations during economic distress periods, this study is necessary to understand the financial health status when the economic conditions are not good. Our study focuses on Vietnam as an example of a transition economy in which the legal regulations, economic conditions, and availability of data are very different compared to that in many advanced economies.

Figure 3.1: GDP growth of Vietnam's economy between 2007 and 2022




*Source*: The World Bank

[1] A transition economy is an economy that is transforming from a centrally planned economy to a market economy. In East and South-East Asia, there are four transition economies including Cambodia, China, Laos, and Vietnam. There are other 25 transition economies in other areas on over the world. For more details, see https://www.imf.org/external/np/exr/ib/2000/110300.htm.

Since the bankruptcy data in many developing and transition economies like Vietnam is usually either not available or not reliable as in many developed economies like the United States, this study focuses on the corporate financial distress probability rather than the corporate bankruptcy probability. We consider firms delisted on the stock markets due to financial reasons as financial distress firms and firms not delisted on the stock markets due to financial reasons as non-financial distress firms.

### 3.1.2 Related literature

**Financial distress vs. bankruptcy prediction**

Although existing research efforts with solid data have been put on financial distress in advanced economies (e.g. Opler and Titman 1994; Altman et al. 2019), little reliable notice has been paid to financial distress in transition economies and developing countries, where normally there is a lack of reliable data and unstable bankruptcy law, especially under economic shocks. To the best of our knowledge, this study is the first of its kind in Vietnam. It is noticeable that the bankruptcy procedure of Vietnam's Law on Bankruptcy in 2014 is clear, but it has rarely been implemented in reality. Therefore, it is infeasible to collect the data of bankruptcy cases in Vietnam. That is why this study focuses on corporate financial distress rather than corporate bankruptcy. Since financial distress is more realistic and practical in transition economies like Vietnam than bankruptcy, it is more meaningful to find suitable models to predict financial distress rather than bankruptcy. Moreover, financial distress normally is the status before bankruptcy, therefore financial distress prediction is also meaningful in the context that firms can be aware of it as the early warning signal to avoid going bankrupt. Note that since prolonged corporate financial distress might eventually lead to corporate bankruptcy this research is based on the framework of default risk or default probability.

**Accounting-based models**

There are two main approaches in studying corporate bankruptcy prediction which uses accounting-based variables (e.g Altman 1968; Ohlson 1980) and market-based variables (e.g. Merton 1974)[2]. Each method has its own strengths and weaknesses. Indeed, choosing which to build the prediction model is not an easy task.

Altman (1968) pioneers in studying the prediction of corporate bankruptcy and financial distress by using the method of multivariate discriminant analysis (MDA). This statistical technique categorizes variables into groups (or indicators) which reduces the dimensionality

[2]Note that we consider the machine learning models that base on the accounting-based variables as the accounting-based models. That is, in this study, the accounting-based models include both accounting models like Altman (1968) and Ohlson (1980), and also the machine learning models using accounting-based variables.

of high-dimension datasets[3]. In this study, Altman discriminates firms into three zones including "non-bankrupt", "zone of ignorance", and "grey area" (Altman 1968) He argues that the financial health of a company is reflected by five financial indicators, four of which can only be constructed from one-year financial statements and one needed equity market values. The subsequent studies expand Altman (1968) in many various ways. Systematically, one can see the evolution of the Altman Z-score family in Chapter 10 in Altman et al. (2019).

There are many accounting-based studies about financial distress and bankruptcy prediction after Altman (1968) such as Ohlson (1980), Zmijewski (1984), Altman et al. (2017), and many others. Among these studies, Ohlson (1980) is the most typical accounting-based work after Altman (1968). In this study, Ohlson uses the logistic regression rather than the multivariate discriminant analysis method as that in Altman (1968).

**Market-based models**

The Black-Scholes-Merton (BSM) model is a structural credit risk model and is a continuous market model. In a seminal work, Black and Scholes (1973) derives a theoretical valuation formula for options called the Black-Scholes option pricing equation. But it can also be applied to corporate liabilities such as common stock, corporate bonds, and warrants. Inspired by this Black and Scholes's pioneer work, Merton (1974) builds a structural model to estimate the continuous probability of default which is derived from the distance-to-default calculation. Note that in this study the BSM model and Merton model are interchangeable. But the Merton model is mostly used. One of the underlying assumptions of the Merton model is that the total value of a firm follows a geometric Brownian motion of the form

$$dV = \mu_V V dt + \sigma_V V dW,$$

where $V$ is the total value of a firm, $\mu_V$ and $\sigma_V$ are the expected return and the volatility, and $dW$ is the standard Wiener process. Many recent studies about the distance to default admit this standard assumption, for example, Bharath and Shumway (2008) and Vassalou and Xing (2004). Our study also bases on this fundamental assumption.

Based on the BSM model, the KMV-Merton model[4] focuses on the default probability of corporate liabilities (e.g. debt) in the canonical Merton model, the KMV-Merton model investigates the probability of default of an individual company as a whole (Kealhofer 2003).

[3]Indeed, in the modern economy, it is complicated to categorize millions of firms into millions of financial health levels. In this case, categorizing all firms into a few groups is a common and useful statistical technique to analyze high-dimensional datasets that consist of many variables. This technique enables one to investigate whether there are significant differences among groups.

[4]The term KMV stands for Kealhofer, McQuown and Vasicek, who are the founders of the KMV corporation. KMV Corporation was acquired by Moody's Corporation in 2002.

In this model, the value of the company's stock equity and debt are modeled as a long call option and a short put option on the firm's assets, respectively.

When the expected value (unobservable) of the firm is less than the firm's debt, shareholders would choose not to repay (not exercise their call option) while when the firm's value is greater than the firm's debt, they would choose to repay (exercise their call option). That is, ultimately the shareholders' gain equals $\max\{\text{Value}(T) - \text{Debt}(T), 0\}$, where $\text{Value}(T)$ and $\text{Debt}(T)$ are the expected value of the firm and the firm's debt at time $T$, respectively.

The KMV-Merton model is probably the most famous application of the Merton model in the financial sector. Our study will explore the KMV-Merton model and apply it in a transition economy. The KMV-Merton model is used widely in practice and is used broadly by the Big-Three credit rating agencies including Standard & Poor's, Fitch, and Moody's. Some recent studies related to the KMV-Merton model are Kealhofer (2003), Tudela and Young (2005), Benos and Papanastasopoulos (2007), Bharath and Shumway (2008), Lee (2011), and Yeh et al. (2012).

#### Accounting-based vs. Market-based models

The accounting-based models (e.g Altman 1968) and market-based models (e.g KMV-Merton model) are different in many aspects. Table 3.1 shows some differences between these two types of models.

Table 3.1: Comparison the accounting-based and market-based models

| | Accounting-based model | Market-based model |
|---|---|---|
| Firm's equity | Book value | Market value |
| Firm's debt | Book values | (Estimated) market values |
| Volatility of firm's assets | No | Yes |
| Frequency of data | Quarterly, yearly | Daily |

Regarding the ability to estimate the default probability, some studies find that the Merton-type model outperforms the Altman-type model (Hillegeist et al. 2004; Tanthanongsakkun et al. 2011). Specifically, Hillegeist et al. (2004) discover that the BSM option-pricing model provides significantly more information than Altman's (1968) Z-Score and Ohlson's (1980) O-Score. Also, Tanthanongsakkun et al. (2011) detects that the Merton model is the most informative model in explaining corporate bankruptcy compared to other models, which include the Altman model. However, some studies find that the Altman-type model outperforms the Merton model. For example, Reisz and Perlich (2007) estimate probabilities of bankruptcy for 5,784 industrial firms in the period 1988–2002. They find that Altman $Z$-and $Z$"-scores outperform structural models in 1-year-ahead bankruptcy predictions compared to BSM models.

Some studies show that the hybrid model (combine accounting-based and market-based models) outperforms either separate accounting-based or market-based models. For example, Shumway (2001) shows that combining accounting and market variables results in the most accurate model.

### Financial distress prediction in Vietnam

While financial distress prediction studies, especially those studies using machine learning methods, are popular in many other transition economies in decades (Gruszczynski 2004; Chen et al. 2006; Wang and Li 2007; Zheng and Yanhui 2007; Li and Sun 2008; Xie et al. 2011; Sun et al. 2011), there are relatively few and late studies in Vietnam.

For Vietnam's stock market, the previous studies about financial distress or default probability can be classified into four groups. The first group uses the Altman model approach (e.g. Z-score or Z"-score model), the second group utilizes the Ohlson-type approach (i.e. logistic regression or logit), the third group employs the KMV-Merton model approach, and the fourth group examines other models (e.g. mixed models and machine learning models). Despite some effort has been made, there is a scarcity of reliable empirical results in many studies in all four groups.

All studies in the first group that we found use the given coefficients and thresholds from Altman's and related studies, especially Altman (1968) (Lieu 2014; Vo and Nguyen 2014; Nguyen 2015; Hoang 2020). This is clearly a big limitation since, for example, the coefficients of variables and the $Z$-score thresholds derived from 66 firms in one sector (manufacturing) in the 1960s in a developed country like the United States Altman (1968) are not necessarily the same as those derived from firms at the moment and for the whole stock market in a transition economy like Vietnam. It is vital to note that the Z-score model (Altman 1968) is a data-dependent formula, which is definitely not like a universal formula in mathematics. Therefore, that formula works in a particular market or country in a specific period does not necessarily mean it works in another (or even the same) market or country in another period. Using coefficients and thresholds of Altman (1968) or of any other studies to apply to other datasets might lead to misleading and unreliable conclusions.

For the second approach, there are relatively few studies compared to the first approach. Some studies using logistic regression include Vo (2015) and Vo et al. (2019). Both work on highly imbalanced data. Vo (2015) works on the data including 946 non-distress observations and 36 distress observations and Vo et al. (2019) work on the data including 1,572 non-distress observations and 3,718 distress observations. Unfortunately, the authors in both studies do not do anything to take imbalanced data into account. Their results, therefore, tend to be biased toward the majority group (non-distress firms). This implies that their results, especially accuracy in Vo et al. (2019), seem unreliable to evaluate the predictive performances of their models.

There are a number of studies investigating the probability of default using KMV-Merton model in Vietnam (Lam and Phan 2009; Le and Le 2012; Nguyen and Pham 2014; Nguyen and Nguyen 2017; Vu et al. 2019). Unfortunately, the results in these studies lack reliability and/or verification. To the best of our knowledge, Lam and Phan (2009) is the first study considering the KMV-Merton model in Vietnam but this paper actually discusses very little about the KMV-Merton model. Le and Le (2012) claim that they combine CVaR and KMV-Merton model to examine the corporate default probability. However, the paper clearly did not derive any results related to the KMV-Merton model. Also, the authors claim that CVaR is better VaR to estimate the default probability under shock. This is an overclaim since they did not compare CVaR and VaR explicitly in their study. Nguyen and Pham (2014) analyze 6,398 listed and unlisted firms as corporate customers of a bank named Vietcombank in Vietnam and find that the probability of default of the whole portfolio is 2.6%. This study, however, does not verify the results or show the accuracy of the prediction.

Regarding the other approaches, Ninh et al. (2018) uses a combination of accounting, market, and macroeconomic variables to examine the corporate financial distress of 800 listed firms on Vietnam's stock market from 3003 to 2016. Unfortunately, the accounting-based model borrows the coefficients and thresholds in the Z"-score model rather than constructing from their own datasets. Moreover, for the Merton model, it is not clear from the paper whether the authors estimate the firm value, expected rate of return, and volatility or not. Hence, the results related to the Merton model are inconclusive.

**Delisted firms due to financial distress**

Geng et al. (2015) study the "special treatment" firms on China's stock markets. These firms are considered as special ones due to several reasons including two years of losses, damaged business, or financially bankrupt. Hosaka (2019) considers delisted firms with the reasons of (1) bankruptcy or rehabilitation or reorganization procedures, (2) excessive debt, (3) suspension of bank transactions, (4) termination of business activities (excluding mergers) as bankrupt firms in this study.

Table 3.2: Delisted reasons for bankruptcy in several studies

| Country | Study | Delisted reason |
|---|---|---|
| United States | Mai et al. (2019) | Bankruptcy and liquidation |
| Japan | Hosaka (2019) | Bankruptcy/rehabilitation/reorganization procedures, excessive debt, suspension of bank transactions, and termination of business activities |
| China | Geng et al. (2015) | Two years of losses, damaged business, financially bankrupt |
| Vietnam | This study | Negative profit in three consecutive years or cummulative loss (negative profit) is greater than authorized capital |

### Sample size of bankrupt vs. non-bankrupt firms

In the literature on financial distress and bankruptcy prediction, there are both studies with equal bankrupt (or distress) and non-bankrupt (or non-distress) firms. The good news for imbalanced data is that there are several techniques for balancing data (e.g. synthetic minority oversampling technique or SMOTE) or balancing results (e.g. balanced accuracy). In this study, we use SMOTE as a technique for balancing train data[5]. That is, by using SMOTE, the number of observations of delisted firms is equal to that of non-delisted firms. We will discuss more about SMOTE and how to use it for balancing data later.

Table 3.3: Number of bankrupt and non-bankrupt firms in the related literature

| BR | NB | Matched-pairs design | Period | Years prior to bankruptcy | Method | Source |
|---|---|---|---|---|---|---|
| 79 | 79 | Yes | 1954-1964 | 1-5 | UDA | Beaver (1966) |
| 33 | 33 | Yes | 1946-1965 | 1-2 | MDA | Altman (1968) |
| 162 | 162 | Yes | 1956-1976 | 1 | LR | Collins (1980) |
| 105 | 2058 | No | 1970-1976 | 1-2 | Logit model | Ohlson (1980) |
| 1600 | 81 | No | 1972-1978 | NA | Probit model | Zmijewski (1984) |
| 65 | 64 | Yes | 1975-1982 | 1 | NN | Odom and Sharda (1990) |
| 65 | 64 | Yes | 1975-1982 | 1 | NN | Wilson and Sharda (1994) |
| 1160 | 1160 | Yes | 1996-1999 | NA | SVM | Shin et al. (2005) |
| 58 | 142 | No | 1971-1981 | NA | Decision trees | Gepp et al. (2010) |
| 107 | 107 | Yes | 2008-2011 | 3-5 | DT, SVM, NN | Geng et al. (2015) |
| 102 | 2062 | No | 2002-2016 | 4 | CNN | Hosaka (2019) |
| 76 | 983 | No | 2011-2021 | 1-3 | Logit, Merton, LDA, SVM, NN | This study |

*Notes*: BR and NB stand for the number of bankrupt and non-bankrupt firms, respectively; UDA stands for univariate discriminant analysis; MDA stands for multivariate discriminant analysis; NN stands for neural networks; SVM stands for support vector machines; DT stands for decision trees; DA stands for discriminant analysis; LR stands for logistic regression.

Overall, our goal is to explore and compare the predictive performance of various models in predicting corporate financial distress for public firms in Vietnam. The main research question is: What are the differences between accounting-based, market-based, and machine-learning models in predicting corporate financial distress in Vietnam?

**Contributions**. Our paper has three main contributions. First, this is the first study employing three methods, including accounting-based, market-based, and machine learning models, to investigate the financial distress of public firms in Vietnam. Second, this is also the first study employing the synthetic minority oversampling technique (SMOTE) to tackle

[5]This technique was first proposed by Chawla et al. (2002).

the biased results caused by imbalanced data on financial distress in Vietnam. Third, we provide explicit and new data for the Merton model, which is helpful for future research.

The remainder of this chapter is organized as follows. Section 3.2 present all methods that we use in this chapter. Section 3.3 describes the data and variables of all models. The results of all models are presented in Section 3.4. We then investigate the effects of firm size and macroeconomic situation on predictive performances in Section 3.5. Hence, we discuss some noticeable points of the chapter in Section 3.6 before concluding in Section 3.7.

## 3.2 Methods

In nature, the corporate financial distress prediction is a classification problem. Therefore, the goal of the models employed to predict corporate financial distress is to classify the firms into a bankrupt and non-bankrupt class. In order to do that this study employs various models including logistic regression, support vector machines (SVM), linear discriminant analysis (LDA), neural networks, and Merton model. We choose logistic regression and SVM since these are the benchmark methods for classification problems. We choose LDA since this method is relevant to the classic method of multiple discriminant analysis (MDA) used in the seminal work (Altman 1968). We choose neural networks since this is the typical method in machine learning and deep learning. Finally, the Merton model is the standard market-based method in the bankruptcy and financial distress prediction literature.

Note that the first four methods will work with the accounting-based variables. Therefore, we can consider these four methods as accounting-based models. However, one can consider logistic regression as the accounting-based method in the context of Ohlson (1980) and the other three (LDA, SVM, and NN) as machine learning (ML) methods. In this context, this study employs three kinds of models, including accounting-based, market-based, and machine-learning models.

Note also that this study focuses on a transition economy therefore we do not use the standard coefficients of Altman's Z-Score and Ohlson O-Score model since these models are built in the context of developed economies. Using the coefficients in Altman (1968) or Ohlson (1980) as the formulas and then applying them to data in transition economies (and also developing countries) might not be a good idea. The reason is that the MDA method used in Altman (1968) and logistic regression used in Ohlson (1980) are statistical models in nature and the coefficients of these models surely depend on the input data. Therefore, for example, using coefficients obtained from a sample of 66 manufacturing firms in the United States in the 1960s (Altman 1968) and then applying for example in the retail industry in Vietnam in the 2020s might be a misleading idea since the data and context are totally different. Surprisingly, there are many studies like that out there, including most of the studies in the bankruptcy and financial distress literature in Vietnam.

The good news is that while using the coefficients from Altman's and Ohlson's models might not be a good idea, one can employ the variables used in these two standard models. The reason is that the list of Altman's variables (working capital/total assets, retained earnings/total assets, earnings before interest and taxes/total assets, total sales/total assets, and market value of equity/total liabilities) and Ohlson's variables (nine-factor financial and economic variables) represent for the fundamental factors that affect the firm's activities and performances, which then affect the probabilities of going bankrupt or being financial distress. This study employs Altman's and/or Ohlson's variables as the independent variables for our chosen methods.

### 3.2.1 Logistic regression

Logistic regression is a probabilistic model that takes the linear combination of a list of variables (linear regression) and maps it to a sigmoid function. For example, given $n$ independent variables, says $X_1, ..., X_n$, the logistic regression given by

$$p(X_1, ..., X_n) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + ... + \beta_n X_n)}},$$

where $\beta_0, \beta_1, ..., \beta_n$ are the coefficients of the corresponding linear regression. One can consider logistic regression as the benchmark for the classification problem.

### 3.2.2 Linear discriminant analysis

Linear discriminant analysis (LDA) was pioneered by Ronald Fisher in 1936 (Fisher 1936) for discriminating data with two classes. Thus, it is generalized by C. R. Rao, a Fisher's student, for more than two classes of the model in Rao (1948).

This is a linear classification model and is used to find linear combinations of variables (or features) that classify two or more classes or groups. In the LDA model, input variables are assumed to be normally distributed. Another assumption is that the input variables should not be correlated with each other. That is, we should deal with the severe multicollinearity if it appears. Note that, in practice, data should be transformed (standardized and normalized) before training with the LDA model.

It is important to note that the normally distributed assumption is crucial when we want to estimate the magnitude of the marginal effects of the predictor variables. However, for prediction or classification purposes (as in this study), this assumption is less sensitive (Shayan et al. 2016). That is why, in practice, the LDA model might perform well even though the normally distributed assumption is violated.

It is necessary to note that LDA is different compared to both linear regression and logistic regression. LDA and logistics are techniques used for classification problems where the dependent variable is discrete (e.g. binary variable) while linear regression is used for prediction/causal inference where the dependent variable is continuous. LDA requires the multivariate normality and equality of covariance matrices among groups whereas logistic regression does not.

### 3.2.3 Support vector machines

Support Vector Machines (SVM) is a type of supervised learning algorithm that can be used for classification and regression tasks. In the context of bankruptcy or financial distress prediction, SVM can be used to classify firms as either bankrupt (or distressed) or non-bankrupt (or non-distress) based on a set of variables such as the firm's financial characteristics, market and economic variables.

SVM is pioneered by Boser et al. (1992) and Cortes and Vapnik (1995). There are several types of SVM that have different kernels such as linear, polynomial, radial basis function (RBF), and sigmoid. In this study, we employ RBF SVM since it has a number of advantages compared to other types of SVM such as handling nonlinear relation between class labels, few hyperparameters, and few numerical difficulties (Hsu et al. 2016).

Formally, as shown in Hsu et al. (2016), given a train data $(\mathbf{x}_i, y_i), i = 1, ..., l$ where $\mathbf{x}_i \in \mathbb{R}^n$ and $y \in \{1, -1\}^l$. RBF SVM requires the solution of the following optimization problem

$$\begin{aligned} \min_{\mathbf{w},b,\boldsymbol{\xi}} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i, \\ \text{subject to} \quad & y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq 1 + \xi_i, \\ & \xi_i \geq 0, \end{aligned}$$

where $C > 0$ is the penalty parameter of the error term, $\mathbf{w}$ is the vector of weights, and $b$ is a scalar, and with RBF kernel given by $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$, where $\gamma > 0$.

### 3.2.4 Neural networks

One of the most limitations of the MDA-based studies, for example, Altman (1968), is that this method requires the discriminating variables to be jointly multivariate normal (Odom and Sharda 1990). We can relax this restriction of multivariate normality by using several machine learning techniques including neural networks and decision trees. Neural networks are flexible since we can adjust the number of inputs, hidden layers, and neurons in each hidden layer. Neural networks can learn complex relationships between variables (predictors) and make predictions with a high degree of accuracy.

Note that, we choose threshold=0.5 to determine whether a firm predicted financial distress or not. It is clear that 0.5 is the standard threshold in the fields of forecasting and machine learning. In fact, we also tried threshold=0.4 since we thought the firms in transition economies like Vietnam have weak financial structures and depend more on debt than firms in developed countries. In preparing the models, we find that 0.5 is a better threshold compared to 0.4 as most of the accuracy and Matthews coefficient correlations are higher than in the former case. One technical reason that we choose 0.5 as the threshold is that by choosing this threshold the accuracy from Keras and the accuracy from Sk-learn are the same. Another threshold, for example, 0.4, will lead to different accuracy when using Keras and when using Sk-learn. In that case, one needs to discuss which libraries that he or she prefers. Fortunately, by choosing threshold=0.5 we do not need to do so.

### 3.2.5 Merton model

Denote $V_t^A$ as the dynamics for the value of the firm's assets at time $t \in [0, T_t]$. Thus, $V_t^A$ is described by the following geometric Brownian motion

$$dV_t^A = (\mu_t^A V_t^A - D_t)dt + \sigma_t^A V_t^A dW_t, \tag{3.1}$$

where $\mu_t^A$ is the instantaneous expected rate of return at time $t$, $\sigma_t^A$ is the instantaneous estimated volatility of the return on the firm per unit time, $D_t$ is the amount of money that the firm pays for liabilities-holders (e.g. interest rates of corporate bonds) per unit time, $dW_t$ is the standard Wiener process. The formal definition of default probability is defined as follows.

**Definition 3.2.1.** *(default probability) The probability of default of a firm is defined as*

$$PD = P(V_{t+m}^A < D_t), \tag{3.2}$$

*where $D_t$ is the financial obligations (or debts) that the firm needs to pay at time $t$, and $V_{t+m}^A$ is the estimated market value of the firm at time $t+m$, where $t \in [0, T_t]$, $m \in [0, T_m]$, and $0 \leq T_t, T_m < \infty$.*

In practice, it is usually to assume $m = 1$ year (Afik et al. 2016). That is, one might examine the debt at the current time $t$ and estimate the value of a firm one year after that, i.e. at time $t+1$. Therefore, we propose the following definition.

**Definition 3.2.2.** *(one-year probability of default) The probability of default of a firm in one year is defined as*

$$PD = P(V_{t+1}^A < D_t), \tag{3.3}$$

*where $D_t$ is the financial obligations (or debts) that the firm needs to pay at time $t+1$, and $V_{t+1}^A$ is the estimated value of the firm at time $t+1$.*

Note that the market value $V_{t+m}^A$ in Definition 3.2.1 and $V_{t+1}^A$ in Definition 3.2.2 are unknown. Therefore, we need to estimate it. Since we are at the current time $t$ and want to estimate the probability of default in time $t+m$, we actually investigate the *implied* probability of default. Assume the forecasting horizon is one year, i.e. $m=1$. For simplicity, we call the one-year probability of default as the probability of default from now on. Note that one year or less is the maturity date of short-term debt. Theoretically, if a firm cannot pay the short-term debt within a year to creditors, it will go bankrupt. In the standard finance literature (e.g. Vassalou and Xing 2004; Bharath and Shumway 2008), one normally takes into account total short-term debt and half of long-term debt to compute the book value of debt $D_t$. In this study, we will consider the book value of debts as a function of short-term debt and long-term debt

$$D_t = \text{Short-term debt} + k(\text{Long-term debt}), \tag{3.4}$$

where $k \in [0,1]$. In this study, we choose $k=0.5$.

Denote $V_t^E$ be the market value of the firm's equity at time $t$, the Black-Scholes formula for call options is given by the following equation

$$V_t^E = V_t^A N(d_1) - D_t e^{-r_t m} N(d_2), \tag{3.5}$$

where $r_t$ is the instantaneous risk-free interest rate at time $t$, $D_t$ is the book value of the firm's debt at time $t$, $N$ is the standard normal cumulative density function, and

$$d_1 = \frac{\ln(V_t^A/D_t) + \left(r + 0.5(\sigma_t^A)^2\right)m}{\sigma_t^A\sqrt{m}}, \quad d_2 = d_1 - \sigma_t^A\sqrt{m}.$$

Equation (3.5) implies that the value of a firm's equity $V_t^E$, which we can observe on the stock market by multiplying the current stock price by the total number of firm's shares outstanding, is a nonlinear function of the expected value of the firm $V_t^A$, which we cannot observe directly. Note that in equation (3.5), both the value of the firm $V_t^A$ and the (implied) volatility of the underlying asset price $\sigma_t^A$ are unobservable. Therefore, we need to find $\sigma_t^A$ and $V_t^A$ by solving equation (3.5).

Since the lack of comprehensive bankruptcy data in many transition economies like Vietnam, this study will work on corporate financial distress probability but still be based on the framework of the default probability.

## 3.3 The data

Since the data and variables for models using accounting-based variables are distinct from those for the Merton model, we describe the procedure of collecting data and preparing variables separately.

### 3.3.1 For accounting-based and machine learning models

#### Data collection

For non-delisted firms, we collect the data of the non-financial listed firms in Vietnam's stock market from 2009 to 2021. Although Vietnam's stock market was established in 2000, the scale of the market and the number of listed firms were pretty limited before 2009. Thus, this paper focused on the period from 2009. Note that since financial institutions such as banks, insurance, and hedge funds have different financial characteristics, therefore, are excluded from our data.

For the accounting-based and machine learning models[6], the financial reports of both delisted and non-delisted firms are obtained from the Refinitiv Eikon database. We collect the delisted stocks due to financial reasons as the proxies for financial distress on the Ho Chi Minh Stock Exchange (HOSE) and Hanoi Stock Exchange (HNX). Note that financial reasons, according to Decree 155/2020/ND-CP, include (1) negative profit in three consecutive years and (2) cumulative loss (negative profit) is greater than authorized capital.

Table 3.4 summarizes the datasets and the corresponding sources that we use in this study. We employ data from five sources. For the variables of firms' characteristics and market value of equity of each firm, we get the data from the Refinitiv Eikon database. For Gross National Product (GNP), we get the data from the World Bank's website. We obtain the risk-free interest rate of Vietnam's 1-year government bond from `www.investing.com`, which is a reliable source for financial markets and investment. Finally, we get the list of delisted stocks from the Ho Chi Minh City Stock Exchange (HOSE) and the Hanoi Stock Exchange (HNX).

Table 3.5 shows, in the final sample of accounting-based and machine learning models, that there are 1,097 firms and 12,685 observations in total including 1,021 non-listed firms (with 12,154 observations) and 76 delisted firms (with 531 observations). It is clear that our sample is an imbalanced data with 12,154 observations (approximately 95.8%) of the majority group (non-delisted firms) and 531 observations (approximately 4.2%) of the minority group (delisted firms). The final sample of the Merton model has

[6] Note that the accounting-based model in this study means logistic regression that is widely used in accounting literature, for example, the seminal work of Ohlson (1980) and subsequent studies. Machine learning models in this study include LDA, SVM, and neural networks. In the next subsection, we will discuss the data and variables for the Merton model, which is a market-based model.

Table 3.4: The data and corresponding sources

| Data | Source |
|---|---|
| Firm's characteristics | Refinitiv Eikon database |
| Gross National Product (GNP) | World Bank |
| Market value of equity | Refinitiv Eikon database |
| Risk-free interest rate | www.investing.com |
| Delisted stocks | HOSE, HNX |

7,707 observations with 6,967 observations (about 90.4%) of non-delisted firms and 740 observations (about 9.6%) of delisted firms. There are 1,059 firms in the final sample of the Merton model with 76 delisted firms and 983 non-delisted firms.

Table 3.5: Number of firms and observations in the final samples

| | Accounting-based and ML models | | Merton model | |
|---|---|---|---|---|
| | Number of firms | Number of obs | Number of firms | Number of obs |
| Delisted firms | 76 | 531 | 76 | 740 |
| Non-delisted firms | 1,021 | 12,154 | 983 | 6,967 |
| Total | 1,097 | 12,685 | 1,059 | 7,707 |

*Notes.* Initially, there are 108 delisted stocks due to financial reasons on HOSE and HNX between 2011 and 2021 as shown in Appendix 2. However, there are only 76 firms (corresponding to 76 delisted stocks) having full data for estimating the financial distress models.

**Data preprocessing**

From the data collected from the Refinitiv Eikon database, we generate the independent variables as those in Altman (1968) and Ohlson (1980) as these are the standard accounting-based models of bankruptcy prediction. Altman's variables include working capital on total assets (wcapat), retained earnings on total assets (reat), earnings before interest and taxes on total assets (ebitat), sales on total assets (saleat), and market value of equity on total liabilities (mvdebt). The Ohlson's variables include size of the firm (size) (proxy by total assets), total liabilities on total assets (ltat), working capital on total assets (wcapat), total current liabilities on total current assets (lctact), dummy variable on whether total liabilities exceed total assets (OENEG), return on assets, the fund provided by operations (proxy by the cash flow of operating) on total liabilities (cfodebt), dummy variable whether net income is negative in the last two years (INTWO), and income growth (CHIN). For more details, please see Table 3.6.

We fill the missing values of each column with the mean of that column group by stock then we delete any left missing values. Hence, we winsorize the top 5% and bottom 5% of the data points by using the Winsorizing technique. Specifically, the data points in the top 5% of each column are replaced by the value at the 95th percentile and the data points in the bottom 5% of each column are replaced by the value at 5% percentile. Our final data includes 12,685 firm-year observations and 1,097 firms as shown in Table 3.5.

### The variables

There are five variables in Altman's variables and nine variables in Ohlson's variables. Since there is one common variable (i.e. wcapat) in two lists, the combination has only thirteen variables. However, we have to remove two variables (ebitat and lctact) to avoid the multicollinearity problems in the combination model. We will present how to detect multicollinearity in Subsection 3.3.1. Finally, in the case of Altman's and Ohlson's variables (i.e. combination), there are only eleven variables as the inputs for bankruptcy prediction models. These indicators reflected five financial aspects of firms including liquidity, profitability, productivity, solvency, and management's capability. All variables used for calculating these financial ratios are collected for financial statements, balance sheets, and cash flow statements of the firms.

Table 3.6: List of all independent variables in accounting-based and ML models

| Variable | Type | Definition |
|---|---|---|
| `wcapat` | Liquidity | Working Capital/Total Assets |
| `reat` | Profitability | Retained Earnings/Total Assets |
| `ebitat` | Profitability | Earnings Before Interest and Taxes/Total Assets |
| `mvdebt` | Solvency | Market Value of Equity/Total debt |
| `saleat` | Management's capability | Sales/Total Assets |
| `size` | Size | log(total assets/GNP price-level index) |
| `ltat` | Solvency | Total liabilities/ Total assets |
| `lctact` | Solvency | Total current liabilities/ Total current assets |
| `OENEG` | Solvency | Dummy: Total liabilities exceed total assets |
| `roa` | Profitability | Net income/Total assets |
| `cfodebt` | Solvency | Funds provided by operations/ Total liabilities |
| `INTWO` | Solvency | Net income $<0$ for in the last 2 years |
| `CHIN` | Management's capability | $\frac{\text{Net income}(t)-\text{Net income}(t-1)}{\lvert\text{Net income}(t)\rvert+\lvert\text{Net income}(t-1)\rvert}$ |

*Notes*: In most cases, we keep the names of variables in this table as that in Altman (1968) and Ohlson (1980). For the variable `CHIN`, in the next sections of the paper we use income growth (denoted by `incgrowth`) as the alternative (yet more intuitive) variable name.

In order to determine the financial distress status, we create a dummy variable named bankrupt_period which takes the value of 1 from the delisted year to the current year and 0 for non-listed firms. Note that we define bankrupt_period equals to 1 from delisted year to the current year because the fact that it is very rare that once a firm is delisted from the two main stock markets (Ho Chi Minh City Stock Exchange - HOSE and Hanoi Stock Exchange - HNX), that firm would be listed again on these two stock markets after that[7]. We actually do not observe any cases like that at the moment.

[7]Those stocks are delisted from HOSE and HNX that will be moved to an informal market called Unlisted Public Company Market (or Upcom).

Table 3.7 illustrates the differences between delisted and non-delisted firms in several other factors including total debt, market values of equity, total assets, and cash flow from operations. The $t$-test method is normally used to determine whether there are significant differences between the means of two groups. The $p$-values of the $t$-test indicate that the differences of the four factors between two groups of firms are statistically significant at the significant level of $\alpha = 0.05$. Note that these four factors shown in Table 3.7 are not included in the list of variables of any models. However, we use them as the intermediate inputs to generate the final list of variables. The differences in these four factors imply that they might be the good features to predict bankruptcy then so are the final list of variables.

Table 3.7: Means of some factors of delisted vs. non-delisted firms

| | debt | equity | at | cf_operating |
|---|---|---|---|---|
| Non-delisted firms | 511.2294 | 657.3682 | 1716.3659 | 93.7074 |
| Delisted firms | 813.5192 | 111.8401 | 1441.8224 | 39.7954 |
| $p$-value ($t$-test) | 0.0000 | 0.0000 | 0.0256 | 0.0000 |

*Notes*: Debt stands for the total debt of firms in the balance sheet. Equity stands for the total equity in the balance sheet. ta is denoted for the total assets of firms, and cf_operating stands for the operating cash flow from the cash flow statement. In Python, one can use the `ttest_ind` method in the `scipy.stats` package to calculate two-sampled $t$-test and get the corresponding $p$-value. This method calculates the $t$-test for the means of two independent samples. The null hypothesis is that two independent samples have identical averages. The t statistic is given by $t = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}}$, where $s_p = \sqrt{\frac{(n-1)s_1^2+(m-1)s_2^2}{n+m-2}}$. Note that, by default, $t$-test assumes that the populations have identical variances.

Table 3.8 shows the differences between delisted and non-delisted firms in terms of the mean of Altman's variables. It can be seen that these two groups are significantly different in terms of wcapt, reat, ebitat, mvdebt, and saleat. The $p$-values of the $t$-test indicate that the differences of these five variables between two groups of firms are statistically significant. Note that in the $t$-test, the null hypothesis is that the means of variables in two groups are equal. Choosing the significant level of $\alpha = 0.05$. Since all $p$-values are smaller than 0.05, we can reject the null hypothesis. This indicates there are significant differences in characteristics between delisted and non-delisted firms. The differences between these two groups in terms of Altman's variables imply that these five variables might be good factors to predict whether a firm goes bankrupt or not. In this case, the $t$-test plays the role of a good selection method.

Table 3.9 shows the summary statistics of all Altman's variables and all Ohlson's variables. Note that in the combination of Altman's and Ohlson's variables, we do not use all variables in this table. Rather, we will eliminate some variables to avoid the multicollinearity issue.

Table 3.8: Means of the Altman's variables of delisted vs. non-delisted firms

| | `wcapat` | `reat` | `ebitat` | `mvdebt` | `saleat` |
|---|---|---|---|---|---|
| Non-delisted firms | 0.1779 | 0.0440 | 0.0683 | 14.8720 | 1.0204 |
| Delisted firms | -0.0423 | -0.1913 | 0.0063 | 7.1502 | 0.5001 |
| $p$-value ($t$-test) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

Table 3.9: Summary statistics

| Variable | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| `bankrupt_period` | 12685 | 0.0418 | 0.2002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 |
| `wcapat` | 12685 | 0.1687 | 0.2106 | -0.2234 | 0.0266 | 0.1473 | 0.3123 | 0.5800 |
| `reat` | 12685 | 0.0342 | 0.1173 | -0.3163 | 0.0102 | 0.0435 | 0.0924 | 0.2209 |
| `ebitat` | 12685 | 0.0657 | 0.0623 | -0.0426 | 0.0225 | 0.0578 | 0.1023 | 0.2022 |
| `mvdebt` | 12685 | 14.5488 | 32.7578 | 0.1041 | 0.5837 | 1.9153 | 8.3740 | 135.0364 |
| `saleat` | 12685 | 0.9986 | 0.7752 | 0.0780 | 0.3755 | 0.7934 | 1.4135 | 2.8690 |
| `size` | 12685 | 4.8465 | 1.4981 | 2.3140 | 3.7416 | 4.7473 | 5.8391 | 7.8813 |
| `ltat` | 12685 | 0.5279 | 0.2351 | 0.1166 | 0.3394 | 0.5397 | 0.7104 | 0.9375 |
| `lctact` | 12685 | 0.7929 | 0.4743 | 0.1604 | 0.4721 | 0.7316 | 0.9485 | 2.1409 |
| `OENEG` | 12685 | 0.0342 | 0.1819 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 |
| `roa` | 12685 | 0.0438 | 0.0557 | -0.0675 | 0.0088 | 0.0353 | 0.0751 | 0.1662 |
| `INTWO` | 12685 | 0.0732 | 0.2605 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 |
| `incgrowth` | 12685 | -0.0029 | 0.4921 | -1.0000 | -0.2197 | 0.0000 | 0.2049 | 1.0000 |
| `cfodebt` | 12685 | 1.2403 | 2.8983 | -0.8524 | -0.0311 | 0.2042 | 0.8976 | 11.529 |

*Notes*: This table shows the summary statistics of variables in this study. `wcapat` is the ratio of working capital to total assets. `reat` is the ratio of retained earnings to total assets. `ebitat` is the earnings before interest and taxes to total assets. `mvdebt` is the market value of equity to total debt. `saleat` is the ratio of sales to total assets. `size` is the logarithm of the ratio of total assets to GNP adjusted to the price-level index. `ltat` is the ratio of total liabilities to total assets. `lctact` is the total current liabilities to total current assets. `OENEG` is a dummy variable that takes value 1 when total liabilities exceed total assets and 0 otherwise. `roa` is the ratio of net income to total assets. `INTWO` is a dummy variable that takes value 1 when net income is negative in the last 2 years. `incgrowth` is the ratio between the change in net income between time $t-1$ and $t$ over the summation of net income in time $t-1$ and $t$. `cfodebt` is the ratio of funds provided by operations to total liabilities.

### Correlation matrix

We will show the correlation matrix of each list of variables using a heat map. Figure 3.2 shows the heat map of Altman's variables. It is clear that all correlation coefficients of Altman's variables are less than 0.8, which is usually used as an empirical threshold to detect highly correlation between each pair of variables. A striking point from the heat map is that all correlation coefficients of each pair of independent variables and dependent variables are negative.

Figure 3.2: The heat map for the Altman's variables

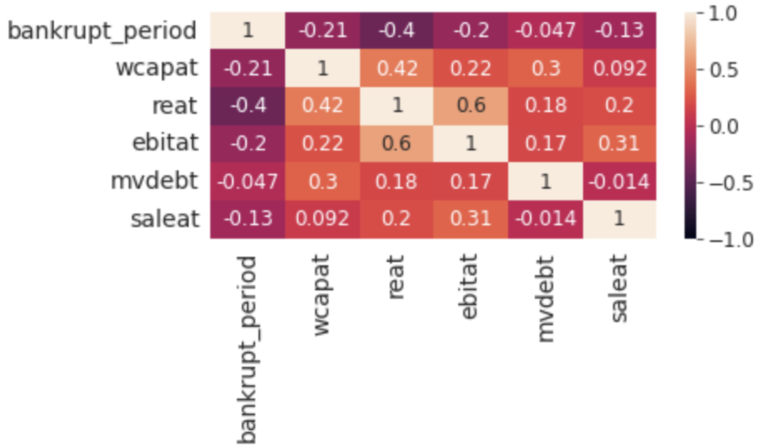



Figure 3.2 shows the heat map of Ohlson's variables. The absolute values of all correlation coefficients are less than 0.8, except that between lctact and wcapat. Since the completeness of the list of Ohlson's variables, we decide to keep both lctact and wcapat in the models for Ohlson's variables even though they are highly negatively correlated with each other.

Figure 3.3: The heat map for the Ohlson's variables

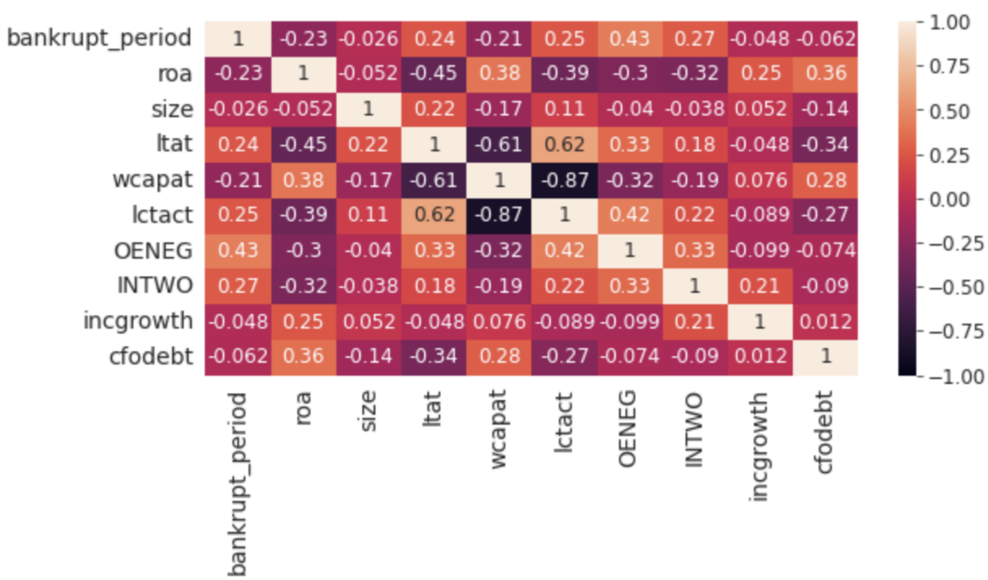
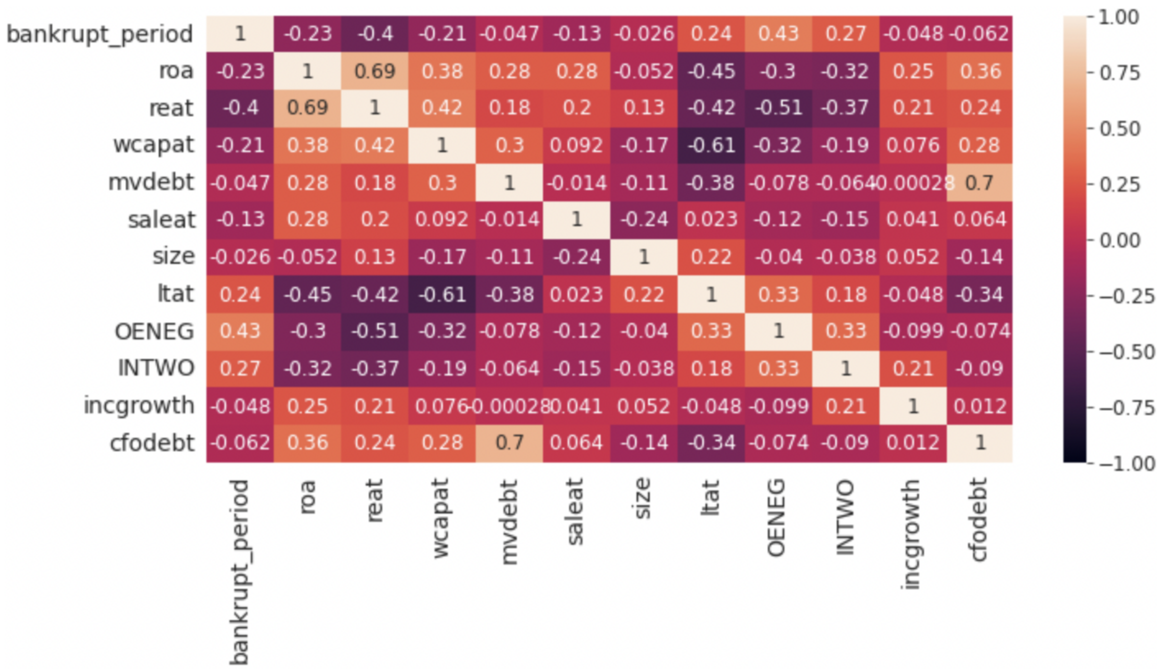



Figure 3.4 shows the correlation matrix among all eleven variables in the combination list of Altman's and Ohlson's variables as well as the dependent variable (i.e. bankrupt_period). We can see that all correlation coefficients are less than 0.8. Note that we have to remove ebitat and lctact since these variables have VIF's values greater than 10. Also,

the correlation coefficient between ebitat and roa is 0.87, and the correlation coefficient between lctact and wcapat is -0.87. We do not show the values of these two pair correlations in Figure 3.4 to avoid redundancy.

Figure 3.4: The heat map for the Altman's and Ohlson's variables




It can be seen that bankrupt_period has the negative correlation with roa, reat, wcapat, mvdebt, saleat, size, incgrowth, and cfodebt. This is rational and as our expectations. For example, the correlation coefficient between size and bankrupt_period is negative, $-0.026$, which indicates that big firms are less likely to be delisted due to financial reasons than small firms. In addition, the correlation coefficient between incgrowth and bankrupt_period is $-0.048$. This means that if income growth increases then bankrupt_period tends to decrease, i.e. likely to take value 0 rather than 1. The figure also shows that the correlation coefficients between ltat, OENEG, and INTWO and bankrupt_period are positive, which make sense too. For example, the correlation coefficient between ltat and bankrupt_period is positive, 0.24, which indicates that firms with more debt relatively to total assets tend to be delisted due to financial reasons than those with small debt to total assets.

### VIF test

This subsection will check the multicollinearity problem among independent variables in Table 3.6. Table 3.10 illustrates the VIF's values of all independent variables of the Altman's, Ohlson's, and combination variables. As the empirical rule, if VIF is greater than 10, then we will remove those variables.

Table 3.10 shows that all VIF's values are less than 10 except ltat in the list of Oblson's variables. We decided to keep ltat due to two reasons. First, the VIF's value of this variable is 10.1650, which is just slightly greater than 10. Second, more importantly, we decide to keep this variable due to the completeness of the list of variables as in Ohlson (1980). Note that the multicollinearity issue is only a concern in logistic regression and LDA[8]. This is not a problem in other models, including SVM, neural networks, and the Merton model.

Table 3.10: The VIF's values

| Altman's variables | | Ohlson's variables | | Combination | |
|---|---|---|---|---|---|
| Variable | VIF | Variable | VIF | Variable | VIF |
| `ebitat` | 2.9395 | `ltat` | 10.1729 | `ltat` | 9.1657 |
| `saleat` | 2.1885 | `lctact` | 9.7887 | `size` | 9.0075 |
| `wcapat` | 1.8831 | `size` | 9.6289 | `roa` | 3.6757 |
| `reat` | 1.7289 | `wcapat` | 3.0244 | `saleat` | 3.0947 |
| `mvdebt` | 1.3321 | `roa` | 2.5652 | `reat` | 2.7760 |
| | | `cfodebt` | 1.4525 | `cfodebt` | 2.4879 |
| | | `OENEG` | 1.4449 | `mvdebt` | 2.4458 |
| | | `INTWO` | 1.4092 | `wcapat` | 2.1995 |
| | | `incgrowth` | 1.2177 | `OENEG` | 1.5290 |
| | | | | `INTWO` | 1.4633 |
| | | | | `incgrowth` | 1.2174 |

### Scaling

Scaling variables before applying the models are very crucial in the cases of SVM and neural networks (Hsu et al. 2016). Since we want to compare the predictive results among all models, we normalize all features/independent variables for all models (except the Merton model since its inputs are different). In Python, we use the package `preprocessing.normalize` from scikit-learn to normalize the variables[9].

## 3.3.2 For the Merton model

### Data collecting

In the Refinitiv Eikon database, the market value of firms on the Vietnam's stock market is available from May 2009. Due to the missing data from January to April of 2009, our study period started from 2010 to 2021 for the Merton model.

[8] Note that the effects of multicollinearity issue on linear regression and LDA are not the same. In LDA, the goal is to find the linear combination of predictor variables (features) that maximizes the separation between different groups or classes. Multicollinearity can still be an issue if there is too much overlap between the predictor variables, but it is less of an issue than in linear regression because LDA is focused on the differences between groups rather than the individual effects of predictors.

[9] For more details, see the documentation at `https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.normalize.html`

### Data processing

For the Merton model, the variables include the number of common shares, market value, debt, and risk-free interest rate. To represent the risk-free interest rate, we use the Vietnam 1-year bond yield[10]. The remaining variables are collected from the Refinitiv Eikon database. In the Merton model, bankruptcy probability is calculated daily and then averaged to get the yearly data. The missing values and the firms which have less than 200 trading days are excluded from the data.

Note that the Merton model works with daily datasets. Thus, we convert the debt, risk-free interest rates into the daily format. For the missing data in the market value of firms, we replace by the stock price multiplied by the common shares. In case the stock prices are also missing, the rolling mean of five periods is used to fill in the market value. A 5-period rolling mean is used with the meaning that today's market value is related to the market value within five trading days of the week.

### The variables

For the Merton model, the variables include (1) market value, (2) short-term and long-term debt, and (3) risk-free interest rate. Note that the market value of each firm is calculated by multiplying the number of common shares (also known as outstanding shares or shares outstanding) and the price of the stock of that firm. In order to represent for a risk-free interest rate, we use the Vietnam 1-year bond yield data. The remaining variables are collected from the Refinitive Eikon database. Table 3.11 shows all variables and the corresponding data sources for the Merton model.

Table 3.11: The Merton's variables and corresponding sources

| Variable | Source |
|---|---|
| Market value of equity | Refinitive Eikon database |
| Debt | Refinitive Eikon database |
| Vietnam 1-year bond yield | www.investing.com |

Notably, the Merton model bankruptcy probability is calculated daily and then converted to yearly data by mean. The missing data and the firms which have less than 200 trading days will be excluded from the data. There are 7,707 firm-year observations with 1,059 firms including 76 delisted firms and 983 non-delisted firms.

[10]See https://www.investing.com/rates-bonds/vietnam-1-year-bond-yield-historical-data

### 3.3.3 Synthetic minority oversampling technique

In order to get balanced data for training models we use the synthetic minority oversampling technique (SMOTE)[11]. This is a data augmentation technique for oversampling the minority group. This technique was first introduced by Chawla et al. (2002). The idea of this technique is that first a random data point $a$ of the minority class is chosen. Then one finds a $k$ nearest data points (e.g. $k = 5$) of minority class. A randomly selected data point $b$ is chosen among $k$ nearest data points. Thus, a synthetic data point is generated at a randomly selected point between $a$ and $b$.

There are two steps for using SMOTE in our study.

- Step 1: Split the data into train (80%) and test set (20%).
- Step 2: Apply SMOTE on the train set.

In Step 1, we split data into 80% of the train set and 20% of the test set. After this step, there are 10,148 firm-year observations (with 414 observations of delisted firms) in the train set and 2,537 firm-year observations (with 117 observations of delisted firms) in the test set.

In Step 2, we only apply SMOTE on the train set to get the balanced training data between non-delisted and delisted firms. After this step, our train set includes $9,734$ firm-year observations of non-delisted firms (50%) and $9,734$ firm-year observations of delisted firms (50%). The total observations in the train set increased from 10,148 to 19,468 after using SMOTE. Note that we keep the test set separate and only use it to validate the models. Notably, SMOTE technique is applied only for the accounting-based and machine-learning models but not for the Merton model. Since the Merton model is calculated independently for each firm, it is not impacted by imbalanced data.

Table 3.12 illustrates the number of observations of train data (Train), test data (Test), and their corresponding proportions before and after using SMOTE of delisted and non-delisted firms. Obviously, for the train data, the data before SMOTE (original data) is highly imbalanced with 4.08% number of observations of delisted firms and 95.92% number of observations of non-delisted firms. After using SMOTE, we get 50% observations of delisted firms and 50% observations of non-delisted firms in the train data. Meanwhile, the test data is the same before and after using SMOTE since this technique only applies on the train data.

[11] Note that one can use other ways to deal with imbalanced data. One way is to balance the weights of the observations in two groups by adding `class_weight = 'balanced'` into the models. While this parameter now is available in logistic regression, SVM, and neural networks (one can calculate the weights from the number of observations in each group and then add these manual weights to the neural networks), it is not available for LDA and the Merton model.

Table 3.12: Number of observations before and after using SMOTE

| | Before SMOTE | | | | After SMOTE | | | |
|---|---|---|---|---|---|---|---|---|
| | Train | Percent | Test | Percent | Train | Percent | Test | Percent |
| Delisted firms | 414 | 4.08% | 117 | 4.61% | 9,734 | 50% | 117 | 4.61% |
| Non-delisted firms | 9,734 | 95.92% | 2,420 | 95.39% | 9,734 | 50% | 2,420 | 95.39% |
| All firms | 10,148 | 100% | 2,537 | 100% | 19,468 | 100% | 2,537 | 100% |

## 3.4 Results

All the results in this section are based on the confusion matrices of relevant models. For more details about how to calculate accuracy, balanced accuracy, Matthew correlation coefficient, precision, recall, and $F_1$ score please see Appendix 1.

Note that in order to obtain good results for neural networks in this section we do hyperparameter tuning with `epochs` = 10, 20, 50,100, 200, 300 and `batch size` = 1, 2, 4, 8, 16, 32, 64[12]. We do not run models with epochs more than 300 to avoid overfitting problem. Also, we do not run models with batch size more than 64 since our final sample only contains 531 observations of delisted-firms, which is relatively small compared to the size of datasets in many neural networks models in the literature. We find that `epochs` = 200 and `batch size` = 32 are the optimal ones.

For the architects of the neural networks, we choose `optimizer='adam'` since this is popular and is one of the most powerful optimizers in machine learning and we choose `loss='binary_crossentropy'` since our target feature is binary one and this is also one of the most popular loss functions in machine learning. Note that we choose these two hyperparameters (optimizer adam and loss function binary_crossentropy) for convenience. But, of course, one can do hyperparameter tuning for these parameters to choose the optimal ones in other studies.

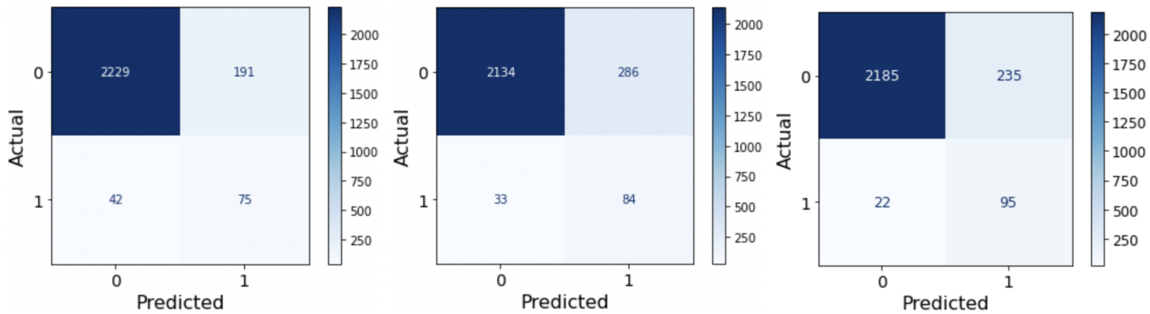
### 3.4.1 Confusion matrix

We denote 0 as non-delisted firms and 1 as delisted firms. Then, we derive the confusion matrix for all models. For example, the confusion matrices of the LDA models for the case of Altman's variables, Ohlson's variables, and the combination are respectively shown Figure 3.5. The top-left numbers mean that the firms are not delisted and the models correctly predict that they are not delisted. The bottom-right numbers mean that the firms are delisted and the models correctly predict that the firms are delisted. The top-right numbers mean that the firms are non-delisted but the models incorrectly predict that

[12]In convention, we choose the `batch size` $= 2^k$, where $k = 0, 1, 2, ....$

the firms are delisted. Finally, the bottom-left numbers mean that the firms are actually delisted but the models incorrectly predict that the firms are non-delisted.

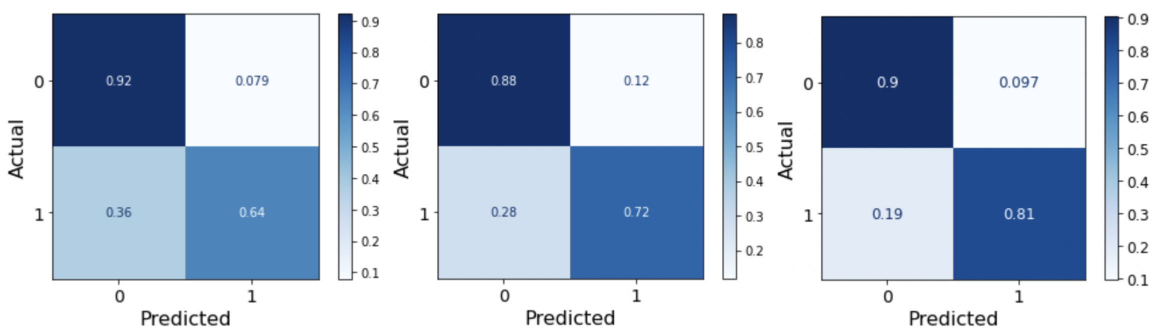
Figure 3.5: Confusion matrices of the LDA models




(a) Altman's variables (b) Ohlson's variables (c) Combination

Since we want to compare the confusion matrices among models (among LDA models and also among the LDA with other accounting-based models), it is useful to normalize the confusion matrices. Figure 3.6 shows the normalized confusion matrices of the LDA models, where we normalize by the sum of the horizontal values in Figure 3.5.

Figure 3.6: Normalized confusion matrices of the LDA models




(a) Altman's variables (b) Ohlson's variables (c) Combination

We also derive the confusion matrix for other models (logistic regression, SVM, neural networks, and the Merton model). Note that for the SVM model, we choose the radial basis function kernel (or RBF kernel). Note also that one can choose other kernels, for example, linear kernel.

The confusion matrix of the Merton model is given in Table 3.13. From this table, we can calculate that the general accuracy is 0.5864, the accuracy for non-delisted firms and delisted firms are 0.5804 and 0.7168, respectively. Note that this general accuracy is also the balanced accuracy of the Merton model. The reason is that in the Merton model, we compute the bankruptcy probability for each firm independently compared to that for the other firms. That is the estimated probability does not depend on the situation of the imbalanced data of the sample size.

Table 3.13: Confusion matrix of the Merton model

| | | Predicted | |
|---|---|---|---|
| | | Non-delisted | Delisted |
| Actual | Non-delisted | 4277 | 3091 |
| | Delisted | 96 | 243 |

### 3.4.2 Balanced accuracy

One important reason that we use balanced accuracy besides accuracy is that machine learning algorithms tend to get biased toward the majority group. That is the algorithms perform well on the majority group and perform poorly on the minority group. In the case of imbalanced data, high accuracy is not necessarily a good score in evaluating models. Because of this reason balanced accuracy might be a better score compared to accuracy in evaluating models with imbalanced data since it captures the performances of both majority and minority groups.

Since our data is imbalanced (skewed distribution), each accuracy in the above section is actually imbalanced accuracy. One can use a balancing technique to generate another accuracy score, which is the so-called balanced accuracy. By definition, it equals the average of the recall of two groups. Similar to accuracy, the values of the balanced accuracy range from 0 to 1.

Table 3.14: Accuracy vs. Balanced accuracy

| Model | Accuracy | | | Balanced accuracy | | |
|---|---|---|---|---|---|---|
| | Altman's | Ohlson's | Combination | Altman's | Ohlson's | Combination |
| LDA | 0.9082 | 0.8715 | 0.8967 | 0.7810 | 0.7984 | 0.8523 |
| LR | 0.9018 | 0.8746 | 0.9011 | 0.7940 | 0.8082 | 0.8790 |
| SVM | 0.9018 | 0.8916 | 0.8975 | 0.8590 | 0.8130 | 0.8771 |
| NN | 0.8849 | 0.8971 | 0.9373 | 0.8949 | 0.8444 | 0.9021 |

*Note*: Altman's stands for Altman's variables and Ohlson's stands for Ohlson's variables.

Table 3.14 shows that all models with both Altman's and Ohlson's variables (the combination case) consistently have the highest balanced accuracy compared to those with only Altman's variables or only Ohlson's variables. Note that we present accuracy in this table just for the purpose of comparison to balanced accuracy. We do not compare or explain the performances of the models in terms of accuracy.

The predictive accuracy of the Merton model is 0.5804[13], which is lower than all given accounting-based models. The poor performance of the Merton model might be because

[13]Note that, similarly to the accounting-based and machine learning models, in order to evaluate the predictive power of the Merton model, we choose 0.5 as the threshold for predicted probability. That is, if

the efficiency of Vietnam's stock market is not clear. Further studies to verify the efficiency of Vietnam's stock market, as well as the predictive performance of the Merton model, are surely necessary.

The next natural step is to separately look at the balanced accuracy in the case of non-delisted and delisted firms in comparison with the case of all firms. It can be seen from Table 3.15 that the balanced accuracy performances for non-delisted firms are higher than those for delisted firms in all models, except in neural networks with Altman's variables.

Table 3.15: Balanced accuracy of the models for non-delisted vs. delisted firms

| Model | All | Non-delisted | Delisted | All | Non-delisted | Delisted | All | Non-delisted | Delisted |
|---|---|---|---|---|---|---|---|---|---|
| | | Altman's | | | Ohlson's | | | Combination | |
| LDA | 0.7810 | 0.9211 | 0.6410 | 0.7984 | 0.8789 | 0.7179 | 0.8523 | 0.9012 | 0.8034 |
| LR | 0.7940 | 0.9128 | 0.6752 | 0.8082 | 0.8814 | 0.7350 | 0.8790 | 0.9033 | 0.8547 |
| SVM | 0.8590 | 0.9062 | 0.8119 | 0.8130 | 0.8996 | 0.7265 | 0.8771 | 0.8996 | 0.8547 |
| NN | 0.8949 | 0.8839 | 0.9059 | 0.8444 | 0.9025 | 0.7863 | 0.9021 | 0.9409 | 0.8632 |

In order to get metrics (e.g accuracy and balanced accuracy) to evaluate neural networks models, we need to set up the architects for neural networks models. In all the neural networks, we use the architects with two hidden layers, one input layer, and one output layer. Table 3.16, Table 3.17, and Table 3.18 respectively show the architect of the neural networks with the Altman's variables, Ohlson's variables, and the combination of both. Note that the number of inputs, which are not the same as the number of neurons in the input layer, of each neural network equals the number of variables in each model. We choose the number of neurons in the input layer that is double the number of neurons in the first hidden layer. Note that one can choose the number of neurons in the input layer differently.

Table 3.16: The architect of the neural networks for Altman's variables

| Layer | Number of neurons | Number of parameters | Activation function |
|---|---|---|---|
| Input layer | 10 | 60 | ReLU |
| Hidden layer 1 | 5 | 55 | ReLU |
| Hidden layer 2 | 5 | 30 | ReLU |
| Output layer | 1 | 6 | Sigmoid |
| Total parameters | | 151 | |

Note that in the above architects for neural networks, we choose ReLU as the activation function in the input and hidden layers. For the output layer, we choose Sigmoid as the activation function. Note also that since the financial distress prediction is a classification problem in nature, one should not use linear as the activation function in the output layer. Linear activation function is suitable when the target variable is continuous while in our case the target variable is binary (distress or non-distress).

the predicted financial distress probability of a firm is greater than 0.5 then we label that firm as a financial distress firm. Otherwise, we label that firm as a non-financial distress firm.

Table 3.17: The architect of the neural networks for Ohlson's variables

| Layer | Number of neurons | Number of parameters | Activation function |
|---|---|---|---|
| Input layer | 18 | 180 | ReLU |
| Hidden layer 1 | 9 | 171 | ReLU |
| Hidden layer 2 | 9 | 90 | ReLU |
| Output layer | 1 | 10 | Sigmoid |
| Total parameters | | 451 | |

Table 3.18: The architect of the neural networks for Altman's and Ohlson's variables

| Layer | Number of neurons | Number of parameters | Activation function |
|---|---|---|---|
| Input layer | 26 | 364 | ReLU |
| Hidden layer 1 | 13 | 351 | ReLU |
| Hidden layer 2 | 13 | 182 | ReLU |
| Output layer | 1 | 14 | Sigmoid |
| Total parameters | | 911 | |

### 3.4.3 Matthews correlation coefficient

Table 3.19 illustrates the Matthews correlation coefficient (MCC) of several models with accounting-based variables. It is clear from Table 3.19 that the models with the combination variables are consistently the best compared to those with either Altman's or Ohlson's variables. Also, the models with Ohlson's variables have the least performance compared to those with either Altman's variables or the combination. Note that MCC's values range from -1 to 1. The higher MCC's values show a higher agreement between the predicted and actual values and vice versa. Note that, in contrast to accuracy, MCC itself is a measure that is less affected by the imbalanced dataset issue. However, we even do not need to worry about this since by using SMOTE our data is balanced already.

Another striking result is that neural networks consistently outperform other models in all of three cases, including Altman's variables, Ohlson's variables, and a combination of the two.

Table 3.19: MCC of the accounting-based and machine learning models

| | Altman's variables | Ohlson's variables | Combination |
|---|---|---|---|
| LDA | 0.3848 | 0.3519 | 0.4377 |
| LR | 0.3876 | 0.3651 | 0.4702 |
| SVM | 0.4524 | 0.3914 | 0.4627 |
| NN | 0.4607 | 0.4306 | 0.5721 |

In order to derive the Table 3.19 we choose 0.5 as the threshold to determine financial distress. Note that we choose 0.5 as the threshold in this study not only because this is the most popular threshold but also since we compared the results for the case of threshold=0.5 and 0.4 (we checked in the code but do not show the comparison here to avoid unnecessary redundancy) and we find that threshold 0.5 generating better results in most models[14]. Note that choosing threshold between 0 and 1 (excluding 0 and 1) only affects neural networks models but not LDA, logistic regression, and SVM since the predicted values in the last three models are either 0 or 1 while that of neural networks can be any real values between 0 and 1.

The MCC for the Merton model is 0.1231, which is lower than any MCC of the accounting-based and machine-learning models. The MCC together with the predictive accuracy consistently shows that the Merton model underperforms any given model, including logistic regression, LDA, SVM, and neural networks. One possible reason can be explained for the poor predictive performances of the Merton model is that Vietnam's stock markets might not be efficient at the current time. Indeed, Loc et al. (2010) finds that Vietnam's stock market is not efficient in the weak form. Moreover, Vietnam's stock market is progressing towards weak form efficiency but the speed of transmission of information is slow (Gupta et al. 2014).

### 3.4.4 Precision, recall and $F_1$ score

Precision, recall, and $F_1$ score are metrics used to evaluate the performance of classification models. Precision is a metric that measures the percentage of correctly identified positive results out of all results that are classified as positive. Recall is a metric that measures the percentage of correctly identified positive results out of all actual positive results. The F1 score is the harmonic mean of precision and recall. For more details, please see Appendix 1.

Table 3.20 compares the precision, recall, and $F_1$ score of the LDA and logistic regression models for three different cases with Altman's variables, Ohlson's variables, and its combination. Note that 0 and 1 in the second column of this table stand for the non-delisted, delisted, and all firms, respectively. Table 3.21 compares the precision, recall, and $F_1$ score of the SVM and neural networks models for three different cases with Altman's variables, Ohlson's variables, and its combination.

One striking result from Table 3.20 and Table 3.21 is that in most cases the predictive performances for non-delisted firms obviously outperform those for delisted firms in terms of precision, recall, $F_1$ score. One possible reason that can explain the differences in predictive performances between delisted and non-delisted firms is that delisted firms might report relatively unreliable financial ratios in their financial statements. The unreliable data, in turn, possibly affect the predictive performances of non-delisted firms. Further studies are needed to test this hypothesis (unreliable data) and possibly other reasons.

[14] The reason why we compare the results with different thresholds is that some studies, for example Staňková (2022), indicates that thresholds should be considered depending on the economic situation and estimation methods.

Table 3.20: Precision, recall, and $F_1$ score of the LDA and logistic regression models

| | | LDA | | | Logistic regression | | |
|---|---|---|---|---|---|---|---|
| | Firms | Precision | Recall | $F_1$ | Precision | Recall | $F_1$ |
| Altman | 0 | 0.98 | 0.92 | 0.95 | 0.98 | 0.91 | 0.95 |
| | 1 | 0.28 | 0.64 | 0.39 | 0.27 | 0.68 | 0.39 |
| | All | 0.63 | 0.78 | 0.67 | 0.63 | 0.79 | 0.67 |
| Ohlson | 0 | 0.98 | 0.88 | 0.93 | 0.99 | 0.88 | 0.93 |
| | 1 | 0.22 | 0.72 | 0.34 | 0.23 | 0.74 | 0.35 |
| | All | 0.60 | 0.80 | 0.63 | 0.61 | 0.81 | 0.64 |
| Combination | 0 | 0.99 | 0.90 | 0.94 | 0.99 | 0.90 | 0.95 |
| | 1 | 0.28 | 0.80 | 0.42 | 0.30 | 0.85 | 0.44 |
| | All | 0.64 | 0.85 | 0.68 | 0.65 | 0.88 | 0.69 |

*Notes*: The precision, recall, and $F_1$ measure of all firms (denoted by All) is the macro average precision, recall, and $F_1$ score. This is actually the average of the precision, recall, and $F_1$ score of both non-delisted and delisted groups, respectively.

Table 3.21: Precision, recall, and $F_1$ score of the SVM and neural networks

| | | SVM | | | Neural networks | | |
|---|---|---|---|---|---|---|---|
| | Firms | Precision | Recall | $F_1$ | Precision | Recall | $F_1$ |
| Altman | 0 | 0.99 | 0.91 | 0.95 | 0.99 | 0.88 | 0.94 |
| | 1 | 0.30 | 0.81 | 0.43 | 0.27 | 0.91 | 0.42 |
| | All | 0.64 | 0.86 | 0.69 | 0.63 | 0.89 | 0.68 |
| Ohlson | 0 | 0.99 | 0.90 | 0.94 | 0.99 | 0.90 | 0.94 |
| | 1 | 0.26 | 0.73 | 0.38 | 0.28 | 0.79 | 0.41 |
| | All | 0.62 | 0.81 | 0.66 | 0.63 | 0.84 | 0.68 |
| Combination | 0 | 0.99 | 0.90 | 0.94 | 0.99 | 0.94 | 0.97 |
| | 1 | 0.29 | 0.85 | 0.43 | 0.41 | 0.86 | 0.56 |
| | All | 0.64 | 0.88 | 0.69 | 0.70 | 0.90 | 0.76 |

### 3.4.5 Feature importance

We already quantified the predictive performances of several models. Now we go further to see which variables are the most important. We consider the coefficients in our logistic regression models as feature importance[15]. Since all variables are normalized, the coefficient values of these variables in the logistic regression models can be considered as feature importance.

Figure 3.7 shows the feature importance for the Altman's variables. The variables with negative coefficient values tend to lead the predicted outcome to be 0 (financially non-distress). Meanwhile, the variables with positive coefficient values tend to lead the predicted

[15] Note that coefficients only can be considered as feature importance in the case of linear models such as linear regression, logistic regression, and their extensions. Since there are no coefficients in some machine learning models (e.g. SVM and neural networks), we do not quantify coefficients as feature importance in these models.

outcome to be 1 (financial distress). All the signs of the coefficient values in this figure are consistent with the correlation coefficients in Figure 3.2, except that of two variables: ebitat and wcapat. One possible reason why the coefficient values of ebitat and wcapat are negative is because the minimum values of ebitat and wcapat in Table 3.9 are negative and the number of observations with negative ebitat and wcapat might be many, which thus affect the sign of the coefficient values. One striking point from Figure 3.7 is that reat is the most important variable among all Altman's variables.

Figure 3.7: Feature importance for the Altman's variables


Figure 3.8 shows the feature importance for Ohlson's variables. It is evident that the coefficient values of ltat, OENEG, and INTWO are positive while the rest of the coefficient values are negative. The signs of the coefficient values of feature importance in Figure 3.8 are almost consistent with the signs of correlation coefficients in Figure 3.3. The only unexpected sign is the case of lctact. This is possibly caused by the high correlation between lctact and wcapat.

Figure 3.8: Feature importance for the Ohlson's variables

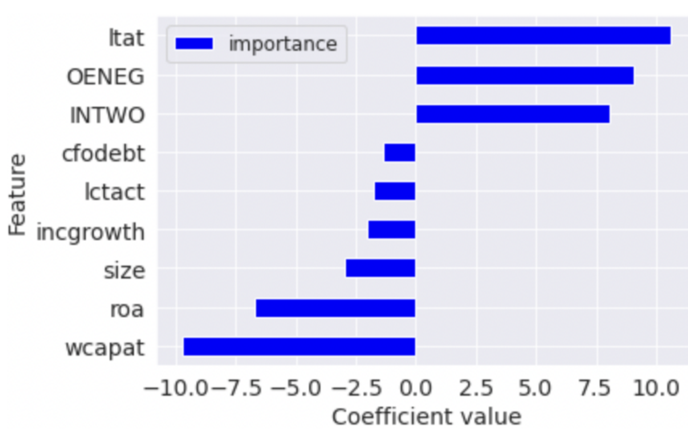


Figure 3.9 shows the feature importance for Altman's and Ohlson's variables. All of the signs of coefficient values in this figure are consistent with the correlation coefficients in Figure 3.4, except that of two variables: size and roa. The negative coefficient value of the variable "size" seems to contradict the fact in Table 3.22. However, this table is just a simple summary statistics. Also, the difference in delisted proportion between small and big firms is small (about 4.7% vs. 3.7%). Further research will be surely needed to quantify the effect of firm size on financial distress probability. One might consider the median of total assets to classify small and big firms rather than the mean to avoid the effects of extreme total assets. For "roa", its positive coefficient value might be because of the fact that the minimum value of roa is negative in Table 3.9 and those negative roa might be many, which thus affects the sign of coefficient value. Overall, strikingly, reat is the most important variable among the list of Altman's and Ohlson's variables.

Figure 3.9: Feature importance for the Altman's and Ohlson's variables


An important result from feature importance is that reat is the most important variable in Altman's variables as well as the combination of Altman's and Ohlson's variables. Also, ltat and wcapat are the most important variables in Ohlson's variables.

## 3.5 The effects of firm size and macroeconomic situation

The outcomes of financial distress prediction might be different depending on firm size and whether firms are in an economic distress situation or not. It is worth exploring the effects of these aspects on predictive outcomes especially in Vietnam, where firms are significantly different in size and where as a transition economy the macroeconomic situations are often unstable.

### 3.5.1 Small firms vs. big firms

Since many public firms in Vietnam are small and medium ones, it is necessary to split all firms into two groups: small and big ones based on the median of the total assets. Small firms are those having total assets less than the median of the total assets of all firms while big firms are those having total assets more than the median of the total assets of all firms. Note that the median of total assets of all firms is approximately 588 billion VND (equivalent to 24.7 million USD).

After splitting, there are 6,344 big firms and 6,341 small firms. Table 3.22 compares the mean and median of variables between small and big firms. The $p$-values of the $t$-test indicate that most of the differences are statistically significant at the significance level of 1%, except INTWO. That is, the means of most variables are statistically different while for INTWO we cannot reject the null hypothesis, which states that the mean of INTWO of small firms equals the corresponding one of big firms. Note that since the means of most variables in Table 3.22 between small and big firms are different, this implies that our list of variables is the good one to potentially capture the differences in predictive performances between small and big firms. Note also that this is not necessarily the case that there are no more good variables that can explain the differences in predictive performances between small and big firms. Possibly, there are some more but we stick with this list of variables since they are the combination of the Altman's and Ohlson's variables.

Table 3.22 shows that the proportion of small firms, that are delisted due to financial reasons, is higher than that of big firms with 4.7% and 3.7%, respectively. This is consistent with the fact from Figure 3.4 that in general the size of a firm has a negative correlation with the likelihood that the firm can be delisted due to financial reasons. The difference in delisted proportion might be because of the difference in income growth. Table 3.22 indicates that the average income growth of the small firms is negative (about $-3.14\%$ per year) while that of big firms is positive (about 2.55% per year). Another possible explanation for the difference in the proportions of delisted firms between small and big firms is the sensitivity to long-term and short-term financial distress. Indeed, Cathcart et al. (2020) indicates that short-term debt is more sensitive to default probability than long-term debt. It can be seen from Table 3.22 that small firms have a proportion of short-term debt over total debt (lctlt) is approximately 86.62% which is moderately higher than that of big firms with the proportion around 72.97%. This difference is consistent with that in Cathcart et al. (2020).

Figure 3.10 compares the performances of the confusion matrices between small and big firms using Altman's variables, Ohlson's variables, and the combination of both. For non-delisted firms, there are no clear different patterns in predictive performances between small and big firms. For delisted firms, the predictive performances between small and big firms are clearer. Specifically, for delisted firms, the predictive performances of big firms consistently outperform those of small firms. Specifically, the proportions of

Table 3.22: Compare means of variables between big and small firms

| | Small firms | | | Big firms | | | |
|---|---|---|---|---|---|---|---|
| Variable | Obs | Mean | Std | Obs | Mean | Std | $p$-value ($t$-test) |
| bankrupt_period | 6,341 | 0.0469 | 0.2116 | 6,344 | 0.0367 | 0.1881 | 0.0038 |
| at | 6,341 | 246.63 | 152.0317 | 6,344 | 3162.42 | 3335.2545 | 0.0000 |
| roa | 6,341 | 0.0463 | 0.0614 | 6,344 | 0.0412 | 0.0492 | 0.0000 |
| reat | 6,341 | 0.0220 | 0.1320 | 6,344 | 0.0464 | 0.0989 | 0.0000 |
| ebitat | 6,341 | 0.0655 | 0.0668 | 6,344 | 0.0660 | 0.0576 | 0.6504 |
| wcapat | 6,341 | 0.2058 | 0.2219 | 6,344 | 0.1316 | 0.1916 | 0.0000 |
| mvdebt | 6,341 | 18.2814 | 36.2832 | 6,344 | 10.8179 | 28.3214 | 0.0000 |
| saleat | 6,341 | 1.1503 | 0.8154 | 6,344 | 0.8469 | 0.7008 | 0.0000 |
| size | 6,341 | 3.6683 | 0.8172 | 6,344 | 6.0240 | 1.0229 | 0.0000 |
| ltat | 6,341 | 0.4737 | 0.2435 | 6,344 | 0.5821 | 0.2130 | 0.0000 |
| lctact | 6,341 | 0.7292 | 0.4690 | 6,344 | 0.8565 | 0.4711 | 0.0000 |
| lctlt | 6,341 | 0.8662 | 0.1955 | 6,344 | 0.7297 | 0.7976 | 0.0000 |
| OENEG | 6,341 | 0.0405 | 0.1972 | 6,344 | 0.0280 | 0.1651 | 0.0001 |
| INTWO | 6,341 | 0.0768 | 0.2662 | 6,344 | 0.0696 | 0.2546 | 0.1233 |
| incgrowth | 6,341 | -0.0314 | 0.5072 | 6,344 | 0.0255 | 0.4749 | 0.0000 |
| cfodebt | 6,341 | 1.6364 | 3.3037 | 6,344 | 0.8445 | 2.3610 | 0.0000 |

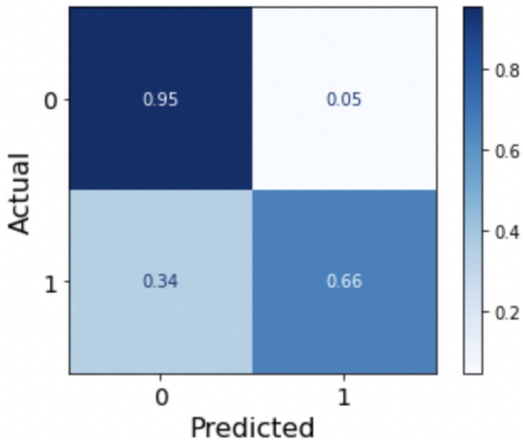
*Notes*: The last column in this table shows the $p$-values of the $t$-test with the null hypothesis that the mean of each variable of the small and big firms are equal. If we choose the significance level at 1%, then we can see that most (except ebitat and INTWO) of the differences in variables between small and big firms are statistically significant. Note that in this table, at and lctlt are not variables in any models. But we show the descriptive statistics of these two variables with the purpose that we might understand the differences between the debt structure of small vs. big firms.

correct predictions of delisted small firms are 66%, 73%, and 75% while the corresponding proportions of correct predictions of delisted big firms are 81%, 94%, and 90%, respectively.

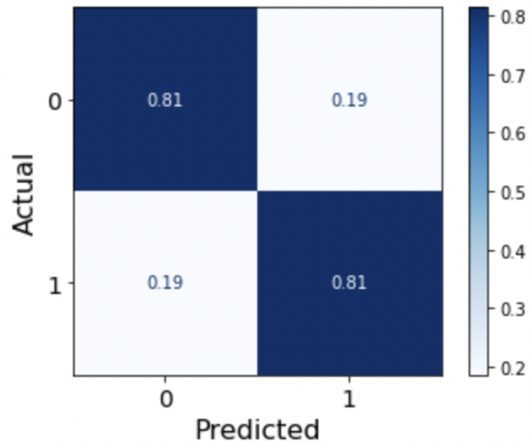
Table 3.23 compares accuracy, balanced accuracy, and Matthews coefficient correlation between small and big firms in the cases of Altman's, Ohlson's, and a combination of variables. In terms of Matthews coefficient correlation, there are no consistent differences in the predictive performances of the models for big vs. small firms. However, for each type of model, the patterns are clearer. For example, for the neural networks models the predictive performances of big firms consistently outperform those of small firms.

In terms of balanced accuracy, the predictive models for big firms are consistently better than those for small firms in most models. An exception is the case for the logistic regression model with Altman's variables. But note that even in this case the balanced accuracy for small and big firms are almost the same with 81.64% and 81.49%, respectively. One possible explanation for the differences in predictive performances between small and big firms is that financial reports and data of the big firms might be more reliable and accurate than those of small firms, thus leading to better predictive performances of big firms compared to small firms.

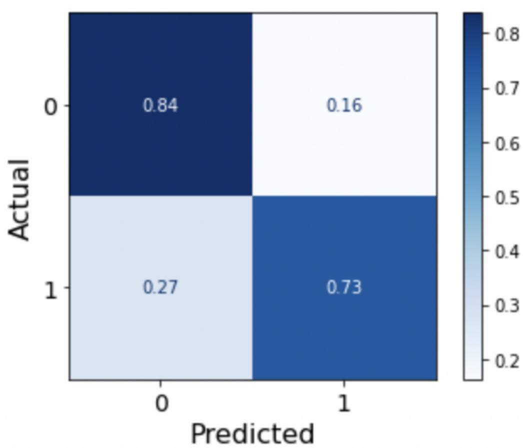
Figure 3.10: Normalized confusion matrices of the LDA models for small and big firms



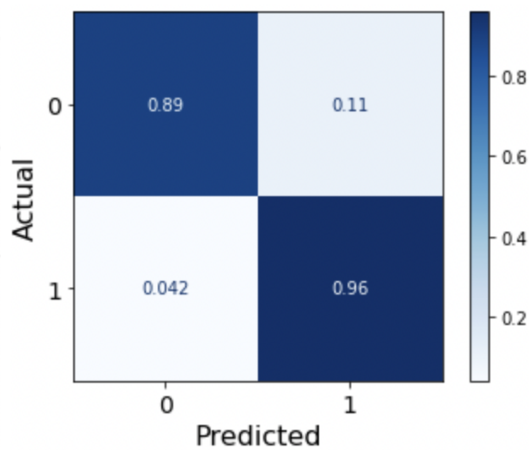

(a) Small firms (Altman's)



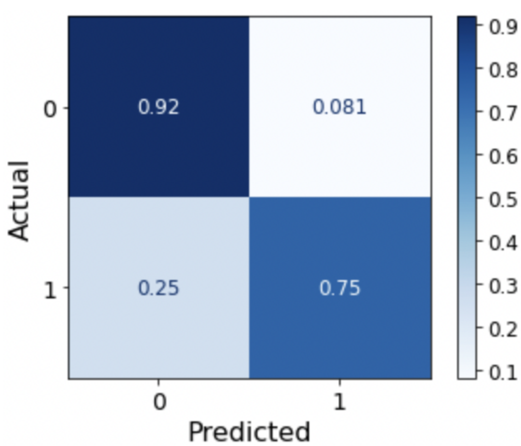

(b) Big firms (Altman's)



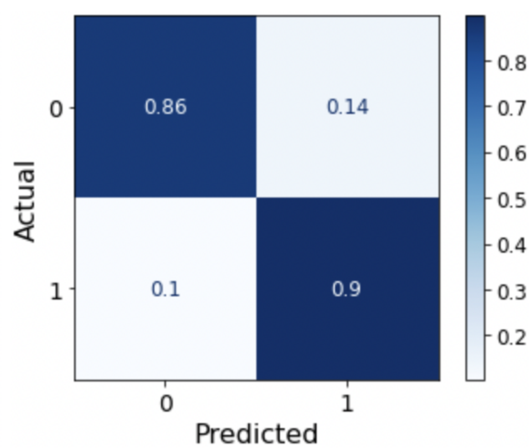

(c) Small firms (Ohlson's)




(d) Big firms (Ohlson's)




(e) Small firms (Combination)




(f) Big firms (Combination)

Table 3.23: Comparison predictive performances between small and big firms

| | | Small firms | | | Big firms | | |
|---|---|---|---|---|---|---|---|
| | | Accuracy | Balanced | MCC | Accuracy | Balanced | MCC |
| LDA | Altman's | 0.9369 | 0.8057 | 0.4800 | 0.8140 | 0.8132 | 0.2936 |
| | Ohlson's | 0.8361 | 0.7851 | 0.3090 | 0.8905 | 0.9131 | 0.4507 |
| | Combination | 0.9117 | 0.8328 | 0.4459 | 0.8597 | 0.8771 | 0.3828 |
| LR | Altman's | 0.9267 | 0.8164 | 0.4624 | 0.8171 | 0.8149 | 0.2968 |
| | Ohlson's | 0.8416 | 0.7557 | 0.2845 | 0.9031 | 0.8796 | 0.4399 |
| | Combination | 0.8968 | 0.8169 | 0.4042 | 0.8692 | 0.8719 | 0.3879 |
| SVM | Altman's | 0.9078 | 0.8146 | 0.4202 | 0.8818 | 0.8585 | 0.3902 |
| | Ohlson's | 0.8503 | 0.7844 | 0.3193 | 0.9086 | 0.9025 | 0.4703 |
| | Combination | 0.9054 | 0.8537 | 0.4548 | 0.8849 | 0.8702 | 0.4048 |
| NN | Altman's | 0.8952 | 0.8725 | 0.4560 | 0.9362 | 0.8868 | 0.5181 |
| | Ohlson's | 0.8810 | 0.8569 | 0.4213 | 0.9251 | 0.9211 | 0.5211 |
| | Combination | 0.9448 | 0.8582 | 0.5573 | 0.9693 | 0.8839 | 0.6551 |

*Notes*: Balanced stands for balanced accuracy and MCC stands for Matthews correlation coefficient.

### 3.5.2 Good years vs. bad years

We choose GDP growth at 6% as the cutoff to classify the study period into two periods: good years and bad years. This is a good cutoff since it captures three economic distress periods including the financial crisis (2008 and 2009), unstable macroeconomic fluctuation (2012, 2013, and 2014), and the Covid-19 pandemic (2020 and 2021) in the bad years. The other years are considered the good years (2010, 2011, 2015, 2016, 2017, 2018, 2019). Note that our delisted data is between 2011 and 2021 but we also have the firm's characteristics data in 2009 and 2010. In total, there are 5,513 observations in the bad years and 7,172 in the good years.

Table 3.24 shows the differences in the mean of variables between good and bad years of some variables. The last column shows the $p$-values of the $t$-test for the null hypothesis that the mean of each variable in bad years equals that in good years. Choose the significance level at 10%, we can observe that more than half of $p$-values are less than 0.1 while the other six $p$-values are greater than 0.1. This implies that for a number of variables including roa, reat, saleat, size, OENEG, INTWO, and incgrowth are statistically significant differences between bad and good years. One striking point in Table 3.24 is that the average income growth of firms during bad years is negative ($-5.41\%$ per year) while that during good years is positive (3.63% per year). It is reasonable to argue that those variables having significant $p$-values of $t$-test (smaller than 0.1) might play more important roles in explaining any differences in predictive performances of firms between bad and good years.

Figure 3.11 compares the confusion matrices of the LDA models for firms in bad and good years. It can be seen that, for delisted firms, the predictive performances for the case of Altman's and Ohlson's variables in the good years are better than the corresponding results

Table 3.24: Comparison means of variables between bad and good years

| | Bad years | | | Good years | | | |
|---|---|---|---|---|---|---|---|
| Variable | Obs | Mean | Std | Obs | Mean | Std | $p$-value ($t$-test) |
| `bankrupt_period` | 5,513 | 0.0417 | 0.1999 | 7,172 | 0.0419 | 0.2005 | 0.9446 |
| `at` | 5,513 | 1694.72 | 2790.70 | 7,172 | 1712.67 | 2762.84 | 0.7180 |
| `roa` | 5,513 | 0.0428 | 0.0569 | 7,172 | 0.0445 | 0.0547 | 0.0735 |
| `reat` | 5,513 | 0.0316 | 0.1197 | 7,172 | 0.0362 | 0.1153 | 0.0271 |
| `ebitat` | 5,513 | 0.0643 | 0.0639 | 7,172 | 0.0668 | 0.0611 | 0.0264 |
| `wcapat` | 5,513 | 0.1715 | 0.2113 | 7,172 | 0.1665 | 0.2100 | 0.1818 |
| `mvdebt` | 5,513 | 14.7565 | 32.8456 | 7,172 | 14.3891 | 32.6915 | 0.5312 |
| `saleat` | 5,513 | 0.9839 | 0.7790 | 7,172 | 1.0099 | 0.7722 | 0.0619 |
| `size` | 5,513 | 5.0956 | 1.5414 | 7,172 | 4.6549 | 1.4350 | 0.0000 |
| `ltat` | 5,513 | 0.5278 | 0.2352 | 7,172 | 0.5280 | 0.2350 | 0.9563 |
| `lctact` | 5,513 | 0.7917 | 0.4798 | 7,172 | 0.7938 | 0.4701 | 0.7998 |
| `OENEG` | 5,513 | 0.0373 | 0.1896 | 7,172 | 0.0319 | 0.1758 | 0.0953 |
| `INTWO` | 5,513 | 0.0440 | 0.2052 | 7,172 | 0.0956 | 0.2941 | 0.0000 |
| `incgrowth` | 5,513 | -0.0541 | 0.5287 | 7,172 | 0.0363 | 0.4581 | 0.0000 |
| `cfodebt` | 5,513 | 1.2642 | 2.9195 | 7,172 | 1.2220 | 2.8819 | 0.4160 |

*Notes*: Bad years indicate the periods with economic distress while good years indicate the periods without economic distress. Note that total assets (denoted by at) is not official variables in any models but we want to do $t$-test to see whether there are significant differences between total assets during bad years compared to good years or not (the $p$-value shows that there is no difference).

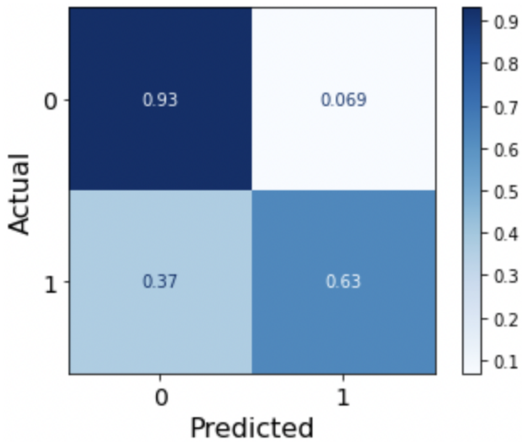
for the bad years. However, interestingly, the predictive performance in bad years is better than that in good years for the combination of Altman's and Ohlson's variables. This implies the role of variables in influencing the predictive ability of the models.

Table 3.25 compares the financial distress prediction performances for bad and good years. In terms of MCC, the predictive performances of most models in the good years are better than those in the bad years. In terms of balanced accuracy, the differences between the predictive performances of LDA and logistic regression models do not have clear patterns. However, for SVM and neural network models, the predictive performances in good years consistently outperform those in bad years. One possible explanation for the outperforms in good years compared to bad years is that during the good years the data is less noisy compared to that in the bad years. Therefore, the predictive performances can be improved.

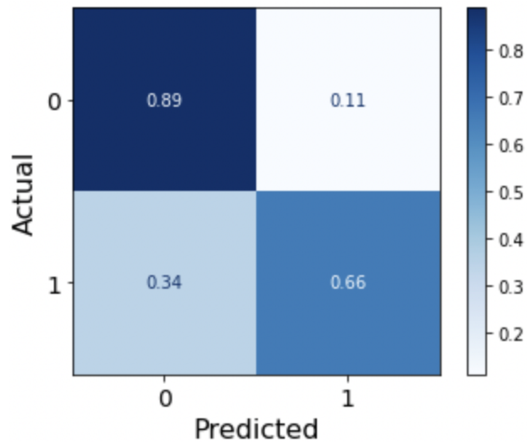
## 3.6 Discussion

In our study, we choose 0.5 as the threshold to predict bankruptcy probability after comparing the predictive performances between the threshold of 0.4 and 0.5. One can choose another threshold to evaluate the performances of neural networks and LDA models. Staňková (2022) argues that although the variables in traditional models are successful in predicting bankruptcy risk, the threshold may not necessarily be the same when carrying financial distress or bankruptcy prediction in different economies. Note that in our study,

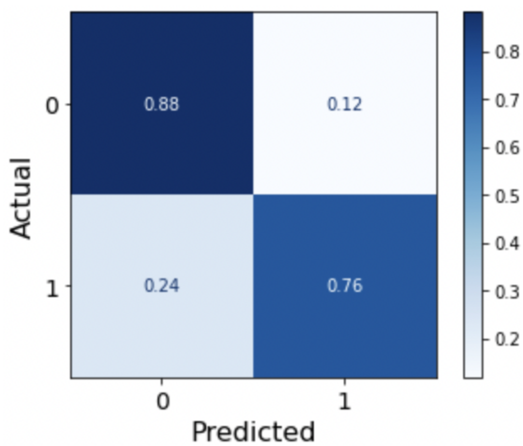
Figure 3.11: Normalized confusion matrices of the LDA models for bad and good years



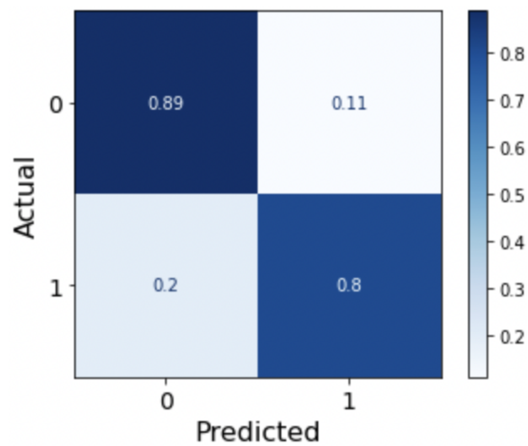

(a) Bad years (Altman's)



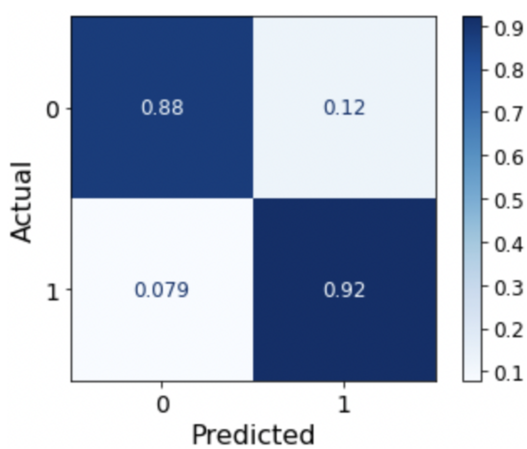

(b) Good years (Altman's)



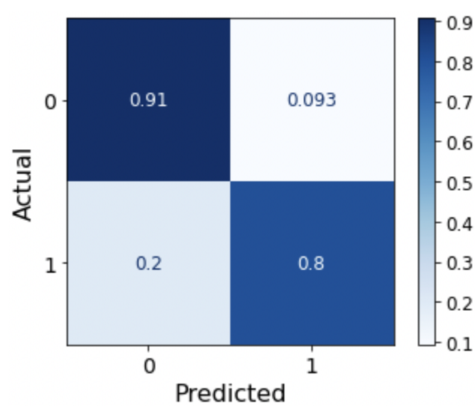

(c) Bad years (Ohlson's)




(d) Good years (Ohlson's)




(e) Bad years (Combination)




(f) Good years (Combination)

Table 3.25: Comparison predictive performances for bad and good years

| | | Bad years | | | Good years | | |
|---|---|---|---|---|---|---|---|
| | | Accuracy | Balanced | MCC | Accuracy | Balanced | MCC |
| LDA | Altman's | 0.9211 | 0.7815 | 0.3626 | 0.8773 | 0.7714 | 0.3196 |
| | Ohlson's | 0.8731 | 0.7947 | 0.3081 | 0.8850 | 0.8459 | 0.4014 |
| | Combination | 0.8766 | 0.8981 | 0.4041 | 0.9003 | 0.8539 | 0.4312 |
| LR | Altman's | 0.9048 | 0.8238 | 0.3752 | 0.8940 | 0.8115 | 0.3804 |
| | Ohlson's | 0.8594 | 0.7749 | 0.2785 | 0.8822 | 0.8132 | 0.3664 |
| | Combination | 0.8758 | 0.8849 | 0.3917 | 0.9108 | 0.8673 | 0.4621 |
| SVM | Altman's | 0.9266 | 0.8478 | 0.4398 | 0.8913 | 0.8727 | 0.4351 |
| | Ohlson's | 0.8867 | 0.7764 | 0.3057 | 0.8934 | 0.8268 | 0.3944 |
| | Combination | 0.8867 | 0.8779 | 0.3988 | 0.9080 | 0.8815 | 0.4704 |
| NN | Altman's | 0.8912 | 0.8548 | 0.3842 | 0.9108 | 0.8986 | 0.4915 |
| | Ohlson's | 0.8948 | 0.8821 | 0.4137 | 0.9192 | 0.9108 | 0.5196 |
| | Combination | 0.9411 | 0.8553 | 0.4859 | 0.9652 | 0.8878 | 0.6557 |

*Notes*: Balanced stands for balanced accuracy and MCC stands for Matthews correlation coefficient.

only neural networks and LDA models are affected by the decision of choosing a threshold between 0 and 1 (excluding 0 and 1). The predictive performances of other models, including logistic regression, LDA, and SVM do not depend on the threshold since their predictive output is either 0 or 1.

We agree with both Abinzano et al. (2020) and Cathcart et al. (2020) that the structure of the sample has an important role on the accuracy of the predictive models. In particular, the firms' size is related to the prediction ability of the model (Abinzano et al. 2020). Also, the predictive performance of models is impacted by the debt structure of firms (Cathcart et al. 2020).

Regarding variables for the prediction models, our results on feature importance not only confirm the significant predictability of variables used in Altman (1968) and Ohlson (1980) but also support Chen et al. (2006) assertion that variables such as reat, ltat, and wcapat provide an overall picture of a firm's performance in terms of profitability, solvency, and liquidity. Indeed, feature importance results provide a general view of a firm's performance in terms of profitability (reat), solvency (ltat), and liquidity (wcapat). Specifically, a company with low profitability, high solvency, and illiquidity indicates early signals of financial distress.

Regarding neural networks, this method normally works well with large datasets, which might not be met in many financial distress and bankruptcy prediction studies due to the limited number of observations of the minority group (i.e. financial distress firms or bankrupt firms). If one can somehow increase the number of observations in the minority group, then the predictive performances of the neural network models might be better. Although with relatively small datasets, we find that neural networks might be more suitable than traditional statistical methods (e.g. logistic regression) in predicting corporate financial distress. Our finding is consistent with Chen and Du (2009).

One important point is that most studies about financial distress or bankruptcy prediction in transition economies (not only in Vietnam) work on imbalanced data but they do not mention in detail their strategies to take into account their imbalanced data. As a result, their results and conclusions might be misleading. For example, Vo et al. (2019) work on two samples, one with about 87% of non-financial distress firms (and 13% of distress firms) and another one with about 73% non-financial distress firms (and 27% financial distress firms). In case they do not take imbalanced data into account their results about accuracy might not be accurate. This is because the models with imbalanced data tend to be biased toward the majority group. In our study, we use balanced accuracy rather than accuracy. Specifically, we assign the weight to one incorrect prediction of the minority group is higher than the weight of one incorrect prediction of the majority group. By doing so, our results are not biased due to imbalanced data.

Note that there are two types of comparisons among models in this study. One is a comparison among the models with the same accounting-based variables (features). The other one is a comparison between models with different input variables. Regarding the comparison between the Merton model and accounting-based models (i.e. logistic regression and machine learning models), the comparability between the results of these models should be considered with caution. One reason is that these models use really different input variables: one base frequent market values while the other ones base on periodic book values. Note that one key variable in the Merton model in our study is a 1-year interest of Vietnam's government bond, which sometimes remains constant for days. Also, the efficiency of Vietnam's stock markets is still not clear at the moment. If the market is inefficient, the Merton model might not predict financial distress well.

## 3.7 Conclusion

This paper compares the performances of the accounting-based, market-based, and machine learning models in predicting corporate financial distress in a transition economy, Vietnam. The results suggest that while all models perform reasonably well in predicting financial distress outcomes for non-delisted companies, their performance is poor when it comes to predicting outcomes for delisted companies, as measured by various accuracy metrics like balanced accuracy, precision, recall, and F1 score. The study shows that models that combine Altman's and Ohlson's variables outperform those that use only one of these variables when it comes to balanced accuracy. Moreover, the study finds that neural networks are consistently the most effective models, as measured by both balanced accuracy and Matthews correlation coefficient (MCC). The variable "reat" (retained earnings over total assets) is the most important variable in Altman's variables as well as in the combination of Altman's and Ohlson's variables, while "ltat" (total liabilities over total assets) and "wcapat" (working capital over total assets) are the most important variables in Ohlson's variables. The study also reveals that the models generally perform better in predicting financial distress outcomes for large companies than for small ones, and their predictive performance is typically better in good years than in bad years, as measured by MCC.

One limitation of this study is that we work on the given list of independent variables (Altman's variables, Ohlson's variables, and the combination). Future research might extend this work by working on a list of variables close to the context of transition economies instead of that base on a developed economy like the United States with Altman's variables and Ohlson's variables in 1968 and 1980, respectively.

Since we only consider the 80/20 split ratio of the train and test set, one potential direction for future research is to examine other split ratios. Also, one can study good years and bad years as a dummy variable to investigate whether this dummy variable affects the predictive performances. Another direction is to compare the predictive results in different sectors or different stock exchanges, including Ho Chi Minh City Stock Exchange (HOSE) and Hanoi Stock Exchange (HNX).

For the Merton model, as we argue in the results section since there is a lack of studies on the efficiency of Vietnam's stock market and the Merton model, further research on both topics is surely needed. Note that in order for the Merton model to work well, the stock market needs to be efficient, including one of weak-form, semi-strong-form, or strong-form efficiency in the sense of Fama (1970). Also, future research needs to verify the assumption of standard normal distribution of some components in the Merton model before carrying it with data. Once the standard distribution assumption and the efficiency of the market are satisfied, the Merton model might be a good model to predict corporate financial distress in a transition economy like Vietnam.

# Appendices

## Appendix 1: Confusion matrix, Matthews correlation coefficient, accuracy, balanced accuracy, recall (or sensitivity), and specificity

Given the following general confusion matrix as follows.

| | | Predicted | | |
|---|---|---|---|---|
| | | Positive (P) | Negative (N) | Total |
| Actual | Positive (P) | True Positive (TP) | False Negative (FN) | TP + FN |
| | Negative (N) | False Positive (FP) | True Negative (TN) | FP + TN |
| | Total | TP + FP | FN + TN | P + N |

The formulas of the Matthews correlation coefficient (MCC), accuracy, sensitivity, and specificity are derived based on the above confusion matrix.

- The Matthews correlation coefficient (MCC)

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}}$$

MCC has values between $-1$ and 1, i.e.

$$-1 \leq \text{MCC} \leq 1.$$

- Accuracy (ACC)

$$\text{ACC} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}.$$

ACC has values between 0 and 1 (or 0% and 100% percent), i.e.

$$0\% \leq \text{ACC} \leq 100\%.$$

- Balanced accuracy for imbalanced data

$$\text{Balanced accuracy} = \frac{\text{Recall (majority group)} + \text{Recall (minority group)}}{2},$$

where the recall of each group is calculated from the confusion matrix of corresponding each group as follows

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}.$$

The values of recall vary from 0 for no recall and 1 for perfect recall. Another score that goes side by side with recall is precision.

• Precision

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

• F1-score: A popular measure based on recall and precision is F1-score (aka F-score or F-measure). The formula of F1-score is given by

$$F_1 = 2\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}.$$

• Sensitivity (or true positive rate - TPR)

$$\text{TPR} = \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

TPR has values between 0 and 1 (or 0% and 100% percent), i.e.

$$0\% \leq \text{TPR} \leq 100\%.$$

Note that the formulas for sensitivity and recall are exactly the same. In that sense, recall and sensitivity are the same.

• Specificity (or true negative rate - TNR)

$$\text{TNR} = \frac{\text{TN}}{\text{N}} = \frac{\text{TN}}{\text{TN} + \text{FP}}.$$

TNR has values between 0 and 1 (or 0% and 100% percent), i.e.

$$0\% \leq \text{TNR} \leq 100\%.$$

Some of the above criteria can be used to evaluate the ability of predictions of the Merton model approach as well as Altman model approach (Z-score).

## Appendix 2: The list of delisted stocks on HOSE and HNX

Table 3.26: Delisted stocks on HOSE & HNX due to financial reasons between 2011-2021

| Symbol | Delisted year | Symbol | Delisted year | Symbol | Delisted year |
|---|---|---|---|---|---|
| FPC | 2011 | HHL | 2014 | CTN | 2016 |
| SHN | 2011 | ILC | 2014 | VC5 | 2016 |
| VTA | 2011 | MIC | 2014 | SRB | 2016 |
| DVD | 2011 | MMC | 2014 | DAC | 2016 |
| VSP | 2012 | NSN | 2014 | SQC | 2016 |
| AGC | 2012 | NVC | 2014 | CYC | 2017 |
| BAS | 2012 | PSG | 2014 | VNA | 2017 |
| CAD | 2012 | PVA | 2014 | VNH | 2017 |
| VKP | 2012 | SDB | 2014 | CPI | 2017 |
| CSG | 2012 | SJM | 2014 | STT | 2018 |
| SBS | 2013 | VCV | 2014 | ICF | 2019 |
| DDM | 2013 | VHH | 2014 | PPI | 2019 |
| IFS | 2013 | YBC | 2014 | VHG | 2019 |
| VES | 2013 | HLA | 2015 | ORS | 2019 |
| VSG | 2013 | HSI | 2015 | SDE | 2019 |
| FBT | 2013 | VNI | 2015 | SCJ | 2019 |
| STL | 2013 | NVN | 2015 | DLR | 2019 |
| THV | 2013 | DCT | 2015 | PCN | 2019 |
| TLC | 2013 | VST | 2015 | PVV | 2019 |
| SVS | 2013 | SSG | 2015 | VPK | 2020 |
| VCH | 2013 | BVG | 2015 | VCR | 2020 |
| SDJ | 2013 | V15 | 2015 | CT6 | 2020 |
| SHC | 2013 | VPC | 2015 | SCL | 2020 |
| SD8 | 2013 | CTM | 2015 | SPP | 2020 |
| S27 | 2013 | LM3 | 2015 | MEC | 2020 |
| SCC | 2013 | PFL | 2015 | PVE | 2020 |
| DHI | 2013 | VNN | 2015 | DID | 2020 |
| AVS | 2013 | TSM | 2015 | NGC | 2021 |
| STL | 2013 | SD1 | 2015 | S74 | 2021 |
| CLP | 2014 | PID | 2015 | ATG | 2021 |
| CNT | 2014 | BTH | 2015 | LO5 | 2021 |
| FDG | 2014 | VLF | 2016 | VTS | 2021 |
| PXM | 2014 | GTT | 2016 | HLY | 2021 |
| BHC | 2014 | PXL | 2016 | CLG | 2021 |
| BHV | 2014 | CID | 2016 | PXT | 2021 |
| GGG | 2014 | S12 | 2016 | DPS | 2021 |

*Source*: HOSE, HNX

# Chapter 4
# Finite and mean field games for optimal investment with HARA utility function and the presence of risk-seeking agents


## Abstract

This study extends the work of Lacker and Zariphopoulou (2019) by considering the financial market with the presence of both risk-averse and risk-seeking agents. Specifically, the $n$-agent (finite) and mean field games for optimal investment with the family of the hyperbolic absolute risk aversion (HARA) utility function under relative performance concern/motivation are studied. Several specific forms of the HARA family, including exponential, power, and logarithmic form are investigated. We prove that there exists a unique constant Nash equilibrium and a unique constant mean field equilibrium in both the $n$-agent and mean field games for the case of strictly concave utility function. For the case of strictly convex utility function, there exists a unique corner solution in these games where agents invest all of their wealth in risky assets (e.g. stock) and invest nothing on riskless assets (e.g. bond). Furthermore, we discuss the qualitative effects of the personal and market coefficients on the optimal investment strategies.



**Keywords**: $n$-agent games, mean field games, optimal investment, risk-seeking agents.

**JEL Codes**: C70, D81, G11.


---


I would like to thank Ruben Juarez, Miroslav Gabrovski, Yuanzhang Xiao, Huy Chau, Nguyen Le Hoang Anh, Malik Younsi, Berkan Yilmaz, Evan Gawlik, Daisuke Takagi, Ekaterina Sherstyuk, Yukihiko Funaki, Anh Tran, Diu Tran, Duc Thach Son Vu, Van Thanh Nguyen, Kang Zheng, Adib Rahman, Chuong Mai, and participants at the Game Theory Seminar and the Applied Math Seminar at the University of Hawaii at Manoa, the 6th PKU-NUS Annual International Conference on Quantitative Finance and Economics, and the 33rd Stony Brook International Conference on Game Theory for helpful comments and discussions. All errors are mine.

## 4.1 Introduction

The most popular assumption in the relevant literature of optimal investment and portfolio optimization usually supposes that agents are uniform risk-averse or their utility functions are concave on the entire domain. The second popular assumption lying in the prospect theory is the S-shape utility function, which is convex in the domain of losses and concave in the domain of gains, and the slope in the domain of losses is steeper than that in the domain of gains. Instead of examining the variation of risk preferences *within* agents like the prospect theory or studying uniform risk-averse behavior like the mainstream economic literature, this study investigates risk preferences which may vary *between* agents depending on their characteristics. Meanwhile, we know a lot about the behavior of risk-averse agents who have uniformly concave or S-shape utility functions, there is a lack of understanding of the behavior of risk-seeking agents and the interaction between risk-averse and risk-seeking agents. In contrast to the prospect theory, this study assumes that the utility functional form is a characteristic of each individual rather than depends on whether she gains or losses. This implies that the individual risk preferences are stable over time and not easy to change during a short period of time.

### 4.1.1 Motivation: The presence of risk-seeking agents

Risk-seeking behavior of agents (e.g. investors) is examined theoretically and empirically by many previous studies in the literature (e.g., Kahneman and Tversky 1979; Wong and Li 1999; Niendorf and Ottaway 2002; Seel and Strack 2013; Bai et al. 2014; Hoang et al. 2015; Guo et al. 2016; Clark et al. 2016; Wu and Jaimungal 2023). Specifically, Seel and Strack (2013) shows that, in a declining industry, risk-loving agents might invest in projects with negative expected returns. In addition, Niendorf and Ottaway (2002) shows that during bear markets, agents have an incentive to act in a risk-seeking manner. Empirically, Bai et al. (2014) find that there exist two types of investors, risk-averse and risk-seeking agents, on the stock market. Moreover, Clark et al. (2016) find the existence of all four investor types corresponding to concave, convex, S-shape, and reverse S-shape on the Taiwan stock and stock index futures markets. Also, they figure out that risk-seeking agents prefer futures markets to spot markets.

In reality, there are many situations where risk-seeking behavior exists. Indeed, we know much about the investing activities on the traditional underlying assets (e.g., stock and bond) on the financial markets; meanwhile, our understanding of risk-seeking investing activities (e.g., playing the lottery, gambling, speculating, sports betting, and short selling) of risk-seeking agents is very limited. It is worth exploring some types of risk-seeking agents and their risk-seeking behavior.

### Lottery players as risk-seeking agents

A lottery is a game involving drawing a number for a prize for which only few buyers win while a much greater number of buyers does not. That means this is risky entertainment. To examine, let's consider the California SuperLotto Plus[1]. To play, a player picks five numbers between 1 and 47 and one Mega number between 1 and 27 and she needs to pay \$1 per play for each SuperLotto Plus ticket. The prize amounts and the game odds (the inverse of the probability of winning a prize) are described in Table 4.1.

Table 4.1: A description of the California SuperLotto Plus lottery

| Matching numbers | Odds 1 in (1/winning probability) | Prize amounts |
|---|---|---|
| All 5 of 5 and Mega | 41,416,353 | \$15,000,000 |
| All 5 of 5 | 1,592,937 | \$33,622 |
| Any 4 of 5 and Mega | 197,221 | \$1,400 |
| Any 4 of 5 | 7,585 | \$89 |
| Any 3 of 5 and Mega | 4,810 | \$52 |
| Any 3 of 5 | 185 | \$9 |
| Any 2 of 5 and Mega | 361 | \$10 |
| Any 1 of 5 and Mega | 74 | \$2 |
| None of 5, only Mega | 49 | \$1 |
| Overall odds of winning | 23 | One (any) prize |

*Source*: Calottery

The probability of winning a prize, for example, the first prize (all 5 of 5 and Mega), is

$$\binom{47}{5}\binom{27}{1} = \frac{1}{41,416,353},$$

and similarly for the probabilities for other prizes. From Table 4.1, the expected payoff of buying a California SuperLotto Plus ticket is

$$\frac{1}{41,416,353}\$15,000,000 + \frac{1}{1,592,937}\$33,622 + ... + \frac{1}{74}\$2 + \frac{1}{49}\$1 \approx \$0.5367 \approx 54 \text{ (cents)}.$$

In general, not only the California SuperLotto Plus ticket that we mention here but it seems almost all (or probably all) real-world private or state lotteries have expected payoffs that are lower than the cost to purchase them.

Why do people play lotteries even though the expected payoff is less than the cost to play? One can use knowledge of behavioral economics to explain by assuming that lottery players

[1] For more details, please see https://www.calottery.com/draw-games/superlotto-plus#section-content-4-3

are irrational. They might overestimate (or overweight) the small probability to win a lottery. Also, they might believe that their subjective confidence in their own ability to win a lottery is greater than the actual probability to win. This is the so-called overconfidence effect. This study, however, assumes that lottery players are rational. They play lotteries because they realize that by doing so they can gain benefits. We assume that lottery players are risk-seeking agents and have convex utility functions. Therefore, even though the expected payoff is negative, the utility they gain might be positive. Moreover, individuals might play lotteries not only because of the monetary value but also the non-monetary value (e.g. pleasure), which is not reflected in the expected payoff but the gaining utility (value). Assume that the lottery players are rational and their purpose is to maximize their utility levels. Thus, if a player prefers the California SuperLotto Plus ticket with the expected payoff is 54 cents than a sure lottery with a payoff \$1 (cost for each California SuperLotto Plus ticket), then she is a risk-seeker.

**Gamblers as risk-seeking agents**

Let's consider *Roulette*, which is a popular casino game. In this game, players choose to place finite bets on either a single integer number (from 1 to 36) or a group of several numbers, which are low (1-18), high (19-36), odd, even, black, or red on a wheel. The number zero might appear once (French/European style roulette) or twice (American style roulette) on the wheel and it is colored in green.

Suppose a European roulette model in which a player chooses the *straight up* (single number) bet strategy with the winning space being any single number. Thus, the probability space is $(\Omega, 2^{\Omega}, \mathbb{P})$, where $\Omega = \{0, 1, ..., 36\}, \mathbb{P}(A) = |A|/37$ for any $A \in 2^{\Omega}$. Define a bet a triple $B = (A, s, p)$, where $A$ is the set of chosen numbers, $s \in^{+}$ is the size of the bet, and $p : \Omega \to \mathbb{R}$ is the payoff function of the bet. Thus, the payoff function of the space $B = (w_0, s, p)$ for some $w_0 \in \Omega$ is defined by

$$p(w) = \begin{cases} -s, & w \neq w_0 \\ 35s, & w = w_0. \end{cases}$$

Thus, the expected payoff is

$$E[p(w)] = \frac{1}{37}(35s) + \frac{36}{37}(-s) = -\frac{s}{37} \approx -0.027s.$$

Suppose $s = \$10$, then $E[p(w)] \approx \$ - 0.27 = -27$ cents. That is, for each time of playing with the cost of \$10, the player loses approximately 27 cents after each game, which is equivalent to 2.7%.

Alternatively, suppose a player chooses even odd bet strategy in a European roulette game. The rule is that if the bet wins then that player gets double, if the bet loses the player gets nothing. Since there are 37 numbers, $\Omega = \{0, 1, \ldots, 36\}$ and the host always win if you choose 0, the probability that a player wins for each game is $p = 18/37 \approx 0.4864$, which is actually less than one-half. Moreover, suppose the size of the bet is $s = \$10$. Thus, the expected payoff for each game equals the sure lost minus expected gain, which is

$$10 - \left[\frac{18}{37}(2 \times 10) + \frac{1}{37}(0)\right] \approx -\$0.27 \quad \text{(or 27 cents)}.$$

To sum up, in either straight up or even odd strategy, the player loses about 27 cents after each game. But why are there still many people playing roulette nowadays?

**Sports betting players as risk-seeking agents**

Sports betting is a gambling-like activity. Football betting, for instance, is now popular with the support of computers and the internet. Table 4.2 shows the betting odds for the top-10 football teams in the 2014 World Cup.

Table 4.2: The 2014 World Cup betting odds for the top-10 teams[2]

| Country | Betting odds | Country | Betting odds |
|---|---|---|---|
| Brazil | 3/1[3] | Italy | 20/1 |
| Argentina | 5/1 | Holland | 25/1 |
| Germany | 5/1 | France | 25/1 |
| Spain | 6/1 | Colombia | 25/1 |
| Belgium | 16/1 | England | 28/1 |

**Short-sellers as risk-seeking agents**

Consider the financial market where short sales are allowed. In reality, approximately 1 of 35 (equivalently about 2.88%) traded stocks on Nasdaq is short selling and the average percentage of shorted shares of 100 most active stocks is 4.17% (Angel et al. 2003). One noticeable result is that short-sellers have preferences for the risky factors (Dechow et al. 2001). We will theoretically show later in this study that those who short stocks are risk-seeking. Short-sellers usually short the stocks that are relatively overvalued relative to its fundamentals or probably the stocks of acquiring firms in the mergers and acquisitions process (Dechow et al. 2001). They may also short the stocks of firms that misrepresent their financial statements (Karpoff and Lou 2010). Short-sellers might borrow an asset, for example, the shares of a stock, from the brokers and sell it and then buy it back at some period in the future. They make a profit when the price drops and make a loss when the price increases. Since the price can only drop to zero, then the profit is bounded.

[2] See more details at https://www.businessinsider.com/world-cup-favorites-2014-4

[3] 3/1 odds means one bet \$1 and get \$4 if Brazil wins, giving her/him a profit of \$3.

Meanwhile, since the price can go up arbitrarily high then the potential loss is unbounded. Therefore, short-selling is a risky activity.

### Speculators as risk-seeking agents

Speculators might utilize leverage to magnify gains (and losses). By using leverage, the potential gains are high but the potential losses are also high. They usually do not hold assets or derivative securities for a long period. Speculators aim at outperforming traditional longer-term investors by making gains that are large enough to offset the high risk. As risk-seeking agents, a common strategy for speculators is to invest in derivatives (e.g. futures). Specifically, Lean et al. (2010) discover in Malaysian markets that, for risk-averse agents, spot dominates futures under second-order stochastic dominance (SSD), while for risk-seeking agents, futures dominate spot under SSD. That is, risk-seeking agents prefer futures while risk-averters prefer spots to maximize their expected utility. Similarly, Qiao et al. (2013) find that, for the emerging markets, futures dominate the spot for risk-seeking agents while the spot dominates futures for risk-averse agents. As a result, there are potential gains for risk-seeking agents if they switch their investment from spot to futures and for risk-averse agents if they switch from futures to spot.

### Entrepreneurs as risk-seeking agents

Entrepreneurs are willing to take many risks, particularly the risk of failure. Indeed, there are about 90% startups failing (Cusumano 2013). This is because of uncertainties and the lack in many aspects of the business at the early state including financing, hiring talents, pricing, knowledge of the market demand, and experience of competition [4].

In summary, Table 4.3 compares this study and standard optimal investment literature, expected utility theory, and prospect theory.

Table 4.3: Comparison the utility functional forms in the literature

| | Standard optimal investment literature | Expected utility theory (EUT) | Prospect theory | This study |
|---|---|---|---|---|
| Utility functional form | concave | concave, convex, or linear | convex-concave (S-shape) | convex, concave |
| Agents | risk-averse | risk-averse, risk-seeking, and risk neutral | risk-seeking and risk-averse | risk-seeking, and risk-averse |
| Variation in risk preferences | NA | between agents | within agents | between agents |

[4] For more details, please see `https://s3-us-west-2.amazonaws.com/cbi-content/research-reports/The-20-Reasons-Startups-Fail.pdf`

### 4.1.2 Mean field games

Mean field games (MFG) are the limiting version (or sometimes called the aggregated version) of the $N$-agent games. This is a kind of stochastic strategic decision game. MFG has been pioneered in Lasry and Lions (2006a,b, 2007a,b) and independently in Huang et al. (2006). MFG has a close relationship with infinite-player games in the game theory literature. Indeed, games with a continuum of infinite players have long been considered in game theory (Shapiro and Shapley 1961; Milnor and Shapley 1961; Shapley 1961; Aumann 1964; Aumann and Shapley 1974; Aumann 1975; Jovanovic and Rosenthal 1988; Myerson 1991; Khan and Sun 2002; Rauh 2003; Huang 2013; Chan and Sircar 2015). In short, Table 4.4 compares the differences between infinite-player games and MFG.

Table 4.4: Comparison between infinite-player games and MFG

| | Infinite-player games | MFG |
|---|---|---|
| Agents | Homogeneous or heterogeneous | Heterogeneous |
| Mean-field parameter | No | Yes |
| Tools for existence proof | Fixed-point theorem (Brouwer or Kakutani) | Optimal control (OC) problem, PDE, SDE |
| Strategies | Pure (discrete) and mixed strategies | Continuous controls (strategies) |
| Setting (model) | Optimization problem or OC problem | OC problem or Coupled system of two PDEs |

### 4.1.3 Brownian motions as stochastic shocks

The price dynamics is described by the following stochastic differential equation, which includes two independent geometric Brownian motions as stochastic shocks: $W_t$ (individual shocks) and $B_t$ (market shocks),

$$\frac{dS_t}{S_t} = \mu dt + \nu dW_t + \sigma dB_t,$$

where $\mu$ is the trend, $\nu$ and $\sigma$ are the individual and market volatility, respectively, and $S_t$ is the price of a particular asset. The market shocks (e.g. pandemic, terrorism attacks, major natural disasters, or financial crisis) and individual shocks (e.g. company restructuring, cost shock, factory fire, or sudden change in dividend policies) can be arbitrarily small or large. We assume that the shocks are random (unpredictable) for typical investors. We also assume that all of the shocks reflect in the price in a short time via two Brownian motions (i.e. the efficient market hypothesis holds). Note that individual shocks mean that each individual investor investing in an individual asset faces different shocks. Even though they face the same market shocks, they universally face different levels of shocks.

### 4.1.4 Relative performance: Individual investors vs. Fund managers

Utility function with relative performance criteria is developed to reflect the human psychology that people normally compare their performance with those of their peers. This kind of difference-like comparison affects people's well-being and behavior. The idea of utility with relative performance is not new in economic literature (Abel 1990; Fehr and Schmidt 1999).

In MFG, typical works assume that fund managers concern about the relative performance of their peers since incentive compensation (e.g. payment incentive, promotion incentive) is popular in the fund managers community (Chevalier and Ellison 1999; Espinosa and Touzi 2015; Huang and Nguyen 2016; Lacker and Zariphopoulou 2019; Fu and Zhou 2020). In contrast with these studies, our study focuses on individual investors rather than fund managers. It is noticeable that, besides incentive compensation, fund managers and individual investors are different in the sense that fund managers are normally involved in agent-principal problems in which the managers' benefits and the principal's benefits might be different. This is not the case for individual investors. As a result, these two groups (individual investors and fund managers) might not share the same investment behavior. Fund managers pay attention to relative performance concern while individual investors might pay attention to relative performance concern as well as relative performance motivation. The differences in the behavior of these two groups of agents are reflected in risk tolerance and relative performance parameter. Table 4.32 compares some recent related work with this study.

Table 4.5: Some studies related to relative performance

| Literature | Espinosa and Touzi (2015) | Lacker and Zariphopoulou (2019) | Fu et al. (2020) | This study |
|---|---|---|---|---|
| Utility function | CARA | CARA, CRRA | Exponential | HARA |
| Agents (or players) | Portfolio managers | Fund managers | Fund managers | Individual investors |
| Risk preference | Risk aversion | Risk aversion | Risk aversion | Risk aversion and risk-seeking |
| Personal risk tolerance | Positive | Positive | Positive | Not equal zero |
| Relative performance ($\theta_i$) | $\theta_i \in [0, 1)$ | $\theta_i \in [0, 1]$ | $\theta_i \in [0, 1)$ | $\theta_i \in [-1, 1]$ |
| Expected value | unclear | positive, unbounded | bounded | non-negative, bounded |
| Method | BSDE | PDE | FBSDE | PDE |
| Well-posedness | Yes | Yes | Yes | Yes |

Why do we assume $\theta_i \in [-1, 1]$? We will explain it based on psychological perspectives. On the one hand, if comparison brings concern or worry or jealousy to agent $i$ (if agent $i$ compares her performance with higher earners, which is called upward social comparison), or arrogance (if agent $i$ compares her performance with lower earners, which is called downward social comparison) then it is not good for her. In this situation, it is described by the so-called *relative performance concern* and denoted by $\theta_i^+ \in [0, 1]$. In an experimental study about the effect of economic inequality on risk behavior, Payne et al. (2017) shows that there is 63% of participants choosing upward comparison while 37% of participants choosing downward comparison. Note that the term "relative performance concern" in this study is equivalent to the term "relative performance" in Espinosa and Touzi (2015) and Lacker and Zariphopoulou (2019).

On the other hand, if comparison brings motivation or inspiration (if upward social comparison), or pride (if downward social comparison) then it is good for agent $i$, which is illustrated by the so-called *relative performance motivation* and denote by $\theta_i^- \in [-1, 0)$. Some previous studies show that peer effects might help improve performance. Eisenkopf (2010) discovers that peers impact performance not just during interaction but also before and afterward. This implies that the performance of an individual might be affected just by comparison with peers. Kaustia and Knüpfer (2012) find that recent stock returns that local peers experience affect an individual's stock market entry decision, especially in areas with better opportunities for social learning.

Briefly, this study aims to characterize the behavior of risk-seeking and risk-averse agents in $n$-agent games and MFG. Specifically, I will tackle the following questions: (i) Are there exist Nash equilibrium (NE) and mean field equilibrium (MFE) in $n$-agent games and MFG with the presence of risk-seeking agents? (ii) If NE and MFE exist, in which conditions are they unique? (iii) How do the NE and MFE depend on the personal and market parameters?

The remainder of the chapter is organized as follows. Section 4.2 will describe a functional form of HARA utility function. We then explore $n$-agent and mean field games for the exponential form in Section 4.3 and the power form in Section 4.4. Hence, we will discuss the effects of personal and market parameters on equilibrium in Section 4.5 and finally conclude the chapter in Section 4.6.

## 4.2 HARA utility function

The family of utility functions called the HARA family is first introduced by Merton (1971) and is studied extensively in the literature of portfolio selection (Merton 1987; Davis and Norman 1990; Kim and Omberg 1996; Duffie et al. 1997; Benth et al. 2001; Çanakoğlu and Özekici 2010; Escobar et al. 2017). Before considering the explicit form of the HARA family for this study, we propose several necessary assumptions.

### 4.2.1 Assumptions

Some necessary assumptions for both $n$-agent games and mean field games are given as follows. These are the common (either implicit or explicit) assumptions of the standard models in mean field games.

**Assumption 4.2.1** (Rationality)**.** *In investing activities, agents are rational.*

**Assumption 4.2.2** (Common knowledge)**.** *The individual preference parameters and market parameters are common knowledge.*

As we will see in the next section, we will investigate models with complete information. Specifically, the personal preferences and market parameters in $n$-agent games and MFG are common knowledge for all agents and each agent knows that each other agent knows a common knowledge.

**Assumption 4.2.3** (Zero transaction costs)**.** *There are no or negligible transaction costs in trading assets.*

**Assumption 4.2.4** (Complete market)**.** *The market is complete, i.e. there exists the price for every asset in any state.*

Note that the Assumption 4.2.1-4.2.4 implies that the financial market in this study is a perfect market. Even though the perfect market might be not realistic in the current uncertain world, "the model may indeed provide the best description of the financial system in the long run" (Merton 1987).

**Assumption 4.2.5** (Heterogeneous agents)**.** *Agents are heterogeneous, i.e. each agent has the same utility functional form and dynamics but with different parameters.*

### 4.2.2 HARA utility functional form

**Definition 4.2.1** (HARA utility function)**.** *A utility function with nonnegative real domain $\mathbb{R}^+$ is said to exhibit the hyperbolic absolute risk aversion (HARA) if and only if the inverse of absolute risk aversion, $r(x) = 1/A(x) = -U'(x)/U''(x)$, is a linear function of wealth (level).*

In this study, the functional form of the family of HARA utility functions with relative performance is given by

$$U_i(X_T^i, \bar{X}_T) = a_i\Big(\frac{1-\alpha}{\alpha}\Big)\Big(\frac{X_T^i * f(\bar{X}_T, \theta_i)}{1-\alpha} - \eta_i\Big)^{\alpha} + b, \quad \alpha \neq 0, \alpha \neq 1 \tag{4.1}$$

where $\alpha, a_i, b, -1 \leq \theta_i \leq 1$, and $\eta_i$ are constant; $*$ is an abstract operation, which can be subtraction or multiplication. Here, $X_T^i$ is the agent $i$'s terminal wealth; $\bar{X}_T$ is the average terminal wealth of all agents including agent $i$. There are two cases: (i) $X_T^i$ and $\bar{X}_T$ have linear relationship (if $*$ is subtraction) and $f(\bar{X}_T, \theta_i)$ describes a linear relationship in terms of wealth, i.e. $f(\bar{X}_T, \theta_i) = \theta_i \bar{X}_T$; (ii) $X_T^i$ and $\bar{X}_T$ have nonlinear relationship (if $*$ is multiplication) and $f(\bar{X}_T, \theta_i)$ describes a nonlinear relationship with wealth with $f(\bar{X}_T, \theta_i) = \bar{X}_T^{-\theta_i}$. We can check that the HARA family of utility function (4.1) satisfies Definition 4.2.1.

If $*$ is subtraction and $f(\bar{X}_T, \theta_i) = \theta_i \bar{X}_T$, we can check that the inverse of absolute risk aversion, $1/A(X_T^i)$, is a linear function of wealth $X_T^i$. Indeed, take the first and second derivative of the utility function (4.1) with respect to $X_T^i$ yields

$$U_i' = a_i \left( \frac{X_T^i - \theta_i \bar{X}_T}{1-\alpha} - \eta_i \right)^{\alpha-1}, \quad U_i'' = -a_i \left( \frac{X_T^i - \theta_i \bar{X}_T}{1-\alpha} - \eta_i \right)^{\alpha-2}.$$

Thus, the linear form of the inverse of absolute risk aversion with respect to $X_T^i$ is given by

$$\frac{1}{A(X_T^i)} = -\frac{U_i'}{U_i''} = \frac{1}{1-\alpha} X_T^i - \left( \frac{\theta_i \bar{X}_T}{1-\alpha} + \eta_i \right).$$

Similarly, if $*$ is multiplication and $f(\bar{X}_T, \theta_i) = \bar{X}_T^{-\theta_i}$, we can derive that

$$U_i' = a_i \left( \frac{X_T^i \bar{X}_T^{-\theta}}{1-\alpha} - \eta_i \right)^{\alpha-1} \bar{X}_T^{-\theta_i}, \quad U_i'' = -a_i \left( \frac{X_T^i \bar{X}_T^{-\theta_i}}{1-\alpha} - \eta_i \right)^{\alpha-2} \left( \bar{X}_T^{-\theta} \right)^2.$$

Thus, obtain the linear expression on the right hand side as follows

$$r(X_T^i) = \frac{1}{A(X_T^i)} = -\frac{U_i'}{U_i''} = \frac{1}{1-\alpha} X_T^i - \frac{\eta_i}{\bar{X}_T^{-\theta_i}}.$$

**Remark 4.2.1.** *By introducing the average terminal wealth $\bar{X}_T$, we can induce the $n$-dimensional $n$-agent games and infinite dimensional MFG to the 2-dimensional optimal control problem. That is, the utility function now only depends on two variables: the agent $i$'s terminal wealth $X_T^i$ and the average terminal wealth of the population $\bar{X}_T$.*

The utility functional form (4.1) is explicitly a function of wealth level (a stock), or wealth for short, rather than a function of return (a flow). We will see in the next sections that the return will be a part of the price dynamics, which then affects the stock of wealth. That is, the return of stocks will indirectly affect the agent $i$'s utility level. We will explain the meaning of the parameters in the next sections. We are going to derive a number of utility functions, which are special cases of (4.1).

**Exponential utility function**. If $*$ is subtraction, and

$$f(\bar{X}_T, \theta_i) = \theta_i \bar{X}_T, \quad \eta_i = \frac{\alpha}{1-\alpha}\frac{1}{\gamma_i}, \quad a_i = \Big(\frac{\alpha-1}{\alpha}\Big)^{\alpha-1}\gamma_i^{\alpha-1}, \quad b = 0$$

then

$$\begin{aligned} U_i(X_T^i, \bar{X}_T) &= \Big(\frac{\alpha-1}{\alpha}\Big)^{\alpha-1}\gamma_i^{\alpha-1}\Big(\frac{1-\alpha}{\alpha}\Big)\Big(\frac{X_T^i - \theta_i\bar{X}_T}{1-\alpha} - \frac{\alpha}{1-\alpha}\frac{1}{\gamma_i}\Big)^{\alpha} \\ &= -\frac{1}{\gamma_i}\Big(\frac{\alpha-1}{\alpha}\Big)^{\alpha}\gamma_i^{\alpha}\Big[\frac{1}{\alpha-1}\Big(\frac{\alpha}{\gamma_i} - (X_T^i - \theta_i\bar{X}_T)\Big)\Big]^{\alpha} \\ &= -\frac{1}{\gamma_i}\Big(1 - \frac{\gamma_i(X_T^i - \theta_i\bar{X}_T)}{\alpha}\Big)^{\alpha}. \end{aligned}$$

Taking limit as $\alpha \to \infty$, we get the exponential functional

$$\begin{aligned} \lim_{\alpha\to\infty} U_i(X_T^i, \bar{X}_T) &= \lim_{\alpha\to\infty} -\frac{1}{\gamma_i}\Big(1 - \frac{\gamma_i(X_T^i - \theta_i\bar{X}_T)}{\alpha}\Big)^{\alpha} \\ &= -\frac{1}{\gamma_i}\lim_{\alpha\to\infty}\Big(1 - \frac{\gamma_i(X_T^i - \theta_i\bar{X}_T)}{\alpha}\Big)^{\alpha} \\ &= \frac{-e^{-\gamma_i(X_T^i - \theta_i\bar{X}_T)}}{\gamma_i}. \end{aligned}$$

For simplicity, we can write

$$U_i(X_T^i, \bar{X}_T) = \frac{-e^{-\gamma_i(X_T^i - \theta_i\bar{X}_T)}}{\gamma_i}. \tag{4.2}$$

Taking the first and second derivative of the utility function (4.2) with respect to $X_T^i$ yields

$$U_i' = e^{-\gamma_i(X_T^i - \theta_i\bar{X}_T)} \quad \text{and} \quad U_i'' = -\gamma_i e^{-\gamma_i(X_T^i - \theta_i\bar{X}_T)}.$$

Thus, we get the constant Arrow-Pratt coefficient of absolute risk aversion (CARA)

$$A(X_T^i) = -\frac{U_i''}{U_i'} = \gamma_i, \quad \text{for all } X_T^i.$$

Therefore, we say (4.2) is a CARA utility function. Note that CARA agents invest a constant amount of dollars in risky assets regardless of wealth level. Also note that in this case the *inverse* of absolute risk aversion is also constant, $r(X_T^i) = 1/A(X_T^i) = -U_i'/U_i'' = 1/\gamma_i$, which is mathematically still a linear function with respect to wealth.

**Power utility function**: If $*$ is multiplication, and for all $i = 1, ..., n$

$$f(\bar{X}_T, \theta_i) = \bar{X}_T^{-\theta_i}, \quad \gamma_i = 1-\alpha, \quad a_i = (1-\alpha)^{\alpha-1}, \quad b = 0, \quad \eta_i = 0$$

then

$$\begin{aligned} U_i(X_T^i, \bar{X}_T) &= (1-\alpha)^{\alpha-1}\frac{1-\alpha}{\alpha}\Big(\frac{X_T^i \bar{X}_T^{-\theta_i}}{1-\alpha}\Big)^{\alpha} \\ &= \frac{1}{\alpha}(X_T^i \bar{X}_T^{-\theta_i})^{\alpha} \\ &= \frac{(X_T^i \bar{X}_T^{-\theta_i})^{1-\gamma_i}}{1-\gamma_i}, \quad \gamma_i \neq 1. \end{aligned}$$

Thus,

$$U_i(X_T^i, \bar{X}_T) = \frac{(X_T^i \bar{X}_T^{-\theta_i})^{1-\gamma_i}}{1-\gamma_i}, \quad \text{where} \quad \bar{X}_T = \Big(\prod_{i=1}^{n} X_T^i\Big)^{1/n}. \tag{4.3}$$

Here, $\bar{X}_T$ is the geometric mean. Note that together with the arithmetic mean, geometric mean is also a popular measure for investors to evaluate a portfolio selection. Since arithmetic mean is always greater than or equal to geometric mean, using the later one might be a better tool in the sense of safety for investors.

From (4.3), taking the first and second derivatives with respect to $X_T^i$ yields

$$U_i' = (X_T^i \bar{X}_T^{-\theta_i})^{-\gamma_i} \bar{X}_T^{-\theta_i} \quad \text{and} \quad U_i'' = -\gamma_i (X_T^i \bar{X}_T^{-\theta_i})^{-\gamma_i - 1} \bar{X}_T^{-2\theta_i}.$$

Thus, we can check that the power functional form (4.3) exhibits decreasing Arrow-Pratt measure of absolute risk aversion (DARA) and constant Arrow-Pratt measure of relative risk aversion (CRRA), respectively,

$$A(X_T^i) = -\frac{U_i''}{U_i'} = \frac{\gamma_i}{X_T^i} \quad \text{and} \quad R(X_T^i) = -X_T^i \frac{U_i''}{U_i'} = \gamma_i, \quad \text{for all } X_T^i.$$

Thus, the inverse of absolute risk aversion is

$$r(X_T^i) = \frac{1}{A(X_T^i)} = -\frac{U_i'}{U_i''} = \frac{1}{\gamma_i} X_T^i,$$

which is clearly a linear function of $X_T^i$. Therefore, it satisfies Definition 4.2.1 of the HARA family. Further, note that a DARA agent invests more amount of dollars in risky assets as wealth increases and a CRRA agent invests a constant fraction in risky assets regardless of the wealth level, which are two reasonable assumptions and are seemingly supported in the literature (Friend and Blume 1975; Graves 1979).

**Logarithmic utility function**. In this study, the logarithmic utility form is a special case of power utility form (4.3) when $\gamma$ tends to 1. Explicitly, it has the form

$$U_i(X_T^i, \bar{X}_T) = \log(X_T^i \bar{X}_T^{-\theta_i}). \tag{4.4}$$

Taking the first and second derivative of (4.4) with respect to $X_T^i$ yields

$$U_i' = \frac{1}{X_T^i} \quad \text{and} \quad U_i'' = -\frac{1}{(X_T^i)^2}.$$

Hence, we can check for all $X_T^i$ that the Arrow-Pratt coefficient of absolute risk aversion is decreasing and the coefficient of relative risk aversion is constant as terminal wealth increases, respectively,

$$A(X_T^i) = -\frac{U_i''}{U_i'} = \frac{1}{X_T^i} \quad \text{and} \quad R(X_T^i) = -X_T^i \frac{U_i''}{U_i'} = 1 \quad \text{(constant)}.$$

Thus, the *inverse* of absolute risk aversion is

$$r(X_T^i) = \frac{1}{A(X_T^i)} = X_T^i,$$

which is a linear function of $X_T^i$. So, the logarithmic utility function (4.4) satisfies Definition 4.2.1 of HARA utility function. Further, note that the logarithmic form (4.4) is a DARA as well as CRRA utility function.

Note that the power and logarithmic utility functions satisfy the Inada-type condition on the behavior at infinity, i.e. $U_i'(\infty) = 0$ but not zero. Zero is out of our domain since the assumption that $X_T^i \geq \varepsilon$ for some small $\varepsilon > 0$ so we do not need the condition $U_i'(0) = \infty$.

## 4.3 Exponential utility function

**Dynamics**. Consider a game with $n$ agents (investors) investing on the financial market. Without loss of generality, one can work with the two-asset case in which each agent trades between a common risk-free asset (e.g. bond) and an individual risky asset (e.g. stock), which indicates asset specialization[5]. Similarly Lacker and Zariphopoulou (2019), here the bond is common to all agents and plays the role of the numeraire, and we assume it yields zero interest rate (zero-coupon bond). Specifically, the agent $i$ invests in stock $i$ whose price process $S_t^i, t \in [0, T]$, where $T$ is the stopping time, with continuous paths, varies according to the following linear stochastic differential equation (SDE)

$$\frac{dS_t^i}{S_t^i} = \mu_i dt + \nu_i dW_t^i + \sigma_i dB_t, \tag{4.5}$$

where $0 \leq \mu_i \leq \bar{\mu}_i$ and $0 \leq \nu_i \leq \bar{\nu}_i$ and $0 \leq \sigma_i \leq \bar{\sigma}_i$ for some constants $\bar{\mu}_i, \bar{\nu}_i, \bar{\sigma}_i > 0$. Here, $\mu_i, \nu_i$, and $\sigma_i$ are constant market coefficients. The parameter $\mu_i$ denotes the expected rate

[5]Many previous studies examine asset specialization, for example, Brennan (1975), Merton (1987), Coval and Moskowitz (1999), Liu (2014), Basak and Makarov (2015).

of return of the stock $i$ which is assumed to be non-negative. The parameters $\nu_i$ and $\sigma_i$ denote the volatility of some individual asset and the whole market, respectively.

Similarly to Lacker and Zariphopoulou (2019), in this study, the price process is driven by two independent 1-dimension standard Brownian motions $W_t^i$ and $B_t$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ is the complete filtration generated by $n+1$ Brownian motions $W_t^i$, where $i = 1, ..., n$, and $B_t$. Here, each $W_t^i$ is called the idiosyncratic noise (idiosyncratic uncertainty) and $B_t$ is called the common noise (aggregate uncertainty).

### 4.3.1 The $n$-agent games

We examine the games where $n$ finite players trade between a common risk-free asset (e.g. bond) and an individual risky asset (e.g. stock) in a common time horizon $[0, T]$. Suppose that all agents have the same finite time horizon $[0, T]$. The exponential utility function[6] for agent $i$ at the terminal time $T$ is given by

$$U_i(X_T^i, \bar{X}_T) = \frac{-e^{-\gamma_i(X_T^i - \theta_i \bar{X}_T)}}{\gamma_i}, \qquad \gamma_i \neq 0, i = 1, 2, ..., n \tag{4.6}$$

where $\bar{X}_T = \frac{1}{n}\sum_{k=1}^{n} X_T^k$[7] is the arithmetic average terminal wealth invested in assets of the whole population. In the above utility function, $X_T^i \geq 0$ is the agent $i$'s individual wealth invested in assets (or just simply wealth for short) at the terminal period $T$[8], and $\bar{X}_T$ is the average wealth of all agents at the terminal period $T$. Here, $\gamma_i \neq 0$ is a constant (but random) which represents the agent $i$'s Arrow-Pratt measure of absolute risk aversion. It does not depend on the types of assets and her wealth level. Note that the parameter $\gamma_i$ can be positive or negative: $\gamma_i > 0$ represents the case of risk aversion while $\gamma_i < 0$ captures the case of risk-seeking behavior. Here, the relative performance parameter $\theta_i \in [-1, 1]$ for all $i = 1, ..., n$ is a constant and it indicates that agent $i$ takes into account her performance by comparison to her peers[9]) in which they usually assume $\theta_i \in [0, 1]$.

[6] This exponential utility function is more general than that in Lacker and Zariphopoulou (2019). In our study, $\gamma_i$ can be positive or negative. Therefore, the utility function can be concave or convex to capture the case of risk-averse or risk-seeking agents, respectively. In Lacker and Zariphopoulou (2019), however, since the parameter $\delta_i > 0$ so it only captures the case that agents are risk-averse.

[7] Similarly to Lacker and Zariphopoulou (2019), here we consider the case that the average wealth of the population including the agent $i$ for simplicity. The reason is that there is a one-to-one mapping between the case of including and excluding the agent $i$ of the optimization problem (4.8).

[8] Many previous studies in the literature examine long-term investment strategies, for example, Siegel (2014).

[9] Note that in the utility function (4.6) the relative performance parameter $\theta_i$ can be negative, positive, or zero. This is different compared to many previous studies (e.g. Espinosa and Touzi (2015) and Lacker and Zariphopoulou (2019)

Taking the first and second-order derivative of the utility function (4.6) with respect to $X_T^i$, we obtain

$$U_i' = e^{-\gamma_i(X_T^i - \theta_i \bar{X}_T)} \quad \text{and} \quad U_i'' = -\gamma_i e^{-\gamma_i(X_T^i - \theta_i \bar{X}_T)}.$$

We can check whether the utility function $U_i$ is strictly concave or convex depending on the sign of parameter $\gamma_i$. Specifically,

$$U_i(X_T^i, \bar{X}_T) \text{ is } \begin{cases} \text{strictly concave} & \text{if } \gamma_i > 0 \\ \text{strictly convex} & \text{if } \gamma_i < 0. \end{cases}$$

**Definition 4.3.1** (Strategy)**.** *A (trading) strategy (or portfolio) $\pi_t$ at time $t$ corresponding to the exponential utility function (4.6) is defined as $\pi_t := (x_t, y_t) \in \mathcal{F}_s, t < s \leq T$, where $x_t$ and $y_t$ denote the absolute value of wealth invested in stock $i$ and in the numeraire, respectively.*

Note that the above definition is for the exponential utility function. For the case of the power and logarithmic utility function, as we will see in the next section, $x_t$ and $y_t$ denote the fraction of wealth invested in stock $i$ and in the numeraire, respectively, instead of the absolute value of wealth in this section. Note also that $\pi_t \in \mathcal{F}_s, t < s \leq T$ means that investment strategy is adapted to the available information at any time $s$ between the current time $t$ and the terminal time $T$.

The wealth of agent $i$ is described by the following dynamics

$$dX_t^i = \pi_t^i(\mu_i dt + \nu_i dW_t^i + \sigma_i dB_t), \tag{4.7}$$

where $0 \leq \mu_i \leq \bar{\mu}_i, 0 \leq \nu_i \leq \bar{\nu}_i$, and $0 \leq \sigma_i \leq \bar{\sigma}_i$ for all $i = 1, ..., n$; $X_0^i = x_0 \in \mathbb{R}$ is the initial wealth, $\pi_t^i$ is the agent $i$'s investment strategy which indicates the amount of money that the agent $i$ invests in stock $i$. Here, the investment strategy $\pi_t^i$ only influences the change of wealth level $X_t^i$ at any time $t \in [0, T]$. It thus indirectly (but not directly[10]) influences the utility level in the utility function (4.6).

**Definition 4.3.2** (Self-financing strategy)**.** *A self-financing strategy at time $t$ is a strategy $\pi_t = (x_t, y_t) \in \mathcal{F}_s, t < s \leq T$, where the changes in the value of the portfolio are entirely due to trading gains and losses, rather than changes in the external sources.*

Similarly to Lacker and Zariphopoulou (2019) and dos Reis and Platonov (2021), we assume that each agent $i = 1, ..., n$ trades using a self-financing strategy $\pi_t^i, t \in [0, T]$, which satisfies the mean square integrable property, i.e. $\mathbb{E}(\int_0^T |\pi_t^i|^2 dt) < \infty, \forall t \in [0, T]$. In other words,

[10] In the literature, the control might directly affect both the dynamics and the objective function, for example, see Dorfman (1969).

there is no borrowing and lending in this study. Note that $\pi_t^i$ can be positive, negative, or zero which means the agent $i$ buys, sells, or does nothing with asset $i$, respectively. Note also that if $\pi_t^i = 0$ then $dX_t^i = 0$. That is, if agent $i$ does not invest in any assets then her level of wealth will remain unchanged. But in the case that she invests in some asset, i.e. $\pi_t^i \neq 0$, her wealth level might still remain unchanged if $\mu_i dt + \nu_i dW_t^i + \sigma_i dB_t = 0$, which means the asset market is absorbed itself.

Suppose agent $i = 1, ..., n$ chooses admissible trajectories $\pi^1, ..., \pi^n \in \mathcal{A}$, where $\mathcal{A}$ is the set of feasible strategies. The portfolio optimization problem is that the agent $i$ aims to maximize her utility function (4.6) at the terminal period $T$

$$U_i(X_T^i, \bar{X}_T) = \sup_{\pi^i} \mathbb{E}\Big[\frac{-e^{-\gamma_i(X_T^i - \theta_i \bar{X}_T)}}{\gamma_i}\Big] \tag{4.8}$$

given by the dynamics constraint (4.7).

Rewriting the utility function (4.8), we get

$$U_i(\pi^1, ..., \pi^n) = \sup_{\pi^i} \mathbb{E}\left[\frac{-e^{-\gamma_i\Big((1-\theta_i)X_T^i + \theta_i(X_T^i - \bar{X}_T)\Big)}}{\gamma_i}\right]. \tag{4.9}$$

**Definition 4.3.3** (Nash equilibrium)**.** *A vector of strategy profile* $(\pi^{1,*}, ..., \pi^{n,*})$ *of admissible strategies* $\mathcal{A}$ *constitutes a Nash equilibrium if for every* $i = 1, ..., n$ *and for every* $t \in [0, T]$,

$$U_i(\pi_t^{1,*}, ..., \pi_t^{i,*}, ..., \pi_t^{n,*}) \geq U_i(\pi_t^{1,*}, ..., \pi_t^{i-1,*}, \pi^i, \pi_t^{i+1,*}, ..., \pi_t^{n,*})$$

*for all* $\pi^i \in \mathcal{A}$.

Here, we assume that the space of admissible strategies $\mathcal{A}$ is a compact metric space. If the strategy is constant (but random) in the whole time horizon $[0, T]$, then we get the constant (but random) Nash equilibrium.

**Definition 4.3.4** (Constant Nash equilibrium)**.** *A vector of strategy profile* $(\pi^{1,*}, ..., \pi^{n,*})$ *of the set of admissible strategies* $\mathcal{A}$ *constitutes a constant Nash equilibrium if for every* $i = 1, ..., n$,

$$U_i(\pi^{1,*}, ..., \pi^{i,*}, ..., \pi^{n,*}) \geq U_i(\pi^{1,*}, ..., \pi^{i-1,*}, \pi^i, \pi_t^{i+1,*}, ..., \pi^{n,*})$$

*for all* $\pi^i \in \mathcal{A}$.

The following theorem provides the conditions for the existence and uniqueness of a constant Nash equilibrium.

**Theorem 4.3.1** (Existence and uniqueness of constant Nash equilibrium)**.** *Given $U_i$ is strictly concave utility function as in (4.6) with $\gamma_i > 0$. Define two constants*

$$\varphi_n = \frac{1}{n}\sum_{i=1}^{n} \frac{\mu_i}{\gamma_i} \frac{\sigma_i}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} \quad \text{and} \quad \psi_n = \frac{1}{n}\sum_{i=1}^{n} \theta_i \frac{\sigma_i^2}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)}. \tag{4.10}$$

*There are three cases:*

*(i) If $\psi_n \neq 1$, then there exists a unique constant Nash equilibrium given by*

$$\pi^{i,*} = \frac{\mu_i}{\gamma_i} \frac{1}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} + \theta_i \frac{\sigma_i}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} \frac{\varphi_n}{1-\psi_n}. \tag{4.11}$$

*Moreover, we have*

$$\frac{1}{n}\sum_{k=1}^{n} \sigma_k \pi^{k,*} = \frac{\varphi_n}{1-\psi_n}.$$

*(ii) If $\psi_n = 1$ and $\varphi_n \neq 0$, there is no constant Nash equilibrium.*
*(iii) If $\psi_n = 1$ and $\varphi_n = 0$, there are infinitely many constant Nash equilibria.*

*Proof.* We will prove the theorem by using PDE/optimal control approach[11]. Let $i$ be fixed. We denote $\alpha_k \in \mathbb{R}$ be the constant investment strategies followed by all other agents $k, k \neq i$. Let $X_t^k, t \in [0,T]$, be the associated wealth processes, given by

$$X_t^k = \alpha_k(\mu_k t + \nu_k W_t^k + \sigma_k B_t), \quad X_0^k = x_0^k$$

where $x_0^k$ is the initial wealth of the agent $k$. Define

$$Y_t := \frac{1}{n}\sum_{k\neq i} X_t^k, \quad t \in [0,T].$$

Thus, we have

$$U_i(X_t^i, \bar{X}_t) = \frac{-e^{-\gamma_i(X_t^i - \theta_i \bar{X}_t)}}{\gamma_i} = \frac{-e^{-\gamma_i(X_t^i - \frac{\theta_i}{n}(X_t^i + \sum_{k\neq i} X_t^k))}}{\gamma_i} = \frac{-e^{-\gamma_i\left((1-\frac{\theta_i}{n})X_t^i - \theta_i Y_t\right)}}{\gamma_i}.$$

The goal of the agent $i$ is to solve the following optimization (optimal control) problem

$$\sup_{\pi^i \in \mathcal{A}} \mathbb{E}\left[\frac{-e^{-\gamma_i\left((1-\frac{\theta_i}{n})X_t^i - \theta_i Y_t\right)}}{\gamma_i}\right] \tag{4.12}$$

[11]This proof is largely inspired by the proof of Theorem 2.3 in Lacker and Zariphopoulou (2019).

for $t \in [0,T]$, with respect to the following two dynamics (one of agent $i$ and one of all the other agents $k, k \neq i$)

$$
\begin{aligned}
dX_t^i &= \pi_t^i(\mu_i dt + \nu_i dW_t^i + \sigma_i dB_t), && X_0^i = x_0^i \\
dY_t &= \frac{1}{n}\sum_{k\neq i} \alpha_k(\mu_k dt + \nu_k dW_t^k + \sigma_k dB_t), && Y_0 = \frac{1}{n}\sum_{k\neq i} x_0^k.
\end{aligned}
$$

By denote

$$
\widehat{\mu\alpha} =: \frac{1}{n}\sum_{k\neq i}\mu_k\alpha_k \quad \text{and} \quad \widehat{\sigma\alpha} =: \frac{1}{n}\sum_{k\neq i}\sigma_k\alpha_k,
$$

we get

$$
dY_t = \widehat{\mu\alpha}dt + \widehat{\sigma\alpha}dB_t + \frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k.
$$

Also, we denote

$$
\widehat{(\nu\alpha)^2} =: \frac{1}{n}\sum_{k\neq i}\nu_k^2\alpha_k^2.
$$

Making the ansatz of the value function $V_i(X_t^i, Y_t, t)$, which we hope later on will solve the Hamilton-Jacobi-Bellman (HJB) equation, as follows

$$
V_i(X_t^i, Y_t, t) = f(t)\frac{-e^{-\gamma_i\left((1-\frac{\theta_i}{n})X_t^i - \theta_i Y_t\right)}}{\gamma_i}, \quad \text{for } t \in [0,T]. \tag{4.13}
$$

For simplicity, we write $V(X_t^i, Y_t, t) = V_i(X_t^i, Y_t, t)$ for all $i = 1, ..., n$. Suppose $V(X_t^i, Y_t, t)$ is differentiable at least once in $t$ and twice in $X_t^i$ and $Y_t$. For simplicity, denote $x = X_t^i$ and $y = Y_t$, and

$$
V_t = \frac{\partial V}{\partial t}, V_x = \frac{\partial V}{\partial X_t^i}, V_y = \frac{\partial V}{\partial Y_t}, V_{xx} = \frac{\partial^2 V}{\partial (X_t^i)^2}, V_{yy} = \frac{\partial^2 V}{\partial Y_t^2}, V_{xy} = \frac{\partial^2 V}{\partial X_t^i \partial Y_t}
$$

Also, since we are looking for constant Nash equilibrium, we write $\pi_t^i = \pi^i$ for all $t \in [0,T]$. Denote $dV$ as the total differential of $V(X_t^i, Y_t, t)$. Then, $dV$ can be approximated by the multivariate Taylor series expansion

$$
\begin{aligned}
dV &= V_t dt + V_x dx + V_y dy + \frac{1}{2}V_{xx}(dx)^2 + \frac{1}{2}V_{yy}(dy)^2 + V_{xy}dxdy \\
&= V_t dt + V_x\Big[\pi^i(\mu_i dt + \nu_i dW_t^i + \sigma_i dB_t)\Big] + V_y\Big[\widehat{\mu\alpha}dt + \widehat{\sigma\alpha}dB_t + \frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k\Big] \\
&+ \frac{1}{2}V_{xx}\Big[\pi^i(\mu_i dt + \nu_i dW_t^i + \sigma_i dB_t)\Big]^2 + \frac{1}{2}V_{yy}\Big[\widehat{\mu\alpha}dt + \widehat{\sigma\alpha}dB_t + \frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k]^2 \\
&+ V_{xy}\Big[\pi^i(\mu_i dt + \nu_i dW_t^i + \sigma_i dB_t)\Big]\Big[\widehat{\mu\alpha}dt + \widehat{\sigma\alpha}dB_t + \frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k\Big] \\
&= V_t dt + V_x\Big[\pi^i\mu_i dt + \pi^i\nu_i dW_t^i + \pi^i\sigma_i dB_t\Big] + V_y\Big[\widehat{\mu\alpha}dt + \widehat{\sigma\alpha}dB_t + \frac{1}{n}\sum_{k\neq i}^{n}\nu_k\alpha_k dW_t^k\Big] \\
&+ \frac{1}{2}V_{xx}\Big[(\pi^i\mu_i)^2(dt)^2 + (\pi^i\nu_i)^2(dW_t^i)^2 + (\pi^i\sigma_i)^2(dB_t)^2 + 2(\pi^i)^2\mu_i\nu_i dt dW_t^i \\
&+ 2(\pi^i)^2\nu_i\sigma_i dW_t^i dB_t + 2(\pi^i)^2\sigma_i\mu_i dB_t dt\Big] + \frac{1}{2}V_{yy}\Big[(\widehat{\mu\alpha})^2(dt)^2 + (\widehat{\sigma\alpha})^2(dB_t)^2 \\
&+ \frac{1}{n^2}\sum_{k\neq i}\alpha_k^2\nu_k^2(dW_t^i)^2 + 2\widehat{\mu\alpha}\widehat{\sigma\alpha}dt dB_t + 2\widehat{\mu\alpha}\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k dt + 2\widehat{\sigma\alpha}\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k dB_t\Big] \\
&+ V_{xy}\pi^i\sigma_i\frac{1}{n}\sum_{k\neq i}\alpha_k\sigma_k(dB_t)^2.
\end{aligned}
$$

According to the rules $dt.dt = dt.dB_t = dB_t.dt = 0$, we obtain

$$
\begin{aligned}
dV =& V_t dt + V_x\Big[\pi^i\mu_i dt + \pi^i\nu_i dW_t^i + \pi^i\sigma_i dB_t\Big] + V_y\Big[\widehat{\mu\alpha}dt + \widehat{\sigma\alpha}dB_t + \frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k\Big] \\
&+ \frac{1}{2}V_{xx}\Big[(\pi^i\nu_i)^2(dW_t^i)^2 + (\pi^i\sigma_i)^2(dB_t)^2 + 2(\pi^i)^2\nu_i\sigma_i dW_t^i dB_t\Big] + \frac{1}{2}V_{yy}\Big[(\widehat{\sigma\alpha})^2(dB_t)^2 \\
&+ \frac{1}{n}\Big(\frac{1}{n}\sum_{k\neq i}\alpha_k^2\nu_k^2\Big)(dW_t^i)^2 + 2\widehat{\sigma\alpha}\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k dB_t\Big] + V_{xy}\pi^i\sigma_i\frac{1}{n}\sum_{k\neq i}\alpha_k\sigma_k(dB_t)^2.
\end{aligned}
$$

Since $E[dW_t^i] = E[dB_t] = 0, E[(dW_t^i)^2] = E[(dB_t)^2] = dt$, and the fact that $W_t^i, W_t^k$ and $B_t$ are independent, taking expectations and rearranging the equality above, we get

$$
dV = V_t dt + \frac{1}{2}[\sigma_i^2 + \nu_i^2](\pi^i)^2 V_{xx}dt + \pi^i[\mu_i V_x + \sigma_i\widehat{\alpha\sigma}V_{xy}]dt + \frac{1}{2}\Big[\widehat{\alpha\sigma}^2 + \frac{1}{n}\widehat{(\nu\alpha)^2}\Big]V_{yy}dt + \widehat{\alpha\mu}V_y dt.
$$

Dividing both sides of the equality by $dt$ and note the notation that $\dot{V} = \partial V/\partial t$, we get

$$
\dot{V} = V_t + \frac{1}{2}[\sigma_i^2 + \nu_i^2](\pi^i)^2 V_{xx} + \pi^i[\mu_i V_x + \sigma_i\widehat{\alpha\sigma}V_{xy}] + \frac{1}{2}\Big[\widehat{\alpha\sigma}^2 + \frac{1}{n}\widehat{(\nu\alpha)^2}\Big]V_{yy} + \widehat{\alpha\mu}V_y.
$$

At equilibrium, the value of $V$ is optimal, and then is constant, thus $\dot{V} = 0$. Hence, we get the following HJB equation

$$V_t + \frac{1}{2}[\sigma_i^2 + \nu_i^2](\pi^i)^2 V_{xx} + \pi^i[\mu_i V_x + \sigma_i \widehat{\alpha\sigma} V_{xy}] + \frac{1}{2}\Big[\widehat{\alpha\sigma}^2 + \frac{1}{n}\widehat{(\nu\alpha)^2}\Big]V_{yy} + \widehat{\alpha\mu} V_y = 0 \quad (4.14)$$

for $(x, y, t) \in \mathbb{R} \times \mathbb{R} \times [0, T]$, with the terminal condition

$$V(x, y, T) = f(T)\frac{-e^{-\gamma_i\left((1-\frac{\theta_i}{n})X_T^i - \theta_i Y_T\right)}}{\gamma_i}.$$

Taking the first-order condition of (4.14) with respect to $\pi^i$, then solving for $\pi^i$, we obtain the agent $i$'s implicit optimal investment strategy

$$\pi^{i,*} = -\frac{\mu_i V_x + \sigma_i \widehat{\sigma\alpha} V_{xy}}{(\sigma_i^2 + \nu_i^2)V_{xx}}. \quad (4.15)$$

Then, plug (4.15) into (4.14) yields

$$V_t - \frac{1}{2}\frac{(\mu_i V_x + \sigma_i \widehat{\sigma\alpha} V_{xy})^2}{(\sigma_i^2 + \nu_i^2)V_{xx}} + \frac{1}{2}\Big(\widehat{\sigma\alpha}^2 + \frac{1}{n}\widehat{(\nu\alpha)^2}\Big)V_{yy} + \widehat{\mu\alpha} V_y = 0. \quad (4.16)$$

Thus, taking derivatives the ansatz in (4.13) $V(X_t^i, Y_t, t)$ to get $V_t, V_x, V_{xy}, V_{xx}, V_y$, and $V_{yy}$, then plug into (4.16), we get

$$f'(t) - \rho f(t) = 0, \quad \forall t \in [0, T]$$

with $f(T) = 1$ and $\rho$ is given by

$$\rho := \frac{(\mu_i + \theta_i\gamma_i\sigma_i\widehat{\sigma\alpha})^2}{2(\sigma_i^2 + \nu_i^2)} - \theta_i\gamma_i\widehat{\mu\alpha} - \frac{\theta_i^2\gamma_i^2}{2}\Big(\widehat{\sigma\alpha}^2 + \frac{2}{n}\widehat{(\nu\alpha)^2}\Big).$$

Consequently, we get $f(t) = e^{-\rho(T-t)}$. Hence,

$$V(x, y, t) = e^{-\rho(T-t)}\frac{-e^{-\gamma_i[(1-\frac{\theta_i}{n})x - \theta_i y]}}{\gamma_i} = \frac{-e^{-\gamma_i[(1-\frac{\theta_i}{n})x - \theta_i y] - \rho(T-t)}}{\gamma_i}. \quad (4.17)$$

Taking derivatives $V(x, y, t)$ to get $V_x, V_{xy}$, and $V_{xx}$, then plug into (4.15), we get the agent $i$'s explicit optimal investment strategy of the optimal control problem (4.12)

$$\pi^{i,*} = \frac{\mu_i/\gamma_i + \theta_i\sigma_i\widehat{\sigma\alpha}}{(\sigma_i^2 + \nu_i^2)(1 - \theta_i/n)}, \quad (4.18)$$

which does not depend on $t$.

For an admissible portfolio vector $(\alpha_1, ..., \alpha_n)$ to be a constant Nash equilibrium, we need $\alpha_i = \pi^{i,*}, \forall i = 1, ..., n$. Define

$$\bar{\sigma\alpha} := \frac{1}{n}\sum_{k=1}^{n} \sigma_k \alpha_k = \frac{1}{n}\sum_{k\neq i}^{n} \sigma_k \alpha_k + \frac{1}{n}\sigma_i\alpha_i = \widehat{\sigma\alpha} + \frac{1}{n}\sigma_i\alpha_i.$$

Then,

$$\widehat{\sigma\alpha} = \bar{\sigma\alpha} - \frac{1}{n}\sigma_i\alpha_i.$$

Thus, plug into (4.18), we have

$$\alpha_i = \pi^{i,*} = \frac{\mu_i/\gamma_i + \theta_i\sigma_i\bar{\sigma\alpha}}{(\sigma_i^2 + \nu_i^2)(1 - \theta_i/n)} - \frac{\theta_i\sigma_i^2}{n(\sigma_i^2 + \nu_i^2)(1 - \theta_i/n)}\alpha_i.$$

Hence, we get

$$\begin{aligned}
\alpha_i &= \frac{\mu_i/\gamma_i + \theta_i\sigma_i\bar{\sigma\alpha}}{(\sigma_i^2 + \nu_i^2)(1 - \theta_i/n)}\Big(1 + \frac{\theta_i\sigma_i^2}{n(\sigma_i^2 + \nu_i^2)(1 - \theta_i/n)}\Big)^{-1} \\
&= \frac{\mu_i/\gamma_i + \theta_i\sigma_i\bar{\sigma\alpha}}{(\sigma_i^2 + \nu_i^2)(1 - \theta_i/n) + \sigma_i^2\theta_i/n} \\
&= \frac{\mu_i/\gamma_i + \theta_i\sigma_i\bar{\sigma\alpha}}{\sigma_i^2 + \nu_i^2(1 - \theta_i/n)}.
\end{aligned}$$

Multiplying both sides by $\sigma_i$ yields

$$\sigma_i\alpha_i = \frac{(\sigma_i\mu_i)/\gamma_i + \theta_i\sigma_i^2\bar{\sigma\alpha}}{\sigma_i^2 + \nu_i^2(1 - \theta_i/n)}.$$

Thus, averaging over $i = 1, ..., n$, we get

$$\frac{1}{n}\sum_{i=1}^{n} \sigma_i\alpha_i = \frac{1}{n}\sum_{i=1}^{n} \frac{\mu_i}{\gamma_i}\frac{\sigma_i}{(\sigma_i^2 + \nu_i^2)(1 - \theta_i/n)} + \frac{1}{n}\sum_{i=1}^{n} \theta_i\frac{\sigma_i^2}{(\sigma_i^2 + \nu_i^2)(1 - \theta_i/n)}\bar{\sigma\alpha}$$

and then

$$\bar{\sigma\alpha} = \varphi_n + \psi_n\bar{\sigma\alpha}, \tag{4.19}$$

where

$$\varphi_n = \frac{1}{n}\sum_{i=1}^{n} \frac{\mu_i}{\gamma_i}\frac{\sigma_i}{\sigma_i^2 + \nu_i^2(1 - \theta_i/n)} \quad \text{and} \quad \psi_n = \frac{1}{n}\sum_{i=1}^{n} \theta_i\frac{\sigma_i^2}{\sigma_i^2 + \nu_i^2(1 - \theta_i/n)}.$$

Note that since $\theta_i \in [-1, 1], \sigma_i \geq 0, \nu_i \geq 0$ and $n$ is finite but large thus $-1 \leq \psi_n \leq 1$. There are three cases.

(i) If $\psi_n \neq 1$, then the equation (4.19) yields $\bar{\sigma\alpha} = \varphi_n/(1-\psi_n)$. Thus, there exists a unique constant Nash equilibrium given by

$$\pi^{i,*} = \frac{\mu_i}{\gamma_i} \frac{1}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} + \theta_i \frac{\sigma_i}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} \frac{\varphi_n}{1-\psi_n}.$$

(ii) If $\psi_n = 1$ (if and only if $\theta_i = 1$ and $\nu_i = 0$ for all $i = 1, ..., n$) and $\varphi_n \neq 0$, thus the equation (4.19) has no solution and then no constant Nash equilibria exist.

(iii) If $\psi_n = 1$ and $\varphi_n = 0$, thus the equation (4.19) has infinitely many solutions and then there are infinitely many constant Nash equilibria. □

**Remark 4.3.1.** *One necessary condition for Case (ii) and (iii) in Theorem 3.2 is $\psi_n = 1$. This only happens if and only if $\theta_i = 1$ and $\nu_i = 0$ for all $i = 1, 2, ..., n$. That means each investor is extremely concerned about the performance of the rest of other investors and also the volatility of each individual stock is zero.*

One real-world example to explain the case that $\theta_i = 1$ and $\nu_i = 1$ for all $i = 1, ..., n$ is the (centrally) planned economy where prices are fixed ($\nu_i = 0, \forall i$) and agents are totally concerned with others' wealth ($\theta_i = 1, \forall i$), thus $\psi_n = 1$ holds. Then, it does not matter if agents invest more or less to a specific asset. That means there are infinitely many solutions in this economic setting.

**Remark 4.3.2.** *The constant $\varphi_n$ can be positive or negative depending on the sign of $\mu_i/\gamma_i$. If $\mu_i/\gamma_i > 0$ for all $i$ then $\varphi_n > 0$ and if $\mu_i/\gamma_i < 0$ for all $i$ then $\varphi_n < 0$.*

**Remark 4.3.3.** *The case that $\varphi_n = 0$ can also happen. This might be the situation that $\gamma_i > 0$ for some $i$ and $\gamma_i < 0$ for some $i$. This implies that there must be some risk-averse agents ($\gamma_i > 0$) and some risk-seeking agents ($\gamma_i < 0$) interacting on the same market.*

**Theorem 4.3.2** (Unique corner solution for strictly convex exponential utility function)**.** *Given $U_i$ is utility function of agent $i$ as in (4.6) with $\gamma_i < 0$. For the exponential case, the optimal solution is $\pi^{i,*} = C_i$ for risky asset (stock) and $\pi^{i,*} = 0$ for riskless asset (bond), where $C_i$ is the total asset value that agent $i$ has.*

*Proof.* First we need to show that the domain of $U_i$ is bounded and closed. Since $t \in [0, T]$, where $T$ is stopping time, $W_t^i$ and $B_t$ in dynamics (4.7) are bounded. Also, since $\mu_i, \nu_i$, and $\sigma_i$ are bounded so $X_t^i$ is bounded. Moreover, since $t$ is closed, $W_t^i$ and $B_t$ are closed. In addition, since $\mu_i, \nu_i$, and $\sigma_i$ are in closed intervals so $X_t^i$ is closed.

For $\gamma_i < 0$, we can check the second order sufficient condition that

$$U''_{X_T^i} = -\gamma_i e^{-\gamma_i(X_T^i - \theta_i \bar{X}_T)} > 0,$$

which implies that $U_i$ is a strictly convex utility function. Since $U_i$ is a strictly convex function and the domain of $U_i$ that is bounded and closed, there exists $m \in S$ such that $U_i(m)$ is global minimum. Suppose $U_i$ attains maximum at $M \in S$ and $M \notin \partial S$. Thus, there exists $s \in S$ such that the line connecting $m$ and $M$ can be extended to $s$ and $U_i(M) \neq U_i(s)$. If not, then $M$ is on the boundary $\partial S$, which is contradict to the above assumption that $M \notin \partial S$.

Since $m, M, s \in S$, there exists $\lambda \in [0,1]$ such that $M = \lambda m + (1-\lambda)s$. Since $U_i$ is strictly convex function,

$$U_i(M) = U_i(\lambda m + (1-\lambda)s) < \lambda U_i(m) + (1-\lambda)U_i(s) < \lambda U_i(s) + (1-\lambda)U_i(s) = U_i(s),$$

which is contradict to the fact that $M$ is maximum of $U_i$. Hence, $M \in \partial S$. That is, the maximum is on the boundary of the domain.

We have $U'_{X_T^i} = e^{-\gamma_i(X_T^i - \theta_i \bar{X}_T)} > 0$, which implies that $U_i$ is a strictly increasing function in $X_T^i$. Thus, $M$ must be on the upper boundary of the domain and $M$ must be unique. This implies that $\pi^{i,*} = C_i$ for the risky asset (stock) and $\pi^{i,*} = 0$ for riskless asset (bond), where $C_i$ is the total asset that agent $i$ has. For agent $i$, $C_i$ is unique so the corner solution is also unique. $\square$

**Definition 4.3.5.** *An agent shorts a stock if and only if $\pi^{i,*} < 0$.*

**Remark 4.3.4.** *Consider the case of relative performance concern, i.e. $\theta_i \in [0,1]$. The constant Nash equilibrium now is*

$$\pi^{i,*} = \frac{\mu_i}{\gamma_i} \frac{1}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} + \theta_i^+ \frac{\sigma_i}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} \frac{\varphi_n}{1-\psi_n}.$$

*If $\mu_i/\gamma_i > 0$ ($\mu_i > 0$ and $\gamma_i > 0$) for all $i$, then agent $i$ does not short sell. If $\mu_i/\gamma_i < 0$ ($\mu_i > 0$ and $\gamma_i < 0$) for all $i$, then agent $i$ short sells. This implies that risk-averse agents are unlikely to short sell stock while risk-seeking agents are likely to short sell stock.*

**Corollary 4.3.1** (Existence and uniqueness equilibrium for single stock)**.** *Assume for all $i = 1, ..., n$ that $0 \leq \mu_i = \mu < \bar{\mu}, 0 \leq \sigma_i = \sigma \leq \bar{\sigma}$, and $\nu_i = 0$. Define $\delta_i$ and two constants*

$$\delta_i := \frac{1}{\gamma_i}, \quad \bar{\delta} := \frac{1}{n}\sum_{i=1}^{n} \delta_i \quad \text{and} \quad \bar{\theta} := \frac{1}{n}\sum_{i=i}^{n} \theta_i.$$

*There are three cases:*

*(i) If $\bar{\theta} \neq 1$, then there exists a unique constant Nash equilibrium given by*

$$\pi^{i,*} = \Big(\delta_i + \theta_i \frac{\bar{\delta}}{1-\bar{\theta}}\Big) \frac{\mu}{\sigma^2}. \tag{4.20}$$

*(ii) If $\bar{\theta} = 1$ and $\mu\bar{\delta} \neq 0$, there is no constant Nash equilibria.*
*(iii) If $\bar{\theta} = 1$ and $\mu\bar{\delta} = 0$, there exist infinitely many constant Nash equilibria.*

*Proof.* With the assumptions of parameters that $\mu_i = \mu, \sigma_i = \sigma \geq 0$, and $\nu_i = 0$ for all $i = 1, ....n$, we now have

$$\varphi_n = \frac{1}{n} \sum_{k=1}^{n} \delta_k \frac{\mu}{\sigma} = \frac{\mu}{\sigma}\bar{\delta} \quad \text{and} \quad \psi_n = \frac{1}{n} \sum_{k=1}^{n} \theta_k \frac{\sigma^2}{\sigma^2} = \bar{\theta}.$$

By applying Theorem 4.3.1, (i) holds since

$$\pi^{i,*} = \delta_i \frac{\mu}{\sigma^2} + \frac{\theta_i}{\sigma} \frac{\varphi_n}{1-\psi_n} = \Big(\delta_i + \theta_i \frac{\bar{\delta}}{1-\bar{\theta}}\Big) \frac{\mu}{\sigma^2}.$$

Since $\psi_n = 1$ if and only if $\bar{\theta} = 1$ and $\varphi_n \neq 0$ if and only if $\mu\bar{\delta} \neq 0$, thus (ii) holds. Moreover, since $\varphi_n = 0$ if and only if $\mu\bar{\delta} = 0$, thus (iii) holds. □

### 4.3.2 The mean field games

In this section, we examine the limiting behavior of $n$-player games as the number of agents tends to infinity, i.e. $n \to \infty$.

**Assumption 4.3.1** (Continuum)**.** *The population is made by the continuum of agents.*

Since the number of agents is continuous, we now no longer consider the agent $i = 1, ..., n$ discretely as in the previous section. Instead, in this section, we are going to investigate the continuous representative agent.

**Definition 4.3.6** (MFG)**.** *The MFG corresponding to the exponential utility function* (4.6) *is defined by a type vector $\zeta = (\xi, \gamma, \theta, \mu, \nu, \sigma)$ of a representative agent $i$ where $\xi$ is the initial wealth of the representative agent.*

Here, since the parameters of risk preferences $\gamma$ and $\theta$ are random, agents are not necessarily homogeneous.

The dynamics of wealth of the representative agent, for $t \in [0, T]$, is given by

$$dX_t = \pi_t(\mu dt + \nu dW_t + \sigma dB_t), \quad X_0 = \xi. \tag{4.21}$$

Given $\bar{X}$ denoting the arithmetic average wealth of the continuum of agents and the dynamics (4.21), the goal of the representative agent is to maximize the expected utility in the following optimization problem

$$\sup_{\pi_t \in \mathcal{A}_{MF}} \mathbb{E}\Big[\frac{-e^{-\gamma(X_T-\theta\bar{X})}}{\gamma}\Big] \tag{4.22}$$

where $\mathcal{A}_{MF}$ denotes the admissible set containing feasible strategies $\pi_t, t \in [0,T]$, of the mean field games.

The main goal of this section is to prove the existence of the mean field equilibrium (MFE), which is stated in the following definitions.

**Definition 4.3.7** (MFE)**.** *Given $\bar{X}$ and let $\pi_t^* \in \mathcal{A}$ be an admissible strategy. Then, $\pi_t^*$ is a MFE if it solves the optimization problem (4.22) given the dynamics (4.21).*

**Definition 4.3.8** (Constant MFE)**.** *The optimal strategy $\pi^*$ is called a constant MFE if $\pi^* = \pi_t^*$ for all $t \in [0,T]$, where $\pi_t^*$ is a MFE defined in Definition 4.3.7.*

**Lemma 4.3.1.** *Define two constants*

$$\varphi := \mathbb{E}\Big(\frac{\mu}{\gamma}\frac{\sigma}{\sigma^2+\nu^2}\Big) \quad \text{and} \quad \psi := \mathbb{E}\Big(\theta\frac{\sigma^2}{\sigma^2+\nu^2}\Big). \tag{4.23}$$

*Then these two expectations exist and are finite.*

*Proof.* Denote two constant but random variables

$$\Xi = \frac{\mu}{\gamma}\frac{\sigma}{\sigma^2+\nu^2} \quad \text{and} \quad \Lambda = \theta\frac{\sigma^2}{\sigma^2+\nu^2}.$$

By assumptions $\underline{\gamma} \leq \gamma \leq \bar{\gamma}, 0 \leq \mu \leq \bar{\mu}, 0 \leq \sigma \leq \bar{\sigma}$, and $0 \leq \nu \leq \bar{\nu}$. Then, $\Xi$ must exist and be bounded. This implies that $\Xi$ is finite, and so is $\varphi$.

Now, we will show that $\psi$ exists and is finite. Since $\theta \in [-1,1]$ and $\sigma^2/(\sigma^2+\nu^2) \in [0,1]$, then $-1 \leq \Lambda \leq 1$. We will show that $-1 \leq \mathbb{E}(\Lambda) \leq 1$. In fact, if one of two these inequalities does not hold, that is

$$\mathbb{E}(\Lambda) < -1 \quad \text{or} \quad \mathbb{E}(\Lambda) > 1.$$

Then, there must exist $\theta, \sigma$, and $\nu$ such that

$$\Lambda < -1 \quad \text{or} \quad \Lambda > 1.$$

But it contradicts the fact that $-1 \leq \Lambda \leq 1$. So, there must exist $\psi$, which is finite and bounded in the closed interval $-1 \leq \psi \leq 1$. So, the proof is complete. $\square$

**Theorem 4.3.3** (Existence and uniqueness of constant MFE). *Given strictly concave utility function $U$. Assume for a.s. that $-1 \leq \theta \leq 1, 0 \leq \mu \leq \bar{\mu}, 0 \leq \nu \leq \bar{\nu}$, and $0 \leq \sigma \leq \bar{\sigma}$. Define two constants $\varphi$ and $\psi$ as that in Lemma 4.3.1.*

*There are three cases:*

*(i) If $\psi \neq 1$, there exists a unique constant MFE, given by*

$$\pi^* = \frac{\mu}{\gamma} \frac{1}{\sigma^2 + \nu^2} + \theta \frac{\sigma}{\sigma^2 + \nu^2} \frac{\varphi}{1 - \psi}. \tag{4.24}$$

*Moreover, we have*

$$\mathbb{E}(\sigma \pi^*) = \frac{\varphi}{1 - \psi}.$$

*(ii) If $\psi = 1$ and $\varphi \neq 0$, there is no constant MFE.*
*(iii) If $\psi = 1$ and $\varphi = 0$, there are infinitely many constant MFE.*

*Proof.* A representative agent will solve the optimization problem (4.22) with respect to the dynamics (4.21), given the average wealth of the population $\bar{X}$. Define $\bar{X}_t := \mathbb{E}(X_t^{\pi_t} | \mathcal{F}_t^B), \forall t \in [0, T]$, where $\mathcal{F}_t^B$ is the filtration generated by the common noise $B_t$.

The integral form of the dynamics (4.21) is

$$X_t = \xi + \mu \pi_t \int_0^t ds + \nu \pi_t \int_0^t dW_s + \sigma \pi_t \int_0^t dB_s, \quad \forall t \in [0, T] \tag{4.25}$$

Since we want to find the constant MFE, we need to find the corresponding constant strategy $\alpha = \pi_t, \forall t \in [0, T]$. Given $\alpha = \pi_t, \forall t \in [0, T]$, taking expectation (4.25) with the note that $(\xi, \mu, \nu, \sigma, \alpha), W_t$ and $B_t$ are independent, we get

$$\bar{X}_t = \bar{\xi} + \bar{\mu \alpha} t + \bar{\sigma \alpha} B_t$$

where we use the following facts, for all $t \in [0, T]$,

$$B_0 = 0 \text{ a.s.}, \int_0^t ds = t, \mathbb{E}\Big(\int_0^t dW_s\Big) = \mathbb{E}(W_t) = 0, \int_0^t dB_s = B_t.$$

and the expectation of an integrable random variable $S$ is denoted by $\bar{S} = \mathbb{E}(S)$.

Denote

$$Z_t^\pi := X_t^\pi - \theta\bar{X}_t, \quad \forall t \in [0,T].$$

Thus,

$$\begin{aligned}
dZ_t^\pi = dX_t^\pi - \theta d\bar{X}_t &= \pi_t(\mu dt + \nu dW_t + \sigma dB_t) - \theta(\bar{\mu\alpha} t + \bar{\sigma\alpha} B_t) \\
&= (\mu\pi - \theta\bar{\mu\alpha})dt + \nu\pi dW_t + (\sigma\pi - \theta\bar{\sigma\alpha})dB_t
\end{aligned}$$

The problem (4.22) now is equivalent to

$$V(Z_t^\pi, t) = \sup_{\pi\in\mathcal{A}_{MF}} \mathbb{E}\Big[\frac{-e^{-\gamma Z_T^\pi}}{\gamma}\Big] \tag{4.26}$$

where $V : \mathbb{R} \times [0,T] \to \mathbb{R}$ denotes the smooth and strictly increasing value function, which is expected to solve the HJB equation later on. Suppose $V(Z_t^\pi, t)$ is continuously at least twice differentiable in $Z_t^\pi$ and once in $t$. For simplicity, denote

$$V_t = \frac{\partial V}{\partial t}, V_x = \frac{\partial V}{\partial X_t^\pi}, V_z = \frac{\partial V}{\partial Z_t^\pi}, V_{xx} = \frac{\partial^2 V}{\partial (X_t^\pi)^2}, V_{zz} = \frac{\partial^2 V}{\partial (Z_t^\pi)^2}$$

It is easy to see that $V_x = V_z$ and $V_{xx} = V_{zz}$. Now, the total differential form $dV$ can be approximated by a Taylor series expansion with terms of order equal or less than $dt$ or $(dW_t)^2$ or $(dB_t)^2$

$$\begin{aligned}
dV =& V_t dt + V_x dZ_t^\pi + \frac{1}{2}V_{xx}(dZ_t^\pi)^2 \\
=& V_t dt + V_z[(\mu\pi - \theta\bar{\mu\alpha})dt + \nu\pi dW_t + (\sigma\pi - \theta\bar{\sigma\alpha})dB_t] \\
&+ \frac{1}{2}V_{xx}[(\mu\pi - \theta\bar{\mu\alpha})dt + \nu\pi dW_t + (\sigma\pi - \theta\bar{\sigma\alpha})dB_t]^2 \\
=& V_t dt + V_x[(\mu\pi - \theta\bar{\mu\alpha})dt + \nu\pi dW_t + (\sigma\pi - \theta\bar{\sigma\alpha})dB_t] \\
&+ \frac{1}{2}V_{xx}[(\nu\pi)^2(dW_t)^2 + (\sigma\pi - \theta\bar{\sigma\alpha})^2(dB_t)^2 + 2(\mu\pi - \theta\bar{\mu\alpha})\nu\pi dt dW_t \\
&+ 2(\mu\pi - \theta\bar{\mu\alpha})(\sigma\pi - \theta\bar{\sigma\alpha})dt dB_t + 2\nu\pi(\sigma\pi - \theta\bar{\sigma\alpha})dW_t dB_t].
\end{aligned}$$

Since $E[dW_t] = E[dB_t] = 0$ and $(dW_t)^2 = (dB_t)^2 = dt$, taking expectation, we get

$$dV = V_t dt + (\mu\pi - \theta\bar{\mu\alpha})V_x dt + \frac{1}{2}[(\nu\pi)^2 + (\sigma\pi - \theta\bar{\sigma\alpha})^2]V_{xx}dt.$$

Dividing both sides of the equality above by $dt$ and note that $\dot{V} = dV/dt$, we get

$$\dot{V} = V_t + (\mu\pi - \theta\bar{\mu\alpha})V_x + \frac{1}{2}[(\nu\pi)^2 + (\sigma\pi - \theta\bar{\sigma\alpha})^2]V_{xx}.$$

At equilibrium point, $\dot{V} = 0$, then we obtain the HJB equation

$$V_t + (\mu\pi - \theta\mu\bar{\alpha})V_x + \frac{1}{2}[(\nu\pi)^2 + (\sigma\pi - \theta\sigma\bar{\alpha})^2]V_{xx} = 0 \tag{4.27}$$

with the terminal condition $V(z, T) = -e^{-\gamma z}/\gamma$.

Taking the first-order condition (4.27) with respect to $\pi$, then after several manipulations, we get the representative agent's implicit optimal investment strategy

$$\pi = -\frac{\mu V_x - \theta\sigma\sigma\bar{\alpha}V_{xx}}{(\sigma^2 + \nu^2)V_{xx}}. \tag{4.28}$$

Plugging back into (4.27), it derives to

$$V_t - \frac{1}{2}\frac{(\mu V_x - \theta\sigma\sigma\bar{\alpha}V_{xx})^2}{(\sigma^2 + \nu^2)V_{xx}} - \theta\mu\bar{\alpha}V_x + \frac{1}{2}(\theta\sigma\bar{\alpha})^2 V_{xx} = 0. \tag{4.29}$$

Making the ansatz $V(z, t) = f(t)\frac{-e^{-\gamma z}}{\gamma}$, and then taking derivatives with the note that $V_x = V_z$ and $V_{xx} = V_{zz}$. Thus, plug $V_t, V_x$ and $V_{xx}$ into (4.29), we get

$$f'(t) - \rho f(t) = 0, \quad \forall t \in [0, T]$$

with $f(T) = 1$ and $\rho$ is given by

$$\rho := \frac{(\mu + \theta\gamma\sigma\sigma\bar{\alpha})^2}{2(\sigma^2 + \nu^2)} - \theta\gamma\mu\bar{\alpha} - \frac{(\theta\gamma\sigma\bar{\alpha})^2}{2}.$$

Hence, we obtain

$$f(t) = e^{-\rho(T-t)}.$$

Plugging into the value function, we get

$$V(z, t) = e^{-\rho(T-t)}\frac{-e^{-\gamma z}}{\gamma} = -\frac{1}{\gamma}e^{-\rho(T-t)-\gamma z}.$$

Taking derivatives to get $V_z$ and $V_{zz}$ and note that $V_z = V_x$ and $V_{zz} = V_{xx}$. Plugging into (4.28), we get the explicit optimal investment strategy of the optimal control problem (4.26)

$$\pi^* = \frac{\mu}{\gamma}\frac{1}{(\sigma^2 + \nu^2)} + \theta\frac{\sigma}{\sigma^2 + \nu^2}\sigma\bar{\alpha}. \tag{4.30}$$

Note that $\pi^*$ is a constant and does not depend on either $z$ or $t$. The feasible control $\alpha$ is a constant MFE if $\alpha = \pi^*$, i.e. we need

$$\pi^* = \frac{\mu}{\gamma}\frac{1}{\sigma^2+\nu^2} + \theta\frac{\sigma}{\sigma^2+\nu^2}\bar{\sigma\pi^*}.$$

Multiplying both sides by $\sigma$ and then taking expectation, we get

$$\bar{\sigma\pi^*} = \mathbb{E}\Big[\frac{\mu}{\gamma}\frac{\sigma}{\sigma^2+\nu^2}\Big] + \mathbb{E}\Big[\theta\frac{\sigma^2}{\sigma^2+\nu^2}\Big]\bar{\sigma\pi^*} = \varphi + \psi\bar{\sigma\pi^*} \tag{4.31}$$

where $\bar{\sigma\pi^*} = \mathbb{E}(\sigma\pi^*)$. Note that since $0 \le \theta \le 1, \sigma \ge 0, \nu \ge 0$ then $0 \le \psi \le 1$. There are three cases:

(i) If $\psi \ne 1$, thus from (4.31) we get $\bar{\sigma\pi^*} = \varphi/(1-\psi)$. Hence, there exists a unique constant MFE given by

$$\pi^* = \frac{\mu}{\gamma}\frac{1}{\sigma^2+\nu^2} + \theta\frac{\sigma}{\sigma^2+\nu^2}\frac{\varphi}{1-\psi}.$$

Note that $\bar{\sigma\pi^*}$ can be equals to zero since $\sigma$ can be equal zero or $\pi^*$ can be equal to zero.
(ii) If $\psi = 1$ (if and only if $\theta = 1$ and $\nu = 0$) and $\varphi \ne 0$, thus (4.31) has no solution and then no constant MFE exist.
(iii) If $\psi = 1$ and $\varphi = 0$, thus (4.31) has infinitely many solutions and then there are infinitely many constant MFE. □

**Remark 4.3.5** (No relative performance)**.** *In the case there is no relative performance, i.e.* $\theta = 0$*, the MFE* (4.24) *and the NE* (4.11) *equal to*

$$\pi^* = \frac{\mu}{\gamma}\frac{1}{\sigma^2+\nu^2}. \tag{4.32}$$

*Note that* (4.32) *is exactly the classical Merton portfolio in Merton (1969) when there is only one Brownian process and bond (sure asset) offers zero interest rate.*

**Corollary 4.3.2** (Existence and uniqueness constant MFE for single stock)**.** *Assume that* $(\mu, \sigma, \nu)$ *are deterministic with* $0 \le \mu \le \bar{\mu}, 0 \le \nu \le \bar{\nu}$*, and* $0 \le \sigma \le \bar{\sigma}$*. Define three constants*

$$\delta := \frac{1}{\gamma}, \quad \bar{\delta} := \mathbb{E}(\delta) \quad \text{and} \quad \bar{\theta} := \mathbb{E}(\theta).$$

*There are three cases:*

*(i) If* $\bar{\theta} \ne 1$*, then there exists a unique constant MFE given by*

$$\pi^* = \Big(\delta + \theta\frac{\bar{\delta}}{1-\bar{\theta}}\Big)\frac{\mu}{\sigma^2}. \tag{4.33}$$

*(ii) If $\bar{\theta} = 1$ and $\mu\bar{\delta} \neq 0$, there is no constant MFE.*
*(iii) If $\bar{\theta} = 1$ and $\mu\bar{\delta} = 0$, there are infinitely many constant MFE.*

*Proof.* With the assumption $\nu = 0$, then the two later constants become

$$\varphi = \frac{\mu\sigma}{\sigma^2 + \nu^2}\bar{\delta} = \frac{\mu}{\sigma}\bar{\delta} \quad \text{and} \quad \psi = \frac{\sigma^2}{\sigma^2 + \nu^2}\bar{\theta} = \bar{\theta}.$$

Applying Theorem 4.3.3 and note that $\psi \neq 0$ if and only if $\bar{\theta} \neq 1$, thus (i) holds since

$$\pi^* = \frac{\mu}{\gamma}\frac{1}{\sigma^2 + \nu^2} + \theta\frac{\sigma}{\sigma^2 + \nu^2}\frac{\varphi}{1 - \psi} = \delta\frac{\mu}{\sigma^2} + \frac{\theta}{\sigma}\frac{\varphi}{1 - \psi} = \left(\delta + \theta\frac{\bar{\delta}}{1 - \bar{\theta}}\right)\frac{\mu}{\sigma^2}.$$

Since $\psi = 1$ if and only if $\bar{\theta} = 1$ and $\varphi \neq 0$ if and only if $\mu\bar{\delta} \neq 0$, thus (ii) holds. Moreover, since $\varphi = 0$ if and only if $\mu\bar{\delta} = 0$, thus (iii) holds. □

**Remark 4.3.6.** *(Compare NE and MFE in our study with those in Lacker and Zariphopoulou (2019) for the case of strictly concave utility function.)*

- *Our setting take into account the case of relative performance motivation. This is not the case in Lacker and Zariphopoulou (2019).*
- *One difference between Theorem 4.3.1 and Theorem 4.3.3 in our study with the corresponding Theorem 2.3 and Theorem 2.10 in Lacker and Zariphopoulou (2019), respectively, is that the case $\psi = 1$ and $\varphi = 0$ cannot happen in the former study but it happens in our study. As a result, there are infinitely many constant MFE in such case.*
- *By assuming of the signs of parameters in Lacker and Zariphopoulou (2019), the optimal strategies must be positive. It contradicts to their comment on page 2 of their paper that "the value $\pi_t^{i,*}$ may be negative, indicating that the agent shorts the stock". Indeed, the optimal control $\pi^{i,*}$ and $\pi^*$ in their study cannot be negative. Consequently, it cannot take into account the case of short selling. In contrast, in our study, with several different assumptions of the signs of parameters, short selling can happen. That means the optimal strategy $\pi^{i,*}, \pi^*$ can be negative, i.e. $\pi^{i,*} < 0, \pi^* < 0$, which capture the popular transaction in the stock market that agents can short their stocks.*

**MFE for strictly convex exponential utility function**. The optimal strategy for Theorem 4.3.2 does not depend on number of players $n$. Therefore, the optimal strategy for the case of $n$-agent games (Nash equilibrium) and that for the case of MFG (mean field equilibrium) should be similar. Only different point is that the optimal strategy in the case of MFG does not depend on $i$. That is, $\pi^* = C$ for risky asset (stock) and $\pi^* = 0$ for riskless asset (bond), where $C$ is the total asset value that the representative agent has. For each representative agent, since $C$ is unique so is the corner solution.

## 4.4 Power and logarithmic utility function

The power and logarithmic utility functions belong to the family of constant relative risk aversion (CRRA) utility functions. In this section, the CRRA utility function is given by

$$U_i(X_T^i, \bar{X}_T) = \begin{cases} \frac{1}{1-\gamma_i}\left(X_T^i \bar{X}_T^{-\theta_i}\right)^{1-\gamma_i} & \text{if } \gamma_i \neq 1 \\ \log(X_T^i \bar{X}_T^{-\theta_i}) & \text{if } \gamma_i = 1 \end{cases} \tag{4.34}$$

where $\bar{X}_T = \left(\prod_{i=1}^{n-1} X_T^i\right)^{\frac{1}{n-1}}$ is the geometric average wealth of the population excluding agent $i$. Note that, here, we consider the case of excluding agent $i$ since, similar to the case of exponential form, there is a one-to-one map from excluding to including agent $i$ case. Thus, the optimal strategy, if it exists, in both cases are the same. The model with including agent $i$ is more tractable and simpler. Therefore, we employ this model in this study. Literally, the given CRRA utility function (4.34) has the power form if $\gamma_i \neq 1$ and logarithmic form if $\gamma_i = 1$. Since the logarithmic form is a special case of power form when $\gamma_i$ tends to 1, we will not consider these two utility forms in two separate sections but rather consider the former case first, then naturally move to the later case within one section. Moreover, we assume that $\underline{\gamma_i} \leq \gamma_i \leq \bar{\gamma}_i$ for some $\underline{\gamma_i} < 0$ and $\underline{\gamma_i} > 0$.

### 4.4.1 The $n$-agent games

Suppose there are $n$ agents trading on the stock market. They have the same investment horizon $T$, where $0 < T < \infty$. The dynamics of wealth of agent $i$ at time $t$, $0 \leq t \leq T$, is given by

$$dX_t^i = \pi_t^i X_t^i(\mu_i dt + \nu_i dW_t^i + \sigma_i B_t), \quad X_0^i = x_0^i, \tag{4.35}$$

where $X_0^i = x_0 \in \mathbb{R}$ is the initial wealth, $\pi_t^i$ is the fraction of wealth, rather than the absolute value of wealth in the case of exponential form, that agent $i$ invests in the individual stock $S^i$ at time $t$. Here, $\pi_t^i$ satisfies the mean square integrable property, i.e. $\mathbb{E}(\int_0^T |\pi_t^i|^2 dt) < \infty$.

**Definition 4.4.1** (Trading strategy)**.** *A trading strategy (or portfolio) $\pi_t$ at time $t$ corresponding to the utility function (4.34) is defined as $\pi_t := (x_t, y_t) \in \mathcal{F}_s, t < s \leq T$, where $x_t$ and $y_t$ denote the fraction of wealth invested an individual stock $i$ and a riskless bond, respectively.*

**Theorem 4.4.1** (Existence and uniqueness of Nash equilibrium)**.** *Given strictly concave utility function $U_i$ as in (4.34) with $\gamma_i > 0$. Assume for all $i = 1, ..., n$ that $-1 \leq \theta_i \leq 1, 0 \leq$*

*$\mu_i \leq \bar{\mu}_i$ where $\bar{\mu}_i > 0$, $0 \leq \nu_i \leq \bar{\nu}_i$, and $0 \leq \sigma_i \leq \bar{\sigma}_i$. Define two constants*

$$\varphi_n := \frac{1}{n} \sum_{i=1}^{n} \frac{\mu_i}{\gamma_i} \frac{\sigma_i}{\sigma_i^2 + \nu_i^2 \left[1 + \left(\frac{1}{\gamma_i} - 1\right) \frac{\theta_i}{n}\right]} \tag{4.36}$$

*and*

$$\psi_n := \frac{1}{n} \sum_{i=1}^{n} \theta_i \left(\frac{1}{\gamma_i} - 1\right) \frac{\sigma_i^2}{\sigma_i^2 + \nu_i^2 \left[1 + \left(\frac{1}{\gamma_i} - 1\right) \frac{\theta_i}{n}\right]}. \tag{4.37}$$

*There are four cases:*

*(i) If $\psi_n \neq 1$, $\gamma_i \neq 1$, and $\gamma_i \neq 0$, there exists a unique constant Nash equilibrium given by*

$$\pi^{i,*} = \frac{\mu_i}{\gamma_i} \frac{1}{\sigma_i^2 + \nu_i^2 \left(1 - \frac{\theta_i}{n} + \frac{\theta_i}{n\gamma_i}\right)} + \theta_i \left(1 - \frac{1}{\gamma_i}\right) \frac{\sigma_i}{\sigma_i^2 + \nu_i^2 \left(1 - \frac{\theta_i}{n} + \frac{\theta_i}{n\gamma_i}\right)} \frac{\varphi_n}{1 - \psi_n}. \tag{4.38}$$

*Moreover, we have the identity given by*

$$\mathbb{E}(\sigma_k \pi^{k,*}) = \bar{\sigma\alpha} = \frac{\varphi_n}{1 - \psi_n}.$$

*(ii) If $\psi_n \neq 1$ and $\gamma_i = 1$, there exists a unique constant Nash equilibrium given by*

$$\pi^{i,*} = \frac{\mu_i}{\sigma_i^2 + \nu_i^2}. \tag{4.39}$$

*(iii) If $\psi_n = 1$ and $\varphi_n \neq 0$, there is no constant Nash equilibrium.*
*(iv) If $\psi_n = 1$ and $\varphi_n = 0$, there are infinitely many constant Nash equilibria.*

*Proof.* Fix $i$. Denote $\alpha_k \in \mathbb{R}$, be the constant investment strategies followed by all other agents $k, k \neq i$. Let $X_t^k$ be the associated wealth processes, given by

$$dX_t^k = \alpha_k X_t^k (\mu_k dt + \nu_k dW_t^k + \sigma_k dB_t), \qquad X_0^k = x_0^k$$

where $x_0^k$ is the initial wealth of agent $k$. Using Itô's formula, we get

$$\begin{aligned} d(\log X_t^k) &= \left(\mu_k \alpha_k - \frac{1}{2}(\nu_k^2 \alpha_k^2 + \sigma_k^2 \alpha_k^2)\right) dt + \nu_k \alpha_k dW_t^k + \sigma_k \alpha_k dB_t \\ &= \left(\mu_k \alpha_k - \frac{1}{2}\omega_k \alpha_k^2\right) + \nu_k \alpha_k dW_t^k + \sigma_k \alpha_k dB_t \end{aligned}$$

where $\omega_k = \nu_k^2 + \sigma_k^2$. Define

$$Y_t := \left(\prod_{k\neq i} X_t^k\right)^{1/n}$$

where $X_t^k$ solves (4.35) with constant fractions $\alpha_k$. Thus, we have

$$\log Y_t = \frac{1}{n}\log\left(\prod_{k\neq i} X_t^k\right) = \frac{1}{n}\sum_{k\neq i}\log X_t^k.$$

Then,

$$\begin{aligned}
d\log Y_t &= \frac{1}{n}\sum_{k\neq i} d\log X_t^k \\
&= \frac{1}{n}\sum_{k\neq i}\left[\left(\mu_k\alpha_k - \frac{1}{2}\omega_k\alpha_k^2\right)dt + \nu_k\alpha_k dW_t^k + \sigma_k\alpha_k dB_t\right] \\
&= \left(\widehat{\mu\alpha} - \frac{1}{2}\widehat{\omega\alpha^2}\right) + \frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k + \widehat{\sigma\alpha}dB_t,
\end{aligned}$$

where we denote

$$\begin{aligned}
\widehat{\mu\alpha} &:= \frac{1}{n}\sum_{k\neq i}\mu_k\alpha_k, & \widehat{\sigma\alpha} &:= \frac{1}{n}\sum_{k\neq i}\sigma_k\alpha_k \\
\widehat{\omega\alpha^2} &:= \frac{1}{n}\sum_{k\neq i}\omega_k\alpha_k^2 \quad \text{and} & \widehat{(\nu\alpha)^2} &:= \frac{1}{n}\sum_{k\neq i}\nu_k^2\alpha_k^2
\end{aligned}$$

Hence, the process $Y_t$ solves

$$\frac{dY_t}{Y_t} = \eta dt + \frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k + \widehat{\sigma\alpha}dB_t, \quad Y_0 = \left(\prod_{k\neq i} x_0^k\right)^{1/n},$$

where

$$\eta = \widehat{\mu\alpha} - \frac{1}{2}\left[\widehat{\omega\alpha^2} - \widehat{\sigma\alpha}^2 - \frac{1}{n}\widehat{(\nu\alpha)^2}\right].$$

From the first case of (4.34), we have

$$
\begin{aligned}
U_i(X_T^i, \bar{X}_T) &= \frac{1}{1-\gamma_i}\left(X_T^i \bar{X}_T^{-\theta_i}\right)^{1-\gamma_i} \\
&= \frac{1}{1-\gamma_i}\left(X_T^i\left[\left(X_T^i \prod_{k\neq i} X_T^k\right)^{1/n}\right]^{-\theta_i}\right)^{1-\gamma_i} \\
&= \frac{1}{1-\gamma_i}\left(\left(X_T^i\right)^{1-\theta_i/n}\left[\left(\prod_{k\neq i} X_T^k\right)^{1/n}\right]^{-\theta_i}\right)^{1-\gamma_i} \\
&= \frac{1}{1-\gamma_i}\left[\left(X_T^i\right)^{1-\theta_i/n} Y_T^{-\theta_i}\right]^{1-\gamma_i}.
\end{aligned}
$$

The goal of the agent $i$ is to solve the following optimization problem

$$
\sup_{\pi^i\in\mathcal{A}} \mathbb{E}\frac{1}{1-\gamma_i}\left[\left(X_T^i\right)^{1-\theta_i/n} Y_T^{-\theta_i}\right]^{1-\gamma_i}
$$

with respect to the following two dynamics (one of agent $i$ and one of all the other agents $k, k \neq i$)

$$
\begin{aligned}
dX_t^i &= \pi_t^i X_t^i(\mu_i dt + \nu_i dW_t^i + \sigma_i dB_t), \quad X_0^i = x_0^i \\
dY_t &= Y_t\left(\eta dt + \frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k + \widehat{\sigma\alpha} dB_t\right), \quad Y_0 = \left(\prod_{k\neq i} x_0^k\right)^{1/n}.
\end{aligned}
$$

**Case 1**: Suppose $\gamma_i \neq 1$. Making the ansatz of the value function $V(X_t^i, Y_t, t)$

$$
V(X_t^i, Y_t, t) = f(t)\frac{1}{1-\gamma_i}\left[\left(X_t^i\right)^{1-\theta_i/n} Y_t^{-\theta_i}\right]^{1-\gamma_i}, \quad t\in[0,T]. \tag{4.40}
$$

For simplicity, denote $x = X_t^i$ and $y = Y_t$. Since we are looking for the constant Nash equilibrium, then $\pi_t^i = \pi^i$ for all $t \in [0,T]$. Similarly to that of Theorem 4.3.1, by using the multivariable Taylor series expansion

$$
\begin{aligned}
dV &= V_t dt + V_x dx + V_y dy + \frac{1}{2}V_{xx}(dx)^2 + \frac{1}{2}V_{yy}(dy)^2 + V_{xy}dxdy \\
&= V_t dt + V_x \left[\pi^i x(\mu_i dt + \nu_i dW_t^i + \sigma_i dB_t)\right] + V_y \left[ y\Big(\eta dt + \frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k + \widehat{\sigma\alpha}dB_t\Big)\right] \\
&+ \frac{1}{2}V_{xx}\left[\pi^i x(\mu_i dt + \nu_i dW_t^i + \sigma_i dB_t)\right]^2 + \frac{1}{2}V_{yy}\left[ y\left(\eta dt + \frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k + \widehat{\sigma\alpha}dB_t\right)\right]^2 \\
&+ V_{xy}\left[\pi^i x(\mu_i dt + \nu_i dW_t^i + \sigma_i dB_t)\right]\left[ y\Big(\eta dt + \frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k + \widehat{\sigma\alpha}dB_t\Big)\right] \\
&= V_t dt + \pi^i\mu_i x V_x + V_x(\nu_i dW_t^i + \sigma_i dB_t) + \eta y V_y + V_y\left[ y\left(\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k + \widehat{\sigma\alpha}dB_t\right)\right] \\
&+ \frac{1}{2}V_{xx}\left[\left(\sigma_i^2(dB_t)^2 + \nu_i^2(dW_t^i)^2\right)\right](\pi^i)^2x^2 + \frac{1}{2}V_{xx}(\mu_i\pi^i X_t^i)^2(dt)^2 \\
&+ V_{xx}\left[\left(\pi^i x\right)^2(\mu_i\nu_i dt dW_t^i + \mu_i\sigma_i dt dB_t + \nu_i\sigma_i dW_t^i dB_t)\right] + \frac{1}{2}V_{yy}\left[\widehat{\sigma\alpha}^2 + \frac{1}{n}(\frac{1}{n}\sum_{k\neq i}\nu_k^2\alpha_k^2)\right] y \\
&+ \frac{1}{2}V_{yy}(\eta y)^2(dt)^2 + V_{yy}\left[\eta\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dt dW_t^k + \eta\widehat{\sigma\alpha}dt dB_t + \widehat{\sigma\alpha}\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k dB_t\right] y \\
&+ V_{xy}\pi^i\sigma_i\widehat{\sigma\alpha}xy(dB_t)^2 + V_{xy}\left[(\pi^i xy)\left(\mu_i\eta(dt)^2 + \mu_i\frac{1}{n}\sum_{k\neq i}dt dW_t^i + \mu_i\widehat{\sigma\alpha}dt dB_t + \nu_i\eta dt dW_t^i\right)\right] \\
&+ V_{xy}\left[(\pi^i xy)\left(\nu_i\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^i dW_t^k + \nu_i\widehat{\sigma\alpha}dW_t^i dB_t + \sigma_i\eta dt dB_t + \sigma_i\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dB_t dW_t^k\right)\right]
\end{aligned}
$$

According to the rules $dt.dt = dt.dB_t = dB_t.dt = 0$, we obtain

$$
\begin{aligned}
dV &= V_t dt + \pi^i\mu_i x V_x + V_x(\nu_i dW_t^i + \sigma_i dB_t) + \eta y V_y + V_y\left[ y\left(\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k + \widehat{\sigma\alpha}dB_t\right)\right] \\
&+ \frac{1}{2}V_{xx}\left[\left(\sigma_i^2(dB_t)^2 + \nu_i^2(dW_t^i)^2\right)\right](\pi^i)^2x^2 + V_{xx}\left(\pi^i x\right)^2(\nu_i\sigma_i dW_t^i dB_t) \\
&+ \frac{1}{2}V_{yy}\left[\widehat{\sigma\alpha}^2 + \frac{1}{n}(\frac{1}{n}\sum_{k\neq i}\nu_k^2\alpha_k^2)\right] y^2 + V_{yy}\left[\widehat{\sigma\alpha}\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^k dB_t\right] y^2 + V_{xy}\pi^i\sigma_i\widehat{\sigma\alpha}xy(dB_t)^2 \\
&+ V_{xy}\left[(\pi^i xy)\left(\nu_i\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dW_t^i dW_t^k + \nu_i\widehat{\sigma\alpha}dW_t^i dB_t + \sigma_i\frac{1}{n}\sum_{k\neq i}\nu_k\alpha_k dB_t dW_t^k\right)\right].
\end{aligned}
$$

Since $E[dW_t^i] = E[dB_t] = 0, E[(dW_t^i)^2] = E[(dB_t)^2] = dt$, and the fact that $W_t^i, W_t^k$ and $B_t$ are independent, taking expectations and rearranging the equality above, we get

$$\begin{aligned} dV &= V_t dt + \pi^i \mu_i x V_x + \eta y V_y + \frac{1}{2} V_{xx} \left[ \left( \sigma_i^2 dt + \nu_i^2 dt \right) \right] (\pi^i)^2 x^2 + V_{xy} \pi^i \sigma_i \widehat{\sigma\alpha} xy dt \\ &\quad + \frac{1}{2} \left[ \widehat{\sigma\alpha}^2 + \frac{1}{n} \widehat{(\nu\alpha)^2} \right] y^2 V_{yy}. \end{aligned}$$

Divide both sides of above equality by $dt$ and rearranging, we get

$$\dot{V} = V_t + \frac{1}{2}(\sigma_i^2 + \nu_i^2)(\pi^i)^2 x^2 V_{xx} + \pi^i(\mu_i x V_x + \sigma_i \widehat{\sigma\alpha} xy V_{xy}) + \eta y V_y + \frac{1}{2} \left[ \widehat{\sigma\alpha}^2 + \frac{1}{n} \widehat{(\nu\alpha)^2} \right] y^2 V_{yy}.$$

At equilibrium, $V$ is optimal and then $\dot{V} = 0$. Thus, we get the HJB equation

$$V_t + \frac{1}{2}(\sigma_i^2 + \nu_i^2)(\pi^i)^2 x^2 V_{xx} + \pi^i(\mu_i x V_x + \sigma_i \widehat{\sigma\alpha} xy V_{xy}) + \eta y V_y + \frac{1}{2} \left[ \widehat{\sigma\alpha}^2 + \frac{1}{n} \widehat{(\nu\alpha)^2} \right] y^2 V_{yy} = 0. \tag{4.41}$$

for $(x, y, t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times [0, T]$, with terminal condition

$$V(x, y, T) = f(t) \frac{1}{1 - \gamma_i} \left[ \left( X_T^i \right)^{1-\theta_i/n} Y_T^{-\theta_i} \right]^{1-\gamma_i}.$$

Taking the first-order condition of (4.41) with respect to $\pi^i$, then solving for $\pi^i$, we obtain the agent $i$'s implicit optimal investment strategy

$$\pi^{i,*} = -\frac{\mu_i x V_x + \sigma_i \widehat{\sigma\alpha} xy V_{xy}}{(\sigma_i^2 + \nu_i^2) x^2 V_{xx}}. \tag{4.42}$$

Then, plug $\pi^{i,*}$ in the above equality into the HJB equation (4.41), we get

$$V_t - \frac{1}{2} \frac{(\mu_i x V_x + \sigma_i \widehat{\sigma\alpha} xy V_{xy})^2}{(\sigma_i^2 + \nu_i^2) x^2 V_{xx}} + \frac{1}{2} \left[ \widehat{\sigma\alpha}^2 + \frac{1}{n} \widehat{(\nu\alpha)^2} \right] y^2 V_{yy} + \eta y V_y = 0. \tag{4.43}$$

Taking derivatives the ansatz (4.40) to get $V_t, V_x, V_y, V_{xx}, V_{xy}$, and $V_{yy}$ then plugging into above equality, we get

$$\frac{1}{1 - \gamma_i} f'(t) + \rho f(t) = 0, \quad \forall t \in [0, T]$$

where

$$\rho = \frac{\left[ \mu_i (1 - \frac{\theta_i}{n}) - \sigma_i \widehat{\sigma\alpha} \theta_i (1 - \frac{\theta_i}{n})(1 - \gamma_i) \right]^2}{2(\sigma^2 + \nu_i^2)(1 - \frac{\theta_i}{n}) \left[ 1 - (1 - \frac{\theta_i}{n})(1 - \gamma_i) \right]} + \frac{1}{2} \left( \widehat{\sigma\alpha}^2 + \frac{1}{n} \widehat{(\nu\alpha)^2} \right) \theta_i \left[ 1 + \theta_i (1 - \gamma_i) \right] - \eta \theta_i.$$

Thus, we get $f(t) = e^{\rho(1-\gamma_i)(T-t)}$. Hence,

$$V(x,y,t) = e^{\rho(1-\gamma_i)(T-t)} \frac{1}{1-\gamma_i} \left(x^{1-\theta_i/n} y^{-\theta_i}\right)^{1-\gamma_i}$$

Then, taking derivatives $V_x, V_{xx}$, and $V_{xy}$ and plugging into (4.42), we obtain the agent $i$'s explicit optimal investment strategy

$$\pi^{i,*} = \frac{\mu_i - \sigma_i \widehat{\sigma\alpha} \theta_i (1-\gamma_i)}{(\sigma_i^2 + \nu_i^2)[1 - (1-\theta_i/n)(1-\gamma_i)]}. \tag{4.44}$$

Define

$$\bar{\sigma\alpha} := \frac{1}{n} \sum_{k=1}^{n} \sigma_k \alpha_k.$$

Then, we have

$$\bar{\sigma\alpha} = \frac{1}{n} \sum_{k \neq i} \sigma_k \alpha_k + \frac{1}{n} \sigma_i \alpha_i = \widehat{\sigma\alpha} + \frac{1}{n} \sigma_i \alpha_i \to \widehat{\sigma\alpha} = \bar{\sigma\alpha} - \frac{1}{n} \sigma_i \alpha_i.$$

We want to find the constant vector $(\alpha_1, ..., \alpha_n)$, i.e. a constant strategy $\alpha_i$ for each agent $i$. So there must be $\alpha_i = \pi^{i,*}$ for every $i = 1, ..., n$. Thus, we have

$$\begin{aligned}
\alpha_i &= \frac{\mu_i - \sigma_i \widehat{\sigma\alpha} \theta_i (1-\gamma_i)}{(\sigma_i^2 + \nu_i^2)[1 - (1-\theta_i/n)(1-\gamma_i)]} \\
&= \frac{\mu_i - \sigma_i \bar{\sigma\alpha} \theta_i (1-\gamma_i)}{(\sigma_i^2 + \nu_i^2)[1 - (1-\theta_i/n)(1-\gamma_i)]} + \frac{\sigma_i^2 \alpha_i (\theta_i/n)(1-\gamma_i)}{(\sigma_i^2 + \nu_i^2)[1 - (1-\theta_i/n)(1-\gamma_i)]}
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
\alpha_i &= \frac{\mu_i - \sigma_i \bar{\sigma\alpha} \theta_i (1-\gamma_i)}{(\sigma_i^2 + \nu_i^2)[1 - (1-\theta_i/n)(1-\gamma_i)]} \left(1 - \frac{\sigma_i^2 (\theta_i/n)(1-\gamma_i)}{(\sigma_i^2 + \nu_i^2)[1 - (1-\theta_i/n)(1-\gamma_i)]}\right)^{-1} \\
&= \frac{\mu_i - \sigma_i \bar{\sigma\alpha} \theta_i (1-\gamma_i)}{(\sigma_i^2 + \nu_i^2)[1 - (1-\theta_i/n)(1-\gamma_i)] - \sigma_i^2 (\theta_i/n)(1-\gamma_i)} \\
&= \frac{\mu_i}{\sigma_i^2 \gamma_i + \nu_i^2 [1 - (1-\theta_i/n)(1-\gamma_i)]} - \frac{\sigma_i \bar{\sigma\alpha} \theta_i (1-\gamma_i)}{\sigma_i^2 \gamma_i + \nu_i^2 [1 - (1-\theta_i/n)(1-\gamma_i)]} \\
&= \frac{\mu_i}{\gamma_i} \frac{1}{\sigma_i^2 + \nu_i^2 \left(1 - \frac{\theta_i}{n} + \frac{\theta_i}{n\gamma_i}\right)} + \theta_i \left(1 - \frac{1}{\gamma_i}\right) \frac{\sigma_i}{\sigma_i^2 + \nu_i^2 \left(1 - \frac{\theta_i}{n} + \frac{\theta_i}{n\gamma_i}\right)} \bar{\sigma\alpha}.
\end{aligned} \tag{4.45}$$

Multiplying both sides by $\sigma_i$ and then averaging the above quality for $i = 1, ..., n$, we get

$$\bar{\sigma\alpha} = \varphi_n + \psi_n \bar{\sigma\alpha},$$

where $\varphi_n$ and $\psi_n$ defined in (4.36) and (4.37). There are three cases:

(i) If $\psi_n \neq 1$, $\gamma_i \neq 1$, and $\gamma_i \neq 0$, there exists a unique constant Nash equilibrium given by

$$\pi^{i,*} = \frac{\mu_i}{\gamma_i} \frac{1}{\sigma_i^2 + \nu_i^2 \left(1 - \frac{\theta_i}{n} + \frac{\theta_i}{n\gamma_i}\right)} + \theta_i \left(1 - \frac{1}{\gamma_i}\right) \frac{\sigma_i}{\sigma_i^2 + \nu_i^2 \left(1 - \frac{\theta_i}{n} + \frac{\theta_i}{n\gamma_i}\right)} \frac{\varphi_n}{1 - \psi_n}.$$

Moreover, we have the identity given by

$$\mathbb{E}(\sigma_k \pi^{k,*}) = \bar{\sigma\alpha} = \frac{\varphi_n}{1 - \psi_n}.$$

(ii) If $\psi_n = 1$ and $\varphi_n \neq 0$, there is no constant Nash equilibrium.
(iii) If $\psi = 1$ and $\varphi_n = 0$, there are infinitely many constant Nash equilibria.

**Case 2**: Suppose $\gamma_i = 1$. Since $\gamma_i = 1 > 0$ then $U_{xx} < 0$. That means the logarithmic form utility function describes the behavior of risk-averse agents. Making the ansatz of the value function

$$V(x, y, t) = \log\left(x^{1-\theta_i/n} y^{-\theta_i}\right) + f(t) = \left(1 - \frac{\theta_i}{n}\right) \log x - \theta_i \log y + f(t).$$

Taking derivatives this ansatz to get $V_t, V_x, V_y, V_{xx}, V_{xy}$, and $V_{yy}$ and then plugging into (4.43)[12], we get

$$f'(t) + \rho = 0,$$

where

$$\rho = \frac{1}{2} \frac{\mu_i^2 (1 - \theta_i/n)}{(\sigma_i^2 + \nu_i^2)} + \frac{1}{2}\theta_i \left(\widehat{\sigma\alpha}^2 + \frac{1}{n}\widehat{(\nu\alpha)^2}\right) - \theta_i \eta.$$

Thus, we obtain $f(t) = \rho(T - t)$. Plugging into the above value function, we get

$$V(x, y, t) = \left(1 - \frac{\theta_i}{n}\right) \log x - \theta_i \log y + \rho(T - t).$$

Taking derivatives $V_x, V_{xx}$, and $V_{xy}$ and plugging into (4.42), we get the agent $i$'s explicit optimal investment strategy

$$\pi^{i,*} = \frac{\mu_i}{\sigma_i^2 + \nu_i^2} \qquad \text{for all } i = 1, ..., n,$$

which is a special case of (4.44) when $\gamma_i = 1$ and is exactly as (4.39). □

[12] Note that the equality (4.43) is the common result derived for both cases: power form and logarithmic form in (4.43). Therefore, it does still work for the case that $\gamma_i = 1$.

**Theorem 4.4.2** (Unique corner solution for strictly convex power utility function)**.** *Suppose $U_i$ the utility function of agent $i$ as in (4.34) with $\gamma_i < 0$. For the power case, the optimal solution is $\pi^{i,*} = 1$ for risky asset (stock) and $\pi^{i,*} = 0$ for riskless asset (bond).*

*Proof.* Most arguments in the proof of Theorem 4.3.2 for the exponential case are still correct in the power case. The only different point is the functional form of utility function.

For the power case, we have $U'_{X^i_T} = (X^i_T \bar{X}_T^{-\theta_i})^{-\gamma_i} \bar{X}_T^{-\theta_i} > 0$, which implies that $U_i$ is a strictly increasing function in $X^i_T$. Moreover, for $\gamma_i < 0$, we can check the second order sufficient condition that

$$U''_{X^i_T} = -\gamma_i \left( X^i_T \bar{X}_T^{-\theta_i} \right)^{-\gamma_i - 1} \bar{X}_T^{-2\theta_i} > 0,$$

which implies that $U_i$ is a strictly convex utility function. Thus, the maximum at $M$ must be unique and be on the upper boundary of the domain and $M$ must be unique. This implies that $\pi^{i,*} = 1$ for the risky asset (stock) and thus $\pi^{i,*} = 0$ for riskless asset (bond). Obviously, the corner solution is unique. □

**Remark 4.4.1** (Economic meaning of corner solution for strictly convex utility function)**.** *Theorem 4.3.2 and Theorem 4.4.2 contribute to the literature of asset specialization in the sense that individual risk-seeking investors only invest their whole wealth in one stock as a risky asset (or small portfolio with few stocks as a whole) and do not invest in riskless assets (e.g. bond) in the case of strictly convex utility function.*

### 4.4.2 The mean field games

In this section, we examine the limiting behavior of $n$-player games in the previous section when $n \to \infty$. We use the same assumption of continuum of agents as Assumption 4.3.1 and the same definition of MFG as Definition 4.3.6.

The dynamics of wealth of the representative agent is given by

$$dX_t = \pi_t X_t(\mu dt + \nu dW_t + \sigma dB_t), \quad X_0 = \xi, \tag{4.46}$$

where the investment fraction $\pi_t$ belongs to the admissible set $\mathcal{A}_{MF}$ and satisfies the mean square integrable property $\mathbb{E} \int_0^T |\pi_t|^2 dt < \infty$. The goal of the representative agent is to maximize the following expected utility

$$\sup_{\pi_t \in \mathcal{A}_{MF}} \mathbb{E} \left[ \frac{1}{1-\gamma} \left( X_T \bar{X}_T^{-\theta} \right)^{1-\gamma} \right] \tag{4.47}$$

given the dynamics (4.46). Here, $\bar{X}_T$ is the geometric mean which is defined as follows

$$\bar{X}_T := \exp \mathbb{E}\left[\log X_T \mid \mathcal{F}_T^B\right], \tag{4.48}$$

where $\mathcal{F}_T^B$ denotes the filtration generated by the Brownian motion $B$.

**Definition 4.4.2** (MFG)**.** *Given $\bar{X}_T$ defined as (4.48) and let $\pi_t^* \in \mathcal{A}_{MF}$ be an admissible strategy. Then, $\pi_t^*$ is a mean field equilibrium (MFE) if $\pi_t^*$ solves the optimization problem (4.47).*

**Definition 4.4.3** (Constant MFE)**.** *The optimal control $\pi^*$ is called a constant MFE if $\pi^* = \pi_t^*$ for all $t \in [0, T]$, where $\pi_t^*$ is a MFE defined in Definition 4.4.2.*

**Lemma 4.4.1.** *Define two constants*

$$\varphi := \mathbb{E}\left[\frac{\mu}{\gamma}\frac{\sigma}{\sigma^2+\nu^2}\right] \quad \text{and} \quad \psi := \mathbb{E}\left[\theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma^2}{\sigma^2+\nu^2}\right],$$

*where $\gamma \neq 0$. We claim that these two expectations exist and are finite.*

*Proof.* Denote

$$\Xi = \frac{\mu}{\gamma}\frac{\sigma}{\sigma^2+\nu^2} \quad \text{and} \quad \Lambda = \theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma^2}{\sigma^2+\nu^2}.$$

By assumptions $0 \le \mu \le \bar{\mu}, \gamma \neq 0, 0 \le \nu \le \bar{\nu}$, and $0 \le \sigma \le \bar{\sigma}$ then $\Xi$ must be exist and is finite, and so is $\varphi$.

By assumptions $-1 \le \theta \le 1$ and $\gamma \neq 0$, and note that $\sigma^2/(\sigma^2+\nu^2) \le 1$ then $\Lambda$ must be exist and is finite, and so is $\psi$. $\square$

**Theorem 4.4.3** (Existence and uniqueness of constant MFE)**.** *Given strictly concave utility function $U$. Assume for a.s. that $-1 \le \theta \le 1, 0 \le \mu \le \bar{\mu}, 0 \le \nu \le \bar{\nu}$, and $0 \le \sigma \le \bar{\sigma}$. Define two constants $\varphi$ and $\psi$ as that in Lemma 4.4.1.*

*There are four cases:*

*(i) If $\psi \neq 1$, $\gamma \neq 1$, and $\gamma \neq 0$, then there exists a unique constant MFE given by*

$$\pi^* = \frac{\mu}{\gamma}\frac{1}{\sigma^2+\nu^2} + \theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma}{\sigma^2+\nu^2}\frac{\varphi}{1-\psi}. \tag{4.49}$$

*Moreover, we have the identity given by*

$$\mathbb{E}(\sigma\pi^*) = \frac{\varphi}{1-\psi}.$$

*(ii) If $\psi \neq 1$ and $\gamma = 1$, there is a unique constant MFE given by*

$$\pi^* = \frac{\mu}{\sigma^2 + \nu^2}. \tag{4.50}$$

*(iii) If $\psi = 1$ and $\varphi \neq 0$, there is no constant MFE.*
*(iv) If $\psi = 1$ and $\varphi = 0$, there are infinitely many constant MFE.*

*Proof.* Define

$$Y_t := \exp \mathbb{E}(\log X_t^\alpha \mid \mathcal{F}_T^B),$$

where $\alpha$ is an admissible constant strategy. From the dynamics (4.46), using Itô's formula, we get

$$d(\log X_t^\alpha) = \left(\mu\alpha - \frac{1}{2}(\sigma^2 + \nu^2)\alpha^2\right) dt + \nu\alpha dW_t + \sigma\alpha dB_t.$$

Define $\hat{X}_t^\alpha := \mathbb{E}(\log X_t^\alpha \mid \mathcal{F}_T^B)$. Denote $\omega = \sigma^2 + \nu^2$. Then, since $\zeta = (\xi, \gamma, \theta, \mu, \nu, \sigma), W$, and $B$ are independent, we obtain

$$d\hat{X}_t^\alpha = \left(\bar{\mu\alpha} - \frac{1}{2}\bar{\omega\alpha^2}\right) dt + \bar{\sigma\alpha} dB_t,$$

Using Itô's formula one more time, we get

$$\begin{aligned} dY_t = de^{\hat{X}_t^\alpha} &= Y_t \left(\left[(\bar{\mu\alpha} - \frac{1}{2}\bar{w\alpha^2}) + \frac{1}{2}\bar{\sigma\alpha}^2\right] dt + \bar{\sigma\alpha} dB_t\right), \quad Y_0 = \bar{\xi}, \\ &= Y_t(\eta dt + \bar{\sigma\alpha} dB_t), \quad Y_0 = \bar{\xi}, \end{aligned} \tag{4.51}$$

where $\eta = \bar{\mu\alpha} - \frac{1}{2}(\bar{\omega\alpha^2} - \bar{\sigma\alpha}^2)$.

The goal of the representative agent is to solve the following problem

$$\sup_{\pi \in \mathcal{A}_{MF}} \mathbb{E} \frac{1}{1-\gamma} \left(X_T Y_T^{-\theta}\right)^{1-\gamma}.$$

with the given dynamics

$$dX_t = \pi_t X_t (\mu dt + \nu dW_t + \sigma dB_t),$$

and $Y_t, t \in [0, T]$, solving the dynamics (4.51).

Note that when $n \to \infty$, we have $Y_T \approx \bar{X}_T$. We consider two cases.

**Case 1**: Suppose $\gamma \neq 1$. Denote $x = X_t, y = Y_t$. Making the ansatz of the value function

$$V(x,y,t) = f(t)\frac{1}{1-\gamma}x^{1-\gamma}y^{-\theta(1-\gamma)}, \quad \forall t \in [0,T] \tag{4.52}$$

and note that when $n \to \infty$, we have $\widehat{\sigma\alpha} = \bar{\sigma\alpha} - \frac{1}{n}\sigma_i\alpha_i \approx \bar{\sigma\alpha}$. Then, following the similar steps as the proof of Theorem 4.4.1, we derive the HJB equation

$$V_t + \frac{1}{2}(\sigma^2+\nu^2)\pi^2x^2V_{xx} + \pi(\mu x V_x + \sigma\bar{\sigma\alpha}xyV_{xy}) + \eta y V_y + \frac{1}{2}\bar{\sigma\alpha}^2y^2V_{yy} = 0 \tag{4.53}$$

with the terminal condition $V(x,y,T) = \frac{1}{1-\gamma}(xy^{-\theta})^{1-\gamma}$.

Taking the first-order condition (4.53) with respect to $\pi$ and then manipulating, we get the implicit optimal investment strategy

$$\pi^* = -\frac{\mu x V_x + \sigma\bar{\sigma\alpha}xyV_{xy}}{(\sigma^2+\nu^2)x^2V_{xx}}. \tag{4.54}$$

Plugging back to (4.53), it reduces to

$$V_t - \frac{1}{2}\frac{(\mu x V_x + \sigma\bar{\sigma\alpha}xyV_{xy})^2}{(\sigma^2+\nu^2)x^2V_{xx}} + \eta y V_y + \frac{1}{2}\bar{\sigma\alpha}^2y^2V_{yy} = 0. \tag{4.55}$$

Taking derivatives the ansatz (4.52) to get $V_t, V_x, V_y, V_{xx}, V_{xy}$, and $V_{yy}$ then plugging into (4.55), we get $\frac{1}{1-\gamma}f'(t) + \rho f(t) = 0$, with $f(T) = 1$, where

$$\rho := \frac{[\mu - \theta(1-\gamma)\sigma\bar{\sigma\alpha}]^2}{2(\sigma^2+\nu^2)\gamma} - \eta\theta + \frac{1}{2}\bar{\sigma\alpha}^2\theta(\theta(1-\gamma)+1).$$

Thus, we obtain $f(t) = e^{\rho(1-\gamma)(T-t)}$. Plugging into (4.52), we get

$$V(x,y,t) = e^{\rho(1-\gamma)(T-t)}\frac{1}{1-\gamma}x^{1-\gamma}y^{-\theta(1-\gamma)}.$$

Then, taking derivatives $V_x, V_{xx}$, and $V_{xy}$ and plugging into (4.54), we obtain the explicit optimal investment strategy

$$\pi^* = \frac{\mu}{\gamma}\frac{1}{\sigma^2+\nu^2} + \theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma}{\sigma^2+\nu^2}\bar{\sigma\alpha} \tag{4.56}$$

We want to find the constant vector $(\alpha_1, ..., \alpha_n)$. So there must be $\alpha_i = \pi^*$ for every $i = 1, ..., n$. Multiplying both sides by $\sigma$ and averaging (4.56), we get

$$\bar{\sigma\pi^*} = \mathbb{E}\left[\frac{\mu}{\gamma}\frac{\sigma}{\sigma^2+\nu^2}\right] + \mathbb{E}\left[\theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma}{\sigma^2+\nu^2}\right]\bar{\sigma\pi^*} = \varphi + \psi\bar{\sigma\pi^*}.$$

There are three cases:

(i) If $\psi \neq 1$, $\gamma \neq 1$, and $\gamma \neq 0$, there exists a unique constant MFE given by

$$\pi^* = \frac{\mu}{\gamma} \frac{1}{\sigma^2 + \nu^2} + \theta \left(1 - \frac{1}{\gamma}\right) \frac{\sigma}{\sigma^2 + \nu^2} \frac{\varphi}{1 - \psi}.$$

Moreover, we have the identity given by

$$\mathbb{E}(\sigma\pi^*) = \sigma\bar{\pi}^* = \bar{\sigma\alpha} = \frac{\varphi}{1 - \psi}.$$

(ii) If $\psi = 1$ and $\varphi \neq 0$, there is no constant MFE.
(iii) If $\psi = 1$ and $\varphi = 0$, there are infinitely many constant MFE.

**Case 2**: Suppose $\gamma = 1$. Note that in this case since the logarithmic utility function is concave, it only describes the behavior of risk-averse agents.

Making the ansatz of the value function

$$V(x, y, t) = \log x - \theta \log y + f(t).$$

Following the similar steps as the proof of Theorem 4.4.1, we obtain $f(t) = \rho(T - t)$. Hence, we get

$$V(x, y, t) = \log x - \theta \log y + \rho(T - t).$$

Taking derivatives $V_x, V_{xx}$, and $V_{xy}$ and plugging into (4.54), we get the explicit optimal investment strategy

$$\pi^* = \frac{\mu}{\sigma^2 + \nu^2},$$

which is exactly (4.50). □

**MFE for strictly convex power utility function**. The optimal strategy for Theorem 4.4.2 does not depend on number of players $n$ and also $i$. Therefore, the optimal strategy for the case of $n$-agent games (Nash equilibrium) and that for the case of MFG (mean field equilibrium) should be the same. That is, $\pi^* = 1$ for risky asset (stock) and $\pi^* = 0$ for riskless asset (bond). Obviously, the corner solution is unique.

### 4.4.3 Convergence of MFE to Nash equilibria

The important link between Nash equilibria and MFE in this study is that MFE of mean field games is the convergent result of the Nash equilibrium in the corresponding $N$-agent game when $n \to \infty$.

**Theorem 4.4.4** (Convergence)**.** *The MFE* (4.24) *and* (4.49) *are the limits as* $n \to \infty$ *of* $n$*-player Nash equilibria* (4.11) *and* (4.38)*, respectively.*

*Proof.* It is obvious by the ways we have constructed the MFG based on the corresponding $N$-agent games. $\square$

In the case of the logarithmic utility function, the Nash equilibrium and the MFE have very similar structures, which respectively are

$$\pi^{i,*} = \frac{\mu_i}{\sigma_i^2 + \nu_i^2} \quad \text{and} \quad \pi^* = \frac{\mu}{\sigma^2 + \nu^2}.$$

These two optimal investment strategies are always positive. That means risk-averse agents with logarithmic utility functions always invest a positive fraction of their wealth in their individual stock either in games with finite or infinite agents.

We can observe that the optimal strategy (4.38) and (4.11), and (4.49) and (4.24) have similar structures, respectively. Note that in (4.11), $\nu_i^2$ is scaled by $1 - \theta_i/n$ compared to that in (4.38) and in (4.24), $\nu_i^2$ is scaled by $1 - \frac{\theta_i}{n} + \frac{\theta_i}{n\gamma_i}$ compared to that in (4.49). A noticeable point is that all of the above optimal strategies do not depend on wealth. Specifically, the MFE only depends on the representative individual preferences and market parameters. Meanwhile, the Nash equilibria depend on individual preferences and market parameters, and the number of players.

Note that in the case there is no relative performance, i.e. $\theta = 0$, the optimal investment strategy (4.49) is equal to

$$\pi^* = \frac{\mu}{\gamma} \frac{1}{\sigma^2 + \nu^2},$$

which is the classical Merton portfolio in Merton (1969) when there is only one Brownian process and bond (sure asset) offers zero interest rate.

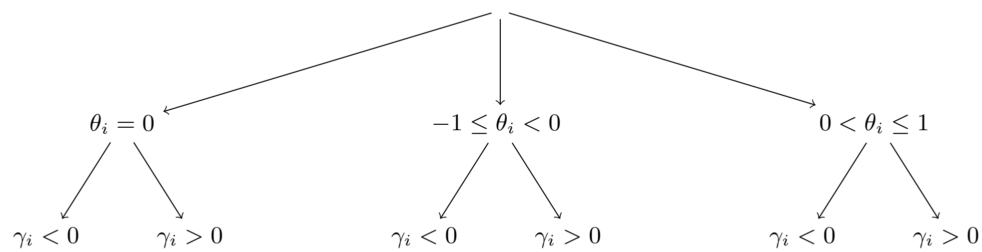
## 4.5 Discussion

Main theorems for the case of strictly concave utility function in this chapter including Theorem 4.3.1, Theorem 4.3.3, Theorem 4.4.1, Theorem 4.4.3 and their proofs are largely inspired by the frameworks of Theorem 2.3, Theorem 2.10, Theorem 3.1, and Theorem 3.6 and their proofs in Lacker and Zariphopoulou (2019), respectively, but our results are more general. The effects of the personal and market parameters in our study are much more complicated than the ideal case of risk aversion in Lacker and Zariphopoulou (2019). We employ this section to discuss the qualitative effects of the number of players in $n$-agent games, relative performance, and risk tolerance parameter on the optimal strategies.

### 4.5.1 The effects of the number agents in $n$-agent games

The stock market is a complex system. Therefore, the effects of the number of agents in $n$-agent games on optimal strategies in both exponential and power/logarithmic cases are not straightforward. There are six cases illustrated in the following tree.

Figure 4.1: Tree with 6 cases depending on personal parameters




Note that Lacker and Zariphopoulou (2019) only considers two cases among these six cases, which are $0 < \theta_i \leq 1, \gamma_i > 0$ and $\theta_i = 0, \gamma_i > 0$. That means, they only study the risk-averse agents. Rather, we analyze the behavior of both risk-averse and risk-seeking agents.

#### Exponential case

For the sake of convenient analysis, we restate the optimal strategy of the $n$-agent games in the exponential case.

$$\pi^{i,*} = \frac{\mu_i}{\gamma_i} \frac{1}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} + \theta_i \frac{\sigma_i}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} \frac{\varphi_n}{1-\psi_n},$$

where

$$\varphi_n = \frac{1}{n}\sum_{i=1}^{n} \frac{\mu_i}{\gamma_i} \frac{\sigma_i}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} \quad \text{and} \quad \psi_n = \frac{1}{n}\sum_{i=1}^{n} \left[\theta_i \frac{\sigma_i^2}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)}\right],$$

and $0 \leq \mu_i \leq \bar{\mu}_i$, $0 \leq \nu_i \leq \bar{\nu}_i, 0 \leq \sigma_i \leq \bar{\sigma}_i$, and $-1 \leq \theta_i \leq 1$ for all $i = 1, ..., n$.

We will only analyze some specific cases. For simplicity, denote

$$M_1 = \frac{\mu_i}{\gamma_i} \frac{1}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} \qquad \text{and} \qquad M_2 = \theta_i \frac{\sigma_i}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} \frac{\varphi_n}{1-\psi_n}.$$

**Case 1**: There is no relative performance, i.e. $\theta_i = 0$. Thus,

$$\pi^{i,*} = \frac{\mu_i}{\gamma_i} \frac{1}{\sigma_i^2 + \nu_i^2}.$$

In this case, the optimal investment strategy does not depend on the number of agents. This makes sense since in this case agent $i$ is not concerned or motivated by other agents (i.e. $\theta_i = 0$). Moreover, if $\gamma_i > 0$ then $\pi^{i,*} > 0$ and if $\gamma_i < 0$ then $\pi^{i,*} < 0$.

**Case 2**: $0 < \theta_i \le 1, \gamma_i > 0$. This is the situation where risk-averse agents with relative performance concern. In this case, we can check that $\pi^{i,*} > 0$. Moreover, an increase in $n$ leads to a decrease in the first term of $\pi^{i,*}$,

$$n \uparrow \longrightarrow \frac{\theta_i}{n} \downarrow \longrightarrow \left(1 - \frac{\theta_i}{n}\right) \uparrow \longrightarrow \frac{1}{\sigma_i^2 + \nu_i^2(1-\theta_i/n)} \downarrow \longrightarrow M_1 \downarrow .$$

Since $\mu_i \ge 0$ and $\gamma_i > 0$, then $\varphi_n \ge 0$. Also, since $0 < \theta_i \le 1$, then $0 < \psi_n < 1$, and then $1 - \psi_n > 0$. Consequently $\varphi_n/(1-\psi_n) \ge 0$. Moreover,

$$\begin{cases} n \uparrow \longrightarrow \frac{\theta_i}{n} \downarrow \longrightarrow \left(1 - \frac{\theta_i}{n}\right) \uparrow \longrightarrow \frac{\sigma_i}{\sigma_i^2+\nu_i^2(1-\theta_i/n)} \downarrow \longrightarrow \sum_{i=1}^n \frac{\mu_i}{\gamma_i} \frac{\sigma_i}{\sigma_i^2+\nu_i^2(1-\theta_i/n)} \downarrow \\ n \uparrow \longrightarrow \frac{1}{n} \downarrow \end{cases} \longrightarrow \varphi_n \downarrow,$$

and

$$\begin{cases} n \uparrow \longrightarrow \frac{\theta_i}{n} \downarrow \longrightarrow \left(1 - \frac{\theta_i}{n}\right) \uparrow \longrightarrow \frac{\sigma_i^2}{\sigma_i^2+\nu_i^2(1-\theta_i/n)} \downarrow \longrightarrow \sum_{i=1}^n \theta_i \frac{\sigma_i^2}{\sigma_i^2+\nu_i^2(1-\theta_i/n)} \downarrow \\ n \uparrow \longrightarrow \frac{1}{n} \downarrow \end{cases} \longrightarrow \psi_n \downarrow \longrightarrow 1 - \psi_n \uparrow .$$

Thus, $n$ increases leading to $\varphi_n/(1-\psi_n)$ decreases. Hence, an increase in $n$ leads to a decrease in the second term of $\pi^{i,*}$ (i.e. $M_2 \downarrow$ ) as well as the first term of $\pi^{i,*}$ (i.e. $M_1 \downarrow$). As a result, an increase in the number of agents $n$ leads to a decrease in the absolute value of wealth that agent $i$ invests in the individual risky stock $i$. This might be because an increase in the number of agents accelerates the level of competition in the market.

**Case 3**: $0 < \theta_i \le 1, \gamma_i < 0$. This is the situation where risk-seeking agents concern with relative performance. Assume further that $\mu > 0$. In this case, since $\mu_i > 0$ and $\gamma_i < 0$ then $\mu_i/\gamma_i < 0$. Thus, $\varphi_n < 0$. Moreover, since $0 < \theta_i \le 1$, then $0 < \psi_n \le 1$, and then $0 \le 1 - \psi_n < 1$. Since the optimal strategy requires that $\psi_n \ne 1$ or $1 - \psi_n \ne 0$, we get $\varphi_n/(1-\psi_n) < 0$. As a result, $M_1 < 0$ and $M_2 < 0$, and then $\pi^{i,*} < 0$. Moreover,

$$n \uparrow \longrightarrow \begin{cases} |\pi^{i,*}| \uparrow \quad \text{if} \quad |\Delta M_1| + |\Delta M_2| > 0 \\ |\pi^{i,*}| \downarrow \quad \text{if} \quad |\Delta M_1| + |\Delta M_2| < 0. \end{cases}$$

**Case 4**:$-1 \leq \theta_i < 0, \gamma_i > 0$. Then, $M_1 > 0$ and $M_2 < 0$. Thus, $\pi^{i,*}$ can be positive or negative,

$$\pi^{i,*} \text{ is } \begin{cases} \text{positive }, & \text{if} \quad \mathrm{M}_1 > |M_2| \\ \text{negative }, & \text{if} \quad \mathrm{M}_1 < |M_2| \\ 0, & \text{if} \quad \mathrm{M}_1 = |M_2|. \end{cases}$$

An increase in $n$ leads to

$$\pi^{i,*} \begin{cases} \text{increases}, & \text{if} \quad \Delta M_1 > |\Delta M_2| \\ \text{decreases}, & \text{if} \quad \Delta M_1 < |\Delta M_2| \\ \text{remains unchanged}, & \text{if} \quad \Delta M_1 = |\Delta M_2|. \end{cases}$$

**Power and logarithmic case**

Recall the optimal strategy of the $n$-agent games in the power/logarithmic case given by

$$\pi^{i,*} = \frac{\mu_i}{\gamma_i} \frac{1}{\sigma_i^2 + \nu_i^2 \left(1 - \frac{\theta_i}{n} + \frac{\theta_i}{n\gamma_i}\right)} + \theta_i \left(1 - \frac{1}{\gamma_i}\right) \frac{\sigma_i}{\sigma_i^2 + \nu_i^2 \left(1 - \frac{\theta_i}{n} + \frac{\theta_i}{n\gamma_i}\right)} \frac{\varphi_n}{1 - \psi_n},$$

where

$$\varphi_n := \frac{1}{n} \sum_{i=1}^{n} \frac{\mu_i}{\gamma_i} \frac{\sigma_i}{\sigma_i^2 + \nu_i^2 \left[1 + \left(\frac{1}{\gamma_i} - 1\right) \frac{\theta_i}{n}\right]}, \quad \psi_n := \frac{1}{n} \sum_{i=1}^{n} \theta_i \left(\frac{1}{\gamma_i} - 1\right) \frac{\sigma_i^2}{\sigma_i^2 + \nu_i^2 \left[1 + \left(\frac{1}{\gamma_i} - 1\right) \frac{\theta_i}{n}\right]},$$

and $-1 \leq \theta_i \leq 1, 0 \leq \mu_i \leq \bar{\mu}_i$, $0 \leq \nu_i \leq \bar{\nu}_i$, and $0 \leq \sigma_i \leq \bar{\sigma}_i$ for all $i = 1, ..., n$, and with the conditions that $\psi_n \neq 1$, $\gamma_i \neq 1$, and $\gamma_i \neq 0$.

**Case 1**: $\theta_i = 0$, i.e. there is no relative performance parameter. Then,

$$\pi^{i,*} = \frac{\mu_i}{\gamma_i} \frac{1}{\sigma_i^2 + \nu_i^2}.$$

Obviously, this optimal strategy does not depend on the number of agents $n$. Moreover, it coincides with that of the exponential utility function case.

**Case 2**: $\gamma_i = 1$. The optimal strategy does not depend on $n$ and is given by

$$\pi^{i,*} = \frac{\mu_i}{\sigma_i^2 + \nu_i^2}.$$

### 4.5.2 The effects of relative performance

In order to examine the effects of $\theta$ on $\pi^*$, we rewrite the MFE in the exponential and power case respectively as follows.

$$\pi^{*,\exp} = \frac{\mu}{\gamma}\frac{1}{\sigma^2+\nu^2} + \frac{\theta}{1-\mathbb{E}\left(\theta\frac{\sigma^2}{\sigma^2+\nu^2}\right)}\frac{\sigma}{\sigma^2+\nu^2}\varphi, \quad \text{where} \quad \varphi := \mathbb{E}\left(\frac{\mu}{\gamma}\frac{\sigma}{\sigma^2+\nu^2}\right)$$

and

$$\pi^{*,\mathrm{pow}} = \frac{\mu}{\gamma}\frac{1}{\sigma^2+\nu^2} + \frac{\theta}{1-\mathbb{E}\left[\theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma^2}{\sigma^2+\nu^2}\right]}\left(1-\frac{1}{\gamma}\right)\frac{\sigma}{\sigma^2+\nu^2}\varphi, \quad \text{where} \quad \varphi := \mathbb{E}\left(\frac{\mu}{\gamma}\frac{\sigma}{\sigma^2+\nu^2}\right).$$

For simplicity, assume

$$\mathbb{E}\left(\theta\frac{\sigma^2}{\sigma^2+\nu^2}\right) = \theta\frac{\sigma^2}{\sigma^2+\nu^2}, \quad \text{and} \quad \mathbb{E}\left[\theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma^2}{\sigma^2+\nu^2}\right] = \theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma^2}{\sigma^2+\nu^2}.$$

Thus, holding other things constant, we can write $\pi^*$ as a function of $\theta$, i.e. $\pi^* = \pi^*(\theta)$. Taking derivative with respect to $\theta$ for both exponential and power cases, respectively, yield

$$\pi_\theta^{*,\exp} = \frac{\sigma}{(1-\theta)\sigma^2+\nu^2}\varphi \quad \text{and} \quad \pi_\theta^{*,\mathrm{pow}} = \frac{1}{1-\theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma^2}{\sigma^2+\nu^2}}\left(1-\frac{1}{\gamma}\right)\frac{\sigma}{\sigma^2+\nu^2}\varphi.$$

Assume further that $\sigma > 0$. Thus,

$$\frac{\sigma}{(1-\theta)\sigma^2+\nu^2} > 0, \quad \frac{1}{1-\theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma^2}{\sigma^2+\nu^2}} > 0, \quad \text{and} \quad \frac{\sigma}{\sigma^2+\nu^2} > 0.$$

Thus, the sign of $\pi_\theta^{*,\exp}$ depends on the sign of $\varphi$ and the sign of $\pi_\theta^{*,\mathrm{pow}}$ depends on the sign of $(1-1/\gamma)\varphi$.

We now analyze some specific cases to see the effects of the relative performance parameter $\theta$ on the optimal investment strategy $\pi^*$ in both exponential and power cases.

**Case 1**: $0 < \theta \leq 1, \gamma > 0$. In this case $\pi_\theta^{*,\exp} > 0$. Thus, an increase in $\theta$ leads to an increase in $\pi^{*,\exp}$ in the exponential case. If $\gamma > 1$, then an increase in $\theta$ leads to an increase in $\pi^{*,\mathrm{pow}}$ in the power case.

**Case 2**: $0 < \theta < 1, \gamma < 0$. In this case, we can check that $\pi_\theta^{*,\exp} < 0$. That is, an increase in $\theta$ leads to a decrease in $\pi^{*,\exp}$.

**Case 3**: $-1 \leq \theta < 0, \gamma > 0$. Assume further that $\mu > 0$. Hence, since $\varphi > 0$, then $\pi_\theta^{*,\exp} > 0$. Thus, an increase in $\theta$ leads to an increase in $\pi^{*,\exp} > 0$. If $\gamma > 1$, then an increase in $\theta$ leads to an increase in $\pi^{*,\mathrm{pow}} > 0$.

### 4.5.3 The effects of personal risk tolerance

We will analyze some specific cases to see the effects of the risk tolerance parameter $\gamma$ on the optimal investment strategy $\pi^*$ in MFG.

**Case 1**: Suppose there is no relative performance, i.e. $\theta = 0$. In this case, $\pi^*$ in both exponential and power cases coincide and become

$$\pi^* = \frac{\mu}{\gamma} \frac{1}{\sigma^2 + \nu^2}.$$

Taking the derivative with respect to $\gamma$ yields

$$\pi^*_\gamma = -\frac{\mu}{\gamma^2} \frac{1}{\sigma^2 + \nu^2},$$

which is negative if $\mu > 0$. This implies that an increase in risk tolerance $\gamma$ tends to decrease the optimal investment strategy $\pi^*$.

**Case 2**: $\gamma > 0, 0 < \theta \leq 1$. In this case $\pi^* > 0$ in exponential case and $\pi^* > 0$ in power case if $\gamma > 1$. Thus, an increase in $\gamma$ leads to a decrease in $\pi^*$. That means an investor with more risk aversion invests less in the risky asset.

**Case 3**: $0 < \gamma < 1, 0 < \theta \leq 1$. For the exponential case, the effect of $\gamma$ on $\pi^*$ is the same as in Case 2. However, for the power case, $\pi^*$ can be positive, negative, or zero since the first summand of $\pi^*$ is positive while the second summand is negative. Consequently, an increase in $\gamma$ leads to

$$\pi^* = \begin{cases} \text{increases }, \text{ if } \Delta\left(\frac{\mu}{\gamma}\frac{1}{\sigma^2+\nu^2}\right) > \Delta\left|\theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma}{\sigma^2+\nu^2}\frac{\varphi}{1-\psi}\right| \\ \text{decreases }, \quad \text{if } \Delta\left(\frac{\mu}{\gamma}\frac{1}{\sigma^2+\nu^2}\right) < \Delta\left|\theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma}{\sigma^2+\nu^2}\frac{\varphi}{1-\psi}\right| \\ 0, \quad \text{if } \Delta\left(\frac{\mu}{\gamma}\frac{1}{\sigma^2+\nu^2}\right) = \Delta\left|\theta\left(1-\frac{1}{\gamma}\right)\frac{\sigma}{\sigma^2+\nu^2}\frac{\varphi}{1-\psi}\right|. \end{cases}$$

## 4.6 Conclusion

In this paper, each individual agent (or investor) faces the problem of allocating her terminal wealth in a portfolio including one risk-free asset (e.g. government bond) and one individual risky asset (e.g. stock). We model the games with finite and infinite agents, namely $n$-agent games and MFG, respectively. Each agent has a utility function that belongs to the family of HARA utility functions. The utility function depends on her terminal wealth as well as the average wealth of the population. The games are modeled in the context of the presence of risk-seeking agents and relative performance. We prove that there exists a unique constant Nash equilibrium and constant MFE for the case of strictly concave utility function. For the case of strictly convex utility function, there exists a corner solution for both $n$-agent and MFG in which risk-seeking agents invest all her asset value on risky asset (e.g. stock) and invest nothing on riskless asset (e.g. bond). The qualitative effects of personal and market parameters on the optimal investment strategy are discussed deeply.

There are several limitations in this chapter which pay ways for future research. First, all personal and market parameters in this chapter are constant over time, one might assume these parameters vary over time. Some parameters depend on time imply that Nash equilibrium or MFE also depend on time rather than constant as in this chapter. Second, this study assumes that each individual investor only invests in one individual stock, which is not always the case in the real world. One might relax this assumption to allow each individual investor to invest in more than one stocks. Third, the setting in this study is the perfect market with no frictions (e.g. transaction costs, and borrowing and lending), one might take one or some frictions into account to make models more realistic. Fourth, this study only constructs and solves an investment-only problem, one might model and tackle an investment-consumption problem.

# Chapter 5
# Conclusions

The thesis comprises three essays on business analytics and game theory, with each chapter analyzing a particular type of shock.

Chapter 2 uses machine learning and text mining techniques to quantify natural disaster risk (climate shocks) effects on the US firm's performance. The first goal of this research is to create a new dictionary of terms associated with natural hazards and disasters. Text mining is used to determine how many of these words appear in Form 10-Ks. This information is then used to create an indicator of the perceived risk of natural disasters, which is combined with data on government-reported damages and other factors to investigate how natural disasters impact the performance of US firms between 1993 and 2021. The study reveals that self-reported perceived risk and government-reported damages of natural disasters in the current year are both linked to reduced profitability in the following year but have no impact on sales growth or Tobin's Q ratio. The services sector is more heavily impacted than the manufacturing sector. Additionally, there is a lag effect on profitability in the services sector related to the perceived risk of natural disasters but not for government-reported damages. The study suggests that CART and neural networks are better than linear regression for predicting firm performance during natural disasters. Ultimately, the study indicates that information from financial reports, such as Form 10-K filings, can be used to evaluate the perceived risk of natural disasters and predict their effects on firm performance, particularly profitability.

Chapter 3 explores firms as microeconomic agents regarding bankruptcy/financial distress probability prediction under periods with good and bad years (i.e., years with economic shocks). In particular, this study compares the effectiveness of three different models - accounting-based, market-based, and machine-learning - in predicting financial distress in Vietnam, a transition economy. The findings indicate that although all models perform reasonably well in predicting outcomes for companies that have not been delisted, they struggle to do so for delisted companies. The study also demonstrates that models combining variables from Altman's and Ohlson's approaches perform better than those using only one set of variables, as measured by balanced accuracy. Furthermore, the study shows that neural networks are consistently the most effective models based on balanced accuracy and Matthews correlation coefficient. Among Altman's variables, the `reat` variable (retained earnings over total assets) is the most important, while `ltat` (total liabilities over total assets) and `wcapat` (working capital over total assets) are the most significant in Ohlson's variables. The study also reveals that the models are generally more effective at predicting outcomes for large companies than for small ones, and their performance tends to be better in good years than in bad ones, based on Matthews correlation coefficient (MCC).

Chapter 4 examines the allocation of wealth in portfolios with potential Brownian motions (as stochastic shocks) consisting of a risk-free asset, such as a government bond, and a single risky asset, like a stock. Each investor is considered an individual agent and must determine how to allocate their terminal wealth in this portfolio. The paper models two types of games: $n$-agent games and mean field games, each with a different number of players. The agents in these games have hyperbolic absolute risk aversion (HARA) utility functions that depend on their terminal wealth and the average wealth of the population. This chapter demonstrates that when the utility function is strictly concave, both the $n$-agent and mean field games have only one Nash equilibrium and one mean field equilibrium. On the other hand, when the utility function is strictly convex, a distinct corner solution exists in which all agents invest their wealth in risky assets and none in riskless assets. The paper also examines how the optimal investment strategy is affected by personal and market coefficients.

# References


Abe, M. and Ye, L. (2013). Building resilient supply chains against natural disasters: The cases of Japan and Thailand. *Global Business Review*, 14(4):567–586.

Abel, A. B. (1990). Asset prices under habit formation and catching up with the Joneses. *The American Economic Review Papers and Proceedings*, 80(2):38–42.

Abinzano, I., Gonzalez-Urteaga, A., Muga, L., and Sanchez, S. (2020). Performance of default-risk measures: the sample matters. *Journal of Banking & Finance*, 120.

Adya, M. and Collopy, F. (1998). How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting*, 17(5-6):481–495.

Afik, Z., Arad, O., and Galil, K. (2016). Using Merton model for default prediction: An empirical assessment of selected alternatives. *Journal of Empirical Finance*, 35:43–67.

Alok, S., Kumar, N., and Wermers, R. (2020). Do fund managers misestimate climatic disaster risk. *The Review of Financial Studies*, 33(3):1146–1183.

Altay, E. and Satman, M. H. (2005). Stock market forecasting: Artificial neural network and linear regression comparison in an emerging market. *Journal of Financial Management & Analysis*, 18(2):18–33.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609.

Altman, E. I., Hotchkiss, I., and Wang, W. (2019). *Corporate Financial Distress, Restructuring, and Bankruptcy: Analyze Leveraged Finance, Distressed Debt, and Bankruptcy*. Wiley, Hoboken, NJ, 4 edition.

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., and Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model. *Journal of International Financial Management & Accounting*, 28(2):131–171.

Amasyali, K. and El-Gohary, N. (2021). Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings. *Renewable and Sustainable Energy Reviews*, 142.

Angel, J. J., Christophe, S. E., and Ferri, M. G. (2003). A close look at short selling on Nasdaq. *Financial Analysts Journal*, 59(6):66–74.

Anyaeche, C. O. and Ighravwe, D. E. (2013). Predicting performance measures using linear regression and neural network: A comparison. *African Journal of Engineering Research*, 1(3):84–89.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297.

Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1):685–725.

Aumann, R. J. (1964). Markets with a continuum of traders. *Econometrica*, 32(1):39–50.

Aumann, R. J. (1975). Values of markets with a continuum of traders. *Econometrica*, 43(4):611–646.

Aumann, R. J. and Shapley, L. S. (1974). *Values of non-atomic games*. Princeton University Press, Princeton, NJ.

Bai, Z., Li, H., McAleer, M., and Wong, W.-K. (2014). Stochastic dominance statistics for risk averters and risk seekers: An analysis of stock preferences for USA and China. *Quantitative Finance*, 15(5):889–900.

Barth, M. E., Li, K., and McClure, C. G. (2023). Evolution in value relevance of accounting information. *The Accounting Review*, 1(98):1–28.

Basak, S. and Makarov, D. (2015). Competition among portfolio managers and asset specialization. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1563567. Accessed on 01/07/2022.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4:71–111.

Bellstam, G., Bhagat, S., and Cookson, J. A. (2021). A text-based analysis of corporate innovation. *Management Science*, 67(7):3985–4642.

Benos, A. and Papanastasopoulos, G. (2007). Extending the Merton model: A hybrid approach to assessing credit quality. *Mathematical and Computer Modelling*, 46(1-2):47–68.

Bensic, M., Sarlija, N., and Zekic-Susac, M. (2005). Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. *Intelligent Systems in Accounting, Finance and Management*, 13(3):133–150.

Benth, F. E., Karlsen, K. H., and Reikvam, K. (2001). Optimal portfolio selection with consumption and nonlinear integro-differential equations with gradient constraint: A viscosity solution approach. *Finance and Stochastics*, 5:275–303.

Bharath, S. T. and Shumway, T. (2008). Forecasting default with the Merton distance to default model. *The Review of Financial Studies*, 21(3):1339–1369.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654.

Bochkay, K. and Levin, C. B. (2019). Using MD&A to improve earnings forecasts. *Journal of Accounting, Auditing Finance*, 34(3):458–482.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152.

Botzen, W. J. W., Deschenes, O., and Sanders, M. (2019). The economic impacts of natural disasters: A review of models and empirical studies. *Review of Environmental Economics and Policy*, 13(2):167–188.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3):199–231.

Breiman, L., Friedman, J. H., Olshen, R., and Stone, C. J. (1984). *Classification and regression trees*. CRC Press, Florida.

Brennan, M. J. (1975). The optimal number of securities in a risky asset portfolio when there are fixed costs of transacting: Theory and some empirical results. *The Journal of Financial and Quantitative Analysis*, 10(3):483–496.

Burton, I. and Kates, R. W. (1963). The perception of natural hazards in resource management. *Natural Resources Journal*, 3(3):412–441.

Cathcart, L., Dufour, A., Rossi, L., and Varotto, S. (2020). The differential impact of leverage on the default risk of small and large firms. *Journal of Corporate Finance*, 60:101541.

Çanakoğlu, E. and Özekici, S. (2010). Portfolio selection in stochastic markets with HARA utility functions. *European Journal of Operational Research*, 201(2):520–536.

Chan, P. and Sircar, R. (2015). Bertrand and Cournot mean field games. *Applied Mathematics and Optimization*, 71(3):533–569.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1):321–357.

Chen, H., Xiao, K., Sun, J., and Wu, S. (2017). A double-layer neural network framework for high-frequency forecasting. *ACM Transactions on Management Information Systems*, 7(4):1–17.

Chen, J., Marshall, B, R., Zhang, J., and Ganesh, S. (2006). Financial distress prediction in China. *Review of Pacific Basin Financial Markets and Policies*, 9(2):317–336.

Chen, W.-S. and Du, Y.-K. (2009). Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*, 36(2 (Part 2)):4075–4086.

Chevalier, J. and Ellison, G. (1999). Career concerns of mutual fund managers. *The Quarterly Journal of Economics*, 114(2):389–432.

Cho, H. and Muslu, V. (2021). How do firms change investments based on MD&A disclosures of peer firms? *The Accounting Review*, 96(2):177–204.

Choi, J., Suh, Y., and Jung, N. (2020). Predicting corporate credit rating based on qualitative information of MD&A transformed using document vectorization techniques. *Data Technologies and Applications*, 54(2):151–168.

Clark, E., Qiao, Z., and Wong, W.-K. (2016). Theories of risk: Testing investor behavior on the Taiwan stock and stock index futures markets. *Economic Enquiry*, 54(2):907–924.

Cohen, L., Malloy, C., and Nguyen, Q. (2020). Lazy prices. *The Journal of Finance*, 75(3):1371–1415.

Collins, R. A. (1980). An empirical comparison of bankruptcy prediction models. *Financial Management*, 9(2):52–57.

Cornett, M. M., Marcus, A. J., and Tehranian, H. (2008). Corporate governance and pay-for-performance: The impact of earnings management. *Journal of Financial Economics*, 87(2):357–373.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20:273–297.

Coval, J. D. and Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, 54(6):2045–2073.

Cusumano, M. A. (2013). Evaluating a startup venture. *Communications of the ACM*, 56(10):26–29.

Davis, M. H. A. and Norman, R. (1990). Portfolio selection with transaction costs. *Mathematics of Operations Research*, 15(4):573–793.

Dechow, P. M., Hutton, A. P., Meulbroek, L., and Sloan, R. G. (2001). Short-sellers, fundamental analysis, and stock returns. *Journal of Financial Economics*, 61(1):77–106.

Desai, V. S. and Bharati, R. (1998). A comparison of linear regression and neural network methods for predicting excess returns on large stocks. *Annals of Operations Research*, 78(127–163).

Dessaint, O. and Matray, A. (2017). Do managers overreact to salient risks? Evidence from hurricane strikes. *Journal of Financial Economics*, 126(1):97–121.

Doganis, P., Alexandridis, A., Patrinos, P., and Sarimveis, H. (2006). Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Journal of Food Engineering*, 75(2):196–204.

Dorfman, R. (1969). An economic interpretation of optimal control theory. *The American Economic Review*, 59(5):817–831.

dos Reis, G. and Platonov, V. (2021). Forward utilities and mean-field games under relative performance concerns. *From Particle Systems to Partial Differential Equations*, page 227–251.

Downey, H. K., Hellriegel, D., and Slocum, J. W. (1975). Environmental uncertainty: The construct and its application. *Administrative Science Quarterly*, 20(4):613–629.

Duffie, D., Fleming, W., Soner, H. M., and Zariphopoulou, T. (1997). Hedging in incomplete markets with HARA utility. *Journal of Economic Dynamics and Control*, 21(4-5):753–782.

Duncan, R. B. (1972). Characteristics of organizational environments and perceived environmental uncertainty. *Administrative Science Quarterly*, 17(3):313–327.

Durbin, J. (1970). Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica*, 38(3):410–421.

Eisenkopf, G. (2010). Peer effects, motivation, and learning. *Economics of Education Review*, 29(3):364–374.

Escobar, M., Neykova, D., and Zagst, R. (2017). HARA utility maximization in a markov-switching bond–stock market. *Quantitative Finance*, 17(11):1715–1733.

Espinosa, G.-E. and Touzi, N. (2015). Optimal investment under relative performance concerns. *Mathematical Finance*, 25(2):221–257.

Fadlalla, A. and Lin, C.-H. (2001). An analysis of the applications of neural networks in finance. *INFORMS Journal on Applied Analytics*, 31(4).

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Faucheux, S. and Froger, G. (1995). Decision-making under environmental uncertainty. *Ecological Economics*, 15(1):29–42.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Felbermayr, G. and Gröschl, J. (2014). Naturally negative: The growth effects of natural disasters. *Journal of Development Economics*, 111:92–106.

Feldman, R., Govindaraj, S., Livnat, J., and Segal, B. (2010). Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, 15:915–953.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188.

Freel, M. S. (2005). Perceived environmental uncertainty and innovation in small firms. *Small Business Economics*, 25:49–64.

Friend, I. and Blume, M. E. (1975). The demand for risky assets. *The American Economic Review*, 65(5):900– 922.

Fu, G., S. X. and Zhou, C. (2020). Mean field exponential utility game: A probabilistic approach. Available at https://arxiv.org/abs/2006.07684. Accessed on 07/20/2020.

Geng, R., Bose, I., and Chen, X. (2015). Prediction of financial distress: An empirical study of listed chinese companies using data mining. *European Journal of Operational Research*, 241(1):236–247.

Gepp, A., Kumar, K., and Bhattacharya, S. (2010). Business failure prediction using decision trees. *Journal of Forecasting*, 29(6):536–555.

Gocheva-Ilieva, S. G., Voynikova, D. S., Stoimenova, M. P., Ivanov, A. V., and Iliev, I. P. (2019). Regression trees modeling of time series for air pollution analysis and forecasting. *Neural Computing and Applications*, 31:9023–9039.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, 46(6):1303–1310.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. The MIT Press, Cambridge, MA.

Graves, P. E. (1979). Relative risk aversion: Increasing or decreasing? *The Journal of Financial and Quantitative Analysis*, 14(2):205–214.

Gruszczynski, M. (2004). Financial distress of companies in Poland. *International Advances in Economic Research*, 10:249–256.

Guo, X., Wong, K.-W., and Zhu, L. (2016). Almost stochastic dominance for risk averters and risk seeker. *Finance Research Letters*, 19:15–21.

Gupta, R., Yang, J., and Basu, P. K. (2014). Market efficiency in emerging economies – Case of Vietnam. *International Journal of Business and Globalisation*, 13(1):25–40.

Hill, T., Marquez, L., O'Connor, M., and Remus, W. (1994). Artificial neural network models for forecasting and decision making. *International Journal of Forecasting*, 10(1):5–15.

Hill, T., O'Connor, M., and Remus, W. (1996). Neural network models for time series forecasts. *Management Science*, 42(7):1082–1092.

Hillegeist, S. A., Keating, E. K., Cram, D. P., and Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9:5–34.

Hoang, T. H. V. (2020). Applying Z-score model in predicting bankruptcy of enterprises in Vietnam (in Vietnamese). *Banking Science & Training Review*, (217):43–51.

Hoang, T.-H.-V., Wong, W.-K., and Zhu, Z. (2015). Is gold different for risk-averse and risk-seeking investors? An empirical analysis of the Shanghai Gold Exchange. *Economic Modeling*, 50:200–211.

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, 117(1):287–299.

Hosono, K., Miyakawa, D., Uchino, T., M, H., Ono, A., Uchida, H., and Uesugi, I. (2016). Natural disasters, damage to banks, and firm investment. *International Economic Review*, 57(4):1335–1370.

Hsu, C.-W., Chang, C.-C., and Lin, C.-J. (2016). A practical guide to support vector classification. Technical report. Available at https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf. Accessed on 03/17/2023.

Hsu, P.-H., Lee, H.-H., Peng, S.-C., and Yi, L. (2018). Natural disasters, technology diversity, and operating performance. *The Review of Economics and Statistics*, 100(4):619–630.

Huang, H. H., Kerstein, J., and Wang, C. (2018). The impact of climate risk on firm performance and financing choices: An international comparison. *Journal of International Business Studies*, 49:633–656.

Huang, M. (2013). A mean field capital accumulation game with HARA utility. *Dynamic Games and Applications: special issue on mean field games*, 3:446–472.

Huang, M., Malhamé, R. P., and Caines, P. E. (2006). Large population stochastic dynamic games: Close-loop Mckean-Vlasov systems and the Nash Certainty Equivalence principle. *Communication in Information and Systems*, 6(3):221–252.

Huang, M. and Nguyen, S. L. (2016). Mean field games for stochastic growth with relative utility. *Applied Mathematics and Optimization*, 74(3):643–668.

Huang, Q., Li, Y., Lin, M., and McBrayer, G. A. (2022). Natural disasters, risk salience, and corporate ESG disclosure. *Journal of Corporate Finance*, 72.

Huang, W., Lai, K. K., Nakamori, Y., Wang, S., and Yu, L. (2007). Neural networks in finance and economics forecasting. *International Journal of Information Technology Decision Making*, 6(1):113–140.

Huynh, T. and Xia, Y. (2022). Panic selling when disaster strikes: Evidence in the bond and stock markets. *Management Science, Forthcoming.*

Inder, B. (1984). Finite-sample power of tests for autocorrelation in models containing lagged dependent variables. *Economics Letters*, 14(2-3):179–185.

Jovanovic, B. and Rosenthal, R. W. (1988). Anonymous sequential games. *Journal of Mathematical Economics*, 17(1):77–87.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47(2):263–292.

Kaldor, N. (1966). Marginal productivity and the macro-economic theories of distribution: Comment on samuelson and modigliani. *The Review of Economic Studies*, 33(4):309–319.

Kang, T., Park, D.-H., and I., H. (2018). Beyond the numbers: The effect of 10-K tone on firms' performance predictions using text analytics. *Telematics and Informatics*, 35(2):370–381.

Karpoff, J. M. and Lou, X. (2010). Short sellers and financial misconduct. *The Journal of Finance*, 65(5):1879–1913.

Kaustia, M. and Knüpfer, S. (2012). Peer performance and stock market entry. *Journal of Financial Economics*, 104(2):321–338.

Kealhofer, S. (2003). Quantifying credit risk I: Default prediction. *Financial Analysts Journal*, 59(1):30–44.

Keele, L. and Kelly, N. J. (2006). Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political Analysis*, 14(2):186–205.

Khan, M. A. and Sun, Y. (2002). *Handbook of Game Theory with Economic Applications*, chapter Non-cooperative games with many players, pages 1761–1808. North-Holland, Amsterdam.

Kim, T. S. and Omberg, E. (1996). Dynamic nonmyopic portfolio behavior. *The Review of Financial Studies*, 9(1):141–161.

Kim, Y. S. (2008). Comparison of the decision tree, artificial neural network, and linear regression methods based on the number and types of independent variables and sample size. *Expert Systems with Applications*, 34:1227–1234.

Kreiser, P. and Marino, L. (2002). Analyzing the historical development of the environmental uncertainty construct. *Management Decision*, 40(9):895–905.

Lacker, D. and Zariphopoulou, T. (2019). Mean field and $n$-agent games for optimal investment under relative performance criteria. *Mathematical Finance*, 29(4):1003–1038.

Ladas, A. C. (2021). Future firm performance and textual characteristics of 10-K disclosures. *International Journal of Banking, Accounting and Finance*, 12(4):333–346.

Lam, C. D. and Phan, D. A. (2009). Application of KMV-Merton model to the quantification of relations between the assets guarantee, the borrower's funds appointment and the credit risk (in Vietnamese). *Journal of Science and Technology*, 2(31):1–6.

Lasry, J.-M. and Lions, P.-L. (2006a). Jeux a champ moyen, I. Le cas stationnaire. *C. R. Acad. Sci. Paris*, 343:619–625.

Lasry, J.-M. and Lions, P.-L. (2006b). Jeux a champ moyen, II. Horizon fini et controle optimal. *C. R. Acad. Sci. Paris*, 343:679–684.

Lasry, J.-M. and Lions, P.-L. (2007a). Large investor trading impacts on volatility. *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*, 24(2):311–323.

Lasry, J.-M. and Lions, P.-L. (2007b). Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260.

Lawrence, P. R. and Lorsch, J. W. (1967). *Organization and Environment. Managing Differentiation and Integration*. Harvard University Press, Cambridge, MA.

Lazzaroni, S. and Bergeijk, P. A. G. V. (2014). Natural disasters' impact, factors of resilience and development: A meta-analysis of the macroeconomic literature. *Ecological Economics*, 107:333–346.

Le, D. C. and Le, T. A. (2012). Using CVaR and KMV-Merton model to estimate the default probability: Empirical evidence from Vietnam (in Vietnamese). *Journal of Development and Integration*, 5(15):10–15.

Lean, H. H., Lien, D.-H. D., and Wong, W.-K. (2010). Futures versus stocks: A stochastic dominance study in Malaysian markets. *Advances in Investment Analysis and Portfolio Management*, 4:49–80.

Lee, W.-C. (2011). Redefinition of the KMV model's optimal default point based on genetic algorithms – Evidence from Taiwan. *Expert Systems with Applications*, 38(8):10107–10113.

Leiter, A. M., Oberhofer, H., and Raschky, P. A. (2009). Creative disasters? Flooding effects on capital, labour and productivity within European firms. *Environmental and Resource Economics*, 43:333–350.

Lewis, G. J. and Harvey, B. (2001). Perceived environmental uncertainty: The extension of miller's scale to the natural environment. *Journal of Management Studies*, 38(2):201–234.

Li, F. a. (2011). Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature*, 29:143–165.

Li, H. and Sun, J. (2008). Ranking-order case-based reasoning for financial distress prediction. *Knowledge-Based Systems*, 21(8):868–878.

Li, K., Liu, X., Mai, F., and Zhang, T. (2021). The role of corporate culture in bad times: Evidence from the COVID-19 pandemic. *Journal of Financial and Quantitative Analysis*, 56(7):2545–2583.

Li, S., Jamieson, K., DeSalvo, G. Rostamizadeh, A., and Talwalkar, A. (2018). Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18:1–52.

Li, X., Wu, C., and Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information & Management*, 56(2):172–184.

Li, Y. and Ma, W. (2010). Applications of artificial neural networks in financial economics: A survey. *2010 International Symposium on Computational Intelligence and Design*, pages 211–214.

Lieu, M. L. (2014). Estimation of default probability by using $Z$-score and $H$-score models for listed companies in Vietnam (in Vietnamese). *Banking Technology Review*, (105):21–28.

Lindenberg, E. B. and Ross, A. (1981). Tobin's q ratio and industrial organization. *The Journal of Business*, 54(1):1–32.

Liu, H. (2014). Solvency constraint, underdiversification, and idiosyncratic risks. *The Journal of Financial and Quantitative Analysis*, 49(2):409–430.

Liu, J. (2000). *Portfolio selection in stochastic environments.* PhD dissertation, Stanford University.

Loc, T. D., Lanjouw, G., and Lensink, R. (2010). Stock-market efficiency in thin-trading markets: the case of the Vietnamese stock market. *Applied Economics*, 42:3519–3532.

López-Gamero, M. D., Molina-Azorín, J. F., and Claver-Cortés, E. (2011). Environmental uncertainty and environmental management perception: A multiple case study. *Journal of Business Research*, 64(4):427–435.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65.

Mai, F., Tian, S., Lee, C., and Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274(2):743–758.

Margaritis, D. and Psillaki, M. (2010). Capital structure, equity ownership and firm performance. *Journal of Banking & Finance*, 34(3):621–632.

Marquez, L., Hill, T., Worthley, R., and Remus, W. (1991). Neural network models as an alternative to regression. *Proceedings of the 24th Annual Hawaii International Conference on Systems Sciences*, IV:129–146.

McKnight, B. and Linnenluecke, M. K. (2017). Patterns of firm responses to different types of natural disasters. *Business & Society*, 58(4):813–840.

Merkel, G. D., Povinelli, R. J., and Brown, R. H. (2018). Short-term load forecasting of natural gas with deep neural network regression. *Energies*, 11(8).

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3):247–257.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2):449–470.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42(3):483–510.

Miles, R. E., Snow, C. C., Meyer, A. D., and Coleman, H. J. (1978). Organizational strategy, structure, and process. *The Academy of Management Review*, 3(3):546–562.

Miller, K. D. (1993). Industry and country effects on managers' perceptions of environmental uncertainties. *Journal of International Business Studies*, 24(4):693–714.

Milliken, F. J. (1987). Three types of perceived uncertainty about the environment: State, effect, and response uncertainty. *The Academy of Management Review*, 12(1):133–143.

Milnor, J. W. and Shapley, L. S. (1961). Values of large games, II: Oceanic games. *Mathematics of Operations Research*, 3(4):290–307.

Mironov, M. (2013). Taxes, theft, and firm performance. *The Journal of Finance*, 68(4):1441–1472.

Mithas, S., Tafti, A., Bardhan, I., and Goh, J. M. (2012). Information technology and firm profitability: Mechanisms and empirical evidence. *MIS Quarterly*, 36(1):205–224.

Myerson, R. B. (1991). *Game theory: analysis of conflict*. Harvard University Press, Cambridge, MA.

Nguyen, D. T. and Nguyen, C. M. (2017). Measuring credit risk of listed firms in Vietnam (in Vietnamese). *Review of Finance*, (654):52–54.

Nguyen, T. C. and Pham, C. K. (2014). Using KMV-Merton to estimate credit risk of corporate customers and potential loss for banks (in Vietnamese). *Journal of Economic Development*, (289):39–57.

Nguyen, V. H. (2015). Using Z-score model to test internal credit risk ranking and estimate probability of default (in Vietnamese). *Journal of Finance and Accounting Research*, 8(145):26–30.

Niendorf, B. and Ottaway, T. (2002). Wealth effects of time variation in investor risk preferences. *Journal of Economics and Finance*, 26(1):77–87.

Ninh, B. P. V., Thanh, T. D., and Hong, D. V. (2018). Financial distress and bankruptcy prediction: An appropriate model for listed firms in Vietnam. *Economic Systems*, 42(4):616–624.

Noth, F. and Rehbein, O. (2019). Badly hurt? Natural disasters and direct firm effects. *Finance Research Letters*, 28:254–258.

Noy, I. and Vu, B. T. (2010). The economics of natural disasters in a developing country: The case of Vietnam. *Journal of Asian Economics*, 21(4):345–354.

Odom, M. D. and Sharda, R. (1990). A neural network model for bankruptcy prediction. *The 1990 IJCNN International Joint Conference on Neural Networks*.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1):109–131.

Opler, T. C. and Titman, S. (1994). Financial distress and corporate performance. *The Journal of Finance*, 49(3):1015–1040.

Ou, J., Xia, J., Wu, Y.-J., and Rao, W. (2017). Short-term traffic flow forecasting for urban roads using data-driven feature selection strategy and bias-corrected random forests. *Transportation Research Record: Journal of the Transportation Research Board*, 2645(1):157–167.

Pankratz, N. M. C., Bauer, R., and Derwall, J. (2023). Climate change, firm performance, and investor surprises. *Management Science (forthcoming)*.

Payne, B. K., Brown-Iannuzzi, J. L., and Hannay, J. W. (2017). Economic inequality increases risk taking. *Proceedings of the National Academy of Sciences of the United States of America*, 114(18):4643–4648.

Pombeiro, H., Santos, R., Carreira, P., Silva, C., and Sousa, J. M. C. (2017). Comparative assessment of low-complexity models to predict electricity consumption in an institutional building: Linear regression vs. fuzzy modeling vs. neural networks. *Energy and Buildings*, 146(1):141–151.

Qiao, Z., Wong, W.-K., and Fung, J. K. W. (2013). Stochastic dominance relationships between stock and stock index futures markets: International evidence. *Economic Modelling*, 33:552–559.

Qiu, X. Y., Srinivasan, P., and Hu.Y (2014). Supervised learning models to predict firm performance with annual reports: An empirical study. *Journal of the Association for Information Science and Technology*, 65(2):400–413.

Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):159–203.

Rauh, M. T. (2003). Non-cooperative games with a continuum of players whose payoffs depend on statistics. *Economic Theory*, 21:901–906.

Razi, M. A. and Athappilly, K. (2005). A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*, 29(1):65–74.

Reisz, A. S. and Perlich, C. (2007). A market-based framework for bankruptcy prediction. *Journal of Financial Stability*, 3(2):85–131.

Schumacher, P., Olinsky, A., Quinn, J., and Smith, R. (2010). A comparison of logistic regression, neural networks, and classification trees predicting success of actuarial students. *Journal of Education for Business*, 85(5):258–263.

Seel, C. and Strack, P. (2013). Gambling in contests. *Journal of Economic Theory*, 148(5):2033–2048.

Shapiro, N. Z. and Shapley, L. S. (1961). Values of large games, I: A limit theorem. *Mathematics of Operations Research*, 3(1):1–9.

Shapley, L. S. (1961). Values of large games, III: A corporation with two large stockholders. *RAND Corporation Research Memorandum*.

Sharda, R. (1994). Neural networks for the MS/OR analyst: An application bibliography. *Interfaces*, 24(2):116–130.

Sharda, R. and Patil, R. B. (1992). Connectionist approach to time series prediction: An empirical test. *Journal of Intelligent Manufacturing*, 3(5):317–323.

Shayan, Z., Mezerji, N. M. G., Shayan, L., and Naseri, P. (2016). Prediction of depression in cancer patients with different classification criteria, linear discriminant analysis versus logistic regression. *Global Journal of Health Science*, 8(7):41–46.

Shin, K.-S., Lee, T. S., and Kim, H.-J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28(1):127–135.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1):101–124.

Siegel, J. J. (2014). *Stocks for the Long Run: The Definitive Guide to Financial Market Returns & Long-Term Investment Strategies*. McGraw-Hill, New York, NY.

Smirlock, M., Gilligan, T., and Marshall, W. (1984). Tobin's q and the structure-performance relationship. *The American Economic Review*, 74(5):1051–1060.

Staňková, M. (2022). Threshold moving approach with logit models for bankruptcy prediction. *Computational Economics*, pages 1–22.

Sun, J., Jia, M.-Y., and Li, H. (2011). AdaBoost ensemble for financial distress prediction: An empirical comparison with data from Chinese listed companies. *Expert Systems with Applications*, 38(8):9305–9312.

Swamidass, P. M. and Newell, W. T. (1987). Manufacturing strategy, environmental uncertainty and performance: A path analytic model. *Management Science*, 33(4):427–546.

Swanson, N. R. and White, H. (1995). A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics*, 13(3):265–275.

Swanson, N. R. and White, H. (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *The Review of Economics and Statistics*, 79(4):540–550.

Tanthanongsakkun, S., Pitt, D., and Treepongkaruna, S. (2011). A comparison of corporate bankruptcy models in Australia: The Merton vs. Accounting-based models. *Asia-Pacific Journal of Risk and Insurance*, 3(2):1–21.

Taylor, L. D. and Wilson, T. A. (1964). Three-pass least squares: A method for estimating models with a lagged dependent variable. *The Review of Economics and Statistics*, 46(4):329–346.

National Research Council (1999). *The Impacts of Natural Disasters: A Framework for Loss Estimation*. The National Academies Press, Washington, DC.

Thornhill, S. (2006). Knowledge, innovation and firm performance in high- and low-technology regimes. *Journal of Business Venturing*, 21(5):687–703.

Tkac, M. and Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, 38:788–804.

Tobin, J. (1969). A general equilibrium approach to monetary theory. *Journal of Money, Credit and Banking*, 1(1):15–29.

Tosi, H., Aldag, R., and Storey, R. (1973). On the measurement of the environment: An assessment of the Lawrence and Lorsch environmental uncertainty subscale. *Administrative Science Quarterly*, 18(1):27–36.

Tsoumakas, G. (2019). A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review*, 52:441–447.

Tudela, M. and Young, G. (2005). A Merton-model approach to assessing the default risk of UK public companies. *International Journal of Theoretical and Applied Finance*, 8(6):737–761.

Tung, R. L. (1979). Dimensions of organizational environments: An exploratory study of their impact on organization structure. *Academy of Management Journal*, 22(4):672–693.

Vassalou, M. and Xing, Y. (2004). Default risk in equity returns. *The Journal of Finance*, 59(2):831–868.

Vo, D. H. and Nguyen, T. M. (2014). The impact of corporate governance on firm performance: Empirical study in Vietnam. *International Journal of Economics and Finance*, 6(6).

Vo, D. H., Pham, B. N. V., Ho, C. M., and McAleer, M. (2019). Corporate financial distress of industry level listings in Vietnam. *Journal of Risk and Financial Management*, 12(4):155.

Vo, X. V. (2015). Using accounting ratios in predicting financial distress: An empirical investigation in the Vietnam stock market. *Journal of Economics and Development*, 17(1):41–49.

Vu, V. T. T., Do, N. H., Dang, H. N., and N., N. T. (2019). Profitability and the distance to default: Evidence from Vietnam securities market. *Journal of Asian Finance, Economics and Business*, 6(4):53–63.

Wang, L.-J., Guo, M., Sawada, K., Lin, J., and Zhang, J. (2016). A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. *Geosciences Journal*, 20:117–136.

Wang, Z. and Li, H. (2007). Financial distress prediction of Chinese listed companies: a rough set methodology. *Chinese Management Studies*, 1(2):93–110.

Weatherford, L. R., Gentry, T. W., and Wilamowski, B. (2003). Neural network forecasting for airlines: A comparative analysis. *Journal of Revenue and Pricing Management*, 1:319–331.

White, H. (1989). Learning in artificial neural networks: A statistical perspective. *Neural Computation*, 1(4):425–464.

Wilkins, A. S. (2018). To lag or not to lag?: Re-evaluating the use of lagged dependent variables in regression analysis. *Political Science Research and Methods*, 6(2):393–411.

Wilson, R. L. and Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems*, 11(5):545–557.

Wong, W.-K. and Li, C.-K. (1999). A note on convex stochastic dominance. *Economics Letters*, 62(3):293–300.

Wu, D. and Jaimungal, S. (2023). Robust risk-aware option hedging. Available at https://arxiv.org/abs/2303.15216. Accessed on 05/07/2023.

Xie, C., Luo, C., and Yu, X. (2011). Financial distress prediction based on SVM and MDA methods: the case of Chinese listed companies. *Quality & Quantity*, 45:671–686.

Ye, L. and Abe, M. (2012). The impacts of natural disasters on global supply chains. *ARTNeT Working Paper Series*, 115.

Yeh, C.-C., Lin, F., and Hsu, C.-Y. (2012). A hybrid KMV model, random forests and rough set theory approach for credit rating. *Knowledge-Based Systems*, 33:166–172.

Yildiz, B., Bilbao, J. I., and Sproul, A. B. (2017). A review and analysis of regression and machine learning models on commercial building electricity load forecasting. *Renewable and Sustainable Energy Reviews*, 73:1104–1122.

Yu, K., Cadeaux, J., Luo, N., Qian, C., and Chen, Z. (2018). The role of the consistency between objective and perceived environmental uncertainty in supply chain risk management. *Industrial Management & Data Systems*, 118(7):1365–1387.

Yu, V., Ting, H.-I., and Wu, Y.-C. J. (2009). Assessing the greenness effort for European firms: A resource efficiency perspective. *Management Decision*, 47(7):1065–1079.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1):35–62.

Zheng, Q. and Yanhui, J. (2007). Financial distress prediction based on decision tree models. *2007 IEEE International Conference on Service Operations and Logistics, and Informatics*.

Zhou, F. and Botzen, W. (2021). Firm level evidence of disaster impacts on growth in Vietnam. *Environmental and Resource Economics*, 79:277–322.

Zhu, M., Philpotts, D., Sparks, R., and Stevenson, M. J. (2011). A hybrid approach to combining cart and logistic regression for stock ranking. *The Journal of Portfolio Management*, 38(1):100–109.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22:59–82.